NEW DEVELOPMENTS IN QUANTITATIVE TRADING AND INVESTMENT

CHRISTIAN L. DUNIS PETER W. MIDDLETON ANDREAS KARATHANASOPOULOS KONSTANTINOS THEOFILATOS

ARTIFICIAL INTELLIGENCE IN FINANCIAL MARKETS

Cutting-Edge Applications for Risk Management, Portfolio Optimization and Economics

Contents

Pa	rt I Introduction to Artificial Intelligence	1
1	A Review of Artificially Intelligent Applications in the Financial Domain Swapnaja Gadre Patwardhan, Vivek V. Katdare, and Manish R. Joshi	3
Pa	rt II Financial Forecasting and Trading	45
2	Trading the FTSE100 Index: 'Adaptive' Modelling and Optimization Techniques Peter W. Middleton, Konstantinos Theofilatos, and Andreas Karathanasopoulos	47
3	Modelling, Forecasting and Trading the Crack: A Sliding Window Approach to Training Neural Networks Christian L. Dunis, Peter W. Middleton, Konstantinos Theofilatos, and Andreas Karathanasopoulos	69
4	GEPTrader: A New Standalone Tool for Constructing Trading Strategies with Gene Expression Programming Andreas Karathanasopoulos, Peter W. Middleton, Konstantinos Theofilatos, and Efstratios Georgopoulos	107

 $See \ discussions, stats, and author \ profiles \ for \ this \ publication \ at: \ https://www.researchgate.net/publication/310624227$

A Review of Artificially Intelligent Applications in the Financial Domain

READS 169

Chapter · January 2016

DOI: 10.1057/978-1-137-48880-0_1

CITATIONS 3	
3 author	s, including:
	Swapnaja Patwardhan Savitribai Phule Pune University 3 PUBLICATIONS 3 CITATIONS SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Portfolio Management View project

All content following this page was uploaded by Swapnaja Patwardhan on 12 July 2022.

A Review of Applications of Artificial Intelligence in Financial Domain

SwapnajaGadrePatwardhan

Research Scholar, North Maharashtra University, Jalgaon, India Swapnaja.gadre24@gmail.com

Vivek V. Katdare,

Director, IMR College, Jalgaon, India. vvkatdare@rediffmail.com

Manish R. Joshi School of Computer Sciences, North Maharashtra University Jalgaon, India joshmanish@gmail.com

Abstract: One of the major problems finance domain faces is the uncertain and nonlinear changes in data with respect to time. There is utmost necessity to tackle such time dependent uncertainties. The traditional models are incapable of resolving these problems to a certain extent. Contemporary models using artificial intelligence techniques are found to be a better solution to the problem. Artificial intelligence techniques such as Artificial neural network (ANN), Expert systems (ES) and Hybrid Intelligence System (HIS) are being applied to many verticals of finance domain including portfolio management, fraud detection, bankruptcy, stock management, risk management etc.

In this chapter, we review research work depicting use of artificial intelligence in finance management by comparing various artificial intelligence techniques. It is observed that artificial intelligent methods are more accurate as compared to traditional statistical methods. This review would be helpful to the researchers planning to explore the interdisciplinary field of computational finance.

Keywords: Portfolio Management, Stock Management, Artificial Neural Network, Expert System, Hybrid Intelligent System, Computational Finance.

I. Introduction:

Undoubtedly, the utmost challenge faced by many researchers and managers in the field of finance is uncertainty. Consequently, such uncertainty introduces an unavoidable risk factor that is the integral part of finance theory. The manifestation of risk not only complicates financial decision making but also creates opportunities for the investors who can manage and analyze risk efficiently and effectively. In order to handle complex nature of problem an interdisciplinary approach is advocated.

Computational finance is a division of applied computer science that deals with practical problems in finance. It can also be defined as the study of data and algorithms used in finance. This is an interdisciplinary field that combines numerical methods and mathematical finance. Computational finance uses mathematical proofs that can be applied for economic analyses. With the help of this finance models and systems are developed. These models are employed for portfolio management, stock prediction and risk management that play an important role in finance management.

During last few years, researchers are aiming to facilitate financial sector by financial trend prediction, identifying investor's behavior, portfolio management, fraud detection, risk management, bankruptcy, stock prediction, financial goal evaluation, finding regularities in security price movement etc. To achieve this, different methods like parametric statistical methods, non-parametric statistical methods and soft computing methods have been used as shown in figure(1). It is observed that many researchers are exploring and comparing soft computing techniques with parametric statistical techniques. Soft computing techniques such as Artificial Neural

Network, Fuzzy Logic, Support Vector Machine, Genetic Algorithm etc. are widely applied and accepted techniques in the field of finance and hence are in the scope of this review.



Figure (1): Techniques for analysis of financial applications

A] Parametric Statistical Methods: Parametric statistics is a division of statistics. It assumes that data is collected from various distributed systems and implemented to draw inferences about the parameters of the distribution. There are two types of parametric statistical methods namely discriminant analysis and logistic regression described as follows-

I] Discriminant Analysis: Discriminant Analysis is a statistical analysis done with the help of discriminant function to assign data to one of two or more naturally occurring groups. Discriminant analysis is used of determining the set of variables for prediction of category membership. Discriminant function analysis is a type of classification that distribute of things into classes or groups or categories of the same type.

II] Logistic Regression: Logistic regression is a method of prediction, which models the relationship between dependent and independent variables. It allows us to find the best fit model and the significance of relationships between dependent and independent variables. Logistic regression is used to estimate the probability of occurrence of an event.

B] Non-Parametric Statistical Methods: These are the methods in which data is not required to fit a normal distribution. Non-parametric method provides series of an alternative statistical methods that requires no or limited assumptions to be made about the data. Following are the techniques of non-parametric statistical methods-

I] Decision Tree: Decision tree is a classifier that is a tree like tool that supports decision making process. It is a tool which is employed for multiple variable analysis. A decision tree consists of nodes that form a rooted tree. All the nodes have only one incoming edge. Remaining nodes are referred as leaves. A node with an outgoing edge is termed as test node or an internal node. In decision tree, test node splits the instance space in two or more sub-spaces according to the discrete function.

II] Nearest Neighbour: Nearest neighbour is a non-parametric method applied for regression and classification. Nearest neighbour can also be referred as similar search, proximity search or closest point search which is used to find the nearest or closest points in the feature space. K-Nearest Neighbour algorithm is a technique used for classification and regression.

C] Soft Computing: Soft computing is a set of methods that aims to handle uncertainty, partial truth, imprecision and approximation. Basically these are based on human minds. Soft computing employs techniques like Artificial Neural Network, Fuzzy logic, Support Vector Machine, Genetic Algorithm etc. [1].

I] Artificial Neural Network: Neuron is a fundamental element of ANN. These neurons are connected to form a graph like structure also referred as networks. These neurons are like biological neurons. A neuron has small branches i.e. dendrites that are used for receiving inputs and axons carry the output and connects to the other neuron. Every neuron carries a signal received from dendrites as shown in Fig.(2) [115]. When the strength of signal exceeds a particular threshold value, an impulse is generated as an output which is known as action signal.



Figure (2): Structure of Artificial Neuron

Like biological neurons, artificial neurons accept input and generate output but are not able to model automatically. In ANN information or data is distributed and stored throughout the network in the form of weighted interconnections. Simulation of biological neuron is done with the help of non-linear function. Interconnections of artificial neuron are referred as weights. Diagram below shows the structure of artificial neuron in which *xi* is the input to the neuron and *wi* is the weight of the neuron. The average input is calculated by the formula [115].

$$a = \sum_{i=0}^{n} xiwi$$
 ----- (1)

ANN has minimum three layers of artificial neuron viz. input, hidden and output as shown in Fig.(3) [114] below. The input layer accepts the input and passes it to the hidden layer. Hidden layer is the most important layer from computation point of view. All the complex functions reside in this layer.



Figure (3): Three Layer Architecture of ANN

II] Fuzzy Logic: Fuzzy Logic is a type of many valued logic which deals with approximate values instead of exact or fixed reasoning. Fuzzy logic is a method of computing based on the degree of truth rather than its true or false value. Its truth value ranges in between 0 and 1.

III] Support Vector Machine: Support Vector Machine is the supervised learning models with related learning algorithms that are used for data analysis and pattern recognition applied in classification and regression. SVM uses the concept of hyper-plane which defines boundaries of decision. Decision plane separates the objects based on the class membership and are able to handle categorical and continuous variables.

IV] Genetic Algorithm: A Genetic Algorithm is an AI technique that mimics natural selection process. This technique is mostly used for optimization and search problems using selection, crossover, mutation and inheritance operations.

This paper accentuates application of Soft Computing Techniques namely Artificial Neural Network (ANN), Expert System (ES) and Hybrid Intelligence System (HIS) in finance management.

In recent years, it has been observed that array of computer technologies can be applied in the field of finance. Artificial Neural Network (ANN) is one of the contemporary methods of artificial intelligence which is used in financial markets. From the array of AI techniques, financial uncertainties are handled in a better manner by Artificial Neural Network. These uncertainties are handled by pattern recognition and future trend analysis. The most difficult exercises in finance are changes in the interest rates and currency movements. Large noisy data can be handled well by ANN. ANNs are characterized as numeric in nature. In statistical techniques like discriminant analysis or regression analysis, data distribution assumptions are required for input data. But ANN does not require any data distribution assumption and hence could be applied to wider range of problems than statistical techniques. Statistical techniques and symbolic manipulation techniques are batch oriented in which old and new data are to be submitted in a single batch to the model and later new mining results are generated. In contrast with these, in ANN it is possible to add new data to a trained ANN so as to update the training result. Since financial market is dynamic in nature, ANN can accommodate new data without reprocessing old data and hence used in finance management [3].

An expert system (ES) is knowledge based system used to solve critical problems in a particular domain. These are the rule-based systems with predefined set knowledge used for decision-making. Generic ES contains two modulesinference engine and knowledge base. The inference engine combines and processes the facts associated to the specific problem using the chunk of knowledge base relevant to it. The knowledge base is coded in the form of rules, semantic nets, predicates and objects in the system. ES are characterized as efficient, permanent, consistent, timely, complete decision making systems and hence used for finance management. Expert Systems are characterized as intelligent, capable of reasoning, drawing conclusions from relationships, dealing with uncertainties etc. Expert systems are capable of reproducing efficient, consistent and timely information so as to facilitate decision making [2]. Further Rich & Knight (1991) specified long ago that financial analysis is an expert's task.

Hybrid intelligent systems (HIS) are the software systems which combine methods and techniques of artificial intelligence like fuzzy expert systems, neuro-fuzzy systems, genetic-fuzzy systems etc. The integration of various learning techniques is combined to overcome the limitation of individual. Because of its ability to combine techniques together, it can be effectively used for finance management.

With reference to financial market, we identified Portfolio Management, Stock Market Prediction and Risk management as three most important AI application domains. As investment is an important aspect of finance management hence these three verticals are considered. In this study, we highlighted contribution of researchers in financial domain from last 20 years. Basic aim of this paper is to study and compare the applications of ANN, ES and hybrid intelligent system with traditional methods. The organization of the paper is – Second, third and fourth section deals with application of ANN, ES and HIS respectively in financial domain. In the fifth section conclusions are put forth.

II. Applications of ANN in Finance:

ANNs are the computational tools and are effectively used in various disciplines for modeling real world complex problem [4]. ANNs resembles biological neuron acting as source inspiration for a variety of techniques covering vast field of application [5]. In general, ANNs are referred as information processing system which uses earning and generalization capabilities which are adaptive in nature. Because of their adaptive nature, ANNs can provide solutions to the problems like forecasting, decision making, information processing etc. In recent years ANNs are proved as a powerful tool for handling dynamic financial market in terms of prediction [6], panning, forecasting [7] and decision-making [8].

With reference to this various studies have been done in order to classify and review the application of ANN in finance domain [9, 10]. Mixed results are obtained on the ability of ANN in finance domain. It has been observed that financial classification like financial evaluation, portfolio management, credit evaluation and prediction are significantly improved with the application of ANN in finance domain. We further classify application of ANN in finance domain as ANN for Portfolio Management, ANN for Stock Market Prediction and ANN for Risk Management. The details of these applications are presented in subsequent subsections –

A. Portfolio Management:

Determination of optimal allocation of assets in the broad categories like mutual funds, bonds, stocks etc. which suits investment of financial institutions across the specific time horizon with tolerance of risk is a crucial task. Nowadays investors prefer diversified portfolios that contain variety of securities.

Motiwalla et al.[11] applied ANN and regression analysis to study the predictable variations in US stock returns and concluded that ANN models are better than regression. Yamamoto et al [12]. designed a multilayer Back Propagation Neural Network (BPNN) for prediction of prepayment rate of mortgage with the help of correlation learning algorithm. Lowe et al. [13] developed analog NN for construction of portfolio under specified constraints. They also developed a feed forward NN for prediction of short term equities as a problem in nonlinear multichannel time series forecasting. Addedji et al. [14] applied ANN for economic analysis of risky projects. For the prediction of the potential returns on investment, NN model is used. On the basis of results obtained from network, financial managers could select the financial project by comparing it to conventional models. Survey conducted in this paper for portfolio management conclude that ANN performs better compared to other traditional methods.

Research papers surveyed in this article for portfolio management demonstrates that as compared to other traditional methods, ANN performs better especially BPNN. Zimmermann et al. [15] demonstrated the application of Back/Litterman portfolio optimization algorithm with the help of an error correction NN. Optimization of portfolio is done with 1) allocation that comply investors constraints and 2) controlled risk of portfolio. The method is tested by internationally diversified portfolios athwart 21 financial markets from G7 countries. They stated that their approach surpass conventional portfolios like Markowitz's mean-variance framework. Ellis et al. [16] performed portfolio analysis by comparing BPNN with randomly selected portfolio method and general property method and concluded that ANN performs better.

B. Stock Market Prediction:

In recent years with the help of online trading, stock market is one of the avenues where individual investors can earn sizeable profits. Hence there is a need to predict stock market behavior accurately. With this prediction investors can take decision about where and when to invest the money. Because of the volatility of financial market, it is challenging task to build a forecasting model.

ANN is a widely used soft computing method for stock market prediction and forecasting applied on IBM daily stock returns and concluded that NN performs better [17]. Kimoto et al. [18] reported the effectiveness of learning algorithms and prediction methods with ANN for Tokyo Stock Exchange price index prediction system. Kazuhiro et al. [19] investigated application of prior knowledge & neural networks for the improvement of prediction ability. Prediction of daily stock prices was considered as a real world problem. They considered some non-numerical features such as political and international events. They studied variety of prior knowledge that was difficult to use into network structure. Prior knowledge of stock price and information presented in newspaper in the form of foreign or domestic events are used by them. It was observed that event knowledge along with neural network was more effective for prediction with the significance level of 5%. Pai et al. [20] stated that ARIMA (Autoregressive integrated moving average) along with SVMs (Support vector machines) can be combined to deal with non-linear data. Unique strengths of ARIMA and SVMs are used for stock price forecasting with more strength. Thawornwong et al. [21] demonstrated that neural network model with feed-forward and probabilistic network (FNN) with a specific structure for realizing fuzzy inference system. Every membership function consists of one or two sigmoid function for inference rule. They concluded that FNN performs better. Duke et al. [23] used BPN for future prediction of German government's bonds.

C. Risk Management:

Financial risk management (FRM) is the process of managing economic value in a firm with the help of financial instruments to manage risk exposure especially market risk and credit risk. Financial Risk Management (FRM) is the process of identification of risk associated to the investments and possibly mitigating them. FRM can be qualitative or quantitative. FRM focuses on how and when hedging is to be done with the help of financial instruments to manage exposure to risk.

Treacy et al. [24] stated that traditional approach of banks for credit risk assessment is to generate an internal rating that considers subjective as well as qualitative factors like earning, leverage, reputation etc. Zhang et al. [25] compared LR and NN and five-fold cross validation procedure on the database of manufacturing firms. They employed Altman's five functional ratios along with the ratio current assets/current liabilities as an input to NN. They concluded that NN outperforms with accuracy 88.2%. Tam et al. [26] introduced a neural network approach to implement discriminant analysis in business research. Using bank data, linear classifier is compared with neural approach. Empirical result

concluded that neural model is more promising for the evaluation of bank condition in terms of adaptability, robustness and predictive accuracy. Huang et al. [27] introduced SVM (Support Vector Machine) to build a model with better explanatory ability. They used BPNN as a benchmark and obtained around 80% prediction accuracy for both SVM and BPNN for Taiwan and United States markets.

Author	Objective	Data Set	Preprocessing	Approach Used	Compared With	Evaluation Metric
Stoppiglia H, Idan Y, Dreyfus G	To develop a neural network - aided model for portfolio management	The data base comprises 398 companies, with 172 A companies, 172 B companies, & 54 C companies	Fifteen financial ratios like working capital/fixed assets, profit after taxes and interest/net worth, etc. per year	ANN	Statistical Method	Classificati on
Hans Georg Zimmermann, Ralph Neuneier and Ralph Grothmann Siemens AG	Portfolio Optimizatio n	Financial markets of the G7 countries	Monthly data I extracted from all databases	ANN	Mean- variance theory	Forecasting
Ellis C, Willson P	To select portfolio	Australian Property sector stocks	Not mentioned	BPNN	Random selection portfolio	Performanc e measure
Fernandez A, Gomez S	Portfolio selection and Portfolio Managemen t	Hang Seng in Hong Kong, DAX 100 in Germany, FTSE 100 in UK, S&P 100 in USA and Nikkei 225 in Japan	Weekly prices from data sets are extracted	ANN, GA and SA	Heuristic methods	Portfolio selection and optimizatio n
Freitas FD, De Souza AF, De Almeida AR	Portfolio selection and Portfolio optimization	IBOVESPA	Selected a subset of 52 stocks with long enough time series for training the neural networks	BPNN	Mean- variance model	Prediction
Po-Chang Ko, Ping-Chen Lin	Portfolio selection and Portfolio optimization	Taiwan stock exchange	Not mentioned	ANN	Traditional ANN model	Portfolio optimizatio n
Chiang W-C, Urban TL, Baldridge GW	Asset Forecasting	US mutual fund	15 economic variables are identified	BPNN	Regression model	NAV prediction
Chen A-S, Leung MT, Daouk H	Stock index forecasting	Taiwan Stock Exchange	Data is extracted on the basis of length of	PNN	Random walk model and the parametric	Return on investment

 Table (1): Brief review of application of ANN for Portfolio Management, Stock Market Prediction and Risk

 Management

			•		0.04	
			investment		GMM	
			horizon		models	
O'Connor N, Madden MG	To predict stock market movement	New York Stock Exchange and NASDAQ	Daily opening and closing values of DJIA index	BPNN	Simple benchmark functions	Accuracy
De Faria, E. L., Marcelo P. Albuquerque, J. L. Gonzalez, J. T. P. Cavalcante, and Marcio P. Albuquerque	To predict stock market movement forecasting	Brazilian stock market	Not mentioned	BPNN	Adaptive exponential smoothing method	Accuracy
Liao A, Wang J	Stock index forecasting	SP500, SAI, SBI, DJI, HIS and IXIC	Data normalization and adjusted to remove the noise	BP stochastic time effective NN	Brownain motion	Forecasting
Hyun-jung Kim , Kyung-shik Shin	To detect patterns in stock market	Korea Stock Price Index 200	Daily stock data is extracted	ATNN	TDNN	Accuracy
R.J. Kuo, C.H. Chen, Y.C. Hwang	Stock market forecasting	Taiwan stock market	Not mentioned	ANN, GFNN	Qualitative and quantitative factors of NN	Performanc e evaluation
Md. Rafiul Hassan ,Baikunth Nath, Michael Kirley	Stock market forecasting	www.finance.yahoo .com	Daily data is extracted	ANN, GA, HMM	ARIMA model	Forecasting
Chye KH, Tan WC, Goh CP	Credit risk assessment	Australian and German credit data sets	For each applicant 24 variables are selected	SVM classifier	Neural networks, genetic programmin g, and decision tree classifiers	Accuracy
Eliana Angelini, Giacomo di Tollo, Andrea Roli	Credit risk evaluation	Bank in Italy	Sample group is categorized into two groups i.e. boins and default	ANN	classical feed forward neural network and special purpose feed forward architecture	Classificati on

Fanning, Cogger and Shrivastava	To develop a model using neural network to find managerial fraud	Management database	Not mentioned	Artificial Neural Network	Generalized adaptive neural network architectures (GANNA) and the Adaptive Logic Network (ALN)	Accuracy
E. H. Feroz, M. K. Taek, V. S. Pastena, and K. Park,	To test the ability of selected red flags for prediction of the targets of the investigatio ns.	AAER	Not mentioned	Artificial Neural Network	Investigated versus Non- investigated forms	Prediction
Kurt Fanning , Kenneth O. Cogger	To develop a model for detection of management fraud	FFS.	Not mentioned	Artificial Neural Network	Statistical methods	Efficiency
R. Brause , T. Langsdorf , M. Hepp	To detect credit card fraud	GZS	38 field per transaction are extracted	Artificial Neural Network	Traditional systems	Prediction
Koskivaara	To investigate the impact of various pre- processing models on the forecast capability of neural network for financial auditing.	Manufacturing firm	Monthly balances	Artificial Neural Network	Traditional systems	Prediction
Fen-May Liou	To detect fraudulent financial reporting	Taiwan Economic Journal data bank and Taiwan Stock Exchange	Not mentioned	Artificial Neural Network	Logistic regression, neural networks, and classification trees.	Prediction
Chu, Jung	To compare statistical methods and ANN in bankruptcy prediction.	International Stock Exchange Official Year Book from a Data stream database	data distribution, group dispersion and orientation scheme	ANN with MDA	Logit, generalized delta rule	Prediction

Zhang et al.	Bankruptcy	COMPUSTAT	(1) have	ANN	MDA, ID3	Classificati
	prediction	database	operated in a			on
			regulated			
			industry; (2)			
			are foreign			
			based and			
			traded publicly			
			in the US; and			
			(3) have failed			
			bankruptcy			
			previously are			
			excluded from			
			the sample.			
Tam KY, Kiang	To perform	Bank default data	Not mentioned	Neural	Linear	predictive
MY	discriminant			Network	classifier,	accuracy,
	analysis in				logistic	adaptability
	business				regression, k	
	research				NN, and ID3	

III. Application of Expert Systems in Finance:

An Expert System is a computer system which is composed of well-organized body of knowledge that emulates expert problem-solving abilities in a limited domain of expertise. Matsatsinis et al. [28] presented a methodology of acquisition of knowledge and representation of knowledge for the development of expert system for financial analysis. Development of FINEVA, a multi-criterion knowledge base DSS for assessment of viability and corporate performance and the application of FINEVA was discussed in the paper. For a particular domain, set of inference rues are provided by human expert. Knowledge base is a collection of relevant facts, data, outcome and judgments [29]. Components of expert system includes knowledge base, User interface and inference engine. Knowledge is represented through the techniques like predicate logic, frames, semantic net etc. but most popular and widely used technique is IF-THEN rules also referred as production rules.

Liao et al. [30] carried out a review of use of Expert System in various areas including finance over the period 1995 to 2004. They observed that expert systems are flexible and provide a powerful method of solving variety of problems and can be used as and when required. Application of Expert System in finance domain is as follows –

A. Portfolio Management:

To explore the portfolio that meets requirements and objectives of fund manager are quite difficult and time consuming task. Ellis et al. [31] examined application of rule-based expert system in the property market and portfolios that are randomly constructed from the market. They observed that portfolios outperform on risk adjusted return basis.

Bohanec et al. [32] developed a knowledge-based tool for portfolio analysis for evaluation of project. This ES was developed for the republic of Solvenia. The model is demonstrated with tree structure supplemented by IF-THEN rules. Sanja Vraneš et al. [33] developed BEST (Blackboard-based Expert Systems Toolkit) for combining knowledge from different sources, use different methodologies for knowledge acquisition. As far as investment decision-making is concern, assortment from proficient economist critical investment ranking might be combined with decision knowledge evolved from operational research methods. When decisions are combined from many sources, there is a probability of redundancy reduction and proved more promising. Varnes et al. [34] suggested ES INVEX for investment management. This system assists investors and project analyst to select a project for investment. Mogharreban et al. [35] developed an ES PROSEL i.e. PORtfolioSELection system that uses set of rules for stock selection. PROSEL consists of three parts a] information center b] fuzzy stock selector and c] portfolio constructor. User friendly interface is available in PROSEL to change rules at run time. The result stated that PROSEL performs well.

B. Stock Market Prediction:

One more promising area of Expert System is Stock Market Prediction. Many investment consultants use these types of systems to improve financial and trading activities. Midland Bank of Landon use ES for interest rate swap portfolios and currency management [31].

Grosan et al. [36] applied MEP (Multi-Expression Programming), a genetic programming technique for prediction of NASDAQ index of NASDAQ stock market and NIFTY stock index. The performance is compared with the help of SVM, Levenberg-Marquardt algorithm, Takagi–Sugenoneuro-fuzzy inference system and concluded that MEP performs outstandingly. Quek [37] applied neuro-fuzzy network and ANFIS investor's measures forecasting for US Stock Exchange and proved best for stock price prediction of US Stock Exchange. Trinkle [38] used ANFIS and NN for forecasting annual excess return of three companies. The predictive ability of ANFIS and NN is compared with ARMA (Autoregressive Moving Average). The result stated that ANFIS and NN are able to forecast significantly better. Afolabi et al. [39] used neuro-fuzzy network, fuzzy logic and Kohonen'sself-organizing plan for stock price forecasting. They concluded that as compared to other techniques, deviation of Kohonen'sself-organizing plan is less. Yunos et al. [40] built a hybrid neuro-fuzzy model with the help of ANFIS for prediction of daily movements of KLCI (Kuala Lumpur Composite Index). For data analysis four technical indicators were chosen and concluded that ANFIS performs better. Atsalakis et al. [41] developed a neuro-fuzzy adaptive control system for forecasting the price trends of stock for the next day of the NYSE and ASE index. The experimental analysis stated that proposed system performs well.

C. Risk Management:

There is a vast scope in financial risk prediction and management using Expert systems. Matsatsinis et al. [42] presented a methodology for acquisition and representation of knowledge for the development of an Expert System. FINEVA (FINancianEVAluation) is a multi-criteria knowledge base expert system for the assessment of viability and performance using M4 ES shell. The interface use forward and backward chaining method. They concluded that ranking of analyzed firms depends upon the class of risk.

Shue et al. [43] build an ES for financial rating of corporate companies. This ES is developed by integrating two knowledge base a] Portege – domain knowledge base and b] JESS- operational knowledge base. The model is tested and verified by applying on financial statements of various companies of Taiwan stock market. Luke et al. [44] presented an ES, CEEES (Credit Evaluation and Explanation Expert System) to take decision about whether to allow credit lines to the firms. CEEES used rule based language for decision making process. They concluded that CEEES will recommend whether to consider or reject the application of firm.

Author	Objective	Data Set	Preprocessing	Approach Used	Compared With	Evaluation Metric
Shaw M, Gentry J	To manage business loan portfolios	Commercial Bank Data	Not mentioned	MARBLE	Conventiona 1 Method	Decision Rules, Classification
Bohanec M, Rajkovic V, Semoil B, Pogacnik	Portfolio selection and Portfolio Manageme nt	Ministry of Science and Technology of the Republic of Slovenia	Portfolio attribute selection and evaluation	expert system shell: DEX	Expert reviews	Rule generation
Ellis C, Willson P	To develop an expert system for portfolio managemen	Australian Stock Exchange (ASX), London Stock	Not mentioned	Expert System	Conventiona 1 Method	Rule generation

Table (2): Brief review of application of ES for Portfolio Management, Stock Market Prediction and R	₹isk
Management	

	t in property investment	Exchange (LSE)				
Kim SH, Chun SH	Stock index forecasting		Not mentioned	BPNN	CBR, APN, PNN	Prediction
YF Wang	Stock price prediction	Taiwan stock market	Every 5 min. data is recorded	Fuzzy grey prediction	GA	Prediction
Quintana et al.	Bankruptcy prediction	552 US companies from COMPUSTA T database	Not mentioned	ANN with ENPC	Naive Bayes (NB), logistic regression (LR),C4.5, PART, support vector machine (SVM), multilayer perceptron (MLP) and ENPC	Classification
Lenard and Alam	To develop a fuzzy logic model for developmen t of clusters to represent red flags in detection of fraud	Not mentioned	Not mentioned	Fuzzy Logic & Expert Reasoning	Statistical methods	Clustering

IV. Applications of Hybrid Intelligence in Finance:

Hybrid Intelligent System is a software system which is formed by combining methods and techniques of AI i.e. Fuzzy Expert system, Neuro-Fuzzy system, Genetic-Fuzzy system etc. Hybrid intelligent system is an effective learning system that combines the features by overcoming the weaknesses of the processing capabilities and representation of learning paradigms. HIS are used for problem solving of various domains [45]. Lertpalangsunti [46] proposed three reasons for creating HIS viz. a] technique enhancement b] multiplicity of application task and c] realizing multifunctionality. The degree of integration between the modules may vary from loosely coupled i.e. standalone modules to fully coupled. Application of Hybrid Intelligent System in finance domain is as follows –

A. Portfolio Management:

Portfolio management is a complex activity which involves crucial decision making process. It is an important activity of many financial institutes and organizations. From last few years HIS is widely applied for portfolio selection [47].

Kosaka et al. [48] applied NN and Fuzzy logic for stock portfolio selection. They concluded that proposed model identify price tuning points with 65% accuracy. Chen et al. [49] developed a portfolio selection model. In the proposed model triangular fuzzy numbers are used to represent future returns rates and risk of mutual funds. Quek et al. [50] developed a fuzzy-neural system for portfolio balancing with the help of GenSoFNN i.e. Generic Self-organizing Fuzzy NN. They applied supervised learning method in the network for detection of inflection points in stock price cycle. Yu l et al. [51] developed NN dependent mean-variance skewness model for portfolio section on the basis of integration of RBF and Lagrange multiplier theory of optimization. Li et al. [52] proposed a hybrid intelligent algorithm by assimilating NN, simulated annealing algorithm and fuzzy simulation techniques for solving portfolio selection problems. IN the proposed model, NN is for approximation of expected value and variance of fuzzy returns. Fuzzy simulation generates the training data for NN. They also compared between their model and genetic algorithm. Quah et al. [53] compared the performance of ANFIS, MLP-NN and GGAP-RBF (General Growing Pruning Radial Basis Function). Author also proposed the method of selection of equities by ROC (Relative Operating Characteristics) curve.

B. Stock Market Prediction:

Volatile nature of stock market needs various computing techniques. As compared to other domains, hybrid AI systems are widely used for financial prediction because hybrid systems are able to combine the capabilities of various systems with their unique abilities.

Kuo et al. [54] developed a system for stock market forecasting. Proposed model deals with qualitative and quantitative factors simultaneously. The system was developed by integrating fuzzy Delphi model and NN for qualitative and quantitative factors respectively. The system is tested on the database of Taiwan Stock Market and found considerably better. Romahi et al. [55] proposed a rule-based Expert System for financial forecasting. They combine rule induction & fuzzy logic and observed that system performs better. Keles et al. [56] developed a model for forecasting domestic debt MFDD. They applied ANFIS for few microeconomic variables of Turkish economy. They observed that MFDD performs better in terms of forecasting. Kuang et al. [57] combined average autoregressive exogenous (ARX) model for prediction with grey system theory and rough set to forecast stock market automatically. They employed GM (1,N) model for data reduction of Taiwan stock exchange. They observed hybrid method has greater forecasting ability.

C. Risk Management:

Risk management is decision making activity that involves social, political, engineering and economic factors. Risk could arise in the form of fraud, bankruptcy etc. Elmer et al. proposed hybrid fuzzy logic and neural network algorithm for credit risk management. HFNN (Hybrid Fuzzy logic-Neural Network) model is used for credit risk evaluation and concluded that HFNN model is robust, accurate and reliable [58]. Lean et al. [59] proposed a HIS for credit risk evaluation and analysis using rough set and support vector machine (SVM). SVMs are used to extract the features and for noise filtration. Rough set acted as a preprocessor for support vector machine. They concluded that proposed model performs better.

Hyunchul et al. [60] focused most important issue of corporate bankruptcy prediction. Various data driven approaches are applied to enhance prediction performance using statistical and AI techniques. Case based reasoning (CBR) is the most widely used data driven approach. Model is developed by combining CBR with genetic algorithm (Gas) and observed that model generates accurate results along with reasonable explanation. Zopounidis et al. [61] presented a review on application of Knowledge Base Decision Support System (KBDSS) in finance and management. KBDSS is developed by combining the features of an Expert System and Decision Support System in many fields like financial analysis, bankruptcy risk assessment, financial planning. Authors described KBDSS for portfolio management, financial analysis, and credit gaining problems. They observed that KBDSS improvise the decision-making process by explaining the operations and result generated by the system. Hua et al. [62] applied Support Vector Machine (SVM) for bankruptcy prediction and proved competing against neural network, logistic regression and linear multiple discriminant analysis. Authors developed an Integrated Binary Discriminant Rule (IBDR) for financial distress prediction. The experimental results proved that IDBR performs better as compared to the conventional SVM.

Table (3): Brief review of application of HIS for Portfolio Management, Stock Market Prediction and Risk Management

Author	Objective	Data Set	Preprocessing	Preprocessing Approach Used		Evaluation Metric
Li X, Zhang Y, Wong H-S, Qin Z	To develop a model for portfolio selection.	Data from B. Liu, Theory and Practice of Uncertain Programming, Physica-Verlag, Heidelberg, 2002	Not mentioned	ANN, GA and Fuzzy simulation	GA GA and	Robust and effective
Ohkawa E, Mabu S, Shimada K, Hirasaa K	a model for portfolio optimizatio n.	market show	Not memoried	controlled nodes, Technical analysis rules	B&H method, Conventional GNP-based methods	SS
Shian-Chang Huang	Stock indices Forecasting	NASDAQ (US), NK225 Japan), TWSI (Taiwan) and KOSPI (South Korea).	stock indices are transformed into daily returns	Hybrid Model	Traditional time domain models	Forecasting
YF Wang	Stock price prediction	Taiwan stock market	Every 5 min. data is recorded	Fuzzy grey prediction	GA	Prediction
Lin S-H	Credit risk assessment	Taiwan Banks	Not mentioned	LR, ANN	ANN, logarithm LR and LR	Prediction
Arminger, Gerhard, Daniel Enache, and Thorsten Bonne	To analyze credit risk		Data is extracted on the basis some factors like age, sex, job etc.	BPNN,LDA	Discriminant Techniques	Classificati on
Kun Chang Lee, Ingoo Han, Youngsig Kwon	To develop the hybrid neural network models for bankruptcy prediction	Korea Stock Exchange	Divided data set into two categories i.e. Failed firms and Non-failed firms.	Hybrid Neural Network	MDA, ID3, SOFM	Accuracy and Adaptabilit y
Garcia- Almanza, Alma Lilia, Biliana Alexandrova- Kabadjova, and Serafin Martinez- Jaramillo	Bankruptcy prediction for banks	FDIC Data	Not mentioned	MP-EDR	Machine Learning	Prediction
Kyung-Shik Shin , Yong-Joo Lee	To predict corporate failure		528 externally audited mid- sized manufacturing firms.	GA	Traditional Statistical Methods	Rule Extraction

Hsueh-Ju Chen, Shaio Yan Huang , Chin- Shien Lin	To develop a model for bankruptcy prediction.	Securities and Exchange Commission(SE C)	Not mentioned	Fuzzy Logic	Traditional methods	Accuracy
Sung-Hwan Mina, Jumin Leeb, Ingoo Hanb	Bankruptcy prediction	Commercial bank in Korea	Financial ratios categorized as stability, Profitability, Growth, activity and Cash flow	SVM-based model	neural network (NN) and logistic regression	Feature subset selection and parameter optimizatio n
Lenard Watkins and Alam	To detect financial statement frauds.	Retail and manufacturing industries	Not mentioned	Fuzzy Logic	External parties	Accuracy
Deshmukh, Romine and Siegel	To provide a fuzzy sets model to assess the risk of managerial fraud	Business accounts	Not mentioned	Fuzzy Logic	Traditional systems	Assessment
Juszczak et al.	To detect fraud in financial statement.	D1 and D2	Not mentioned	Supervised and semi- Supervised Classificatio n	Supervised classification versus unsupervised classification	Classificati on
Lenard and Alam	To develop a fuzzy logic model for developmen t of clusters to represent red flags in detection of fraud	Not mentioned	Not mentioned	Fuzzy Logic & Expert Reasoning	Statistical methods	Clustering

V. Conclusion:

A comprehensive review is conducted on applications of Artificial Intelligence in Finance management. The review is organized by considering the type of techniques, their application domain, objectives and evaluation metrics. The review indicates that traditional approach is not sufficient enough to tackle and analyze huge financial data. Hence contemporary methods need to be applied.

Artificial Intelligence technique is an important dimension of present research. An important conclusion which is drawn from this research is that researchers employ various AI techniques to solve the problems associated with finance management effectively. Comparative study illustrates that ANNs are successful in financial prediction. But it is quite difficult to determine the required structure and size of a neural network to solve the given problem. The major difficulty with ANN is that they are trained using past data which may not be repeated further. Alternative method for this could be an expert system. These expert systems generate predictions. The problem of ES is, they couldn't learn through experience and are unable to handle non-linear data. To overcome these problems hybrid

intelligent systems could be implemented that are able to handle linear and non-linear data. HIS can combine the capabilities of various systems to overcome the limitations of individual techniques. It is observed that limited literature is available on ES and HIS in finance domain that researchers can target.

Computational Finance is a blending of computational power and machine learning techniques to cope up with problems of practical interest in financial domain. This is an application oriented study which proposes innovative techniques to solve the financial domain problem. Although we focus on portfolio management, stock market prediction and risk management related problems only in these study practically all types of financial problems can be addressed computes and with AI techniques in particular. AI techniques are also used for building models of financial markets. Novel approach in the field of research could have a fusion of different techniques [116]. Numbers of researchers in various research institutes are working in this area such as University of Essex, Illinois Institute of Technology, University of Washington and many more.

The objective of this review is to provide introduction of the field of Computational Finance and how AI techniques are being used to deal with problems of practical interest in financial domain. Furthermore, we wish to elaborate how researchers are exploring the strengths of certain AI techniques to overcome the problems that mere statistical techniques cannot deal with. We feel that this review could be a helpful guideline to study various AI techniques and we wish that the researchers would overcome the drawbacks of some techniques and try to develop an influential integrated system by utilizing strengths and complementary features of different techniques.

References:

- 1. Pal, S. K. and Srimani, P. K., "Neurocomputing: Motivation, Models, and Hybridization", *Computer*, ISSN 0018-9162, Vol. 29 No. 3, IEEE Computer Society, NY, USA, pp. 24-28, March 1996.
- 2. Rich, E. & Knight, K., Artificial Intelligence, Second Edition, McGraw Hill, pp. 4-6, 1991.
- 3. Lam M (2004) Neural network techniques for financial performance prediction, integrating fundamental and technical analysis. Decision Support Syst 37:567-581.
- 4. Basheer IA, Hajmeer M (2000) Artificial neural networks: fundamentals, computing, design, and application. J Microbiol Meth 43:3–31.
- Anderson CW, Devulapalli SV, Stolz EA (1995) Determining mental state from EEG signals using parallel implementations of neural networks. Scientific programming, special issue on applications analysis. Fall 4(3):171–183.
- 6. Touzet CF (1997) Neural reinforcement learning for behaviour synthesis. Robotics Autonom Sys 22:251–281.
- 7. Faller W, Schreck S (1995) Real-time prediction of unsteady aerodynamics: application for aircraft control and maneuverability enhancement. IEEE Trans Neural Net 6(6):1461–1468.
- 8. Sang KK, Niyogi P (1995) Active learning for function approximation. In: Tesauro G (ed) Neural information processing systems, vol 7, The MIT Press, Cambridge, pp 497–504.
- 9. Vellido A, Lisboa PJG, Vaughan J (1999) Neural networks in business: a survey of applications (1992–1998). Expert SystAppl 17:51–70.
- 10. Atiya AF (2001) Bankruptcy prediction for credit risk using neural networks: a survey and new results. IEEE Trans Neural Net 12(4):929–935.
- 11. Motiwalla L, Wahab M (2000) Predictable variation and profitable trading of US equities: a trading simulation using neural networks. ComputOper Res 27:1111–1129.
- 12. Yamamoto Y, Zenios SA (1993) Predicting prepayment rates for mortgages using the cascade correlation learning algorithm. J Fixed Income 2(4):86–96.
- 13. Lowe, David. "Novel exploitation of neural network methods in financial markets." *Neural Networks, 1994. IEEE World Congress on Computational Intelligence, 1994 IEEE International Conference on*. Vol. 6. IEEE, 1994.
- 14. Badiru, Adedeji B., and David B. Sieger. "Neural network as a simulation metamodel in economic analysis of risky projects." *European Journal of Operational Research* 105.1 (1998): 130-142.
- 15. Zimmermann HJ, Neuneier R, Grothmann R (2001) Active portfolio-management based on error correction neural networks, Advances in neural information processing systems, NIPS.
- 16. Ellis C, Willson P (2005) Can a neural network property portfolio selection process outperform the property market? J Real Estate Portfolio Manage 11(2):105–121

- 17. White, H. (1988). Economic prediction using neural networks: the case of IBM daily stock returns. In Proceedings of the second IEEE annual conference on neural networks, II (pp. 451–458).
- Kimoto, T., Asakawa, K., Yoda, M., &Takeoka, M. (1990). Stock market prediction system with modular neural networks. In Proceedingof the international joint conference on neural networks (IJCNN) (Vol.1, pp. 1–6.) San Diego.
- 19. Kohara, Kazuhiro, et al. "Stock price prediction using prior knowledge and neural networks." Intelligent systems in accounting, finance and management 6.1 (1997): 11-22.
- Pai, Ping-Feng, and Chih-Sheng Lin. "A hybrid ARIMA and support vector machines model in stock price forecasting." Omega 33.6 (2005): 497-505.
- S. Thawornwong, D. Enke, The adaptive selection of financial and economic variables for use with artificial neural networks, Neurocomputing 56 (2004) 205–232.
- 22. S. Nakayama, S. Horikawa, T. Furuhashi, Y. Uchikawa, Knowledge acquisition of strategy and tactics using fuzzyneural networks, Proc. IJCNN'92, 1992, pp. II-751{II-756.
- L. Duke, J. Long, Neural network futures trading—a feasibility study, in: Society for Worldwide Interbank Financial Telecommunications (Ed.), Adaptive Intelligent Systems, Society for Worldwide Interbank Financial Telecommunications, Elsevier Science Publishers, Amsterdam, 1993, pp. 121–132.
- 24. W. Treacy and M. Carey, "Credit risk rating at large US banks," J.Banking& Finance, vol. 24, pp. 167–201, 2000.
- 25. Z. Yang, M. Platt, and H. Platt, "Probabilistic neural networks in bankruptcy prediction," J. Business Res., vol. 44, pp. 67–74, 1999.
- 26. K. Y. Tam and M. Kiang, "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions," Management Science, Vol. 38, No. 7, 1992, pp. 926-947.
- Z. Huang, H. Chen, C. J. Hsu, W. H. Chen and S. Wu, "Credit Rating Analysis with Support Vector Machines and Neural Networks: A Market Comparative Study," Decision Support System, Vol. 37, No. 4, 2004, pp. 543-558.
- 28. Matsatsinis, N. F., Doumpos, M., & Zopounidis, C. (1997). Knowledge acquisition and representation for expert systems in the field of financial analysis. Expert Systems with Applications, 12(2), 247-262.
- 29. Ellis C, Willson P (2005) Expert system portfolios of Australian and UK securitized property investments. Pacific Rim Property Res J 12(1):107–127.
- 30. Liao S-H (2005) Expert system methodologies and applications a decade review from 1995 to 2004. Expert SystAppl 28:93–103.
- 31. Ellis, Craig, and Patrick Wilson. "Expert system portfolios of Australian and UK securitised property investments." *Pacific Rim Property Research Journal* 12.1 (2006): 107-127.
- 32. Bohanec M, Rajkovic V, Semoil B, Pogacnik A (1995) Knowledge-based portfolio analysis for project evaluation. Information Manage 28(5):293–302.
- 33. Vraneš, Sanja, et al. "INVEX: Investment advisory expert system." Expert Systems 13.2 (1996): 105-119.
- 34. Vranes S, Stanojevic M, Stevanovic V, Lucin M (1996) INVEX: investment advisory expert system. Expert SystAppl 13(2):105–119.
- 35. Mogharreban N, Zargham R (2005) PORSEL: an expert system for assisting in investment analysis and valuation, soft computing a fusion of foundations. MethodolAppl 9(10):742–748.
- Grosan, C., Abraham, A., Ramos, V., & Han, S. Y. (2005). Stock market prediction using multi expression programming. In Proceedings of Portuguese conference of artificial intelligence, workshop on artificial life and evolutionary algorithms (pp.73–78). Portuguese: IEEE Press.
- 37. Quek, C. (2005). Predicting the impact of anticipator action on US stock market An event study using ANFIS (a neural fuzzy model). Computational Intelligence, 23,117–141.
- 38. Trinkle, B. S. (2006). Forecasting annual excess stock returns via an adaptive network-based fuzzy inference system. Intelligent Systems in Accounting, Finance and Management, 13(3), 165–177.
- Afolabi, M., &Olatoyosi, O. (2007). Predicting stock prices using a hybrid Kohonen self-organizing map (SOM). In 40th Annual Hawaii international conference on system sciences (pp. 1–8).
- Yunos, Z. M., Shamsuddin, S. M., & Sallehuddin, R. (2008). Data Modeling for Kuala Lumpur Composite Index with ANFIS. In Second Asia international conference on modeling and simulation, AICMS 08, Kuala Lumpur (pp. 609–614).
- 41. Atsalakis, G. S., & Valavanis, K. P. (2009). Forecasting stock market short-term trends using a neuro-fuzzy based methodology. Expert Systems with Applications, 36(7), 10696-10707.

- 42. Matsatsinis NF, Doumpos M, Zopounidis C (1997) Knowledge acquisition and representation for expert system in the field offinancial analysis. Expert SystAppl 12(2):247–262.
- 43. Shue L-Y, Chen C-W, Shiue W (2009) The development of anontology-based expert system for corporate financial rating.ExpertSystAppl 36(2):2130–2142.
- 44. Hodgkinson, Luke, and Ellen Walker. "An expert system for credit evaluation and explanation." Journal of Computing Sciences in Colleges 19.1 (2003): 62-72.
- 45. Taha IE (1997) A hybrid intelligent architecture for revising domain knowledge. Ph. D. Report. University of Texas at Austin.
- 46. Lertpalangsunti N (1997) An implemented framework for the construction of hybrid intelligent forecasting systems. Ph. D.Report. University of Regina
- 47. Lin C, Hsieh P-J (2004) A fuzzy decision support system for strategic portfolio management. Decision Support Syst 38:383–398.
- 48. Kosaka M, Mizuno H, Sasaki T, Someya T, Hamada N (1991)Applications of fuzzy logic/neural network to securities trading decision support system. Proc IEEE ConfSyst, Man, Cyber3:1913–1918.
- 49. Chen L-H, Huang L (2009) Portfolio optimization of equity mutual funds with fuzzy return rates and risks. Expert SystAppl36(2):3720–3727.
- Quek C, Yow KC, Cheng PYK, Tan CC (2009) Investment portfolio balancing: application of a generic selforganizing fuzzy neural network (GenSoFNN). Risk Anal Complex Syst:IntelligentSyst Finance 16(1/2):147–164.
- 51. Yu L, Wang S, Lai KK (2008) Neural network-based mean-variance-skewness model for portfolio selection. ComputOperRes 35:34-46.
- 52. Li X, Zhang Y, Wong H-S, Qin Z (2009) A hybrid intelligent algorithm for portfolio selection problem with fuzzy returns. JComputAppl Math 233(2):264–278
- 53. Quah T-S (2008) DJIA stock selection assisted by neural network. ExpertSystAppl 35(1–2):50–58.
- 54. Kuo R J, Lee LC, Lee CF (1996) Integration of artificial neural networks and fuzzy delphi for stock market forecasting. IEEE,pp 1073–1078.
- 55. Romahi Y, Shen Q (2000) Dynamic financial forecasting with automatically induced fuzzy associations. IEEE Syst 1:493–498.
- 56. Keles A, Kolcak M, Keles A (2008) The adaptive neuro-fuzzymodel for forecasting the domestic debt. Knowledge-Based Syst21(8):951–957.
- Huang, Kuang Yu, and Chuen-Jiuan Jane. "A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories." *Expert Systems with Applications* 36.3 (2009): 5387-5392.
- 58. Dadios, E. P., & Solis, J. (2012). Fuzzy-neuro model for intelligent credit risk management. Intelligent Information Management, 4(05), 251.
- 59. Lean, Y. U., et al. "Designing a hybrid intelligent mining system for credit risk evaluation." *Journal of Systems Science and Complexity* 21.4 (2008): 527-539.
- 60. Ahn, Hyunchul, and Kyoung-jae Kim. "Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach." *Applied Soft Computing* 9.2 (2009): 599-607.
- Zopounidis, Constantin, Michael Doumpos, and Nikolaos F. Matsatsinis. "On the use of knowledge-based decision support systems in financial management: a survey." *Decision Support Systems* 20.3 (1997): 259-277.
- 62. Hua, Zhongsheng, et al. "Predicting corporate financial distress based on integration of support vector machine and logistic regression." *Expert Systems with Applications* 33.2 (2007): 434-440.
- 63. Shaw M, Gentry J "Using an expert system with inductive learning to evaluate business loans."FinancManag 17(3):45–56
- 64. Bohanec M, Rajkovic V, Semoil B, Pogacnik A (1995) Knowledge-based portfolio analysis for project evaluation. Information Manage 28(5):293–302
- 65. *Ellis C, Willson P* (2005) Expert system portfolios of Australian and UK securitized property investments. Pacific Rim Property Res J 12(1):107–127.
- 66. Casqueiro PX, Rodrigues AJL (2006) Neuro-dynamic trading methods. Eur J Oper Res 175:1400–1412
- 67. *Stoppiglia H, Idan Y, Dreyfus G* (1996) Neural-network-aided portfolio management, Industrial applications of neural networks. In: Fogelman F, Gallinari P (eds) Proceedings of ICNN95, conference of European Union Neural Network, World Scientific, Paris.
- 68. *Yu L, Wang S, Lai KK* (2008) Neural network-based mean-variance-skewness model for portfolio selection. ComputOper Res 35:34–46

- 69. *Li X, Zhang Y, Wong H-S, Qin Z* (2009) A hybrid intelligent algorithm for portfolio selection problem with fuzzy returns. J ComputAppl Math 233(2):264–278
- 70. Chen Y, Ohkawa E, Mabu S, Shimada K, Hirasaa K (2009) A portfolio optimization model using genetic network programming with control nodes. Expert SystAppl 36(7):10735–10745
- Financial news analysis for intelligent portfolio management *Young-Woo Seo* Carnegie Mellon University, *Joseph A. Giampapa,KatiaSycara* Carnegie Mellon University Research Showcase @ CMU Robotics Institute School of Computer Science.
- 72. Active Portfolio-Management based on Error Correction Neural Networks *Hans Georg Zimmermann, Ralph Neuneier and Ralph GrothmannSiemensAG*Corporate TechnologyD-81730M⁻⁻unchen, Germany
- 73. Ellis C, Willson P (2005) Can a neural network property portfolio selection process outperform the property market? J Real Estate Portfolio Manage 11(2):105–121
- 74. Fernandez A, Gomez S (2007) Portfolio selection using neural networks. ComputOper Res 34:1177-1191
- 75. *Freitas FD, De Souza AF, de Almeida AR* (2009) Prediction based portfolio optimization model using neural networks. Neurocomputing 72(10–12):2155–2170
- 76. Resource allocation neural network in portfolio selection Po-Chang Ko, Ping-Chen Lin[,] Expert Systems with Applications Volume 35, Issues 1–2, July–August 2008, Pages 330–337
- 77. Huang, Shian-Chang. "Forecasting stock indices with wavelet domain kernel partial least square regressions." Applied soft computing 11.8 (2011): 5433-5443.
- 78. *Chiang W-C, Urban TL, Baldridge GW* (1996) A neural network approach to mutual fund net asset value forecasting. Int J Manage Sci 24(2):205–215
- 79. *Chen A-S, Leung MT, Daouk H* (2003) Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. ComputOper Res 30:901–923
- 80. O'Connor N, Madden MG (2006) A neural network approach to predicting stock exchange movements using external factors. Knowledge-Based Syst 19:371–378
- Faria EL, Albuquerque MP, Gonzalez JL, Cayalcante JTP (2009) Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods. Expert SystAppl 36(10):12506– 12509
- 82. *Liao A, Wang J* (2010) Forecasting model of global stock index by stochastic time effective neural network. Expert SystAppl 37:834–841
- 83. *Kim SH, Chun SH* (1998) Graded forecasting using an array of bipolar predictions: application of probabilistic neural networks to a stock market index. Int J Forecast 14:323–337
- 84. Wang ZB (2004) Prediction of stock market prices using neural network techniques, Ph. D. Thesis, University of Ottawa
- A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets *Hyun-jung Kim, Kyung-shikShin*, Applied Soft Computing Volume 7, Issue 2, March 2007, Pages 569–576
- 86. An intelligent stock trading decision support system through integration of genetic algorithm based fuzzy neural network and artificial neural network *R.J. Kuo, C.H. Chen, Y.C. Hwang*, Fuzzy Sets and Systems 118 (2001) 21{45 0165-0114/01/\$ see front matter c 2001 Elsevier Science.
- 87. A fusion model of HMM, ANN and GA for stock market forecasting *Md. RafiulHassan, BaikunthNath, Michael Kirley*, Expert Systems with Applications Volume 33, Issue 1, July 2007, Pages 171–180.
- 88. Predicting stock price using fuzzy grey prediction system *Y.-F Wang*, Expert Systems with Applications Volume 22, Issue 1, January 2002, Pages 33–38.
- 89. Jerry W. Lin, Mark I. Hwang, Jack D. Becker, (2003) "A fuzzy neural network for assessing the risk of fraudulent financial reporting", Managerial Auditing Journal, Vol. 18 Iss: 8, pp.657 665
- 90. Arminger G, Enache D, Bonne T (1997) Analyzing credit risk data: a comparison of logistic discrimination classification tree analysis and feedforward networks. Comput Stat 12:293–310
- Chye KH, Tan WC, Goh CP (2004) Credit scoring using data mining techniques. Singapore Manage Rev 26(2):25–47
- 92. Lin S-H (2009) A new two-stage hybrid approach of credit risk in banking industry. Expert SystAppl 36(4):8333-8341
- 93. A neural network approach for credit risk evaluation ElianaAngelini, Giacomo di Tollo, Andrea Roli, The Quarterly Review of Economics and Finance Volume 48, Issue 4, November 2008, Pages 733–755

- 94. Hybrid neural network models for bankruptcy predictions, *Kun Chang Lee, Ingoo Han, Youngsig Kwon* Decision Support Systems Volume 18, Issue 1, September 1996, Pages 63–72 Artificial Intelligence and Expert Systems.
- 95. Comparing statistical methods and artificial neural networks in bankruptcy prediction Chu, Jung, Ph.D. (1997) *Comparing statistical methods and artificial neural networks in bankruptcy prediction.* PhD thesis, University of Warwick.
- 96. Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis Guoqiang Zhang a, Michael Y. Hu b,*, B. Eddy Patuwo b, Daniel C. Indro b a Department of Decision Sciences, College of Business, Georgia State University, Atlanta, GA 30303, USA b Graduate School of Management, College of Business Administration, Kent State University, Kent, OH 44240-0001, USA Received 10 March 1997; accepted 22 December 1997. European Journal of Operational Research 116 (1999) 16±32
- 97. Tam KY, Kiang MY (1992) Managerial applications of the neural networks: the case of bank failure predictions. Manage Sci 38(7):926–947
- 98. Early bankruptcy prediction using ENPCDavid Quintana · YagoSaez · Asuncion MochonPedroIsasi D Quintana, Y Saez, A Mochon, P Isasi Applied Intelligence, 2008 Springer.
- 99. Predicting Financial Distress of Companies: Revisiting the Z-Score & ZETA Models Edward I Altman
- 100.Garcia-Almanza, Alma Lilia, Biliana Alexandrova-Kabadjova, and Serafin Martinez-Jaramillo. "Bankruptcy Prediction for Banks: An Artificial Intelligence Approach to Improve Understandability." Artificial Intelligence, Evolutionary Computing and Metaheuristics. Springer Berlin Heidelberg, 2013. 633-656.
- 101.Selecting the neural network topology for student modelling of prediction of corporate bankruptcy M.L. Nasir, R.I. John, S.C. Bennett, D.M. Russell, (2001) "Selecting the neural network topology for student modelling of prediction of corporate bankruptcy", Campus-Wide Information Systems, Vol. 18 Iss: 1, pp.13 - 22
- 102.A genetic algorithm application in bankruptcy prediction modeling *Kyung-Shik Shin*, *Yong-JooLee*European Journal of Operational Research Volume 180, Issue 1, 1 July 2007, Pages 1–28
- 103.Alternative diagnosis of corporate bankruptcy: A neuro fuzzy approachHsueh-Ju Chen, Shaio Yan Huang, Chin-ShienLinExpert Systems with Applications Volume 36, Issue 4, May 2009, Pages 7710–7720 Copyright © 2014 Elsevier.
- 104.Hybrid genetic algorithms and support vector machines for bankruptcy prediction Sung-Hwan Mina, JuminLeeb, IngooHanb Expert Systems with Applications Volume 31, Issue 3, October 2006, Pages 652–660
- 105.Fanning, Kurt, Kenneth O. Cogger, and RajendraSrivastava. "Detection of management fraud: A neural network approach." Intelligent Systems in Accounting, Finance and Management 4.2 (1995): 113-126.
- 106.Feroz, EhsanHabib, et al. "The efficacy of red flags in predicting the SEC's targets: An artificial neural networks approach." International Journal of Intelligent Systems in Accounting, Finance & Management 9.3 (2000): 145-157.
- 107.Fanning, Kurt M., and Kenneth O. Cogger. "Neural network detection of management fraud using published financial data." International Journal of Intelligent Systems in Accounting, Finance & Management 7.1 (1998): 21-41.
- 108.Koskivaara, Eija, and Barbro Back." Artificial neural network assistant (ANNA) for continuous auditing and monitoring of financial data." Journal of Emerging Technologies in Accounting 4.1 (2007): 29-45.
- 109.Liou, Fen-May. "Fraudulent financial reporting detection and business failure prediction models: a comparison." Managerial Auditing Journal 23.7 (2008): 650-662.
- 110. Application of Fuzzy Logic to Fraud Detection Mary Jane Lenard University of North Carolina Greensboro, USA PervaizAlam Kent State University, USA
- 111.Lenard, Mary Jane, Ann L. Watkins, and PervaizAlam. "Effective use of integrated decision making: An advanced technology model for evaluating fraud in service-based computer and technology firms." Journal of Emerging Technologies in Accounting 4.1 (2007): 123-137.
- 112.Deshmukh, Ashutosh, Jeff Romine, and Philip H. Siegel. "Measurement and combination of red flags to assess the risk of management fraud: a fuzzy set approach." Managerial Finance 23.6 (1997): 35-48.
- 113.Juszczak, Piotr, et al. "Off-the-peg and bespoke classifiers for fraud detection." Computational Statistics & Data Analysis 52.9 (2008): 4521-4532.
- 114. https://en.wikipedia.org/wiki/Artificial_neural_network.
- 115. Mat Buckland & Andre LaMothe, AI Techniques for Game Programming, 2002, Premier Press.

116.Hassan, Md Rafiul, Baikunth Nath, and Michael Kirley. "A fusion model of HMM, ANN and GA for stock market forecasting." Expert Systems with Applications 33.1 (2007): 171-180.



ROLE OF UNMANNED ARIEL VEHICLE (DRONES) IN AGRICULTURE SECTOR IN CONCERN WITH PRECISION AGRICULTURE Dr. Santosh Deshpande¹, Mr. Avinash Satish Pagare²

¹Department of Management, DYPIMR, Pune – 18,

²Department of Master of Computer Applications, MES IMCC, Pune – 38

ABSTRACT

Since the beginning of agriculture, the big paradigm shifts are caused by newly developed techniques and tools to cultivate arable land. There is constant improvement and change in methods of farming according to new studies and technologies. As we now have modern sensing technology that expands our sensing capabilities along with the birds view perspective, we thrive to utilize the possibilities it gives us. The introduction of Precision Agriculture has changed the way of farming. "Precision Agriculture generally involves management of farm inputs such as herbicides, fertilizers, seed, fuel (used for, planting spraying and tillage) by doing the right management practice at the right place and right time". This study is focusing on the latest technology trend in the agricultural sector, an Unmanned Ariel Vehicle (UAV) - Drone. The UAV technology can perform aerial mapping, giving a clear image of the total size of a crop field, as well as showing possibly underutilized areas of land. By deploying sensors and mapping fields, farmers can begin to understand their crops at a micro scale, conserve resources, and reduce impacts on the environment. They can also able to monitor the health of plants in terms of temperature, chlorophyll levels, foreign contaminants and even leaf thickness. The practice of precision agriculture, allows farmers to maximize yields using minimal resources such as water, fertilizer, and seeds also to address the problems by adjusting necessary parameters before they become more widespread.

Keywords: Precision Agriculture, Unmanned Aerial Vehicle, DSS, Sensors.

[1] INTRODUCTION

ROLE OF UNMANNED ARIEL VEHICLE (DRONES) IN AGRICULTURE SECTOR IN CONCERN WITH PRECISION AGRICULTURE

As technology has infiltrated our lives through smartphones, computers and the internet, agriculture has remained the last bastion of tradition as an enterprise of human labor and intellect. Despite mechanization, much of farming involves human labor and legwork. Farmers walk down their fields checking for plant health, the presence of weeds, bugs, parched soil and other overt signs of distress. However, even with a trained eye, it is often hard to notice diminished photosynthesis or signs of pathogen infestation early enough to avoid large losses. This, in turn, has led to a culture of prevention where farmers have resorted to the excessive use of water, fertilizers, pesticides, etc. Otherwise, the cost of non-intervention, or even late intervention, is massive crop failure.

[1.1] PRECISION AGRICULTURE

A key component of this agriculture management approach is the use of information technology, thus, Precision Farming can make a difference to food production by using a wide array of technology such as GPS guidance, control systems, sensors, robotics, drones, autonomous vehicles, variable rate technology, GPS-based soil sampling, automated hardware, telematics, and software.

A Precision agriculture is a viable solution to agricultural production due to:

- i) The ongoing reduction in the amount of arable land;
- ii) Projected increase in global population, and;
- iii) Reduction in cost of agriculture by avoiding wastage in the application of chemicals or water.

Precision Farming is about managing variations in the field accurately to grow more food using fewer resources and reducing production costs.

[1.2] CYCLE OF PRECISION AGRICULTURE

The cycle of precision farming involves obtaining images or data to generate map yields, weeds and topography, before application of the herbicides or fertilizer, water and finally to obtain results for implementation. Before an agriculturist starts the process of precision farming, an understanding of the soil types, hydrology, micro-climates, aerial photography as well as an understanding of the variable factors within the fields that affect a yield map, are required.



Remote sensing used in implementing the technology include platforms such as satellites, aircraft, balloons and helicopters, and a variety of sensors such as optical and near-infrared and RADAR (Radio Detection and Ranging) installed on these platforms for its applications. Diagnostic information derived from images collected from these on-board sensors,

fig. Precision Farming techniques employed in the crop growth cycle:

such as biomass, Leaf Area Index (LAI), disease, water stress and lodging, can thus assist in crop management, yield forecasting, and environmental protection.

[1.3] BENEFITS OF PRECISION AGRICULTURE

Precision agriculture helps to monitor the vegetation's physical and chemical parameters. These ensure that the optimal conditions for plant growth are achieved. This makes the field administration automated by consolidating a Decision Support System (DSS) in the precision agriculture environment.

Information from the sensors is naturally transmitted to a central server and this can be consulted utilizing a Smartphone or Laptop. On the other hand, email or SMS alerts can be modified to tell the field holder when there is a need to flood, prepare or address any issue in their properties. Also, costs regarding water, pesticides and others are improved and can undoubtedly be decreased.

[2] UAV'S IN PRECISION AGRICULTURE

The use of UAVs in agriculture is fast becoming widespread, while the implementation of aerospace engineering and sensor technology are reducing in cost. Drones, or Unmanned Aerial Vehicles (UAVs), have been adopted and can have a significant impact on the agricultural industry, as they can monitor several aspects of farming that humans cannot accomplish on their own.

ROLE OF UNMANNED ARIEL VEHICLE (DRONES) IN AGRICULTURE SECTOR IN CONCERN WITH PRECISION AGRICULTURE

UAVs employ cameras to collect images and sensors to compile a set of data to help with monitoring and decision-making on-farm. UAVs collect data at high spatial resolutions enabling differences in crops to be compared by the centimeter rather than the meter, as in the case with satellites. They also provide immediate visual information about large areas of crops, which help farmers with fast decision-making.

UAVs are small-sized electronically controlled devices launched from the side of a field or area of interest to the user. The size of UAVs usually affects the size of the payload, respectively. UAVs are guided by either a radio controller. Cameras attached to a UAV takes pictures that can be two dimensional, with every pixel linked to a GPS location on the ground.

UAVs can also transmit live videos from flight to the receiving station on the ground. Depending on the type of camera employed, thermal images can be obtained from the UAV.

[2.1] IMPORTANCE OF UAV

It can be understood from the fact that it can provide farmers with three detailed views:

- 1. First, keeping an eye on crop from the air can help reveal patterns that show a problem related to irrigation, soil variation and fungal infestations.
- 2. Secondly, drone uses Satellite remote sensing method which is used to identify the crop growth by comparing multiple images taken by the satellite.
- 3. Third, airborne cameras can take multispectral images, capturing data using visual spectrum as well as infrared, which shows the difference between the distressed and healthy plants which can't be viewed with naked eyes.

This, in turn, results in higher crop yields, as many of the common agricultural problemscan be identified and dealt with earlier than with conventional means of detection.

[2.2] LIMITATIONS OF UAVS

• The low costs of UAVs (the cost of purchasing a drone is just a part and not inclusive of the cost of processing, collecting and producing images) limit the sensor payload, both in weight and dimension, and they are also not as stable as the high-end sensors resulting in reduced image quality. Since low cost UAVs are usually equipped with light weight engines, this limits their reachable altitude.

- Also, the available commercial software packages applied for photogrammetric data processing are set up to support UAV images, as there are no standardized work flows and sensor models currently being implemented.
- UAV images do not also benefit from the sensing and intelligent human feature that limits them from taking actions based on unexpected situations, like avoiding collisions with an oncoming flying object and also requires full understanding before it can be deployed with maximum interaction with its receiving station, usually a laptop.
- The frequency of flying UAVs also has to be carefully selected and, finally, there are insufficient regulations on flying UAVs and they are restricted in certain regions as a security precaution.
- Some other challenges are their inability to take readings during extremeweather conditions like rain.

[3] CONCLUSION

Unmanned Ariel Vehicle (UAV) based Precision Agriculture will reap many benefits to the farmers and potential users, in the form of cultivation to consumption. It would be possible to provide timely and accurate inputs could be given to processes. It will also help the environment by reducing usage of fertilizers, pesticides etc. The pesticide free vegetation attracts more consumers which also increases the health benefit of farmers and users of farm products.Therefore, the technological application like UAV holds a great potential for a more sustainable and precise agriculture.

ROLE OF UNMANNED ARIEL VEHICLE (DRONES) IN AGRICULTURE SECTOR IN CONCERN WITH PRECISION AGRICULTURE

REFERENCES

- C. Zhang and J. M. Kovacs, "The application of small unmanned aerial systems for precision agriculture: A review," Precis. Agric., vol. 13, pp. 693–712, 2012.
- [2] Lambert, D., &Lowenberg-Deboer, J. (2000). Precision agriculture profitability review (p. 154).
 Purdue, USA: Site Specific Management Center.
- [3] Lelong, C. C. D., Burger, P., Jubelin, G., Roux, B., Labbe, S., &Barett, F. (2008). Assessment of unmanned aerial vehicles imagery for quantitative monitoring of wheat crop in small plots. Sensors, 8, 3557–3585.
- [4] Zhou, G. (2009). Near real-time ortho rectification and mosaic of small UAV flow for time-critical event response. IEEE Transactions on Geoscience and Remote Sensing, 47, 739–747.
- [5] HalimatuSadiyahAbdullahi and Ray E Sheriff, "Technology Impact on Agricultural Productivity: A Review of Precision Agriculture Using Unmanned Aerial Vehicles"
- [6] Moran, M. S., Inoue, Y., & Barnes, E. M. (1997). Opportunities and limitation for image-based remote sensing in precision crop Management. Remote Sensing of Environment, 61, 319–346.
- [7] Schmale, D. G., Dingus, B. R., &Reinholtz, C. (2008). Development and application of an autonomous aerial vehicle for precise aerobiological sampling above agricultural fields. Journal of Field Robotics, 25, 133–147.

CURRENT ASPECTS IN HANDWRITING BIOMETRICS

Dr. Santosh Deshpande¹, Mukul Kulkarni²

¹Additional Director, MES'S Institute of Management and Career Courses Pune, India

²Research scholar, Savitribai Phule Pune University, Pune, India

ABSTRACT:

The handwriting biometrics is new branch for user identification and analysis. New techniques for understanding of user and authentication of user through online and offline databases using repository are emerging day by day. Traditional handwriting analysis system is being evolved using different techniques like graphometry, signature segmentation and verification by image processing techniques. Many neuroscientists, psychologies, handwriting experts are defining and redefining aspects of handwriting biometrics everyday This paper focuses on all such aspects in handwriting biometrics with their proposed outcomes in detail.

Keywords: Graphology, Trait and Gestalt method, Conscience, Pseudoscience, Personality, Brain writing, Neurological writing pattern

[1] INTRODUCTION

In the current era, human analysis is becoming important. For analysis user can observe special body movements, changes in voice tone, body gestures and posture, sometimes body structure and dressing sense etc. for interpretation of behavior and psychological understanding. Sometimes it's possible for masqueraded behavior or suppress behavior to be inaccurately analyzed. As certain fascial expression is associated with different moods like happy, sad etc. We can observe similar trends in handwriting of various writers. [1] [4]

Figure: 1. Physical Process of writing

Handwriting is conscious efforts made by writer by the instructions of brain. The process of handwriting is made on different levels consciously, unconsciously as well as due to nervous system and body movements. [9]. The actual writing process could be explained as follows: Every individual's handwriting depicts similar features like faces, body, fingers, voices, anger or warmth. Basically, that's the reflection of person with characteristics, special signature and writers state of mind. Major influence on writing is due to the emotional and circumstantial happening in and around the writer. Handwriting analysis is scientific method which reveals true personality traits. It is often referred as brain writing due to neurological brain pattern produced by neuromuscular movement creating writing strokes. It's basically the science of identifying these strokes and describing associated personality traits. [3] [5] [7]

Many organizations deny use of graphology due to the fact that according to 'Disability Discrimination Act 1995 recruitment based on graphology is a violation. Some due to confidentiality of clients and employer's organization deny use of services like graphology etc. [13].

[2] RELATED WORK

Graphologist analyze persons handwriting and study personality of writer. Some of the features they analyze include margins of the page, legibility and regularity, directions of lines called baseline, size and height of characters i.e. zones, inclination of letters called slants, pressure and speed of writing, ratio of upper, middle and lower zones, beginning and ending strokes, connections between characters, positions and shapes of certain alphabets like i and t etc. However, individuals handwriting may be consistent or distinct but should be unique to identified individuals from others. [8] [10] [11] [12] [13] [16][17][20]

[3] EXISTING METHODS

Currently professional graphologist/ handwriting expert's writer with piece of paper and their analysis is mostly based on their skill level. To do this kind of predictive analysis they mostly focus on common features like baseline, letters size, writers pressure, connected strokes, spacing between letters, words and lines, starting and ending strokes, speed, word slants, with of margins and many others. [2] [3] [5] [14] [15]][18][19]

[4] ASPECTS OF HANDWRITING BIOMETRICS

In this paper the major focus on three aspects viz. 1.Zones, 2. Baseline and 3. Pressure. All the sub aspects and their interpretations are listed and explained below:

Sr. No.	Features	Associated with	Characteristic/Traits Related with		
1	Upper Zone	Zone of abstract	1.	Realm of thoughts	
	[3]		2.	Dreams	
			3.	Hopes	
			4.	Plans	
			5.	Ideas and idols	
			6.	Open/ close/ narrow/ broad mindedness	
2	Middle Zone	Related to day to	1.	Here and now	
	[3]	day emotions	2.	Social interaction	
			3.	Behavior with people	
			4.	Extroverted/introverted	
			5.	Self confidence	
3	Lower Zone	Materialistic	1.	Food	
	[3]	Aspects	2.	Money	
			3.	Physical activity	
			4.	Basic instincts	

Table 1 Table of Zones

Zones: Like in the real world we live in three dimensions, height, width and depth. Similarly, in grahphology the dimension of height is studied using zones. This is vertical dimension which is further divided into three section upper zone, middle zone and lower zone. [1]

Considering timeline, upper zone represents upcoming or future, middle zone represents current state or present and lower zone represents the past. The head is considered as upper zone hence traits are related with intelligence, imagination, moral sciences etc. The body part is considered as middle zone expressing once ego, self-concious, social aspects etc. Lastly the lower part is considered lower zone which talks more about organic or physical needs of person in major.

1. Upper Zone – The Zone of Abstract

Graphology consider upper zone as the realm of mind as well as the guardian of spiritual aspirations. It is used to study intellect, imagination, illusion, fantasy and drive for power.

2. Middle Zone – Zone of day to day emotions

To express writer's self-image onto others, his/her role in life and thought about oneself middle zone is majorly studied. It major emphasis is on the quality and persistent efforts towards goals. Also, social and work-related activity showing the progress in surrounding environment.

3. Lower Zone – Zone of materialistic aspects

The zone basically focuses on all the materialistic aspect. They shows foody nature, money minded aspect, physical acitivities and basic instincts of person. Lower zone is used to study the balance between baseline and lower enlargement and its effect.

CURRENT ASPECTS IN HANDWRITING BIOMETRICS

Sr. No.	Features	Characteristics	Associated Traits	
1	Baseline	Straight Baseline	1.	Stable
		[3]	2.	Reliable
			3.	Even Temper
			4.	Self-confident
			5.	Do not get easily happy/depress
2	Baseline	Overly Straight	1.	Overly Control
		Baseline [3]	2.	Inflexibility
			3.	Defensive attitude
			4.	Lacks in self confidence
3	Baseline	True Ascending	1.	Enthusiastic person
		Baseline [3]	2.	Has faith in future
			3.	Full evolvement in everything
			4.	Don't easily get discourage
			5.	Eagerness for new things
4	Baseline	Descending	1.	Feeling of temporary weakness
		Baseline [3]	2.	Depression, discouragement, illness
			3.	State of grief
			4.	Emotional
			5.	Sceptics/Critics
5	Baseline	Wavy-Bouncy	1.	Have diffused objective
		Baseline [3]	2.	Not goal oriented
			3.	Moody, unsteady and inconsistent
			4.	Jack of all type
	Baseline	Erratic Baseline	1.	Casual
6		[3]	2.	Laugh and Cry very easily
			3.	Lot of unplanned work and drifters
			4.	Need constant attention
7	Baseline	Concave	1.	Excellent team players
		Baseline [3]	2.	Good crisis mangers
			3.	Self-motivated and self-directed
			4.	Strong fighting spirit
			5.	Change circumstances with hard work
8	Baseline	Convex Baseline	1.	Excellent enthusiastic people
		[3]	2.	Classic quitters
			3.	Never complete the work
			4.	Easily disappointed
			5.	Can't sustain work/pressure

Table 2 Table of Baseline

Baseline of a handwriting forms an imaginary line between middle and lower zone on which letters reside. Through the steadiness of baseline and balance of three zones personality is reveled. Baseline is considered as linear graph balancing ego and consious by materialistic drive.. Hence, it is also called as ego-adjustmenmt line. Thus baseline is used to study mood changes, moral and social balance, disposition and flexibility. .[1]

Some of the major baselines and studied and explained below:

1. Straight Baseline

The straight baseline indicates personality traits like stability, reliability and confident personality. Writer shows even temper and do not get easily depress or excited.

2. Overly Straight Baseline

The overly straight baseline is excessive effort baseline created with more efforts. This indicates writer have overly control and shows less flexibility. This shows build defensive attitude and lack of self confidence.

3. True Ascending Baseline

True Ascending baseline represents most positive traits. The writer is enthusiastic person with postive faith about future. Writer puts full envolvement in eveything does and don't get discourage due to small things.

4. Descending Baseline

Descending baseline shows negative traits on writer. The writer might feel weakness or might be depressed due illness, discouragement or environment. This baselines shows writers state of grief at that moment. Emotional effects and medical effects can be studied for more details.

5. Wavy-Bouncy Baseline

Most of the times writer changes goals or objective many times and does not show will or determination for its achievement. This baselines indicates witers moody nature, unsteadyness and inconsistent behavior. This writer migh be jack of all type.

6. Erratic Baseline

This baseline is important from psychological analysis point. This baseline shows writer's casual approach towards life. Writer shows emotional frugile nature and unplanned work. This writer might need constant attention and care.

7. Concave Baseline

Another postive baseline. This writer show positive traits like self motivation, good crisis mangment skills, excellent team player, strong fighting spirit and hard working nature to change circumstances postitively.

8. Convex Baseline

These writer shows excellent enthusiasm but lacks in hard work to complete the work. They give up their work in between hence this baseline is called as classic quitters. They get disappointment easily and unable to handle work pressure.

CURRENT ASPECTS IN HANDWRITING BIOMETRICS

Sr. No.	Features	Associated with	Characteristic/Traits Related with	
1	Pressure	Heavy [3]	1.	Deep intense personality
			2.	Creative power
			3.	Dominance
			4.	Enthusiasm & involvement
			5.	Endurance
			6	Deep commitments
2	Pressure	Medium [3]	1.	Good drive and determination
			2.	Energetic and resourceful
			3.	Shoulder their responsibility carefully
			4.	Dynamic, well inspired and motivated.
3	Pressure	Light [3]	1.	lack of driving power and intensity
			2.	Forgive easily
			3.	Spiritual
			4.	Soft spoken
			5.	Low physical drive
4	Pressure	Uneven	1.	Depression and anger
			2.	Sign of nervousness
			3.	Energetic but become tired very soon
			4.	Parental problem in childhood
			5.	Loses motivation very fast

Table 3 Table of Pressure

Pressure is body force excerted by muscle using pen on the paper. Pressure strokes changes everytime with upward and downword strokes. Higher the pressure shows physical fitness and mental energy of writer. As the pressure decreases the writer energy and intensity decreases. Diseases like parkinson can be identified by using pressure.

1. Heavy Pressure

This writer shows characteristics like deep intense personality, dominant nature, deep commitments, enthusiasm and hight involvement.

2. Medium Pressure

The is mostly found pressure overall. Writer with heavy pressure are driven and determined. They are energetic, resourceful, responsible, dynamic and motivated.

3. Light Pressure

Writer with low intensity and physical drive show light pressure. They forgive easily and are soft spoken. Some of the writer shows spirituality. Medical illness could also reduce the pressure.

[5] SUMMARY

Handwriting biometrics is still unexplored field. Various personality analysis techniques are being developed and many aspectes are being studied for their correlation and correctness. In

this we have majorly focused on current aspects in graphology namely zones, baselines and pressure. The details of aspects at their sub level are mentioned with personality traits and tried to cover major collabrative studies all together in this paper.

REFERENCES

- Omar Santana, Carlos M. Travieso, Jesus B. Alonso, Miguel A. Ferrer, "Writer Identification Based on Graphology Techniques", IEEE A&E SYSTEMS MAGAZINE, JUNE 2010
- [2] Carlos F. Romero, Carlos M. Travieso, Jesus B. Alonso and Miguel A. Ferrer, "Using off-line handwritten text for writer identification", WSEAS Transactions on Signal Processing, Issue I, Vol. 3, pp. 56-61, January 2007.
- [3] Prof. Seema Kedar, Ms. Vaishnavi Nair, Ms. Shweta Kulkarni, "Personality Identification through Handwriting Analysis", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 1, January 2015
- [4] D. John Antony, O. F. M. Cap, "Personality profile through handwriting analysis".
- [5] Champa H N, K R Ananda Kumar, "Artificial neural network for human behavior prediction through handwriting analysis". International Journal of Computer Applications, Vol. 2, May 2010, pp. 36-41
- [6] Sheikholeslami, S. N. Srihari, V. Govindaraju, "Computer aided graphology Center of Excellence for Document Analysis and Recognition" (New York).
- [7] Champa H. N. and K. R. Ananda Kumar, 2010, "Automated Human Behavior Prediction through Handwriting Analysis", First International Conference on Integrated Intelligent Computing.
- [8] Abdul Rahiman M, Diana Varghese, Manoj Kumar G, "HABIT- Handwriting Analysis Based Individualistic Traits Prediction".
- [9] Parmeet Kaur Garewal, Deepak Prashar, "Behavior Prediction through Handwriting Analysis".
- [10] Vlad Atanasiu, "Surveys and projects in handwriting analysis", MIT Postdoctoral fellow, Department of Brain and Cognitive Sciences 77, Massachusetts Av., Cambridge MA 02139, USA.
- [11] Bin Zhang Sargur N. Srihari, "Analysis of Handwriting Individuality Using Word Features", State University of New York at Buffalo, Buffalo, NY 14228.
- [12] Reâ Jean Plamondon, Sargur N. Srihari," On-Line And Off-Line Handwriting Recognition: A Comprehensive Survey", Ieee Transactions On Pattern Analysis And Machine Intelligence, Vol. 22, No. 1, January 2000.
- [13] Dharam Veer Sharma, Geeta Saini, Mohit Joshi, "Statistical Feature Extraction Methods for Isolated Handwritten Gurumukhi Script", Vol. 2, Issue 4, July-August 2012, pp.380-384.
- [14] Vikram Kamath, Nikhil Ramaswamy, P. Navin Karanth, Vijay Desai And S. M. Kulkarni, "Development Of An Automated Handwriting Analysis System", Vol. 6, No. 9, September 2011
- [15] Imran Siddiqi Chawki Djeddi Ahsen Raza •Labiba Souici-meslati, "Automatic analysis of handwriting for gender classification", Pattern Analysis and Applications • May 2014
- [16] Djeddi C, Siddiqi I, Souici-Meslati L, Ennaji A (2013) Codebook for writer characterization: A vocabulary of patterns or a mere representation space? In: Proceedings of the 12th Int'l Conference on Document Analysis and Recognition
- [17] HuangD,ShanC,ArdabilianM,WangY,ChenL(2011) Local binary patterns and its application to facial image analysis: a survey. IEEE Trans Systems Man Cybern Part C Appl Reviews 41(6):765–781
- [18] Liwicki M, Schlapbach A, Bunke H (2011) Automatic gender detection using on-line and offline information. Pattern Anal Appl 14(1):87–92
- [19] Siddiqi I, Vincent N (2010) Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. Pattern Recognit 43(11):3853– 3865
- [20] Sokic E, Salihbegovic A, Ahic-Djokic M (2012) Analysis of offline handwritten text samples of different gender using shape descriptors. In: Proceedings of IX International Symposium on Telecommunications (BIHTEL), pp 1–6

(An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861 Volume XI, Issue I, Jan-June 2018



REVIEW OF MACHINE LEARNING ALGORITHMS FOR IT OPERATIONS

Dr. Santosh Deshpande¹, Rahul Gaikwad²

¹Director, MES'S Institute of Management and Career Courses Pune, India ²Sr. DevOps Engineer, FireEye and Research scholar, Savitribai Phule Pune University, Pune, India

ABSTRACT

In IT operations, system alerts and failures are very difficult to predict but it's becoming crucial for business. If we are able to predict alerts or system failures it would really help to prevent unexpected and critical system downtimes also it would also assures reliability and high availability of business services for end users. Depends on system architecture , there are several types of logs generates like Application logs, database logs, network logs, server and business process logs etc. These logs are footprint of any activity performed by the user or any internal system activity. Logs includes critical and huge amount of information. In most of production environment, huge amount of logs produced per second. As human being , it's not feasible to read these logs line by line . To resolve any system alerts or failures , there is strong need of a mechanism which can go analyse all these logs, events and alerts and provide more insights. By applying Machine learning algorithms we can automatically analyze and process these logs and alerts to get more insights which helps to understand and reduce the system downtimes. This paper focuses on analysis of different machine learning algorithms for IT system logs and alerts with their proposed outcomes in detail.

Keywords -Log Analysis, Failure Prediction, Text Mining, Machine Learning Algorithms

[1] INTRODUCTION

IT operations teams always need to focus on infrastructure (on-premise, virtual/cloud) monitoring, such as network, server, database, application, devices and storage. IT application systems gets affected by many factors such as network dis-connectivity, software application bugs or administrator mistakes. To diagnose these production failures we may need to reproduce them in other non-production environments which is very difficult because of various reasons like difficulty of building same environment setup, time constraint, unavailability of user's input etc. So to diagnose these production alerts or failures, system admins or programmer need to be depend upon different system logs [1]. Typical logs may have fields like IP Address, Timestamp, time zone, http method, client request, protocol, request status code, response size, request execution time or custom fields.

Basically log analysis should focus on major operations problems like forensic analysis, fault detection, system alert or failure prediction and remediation recommendation. To find out the root cause of system failure, need to perform post-analysis of system logs known as

(An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861 Volume XI, Issue I, Jan-June 2018

Forensic analysis. To quickly detect the symptoms of critical failures when they appear like anomaly detection is known as Fault detection. To analyse and predict early sign of potential system alert or failure is known as failure prediction and it is a proactive approach [2]. Suggesting healing actions, based on alert history and performed manual steps, is known as remediation recommendation.

Machine Learning algorithms can play a vital role to reduce complexity and minimizes manual efforts in these major IT operations. Also we can leverage Machine Learning algorithms for better system health monitoring, to reduce alert noise and operations cost, to reduce mean Time to detect (MTTD) and faster mean time to recovery (MTTR).

[2] RELATED WORK

Log-based System Administration:

System logs keep records of an IT system's operational states, events and alerts over the time. For better system administration, log analytics should focused on system post-analysis (forensic), system failure or fault detection, and system failure prediction.[2]

Log clustering and pattern learning:

Log clustering can be done using supervised and unsupervised machine learning algorithms. In Supervised machine learning method, firstly users need to manually label a set of log patterns or categories and use classifier models such as Naive Bayes to perform log text classification [8]. Unsupervised machine learning can be use for log clustering, like hierarchical partitioning method [9] and multi-pass data summarization process [10]. After obtaining cluster structures from the system logs, we can extract the common patterns or features from multiple system logs from the cluster. [2]

Sequence modelling with deep neural networks:

Recurrent Neural Networks (RNN) such as Long Short-Term Memory (LSTM) are powerful models that are capable of learning effective feature representations of sequences when given enough training data.

[3] Analysis of Machine Learning Algorithms for IT Operations

Failure prediction:

A general log pattern can be extracted by using clustering method based on matching features, identical format and content details. Apply Term frequency - inverse document frequency (TFIDF) algorithm to consider each pattern as a word and the set of patterns in each discretized epoch as a document. As system failures occur rarely so Long Short-Term Memory (LSTM) a type of recurrent neural network(RNN) can apply to deal with the "rarity" of labelled data in the model training process.[2]

Alert Classification

The labelled alerts can be use for training the automatic classifier which uses RIPPER algorithm for learning the classification rules. This automatic alert classifier reduces the system admin's workload. Alert classifier model classifies the alerts as true and false positive alerts.

(An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861 Volume XI, Issue I, Jan-June 2018

The alerts which classified as false positive by the system admin can be considered for training purpose. For the Alert classification, we can implement and compare different classifier algorithms like Random Forest, Decision Stump, RIPPER, NNge, oneR, PART. [3]

Detection of System Failure Time

The timestamp and the features are exists in the system logs within timeframe. As per these features, using the clustering algorithm we can obtain the cluster of the corresponding time and identify the fault timeframe according to the cluster. As a classifier for Detection of System Failure Time, K-means Clustering can be implement with tuning of K element. [4]

Fault Location

Once fault time detected , we can refer corresponding logs for further analysis with respect to timestamp. TFIDF (Term Frequency-Inverse Document Frequency) algorithm is a commonly use to extract keywords from inputs and weighting technique for information retrieval and mining [5]. Term frequency consider the number of occurrences of a targeted word in the input file. To prevent from biasing towards long files , these number of occurrence is normalized.[4]

Alert or Alarm handling and Noise reduction :

The root cause of any system alarm or alert is the reason for which it occurs. Most of the time system configuration problems is the root cause and does not resolve unless system admin fixes them. Alarm clustering algorithm that groups similar system alarms together and consider them as a single generalized alarm. We can drastically reduce the number of newly generating system alarms , by removing these root causes.

Alert or Alarm Correlation

Manual system alarm correlation is difficult and very time consuming so by using Machine learning algorithms like clustering algorithms we can build Alarm correlation systems (ACSs) where we identify the correlation between alarms. [13]

Anomaly detection

Anomaly detection is the process of identifying unexpected items or events in a given datasets. Clustering and anomaly detection algorithms can be used to detect and flag statistical outliers or anomaly that can be an indicators of a problem. In

- *Supervised Anomaly Detection* The training and test data sets are fully labelled. Algorithms like decision trees, Support Vector Machines (SVM) or Artificial Neural Networks (ANN) algorithms can be used for Supervised Anomaly Detection.
- *Semi-supervised Anomaly Detection* Only training data consists of normal data without any anomalies but test data can have anomalies. Well known algorithms such as One-class SVMs and autoencoders can be used for Semi-supervised Anomaly Detection.
- *Unsupervised Anomaly Detection* This method does not need any labels for training data. Also there is no separation between training and test datasets.

Statistical Performance analysis

(An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861 Volume XI, Issue I, Jan-June 2018

Statistical analysis algorithms can be used to discover performance trends and predict future behaviour. The researcher tried to predict the future system load conditions of a resources by considering different measures obtained from the load monitoring systems of servers. These raw data is extremely variable and dynamic which makes it difficult to forecast the behaviour of future resource measures and deduce a clear pattern or trend about the system load behaviour of a resource. We can apply load prediction algorithm based on linear regression model to predict system load trends. [18]

Bug-tracking

Bug-Tracking is very crucial part in any software development and IT industry which assures quality of products. Automatically tracking bug can be consider as classification problem, which takes the bug no or title and bug description as the input for processing and mapping it to any available developers (using labels). The major difficult is that the bug description is in text format which usually includes a combination of free unstructured strings, code snippets, which makes input data highly noisy and hard to process and analyse. The bag-of-words (BOW) model does not consider the syntactical and sequential word information available in the unstructured text. In recent research , they proposed bug report representation algorithm using an attention based deep bidirectional recurrent neural network (DBRNN-A) model which learns a syntactic and semantic feature from long word sequences in an unsupervised manner. Instead of just BOW features, the DBRNN-A based bug representation is then used for training the classifier. Using an attention mechanism enables the model to learn the context representation over a long word sequence, as in a bug report. [15]

Bug Assignment

In Quality Assurance or Software Testing, for effective bug resolution, it is very important to assign the reported bug to a respective developer or engineer. Bug assignment is an initial part of bug tracking whose objective is to assign a respective developer or engineer to the reported bug. The assigned developer or engineer can perform various checks, troubleshoots the issue and do the changes in the source code to resolve the reported issue. The selection and assignment of a respective developer for specific bug is a challenging, time and cost consuming process in the project. From recent research, we can leverage multiple Machine Learning algorithms such as Naïve Bayes, Support Vector Machine (SVM), C4.5, Expectation Maximization, Conjunctive Rules and the Nearest Neighbour (NN) algorithm for the selection and assignment of developer or engineer resources to specific bugs. [16]

[5] SUMMARY

In this paper, we analyse various machine learning algorithms and their application in a complex and critical IT systems based on logs, events and alerts. Recent research work focuses on anomaly detection in running system, alert or failure prediction, finding root cause of system failure etc. using different system logs. Also system logs can be leverage to understand the system behaviour, common features, correlation between events and recommendations for remediations based on historical resolutions. There is need of Domain based Algorithms that leverages IT Operation domain expertise (specific to any environment) to intelligently analyse, process, interpret and implement the rules , patterns and models , as directed by an organization's domain specific data and its expected outcomes. These algorithms should target IT Operations specific goals like eliminating alert or alarm noise, correlating unstructured log

(An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861 Volume XI, Issue I, Jan-June 2018

or event data, setup baselines or thresholds, alarming on abnormalities , try to identify possible root causes and predict remediations.

REFERENCES

[1] Ding Yuan, Haohui Mai, Weiwei Xiong, Lin Tan, Yuanyuan Zhou, Shankar Pasupathy "SherLog: Error Diagnosis by Connecting Clues from Run-time Logs", Association for Computing Machinery, New York, NY, USA 2010

[2] Ke Zhang , Jianwu Xu , Martin Renqiang Min , Guofei Jiang , Konstantinos Pelechrinis and Hui Zhang "Automated IT System Failure Prediction: A Deep Learning Approach", 2016 IEEE International Conference on Big Data (Big Data)

[3] Dr T Subbulakshmi, S. Mercy Shalinie "Real Time Classification and Clustering Of IDS Alerts Using Machine Learning Algorithms", International Journal of Artificial Intelligence & Applications (IJAIA), Vol. 1, No.1, January 2010

[4] Leyi Zhang, Lei Fan and Naiwang Guo ," Log-Based OpenStack Fault Diagnosis by Machine Learning". Journal of Physics: Conference Series, Volume 1069, conference1.

[5] Chum O, Philbin J, Zisserman A. "Near Duplicate Image Detection: min-Hash and TFIDF Weighting [J]". 2008.

[6] Ruben Sipos, Cornel U. Dmitriy Fradkin, Siemens Fabian Moerchen, Amazon Zhuang (John) Wang, Skytree, "Log-based Predictive Maintenance", ACM New York, NY, USA KDD (Knowledge Discovery and Data Mining)

[7] Li, T., Liang, F., Ma, S., Peng, W.: "An integrated framework on mining logs files for computing system management." ACM New York, NY, USA KDD (Knowledge Discovery and Data Mining) (2005)

[8] T. Li, F. Liang, S. Ma, and W. Peng, "An integrated framework on mining logs files for computing system management," in Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining. ACM, 2005, pp. 776–781

[9] A. A. Makanju, A. N. Zincir-Heywood, and E. E. Milios, "Clustering event logs using iterative partitioning," in Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2009, pp. 1255–1264.

[10] R. Vaarandi et al., "A data clustering algorithm for mining patterns from event logs," in Proceedings of the 2003 IEEE Workshop on IP Operations and Management (IPOM), 2003, pp. 119–126.

[11] Klaus Julisch, IBM Research, Zurich Research Laboratory, Rüschlikon, Switzerland "Clustering intrusion detection alarms to support root cause analysis", ACM New York, NY, USA, ISSN: 1094-9224 EISSN: 1557-7406

[12] Klaus Julisch IBM Research, "Mining Alarm Clusters to Improve Alarm Handling Efficiency", IEEE Seventeenth Annual Computer Security Applications Conference

(An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861 Volume XI, Issue I, Jan-June 2018

[13] Klaus Julisch IBM Research, Zurich Research Laboratory, Rüschlikon, Switzerland, "Clustering intrusion detection alarms to support root cause analysis", ACM Transactions on Information and System Security (TISSEC)

[14] Markus Goldstein, Seiichi Uchida, "A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data", PLOS ONE DOI:10.1371/journal.pone.0152173

[15] Senthil Mani, Anush Sankaran, Rahul Aralikatte, "DeepTriage: Exploring the Effectiveness of Deep Learning for Bug Triaging" arXiv:1801.01275v1

[16] Anjali Goyal, Neetu Sardana, "Machine Learning or Information Retrieval Techniques for Bug Triaging: Which is Better?" e-Informatica Software Engineering Journal, Volume 11, Issue 1, 2017, pages: 117–141, DOI 10.5277/e-Inf170106

[17] Mario MILICEVIC, Mirta BARANOVIC, Krunoslav ZUBRINIC, "Application of Machine Learning Algorithms for the Query Performance Prediction" Advances in Electrical and Computer Engineering (AECE)

International Journal of Computer Engineering and Applications, Volume XII, Special Issue, May 18, www.ijcea.com ISSN 2321-3469



APPLICATION OF NATURAL LANGUAGE INTERFACE TO DEVELOP INTELLIGENT INFORMATION RESOURCE CENTER (GOVERNMENT SCHEMES)

Ravindra P. Vaidya¹, Manish R. Joshi²

¹ Associate Professor, Department of Computer, MES's Institute of Management and Career Courses (IMCC), Pune, INDIA

² Associate Professor, School of Computer Sciences, North Maharashtra University, Jalgaon. INDIA

ABSTRACT:

Government does come up with various innovative schemes with the motive of betterment of the needy people. But these all schemes prove to be non-effective as the target mass does not aware of the schemes. Sometimes, it is found that, the responsible authorities are reluctant to provide complete and correct information regarding the scheme to those for whom the schemes are intended to. Efforts should be made to evolve the system so that any common person can interact with the computer in non-formal manner without having any predetermined constrains on communication to get knowledge of all governmental beneficiary schemes, which are suitable for him. We are trying to develop a system where computer will act as an 'Information Expert', who will not only provide correct and complete information about various schemes to the user by considering educational, social background and need but also guides in selecting the best scheme

Keywords: Natural Language Interface, Human Computer Interaction, Information Resource Center, Computational behaviorism, knowledge base.

[1] INTRODUCTION

With recent advances in computer and Internet technology, people have access to more information than ever before. As the amount of information grows, so does the problem of finding what one is looking for. The quality of providing significant and appropriate

APPLICATION OF NATURAL LANGUAGE INTERFACE TO DEVELOP INTELLIGENT INFORMATION RESOURCE CENTER (GOVERNMENT SCHEMES)

information can be considered as an important measure of productivity and prosperity of nation. Social, political and economic changes and reforms are possible only after availability of information to common person. Organization and management of resources, their utilization up to the optimum level and the techniques of utilizing those resources, all these measures depends upon the sufficiency and efficiency of the people of the country. The development of common person ultimately leads to the development of nation thus depends on up to what extend the information is percolated to the last levels in the society.

India is one of the developing countries and one of the major democracies in the World. For the betterment of the people, the government of India asserts various schemes. Though Government is putting all its effort to pass through complete and correct information about these schemes to the target mass, it is disappointing fact that, the grants allotted for the beneficiary schemes wipe up because of its non-utilization.

[2] PROBLEM:

Only ill literate and ignorant common people are withhold from getting proper and useful information is not true. Even well educated people may face problems while acquiring significant and adequate information.

Following factors contribute a lot in ceasing the proliferation of information.

- 1. **Information Hiding**: In many cases the authorities hides the information about various schemes either intentionally or un–intentionally.
- 2. **Communication Barrier**: Common person always afraid to directly approach to the authorities. Authorities very rarely speak with common person in informal manner. They always try to impress common people by using technical words and official jargon. These communication barriers prevent propagation of information.
- 3. **Reluctantness of Authorities**: Authorities are not enthusiastic enough nor they show any interest in passing on the information.

[3] INFORMATION RESOURCE CENTER:

Information Resource Center depicts the idea of one centralized center where large volume of information is stored in systematic and efficient manner. The information at the center is so well organized that appropriate information can be extracted out effectively without much delay. Any common man can provide his/her information containing educational, economic and social background in very non-formal way and the center after taking into consideration all these factors will suggest appropriate schemes and guide for selection. To make this Information Resource Center more effective one has to add the flavors of technological advancements into it. Wide accessibility of information to common public is the major thrust of such type of application, which has following additional characteristics-

International Journal of Computer Engineering and Applications, Volume XII, Special Issue, May 18, www.ijcea.com ISSN 2321-3469

- 1. The number of users interacting with the system is large
- 2. The number of queries posed to the system is extensive.
- 3. There are large variations in the queries posed to the system.

[4] PROPOSED TECHNOLOGY FOR ACCESSING INFORMATION:

The standard user interfaces we are forced to use when working with computers are rapidly becoming inadequate in many situations. Sometimes they are cumbersome and inconvenient; in others, they require skill, which may cause hesitation on the user's side.

[4.1] NATURAL LANGUAGE AS AN INTERFACE STYLE

We believe that the most natural form of communication and information access for humans is natural language. We propose to address the information access problem at the center with a uniform natural language interface. Natural language presents a viable choice as an Interface Style that can be used in the dialog between a human user and a computer. The appeal of Natural Language Interfaces is because one is able to address a machine and direct its operations by using the same language that we use in everyday human-to-human interaction.

[4.2] MOTIVATION:

In December 1993, START (SynTactic Analysis using Reversible Transformations) became the first natural language system available for question answering on the World Wide Web. Since then START has been involved in dialogs with users all over the world, answering millions of questions. In response to a question, START can dispatch the user to a weather forecast page, a map collection, a World Fact book database, a personal homepage, etc., all accessed through a uniform natural language interface.

v-Rep is the online virtual Representative, representing the company for a particular product and is trained to provide proper information to the user in response to the query posed by the customer (www.nativemainds.com). The system CHAT (Conversational Hypertext Access Technology) also gives insight to the Natural Language Interface based interactive system.

[4.3] HUMAN-COMPUTER INTERACTION PERSPECTIVE

Following points must be considered while developing effective Human-Computer Interaction



Ravindra P. Vaidya and Manish R. Joshi

APPLICATION OF NATURAL LANGUAGE INTERFACE TO DEVELOP INTELLIGENT INFORMATION RESOURCE CENTER (GOVERNMENT SCHEMES)

Interfaces be made easier?

An unrestricted natural language interface is generally considered an enticing prospect because, if it could be implemented, it would offer many advantages: it would be easy to learn and easy to remember, because its structure and vocabulary are already familiar to the user; because the same language could be used for many application, there might be fewer transfer problems between applications; they are particularly powerful because of the multitude of ways in which to accomplish an action; and they also allow considerable flexibility in executing the steps of a task.

Unfortunately, natural language is often ambiguous and is dependent on a great deal of world knowledge. In order to implement a working natural language system one must usually restrict it to cover only a limited subset of the vocabulary and syntax of a full natural language. This allows ambiguity to be reduced and processing time to be kept within reasonable bounds.

[4.4]INTERFACE TO INTELLIGENT INFORMATION RESOURCE CENTER:

Following figure depicts the general structure of an interface.



Figure 1 : General Structure of Interface

Interface will allow user to communicate with the system in his own style without imposing formal language constraints on him. User may ask any query related to the Government scheme domain like, "How can I start my own business after securing first class in Graduation?" which interface will forward in a proper manner to the computer where all information is properly stored in the Knowledge-base. Based on the query, appropriate answer will again be pass on to the user through an interface.

[4.5] METHODOLOGY:

A different approach to language understanding will be followed which addresses Natural Language interaction as a behavioral problem rather than a linguistic one. Due to this fundamental change in approach, the essential question is no longer "How does language work?" but rather, "What do people say?" More specifically, it can be put, as "What are the

International Journal of Computer Engineering and Applications, Volume XII, Special Issue, May 18, www.ijcea.com ISSN 2321-3469

invariant parts of verbal behavior (i.e. questions) that form a specific operant (i.e. that lead to a specific answer)

In the same way, a person who wants to know about Prime Minister Rojgar Yojana may ask a great variety of different questions, such as, "How to get loan under PMRY scheme?", "Can any graduate is eligible to take benefits of PMRY Scheme?", or "Must one has his own land to be get benefited by PMRY scheme?". All of these questions, and many others that could be generated, have the common characteristic that they include the string " PMRY scheme" and they do not include the strings, "Social Relief Fund" or "National Talent Search". We can study this regularity as a measurable phenomenon and we can use it to predict which answer is likely to be appropriate to a question.

Using this approach of 'computational behaviorism' our interface will have the structure as shown in following figure.



The important issue, then, is not just to find the invariant parts of linguistic behaviors, but also to map them properly after encoding them. Separating out the invariant part (Tokenization) can be carried out using Link Grammar Parser, which identifies noun phrase, verb phrase and other grammatical components from the given sentence. A proper data structure which should be sufficiently general to hold a large variety of different behaviors as well as sufficiently powerful to allow meaningful operations on those behaviors should be selected so as to carry out further mapping and appropriate association.

[5] CONCLUSION

Intelligent Information Resource Center can be developed which will work as 'Information Expert'. Natural Language Interface to this System can be developed using the approach of 'Computational Behaviorism'. The system Conversational Hypertext Access

APPLICATION OF NATURAL LANGUAGE INTERFACE TO DEVELOP INTELLIGENT INFORMATION RESOURCE CENTER (GOVERNMENT SCHEMES)

Technology (CHAT) and Virtual Representative (v-Rep) are the real life examples, which can guide us in this development.

REFERENCES

- [1] Byron Long (May-1994), "Natural Language as an Interface Style"
- [2] Dan Patterson, Introduction to Artificial Intelligence & Expert Systems.
- [3] Elaine Rich, Kevin Knight. Artificial Intelligence Second Edition.
- [4] Kaula P.N., Information and Communication Technology impact and challenges, University News, Vol.35 No.35
- [5] Mahajan, S. L. (2002) Information Communication Technology in distance Education in India and Commonwealth Countries: A Challenge,
- [6] Liddy, Elizabeth D. Date of Access: 1 October 1999. "Enhanced Text Retrieval Using Natural Language Processing."
- [7] Thomas Whalen (April 1996). Computational Behaviourism Applied To Natural Language
- [8] Yagananarayana. Introduction to Neural Networks.
- [9] http:// www.aaai.org
- [10] http:// www.nativeminds.com
- [11] http://www.dfki.de/fluids/Intelligent_User_Interfaces.html
- [12] http://www.ai.mit.edu



SIGNIFICANCE OF INFORMATION TECHNOLOGY IN HR DEPARTMENT

Dr Manasi Bhate

Computer Department, MES IMCC, Savitrihai Phule Pune University, Pune, India

ABSTRACT:

Abstract—Information technology deals with technology and other aspects of managing and processing information, especially in large organizations. Human Resource Development is the framework for helping employees to develop their personal and organizational skills, knowledge, and abilities. For analyzing the Implementation of IT in HR Department, we have conducted a survey with the help of questionnaire considering various objectives. Primary Data was collected from eight Information Technology organizations. After analysis of data, results were presented in the paper.

Keywords: Information Technology (IT), Human Resource (HR)

[1] INTRODUCTION

"Behind Every Successful organization are its Resources." The main important resource of the organization is "HUMAN". An Organizations' success hinges more on 'high touch' than 'high tech'. [2][3]Human Resource Development includes opportunities such as employee training, employee career development, performance management and development, coaching, succession planning, key employee identification, tuition assistance, and ultimately Harnessing Productivity of Organization.

IT deals with the use of electronic computers and computer software to convert, store, protect, process, transmit, and retrieve information. Over the past 20 years, its prevalence has dramatically increased so that it is now a part of nearly every aspect of daily life.

Information Technology (IT) Organizations — These are the Organizations which derive their revenue from software Development. [1] Basically tasks handled by IT companies are Custom Application Development and Maintenance (CADM), System Integration, IT Consulting, Application Management, Infrastructure Management Services, Software testing, Service-oriented architecture etc.

[2] NEED FOR IMPLEMENTATION OF IT IN HR DEPARTMENT

SIGNIFICANCE OF INFORMATION TECHNOLOGY IN HR DEPARTMENT

New Internet-enabled HR enterprise systems pledge to go well beyond empowering employees through information and self-service, while freeing the HR staff of many routine tasks in favor of focusing on true organizational capability, learning, effectiveness and performance.

Following are the reasons of using IT in HR

• HR needs robust systems for its strategic roles - In a knowledge-based economy, HR's roles — in recruiting, retention, employee service, and talent and knowledge management — grow ever more strategic.

• HR systems are all over the place which is possible in execution with use of IT.

• HR and IT need to team up to Maximize the business value and leverage of HR systems rests upon a close and productive relationship between HR and IT.

You need to build a compelling employment "brand".

• With the possibility of opportunities being made accessible more equitably for an increasingly large population in all corners of the world in the shortest possible time, IT has demonstrated high significance for economic development.

• It is important to bring together all relevant information at a single site in order to make it easily accessible to the workforce.

• The information stored in the HR system can and should be used for the planning and simulation of different scenarios.

[3] OBJECTIVE OF THE RESEARCH

- To assign responsibility for common HR tasks (e.g. recruitments, leave approvals) throughout the organization via a customized HR package.
- To make assessment of limitations if any of deployment of Information Technology in HR Department.

Sample Size - Data was collected from eight different IT organizations for above objectives through Questionnaire.

[4] DATA ANALYSIS

To study the objective, Data was collected and converted in to tabular form [4] Table 1: Standard Packages used by HR Departments

Company name	Answe r	If yes any specification	Scope	Hardwa re used	Software used
Company I	Yes	-	-	Standar	Standard
Company 2	Yes	Up to certain degree	Limited scope	-	HR align, Pay pack

Company 3	3 Yes	-	Payroll. Recruitment, Employee appraisal, Training	-	-
Company 4	Yes	-	Payroll, Recruitment, Employee appraisal	-	Resume database, People soft
Company 5	No	-	-	-	-
Company 6	Yes	Semi-automated	Payroll, Recruitment, Employee appraisal		People soft
Company 7	No	-		-	
Company 8	No	-	-	_	

Questionnaire consists of following points

- customization done to the standard software
- Additions to standard package
- Criteria in selection of customized HR package
- Criteria Details
- Training for using customized HR package
- Software tools used instead of customized HR package

Analysis of primary data is represented in Fig I

X- Axis consists of following:

- Use of standard package by HR.
- Customization done to standard software
- Typical Functions used by HR in recruiting(Aptitude, GD, PI etc)
- Addition in standard package.
- Criteria's used for selection of customized package
- Requirement of training to use customized HR Package
- Use of software tools instead of customized HR Package

Dr. Manasi Bhate



Y- Axis consists of following:

For studying limitations in implementation of IT in HR department following points were considered in the Questionnaire

- Problems in using automated system
- Degree of maintenance
- Implementing new techniques for advancement of organization
- On line HR Department
- IT readiness Time required for employee
- Disk space the problem



Fig 2 : Employability issues due to IT Implementation (1 = No, 2 = Yes)



Fig 3 : Disk space Problem (1 = No, 2 = Yes)

Analysis : Data was collected and analyzed for security measures in HR Department. Analysis is done by presenting the facts in the table.

Company Name	Answer	
Company I	Access control logs	
Company 2	Many, cannot be disclosed	
Company 3	-	
Company 4	Security enabled sheet, minimal sharing of confidential data people using laptop instead of desktop	
Company 5	-	
Company 6	Password protected and access control	
Company 7	-	
Company 8	-	

TABLE II: SECURITY MEASURES

Analysis - 50% companies take security measures for data in various ways in HR department.

[6] CONCLUSION

- Information Technology is used in majority of organizations for completing certain tasks of Human Resource Department. It is mostly applied in recruitment and training.
- 30% to 40%(Approximately 37.5%)company's recruitment process in HR department is online to save human resource, time and money.
- 20% to 30% Approximately 25%) companies have problems in their automated system.

Dr. Manasi Bhate



MANET PERFORMANCE ANALYSIS USING DIFFERENT DISTRIBUTION LOAD Dr. Ravikant Zirimite, Dr. Manasi Bhate²

¹Associate Professor, MES'S Institute of Management and Career Courses Pune, India ¹Associate Professor, MES'S Institute of Management and Career Courses Pune, India Affiliated by Savitribai Phule Pune University, Pune, India

ABSTRACT

The paper presents the results of a detailed packet-level simulation comparing four multi-hop wireless ad hoc network routing protocols under the load of different probability distributions, that cover a range of design choices having different protocol viz. DSR, AODV and OLSR. We have extended the OPNET network simulator to accurately model the MAC and physical-layer behavior of the IEEE 802.11 wireless LAN standard, including a realistic wireless transmission channel model. Simulation of 60 mobile nodes has been carried out and the performance optimization is determined.

Keywords - *Simulation, Opnet, Wireless, Statistical probability distribution, IEEE802.11, throughput, delay, retransmission attempt, load, protocol, MAC, LAN.*

[1] INTRODUCTION

Ad-hoc wireless network is that network where no communication is present, in such network; each mobile node operates not only as a host but also as router. Mobile nodes in the network may not be with in range of each other, communication of these nodes perform by discovering "multi-hop" paths through the network to other nodes. This type of network is some time called infrastructure less network [1]. Some examples of the possible uses of ad hoc networking are students using laptop computers to participate in an interactive lecture, business associates sharing information during a meeting, soldiers relaying information for situational awareness on the battlefield [2, 3]. Many different protocols have been proposed to solve the multi hop routing problem in ad hoc networks, each based on different assumptions and intuitions.

Mobile Ad hoc Networks (MANETs)[1] are an emerging technology that allows establishing an instant communication network for civilian and military applications, without relying on pre-existing fixed network infrastructure. The nodes in a MANET can dynamically join and leave the network, frequently, often without warming, and possibly with out disruption to other nodes' communication. Each node in the network also acts as router, forwarding data packet for other nodes. A central challenge in design of Ad hoc network is the development of dynamic routing protocols that can effectively find the route between two communicating nodes. The routing protocol must be able to keep up with the high degree of node mobility that often changes the topology drastically and unpredictably.

The current Mobile Ad Hoc Network (MANET) [2] paradigm as described by the Internet Engineering Task Force (IETF) MANET work group. Routing algorithms are often difficult to

formalize into mathematics; they are instead tested using extensive simulation. A large amount of work has been done in the area of energy efficient routing. This approach attempts to maximize network lifetime by routing through paths, which use the least amount of energy relative to each node. Now a day, more attention has been given to use specific network parameters while specifying routing matrixes. Routing matrixes includes delay of network, link capacity, link stability or identifying low mobility nodes. These schemes are generally based on previous work, which is then enhanced with the new matrix.

The paper is providing a realistic, quantitative analysis comparing the performance of a variety of multi-hop wireless ad hoc network routing protocols. We present results of detailed simulations showing the relative performance of three recently proposed ad hoc routing protocols: DSR [3, 4, 5], AODV [6], OLSR [8]

Our results in this paper are based on simulations of an ad hoc network of 60 wireless mobile nodes moving about and communicating with each other. We analyze the performance of each protocol and explain the design choices that account for their performance.

The section 2 of the paper describes the different types of protocols used in the simulation. The section 3 has given description of different types of statistical probability distribution used for arrival and departure of packets in simulation. The performance analysis is describes in section 4. The section 5 has summaries conclusion of the paper.

[2] DESCRIPTION OF THE PROTOCOLS

2.1 Dynamic Source Routing (DSR) [3, 4, 5]

The DSR is using source routing, i.e. the sender knows the complete hop-by-hop route to the destination. When node sending data packet to the destination, DSR has use route discovery by flooding the network with route request (RREQ) packets. Each node receiving an RREQ, node will rebroadcast it, unless it is the destination or it has a route to the destination in its route cache. Such node replies to the RREQ with a route replay (RREP) packet that is routed back to the original source. If any link on a source route is broken, the source node is notified using a route error (RERR) packet. The source will initialize new discovery process. No special mechanism is required to detect routing loops. Also, any forwarding node caches the source route in a packet it forwards for possible future use. Several additional optimizations have been proposed and have been evaluated to be very effective by the authors of the protocol [5], as described in the following:

• Salvaging: An intermediate node can use an alternate route from its own cache when a data packet meets a failed link on its source route.

• Gratuitous route repair: A source node receiving an RERR packet piggybacks the RERR in the following RREQ. This helps clean up the caches of other nodes in the network that may have the failed link in one of the cached source routes.

• Promiscuous listening: When a node overhears a packet not addressed to itself, it checks whether the packet could be routed via itself to gain a shorter route. If so, the node sends a gratuitous RREP to the source of the route with this new, better route. Aside from this, promiscuous listening helps a node to learn different routes without directly participating in the routing process.

2.2 Ad Hoc on demand Vector (AODV) [6]

AODV discovers routes on demand basis. It uses routing table to maintain routing information, one entry per destination. RREP packet is used to replies back to the source and, subsequently, to route data packets to the destination. AODV uses sequence numbers to maintain at each destination to determine routing information and to prevent routing loops [6]. AODV working on timer- based states in each node. A routing table entry is expired if not used recently. If node link is broken, the all predecessor nodes forward the RERR packets, to effectively erasing all routes using broken link. AODV uses expanding ring search technique initially to discover routes to an unknown destination. AODV algorithm has the ability to quickly adapt to dynamic link conditions with low processing and memory overhead. AODV offers low network utilization and uses destination sequence number to ensure loop freedom AODV keeps the following information with each route table entry.

(i) Destination IP address (IP address for the destination node),

- (ii) Destination sequence number,
- (iii) Valid destination sequence number flag,
- (iv) Network interface,

(v) Hop count, that is, number of hops required to reach the destination,

(vi) Next hop (the next valid node that did not re broadcast the RREQ message),

(vii) List of precursor,

(viii) Life time, that is, expiration or deletion time of a route.

2.3 Optimized Link State Routing (OLSR) [8]

The OLSR model implements the MPR (Multi Point Relay) flooding mechanism to broadcast and flood Topology Control (TC) messages in the network. The algorithm is implemented as suggested in OLSR RFC 3626. This mechanism takes advantage of controlled flooding by allowing only selected nodes (MPR nodes) to flood the TC message. Each node selects an MPR to reach its two-hop neighbors The OLSR model implements the neighbor sensing mechanism through periodic broadcast of Hello messages. These Hello messages are one-hop broadcasts (never forwarded) that carry neighbor type and neighbor quality information. The neighbor sensing mechanism provides information on up to two-hop neighbors. Generation and processing of the Hello messages are implemented as suggested in the OLSR RFC. Periodic and triggered Topology Control (TC) messages implement the topology discovery/diffusion mechanism in the OLSR model. TC messages are diffused throughout the network using controlled flooding, thus helping to form a topology of reachable nodes, previous hop on each node.

[3] INTRODUCTIONS OF STATISTICAL PROBABILITY DISTRIBUTIONS [9]

3.1 Poisson Distribution

A random variable x is said to follow a Poisson distribution if it assume only non – negative values and its probability mass function is given by $p(x, \lambda)$

 $p(x) = e - \lambda \lambda x / x! ; \qquad x = 0, 1, 2, 3 \dots, \lambda > 0$ = 0 otherwise

Here λ is known as the parameter of the distribution.

3.2 Normal Distribution

A random variable x is said to have a normal distribution with parameters μ (mean) and $\sigma 2$ (Variance) if its probability density function is given by the probability law.

f(x) = $1/\sigma(2\Pi)1/2$ X exp{-(x- μ)2/2 σ 2}, - $\infty < x < \infty$, - $\infty < \mu < \infty$, $\sigma > 0$

3.3 Rectangular or Uniform Distribution

A random variable x is said to have a continues rectangular distribution over an interval (a, b) if its probability density function is given by

f(x) = 1/(b-a) if a < x < b

Otherwise.

3.4 Gamma distribution

A random variable x is said to have a gamma distribution with parameter $\lambda > 0$, if its probability density function is given by

 $f(x) = e - xx \lambda - 1/\Gamma(\lambda), \ \lambda \ge 0, \ 0 \le x \le \infty$

3.5 Exponential Distribution

A random variable x is said to have an exponential distribution with parameter $\theta > 0$, if its probability density function is given by

 $\begin{array}{ll} f(x) = \theta \ e\text{-} \ \theta \ ; & x \geq 0 \\ 0 & \text{otherwise} \end{array}$

[4] ANALYSIS OF SIMULATION

4.1 Performance of Adhoc Network with using mobility

As shown in the figure 4.1, Adhoc wireless network model is developed using mobility models. In Adhoc wireless network model 60 nodes are participating in the network. All nodes of the network are not following same mobility models, instead of that, 60 nodes are divided into 10 groups. Every group is following one trajectory; white lines indicated in the figure 4.1, shows moving direction of the nodes.



Figure 4.1 Adhoc Wireless Network Model with Mobility

4.1.1 Performance Analysis of Protocols.

We have constructed a detailed simulation model that accurately follows the details of Wireless routing protocol on random waypoint model. We have performed simulation for specifically inter arrival time probability distributions. In order to verify the accuracy of our model, we set up the simulator to represent a real system for which sufficient details are available in the literature. Our simulation model is based on OPNET 14.5 simulator. The effective parameters with their optimized values are reported here for each of different set of simulation.

- 1. Throughput (bits/sec).
- 2. End-to-End Delay.
- 3. Retransmission Attempts (packets).

4.1.2 Wireless Throughput (bits/sec)

Performance of AODV with reference to throughput was best as compared to the other protocols as shown in the figure 4.1.

Performance of DSR and OLSR was stable but poor throughout the simulation as compared to AODV

Name of Protocol	Average Throughput (bits/sec)
	Model with Mobility
AODV	15038091.11
DSR	4300581.778
OLSR	4176687.111

 Table 4.1 Average throughput (bits/sec)



Figure 4.1 Throughput (bits/sec)

4.1.2 End-to-End Delay (sec) :

AODV and DSR at the starting of simulation have shown delay but as simulation progresses it becomes stable to the lowest value of delay. DSR has shown highest delay of 0.11 sec and AODV has shown 0.067 sec as shown in the figure 4.2. An average delay of OLSR was 0.00031.

Delays of model with mobility conditions and without mobility conditions are shown in following table. As shown in the table 5.2, it is observed that model with mobility conditions has slightly more delay as compared to ad hoc network without mobility conditions. Ad hoc network model with mobility conditions has shown more delay as compared to model without conditions with better throughput.



Figure 4.2 End-to-End Delay (bits/sec)

Fable	4.2	End-to-End	Delav	(sec)	
		Ling to Ling	Dunay	(,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	

Name of Protocol	Average Throughput (bits/sec)
	Model with Mobility
AODV	0.0014
DSR	0.0019
OLSR	0.00031

4.1.3 Retransmission Attempt

Initially, at the start of simulation AODV protocol has shown maximum number of retransmission attempts but as simulation progresses it settled on an average 0.004925 packets. As shown in the figure 4.3 DSR has shown maximum number of retransmission attempts as compared to other protocols. At the start of simulation DSR has also shown maximum retransmission attempts, and then it settled to average 0.002420 packets.

OLSR protocols have shown average retransmission attempts 0.001518.. Comparison between Ad hoc network model with mobility conditions and without mobility conditions are given in table 4.3. When in Adhoc network model node movements are restricted then minimum number of retransmission attempts is required as shown in the table 4.3



Figure 4.3 Retransmission Attempts (Packets)

Name of Protocol	Average Throughput (bits/sec)
	Model with Mobility
AODV	0.004925
DSR	0.002420
OLSR	0.001518

 Table 4.3 Retransmission Attempts (bits/sec)

5.4 CONCLUSIONS:

A MANET simulation model was built-up using parameter which mention in table 3.1. Different statistical distributions load have given to the model and performance of simulation models were observed and discuss in above sections. Performance of routing protocols were considered and given in above sections. Overall performance of routing protocols is as follows:

- 1. Performance of all protocols has improved after putting mobility conditions except OLSR protocol with reference to throughput.
- 2. In terms End-to-End Delay AODV has shown improvement as compared to other protocols.
- 3. Retransmission attempts have been considerably reduced in all protocols after putting mobility conditions.

REFERENCES

[1] F.Borgonovo, A.Capone, M.Cesana, L.Fratta, "ADHOC MAC: new MAC architecture for ad hoc networks providing efficient and reliable point-to-point and broadcast services," *Wireless Networks Journal*, Vol.10, issue 4, July 2004.

[2] Geetha Jayakumar and G. Gopinath, Performance Comparison of Two On-demand Routing Protocols for Ad-hoc Networks based on Random Way Point Mobility Model, American Journal of Applied Sciences 5 (6): 659-664, 2008

[3] K.E.Kannammal and K.E.Eswari Behaviour of Adhoc Routing Protocols in Different Terrain Sizes International Journal of Engineering Science and Technology Vol. 2(5), 2010, 1298-1303.

[4] Georgios Parissidis, Vincent Lenders, Martin May, and Bernhard Plattner Multi-path Routing Protocols in Wireless Mobile Ad Hoc Networks, 2006

[5] Aparna K, Performance Comparison of MANET (Mobile Ad hoc Network) Protocols (ODMRP with AMRIS and MAODV), 2010 International Journal of Computer Applications (0975 – 8887) Volume 1 – No. 10.

[6] Mohd Izuan Mohd Saad and Zuriati Ahmad Zukarnain, Performance Analysis of Random-Based Mobility Models in MANET Routing Protocol, European Journal of Scientific Research ISSN 1450-216X Vol.32 No.4 (2009).

[7] Yasser Kamal Hassan, Mohamed Hashim Abd El-Aziz and Ahmed Safwat Abd El-Radi Performance Evaluation of Mobility Speed over MANET Routing Protocols , *International Journal of Network Security*, *Vol.11, No.3, , Nov. 2010*

[8] Bhavyesh Divecha, Ajith Abraham, Crina Grosan and Sugata Sanyal Impact of Node Mobility on MANET Routing Protocols Models,

[9] User Manual for IMPORTANT Mobility Tool Generator in NS-2 Simulator. http://nile.usc.edu/important/software.htm , Release Date February 2004 .

[10] Mobility Generator (version 1.0) from the site, http://nile.usc.edu/important/software.htm, February 2004
[11] Guolong Lin, Guevara Noubir and Rajmohan Rajaraman, "Mobility Models for Adhoc Network Simulation", In Proceedings of IEEE INFOCOM 2004. Volume 1, pp. 7-11, 2004.



MANET PERFORMANCE ANALYSIS USING DIFFERENT DISTRIBUTION LOAD Dr. Ravikant Zirimite, Dr. Manasi Bhate²

¹Associate Professor, MES'S Institute of Management and Career Courses Pune, India ¹Associate Professor, MES'S Institute of Management and Career Courses Pune, India Affiliated by Savitribai Phule Pune University, Pune, India

ABSTRACT

The paper presents the results of a detailed packet-level simulation comparing four multi-hop wireless ad hoc network routing protocols under the load of different probability distributions, that cover a range of design choices having different protocol viz. DSR, AODV and OLSR. We have extended the OPNET network simulator to accurately model the MAC and physical-layer behavior of the IEEE 802.11 wireless LAN standard, including a realistic wireless transmission channel model. Simulation of 60 mobile nodes has been carried out and the performance optimization is determined.

Keywords - *Simulation, Opnet, Wireless, Statistical probability distribution, IEEE802.11, throughput, delay, retransmission attempt, load, protocol, MAC, LAN.*

[1] INTRODUCTION

Ad-hoc wireless network is that network where no communication is present, in such network; each mobile node operates not only as a host but also as router. Mobile nodes in the network may not be with in range of each other, communication of these nodes perform by discovering "multi-hop" paths through the network to other nodes. This type of network is some time called infrastructure less network [1]. Some examples of the possible uses of ad hoc networking are students using laptop computers to participate in an interactive lecture, business associates sharing information during a meeting, soldiers relaying information for situational awareness on the battlefield [2, 3]. Many different protocols have been proposed to solve the multi hop routing problem in ad hoc networks, each based on different assumptions and intuitions.

Mobile Ad hoc Networks (MANETs)[1] are an emerging technology that allows establishing an instant communication network for civilian and military applications, without relying on pre-existing fixed network infrastructure. The nodes in a MANET can dynamically join and leave the network, frequently, often without warming, and possibly with out disruption to other nodes' communication. Each node in the network also acts as router, forwarding data packet for other nodes. A central challenge in design of Ad hoc network is the development of dynamic routing protocols that can effectively find the route between two communicating nodes. The routing protocol must be able to keep up with the high degree of node mobility that often changes the topology drastically and unpredictably.

The current Mobile Ad Hoc Network (MANET) [2] paradigm as described by the Internet Engineering Task Force (IETF) MANET work group. Routing algorithms are often difficult to

formalize into mathematics; they are instead tested using extensive simulation. A large amount of work has been done in the area of energy efficient routing. This approach attempts to maximize network lifetime by routing through paths, which use the least amount of energy relative to each node. Now a day, more attention has been given to use specific network parameters while specifying routing matrixes. Routing matrixes includes delay of network, link capacity, link stability or identifying low mobility nodes. These schemes are generally based on previous work, which is then enhanced with the new matrix.

The paper is providing a realistic, quantitative analysis comparing the performance of a variety of multi-hop wireless ad hoc network routing protocols. We present results of detailed simulations showing the relative performance of three recently proposed ad hoc routing protocols: DSR [3, 4, 5], AODV [6], OLSR [8]

Our results in this paper are based on simulations of an ad hoc network of 60 wireless mobile nodes moving about and communicating with each other. We analyze the performance of each protocol and explain the design choices that account for their performance.

The section 2 of the paper describes the different types of protocols used in the simulation. The section 3 has given description of different types of statistical probability distribution used for arrival and departure of packets in simulation. The performance analysis is describes in section 4. The section 5 has summaries conclusion of the paper.

[2] DESCRIPTION OF THE PROTOCOLS

2.1 Dynamic Source Routing (DSR) [3, 4, 5]

The DSR is using source routing, i.e. the sender knows the complete hop-by-hop route to the destination. When node sending data packet to the destination, DSR has use route discovery by flooding the network with route request (RREQ) packets. Each node receiving an RREQ, node will rebroadcast it, unless it is the destination or it has a route to the destination in its route cache. Such node replies to the RREQ with a route replay (RREP) packet that is routed back to the original source. If any link on a source route is broken, the source node is notified using a route error (RERR) packet. The source will initialize new discovery process. No special mechanism is required to detect routing loops. Also, any forwarding node caches the source route in a packet it forwards for possible future use. Several additional optimizations have been proposed and have been evaluated to be very effective by the authors of the protocol [5], as described in the following:

• Salvaging: An intermediate node can use an alternate route from its own cache when a data packet meets a failed link on its source route.

• Gratuitous route repair: A source node receiving an RERR packet piggybacks the RERR in the following RREQ. This helps clean up the caches of other nodes in the network that may have the failed link in one of the cached source routes.

• Promiscuous listening: When a node overhears a packet not addressed to itself, it checks whether the packet could be routed via itself to gain a shorter route. If so, the node sends a gratuitous RREP to the source of the route with this new, better route. Aside from this, promiscuous listening helps a node to learn different routes without directly participating in the routing process.

2.2 Ad Hoc on demand Vector (AODV) [6]

AODV discovers routes on demand basis. It uses routing table to maintain routing information, one entry per destination. RREP packet is used to replies back to the source and, subsequently, to route data packets to the destination. AODV uses sequence numbers to maintain at each destination to determine routing information and to prevent routing loops [6]. AODV working on timer- based states in each node. A routing table entry is expired if not used recently. If node link is broken, the all predecessor nodes forward the RERR packets, to effectively erasing all routes using broken link. AODV uses expanding ring search technique initially to discover routes to an unknown destination. AODV algorithm has the ability to quickly adapt to dynamic link conditions with low processing and memory overhead. AODV offers low network utilization and uses destination sequence number to ensure loop freedom AODV keeps the following information with each route table entry.

(i) Destination IP address (IP address for the destination node),

- (ii) Destination sequence number,
- (iii) Valid destination sequence number flag,
- (iv) Network interface,

(v) Hop count, that is, number of hops required to reach the destination,

(vi) Next hop (the next valid node that did not re broadcast the RREQ message),

(vii) List of precursor,

(viii) Life time, that is, expiration or deletion time of a route.

2.3 Optimized Link State Routing (OLSR) [8]

The OLSR model implements the MPR (Multi Point Relay) flooding mechanism to broadcast and flood Topology Control (TC) messages in the network. The algorithm is implemented as suggested in OLSR RFC 3626. This mechanism takes advantage of controlled flooding by allowing only selected nodes (MPR nodes) to flood the TC message. Each node selects an MPR to reach its two-hop neighbors The OLSR model implements the neighbor sensing mechanism through periodic broadcast of Hello messages. These Hello messages are one-hop broadcasts (never forwarded) that carry neighbor type and neighbor quality information. The neighbor sensing mechanism provides information on up to two-hop neighbors. Generation and processing of the Hello messages are implemented as suggested in the OLSR RFC. Periodic and triggered Topology Control (TC) messages implement the topology discovery/diffusion mechanism in the OLSR model. TC messages are diffused throughout the network using controlled flooding, thus helping to form a topology of reachable nodes, previous hop on each node.

[3] INTRODUCTIONS OF STATISTICAL PROBABILITY DISTRIBUTIONS [9]

3.1 Poisson Distribution

A random variable x is said to follow a Poisson distribution if it assume only non – negative values and its probability mass function is given by $p(x, \lambda)$

 $p(x) = e - \lambda \lambda x / x! ; \qquad x = 0, 1, 2, 3 \dots, \lambda > 0$ = 0 otherwise

Here λ is known as the parameter of the distribution.

3.2 Normal Distribution

A random variable x is said to have a normal distribution with parameters μ (mean) and $\sigma 2$ (Variance) if its probability density function is given by the probability law.

f(x) = $1/\sigma(2\Pi)1/2$ X exp{-(x- μ)2/2 σ 2}, - $\infty < x < \infty$, - $\infty < \mu < \infty$, $\sigma > 0$

3.3 Rectangular or Uniform Distribution

A random variable x is said to have a continues rectangular distribution over an interval (a, b) if its probability density function is given by

f(x) = 1/(b-a) if a < x < b

Otherwise.

3.4 Gamma distribution

A random variable x is said to have a gamma distribution with parameter $\lambda > 0$, if its probability density function is given by

 $f(x) = e - xx \lambda - 1/\Gamma(\lambda), \ \lambda \ge 0, \ 0 \le x \le \infty$

3.5 Exponential Distribution

A random variable x is said to have an exponential distribution with parameter $\theta > 0$, if its probability density function is given by

 $\begin{array}{ll} f(x) = \theta \ e\text{-} \ \theta \ ; & x \geq 0 \\ 0 & \text{otherwise} \end{array}$

[4] ANALYSIS OF SIMULATION

4.1 Performance of Adhoc Network with using mobility

As shown in the figure 4.1, Adhoc wireless network model is developed using mobility models. In Adhoc wireless network model 60 nodes are participating in the network. All nodes of the network are not following same mobility models, instead of that, 60 nodes are divided into 10 groups. Every group is following one trajectory; white lines indicated in the figure 4.1, shows moving direction of the nodes.



Figure 4.1 Adhoc Wireless Network Model with Mobility

4.1.1 Performance Analysis of Protocols.

We have constructed a detailed simulation model that accurately follows the details of Wireless routing protocol on random waypoint model. We have performed simulation for specifically inter arrival time probability distributions. In order to verify the accuracy of our model, we set up the simulator to represent a real system for which sufficient details are available in the literature. Our simulation model is based on OPNET 14.5 simulator. The effective parameters with their optimized values are reported here for each of different set of simulation.

- 1. Throughput (bits/sec).
- 2. End-to-End Delay.
- 3. Retransmission Attempts (packets).

4.1.2 Wireless Throughput (bits/sec)

Performance of AODV with reference to throughput was best as compared to the other protocols as shown in the figure 4.1.

Performance of DSR and OLSR was stable but poor throughout the simulation as compared to AODV

Name of Protocol	Average Throughput (bits/sec)
	Model with Mobility
AODV	15038091.11
DSR	4300581.778
OLSR	4176687.111

 Table 4.1 Average throughput (bits/sec)



Figure 4.1 Throughput (bits/sec)

4.1.2 End-to-End Delay (sec) :

AODV and DSR at the starting of simulation have shown delay but as simulation progresses it becomes stable to the lowest value of delay. DSR has shown highest delay of 0.11 sec and AODV has shown 0.067 sec as shown in the figure 4.2. An average delay of OLSR was 0.00031.

Delays of model with mobility conditions and without mobility conditions are shown in following table. As shown in the table 5.2, it is observed that model with mobility conditions has slightly more delay as compared to ad hoc network without mobility conditions. Ad hoc network model with mobility conditions has shown more delay as compared to model without conditions with better throughput.



Figure 4.2 End-to-End Delay (bits/sec)

Fable	4.2	End-to-End	Delav	(sec)	
		Ling to Ling	Dunay	(,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,	

Name of Protocol	Average Throughput (bits/sec)
	Model with Mobility
AODV	0.0014
DSR	0.0019
OLSR	0.00031

4.1.3 Retransmission Attempt

Initially, at the start of simulation AODV protocol has shown maximum number of retransmission attempts but as simulation progresses it settled on an average 0.004925 packets. As shown in the figure 4.3 DSR has shown maximum number of retransmission attempts as compared to other protocols. At the start of simulation DSR has also shown maximum retransmission attempts, and then it settled to average 0.002420 packets.

OLSR protocols have shown average retransmission attempts 0.001518.. Comparison between Ad hoc network model with mobility conditions and without mobility conditions are given in table 4.3. When in Adhoc network model node movements are restricted then minimum number of retransmission attempts is required as shown in the table 4.3



Figure 4.3 Retransmission Attempts (Packets)

Name of Protocol	Average Throughput (bits/sec)
	Model with Mobility
AODV	0.004925
DSR	0.002420
OLSR	0.001518

 Table 4.3 Retransmission Attempts (bits/sec)

5.4 CONCLUSIONS:

A MANET simulation model was built-up using parameter which mention in table 3.1. Different statistical distributions load have given to the model and performance of simulation models were observed and discuss in above sections. Performance of routing protocols were considered and given in above sections. Overall performance of routing protocols is as follows:

- 1. Performance of all protocols has improved after putting mobility conditions except OLSR protocol with reference to throughput.
- 2. In terms End-to-End Delay AODV has shown improvement as compared to other protocols.
- 3. Retransmission attempts have been considerably reduced in all protocols after putting mobility conditions.

REFERENCES

[1] F.Borgonovo, A.Capone, M.Cesana, L.Fratta, "ADHOC MAC: new MAC architecture for ad hoc networks providing efficient and reliable point-to-point and broadcast services," *Wireless Networks Journal*, Vol.10, issue 4, July 2004.

[2] Geetha Jayakumar and G. Gopinath, Performance Comparison of Two On-demand Routing Protocols for Ad-hoc Networks based on Random Way Point Mobility Model, American Journal of Applied Sciences 5 (6): 659-664, 2008

[3] K.E.Kannammal and K.E.Eswari Behaviour of Adhoc Routing Protocols in Different Terrain Sizes International Journal of Engineering Science and Technology Vol. 2(5), 2010, 1298-1303.

[4] Georgios Parissidis, Vincent Lenders, Martin May, and Bernhard Plattner Multi-path Routing Protocols in Wireless Mobile Ad Hoc Networks, 2006

[5] Aparna K, Performance Comparison of MANET (Mobile Ad hoc Network) Protocols (ODMRP with AMRIS and MAODV), 2010 International Journal of Computer Applications (0975 – 8887) Volume 1 – No. 10.

[6] Mohd Izuan Mohd Saad and Zuriati Ahmad Zukarnain, Performance Analysis of Random-Based Mobility Models in MANET Routing Protocol, European Journal of Scientific Research ISSN 1450-216X Vol.32 No.4 (2009).

[7] Yasser Kamal Hassan, Mohamed Hashim Abd El-Aziz and Ahmed Safwat Abd El-Radi Performance Evaluation of Mobility Speed over MANET Routing Protocols , *International Journal of Network Security*, *Vol.11, No.3, , Nov. 2010*

[8] Bhavyesh Divecha, Ajith Abraham, Crina Grosan and Sugata Sanyal Impact of Node Mobility on MANET Routing Protocols Models,

[9] User Manual for IMPORTANT Mobility Tool Generator in NS-2 Simulator. http://nile.usc.edu/important/software.htm , Release Date February 2004 .

[10] Mobility Generator (version 1.0) from the site, http://nile.usc.edu/important/software.htm, February 2004
[11] Guolong Lin, Guevara Noubir and Rajmohan Rajaraman, "Mobility Models for Adhoc Network Simulation", In Proceedings of IEEE INFOCOM 2004. Volume 1, pp. 7-11, 2004.




STUDENTS CENTRIC ONLINE DOCUMENT REPOSITORY WITH E-VERIFICATION (MYDOC)

Dr. Minakshi S. Tumsare

tumsareminakshi@gmail.com Assistant Professor, Department of MCA, IMCC, Pune Savitribai Phule Pune University,Pune

ABSTRACT

Most recently the Government of India took initiative to help the Indians by providing a great opportunity to save their crucial documents on Internet in the DigiLocker. "The initiative is implemented to create 'digital empowerment' among the masses. The digital storage system acts as a helping hand in providing data to government agencies, departments, college admissions and applying for different jobs. Digilocker systems will continue to play a key role in improving the performance and effectiveness of organizations. In this paper, the goal is to provide the facility of online document repository for students or any person to upload there original documents in it. In that case the students or any person s need not carry hardcopy of original documents every time and important point is that the use of papers is decreases.

Keywords: DigiLocker, Digital Empowerment, Document Repository

[1] INTRODUCTION

MyDoc locker is the repository to store the students or any person's academic documents which are required by the university eligibility section or any other government agency and the each user will be assigned the unique identification number. This number will act as reference whenever the agency/authority or university can refer the documents of students or any person instead of asking

for physical documents for admission or exam for validation every time. Students or any person can upload the documents like SSC, HSC, PG mark sheets or degree certificates, self-photo etc. using this system. This will avoid the manual work and duplication of work to submit the documents every time as these documents will be available in the repository so it will not be necessary to submit the documents everywhere.

[2] ORIGIN OF THE RESEARCH PROBLEM

In current system each students or any person has to submit the hardcopy of all the documents to take the admission of any course in college under the university, or job application in any company or for passport application process. After documents submitted, the students or any person has to fill online application form of the respective work. The application process gets completed after original documents are been verified by the government authorities or agencies like university, passport office or college but the original documents are checked through manual document verification process through different government agencies separately. So instead of doing the manual verification process, if the documents are already available in repository along with the photocopies and verification number, then the university, AICTE and college etc can view these documents online and also the system will show the eligibility by validating the data along the details. It is easiest and fastest process to avoid the repetition of uploading the document and verification process every time with form submission as the documents will be available in the repository. The students or any person only need to fill the application form at the time of admission to college, password or caste certification agency, recruitment agencies, multinational companies etc. and fill the required application form only without uploading any other document. To avoid the above problems the researchers propose the MyDoc Locker system which will be useful for faster processing and avoiding the duplication of work for all verification agencies, government agencies and university or college authorities etc..

[3] REVIEW OF LITERATURE

Raghunathan V.S. et.al. (2016)[1] Evaluated the open standard system of unique identification (UID) and verification of any government documents to be secured, verified and track the status of documents which are issued by the government through the Adhaar Card. Dr. T. Jeevan Kumar (2015)[2] focused on the internet and its impact on the society, in various field like Media, Communication, Banking, Government and Business, Education and Medical Politics. The author highlighted "The initiative is implemented to create 'Digital Empowerment' among the masses. The digital storage system will help to provide data to Government agencies, Departments and Applying for different jobs." M. Michael Serbinis; Evan V Chrapko, [3] the patent invention relates to apparatus and methods for use of access tokens in a system for managing electronic documents over open networks, such as the Internet, to permit users to store, retrieve, and collaboratively manipulate files. Martin King et. al [4], focused on Data Gathering In Digital and Rendered Document Environments, this invention relates generally to publishing and, more specifically, to

data gathering systems and methods for use in a published document environment including documents With advertisements.

[4] ANALYSIS OF THE PROBLEM

Now a days following problems are faced by students in case of using physical documents.

- Documents in physical form leading to huge administrative overhead to check student's eligibility.
- Challenges to students in submission of multiple copies of certificates to complete the admission process.
- Sharing of physical documents is a challenge.
- It is a challenge to verify authenticity of physical documents and their copies.

Digital Locker will provide secure access to Students or any person s educational documents. Digital Locker will minimize the use of physical documents by use of the electronic documents. It is easy to validate the authenticity of these documents. Digital Locker will reduce the administrative overhead of AICTE, University and College admin departments created due to paper work. It will also make it easy for the students or any person s to complete admission process by verifying documents online by saving time and effort as their documents will now be available anytime, anywhere and can be shared electronically in MyDoc.

[5] DOCUMENT SECURITY

Another important benefit of online document storage is secure document storage. Authorized person can pick up a document off of a desk. But without permission, no one can access a system that has sensitive documents. Document security goes well beyond who can access the system. With online document storage, users can be granted access to specific documents. Thus you can secure your most sensitive documents to only certain individuals who have a need to access that information. We can take this even further by only allowing some users to access all of a given document and other users to only be able see a portion of the document. In many instances, this type of security can be mandated by government regulations. However, for all companies, securing data is the key to making sure that confidential information does not get leaked. Security is also about ensuring the documents are preserved and do not get lost or damaged. Lost documents are obvious. Who hasn't looked for a file in a file cabinet and not found what they were looking for? Sometimes documents were misfiled. Other times the documents were on someone else's desk. In the worst instances, documents are removed from the premises and then forgotten at an offsite location. Documents can be lost forever, unless they are secured in a document management system.

[6] ONLINE DOCUMENT REPOSITORY SYSTEM

The purpose of an Electronic Document Management System is to provide a central repository for documents. An electronic document management system allows for:

•Secure environment for document storage;

•Immediate web based access to documents by authorized personnel from any location;

•Multiple users can access the same document simultaneously, at anytime;

•Management of versions of documents to minimize inconsistency and redundancy;

•Increased speed of storing, retrieving and modifying documents;

•Reduction of paper, space and staffing requirements associated with paper filing systems.

An online document Repository system offers a means of adding documents to an online repository and associating properties with those documents to make them easier for users to locate and retrieve. Users may upload search, view, check-out and check-in documents according to the user group to which they belong, or the access rights specifically granted by the individual that added the document to the online document Repository system. Usually access to an online document Repository system is provided based on user groups.

[7] PROPOSED INTENTS

- 1. Study and analyze the existing admission, eligibility process and exam form submission process in detail
- 2. Ensure authenticity of the e-documents and thereby eliminate usage of fake documents.
- 3. Anytime, anywhere access to the documents by the students or any person s.
- 4. MyDoc Locker Repository System by developing the application which will be store the documents of the students or any person as well as marks and percentage with secured interface.
- 5. Design new Repository System application for AICT, University and Students or any person s.
- 6. Support a well-structured standard document format to support easy sharing of documents across departments
- 7. Streamline the process of document verification/exchange.
- 8. Ensure privacy and security through user authentication and consented access to documents

[8] COMPONENTS OF DIGITAL LOCKER SYSTEM

Repository is a Collection of e-Documents which are uploaded by issuers in a standard format and exposing a set of standard APIs for secure real-time search and access.

Access Gateway provides a secure online mechanism for requesters to access e-documents from various repositories in real-time using e-Document URI (Uniform Resource Indicator). The URI is

a link to the e-Document uploaded by an issuer in a repository. The gateway will identify the address of the repository where the e-Document is stored based on the URI and will fetch the e-Document from that repository.

Web interface for uploading and downloading the document and provide the authentication to the user.

Issuer any public or private sector entity/organization/department issuing digitally signed edocuments to individuals/entities and making them available within a repository for access through a digital locker of their choice. The issuer is also responsible to revoke/invalidate their own documents.

Requester An entity/organization/department requesting secure access with user approval to specific e-documents stored in Mydoc the Ecosystem to provide paperless service to end users.

[9] BENEFITS OF DIGITAL LOCKER

Access: Students or any person s can access their digital documents anytime, anywhere and can share it online.

Paperless: It reduces the use of paper.

Authenticity: Digilocker makes it easier to validate the authenticity of document as they issued directly by the registered issuers.

E-verification: Self-uploaded documents can be digitally signed using the e-verification facility.

[10] PROCESS OF E-LOCKER

In digital locker students or any person s can upload the personal documents like Mark sheets, Adhar card, Pan Card etc. The uploaded document is verified by the system and then document will be stored in online repository or in E-locker system. When the students or any person 's wants do admission process or any other legal process at that time he needs not to carry the hard copy of documents because all the original documents is uploaded on online repository or Mydoc. While accessing the document from E-locker system the students or any person will get the Unique Identification Key (UIK) to provide authenticity of Document to requester like (AICTE, University and College). After getting the permission from students or any person s to access the document it can be verify through the AICTE, University and College. The students or any person 's eligibility process will be successfully completed by document e-verification process from online repository or Mydoc. The diagrammatic representation of E-Locker process is shown in figure 1.



Figure: 1. Process of E-Locker

[10.1] FLOW OF ONLINE DOCUMENT E-VERIFICATION PROCESS

[10.1.1] STUDENTS OR ANY PERSON REGISTRATION PROCESS

In online document e-verification process students or any person first signup and fill the registration form (fills his/her personal details). The students or any person 's personal details will be validated through Aadhar database and digital document will get create, after creation of digital document students or any person s registered successfully on site as shown in **[Figure-2]**.



Figure: 2. Students or any person Registration Process

[10.1.2] PROCESS OF E-LOCKER AND DOCUMENT E-VERIFICATION

After successful registration of students or any person s, he/she will be able to login in to the account and update his/her profile. After profile updation students or any person s will be able to upload the educational document in online repository or MyDoc. The uploaded educational certificates will verify by E-verification process from competent authority like State board, DTE, AICTE and Universities. If the students or any person 's document is successfully verified then it is accepted and uploaded in Online Document Repository or MyDoc and generate document verification number. If document verification is rejected then again it drives to the e-verification process. The Execution flow of E-Locker and Document E-verification is shown in figure 3.



Figure: 3. Process of E-locker and Document E-verification

[10.1.3] STUDENTS OR ANY PERSON ADMISSION PROCESS

In students or any person admission process, the students or any person s want to take admission in any college or in any University if his/her document is available on online repository or in MyDoc in that case students or any person s need not to carry hard copy of document. Students or any person s will visit on college website, and will complete his/her registration process by giving personal details like Name, Address, Email and contact number. These students or any person

STUDENTS CENTRIC ONLINE DOCUMENT REPOSITORY WITH E-VERIFICATION (MYDOC)

details and Educational documents will verify through Aadhar number and Document Verification number. In e-verification Process educational certificates should match with aadhar no as well as document verification no. if it is match then document verified successfully, otherwise again it goes to verification process as shown in figure 4.



Figure: 4. Process of E-locker and Document E-verification

[11] CONCLUSION:

"Online Document Repository or MyDoc" reduces the unnecessary usage of repetative documents submissions risk of carrying/handling, forgery etc. Sharing and document authentication of the e-documents will be done online. This provides easy way to upload our own online documents and shared with Passport, AICTE, DTE, Universities Colleges or other entities like Government

organizations. In this paper author focused on educational system as students are facing the problems while taking the admission in college, DTE, AICTE, University and Free ship or scholarship process. The proposed system can be used by any person to do the administration works as he/she can share or open his/her documents anywhere at any time. Hence the process of E-locker, E-verification and other government agencies work process increases the functionality and efficiency. We can achieve the objectives to minimize the use of physical documents, ensure authenticity of the e-documents by Gov. Verifier and thereby eliminate usage of fake documents, anytime, anywhere access to the documents by the students or any persons, it ensures privacy and authorized access to students or any person's data from online document Repository or MyDoc.

REFERENCES:

- Raghunathan V.S. Et.al. "Optimization of Digitalized Document Verification Using E-governance Service Delivery Platform (E-SDP)", Department of computer science international journal of applied engineering research ISSN 0973-4562 volume 11, number 4 (2016) pp.2535-2543 © research India publications. Http://www.ripublication.com 2535
- 2. Dr. T. Jeevan Kumar, "Internet and Its Impact on the Society", Indian scholar an international multidisciplinary research e-journal 68 vol. 2 issue ii December, 2015 ISSN 2350-109x www.indianscholar.co.in
- 3. M. Michael Serbinis; Evan v. Chrapko, "Apparatus and Methods for Use of Access Tokens in an Internet Document Management System", Nov 6, 2001.
- 4. Martin king et. al. "Data Gathering in Digital and Rendered Document Environments", Apr. 15, 2010
- Siciarek, J; Smiatacz, M; Wiszniewski, B. "For Your Eyes Only Biometric Protection of PDF Documents" Proceedings of the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE);
- Dr. Mohite Pratapsinh Vitthalrao, Ms. Joshi MugdhaMukund et. al. Digilocker (Digital Locker-Ambitious Spect Of Digital India Programme.), GE-International Journal of Management Research (GE-IJMR), Volume -3, Issue-6, June-2015.
- Vaibhav Vaijawade, Nandkishor Khelkar, Dhiraj Thakare Review on "Cloud Based Digital Locker", IJRISE e-ISSN: 2394-8299 Special Issue: Techno-Xtreme 16 p-ISSN: 2394-8280
- 8. <u>https://digitallocker.gov.in/</u>
- 9. J Sengol Mary, S Usha, "Web based document management systems in life science organization", *Green Engineering and Technologies (IC-GET) 2015 Online International Conference on*, pp. 1-3, 2015.
- 10. Ms.Mehek Gulati & Ms.Kanika Verma, "Digital locker", Journal of Emerging Technologies and Innovative Research (JETIR) www.jetir.org, June 2016, Volume 3, Issue 6, ISSN-2349-5162



INVESTMENT BEHAVIOUR OF INDIVIDUAL INVESTORS IN SHARES, BONDS AND MUTUAL FUNDS: A PRAGMATIC STUDY

Swapnaja Patwardhan¹, Vivek Katdare², Manish Joshi³

¹ MES's IMCC, Pune, <u>swapnaja.patwardhan@gmail.com</u> ² KCE's Institute of Management & Research, Jalgaon, <u>vvkatdare@rediffmail.com</u> ³Dept. of C.S., NMU, Jalgaon_, <u>joshmanish@gmail.com</u>

ABSTRACT

Diversified financial service sector has given wide range of opportunities to individual investor's. The individual's decision of investment is prejudiced by the kind of services rendered and the benefits offered in the financial market. Financial knowledge and experience has an impact in financial investment decision making process. This paper aims to study the investment patterns of individual investors' considering the objective of the investment, composition of portfolio, risk bearing capacity, the market knowledge they possess, demographic characteristics, attitude etc. Duration of the investment is one of the most important factors in decision making process. Quantitative research has been conducted to determine the type of relationship that exists between the variables. To perform this literature review research papers, journal articles, white papers on behaviour of individual investor from various countries are studied. Research papers are analysed thoroughly in order to present excellent conceptual framework. This study proposes that the perception of an individual investor could be a great help to the financial consultants, fund managers and asset management agencies.

Keywords: Investment Management, Individual Investor, Behavioural Patterns, Financial Markets.

1. INTRODUCTION

An investment in diversified financial sector is an interesting task which attracts people during their journey of life. Basic aim behind all investments is capital appreciation for earning income. Wide ranges of opportunities are available for an investor to invest in financial market. The pattern of investment differs from investor to investor on the basis of objective behind the investment. There are several factors that attract and motivate investors' to invest in financial market. Researchers from various countries are concentrating on the behavioural patterns and demographic characteristics of individual investor. Decisions of individual investors' were believed to be based on the Modern Portfolio Theory proposed by Markowitz in 1952. There are three basic criterion's of modern portfolio theory i.e. Standard deviation, Expected return and correlation. However it was proved that individual investor possess very few stocks and are unable to diversify their portfolio [1]. An individual's risk bearing capacity also plays an important role in financial decision making process to achieve desired financial goals. To understand the individual investors' behaviour factors like objective of the investment, composition of portfolio, risk bearing capacity, the market knowledge, demographic characteristics and attitude need to be considered.

The present study aims to understand the investment patterns of individual investors'. To perform this literature review research papers, journal articles, white papers on behaviour of individual investor from various countries are studied. Research papers are analysed thoroughly in order to present excellent conceptual framework. It will be a great help to the financial consultants, fund managers and asset management agencies to understand the requirements and expectations of an investor.

In this study, research papers are reviewed on the basis of parameters like objective, Method used Main finding and sample size, Result obtained etc. It was observed that in most of the papers primary data



collection method is used and responses of investors' are collected through surveys, questionnaires, and interviews. In some papers secondary source of data collection is also used like Stock exchange data and Stock brokering agencies.

2. LITERATURE REVIEW

Nowadays various researchers are exploring behavioural patterns that influence individual investors'. It is observes that there is a dynamic relationship between behaviour of individual investor, Stock price movements, volume of trading and returns gained. It is also observed that up to certain extent there could be region wise cultural difference in investment patterns. Are view of the literature clearly states that individual investors' behavioural pattern is the most important characteristics in stock price movements and the probable returns.

Nagy et al. in 1994 examined the factors that influenced the individual investor's investment pattern in stock market with the help of Factor Analysis. They observed some factors like wealth growth, social relevance; personal financial requirements etc. are the factors that influence the behavioural pattern of individual investors' [2].

Gupta et al. in 2001 inspected and compared the preference patterns of investors' among Mutual Funds. Study was undertaken with the help of Descriptive Analysis. It was observed that UTI owned US-64 was the most widespread mutual fund schemes. Its position with reference to other equity schemes is weaker [3].

Merikas et al. in 2004 identified the factors influencing the individual investors' in the Greek stock exchange. Factor analysis technique was used for the study. They recognized the criteria's like expected corporate earnings, financial statement condition and status of the firm are the most influencing factors [4].

Feng et al. in 2007 examined the behaviour pattern of men and women in stock markets from China. Techniques like descriptive analysis, survival analysis, cross sectional statistics and correlation are implemented and concluded that men and women similar kind of behaviour and performance. Results were verified with returns gained on investments [5].

Walia and Ravikant in 2009 inspected the expectations of individual investor towards mutual funds. Chi square test, rank and rating methodology, average performance scores (APS) and ANOVA are the techniques applied for the study. They detected that the investment patterns vary and added quality dimensions of existing services need to be enhanced [6].

Hood et al. in 2011 studied the factors which affect the decisions of socially responsible investors'. Logistic regression, descriptive statistics, clustered standard error and correlation are the techniques used. Study showed that personal and social characteristics and values have major impact on individual investors' stock composition [7].

S. Lodhi in 2014 investigated individual investor behaviour of Karachi, Pakistan. She observed that there were five independent variables - financial literacy, high experience, use of accounting information, importance of analysing financial statements and age are the factors that affect on the investors' decision making process. She also observed that accounting information and risk aversion are directly related to each other [8].

3. METHODOLOGY

The present study is based upon the primary data collected using questionnaire. Random sampling technique is used for data collection. The sample size is 65 individual investors' from Pune region. The respondents are teachers, managers, businessmens, housewifes, workers, entrepreneurs and executives from various income groups.



3.1 Objectives:

The present study is an attempt to understand behaviour of individual investor in various investment avenues. For this following objectives are considered- 1] To identify the factors that affect upon the investment decision making process. 2] To study the relationship between the investors' profile and their behaviour towards investments.

3.2 Study Sample:

The population for the present study comprises of the investors' from Pune region. Investors' from all age group, education, occupation and income group are considered who are investing in shares, mutual funds and bonds.

3.3 Data Collection Method:

Individual investor's data is collected through questionnaire. Questionnaire comprises of two sections as - Section-I deals with Personal Information and Section-II includes the questions about the composition and risk bearing capacity of individual investor's Portfolio. Section -II is divides into six axes viz. First Axis: Investment objectives, Second Axis: Time Horizon, Third Axis: Level of Satisfaction, Fourth Axis: Factors Influencing Investments, Fifth Axis: Knowledge and Sixth Axis: Risk Tolerance. With the help of 26 questions data is collected and recorded. Answer of each question is assigned some positive weights. Higher weight indicate risky portfolio and lower weight indicate a balanced or Moderate or defensive portfolio. Classification of investors' on the basis of risk bearing capacity is Defensive, Conservative, Balances, Moderate, Aggressive and Very Aggressive is shown in Table-1 below .

Investor Type	Risk profile	Expectation		
Defensive	Risk Averse	Regular Income. Capital		
		Gain		
Conservative	Extremely Risk Averse	Capital Gain		
Balanced	Take necessary Risk	Good Return		
Moderate	Comfortable Level of	Steady Returns		
	Risk			
Aggressive	Take little more Risk	Take little more Risk		
Very Aggressive	Take Maximum Risk	Maximum Returns		

Table-1: CLASSIFICATION OF INVESTORS

3.3 Data Analysis:

It was observed that some factors such as Demographic, Personal, Psychosomatic Characteristics, Ethics, Risk Tolerance, Economic Factors, Political Factors, and few others directly affect on the behaviour of individual investor and are listed in Table -2.

Demographic	Age, Gender, Education, Marital		
Characteristics	status, Occupation, Annual income		
Personal	Knowledge, Confidence, Ability,		
Characteristics	Responsibility, Social Responsibility,		



	Policies and practices for children's,
	NGOs, Women, Monetary
	organizations
Risk Tolerance	Risk and return policies, Safety,
	Liquidity, Capital appreciation, Entry
	and exit load
Economic Factors	Change in Economy, Excess Income,
	Market Risk
Political Factors	Changes in Government policies,
	Political Stability, Interest Rates, Tax
	Rebates
Other Factors	Brand, Price, Payment Policies,
	Quality Management, Settlement
	Practice, Awareness about the Product
	and the Services rendered

Table-2: FACTORS AFFECTING INDIVIDUAL INVESTORS DECISION MAKING PROCESS

Table-3 illustrates the statistical data of demographic characteristics such as Age, Gender, Marital Status, Education, Occupation and Annual Income that affect on investment decision making process.

Percent, Frequency calculation and Chi-square test are used for the data analysis and data interpretation. Knowledge of the investors' regarding the investments is as shown in Table-4. From the table we can observe that 12.3 % respondents have extensive knowledge about the investments, 61.5 % respondents are having moderate knowledge and 23 % have limited knowledge about the market conditions, benefits and loss of investing in shares, mutual funds and bonds and only. Hence knowledge can be considered as a key factor in individual investments. This will help the investors' to take the decision based on the present market conditions either of their own or in consultation with the fund manager.

Demographic	Variables	Count	Percentage
Characteristics			
Age	a) Below 30	20	30.8
	b) Between 31-40	29	44.6
	c) Between 41-50	10	15.4
	d) Between 51-60	3	4.6
	e) Above 61	3	4.6
Gender	a) Male	45	69.2
	b) Female	40	30.8
Education	a) Undergraduate	1	1.5
	b) Graduate	24	36.9
	c) Post Graduate	35	53.8
	d) Professional Degree	5	7.7
Marital Status	a) Married	50	76.9
	b) Single	15	23
Occupation	a) Self-Employed	3	4.6
	b) Retired	4	6.2



	c) Student	1	1.5
	d) Salaried	57	87.7
Annual Income	a) Below 1,00,000	1	1.5
	b) 1,00,001 - 3,00,000	10	15.4
	c) 3,00,001 - 5,00,000	12	18.5
	d) 5,00,001 - 10,00,000	26	40
	e) 10,00,001 - 15,00,000	14	21.5
	f) Above 15,00,001	2	3.1

Table-3: DEMOGRAPHIC CHARACTERISTICS

Knowledge	Frequency	Percent	Cumulative Percent
Extensive	8	12.3	12.3
Moderate	40	61.5	73.8
Limited	15	23	96.9
No knowledge	2	3	100

Table-4: KNOWLEDGE ABOUT INVESTMENTS IN SHARES, MUTUAL FUNDS AND BONDS

Figure-1 shows that around 74 % respondents are ready to take little or more risk with some money they are having, around 17% respondents are not at all ready to bear any kind of risk, and 9.2% are ready to take little risk with all money. As investors are making provisions for the future financial requirements, hence not ready to choose risky investment portfolio. Risk bearing capacity of individual investment totally depends on expected returns on investments.



Fig -1 Attitude towards Risk

In our study, we have focused on Shares, Bonds and Mutual Funds as investment avenues. As investments are meant to financial growth, safety is very much important. Financial market is quite a volatile market and hence our study involves the investments in shares, mutual funds and bonds. Figure 2



e-ISSN: 2394-8299 p-ISSN: 2394-8280

shows the investments in shares. From the figure we can observe that around 34 % investors' prefer investments in shares. Figure 3 illustrates the investments in Bonds. 35% prefer investment in Bonds and figure 4 shows the investments in Mutual Funds. 38% prefer investments in Mutual Funds. It is observed that investments in share market is quite risky but can get good returns on the investments. But investors' like to be in safe zone and this is the reason why they prefer investments in mutual funds.



Fig -2: Investments in Bonds





Fig -3: Investments in Mutual Funds

While investing in any avenue withdrawal time is also an important factor. Table-5 shows the frequency distribution of individual investor's withdrawal. Majority of the investors' are interested in investments in between 3 to 5 years. Some investors' are also interested in long term investments.

Time Duration	Frequency	Percent	Cumulative
			Percent
Less than 3 years	3	4.6	4.6
3 to 5 years	30	46.2	50.8
6 to 10 years	14	21.5	72.3
More than 10 years	18	27.7	100

TABLE-5: WITHDRAWAL PLAN

4. CONCLUSION

Diversified financial service sector has given wide range of opportunities to individual investors'. The individual's decision of investment is prejudiced by the kind of services rendered and the benefits offered in the financial market. Financial knowledge and experience has an impact in financial investment decision making process.

The present study focuses on investment pattern of individual investors' of Pune city by considering the factors like objective of the investment, risk bearing capacity, composition of the portfolio, knowledge they have and time horizon of the investment. Investment decisions are highly dependent on demographic, personal, psychosomatic characteristics, ethics, risk tolerance, economic factors etc. Depending on the risk bearing capacity, investors' are classified as Defensive, Conservative, Balanced, Moderate, Aggressive and Very Aggressive. Reason behind the investments, expected returns, time duration and the type of the investors' are some of the factors that help individual investors' or their fund managers to manage funds effectively.

From the study we can observe that 61.5% participants have moderate knowledge of the investment portfolio management. Attitude towards risk is also studied. 74% respondents are ready to take little or more risk with some money they are having whereas 17% respondents are not at all ready to bear any kind of risk. Withdrawal plan matters a lot when return on investment is expected. 50.8% investors' are comfortable in 3 to 5 years of investments and 72% are comfortable in 6 to 10 years of



investments. Long term investments are less preferred by the respondents. As respondents are ready to take moderate risk hence are interested in Mutual Funds, Bonds and sometimes in Shares.

ACKNOWLEDGEMENT

We would like to thank respondents/participants for providing honest and valuable information. We appreciate their time and input.

REFERENCES

[1] D. Kiran and U. Rao, "Identifying investor group segments based on demographic and psychographic characteristics," in 8th Capital Markets Conference, Indian Institute of Capital Markets Paper, 2005.

[2] R. A. Nagy and R. W. Obenberger, "Factors influencing individual investor behavior," Financial Analysts Journal, pp. 63–68, 1994.

[3] L. Gupta, C. P. Gupta, and N. Jain, Indian Households' Investment Preferences: With Special Reference to Debt Market Instruments Based on the 3rd All India Household Investors Survey. Society for Capital Market Research and Development, 2001.

[4] A. A. Merikas, A. G. Merikas, G. S. Vozikis, and D. Prasad, "Economic factors and individual investor behavior: The case of the greek stock exchange," Journal of Applied Business Research (JABR), vol. 20, no. 4, 2011.

[5] L. Feng and M. S. Seasholes, "Individual investors and gender similarities in an emerging stock market," Pacific-Basin Finance Journal, vol. 16, no. 1, pp. 44–60, 2008.

[6] N. Walia and R. Kiran, "An analysis of investors risk perception towards mutual funds services," International Journal of business and Management, vol. 4, no. 5, p. P106, 2009.

[7] M. Hood, J. R. Nofsinger, and A. Varma, "Conservation, discrimination, and salvation: Investors social concerns in the stock market," Journal of Financial Services Research, vol. 45, no. 1, pp. 5–37, 2014.

[8] S. Lodhi, "Factors influencing individual investor behaviour: An empirical study of city karachi," Journal of Business and Management, vol. 16, no. 2, pp. 68–76, 2014.

Smart Innovation, Systems and Technologies 77

Suresh Chandra Satapathy Vikrant Bhateja -Swagatam Das *Editors*



Smart Computing and Informatics

Proceedings of the First International Conference on SCI 2016, Volume 1





Smart Innovation, Systems and Technologies

Volume 77

Series editors

Robert James Howlett, Bournemouth University and KES International, Shoreham-by-sea, UK e-mail: rjhowlett@kesinternational.org

Lakhmi C. Jain, University of Canberra, Canberra, Australia; Bournemouth University, UK; KES International, UK e-mails: jainlc2002@yahoo.co.uk; Lakhmi.Jain@canberra.edu.au

Scale Space-based Capillary Detection to Determine Capillary Density and Avascular in Nailfold Capillary Images Using USB Digital Microscope	603
H. S. Ajaya, H. R. Shreyas, Vikram Manikandan, K. V. Suma and Bheemsain Rao	005
Augmenting Non-verbal Communication Using a Tangible User Interface	613
An Approach to En Route Environmentally Sustainable Future Through Green Computing Bhubaneswari Bisoyi and Biswajit Das	621
Energy Aware Task Scheduling Algorithms in Cloud Environment:	
A Survey Debojyoti Hazra, Asmita Roy, Sadip Midya and Koushik Majumder	631
Design of Jitter Spectral Shaping as Robust with Various Oversampling Techniques in OFDM Syed Umar, N. Priya, P. Gayathri, T. Subba Reddy and Azeem Mohammed Abdul	641
Architecture for the Strategy-Planning Techniques Using Big DataAnalyticsPuja Shrivastava, Laxman Sahoo and Manjusha Pandey	649
Indian Classical Dance Mudra Classification Using HOG Features andSVM ClassifierK. V. V. Kumar and P. V. V. Kishore	659
Feature Extraction Model for Social Images Seema Wazarkar and Bettahally N. Keshavamurthy	669
Study of Dimensionality Reduction Techniques for Effective Investment Portfolio Data Management Swapnaja Gadre-Patwardhan, Vivek Katdare and Manish Joshi	679
Automatic Text Recognition Using Difference Ratio	691
Sign Language Conversion Tool (SLCTooL) Between 30 World Sign Languages	701
A. S. C. S. Sastry, P. V. V. Kishore, D. Anil Kumar and E. Kiran Kumar	

Study of Dimensionality Reduction Techniques for Effective Investment Portfolio Data Management

Swapnaja Gadre-Patwardhan, Vivek Katdare and Manish Joshi

Abstract The aim of dimensionality reduction is to depict meaningful low-dimensional data of high-dimensional data set. Several new nonlinear methods have been proposed for last many years. But the question of their assessment is still open for the study. Dimensionality reduction is the vital problem in supervised and unsupervised learning. For high-dimensional data, computation becomes heavy if no pre-processing is done before supplying it to any of the classifiers. Because of the constraints like memory and speed, it is not suitable for certain practical applications. As per the method of attribute selection process, attribute sets are provided as an input to the classifier. The attributes that incorrectly classified are supposed to be irrelevant and are removed by obtaining the subset of selected attributes. Thus, accuracy of the classifier is improved, and time is also reduced. Attribute evaluators such as cfsSubset evaluator, information gain ranking filter, chi-squared ranking filter and gain ration feature evaluator are used for the classifiers viz. decision table, decision stump, J48, random forest. Individual investor's investment portfolio data is used for the present study. Twenty-six attributes are obtained from the questionnaire. By applying dimensionality reduction techniques, five major attributes are obtained using information gain ranking filter, chi-squared ranking filter, gain ratio feature evaluation and seven attributes using cfsSubset evaluator. Around 70.7692% accuracy is obtained using three attribute evaluators for all five classification algorithms, whereas cfsSubset evaluator along with random forest classifier gives 81.5385% accuracy. It has been observed that cfsSubset

S. Gadre-Patwardhan (🖂)

MES's Institute of Management and Career Courses, Pune, India e-mail: swapnaja.gadre24@gmail.com

V. Katdare

KCESs Institute of Management and Research, North Maharashtra University, Jalgaon, India e-mail: vvkatdare@rediffmail.com

M. Joshi

School of Computer Sciences, North Maharashtra University, Jalgaon, India e-mail: joshmanish@gmail.com

© Springer Nature Singapore Pte Ltd. 2018

679

S. C. Satapathy et al. (eds.), *Smart Computing and Informatics*, Smart Innovation, Systems and Technologies 77, https://doi.org/10.1007/978-981-10-5544-7_67

evaluator with partition membership as a pre-processing technique and random forest as classification algorithm performs reasonably better in terms of accuracy.

Keywords Dimensionality reduction • Classification • cfsSubset evaluator Random forest • Chi-squared ranking filter • Information gain ranking filter

1 Introduction

With the technological advancement, information technology is used in almost all aspects of daily lives. A huge amount of data is generated such as financial and commercial data, medical data, climate patterns, agricultural and much more. The data is accumulated in several databases and data warehouses. Most of these data have various attributes which are spread over the high-dimensional space. People working with these regularly face the problem of dimensionality reduction which is the process of random variable reduction under consideration for obtaining principal variables [3]. Study of dimensionality reduction depicts meaningful low-dimensional data of high-dimensional data set as shown in Fig. 1. There are two aspects of dimensionality reduction: (1) feature selection and (2) feature extraction. Feature selection is the active and promising research areas in the fields like statistics, data mining, machine learning and pattern recognition. The objective of the feature selection is to determine a subset of input data items by eliminating irrelevant data elements. Feature extraction transforms the data of high-dimensional space to low-dimensional space, thereby reducing the number of attributes [6]. For high-dimensional data, computation becomes heavy if no pre-processing is done before supplying it to any of the classifier. Because of the constraints like memory and speed, it is not suitable for certain practical applications. As per the results of attribute selection methods, attribute sets are provided as an input to the classifier. The attributes that incorrectly classified are supposed to be irrelevant and are removed by obtaining the subset of selected attributes. Attribute evaluators such as

Fig. 1 Dimensionality reduction



cfsSubset evaluator, information gain ranking filter, chi-squared ranking filter and gain ratio feature evaluator are used for reduction of dimension.

Individual investor's portfolio data is used for the present study. Decision of the investment is dependent on some demographic characteristics. The total risk associated with the investment portfolio can be calculated or predicted on the basis of the investment amount, the purpose of the investment returns expected from the investment and the period of the investment [8, 9].

The paper is organized as—Sect. 2 deals with related work. Literature related to dimensionality reduction is studied under this section. Section 3 contains the information about the data set generated and used. Section 4 is based on experimental work. Comparative analysis of attribute selection methods is performed in this. In Sect. 5, results and observations are discussed, and Sect. 6 talks about the conclusion of the present dimensionality reduction study.

2 Related Work

Dimensionality reduction techniques are classified as supervised and unsupervised techniques on the basis of learning process. Supervised algorithms require a training data set with the class label data to learn the lower dimensional representation of data based on some criteria and then predict the class labels for unknown data. Unsupervised approaches project the original data set to a new lower dimensional space without using the label information.

Dimensionality reduction techniques operate either by selecting the subset of the existing data or by transforming the existing features to a reduced data set. Several new nonlinear methods have been proposed for last many years. But the question of their assessment is still open for the study. Since from a long time, researchers are exploring the area of dimensionality reduction in different domains for many years. Stating below the study of some researchers.

Fu et al. [2] have applied dimensionality reduction for simplifying radial basis function (RBF) network structure to improve the classification performance. They have also proposed a novel method, i.e. separability correlation measure (SCM) for ranking the importance of the attributes. In comparison of existing methods such as SUD and Relief-F methods, SCM points to smaller attribute subsets and higher classification accuracies in simulation. They also proposed a better modification for RBF network construction and training by allowing cluster overlap of the same class.

Wang et al. [13] used feature extraction and dimensionality reduction techniques for vowel recognition. They investigated minimum classification error (MCE) for feature extraction. A generalized MCE (GMCE) has been proposed to overcome the shortcoming of MCE algorithm. Linear discriminant analysis (LDA), principal component analysis (PCA), MCE and GMCE are applied for feature extraction. They also investigated support vector machine (SVM) and compared it with linear feature extraction method. They observed that linearly extracted models are fit for training data, whereas SVM has improved generalized properties. Esser et al. [1] have demonstrated a convex model for non-negative matrix factorization and dimensionality reduction on physical space. They have successfully applied it to hyper-spectral end-member detection and separation of blind source in NMR. A model that can handle outliers is also proposed by them.

Joseph et al. [7] have performed dimensionality reduction and classification on hyper-spectral image. The objective was to find and classify the constituent materials for every pixel in a scene and reduce the dimensionality or volume without losing any critical information. Joseph et al. have described a technique that reduces data dimensions simultaneously and suppresses unwanted signature of interest. The main concept was to project each pixel vector on an orthogonal subspace to the undesired signatures. This is an optimal interference suppression process. Using this process signal-to-noise ratio was maximized and resulted in single component image. Orthogonal subspace projection (OSP) operator can handle k signatures, thereby simultaneously classifying and reducing dimensionality.

Kumar [11] studied unsupervised dimensionality reduction techniques for text data retrieval. The analysis was done on the basis of retrieval quality, complexity and approximation error. He concluded that semantic space obtained by singular value decomposition (SVD) and fuzzy k-means (FKM) produced better results as compared to other dimensionality techniques.

Kaushik et al. [10] have performed dimensionality reduction for indexing large time series database. They studied various dimensionality reduction techniques including singular value decomposition (SVD), discrete wavelet transform (DWT) and discrete Fourier transform (DFT). A new technique viz. adaptive piecewise constant approximation (APCA) is proposed by them. They have used two distant measures for fast searching: (1) a lower bounding Euclidean distance approximation and (2) a non-lower bounding and found that APCA is superior to other techniques.

John et al. [12] examined the quality of dimensionality reduction using rank-based criteria. They reviewed some quality measures which were based on k-ray neighbourhood and distance ranking. According to their observation co-ranking matrix can be used for rank comparison for the initial data set and also for some low-dimensional embedding. After that, rank errors and concepts can be associated with various blocks of the co-ranking matrix. They used real and synthetic data for experiments. They concluded that co-ranking matrix is better tool for simple and discriminatory quality criteria with the help of straightforward interpretation.

Amir et al. [4] introduced an information theoretic nonlinear method for informative dimensionality reduction. They selected continuous feature functions from the co-occurrence matrix. The algorithm proposed by them is analogous to the association analysis used in statistics with the help of feature extraction. They used synthetic co-occurrence data for experimental analysis and observed performance improvement in original feature set.

Zhang et al. [15] applied dimensionality reduction technique on hyper-spectral imagery based on clonal selection. The clonal selection was used to describe the basic features. They developed feature weighting algorithm, clonal selection feature selection (CSFS) algorithm, a feature subset search algorithm and clonal selection

feature weighting (CSFW) algorithm. Experimental analysis demonstrated that CSFS and CSFW are much better than other algorithms, and hence found effective for dimensionality reduction of hyper-spectral remote sensing imagery.

Kilian et al. [14] introduced nonlinear dimensionality reduction using maximum variance unfolding. They reviewed maximum variance unfolding algorithm for low-dimensional representation of high-dimensional data. They concluded that algorithm is based on modern tools and found applicable in the field of machine learning.

3 Data Set

The present study is an attempt to understand the behaviour of individual investor in various investment avenues. The investment decision-making depends upon many factors. The objective of the present study is to identify the factors that are affecting on individual investors decision-making process.

3.1 Study Sample

The population for the present study comprises of the investors from Pune (Maharashtra, India) region. Investors from all age group, education level, occupation and income group are considered.

3.2 Data Collection Method

Questionnaire is designed to collect the data from the investors. Overall, there are 26 questions in the questionnaire that are divided in two sections. It consists of two sections as—Section I personal information and section II includes the questions about the portfolio composition and risk-bearing capacity of an individual investor. Section II is divided into six axes viz. first axis: investment objectives, second axis: time horizon, third axis: level of satisfaction, fourth axis: factors influencing investments, fifth axis: knowledge and sixth axis: risk tolerance.

4 Experimental Work

Due to a large amount of data elements in data sets, we applied some filtering approach for feature selection. Attribute selection methods allow us to extract the subset of data elements from the high-dimensional data set, thereby improving the computation accuracy and efficiency. Principal component analysis, factor analysis, backward feature elimination are some of the popular techniques of dimensionality reduction. For our study, we have used cfsSubset evaluator, information gain ranking filter, chi-squared ranking filter and gain ratio for dimensionality reduction.

Information gain ratio is the ratio of information gain to the intrinsic information. When an attribute X splits the set S into subsets S_i , average entropy is calculated, and the sum of the entropy is compared with the original set S. The attributes that maximize the difference are selected.

Chi-squared ranking filter is the statistical method that measures the closeness of actual and expected result. In this method, it is assumed that variables are random and drawn from sufficient sample of independent variable. The result indicates the difference between the actual and expected outcomes. Each feature is independently evaluated with reference to the class label. The larger the value of chi-square, features are more relevant to the class model.

Gain ratio is the modified version of information gain. Information gain selects the features with large number of values, whereas gain ratio maximizes the features information gain by minimizing its value.

Correlation-based feature selection (CFS) subset evaluator uses the search algorithm together with a function to evaluate the merits of the feature subset. The method by which CFS measures the usefulness of feature subset considers the importance of the individual feature for class label prediction along with the level of inter-correlation between them [5].

Diagrammatic representation of the dimensionality reduction process used in the study is shown in Fig. 2.

For dimensionality reduction process, WEKA attribute selection techniques such as information gain ranking filter, chi-squared ranking filter, gain ratio feature evaluation and cfsSubset evaluator are demonstrated. After applying dimensionality reduction techniques, we obtained five attributes in information gain ranking filter, chi-squared ranking filter, gain ratio feature evaluation and seven attributes in cfsSubset evaluator. These techniques are applied on the classifiers like decision tree, decision stump, J48, random forest and logistic model tree (LMT). Comparative study is done for all these classifiers. The accuracy of the classifier can be calculated as follows:

$$\operatorname{accuracy} = (\operatorname{tp} + \operatorname{tn})/(\operatorname{tp} + \operatorname{tn} + \operatorname{fp} + \operatorname{fn})$$
(1)



Fig. 2 Dimensionality reduction process

and precision by following formula

$$precision = tp/(tp + fp)$$
(2)

5 Result and Observations

On the selected five attributes using information gain ranking filter, chi-squared ranking filter, gain ratio feature evaluation techniques and seven attributes using cfsSubset evaluator some pre-processing is done. Pre-processing techniques as follows:

Class-order alters the order of the classes. The values are arranged in the order specified by the user, i.e. either ascending or descending or in random order.

Discretise discretises the numeric attribute data set to nominal attributes.

Partition Generator partitions the membership value.

Class condition converts the values of numeric and/or nominal attributes to class conditional probabilities.

Analysis of all these dimensionality reduction methods along with the pre-processing techniques is as below.

Figure 3 represents the accuracy obtained by using decision tree, decision stump, LMT, J48 and random forest classifiers. By applying these classifiers on selected five attributes, we can observe that LMT gives 69.2308% accuracy, i.e. 45 instances are correctly classified.

Figure 4 represents the graphical view of correctly classified and incorrectly classified instances using decision table, decision stump, LMT, J48 and random forest classifiers with the help of pre-processing technique as class conditional probability. From the graphical representation, it can be seen that LMT gives the 69.2308% accuracy. Thus, we can say that with pre-processing technique as class conditional probabilities min values, maximum obtained accuracy remains the same.







With the help of class-order pre-processing technique along with LMT classifier, 64.6154% accuracy is obtained, i.e. 42 instances are correctly classified as shown in Fig. 5 and whereas from Fig. 6 we can see that Random Forest classifier gives 66.1538% accuracy which means 43 instances are correctly classified.

By applying discretise pre-processing technique, LMT classifier performs better. Using all the techniques, we obtained 67.6923% accuracy which maximum of all. The results are illustrated in Tables 1 and 2.

Classifier used	Decision table	Decision stump	LMT	J48	Random forest
Correctly classified instances	66.1538%	60%	67.6923%	66.1538%	66.1538%
Incorrectly classified instances	33.8462%	40%	32.3077%	33.8462%	33.8462%
Kappa statistics	0.4758	0.3749	0.5051	0.4758	0.4838
MAE	0.1904	0.1813	0.1682	0.1631	0.1632
RMSE	0.2979	0.307	0.2948	0.2999	0.3003
RAE	83.0547	79.0818	73.3937	71.1426	71.1870
RESE	88.7312	91.4458	87.7988	89.3214	89.4303
Cov of clas	93.8462	90.7692	92.3077	87.6923	87.6923
MRRS	85.1282	60.5128	65.6410	45.8974	46.4103
Total inst.	65	65	65	65	65

Table 1 Five attributes using discretise

 Table 2
 Seven using discretise

Classifier used	Decision table	Decision stump	LMT	J48	Random forest
Correctly classified instances	66.1538%	60%	67.6923%	66.1538%	66.1538%
Incorrectly classified instances	33.8462%	40%	32.3077%	33.8462%	33.8462%
Kappa statistics	0.4758	0.3749	0.5051	0.4758	0.4838
MAE	0.1904	0.1813	0.1682	0.1631	0.1631
RMSE	0.2979	0.307	0.2948	0.2999	0.3005
RAE	83.0547	79.0818	73.3937	71.1426	71.1688
RRSE	88.7312	91.4458	87.7988	89.3214	89.4917
Cov of clas	93.8462	90.7692	92.3077	87.6923	87.6923
MRRS	85.1282	60.5128	65.6410	45.8974	46.1538
Total ins	65	65	65	65	65

With the help of partition membership pre-processing technique and by applying information gain ranking filter, chi-squared ranking filter and gain ratio feature evaluation techniques, maximum 70.7692% accuracy is obtained as shown in Table 3.

On the other hand with the help of partition membership pre-processing technique along with cfsSubset evaluator as a dimensionality reduction technique, 81.5385% accuracy is obtained which is the maximum and is shown in Table 4.

Classifier used	Decision table	Decision stump	LMT	J48	Random forest
Correctly classified instances	70.7692%	56.9231%	70.7692%	70.7692%	70.7692%
Incorrectly classified instances	29.2308%	43.0769%	29.2308%	29.2308%	29.2308%
Kappa statistics	0.5489	0.3211	0.5489	0.5489	0.5489
MAE	0.1871	0.187	0.162	0.1604	0.1611
RMSE	0.2936	0.317	0.2862	0.2924	0.2925
RAE	81.6124	81.5680	70.6509	69.9638	70.2857
RESE	87.4384	94.4045	85.2572	87.0882	87.1167
Cov of clas	95.3846	90.7692	96.9231	89.2308	89.2308
MRRS	89.4872	59.4872	83.3333	61.2821	61.2821
Total inst	65	65	65	65	65

Table 3 Five using partition membership

Table 4 Seven attributes using partition membership

Classifier used	Decision table	Decision stump	LMT	J48	Random forest
Correctly classified instances	63.0769%	56.9231%	61.5385%	75.3846%	81.5385%
Incorrectly classified instances	36.9231%	43.0769%	38.4615%	24.6154%	18.4615%
Kappa statistics	0.4413	0.3211	0.4153	0.6077	0.7143
MAE	0.1846	0.187	0.1662	0.1074	0.096
RMSE	0.2933	0.317	0.2928	0.2618	0.236
RAE	80.5680	81.5680	72.5043	46.8648	41.8794
RESE	87.3479	94.4045	87.2007	77.9908	70.2879
Cov of clas	100	90.7692	96.9231	93.8462	90.7692
MRRS	85.6410	59.4872	51.7949	38.2051	31.7949
Total inst	65	65	65	65	65

6 Conclusion

Dimensionality reduction is a vital problem in supervised and unsupervised learning. In this paper, we studied feature selection, feature extraction and classification algorithms for dimensionality reduction. Five classification algorithms are taken into consideration. Individual investor's investment portfolio data is used for the present study. Data is collected through questionnaire, and it is observed that some demographic characteristics affect upon the individual investors investment decision-making process. In all, we have got 26 attributes through the questionnaire. By applying dimensionality reduction techniques, we obtained five major attributes using information gain ranking filter, chi-squared ranking filter, gain ratio

feature evaluation and seven attributes using cfsSubset evaluator. Experimental results show that there is no significant difference in the results obtained by information gain ranking filter, chi-squared ranking filter, gain ratio feature evaluation attribute evaluators. Around 70.7692% accuracy is obtained using three attribute evaluators for all five classification algorithms, whereas cfsSubset evaluator along with random forest classifier gives 81.5385% accuracy. It has been observed that cfsSubset evaluator with partition membership as a pre-processing technique and random forest as classification algorithm performs reasonably better in terms of accuracy.

Declaration Authors have obtained permission to use the data from the investors. Authors take full responsibilities to bear any consequences if any issues arise due to this. Publisher or Editors are not responsible for this.

References

- Esser, E., Moller, M., Osher, S., Sapiro, G., Xin, J.: A convex model for nonnegative matrix factorization and dimensionality reduction on physical space. IEEE Trans. Image Process. 21 (7), 3239–3252 (2012)
- Fu, X., Wang, L.: Data dimensionality reduction with application to simplifying RBF network structure and improving classification performance. IEEE Trans. Syst. Man Cybern. B (Cybern.) 33(3), 399–409 (2003)
- Geng, X., Zhan, D.-C., Zhou, Z.-H.: Supervised nonlinear dimensionality reduction for visualization and classification. IEEE Trans. Syst. Man Cybern. B (Cybern.) 35(6), 1098–1107 (2005)
- Globerson, A., Tishby, N.: Sufficient dimensionality reduction. J. Mach. Learn. Res. 3, 1307–1331 (2003)
- 5. Hall, M.A., Smith, L.A.: Feature subset selection: a correlation based filter approach (1997)
- 6. Han, J., Pei, J., Kamber, M.: Data Mining: Concepts and Techniques. Elsevier, Amsterdam (2011)
- Harsanyi, J.C., Chang, C.-I.: Hyperspectral image classification and dimensionality reduction: an orthogonal subspace projection approach. IEEE Trans. Geosci. Remote Sens. 32(4), 779– 785 (1994)
- Hoffmann, A.O.I., Post, T., Pennings, J.M.E.: Individual investor perceptions and behavior during the financial crisis. J. Bank. Finance 37(1), 60–74 (2013)
- 9. Kaur, M., Vohra, T.: Understanding individual investor's behavior: a review of empirical evidences. Pac. Bus. Int. 5(6), 10 (2012)
- Keogh, E., Chakrabarti, K., Pazzani, M., Mehrotra, S.: Locally adaptive dimensionality reduction for indexing large time series databases. ACM SIGMOD Rec. 30(2), 151–162 (2001)
- Kumar, A.C.: Analysis of unsupervised dimensionality reduction techniques. Comput. Sci. Inf. Syst. 6(2), 217–227 (2009)
- Lee, J.A., Verleysen, M.: Quality assessment of dimensionality reduction: rank-based criteria. Neurocomputing 72(7), 1431–1443 (2009)
- Wang, X., Paliwal, K.K.: Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition. Pattern Recogn. 36(10), 2429–2439 (2003)
- 14. Weinberger, K.Q., Saul, L.K.: An introduction to nonlinear dimensionality reduction by maximum variance unfolding. In: AAAI, vol. 6, pp. 1683–1686 (2006)
- Zhang, L., Zhong, Y., Huang, B., Gong, J., Li, P.: Dimensionality reduction based on clonal selection for hyperspectral imagery. IEEE Trans. Geosci. Remote Sens. 45(12), 4172–4186 (2007)





CONDUCTING ICT RESEARCH IN VOLUNTARY ORGANIZATIONS: CHALLENGES AND DIRECTIONS

Ashwini Patil¹, Dr. R. M. Patil²

¹Research Scholar, Savitribai Phule Pune University, Pune 411007 ¹asp.imcc@mespune.in

²Dr. D. Y. Patil Instituite of Management and Research, Pune 411018, Maharashtra, India ²patil.rm10@gmail.com

ABSTRACT:

As the roles and functions of voluntary organizations have significantly expanded in recent years, there is a growing concern over the need to transform the operation and structure of these organizations. ICT support in voluntary organizations is an interesting emergent field of research. Recent literature has highlighted that many voluntary organizations are still in early stage of ICT adoption, in their organizational settings. Due to the social importance of this sector, it is important that universities and researchers carry on action research projects in such settings to improve ICT usage. These organizations are quite diverse in their organizational structure, scope, application area and working. Furthermore, they differ from traditional organizations in their objectives, rationale, operations and work practices. Appropriate technology design requires a deep understanding of organizational work practices, paving the way for ethnographic action research studies in these settings. The paper reviews the related work and discusses important issues which are worth

The paper reviews the related work and discusses important issues which are worth investigating for the improvement of ICT support in voluntary organizations.

Keywords: Information and Communication Technology, Voluntary Organization, NGOs, ICT support, Knowledge Transfer

[I] INTRODUCTION

The role of voluntary organizations has become more crucial in accelerating the economic and social development in any country. Therefore, a number of voluntary organizations have participated in social welfare, community construction, and other public service activities as an effort to form a new cooperative relationship with the government. [1] There is awareness among community organizations to use technology in their activities but the complexity of technologies and lack of technological knowledge are big hurdles. Voluntary organizations are facing a challenge in deciding which technologies are more suitable for their organization. Compared to the governmental organizations, voluntary organizations are less bureaucratic. They are highly prepared to "speak and act on behalf of the chosen causes". [2] They are very much concerned about the daily lives of people their problems instead of technology they are using.

Voluntary organizations are still in early stage of ICT adoption, in their organizational settings. There is a need of information technology services and applications that can be effectively embedded in organizational settings of voluntary organizations to achieve technological appropriation. ICT support in voluntary organizations is an interesting emergent field of research. There is awareness among community organizations to use technology in their activities but the complexity of technologies and lack of technological knowledge are big obstacles. [3] Voluntary organizations are isolated from the world of advanced technology and thus face a challenge in deciding which technologies are more suitable for their organization. The findings, moreover, indicated that most of the voluntary organizations own very basic ICT equipment and limited hardware/software essential to perform efficient tasks. On the other hand, while some voluntary organizations' senior managers and board members adopted more-advanced ICT, they are still not comfortable with their investment in the ICT, as the connections between the adopted ICT and the organizations' missions and objectives are still not clear. Based on the literature, the use of ICT should be aligned with the goal and mission of each organization in order to add value to the organization's activities and gain a competitive advantage. In this context the appropriation of ICT in the organizational settings of voluntary organizations seems to be an important research area.

[II] CHALLENGES IN ICT USE

1. ICT Investments

Although nearly all voluntary sector organizations use some form of low-cost, commercially available technology as an integral part of daily operations at both the central office and the field level, NGOs do not receive grants for infrastructure investment. Donors would rather see funding spent on direct aid to victims. Due to inadequate infrastructure, NGOs ultimately depend on other organizations for assistance with technology and communications. For example, most NGOs are unable to host their own Internet service in-country and sometimes cannot even manage their own area-wide radio networks. [4]

Thus very few NGOs are in a position to procure, implement, and sustain more advanced systems on an organization-wide basis. They are under constant pressure to keep overhead costs down.

2. Lack of Standards

Although many forms of communications technology are manufactured to industry specifications and standards, there is virtually no standardization among NGOs as to what types of ICT they use or how and when those technologies are used. [4] There are lot of differences in data; systems, styles, data management capacities, communication infrastructure.

3. Organizational Structure

Organizational structure of NGOs and level of professionalism of staff are other important dimensions. Small NGOs normally don't have a stable defined organizational structure. This results in lack of organizational knowledge and inconsistency in decision making, posing difficulty in establishing ICT infrastructures in the NGO sector. Secondly, NGOs normally do not have formal requirements of a specific profile to become a member. So, volunteer will not possess any specific skill set or profile. This typically results in low emphasis on developing ICT infrastructures and using IT capabilities to perform their tasks. [5]

[III] RESEARCH OBJECTIVE

The objective of the research is to analyse the impact of ICT applications on the work practices of voluntary organizations and to identify design guidelines for future systems.

Major Research Questions:

- 1. How do Voluntary Organizations use Information and Communication Technology for Information Management, Knowledge transfer, Communication etc.?
- 2. Which type of technology they use?
- 3. What are the issues and challenges in ICT implementation in voluntary organizations?
- 4. How ICT usage affects efficiency, productivity and performance of voluntary organizations.

[IV] STATE OF THE ART

Jaflah Al-Ammary, Sharifa Hamad describe the need to use IT to manage information, communicate with staff and volunteers, perform accounting, and handle various other tasks continues to grow. They suggest the IT framework for NGOs. [1] This framework illustrates five inhibitors for the adoption of IT by NGOs: lack of financial support, lack of awareness, culture, policies and regulation issues, and lack of IT-skilled volunteers. In addition, five enablers were identified for the adoption of IT. The main enabler is governmental support.

E. Burt, and J. Taylor says ICT could serve voluntary organizations in different ways; it could improve productivity, increase the overall effectiveness through better collaboration, and extend services to new communities in need Moreover, ICT can be a powerful force that opens exciting opportunities for voluntary organizations to achieve their missions and goals in an effective way. They suggest that that ICT plays a vital role in improving the quality and quantity of information, its potential for adoption and innovation is often uncertain. [3]

Authors focus on application of ICT, types of technology NGOs use, challenges in ICT use. Authors conclude that, and approach and strategy will need to be developed to promote expansion of technology and achieve uniformity of technology packages for NGOs.

Although the humanitarian sector has lagged behind other industries in adopting technology, NGOs are increasingly becoming more sophisticated in their use of technology—particularly wireless technology such as mobile phones—and are openly collaborating with a range of partners. To promote the expansion of technology and achieve uniformity of technology packages, some larger NGOs have formally joined consortiums such as NetHope. This type of organizing allows groups to pool resources and coordinate requirements in hopes of saving costs and maximizing benefits. Other NGOs are engaging in variety of partnerships with the private sector, academia, and other NGOs. Authors conclude that, and approach and strategy will need to be developed to promote expansion of technology and achieve uniformity of technology packages for NGOs. [4]

Saqib Saeed, Markus Rohde, and Volker Wulf highlights the importance of ICT support for NGOs and stresses the need for further research to empower this important sector with modern computing technologies. The discussion highlights sporadic efforts to embed information technology in NGOs. [5]

Using the language and concerns of sociology of technology and by focusing on sociotechnical dynamics of everyday ICTs practices, this research seeks to explore and conceptualize the ways such practices are constituted, maintained and re-produced by multiple actors and settings. This research aims to explore ICT-related practices within the specific context of VSOs through an interdisciplinary research, including Science and Technology Studies (contexts), Organization Studies (practices), Computer-Supported Collaboration Work (interactions), Information Systems Research (artifacts), and more importantly Social Informatics (dynamics). The broad research question that this study tackles is: "How everyday ICT practices are constructed, maintained and reshaped in smaller voluntary organizations?" [6]

Theodore E. Zorn, Andrew J. Flanagin, & Mirit Devorah Shoham contributes to the limited research on ICT adoption and use in the NPO sector. The study provides a more complex picture of ICT adoption processes and the complementary roles played by both organizational features and institutional forces. [7]

According to Saeed, S., Rohde, M. and Wulf, V., as these organizations work in vast geographical area the difference in culture, backgrounds, education, work practices, available infrastructure and needs, play a vital role in the acceptance of technology. A generic solution to all scenarios is not effective and these factors make technological appropriation much more complex. They are in the initial stages of starting a research on the usage of web 2.0 by different transnational NGOs in their work settings. [8]

Theoretical study was conducted to examine the role and functions of ICTs in the social welfare NGOs of Hong Kong. The findings were; the deployment and development of ICTs

by Hong Kong NGOs in the social welfare sector are still underway rather than completed and neither NGOs nor the government takes a full initiative to retrieve benefits from ICTs in the social welfare sector. The authors have given following suggestions.

1. ICTs can serve as a means to reengineer the operations and activities of NGOs that can go beyond managerial efficiency. It requires NGOs to take a proactive approach to the use of ICTs.

2. ICT-mediated applications can help NGOs disseminate the financial statement and accounting information at less cost for internal and public monitoring.

3. It is crucial to realize that training and education programs are essentially required for the recipients to fully reap the benefits.

4. Both governmental and individual NGOs should actively participate in publicizing the benefits of ICT usage

5. Much effort should be allotted to develop measurements to assess whether the adoption of ICT applications have actually brought about intended benefits. [9]

Denison, Tom, and Graeme Johanson conducted survey to fill an international gap in knowledge about the adoption of information and communication technologies by third sector organizations with a specific focus on the Australian experience. This paper summarizes findings of recently published surveys, compares their findings, and proposes the use of social network analysis as a more useful lens through which to consider current developments. They suggest that, the nature of networked interactions in relation to community sector organizations needs to be researched systematically, as does the role of networks in facilitating the development of organizational strategies and initiatives. [10]

To guide on policy and practice of organizational implementation of ICT within the voluntary/NGO sector, Ocean and Geoffrey piloted case study in UK and Uganda. This paper offers a new Technology Adoption Model (TAM) for dealing with the issues governing ICT adoption and the factors driving wider diffusion in voluntary and community sector organizations in the UK. The paper reports on website adoption process in 5 small and medium-sized voluntary sector organizations and identify Technology, Organizational and People (TOP) imperatives that provide new conceptual framework for facilitating organizational implementation of ICT.

The author's future scope of work is validation of TAM in Uganda and/or other developing countries. [11]

Burt, Eleanor, and John A, made an attempt to map the uptake and application of ICTs across a sample of UK voluntary organizations whose incomes range from £250,000 to over £11 million. Their objective was to stimulate ideas and debate regarding the application of ICTs within voluntary organizations, and to provide initial benchmarks for managers of voluntary organizations and others associated with the sector. The research method was questionnaire survey.

The research confirms the spread of computer networking in all its forms. The evidence reveals lower levels of uptake of core networking technologies such as external and internal email, portable computing, intranets and extranets, as well as significant disparity in their application to key business activities. Perceptions of the significance of ICTs are also higher among organizations employing IT development staff. [12]
[V] CONCLUSIONS

The objective of this paper is to highlight the importance of ICT support for voluntary organizations and stress the need for further research to empower this important sector with modern computing technologies. This discussion highlights sporadic efforts to embed information and communication technology in voluntary organizations. One central research question which needs to be answered is how ICT services and applications can be effectively embedded in organizational settings of transnational voluntary organizations to achieve technological appropriation. It needs to be investigated, that how funds can be utilized for ICT support, it is also worth investigating methods of standardization, how the organizational structure and application area affect the ICT requirements. There is a need to investigate how knowledge management technologies can overcome existing shortcomings and help in establishing organizational knowledge. There is a need for developing effective ICT applications.

These organizations also need to pay special attention to providing adequate training and support during implementation and use of ICT systems. Managers need to provide a facilitating environment where employees are encouraged to use new technologies.

We are confident that the findings from our studies will help in better understanding the ICT requirements of this sector and more researchers will focus towards helping voluntary organizations by customizing technology for them. In this way these organizations could be better prepared to carry on their roles.

REFERENCES

- Al-Ammary, Jaflah, and Sharifa Hamad. "INFORMATION TECHNOLOGY FOR ENHANCING NGOS'PERFORMANCE IN THE KINGDOM OF BAHRAIN."International Journal of Electronic Commerce Studies 3.1 (2012): 111-120.
- [2] United Nation Environment programme, Final report of Arab NGOs forum in preparation for the world submmit on sustainable development. Manama, Bahrain: WSSD, 2001.
- [3] Burt, Eleanor, and John Taylor. "New technologies, embedded values, and strategic change: Evidence from the UK voluntary sector." Nonprofit and voluntary sector quarterly 32.1 (2003): 115-127.
- [4] Cynthia Barrigan and Ben Hemingway, A Guide to NGOs for the Military, "NGO Use of Information and Communications Technology"
- [5] Saeed, Saqib, Markus Rohde, and Volker Wulf. "Designing IT Systems for NGOs: Issues and Directions." The Open Knowlege Society. A Computer Science and Information Systems Manifesto. Springer Berlin Heidelberg, 2008. 560-565.
- [6] Eshraghi, Ali. "ICT practices by voluntary groups: a multi-setting study."Proceedings of the 2013 conference on Computer supported cooperative work companion. ACM, 2013.
- [7] Zorn, Theodore E., Andrew J. Flanagin, and Mirit Devorah Shoham. "Institutional and noninstitutional influences on information and communication technology adoption and use among nonprofit organizations." Human Communication Research 37.1 (2011): 1-33.
- [8] Saeed, Saqib, Markus Rohde, and Volker Wulf. "A framework towards IT appropriation in voluntary organisations." International Journal of Knowledge and Learning 4.5 (2008): 438-451.

- [9] Choi, Jin-Wook. "Making ICTs Work: A Study of Hong Kong NGOs in the Social Welfare Sector."
- [10] Denison, Tom, and Graeme Johanson. "Surveys of the use of information and communications technologies by community-based organisations." The Journal of Community Informatics 3.2 (2007).
- [11] Ocen, Geoffrey. "Organisational Implementation of ICT: Findings from NGOs in the United Kingdom and Lessons for Developing Countries." Strengthening the Role of ICT in Development (2007): 230.
- [12] Burt, Eleanor, and John A. Taylor. "Information and communication technologies: reshaping voluntary organizations?." Nonprofit Management and Leadership 11.2 (2000): 131-143.





DATA MINING IN CRM

Jayashree Patil¹, Dr. R.M. Patil²

¹Research Scholar, Savitribai Phule Pune University, Pune, Maharshtra, India ²Dr. D.Y Patil Instituite of Management and Research, Pune Maharshtra, India

ABSTRACT:

Customer Relationship Management is crucial in any industry to improve their business and retain their customer. Banking sectors are promisingly uses CRM software. Use of IT in CRM enriches the data analysis and data visualization.

Effectiveness of CRM is proportionate to how the data is classified. Data mining tools play important role in data retrieval and classification. Appropriate classification can lead in proposing good strategies. This paper discusses the various data mining tools used in CRM and its impact on business development of bank.

Keywords: CRM, Banking Sector, Data mining

INTRODUCTION

Customer Relationship Management is important in service oriented industry like banking. Managing customer relationships requires managing customer knowledge. CRM and knowledge management are directed towards improving and continuously delivering good services to customers.

DATA MINING IN CRM

CRM is a co-ordinate approach in business to maintain the relationship between the firm and its customers to satisfy and retain the firm's customer. This will helps the firm to exist in business and to attract more customers by giving promotions and more comfort in doing business with the firm [12].

WHY CRM?

Some of the common objectives of implementing CRM are -

- > Increase the customer service in order to retain them
- Increasing efficiency of the organization which helps the employees to maximize their skills in understanding their customers
- > Reduce the cost for running business which in turn increases the profit
- Support the marketing department to understand the customer needs and make promotions based on the customers' needs [1].

The use of IT has created new ways for firms to exploit vast potentials of customer relationships that have never been exploited before. With growing competition from both traditional and online businesses, keeping customers satisfied, increasing potential sales, and maintaining customer loyalty become strategically important to business success.

Business Intelligence tools are used to assist a CRM system which focuses on decision support, customer service and target marketing.

BENEFITS OF CRM IN BANKING SECTOR

- A CRM system consists of a historical view and analysis of all the acquired or to be acquired customers. This helps in reduced searching and correlating customers and to foresee customer needs effectively and increase business.
- CRM contains each bit of details of a customer, hence it is very easy for track a customer accordingly and can be used to determine which customer can be profitable and which not.
- In CRM system, customers are grouped according to different aspects according to the type of business they do or according to physical location and are allocated to different customer managers often called as account managers. This helps in focusing and concentrating on each customer separately.
- A CRM system is not only used to deal with the existing customers but is also useful in acquiring new customers. The process first starts with identifying a customer and maintaining all the corresponding details into the CRM system. The Sales and Field representatives then try getting business out of these customers by sophistically following up with them and converting them into a winning deal. All this is very easily and efficiently done by an integrated CRM system.

The strongest aspect of Customer Relationship Management is that it is very cost-effective. The advantage of decently implemented CRM system is that there is very less need of paper and manual work, which requires lesser staff to manage and lesser resources to deal with. The technologies used in implementing a CRM system are also very cheap and smooth as compared to the traditional way of business. [10]

Related work

Sanjay Das perform Comparative study of SBI and other nationalized bank and give the suggestions for improvement of CRM. This study is significant to banks as they get information on what are items that are important to customers so as to maintain the relationship.[1]

Application of customer relationship management tool in business gives a new dimension. It proved beneficial but applying data mining in customer relationship management was further more beneficial. Researcher's main focus was on customer retention techniques to enhance our customer relationships via Data Mining. Different data mining techniques are used to retrieve a data from large database. Researcher put a conclusion as data mining would fasten up the process of searching large databases so as to extract customer buying patterns, to classify customers into groups which also make databases to be handled efficiently [4]

The collaborative CRM deals with the communication between bank and its clients, the operational CRM automates certain processes in banking, and the analytical CRM analyze the customer information and generates the business intelligence to operate in the competitive banking industry [1].

Gaurav Gupta, Himanshu Aggarwal discusses the CRM model with all 7 components in detailed. Researcher had thrown a light on what type of information about customer need to be collected for CRM model. The more effectively we can use the information about our customers to meet their needs the more profitable we will be. If CRM is integrated with data mining tools then we can improve the profitability of business. Operational CRM gives better result if we use analytical CRM with predictive mining tools. [9]

The following table describes the summary of implementation of CRM with some technologies and/or tools.

Author	Objective	Methodology	Result
Basar Oztays	Measure performance of CRM	Multiple Criteria Decision Making (MCDM)	Ranking among the alternatives are sensitive to changes in the parameters.
Siavash Emtiyaz	Propose a model to identify customer's behavior	Semi supervised Learning and Neural Network for data visualization	Data mining and CRM integration improves the profitability.
Yong Wang	combines CRM and the data mining tool, and constructs a decision tree suitable for the bank's customer classification by applying the ID3 algorithm to optimize the methods of classifying the bank's customers	ID3 algorithm for classification	Effectively improve service quality provided by the bank to target customers and reduce the operation cost for the bank
A.Kaleeswaran	CRM models responsible for receiving the request and responses to the customers quickly and directly. The working of CRM is simplified and enhanced with the use of different technology like data mining.	Apriori Algorithm	The CRM in commercial banks has become priority tool for banks to carry out various personalized services as a novel operating mechanism aiming at improving the relationship between banks and customers.

Table 1. Summary table CONCLUSION

This paper is an attempt to study the working of CRM. Using a secondary data it is observed that the CRM is widely used to improve the customer satisfaction and increase the customer retention. CRM process can give more effective results if the proper data get retrieved. For this data mining tools can be helpful. Data classification and data processing methods can differ according to queries.

FUTURE WORK

Most of the literature is available on CRM and its application in banking sector. In future the structure and functionality used in CRM can be studied and the appropriate data mining strategy can be identified based on the type of bank.

REFERENCES

- 1. Das, SANJAY KANTI. "Customer relationship management in banking sector: a comparative study of sbi and other nationalised commercial banks in India." A Journal of economics and management 1 (2012).
- Emtiyaz, Siavash, and MohammadReza Keyvanpour. "Customers behavior modeling by semisupervised learning in customer relationship management."arXiv preprint arXiv:1201.1670 (2012).
- Öztayşi, Başar, Tolga Kaya, and Cengiz Kahraman. "Performance comparison based on customer relationship management using analytic network process."Expert Systems with Applications 38.8 (2011): 9788-9798.
- 4. Farooqi, Md, and Khalid Raza. "A Comprehensive Study of CRM through Data Mining Techniques." arXiv preprint arXiv:1205.1126 (2012).
- 5. Wang, Yong, and Dongsheng Wu. "Research of the Bank's CRM Based on Data Mining Technology." Communications in Information Science and Management Engineering (2011).
- Kaleeswaran, A., and V. Ramasamy. "RULE-BASE DATA MINING SYSTEMS FOR CUSTOMER QUERIES." International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) 1.8 (2012): pp-282.
- ManjariTejpal Dr. Dhiraj Sharma." Customer Relationship Management: An Empirical Review." PARIPEX - INDIAN JOURNAL OF RESEARCH 2.7(2013)
- 8. Padhy, Neelamadhab, Dr Mishra, and Rasmita Panigrahi. "The survey of data mining applications and feature scope." arXiv preprint arXiv:1211.5723 (2012).
- 9. Gaurav Gupta, Himanshu Aggarwal. "Improving Customer Relationship Management using Data Mining." International Journal of Machine Learning and Computing.2.6(2012)
- Ramesh, K. "Role of Customer Relationship Management in Indian Banking System." International Journal of Applied Services Marketing Perspectives 2.4 (2014): 645-650.
- 11. Padhy, Neelamadhab, Dr Mishra, and Rasmita Panigrahi. "The survey of data mining applications and feature scope." arXiv preprint arXiv:1211.5723 (2012).
- 12. This little Piggy 2012, what is CRM? http://www.thislittlepiggy.co.uk/crm_definition.htm
- 13. Management Thought 2012, CRM in Banking, http://mansedition.blogspot.in/2012/05/crm-in-banking.http://mansedition.blogspot.in/2012/05/crm-in-banking.html





CURRENT ASPECTS IN HANDWRITING BIOMETRICS

Dr. Santosh Deshpande¹, Mukul Kulkarni²

¹Additional Director, MES'S Institute of Management and Career Courses Pune, India

²Research scholar, Savitribai Phule Pune University, Pune, India

ABSTRACT:

The handwriting biometrics is new branch for user identification and analysis. New techniques for understanding of user and authentication of user through online and offline databases using repository are emerging day by day. Traditional handwriting analysis system is being evolved using different techniques like graphometry, signature segmentation and verification by image processing techniques. Many neuroscientists, psychologies, handwriting experts are defining and redefining aspects of handwriting biometrics everyday This paper focuses on all such aspects in handwriting biometrics with their proposed outcomes in detail.

Keywords: Graphology, Trait and Gestalt method, Conscience, Pseudoscience, Personality, Brain writing, Neurological writing pattern

[1] INTRODUCTION

In the current era, human analysis is becoming important. For analysis user can observe special body movements, changes in voice tone, body gestures and posture, sometimes body structure and dressing sense etc. for interpretation of behavior and psychological understanding. Sometimes it's possible for masqueraded behavior or suppress behavior to be inaccurately analyzed. As certain fascial expression is associated with different moods like happy, sad etc. We can observe similar trends in handwriting of various writers. [1] [4]



Figure: 1. Physical Process of writing

Handwriting is conscious efforts made by writer by the instructions of brain. The process of handwriting is made on different levels consciously, unconsciously as well as due to nervous system and body movements. [9]. The actual writing process could be explained as follows: Every individual's handwriting depicts similar features like faces, body, fingers, voices, anger or warmth. Basically, that's the reflection of person with characteristics, special signature and writers state of mind. Major influence on writing is due to the emotional and circumstantial happening in and around the writer. Handwriting analysis is scientific method which reveals true personality traits. It is often referred as brain writing due to neurological brain pattern produced by neuromuscular movement creating writing strokes. It's basically the science of identifying these strokes and describing associated personality traits. [3] [5] [7]

Many organizations deny use of graphology due to the fact that according to 'Disability Discrimination Act 1995 recruitment based on graphology is a violation. Some due to confidentiality of clients and employer's organization deny use of services like graphology etc. [13].

[2] RELATED WORK

Graphologist analyze persons handwriting and study personality of writer. Some of the features they analyze include margins of the page, legibility and regularity, directions of lines called baseline, size and height of characters i.e. zones, inclination of letters called slants, pressure and speed of writing, ratio of upper, middle and lower zones, beginning and ending strokes, connections between characters, positions and shapes of certain alphabets like i and t etc. However, individuals handwriting may be consistent or distinct but should be unique to identified individuals from others. [8] [10] [11] [12] [13] [16][17][20]

[3] EXISTING METHODS

Currently professional graphologist/ handwriting expert's writer with piece of paper and their analysis is mostly based on their skill level. To do this kind of predictive analysis they mostly focus on common features like baseline, letters size, writers pressure, connected strokes, spacing between letters, words and lines, starting and ending strokes, speed, word slants, with of margins and many others. [2] [3] [5] [14] [15]][18][19]

[4] ASPECTS OF HANDWRITING BIOMETRICS

In this paper the major focus on three aspects viz. 1.Zones, 2. Baseline and 3. Pressure. All the sub aspects and their interpretations are listed and explained below:

Sr. No.	Features	Associated with	Characteristic/Traits Related with			
1	Upper Zone	Zone of abstract	1. Realm of thoughts			
	[3]		2.	Dreams		
			3. Hopes			
			4.	Plans		
			5.	Ideas and idols		
			6. Open/ close/ narrow/ broad mindedness			
2	Middle Zone	Related to day to	1.	Here and now		
	[3]	day emotions	2. Social interaction			
			3. Behavior with people			
			4. Extroverted/introverted			
			5.	Self confidence		
3	Lower Zone	Materialistic	1.	Food		
	[3]	Aspects	2.	Money		
			3.	Physical activity		
			4.	Basic instincts		

Table 1 Table of Zones

Zones: Like in the real world we live in three dimensions, height, width and depth. Similarly, in grahphology the dimension of height is studied using zones. This is vertical dimension which is further divided into three section upper zone, middle zone and lower zone. [1]

Considering timeline, upper zone represents upcoming or future, middle zone represents current state or present and lower zone represents the past. The head is considered as upper zone hence traits are related with intelligence, imagination, moral sciences etc. The body part is considered as middle zone expressing once ego, self-concious, social aspects etc. Lastly the lower part is considered lower zone which talks more about organic or physical needs of person in major.

1. Upper Zone – The Zone of Abstract

Graphology consider upper zone as the realm of mind as well as the guardian of spiritual aspirations. It is used to study intellect, imagination, illusion, fantasy and drive for power.

2. Middle Zone – Zone of day to day emotions

To express writer's self-image onto others, his/her role in life and thought about oneself middle zone is majorly studied. It major emphasis is on the quality and persistent efforts towards goals. Also, social and work-related activity showing the progress in surrounding environment.

3. Lower Zone – Zone of materialistic aspects

The zone basically focuses on all the materialistic aspect. They shows foody nature, money minded aspect, physical acitivities and basic instincts of person. Lower zone is used to study the balance between baseline and lower enlargement and its effect.

CURRENT ASPECTS IN HANDWRITING BIOMETRICS

Sr. No.	Features	Characteristics	Associated Traits		
1	Baseline	Straight Baseline	1. Stable		
		[3]	2.	Reliable	
			3.	Even Temper	
			4.	Self-confident	
			5.	Do not get easily happy/depress	
2	Baseline	Overly Straight	1.	Overly Control	
		Baseline [3]	2.	Inflexibility	
			3.	Defensive attitude	
			4.	Lacks in self confidence	
3	Baseline	True Ascending	1.	Enthusiastic person	
		Baseline [3]	2.	Has faith in future	
			3.	Full evolvement in everything	
			4.	Don't easily get discourage	
			5.	Eagerness for new things	
4	Baseline	Descending	1.	Feeling of temporary weakness	
		Baseline [3]	2.	Depression, discouragement, illness	
			3.	State of grief	
			4.	Emotional	
			5.	Sceptics/Critics	
5	Baseline	Wavy-Bouncy	1.	Have diffused objective	
		Baseline [3]	2.	Not goal oriented	
			3.	Moody, unsteady and inconsistent	
			4.	Jack of all type	
	Baseline	Erratic Baseline	1.	Casual	
6		[3]	2.	Laugh and Cry very easily	
			3.	Lot of unplanned work and drifters	
			4.	Need constant attention	
7	Baseline	Concave	1.	Excellent team players	
		Baseline [3]	2.	Good crisis mangers	
			3.	Self-motivated and self-directed	
			4.	Strong fighting spirit	
			5.	Change circumstances with hard work	
8	Baseline	Convex Baseline	1.	Excellent enthusiastic people	
		[3]	2.	Classic quitters	
			3.	Never complete the work	
			4.	Easily disappointed	
			5.	Can't sustain work/pressure	

Table 2 Table of Baseline

Baseline of a handwriting forms an imaginary line between middle and lower zone on which letters reside. Through the steadiness of baseline and balance of three zones personality is reveled. Baseline is considered as linear graph balancing ego and consious by materialistic drive.. Hence, it is also called as ego-adjustmenmt line. Thus baseline is used to study mood changes, moral and social balance, disposition and flexibility. .[1]

Some of the major baselines and studied and explained below:

1. Straight Baseline

The straight baseline indicates personality traits like stability, reliability and confident personality. Writer shows even temper and do not get easily depress or excited.

2. Overly Straight Baseline

The overly straight baseline is excessive effort baseline created with more efforts. This indicates writer have overly control and shows less flexibility. This shows build defensive attitude and lack of self confidence.

3. True Ascending Baseline

True Ascending baseline represents most positive traits. The writer is enthusiastic person with postive faith about future. Writer puts full envolvement in eveything does and don't get discourage due to small things.

4. Descending Baseline

Descending baseline shows negative traits on writer. The writer might feel weakness or might be depressed due illness, discouragement or environment. This baselines shows writers state of grief at that moment. Emotional effects and medical effects can be studied for more details.

5. Wavy-Bouncy Baseline

Most of the times writer changes goals or objective many times and does not show will or determination for its achievement. This baselines indicates witers moody nature, unsteadyness and inconsistent behavior. This writer migh be jack of all type.

6. Erratic Baseline

This baseline is important from psychological analysis point. This baseline shows writer's casual approach towards life. Writer shows emotional frugile nature and unplanned work. This writer might need constant attention and care.

7. Concave Baseline

Another postive baseline. This writer show positive traits like self motivation, good crisis mangment skills, excellent team player, strong fighting spirit and hard working nature to change circumstances postitively.

8. Convex Baseline

These writer shows excellent enthusiasm but lacks in hard work to complete the work. They give up their work in between hence this baseline is called as classic quitters. They get disappointment easily and unable to handle work pressure.

CURRENT ASPECTS IN HANDWRITING BIOMETRICS

Sr. No.	Features	Associated with	Characteristic/Traits Related with			
1	Pressure	Heavy [3]	1. Deep intense personality			
			2.	Creative power		
			3.	3. Dominance		
			4.	Enthusiasm & involvement		
			5.	Endurance		
			6	Deep commitments		
2	Pressure	Medium [3]	1.	Good drive and determination		
			2.	Energetic and resourceful		
			3.	Shoulder their responsibility carefully		
			4.	Dynamic, well inspired and motivated.		
3	Pressure	Light [3]	1.	lack of driving power and intensity		
			2.	2. Forgive easily		
			3.	Spiritual		
			4.	Soft spoken		
			5.	Low physical drive		
4	Pressure	Uneven	1.	Depression and anger		
			2.	Sign of nervousness		
			3.	Energetic but become tired very soon		
			4. Parental problem in childhood			
			5.	Loses motivation very fast		

Table 3 Table of Pressure

Pressure is body force excerted by muscle using pen on the paper. Pressure strokes changes everytime with upward and downword strokes. Higher the pressure shows physical fitness and mental energy of writer. As the pressure decreases the writer energy and intensity decreases. Diseases like parkinson can be identified by using pressure.

1. Heavy Pressure

This writer shows characteristics like deep intense personality, dominant nature, deep commitments, enthusiasm and hight involvement.

2. Medium Pressure

The is mostly found pressure overall. Writer with heavy pressure are driven and determined. They are energetic, resourceful, responsible, dynamic and motivated.

3. Light Pressure

Writer with low intensity and physical drive show light pressure. They forgive easily and are soft spoken. Some of the writer shows spirituality. Medical illness could also reduce the pressure.

[5] SUMMARY

Handwriting biometrics is still unexplored field. Various personality analysis techniques are being developed and many aspectes are being studied for their correlation and correctness. In

this we have majorly focused on current aspects in graphology namely zones, baselines and pressure. The details of aspects at their sub level are mentioned with personality traits and tried to cover major collabrative studies all together in this paper.

REFERENCES

- Omar Santana, Carlos M. Travieso, Jesus B. Alonso, Miguel A. Ferrer, "Writer Identification Based on Graphology Techniques", IEEE A&E SYSTEMS MAGAZINE, JUNE 2010
- [2] Carlos F. Romero, Carlos M. Travieso, Jesus B. Alonso and Miguel A. Ferrer, "Using off-line handwritten text for writer identification", WSEAS Transactions on Signal Processing, Issue I, Vol. 3, pp. 56-61, January 2007.
- [3] Prof. Seema Kedar, Ms. Vaishnavi Nair, Ms. Shweta Kulkarni, "Personality Identification through Handwriting Analysis", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5, Issue 1, January 2015
- [4] D. John Antony, O. F. M. Cap, "Personality profile through handwriting analysis".
- [5] Champa H N, K R Ananda Kumar, "Artificial neural network for human behavior prediction through handwriting analysis". International Journal of Computer Applications, Vol. 2, May 2010, pp. 36-41
- [6] Sheikholeslami, S. N. Srihari, V. Govindaraju, "Computer aided graphology Center of Excellence for Document Analysis and Recognition" (New York).
- [7] Champa H. N. and K. R. Ananda Kumar, 2010, "Automated Human Behavior Prediction through Handwriting Analysis", First International Conference on Integrated Intelligent Computing.
- [8] Abdul Rahiman M, Diana Varghese, Manoj Kumar G, "HABIT- Handwriting Analysis Based Individualistic Traits Prediction".
- [9] Parmeet Kaur Garewal, Deepak Prashar, "Behavior Prediction through Handwriting Analysis".
- [10] Vlad Atanasiu, "Surveys and projects in handwriting analysis", MIT Postdoctoral fellow, Department of Brain and Cognitive Sciences 77, Massachusetts Av., Cambridge MA 02139, USA.
- [11] Bin Zhang Sargur N. Srihari, "Analysis of Handwriting Individuality Using Word Features", State University of New York at Buffalo, Buffalo, NY 14228.
- [12] Reâ Jean Plamondon, Sargur N. Srihari," On-Line And Off-Line Handwriting Recognition: A Comprehensive Survey", Ieee Transactions On Pattern Analysis And Machine Intelligence, Vol. 22, No. 1, January 2000.
- [13] Dharam Veer Sharma, Geeta Saini, Mohit Joshi, "Statistical Feature Extraction Methods for Isolated Handwritten Gurumukhi Script", Vol. 2, Issue 4, July-August 2012, pp.380-384.
- [14] Vikram Kamath, Nikhil Ramaswamy, P. Navin Karanth, Vijay Desai And S. M. Kulkarni, "Development Of An Automated Handwriting Analysis System", Vol. 6, No. 9, September 2011
- [15] Imran Siddiqi Chawki Djeddi Ahsen Raza •Labiba Souici-meslati, "Automatic analysis of handwriting for gender classification", Pattern Analysis and Applications • May 2014

- [16] Djeddi C, Siddiqi I, Souici-Meslati L, Ennaji A (2013) Codebook for writer characterization: A vocabulary of patterns or a mere representation space? In: Proceedings of the 12th Int'l Conference on Document Analysis and Recognition
- [17] HuangD,ShanC,ArdabilianM,WangY,ChenL(2011) Local binary patterns and its application to facial image analysis: a survey. IEEE Trans Systems Man Cybern Part C Appl Reviews 41(6):765–781
- [18] Liwicki M, Schlapbach A, Bunke H (2011) Automatic gender detection using on-line and offline information. Pattern Anal Appl 14(1):87–92
- [19] Siddiqi I, Vincent N (2010) Text independent writer recognition using redundant writing patterns with contour-based orientation and curvature features. Pattern Recognit 43(11):3853– 3865
- [20] Sokic E, Salihbegovic A, Ahic-Djokic M (2012) Analysis of offline handwritten text samples of different gender using shape descriptors. In: Proceedings of IX International Symposium on Telecommunications (BIHTEL), pp 1–6



CLOUD BASED SECURITY USAGE FOR THE FACE RECOGNITION

44835

Mr. Venugopal S. Narsingoju¹, Dr. Aslam Y. Suriya²

¹Assistant Professor, MES's Institute of Management and Career Courses (IMCC), Pune ²Assistant Professor, Dept. of Computer Science, Janata Mahavidyalaya, Chandrapur

ABSTRACT:

Face recognitions plays a major role in biometrics research which helps to identify the users based on various parameters. In this fast developing technological world face recognition is increasingly used to distinguish the users in mobile. Breach of mobile security is common in this fast paced world which is seriously something to look upon. Cloud gives us significant security as far as data is concerned. Cloud-based storage can assure a certain level of privacy to the end-users.

Keywords: Image capturing, Cloud transfer, Registration, Authentication, Data privacy, Face Detection, Face Recognition

[1] INTRODUCTION

Cloud computing could be a new technology within the market. In cloud computing user will access their files or knowledge from anywhere exploitation net. There square measure many advantages of cloud computing like increase output, scale back prices, improve accessibility and needs less coaching however on the opposite hand it's some security problems. In that, distinctive approved user could be a major issue. The user needing to access information or services must be registered and before each access to data or services; his/her identity should be attested for authorization. There square measure many authentication techniques as well as ancient and biometric however has some drawbacks.

CLOUD BASED SECURITY USAGE FOR THE FACE RECOGNITION

Face recognition has been at the crux of many novel breakthroughs over the past twenty years and has steady proffered many cross-domain applications that vary from thought business package to crucial enforcement applications. The fascinating properties of the same ideas are amalgamated to create an efficient coalition that may augment the performance of Face Recognition, additionally to serving boundlessly during an over plus of different disciplines.

[2] CLOUD COMPUTING

During the past few years, the IT world witnessed the birth and growth of a new paradigm, commonly called Cloud Computing, although it has also been named as Dynamic Computing [1]. It is difficult to assign a precise date for its genesis, since the term "cloud" has already been used in several contexts, describing large ATM networks in the 1990s for instance [2], and is also based upon some already existing technologies like distributed computing, virtualization or utility computing which have been around for several years [3]. However, some [4] claim that the true birth of Cloud Computing happened when IBM and Google announced a partnership in this domain [5], leading to a hype around the subject and lots of popularity.

[3] FACE IDENTIFICATION

The ability to recognize and distinguish one face from thousands of other different faces has been imprinted on the human brain visual cortex for millennia, so that now people usually perform this task subconsciously without realizing the difficult process around it, and what it takes for their brain to accomplish an identification with a very high precision. Nonetheless, however easy it seems for a human to recognize a face quite independently of the context, computers still struggle for high coincidence intervals when presented with the same face in several different environments. During the last 40 years, the problem of face identification by computers has been the subject of extensive research, with several techniques being suggested which, sometimes, can match or exceed humans [6].

During this last decade, the great improvements in computing power have brought new possibilities, enabling techniques that 20 or 30 years ago were simply not feasible, CPU- and memory-wise. Human face identification is a natural activity to the majority of humans, being an essential capability that has evolved and reached a high accuracy percentage. Despite being natural to humans, it is necessary to define clearly what facial identification means for a computer. It is part of a broader subject, which is known as human face perception in [7]. A general notion of this area is given, stating that it is a kind of intelligent behavior for computer to catch and deal with the information on human faces, including human face detection, human face tracking, human face pose detection, human face recognition, facial expression recognition, lip-reading and others. Although

face identification is a specific sub-area of human face perception, it partially encompasses several of the others areas presented

One of the goals of automatic face identification systems is to reach and eventually surpass human performance on that same task, although they are not required to follow the same strategies humans have developed over time. As a starting point, this section describes the strategies the human visual system uses to perform face identification, which might be useful when developing automatic systems. Humans make use of a lot of sensory information to perform recognition (visual, auditory, olfactory, and tactile, etc.) and in some cases can even rely on the context to identify a person. Yet, most of this information is not easily available to computers, which are frequently only presented with 2D-images. However, computers have a clear advantage on the amount of information that can be stored and processed, even if they cannot use it so effectively. A person is typically able to store, or remember, a limited set of faces in the order of thousands, while a computer system can potentially store an infinite number.



Fig. 1 : Cloud Structure

[4] AUTHENTICATION IN CLOUD

As cloud users store their information to various services across the Internet, it can be accessible by unauthorized people [8]. So security is the most important issue in cloud computing. To provide security we require proper authentication technique in cloud computing. Typically, authentication is done based on information about one or more of the following: i. Knowledge of the subject, such as password or secret information. ii. Possession of the user, such as smart card,

CLOUD BASED SECURITY USAGE FOR THE FACE RECOGNITION

passport, driver's license, etc. iii. Biometric traits of the client, such as fingerprint, voice, iris, etc. [9].

The data leakage and security attacks can be caused by insufficient authentication [10]. Cloud services are paid services so to identify authorized user is major concern in cloud computing. In this paper, we focus on the security issues of cloud computing, particularly on authentication. To solve authentication problem in cloud computing, there are different traditional as well as biometric techniques.

[5] FACE RECOGNITION ARCHITECTURE

Face recognition is a biometric security system. As the name suggests the face acts as a password for the systems.



Fig. 2 : Face Recognition Architecture

- 1) User has to fill the registration form which is provided by cloud provider. It contains detail information about the user.
- 2) User has to provide valid Email ID as a username to the face recognition system at the time of registration.
- Face recognition system checks the Email ID against the availability of that username. Username should not repeat or match with existing user's username.
- 4) After checking the availability of username, the password must be created. Face image through web camera is stored in database as a password.
- 5) After providing valid username and storing face image as a password, the registration on cloud server is completed

[6] CONCLUSION

In this paper, we discussed about cloud services which can be used only to the authorized user, secure authentication is necessary in cloud computing. There are so many authentication techniques like password, OTP, Voice recognition, finger recognition, palm recognition etc. but still it has some drawbacks like at times password techniques are not feasible, password can be easily stolen by hacker or if user uses complex password, user may forget that password etc. So it is a

better option to use face recognition system rather than traditional or other biometric authentication techniques. The security level of cloud provider in terms of secure authentication is much improved by using face recognition technique

REFERENCES

[1]. Rajan, S. & A. Jairath (2011, June). Cloud Computing: The Fifth Generation of Computing. In International Conference on Communication Systems and Network Technologies, pp. 665-667. IEEE.

[2]. Zhang, Q., L. Cheng, & R. Boutaba (2010, April). Cloud computing: state-of-the-art and research challenges. Journal of Internet Services and Applications 1 (1), 7-18.

[3]. Vaquero, L. M., L. Rodero-Merino, J. Caceres, & M. Lindner (2008, December). A Break in the Clouds: Towards a Cloud De_nition. ACM SIGCOMM Computer Communication Re- view 39 (1), 50 -55.

[4]. Gong, C., J. Liu, Q. Zhang, H. Chen, & Z. Gong (2010, September). The Characteristics of Cloud Computing. In 39th International Conference on Parallel Processing Workshops, pp. 275-279. IEEE.

[5]. IBM & Google (2007). Google and IBM Announce University Initiative to Address Internet-Computing Challenges.

[6]. O'Toole, A., P. J. Phillips, F. Jiang, J. Ayyad, N. Penard, & H. Abdi (2005). Face recognition algorithms surpass humans. IEEE Trans. PAMI.

[7]. Hongxun, Y., G. Wen, L. Mingbao, & Z. Lizhuang (2000). Eigen features technique and its application. In 5th International Conference in Signal Processing Proceedings, Volume 2, pp.1153-1158.

[8]. R. Kalaichelvi Chandrahasan, S Shanmuga Priya and Dr. L. Arockiam, "Research Challenges and Security Issues in Cloud Computing," International Journal of Computational Intelligence and Information Security, Volume-3, No-3, March 2012.

[9] Minhaz Fahim Zibran, "Biometric Authentication: The Security Issues," University of Saskatchewan, 2012.

[10] S. O. Kuyoro, F. Ibikunle and O. Awodele, "Cloud Computing Security Issues and Challenges," International Journal of Computer Networks (IJCN), Volume-3, Issue-5, 2011.





THE IOT AND BIG DATA FUSION

Dipti N. Belsare MES's Institute of management and career courses, Pune, India

ABSTRACT:

Big data and IoT are the buzzwords everyone talks about. The term of big data was coined under the explosive increase of global data and was mainly used to describe enormous datasets. Compared with traditional datasets, big data generally includes masses of unstructured data that need more real-time analysis. In IoT an enormous amount of network sensors are embedded into devices in the real world. Such sensors deployed in different fields may collect various kinds of data. Many operators of IoT realize the importance of big data since the success of IoT is hinged upon the effective integration of big data and cloud computing [1]. In this paper I am trying to focus these two technologies are inter-dependent and should be jointly developed.

Keywords: Big Data, IoT, Big Data IoT Analytics, IoT and Big Data Analytics Architecture

[1] INTRODUCTION

We live in a world increasingly driven by data. Big Data and Internet of Things (IoT) are having profound impacts on our everyday life. There are many technologies making boom in market.

IoT promises a range of consumer benefits, such as smart cars that proactively keep track of vehicle maintenance, or services that control home alarm systems and thermostats. The wave of big data is coming from IoT and interconnected devices, bringing with it a fresh set of potential integration and security challenges for IT departments.

The integration between Big Data and IoT can produce positive impacts in the nextgeneration of our development in smart cities, national planning and forecasting of our future activities and investments. Big Data and IoT integration can make valuable impact on our daily life in healthcare, finance, security, transportation and education. Big Data and IoT integration can also help high-tech sectors such as weather forecasting, space technology and biotechnology to enable thousands of simulations to be completed in seconds.

In this paper, I tried to explain Big Data Analytics, IoT, IoT and Big Data Analytics, IoT and Big Data Analytics Architecture, Challenges of IoT Big Data.

[2] OVERVIEW OF IOT

In today's world billions of connected devices, from smart vehicles to smart meters, generate ever increasing quantities of data. This worldwide network of connected devices is the Internet of Things (IoT). To derive value from the IoT, enterprises need to securely connect their IoT enabled devices to collect data, then analyze and integrate that data with enterprise applications and processes.

Following are some internet of things characteristics

Things- Anything that can be tagged or connected as such as it's designed to be connected.

Data- Data is the glue of the Internet of Things, the first step towards action and intelligence.

Communication Devices get connected so they can communicate data and this data can be analyzed.

Intelligence- The aspect of intelligence as in the sensing capabilities in IoT devices and the intelligence gathered from data analytics (also artificial intelligence).

Action- This can be manual action, action based upon debates regarding phenomena (for instance in climate change decisions) and automation, often the most important piece.

Ecosystem- The place of the Internet of Things from a perspective of other technologies.

IoT has emerged as a new trend in the last few years, where mobile devices, transportation facilities, public facilities, and home appliances can all be used as data acquisition equipment in IoT. All surrounding electronic equipment to facilitate daily life operations, such as wrist- watches, vending machines, emergency alarms, and garage doors, as well as home appliances, such as refrigerators, microwave ovens, air conditioners, and water heaters are connected to an IoT network and can be controlled remotely.

Ciufo [2] stated that these devices "talk" to one another and to central controlling devices. Such devices deployed in different areas may collect various kinds of data, such as geographical, astronomical, environmental, and logistical data [2].

[3] OVERVIEW OF BIG DATA AND ANALYTICS

Big Data-

What is Big Data?

"Big Data" is often defined in terms of "3V's" i.e.

- Volume the amount of data generated, stored and analyzed. The amount of data stored determines the level of insight that can be obtained from that data;
- **Variety** type and nature of data. Historically data was structured and from a single source in which case it would fit readily into 'columns' and 'rows'. Increasingly data is sourced from a variety of sources with many different formats;
- Velocity the speed at which data is generated and processed.
- Variability Variations in the data sets. For example is a temperature measured in degrees Celsius, Fahrenheit or Kelvin;
- Veracity Quality of the captured data. Where decisions are being made on data you need to be sure that the data is correct.
 - "Big Data" can be broadly defined as being associated with:

• Data sets those are so large or complex that traditional data processing applications/ methods are inadequate;

• Challenges include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying, and updating and information privacy;

- Use of predictive analytics or other advanced methods to extract value from data;
- Any size of data set (though likely to be large);

• Use in decision making for new applications, greater operational efficiency, cost reduction and reduced risk.

Big Data Analytics

Big data analytics aim to immediately extract knowledgeable information that helps in making predictions, identifying recent trends, finding hidden information, and ultimately, making decisions [3]. Big data analytics require technologies and tools that can transform a large amount of structured, unstructured, and semi-structured data into a more understandable data and metadata format for analytical processes. The algorithms used in these analytical tools must discover patterns, trends, and correlations over a variety of time horizons in the data [4].

After analyzing the data, these tools visualize the findings in tables, graphs, and spatial charts for efficient decision making. Thus, big data analysis is a serious challenge for many applications because of data complexity and the scalability of underlying algorithms that support such processes [5].

Different analytic types are used according to the requirements of IoT applications [6]. These analytics types are discussed under real-time, off-line, memory level, business intelligence (BI) level, and massive level analytics categories [7].

Real-time analytics is typically used to analyze the data collected from sensors. Greenplum [8] and Hana [9] are examples of real-time analytics architecture.

Off-line analytics is used when a quick response is not required [6]. SCRIBE [10], Kafka [11], Time-Tunnel [12], and Chukwa [13] are examples of architectures that conduct off-line analytics and can satisfy the demands of data acquisition.

Memory level analytics is applied when the size of data is smaller than the memory of a cluster [6]. To date, the memory of clusters has reached terabyte (TB) level [14]. Memory-level analytics is suitable for conducting real-time analysis. MongoDB [15] is an example of this architecture.

BI analytics is adopted when the size of data is larger than the memory level, but in this case, data may be imported to the BI analysis environment [16].

Massive analytics is applied when the size of data is greater than the entire capacity of the BI analysis product and traditional databases [17]. Massive analytics uses the Hadoop distributed file system for data storage and map/reduce for data analysis.

[4] IOT AND BIG DATA ANALYTICS

The concept of IoT is becoming more pertinent to the realistic world due to the development of mobile devices, embedded and ubiquitous communication technologies, cloud computing, and data analytics. Moreover, IoT presents challenges in combinations of volume, velocity and variety. Several diversified technologies such as computational intelligence, and big-data can be incorporated together to improve the data management and knowledge discovery of large scale automation applications [18].

IoT will enable big data, big data needs analytics, and analytics will improve processes for more IoT devices. IoT and big data can be used to improve various functions and operations in diverse sectors. Both have extended their capabilities to wide range of areas [19].

In IoT as data from sensors and SCADA systems are carried to cloud servers from the remote locations, these data are highly unstructured. The volume and velocity of such data is too huge and the time need to analysis is very short. Hence the stream of Big data analytics comes in IOT. The relationship between Big Data and IoT shown in figure [7]



In figure1 data can be divided into three steps to enable the management of IoT data. The first step comprises managing IoT data sources, where connected sensors devices use applications to interact with one another. For example, the interaction of devices such as CCTV cameras, smart traffic lights, and smart home devices, generates large amounts of data sources with different formats. This data can be stored in low cost commodity storage on the cloud. In the second step, the generated data are called ``big data," which are based on their volume, velocity, and variety. These huge amounts of data are stored in big data files in shared distributed fault-tolerant databases. The last step applies analytics tools such as MapReduce, Spark, Splunk, and Skytree that can analyze the stored big IoT data sets. The four levels of analytics start from training data, then move on to analytics tools, queries, and reports [7].

[5] IOT AND BIG DATA ARCHITECTURE



The diagram below shows the general architecture for delivery of IoT Big Data services [20]

Fig. 2:- General architecture for IoT Big Data

A number of functional units are identified in the IoT Big Data architecture as shown in the

These functional units do not need to be implemented by all organizations delivering IoT Big Data services.

Context Data Layer:- It is concerned with obtaining external non IoT data ("Context data") which is either available to the third party application or used during the processing of IoT data e.g. "mashing up" IoT data with context data. The Context Data Layer is also able to communicate with the external data sources.

IoT Service Layer:- The IoT service layer is concerned with handling the device specific interactions required for obtaining data from IoT devices and sending control commands (where relevant) to those IoT devices. Therefore this layer is required to handle bi-directional communications both to IoT devices and to the upper layers of the architecture.

The IoT Service Layer is expected to handle the lower level interactions with IoT devices. Those devices might be connected using a variety of protocols and low level communication technologies including (but not limited to) oneM2M [3], Hypercat [4], Constrained Application Protocol (CoAP), MQ Telemetry Transport (MQTT), Real Time Streaming Protocol (RTSP), or device specific interfaces such as JavaScript Object Notation (JSON)/Extensible Markup Language (XML) over HTTP.

The IoT Service Layer is expected to handle authentication and security aspects regarding the interfacing with IoT devices.

Data and Protocol Mediator: - The Data and Protocol Mediator is responsible for ingesting information from IoT devices as well as other external sources ("context data"). It ensures that data is transformed to the Harmonized Entity Definition before being stored and published by the Data & Control Broker.

Data & Control Broker: - The Data & Control Broker is responsible for enabling third party application access to harmonized data entities through a query and subscribe API, allowing applications to gain access to such data in a standard way.

Peer API Access Management: - The Peer API Access Management function is responsible for interfacing with its peers in other organizations to receive and publish additional relevant harmonized IoT and context data.

Developer API Access Management: - It implements authentication, authorization and access control using industry standards to ensure privacy and security around the harmonized data.

IoT Big Data Store This function provides data storage for this massive data and it may also provide short to medium term storage capabilities for use by the Data & Control Broker, depending on the specific implementation.

The "Big Data" database is expected to be used to store the harmonized data entities received from the IoT devices and/or the external data sources. As it is expected there could be many millions of IoT devices generating data frequently, the required storage space may be vast (i.e. of the order of many terabytes to many petabytes of data). It is expected the "Big Data" database could be implemented using products such as Apache Cassandra, Apache Hadoop, MongoDB, Neo4j, Titan or DynamoDB. To achieve high performance the database component may employ substantial quantities of memory to hold copies of data that is persistently stored on "hard disk". [20]

[6] CHALLENGES

Some of the key challenges for IoT Big Data, which have a bearing on the design of architectures suitable for service delivery, include [20]

1. The number of IoT devices: With forecasted growth in the number of connected "things" expected into the billions world-wide there will be masses of devices which may be a data source, and which may be subject to third party control;

2. The variety of IoT devices: There will be enormous variety in the devices which may provide data, even in the case of similar devices e.g. an electronic thermostat. Data from any individual device manufacturer or model may be quite dissimilar from that of nominally identical devices in such areas as field names, units, and data structures;

3. Intelligence of IoT devices: IoT devices have more and more compute resources and integrate several technologies like Graphics Processing Unit (GPU) and Solid State Drive (SSD) storage. Simple sensors are evolving to autonomous systems which will be able to manage their own analytics and be part of large analytics networks;

4. Risk of IoT device malfunction: With a great number of IoT devices and manufacturers it is reasonable to assume there will be many occasions where IoT devices malfunction in various ways. In the most drastic situations devices will fail completely but it should be expected that more subtle malfunctions will occur which might result in aberrations of data coming from those devices, or a failure of the device to perform a required control function;

5. Update frequency: Though some devices (e.g. remote sensors) will produce data reports at a low frequency there may be substantial quantities of data streaming from more sophisticated Internet connected things such as cars;

6. Historical data: It is expected that many Big Data insights will derive from historical data recorded from IoT devices. This may be processed alone to derive analytics/ intelligence or considered alongside current data particularly to enable smart monitoring and control;

7. Context data: Much IoT data will make more sense when put in context with other data. Context data might be generally "static" (or at least with a slow update period) such as geographical data, or could be more dynamic e.g. weather forecast data.

8. Privacy issues: With so many expected IoT devices acquiring data there could be a substantial risk relating to the disclosure of data which is considered personal to end users.

When IoT data is stored in a Big Data system and made available to third parties there is a need to implement strong safeguards to ensure end users remain in control of their personal information. Mobile Network Operators (MNOs) are in a strong position to help users remain in control of their data and to make data available in the best way via consent, aggregation.

REFERENCES

- [1] Big Data Related Technologies, Challenges and Future Prospects Min Chen, Shiwen Mao, Yin Zhang, Victor C.M. Leung- 3 Published online: 22 January 2014 © Springer Science+ Business Media New York 2014.
- [2] C. A. Ciufo. (2014).Industrial Equipment Talking on the IoT? Better get a Gateway (Device). [Online]. Available: http://eecatalog.com/caciufo/2014/07/15/iot-gateway-adlink/
- [3] C.-W. Tsai, "Big data analytics: A survey," J. Big Data, vol. 2, no. 1, pp. 1–32, 2015.

- [4] S. Oswal and S.Koul, "Big data analytic and visualization on mobile devices" in Proc. Nat. conf. New Horizons IT-NCNHIT, 2013, p. 223.
- [5] L. Candela and D. P. C. Pagano, "managing big data through hybrid data infrastructures" ERCIM News, vol. 89, pp. 37-38, Jun. 2012.
- [6] C. L. P. Chen and C.-Y. Zhang, "Data-intensive applications, challenges, techniques and technologies: A survey on big data," Inf. Sci., vol. 275, pp. 314_347, Aug. 2014.
- [7] M. Marjani et al, "Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges" IEEE Access, vol5, 2017
- [8] F. M. Waas, "beyond conventional data warehousing-massively parallel data processing with Greenplum database" in International Workshop on Business Intelligence for the realtime Enterprise. Berlin, Germany, Springer, 2008.
- [9] F. Färber, S. K. Cha, J. Primsch, C. Bornhövd, S. Sigg, and W. Lehner, "SAP HANA database: Data management for modern business applications," ACM SIGMOD Rec., vol. 40, no. 4, pp. 45_51, Dec. 2011.
- [10] M. Castro, P. Druschel, A. M. Kermarrec, and A. I. T. Rowstron, "Scribe: A large-scale and decentralized application-level multicast infrastructure," IEEE J. Sel. Areas Commun., vol. 20, no. 8, pp. 1489_1499, Oct. 2002.
- [11] J. Kreps, N. Narkhede, and J. Rao, ``KAFKA: A distributed messaging system for log processing," in Proc. NetDB. 2011, pp. 1_7.
- [12] H. Notsu, Y. Okada, M. Akaishi, and K. Niijima, ``Time-tunnel: Visual analysis tool for time-series numerical data and its extension toward parallel coordinates," in Proc. Int. Conf. Comput. Graph., Imag. Vis. (CGIV), Jul. 2005, pp. 167_172.
- [13] A. Rabkin and R. H. Katz, "Chukwa: A system for reliable large-scale log collection," in Proc. LISA, Nov. 2010, pp. 1_15.
- [14] S. Hong and H. Kim, ``An analytical model for a GPU architecture with memory-level and thread-level parallelism awareness," ACM SIGARCH Comput. Archit., vol. 37, no. 3, pp. 152_163, Jun. 2009.
- [15] K. Chodorow, MongoDB: The Definitive Guide. Newton, MA, USA: O'Reilly Media, Inc, 2014.
- [16] Z. Jourdan and R. K. T. E. Rainer Marshall, "Business intelligence: An analysis of the literature 1," Inf. Syst. Manage., vol. 25, no. 2, pp. 121_131, 2008.
- [17] A. Bifet, G. Holmes, R. Kirkby, and B. Pfahringer, "MOA: Massive online analysis," J. Mach. Learn. Res., vol. 11, pp. 1601_1604, May 2010.
- [18] N. Mishra, C. Lin and H. Chang, A cognitive adopted framework for IoT big data management and knowledge discovery prospective, International Journal of Distributed Sensor Networks, 2015, (2015), pp.1-13
- [19] K. R. Kundhavai, "IoT and Big Data The Current and Future Technologies: A Review" et al, IJCSMC, vol.5, Issue 1, Jan 2016, pg.10-14
- [20] GSM Association Official Document CLP.25 IoT Big Data Framework Architecture 20 Oct. 2016.
- [21] IoT and Big Data Combine Forces the Internet of Things produces big data that can drive improvements in manufacturing and light industrial applications Author: Advantech Technical White Paper
- [22] IoT Analytics: using Big Data to architect IoT solutions By Srinath Perera Vice President Research White Paper WSO2

Impact of stress and work-family conflict on the mental well-being of physcians: Mediation by job and career satisfaction

Meenakshi Aggarwal-Gupta, Poonam Vatharkar

ABSTRACT

The mental well-being (MWB) of physicians is important both for patients and for the physicians themselves, positively influencing their health and patient care. The current study was undertaken to clarify the impact of perceived stress (PS) and work–family conflict on the MWB of physicians. Two satisfaction-related variables, namely, job satisfaction (JS) and career satisfaction (CS), were expected to mediate between these variables. Data were collected through a survey of 102 physicians. The results show strong correlations between all study variables. JS and CS completely mediated the link between PS and MWB. These results highlight the importance of JS and CS in ensuring the MWB of physicians. If physicians feel satisfied and empowered, then PS will not impact their MWB. Directions for future research are discussed at the end of the paper. Organizations must enact measures to enhance the satisfaction levels of physicians by providing a supportive environment, opportunities for career growth, and diversity of work.

Key Words: • Job Satisfaction • Career Satisfaction • Work Family Conflict • Perceived Stress • Mental Well being • Physicians

Meenakshi

Aggarwal-Gupta, Associate Professor, Organizational Behaviour and Human Resource Management, Indian Institute of Management Indore, Mumbai, India

Poonam Vatharkar,

Doctoral Scholar, Organizational Behaviour and Human Resource Management, University of Mumbai, Mumbai, India

Introduction

Ltd

MΑ

Healthcare professions are significant and personally gratifying (Shanafelt et al, 2012). Though the work is demanding, exhausting and sometimes stressful, it was once widely recognised that physicians had good jobs (Williams et al, 2000). Current research in this field, however, suggests that this profession no longer enjoys the respect it used to (Levinson and Pizo, 2011). An emphasis on profits, increased healthcare costs and lawsuits have brought healthcare into the commercial mainstream.

Physicians today must balance various new roles and face concomitant stress and strains. Many occupational stressors, such as long and odd working hours leading to poor sleeping habits, high performance expectations, and time pressure contribute to higher stress levels for physicians (Clever, 2002). Physicians have higher rates of depression and anxiety than the general population (Chambers and Belcher, 1993) and poor mental well-being (MWB) Rout et al, 1996)). Beyond these stressors, social media and the press may make demands on doctors over and above the standards that the profession and society sets for them and further increase their levels of stress and dissatisfaction (Collum et al, 2013).

Excessive stress can reduce MWB (Chandraiah et al, 2003). Studies have shown that physicians have high rates of mental health problems such as anxiety (Kroenke et al, 2007) and depression (Firth-Cozens, 2006). Due to their odd working hours, unpredictability of work commitments and lack of control over their professional lives, doctors also face high levels of work-family conflict (WFC), which affect their MWB (Hughes and Bozionelos, 2007; Elston, 2009).

While much work has been carried out on the impact of stress and WFCs on physicians MWB, not enough attention has been given to the concept of job satisfaction (JS). JS is the extent to which people like or dislike their work (Spector, 1997). The JS of physicians is crucial, since it affects not only their own health but also the health of their patients (DiMatteo et al, 1993). Previous research suggests that the JS of physicians leads to greater productivity, better patient care and better health and MWB for themsevles (Cooper et al, 1989; Judge et al, 2001; Dyrbye et al, 2013). Evidence unfortunately suggests that the dissatisfaction of physicians is increasing (Leigh et al, 2002).

In addition to JS, career satisfaction (CS) Parasuraman et al, 1996)) also forms an important element of the way that health care professionals view their work. The concept of CS includes two important factors, namely, career progression (Stroh et al, 1996) and career involvement (Tenbrunsel et al, 1995). Individuals feel satisfied with their career if they perceive growth opportunities and are able to allot sufficient time and attention to it. The high stress level found among physicians and its impact on their CS has received much attention (Collum et al, 2013; Friedberg et al, 2014). Along with work-related stressors, challenges on the personal front have also increased. While there has been a move towards better work-life

balance and reduced WFC (Collum et al, 2013), this may run contrary to physicians's beliefs about professionalism and the standards they set for themselves. This conflict is likely to reduce satisfaction in the field (Sutherland and Cooper, 1992). This study was undertaken to understand the impact of stress, WFC and the CS and JS of physicians on their MWB.

Literature review

Due to nature of their work, physicians must face a several occupational stressors. It is evident that stressors like workload, time pressure, patient stress and professional relations lead to reduced JS (Phipps, 1988; Asamani et al, 2015). Studies have found that higher perceived stress (PS) is associated with lower satisfaction levels in various spheres, such as work, career and life (Cooper et al, 1989; Williams et al, 2000; Richardson and Rothstein, 2008).

Excessive stress among physicians creates conflict between their various life roles, contributing to the feeling of overload and reduced physical well-being and MWB (Sekaran, 1983; Greenhaus and Parasuraman, 1986). There also appear to be age-and-gender-related differences in the factors that lead to stress: one study found that work demands and patient expectations were the most important predictors of stress among male physicians. While these factors were also relevant for female physicians, the main stressors for them were work-home interface and social life (Cooper et al, 1989). Other studies have determined that younger doctors experience more stress than older doctors (Shanafelt et al, 2009). As the level of stress increases, it paves the way for burnout, mental health problems (Arnetz, et al, 1987), psychosomatic illness (Cooper et. al, 1989) and reduced quality of life and quality of service (Friedberg et al, 2014). High levels of stress thus have a negative impact on the overall MWB of physicians and can also negatively impact the delivery of healthcare. Thus, it is hypothesized:

H1: If the PS of physicians is high, their MWB, JS and CS will be low.

WFCs are created when the demands of work, time devoted to meet these demands and the strains created are so great that they interfere with meeting family responsibilities (Netemeyer et al, 1996). Physicians are expected to work longer unsociable hours and be available for emergency on-call services. This makes them vulnerable to work interference in family life (Shanafelt et al, 2012), resulting in WFCs. WFC has been linked to various career and work related variables, such as career progression, involvement and JS (Martins et al, 2002). This feeling of WFC leads to a reduced sense of JS. Researchers also found that WFCs are negatively related to the JS and CS of physicians (Keeton et al, 2007) and positively related to depression and impaired MWB (Bacharach et al, 1991; Thomas and Ganster, 1995).

When individuals face WFCs, they become exposed to the risk of depression and psychosomatic complaints (e.g., Frone et al, 1996). Work interference with family is an important predictor of MWB for physicians (Rout et al, 1996; Elston 2009). Previous research suggests that WFC leads to stress and creates diminished MWB (Bacharach et al, 1991; Frone et al, 1996). While most studies have addressed the negative impact of WFC on individuals, a recent study has suggested that doctors can struggle with work-life balance issues and yet report high CS, depending on how their career evolves (Keeton et al, 2007). It is thus important to study WFC in association with satisfaction variables.

H2: If the WFC of physicians is high, their MWB, JS and CS will be low.

Research in MWB among physicians is important for both the professionals and the patients they care for (Wallace et al, 2009): when JS is reduced, this boosts the feeling of burnout, which ultimately leads to reduced MWB (Arnetz et al, 1987). Various strategies, such as coping, mindfulness, diet and meditation have been recommended to reduce stress and increase the MWB of physicians (e.g., Firth-Cozens, 2001). While these interventions are required, we feel that the work context also has a critical role to play in the MWB of physicians. Previous studies have found that peer support, supportive work environments and JS are related to the health and MWB of employees (Cooper et al, 1989; Rathi and Rastogi, 2007). Satisfaction with these elements is thought to be critical to the MWB of physicians.

H3: If the JS and CS of physicians are high, MWB will also be high.

JS and CS may have a larger role to play in physician MWB than previously thought. Various stressors, demands and conflicts may reduce the MWB of physicians, but these are balanced out by the sense of achievement and fulfilment that their work may provide. Role theory suggests that, while role demands may lead to feelings of imbalance and stress, they can also provide a sense of enrichment, especially if the individual is able to play out various roles in a satisfactory manner (Haar et al, 2014). If their JS or CS is high, this may help physicians to cope with various stressors and support their MWB (Bliese et al, 2017). It is expected that:

H4: *a*) *JS* and *b*) *CS* will mediate between PS and MWB.

H5: *a*) *JS* and *b*) *CS* will mediate between WFC and MWB.

See *figure 1* for proposed research model.

Methods

Data was collected through a survey of 200



© 2017 MA Healthcare Ltd

physicians working in various fields of medicine. An initial set of physicians was approached during a conference. Prior permission had been granted by the conference organisers to collect the data. Subsequent to this, we used the snowballing method to reach more physicians. The respondents were contacted either in person or via email and the purpose of the questionnaire was explained. They were assured of the confidentiality of their responses. Of the 200 questionnaires that were distributed/ emailed, 102 usable responses were received by the deadline for data analysis. The sample was 70% male, with an average age of 43 years old. Approximately 53% of the respondents had a bachelor's degree in medicine, and the rest had a post-graduate degree in medicine. The average work experience was 17.46 years.

Measures

The following scales were included in the questionnaire to measure the study variables. They have all been used extensively in the past in various contexts and have shown high validity related questions such as gender, age, qualifications and work experience.

MWB: The 14-item Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) developed by Tennant et al (2007) was used to measure MWB. Each item was rated on 5-point scale from 1 (none of the time) to 5 (all the time), with higher scores indicating greater MWB. A sample question is 'I've been dealing with problems well'.

JS *and* CS were measured using items from a scale developed by Linzer et al (2000). This scale measures 13 aspects of JS. For the purposes of this study, two subscales, a 5-item scale on JS

and a 4-item CS scale, were considered. Each item was rated on 5-point scale from 1 (strongly disagree) to 5 (strongly agree), with higher scores indicating greater levels of satisfaction. A sample question for JS is 'I find my present clinical work personally rewarding'.

A sample question from the CS scale is 'All things considered, I am satisfied with my career as a physician'.

PS was measured using a 14-item scale adapted from Cohen et al (1983). Using a 5-point scale from 0 (never) to 4 (very often), respondents indicated the frequency with which they had felt or thought certain way during the previous month. A sample question from this scale is 'felt that you were unable to control the important things in your life'.

WFC was measured using a 5-item scale developed by Netemeyer et al (1996). Each item was rated on 5-point scale from 1 (strongly disagree) to 5 (strongly agree), with higher scores indicating greater WFC. A sample question is 'the demands of my work interfere with my home and family life'.

Analysis and results

Means, standard deviations and correlations for all variables are reported in Table 1. It is interesting to note the low mean values for PS. These values may be explained by the average age and experience of the sample under study. Several studies have found that physicians who are older and have more experience are less likely to have PS (Haar et al, 2014; Shanafelt et al, 2012) and some similar factor may have been operating in this study population. All scales had acceptable reliabilities. The alpha reliabilities are reported along the diagonal in the table.

Table 1. Means, standard deviations and zero-order correlations among study variables								
Variables	Mean	Standard	MWB	JS	CS	PS	WFC	
		deviation						
Mental well being	3.85	.73	0.92					
Job satisfaction	3.80	.79	.507**	0.80				
Career satisfaction	3.55	.79	.382**	.652**	0.73			
Perceived stress	1.58	.48	310**	436**	312**	0.76		
Work-family conflict	3.50	1.75	176	372**	446**	.497**	0.93	
						÷		

** p < .01, Figures along the diagonal represent the alpha reliabilities of the scales in the study

ORIGINAL RESEARCH

	Table 2. Regression results								
	Independent variable	Dependent variable	Unstandardized β	t	R2	Δ R2			
H1	Perceived stress	Mental well being	369	369	-2.329*	.126 .053*			
		Job satisfaction	615	-3.976**	.282	.132**			
		Career satisfaction	506	-2.831**	.200	.075**			
H2	Work–family conflict	Mental well being	064	-1.417	.098	.022			
		Job satisfaction	126	-3.151**	.245	.092**			
		Career satisfaction	182	-4.164**	.286	.155**			
H3	Job satisfaction	Mental well being	.453	4.755**	.252	.188**			
	Career satisfaction	Mental well being	.372	3.968**	.213	.144**			
H4	Perceived stress and	Mental well being	220	-1.286	.275	.127**			
	job satisfaction		.422	3.852**					
H4	Perceived stress and	Mental well being	319	.1.890	.245	.096**			
b	career satisfaction		.320	3.271**					
	Control variables: age work experience *p< 05 ** p < 01								

Moderate to strong correlations were found between MWB, JS, CS, PS and WFC.

We used multiple regression analysis to test the hypotheses. We controlled for age and education and entered these variables into model 1. The results in *Table 2* (controlling for age and educational qualification) show partial support for H2 and complete support for H1 and H3. Surprisingly, WFC and MWB were not significantly correlated.

A precondition of testing for mediation (H4 and H5) is that all variables should have significant bivariate correlations with each other (Baron and Kenny, 1986). Since WFC did not appear to be significantly correlated with MWB, the basic criterion for the test of mediation was not met and we could not test for mediation for H5 (H5 was not supported). We tested for mediation of JS and CS in the link between PS and MWB. We regressed PS on MWB and added JS and CS to the regression equation. When JS and CS were added, PS became insignificant, indicating complete mediation.

Discussion

As expected, we found that PS significantly negatively impacts the satisfaction (both JS and CS) and MWB of physicians, although JS was more significantly impacted than CS. This may occur because a person is less impacted by various concomitant stressors and strain if there is growth and development in his or her career. Similar to PS, WFC also negatively impacted both the JS and CS of physicians. This significant association found between WFC, JS and CS suggests that competing roles may reduce the satisfaction of an individual with one such role, depending on the priorities of the person. Older individuals can better manage competing roles. A longitudinal study is required to understand the impact of various life stages on WFC and CS (Darcy et al, 2012).

PS, JS and CS were also significantly negatively correlated to MWB. Occupational stressors may reduce the sense of achievement and fulfilment gained by working in roles that are otherwise high on meaning and personally rewarding. It is surprising to note the absence of significant association between WFC and MWB. We can only attribute this to the higher age group of the physicians surveyed. It may be that by the time people reach their forties, the challenges of managing conflicting responsibilities such as child care are reduced (Byron, 2005) and they are better able to cope with conflicts.

The full mediation by JS suggests that it may help physicians cope with the stressors inherent in their work. However, this result needs to be interpreted with some caution, given the low levels of PS in the original sample. Future studies must look at samples with relatively higher PS and retest these hypotheses. Though the original research plan had been to assess physicians based on their specialties, we were unable to do so with the limited sample available. Future studies should consider the stress-and-WFC-

Ę

related challenges of various specialists and how CS and JS impact their MWB.

Recommendations

The present study suggests that JS and CS play important roles in reducing the impact of PS on the MWB of physicians. Even though organisations cannot do much about the nature of a physician's work to reduce their level of PS, they can work on improving the factors that enhance JS and CS and on eliminating the factors that reduces them. Based on our findings, we now look at the mediating variables in this study and provide recommendations to enhance the JS and CS of physicians.

Extensive research suggests that factors such as working conditions (Atakli and Kmeta, 2003), work diversity (Kapur, 1999), relations and contacts with colleagues (Buciuniene et al, 2005), freedom/autonomy at work and amount of responsibility given (O'Sullivan, 2005), opportunity to have more control over one's work (McGlone and Chenoweth, 2001) and relationship with patients (Kalda et al, 2000) largely contribute to physician satisfaction. Factors such as long working hours, work overload, low income levels, administrative burdens, lack of recognition and WFC due to inadequate time for personal and family activities reduce physician satisfaction (Van Ham et al, 1998).

Similarly, developmental career prospects are important for the retention of staff (Doherty, 2009). The presence of interesting and challenging work opportunities can contribute to theorganizational loyalty and JS of employees (Sturges and Guest, 2001). Thus, organisations need to work on ways to enhance the JS and CS of their employees. A recent study (Martins et al, 2002) found that the minority gender often has a less supportive relationship with their coworkers than the majority gender which reduces their satisfaction at work. It also found that financial resources did not moderate the relation between WFC and CS. This clearly highlights that a high-paying job alone is not sufficient to reduce WFC. Organizations need to work on providing a socially supportive environment to all its members, especially to the minority gender, to increase their CS.

Conclusion

MWB is important for individuals to live a healthy and balanced life. High positive MWB is associated with many positive workrelated outcomes like creativity, flexibility, better performance of complex tasks, positive judgment of others and positive self-perception (Lyubomirsky et al, 2005; Robertson and Taylor, 2010). Higher levels of PS damage the MWB of physicians. The present study provides suggestions for enhancing the MWB of physicians by improving their JS and CS.

A physician's role is inherently stressful, but a sense of satisfaction in their work and careers provides them with confidence in their skills and a feeling of empowerment that they can influence their work-related outcomes and provide them with meaning while acting out their occupation as a physician to support their MWB. BIHCM

References

Arnetz B, Hörte L, Hedberg A, Theorell T, Allander E, Malker H (1987) Suicide patterns among physicians related to other academics as well as to the general population. Acta Psychiatr Scand 75(2): 139-43.

Asamani J, Amertil N, Chebere M (2015) The influence of workload levels on performance in a rural hospital. British Journal of Healthcare Management 21(12): 577-86.

Atakli A, Kmeta E (2003) Managers' and secretaries' views of job satisfaction. British Journal of Healthcare Management 9(8): 265-270.

Bacharach S, Bamberger P, Conley S (1991) Work□ home conflict among nurses and

engineers: Mediating the impact of role stress on burnout and satisfaction at work. Journal of Organizational Behaviour 12(1): 39-53.

Baron R, Kenny D (1986) The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. Journal of Personality and Social Psychology 51(6): 1173-1182.

Bliese P, Edwards J, Sonnentag S (2017) Stress and well-being at work: A century of empirical trends reflecting theoretical and societal influences. Journal of Applied Psychology 102(3): 389-402.

Buciuniene I, Blazeviciene A, Bliudziute E (2005) Health care reform and job satisfaction of primary health care physicians in Lithuania. BMC Family Practice 6(1): 10 -16.

Byron, K (2005) A meta-analytic review of work-

KEY POINTS

- Perceptions of stress can significantly influence the job and career satisfaction and mental well-being of physicians
- Work-family conflict has a significant impact on job and career satisfaction
- Job and career satisfaction can reduce the negative impact of stress on well-being
- Organisations need to enhance the job and career satisfaction of physicians through greater variety, supportive work environments and opportunities for career growth

family conflict and its antecedents. Journal of Vocational Behavior 67(2):169-198.

Chambers R, Belcher J (1993) Comparison of the health and lifestyle of general practitioners and teachers. British Journal of General Practice 43(374): 378-382.

Chandraiah K, Agrawal S, Marimuthu P, Manoharan N (2003) Occupational stress and job satisfaction among managers. Indian Journal of Occupational and Environmental Medicine 7(2): 6-11.

Clever L (2002) Who is sicker: patients—or residents? Residents' distress and the care of patients. Annals of Internal Medicine 136(5): 391-93.

Cohen S, Kamarck T, Mermelstein R (1983) A global measure of perceived stress. Journal of Health and Social Behaviour 24(4): 385-96.

Collum J, Moreton A, Booth T (2013) Can professionalism and regulation of hours co- exist? British Journal of Healthcare Management, 19(6): 276-83.

Cooper C, Rout U, Faragher B (1989) Mental health, job satisfaction, and job stress among general practitioners. British Medical Journal, 298(6670): 366-70.

Darcy C, McCarthy A, Hill J, Grady G (2012) Work–life balance: One size fits all? An exploratory analysis of the differential effects of career stage. European Management Journal 30(2): 111-20.

DiMatteo M, Sherbourne D, Hays R, Ordway L, Kravitz L, McGlynn E Rogers W (1993) Physicians' characteristics influence patients' adherence to medical treatment: results from the Medical Outcomes Study. Health Psychology 12(2): 93-102.

Doherty C (2009) Modernising nursing careers. British Journal of Healthcare Management 15(2): 81-86.

Dyrbye L, Varkey P, Boone S, Satele D, Sloan J, Shanafelt T (2013) Physician satisfaction and burnout at different career stages. In Mayo Clinic Proceedings, 88(12): 1358-1367. Elsevier.

Elston M (2009) Women and medicine: the future. London: Royal College of Physicians. Firth-Cozens J (2001) Interventions to improve physicians' well-being and patient care. Social Science & Medicine 52(2): 215-22. Firth-Cozens J (2006) A perspective on stress and depression - understanding doctors' performance, Edited by: Cox J, King J, Hutchinson A, and McAvoy P. Oxford: Radcliffe Publishing, pp. 22-37.

Friedberg M, Chen P, Van Busum K (2014) Factors affecting physician professional satisfaction and their implications for patient care, health systems, and health policy. Rand Health Quarterly 3(4): 1.

Frone M, Russell M, Barnes G (1996) Work–family conflict, gender, and health-related outcomes: A study of employed parents in two community samples. Journal of Occupational Health Psychology 1(1): 57-69.

Frone M, Russell M, Cooper M (1992) Antecedents and outcomes of work-family conflict: testing a model of the work-family interface. Journal of Applied Psychology 77(1): 65-78. Greenhaus J, Parasuraman S (1986) Vocational and organizational behaviour, 1985: A review. Journal of Vocational Behaviour 29(2): 115-76.

Haar J, Russo M, Suñe A, Ollier-Malaterre A (2014) Outcomes of work–life balance on job satisfaction, life satisfaction and mental health: A study across seven cultures. Journal of Vocational Behaviour 85(3): 361-73.

Hughes J, Bozionelos N (2007) Work-life balance as source of job dissatisfaction and withdrawal attitudes: An exploratory study on the views of male workers. Personnel Review 36(1): 145-54.

Judge T, Thoresen C, Bono J, Patton G (2001) The job satisfaction–job performance relationship: A qualitative and quantitative review. Psychological Bulletin 127(3): 376-407. Kalda R, Maaroos H, Lember M (2000) Motivation and satisfaction among Estonian family doctors working in different settings. The European Journal of General Practice 6(1): 15-19. Kapur N, Appleton K, Neal R (1999) Sources of job satisfaction and psychological distress in GPs and medical house officers. Family Practice 16(6): 600-601.

Keeton K, Fenner D, Johnson T, Hayward R (2007) Predictors of physician career satisfaction, work–life balance, and burnout. Obstetrics & Gynaecology 109(4): 949-55. Kroenke K, Spitzer R, Williams J, Monahan

đ

P, L we B (2007) Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Annals of Internal Medicine 146(5): 317-25.

Leigh J, Kravitz R, Schembri M, Samuels S, Mobley S (2002) Physician career satisfaction across specialties. Archives of Internal Medicine 162(14): 1577-84.

Levinson W, Pizzo P (2011) Patient-physician communication: it's about time.JAMA 305(17): 1802-03.

Linzer M, Konrad T, Douglas J (2000) Managed care, time pressure, and physician job satisfaction: results from the physician work life study. Journal of General Internal Medicine 15(7): 441-50.

Lyubomirsky S, King L, Diener E (2005) The benefits of frequent positive affect: Does happiness lead to success? Psychological Bulletin 131(6), 803-855.

McGlone S, Chenoweth I (2001) Job demands and control as predictors of occupational satisfaction in general practice. The Medical Journal of Australia 175(2): 88-91.

Martins L, Eddleston K, Veiga J (2002) Moderators of the relationship between work-family conflict and career satisfaction. Academy of Management Journal 45(2): 399-409.

Netemeyer R, Boles J, McMurrian R (1996) Development and validation of work–family conflict and family–work conflict scales. Journal of Applied Psychology 81(4): 400-10 Nylenna M, Gulbrandsen P, Førde R, Aasland O (2005) Job satisfaction among Norwegian general practitioners. Scandinavian Journal of Primary Health Care 23(4): 198-202. O'Sullivan B, Keane A, Murphy A (2005) Job stressors and coping strategies as predictors of mental health and job satisfaction among Irish general practitioners. Irish medical journal 98(7): 199-200.

Parasuraman S, Purohit Y, Godshalk V et al (1996). Work and family variables, entrepreneurial career success and psychological well-being. Journal of Vocational Behaviour 48(3): 275-300.

Phipps L (1988) Stress among doctors and nurses in the emergency department of a general hospital. Canadian Medical Association Journal 139(5): 375-76.

Rathi N, Rastogi R (2007) Meaning in life and psychological well-being in pre-adolescents and adolescents. Journal of the Indian Academy of Applied Psychology 33(1): 31-38. Richardson K, Rothstein H (2008) Effects of occupational stress management intervention programs: a meta-analysis. Journal of Occupational Health Psychology 13(1): 69-93. Robertson I, Taylor J (2010) Wellbeing in healthcare organizations: key issues. British Journal of Healthcare Management 16(1): 19-25.

Rout U, Cooper C, Rout J (1996) Job stress among British general practitioners: predictors of job dissatisfaction and mental ill-health. Stress Medicine 12(3): 155-166.

Sekaran U (1983) How husbands and wives in dualcareer families perceive their family and work worlds. Journal of Vocational Behaviour 22(3): 288-302.

Shanafelt T, Balch C, Bechamps G, et al (2009). Burnout and career satisfaction among American surgeons. Annals of Surgery 250(3): 463-71.

Shanafelt T, Boone S, Tan L, et al (2012) Burnout and satisfaction with work-life balance among US physicians relative to the general US population. Archives of Internal Medicine 172(18): 1377-85.

Spector P (1997) Job satisfaction: Application, assessment, causes, and consequences. Sage publications. Behavior 49(1): 99-118.

Sturges J, Guest D (2001) Don't leave me this way! A qualitative study of influences on the organisational commitment and turnover intentions of graduates early in their career. British Journal of Guidance and Counselling 29(4): 447-462.

Sutherland V, Cooper C (1992) Job stress, satisfaction, and mental health among general practitioners before and after introduction of new contract. British Medical Journal 304(6841): 1545-48.

Tenbrunsel A, Brett J, Maoz E, Stroh L, Reilly A (1995) Dynamic and static work-family relationships. Organizational Behavior and Human Decision Processes 63(3) pp. 233-246 Tennant R, Hiller L, Fishwick R, et al (2007) The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. Health and Quality of life Outcomes 5(1): 63.

Thomas L, Ganster D (1995) Impact of familysupportive work variables on work-family conflict and strain: A control perspective. Journal of Applied Psychology 80(1): 6-15.

Van Ham I, Verhoeven A, Groenier K, Groothoff J. De Haan J (2006) Job satisfaction among general practitioners: a systematic literature review. The European Journal of General Practice 12(4): 174-180.

Wallace J, Lemaire J, Ghali W (2009) Physician wellness: a missing quality indicator. The Lancet 374(9702): 1714-21.

Williams E, Konrad T, Scheckler W (2000) The effects of job satisfaction and perceived stress on the physical and mental health and withdraw intentions of physicians. Academy of Management Proceedings 2000(1): C1-C6.