Knowledge Discovery from Dynamically Evolving User Profiles

Md. Ahemad Pasha
Research Scholar, Dept. of CSE
SR Engineering College
Warangal, Andhra Pradesh

R. Vijaya Prakash
Associate Professor
SR Engineering College
Warangal, Andhra Pradesh

ABSTRACT

Computer users play important role in all organizations. The users are software engineers, data entry or other kind. However having know how about the users is very useful as it helps in predicting their actions and finding whether they are compatible with their job profile, assisting them to do their job in better way and also detect adversaries or masqueraders. In this paper it analyzes computer user behavior based on the commands that they execute as part of their job profile. The commands are compared with sequence of commands associated with their profile in order to generate user behavior profiles. The profiles are evolved over a period of time and they are dynamic as the users get different positions in company and get associated with various job roles. For this reason the user behavior profiles are continually updated on order to ensure that they reflect the true knowledge of the behavior of users with respect to their job roles. It built a prototype application to demonstrate the efficiency of this approach. The empirical results revealed that this approach is useful in helping users and also monitoring their behavior in work environment and take necessary actions.

Index Terms –User behavior profiles, job roles, classifiers, user modeling

1. INTRODUCTION

Analyzing behavior of end users who operates computers in offices as part of their job roles is an important activity which has utility in the real world applications. The user behavior can be used to take necessary steps based on the result of analysis. The behavior profile [1] which evolves over a period of time based on the commands issued by userscan assist making well informed decisions that can be used to assist a new user, give training to a user, avoid attacks, and predict future actions of end users and so on. The availability of user profiles is very essential and that can be used for data mining activities. In this paper it uses various generic algorithms or procedures that can be used for user behavior profile management. According to this proposed system every user in the organization is monitored for extracting knows how pertaining to the behavior of that person. Creating user profiles is also discussed in [2]. In this paper different models are used to manage user behavior profile including EVABCD. Prior research made use of detecting masqueraders while analyzing UNIX commands [3], [4].

In this paper users are assumed to be the users working for some development company where UNIXcommands are to be issued regularly. However, the commands generally used by users are pertaining to their job role. When any user issues commands which are not related to their job role, or else issue commands that are not safe, it is inevitable to discuss the scenario and take necessary steps. The user behavior is not static. Therefore the user profiles once created are updated or evolved continually. It is used various methods to implement the concept of creating evolving user behavior profiles. The remainder of this paper is structured as follows. Section II reviews literature. Section III presents the proposed approach. Section IV presents information about prototype application. Section V presents experimental results while section VI concludes the paper.

2. RELATED WORKS

History of navigation is recorded in WebMemex proposed by Dacado et al. [5].Logistic regression modeling and queuingtheory were proposed by Pepyne et al. [6] for making profiles of computer users. Readable user profiles were generated by Gody and Amandi [7]. Problem of sequence classification is similar to the problem of user behavior profiles [8]. Unlabeledsequence data is used in the experiments by Kaminka [9] for maintaining behavior patterns. IBL (Intance Based Learning) issued to reduce the storage required for processing in [10]. Number of statistical approachedstop work with user profiles is presented in [10]. Evolving fuszzy classifiers was proposed by Zhou [11]. The performance of various classification algorithms are made by Panda and Petra[12]. Classification accuracy of J48, ID3, NBTree, SimpleCART, and NB are presented in [13]. In [14] an incremental algorithm is presented. Decision trees were explored in [15] and [16]. In [17] and [18] artificial neural networks are presented for clustering processes. Fuzzy neural network for evolving user profiles is used in [19]. The classifiers can also use the concept of drift and shift as explored in [20].

SVM is one of the data mining algorithms that can be used for classification. Incremental learning is possible with some sort of clustering algorithm. Such algorithmic proposed by Xiao et al. [21] which make use of SV set and discovers actionable knowledge. An incremental learning algorithm should have the following characteristics.

1. Learning additional information.
2. No need to have access to original data.
3. Preserving knowledge acquired previously.
4. Ability to accommodate new classes

3. PROPOSED APPROACH

Proposed approach in this paper is influenced by [22]. The aim of the approach is to create and evolve user behavior profiles that can give knowledge required to take necessary steps based on the user profiles. User behavior prediction helps in making well informed decisions in organizations. The approach it followed is summarized as follows. More details can be found in [22].

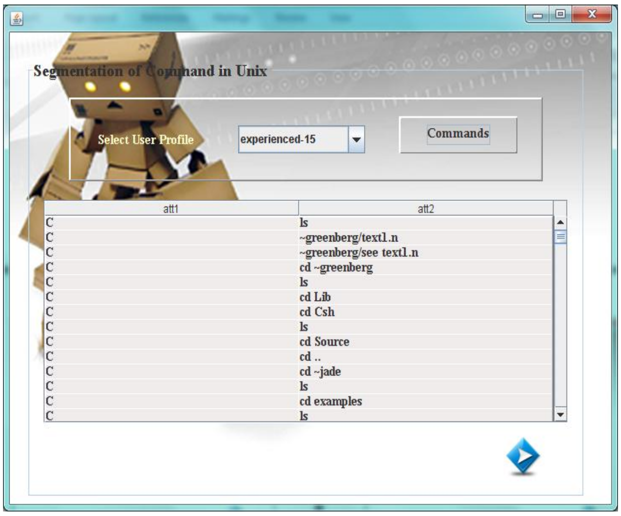
1. Creating and evolving the classifier
    a. Creating the user behavior profiles
    b. Evolving the classifier
2. User classification

The user profile construction needs certain steps to be carried out. It can be done using three important steps which are as follows.

1. Segmentation of sequence of commands
2. Storage of the subsequences in a data structure
3. Creation of the user profile

4. PROTOTYPE APPLICATION

It's built a prototype application to demonstrate the proof of concept. The application is built in Java platform with graphical user interface (GUI). The environment used to build the application is a PC with 4 GB RAM, core 2 dual processor running Windows 7 operating system. The datasets used include the UNIX or Linux commands which are stored in different files based on the job profiles. These files can be updated on the fly with new commands as per the user's changing job profile. Figure 1 illustrates the selection of user profile and shows list of associated commands.



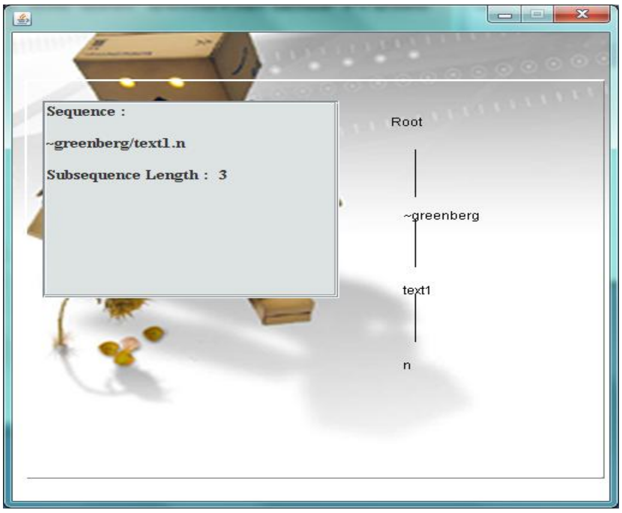
Segmentation of Command in Unix

Select User Profile: experienced-15

att1	att2
C	ls
C	~greenberg/text1.n
C	~greenberg/see text1.n
C	cd ~greenberg
C	ls
C	cd Lib
C	cd Csh
C	ls
C	cd Source
C	cd ..
C	cd ~jade
C	ls
C	cd examples
C	ls

Fig. 1 –Job profile and associated commands

As shown in figure 1, the "experienced – 15" user has job profile which enables him to make use of the commands listed. These commands are his legitimate commands as per the job profile. However, he may run other commands as well. When he executes other commands there might be many reasons. For instance the user is not having knowledge to execute correct commands (training issue); the user is a hacker tying to hack (assume the user is an adversary); and the user is just learning new commands. Based on the commands user executes, the behavior profile is built. This will help to analyze the behavior of user and take necessary steps. Figure 2 shows the evolving user profile graphically.



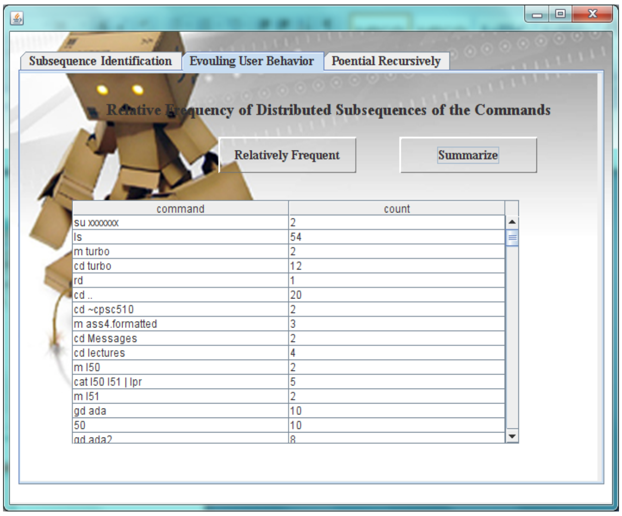
Sequence :

~greenberg/text1.n

Subsequence Length : 3

Fig. 2 –User profile evolving

As seen in figure 2, it is evident that the user's commands are analyzed and the behavior profile is created. Moreover the profile is graphically shown. This behavior is dynamic and it affects the representation of it on figure 2. The proposed system keeps track of the commands executed by the user who logged into the system. Figure 3 shows the breakup of the number of commands that have been issued by the user.

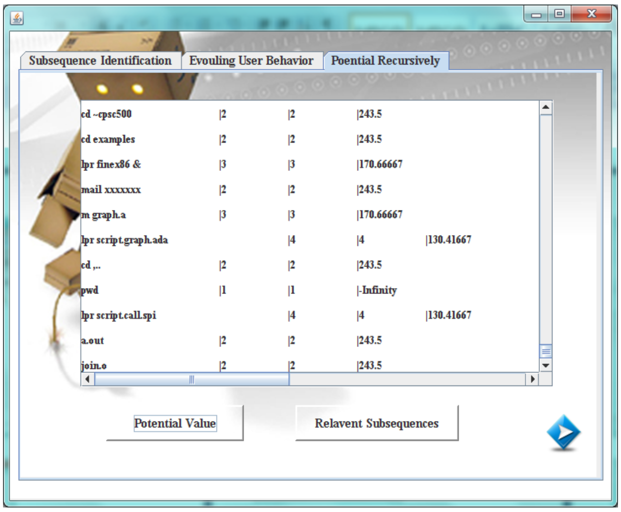


Relative Frequency of Distributed Subsequences of the Commands

command	count
su xxxxxxx	2
ls	54
m turbo	2
cd turbo	12
rd	1
cd ..	20
cd ~cpsc510	2
m ass4.formatted	3
cd Messages	2
cd lectures	4
m l50	2
cat l50 l51 | lpr	5
m l51	2
gd ada	10
50	10
gd ada2	8

Fig. 3 – Breakup of commands

As seen in figure 3, the commands associated with user profile are kept track as user operates the system and issues his commands. As the commands are recorded, their break up is also obtained and presented. This will help in finding user Bahadur towards individual commands. The commands might be used by multiple users' recursively. Figure 4 shows the recursive potential.



```
cd ~cpsc500          |2       |2       |243.5
cd examples          |2       |2       |243.5
lpr finex86 &        |3       |3       |170.66667
mail xxxxxxx         |2       |2       |243.5
m graph.a            |3       |3       |170.66667
lpr script.graph.ada          |4       |4         |130.41667
cd ,..               |2       |2       |243.5
pwd                  |1       |1       |-Infinity
lpr script.call.spi           |4       |4         |130.41667
a.out                |2       |2       |243.5
join.o               |2       |2       |243.5
```

Fig. 4 – Illustrates recursive potential

As shown in figure 4, it is evident that the commands are recursively used by multiple users online. Same command might be part of multiple users' profiles. This is because a job profile is associated with multiple users.

5. EXPERIMENTAL RESUTLS

Experiments are made with the prototype application that it built. Experimental are made with users of various profiles with pre-defined commands associated with their profiles. These profiles are changed dynamically or evolved over a period of time as the user continues using commands as part of their work with respect to given job role. The experiments are made in terms of support of UNIX commands and their sub sequence. Number of users vs. classification rate percentage is considered for experiments. Experiment is also done with various algorithms. Figure 5 shows the distribution of sub sequence of commands.

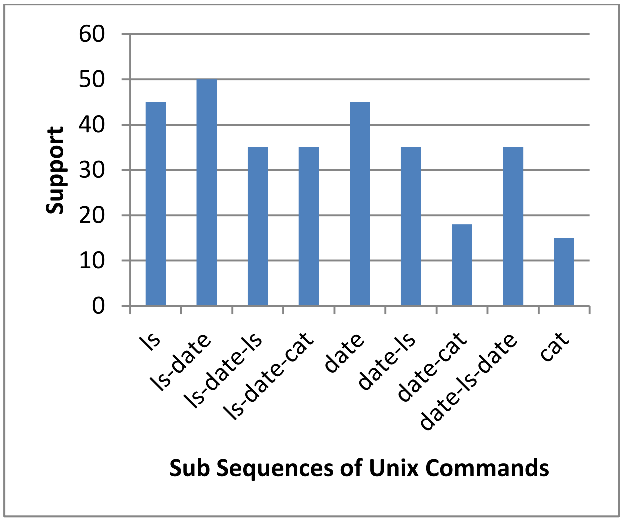


Fig. 5 –Distribution of sub sequence of commands

The UNIX commands are issued by users. The commands issues are tracked and the support of commands that has been computed is presented for various sub sequence commands. The results reveal that ls-date sub sequence has highest usage in the profiles.

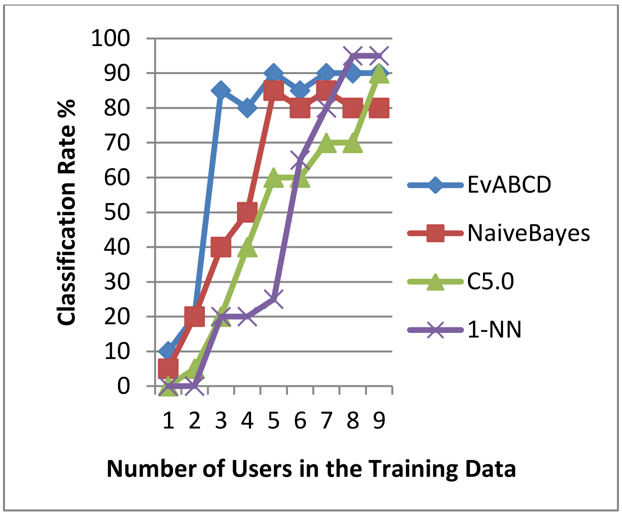


Fig. 6 –Evolution of classification rate

As can be seen in figure 6, the horizontal axis represents number of users in the training data while the vertical axis represents classification rate. As per the results EVABCD quickly evolves and adapts to new class.

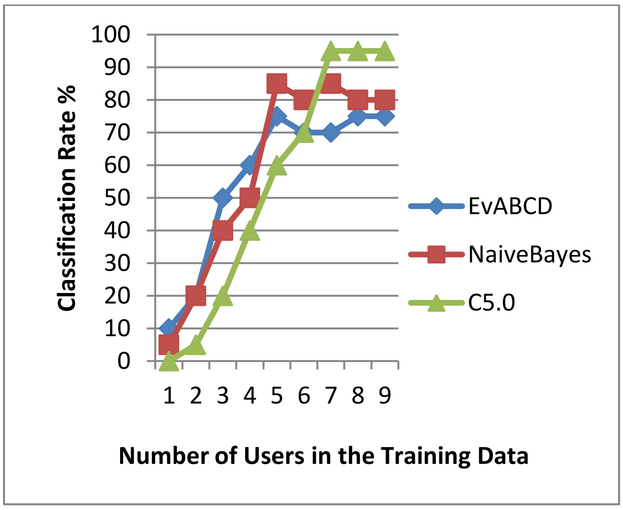


Fig. 7 – Evolution of classification rate

As can be seen in figure 7, the horizontal axis represents number of users in the training data while the vertical axis represents classification rate. As per the results EVABCD quickly evolves and adapts to new class.

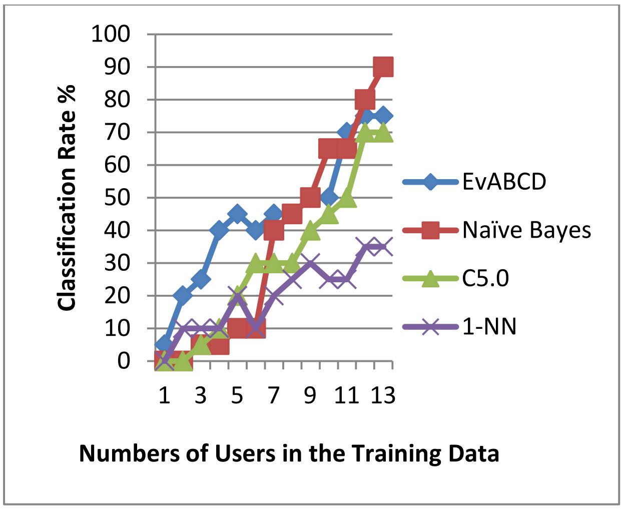


Fig. 8 – Evolution of classification rate

As can be seen in figure 8, the horizontal axis represents number of users in the training data while the vertical axis represents classification rate. As per the results EVABCD quickly evolves and adapts to new class.

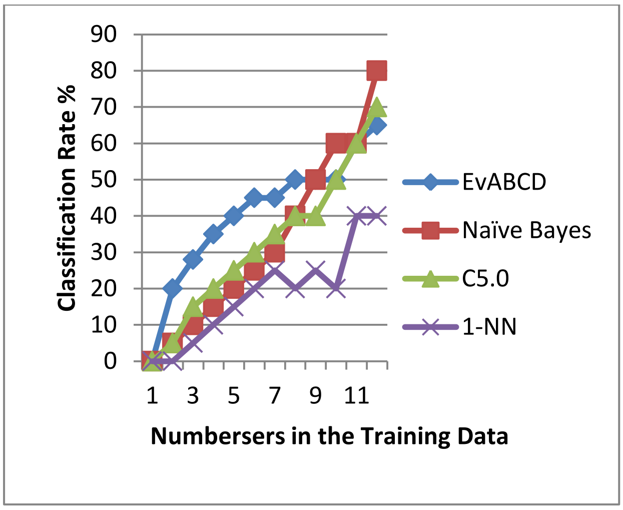


Fig. 9 – Evolution of classification rate

As can be seen in figure 9, the horizontal axis represents number of users in the training data while the vertical axis represents classification rate. As per the results EVABCD quickly evolves and adapts to new class.

6. CONCLUSION

In this paper it is focused on user behavior profile creation and evolving it over a period of time in order to analyze user behavior in an organization. Computer users play different roles in organizations. Each job role needs certain commands. It assumes the operating system as UNIX or Linux. The commands of OS are pertaining to job roles. When job role is changes the kind of commands issues will get changed. This is the basis for the user behavior profile evolving concept. It built user behavior profiles and keeps track of them as user issues various commands with the system. Theses user behavior profiles are used to analyze the behavior of users and take well informed decisions. It also built a prototype application that demonstrates the functionality of the proposed system. The experimental results revealed that the application is very useful in the real world.

7. REFERENCES

[1] D. Godoy and A. Amandi, “User Profiling in Personal InformationAgents: A Survey,” Knowledge Eng. Rev., vol. 20, no. 4, pp. 329-361, 2005.

[2] J.A. Iglesias, A. Ledezma, and A. Sanchis, “Creating User Profilesfrom a Command-Line Interface: A Statistical Approach,” Proc.Int’l Conf. User Modeling, Adaptation, and Personalization (UMAP),pp. 90-101, 2009.

[3] M. Schonlau, W. Dumouchel, W.H. Ju, A.F. Karr, and Theus,“Computer Intrusion: Detecting Masquerades,” Statistical Science,vol. 16, pp. 58-74, 2001.

[4] R.A. Maxion and T.N. Townsend, “Masquerade Detection UsingTruncated Command Lines,” Proc. Int’l Conf. Dependable Systemsand Networks (DSN), pp. 219-228, 2002.

[5] A. Alaniz-Macedo, K.N. Truong, J.A. Camacho-Guerrero, and M.Graca-Pimentel, “Automatically Sharing Web Experiencesthrough a Hyperdocument Recommender System,” Proc. ACMConf. Hypertext and Hypermedia (HYPERTEXT ’03), pp. 48-56, 2003.

[6] D.L. Pepyne, J. Hu, and W. Gong, “User Profiling for ComputerSecurity,” Proc. Am. Control Conf., pp. 982-987, 2004.

[7] D. Godoy and A. Amandi, “User Profiling for Web PageFiltering,” IEEE Internet Computing, vol. 9, no. 4, pp. 56-64, July/Aug. 2005.

[8] J. Anderson, Learning and Memory: An Integrated Approach. JohnWiley and Sons, 1995.

[9] Y. Horman and G.A. Kaminka, “Removing Biases in UnsupervisedLearning of Sequential Patterns,” Intelligent Data Analysis,vol. 11, no. 5, pp. 457-480, 2007.

[10] T. Lane and C.E. Brodley, “Temporal Sequence Learning and DataReduction for Anomaly Detection,” Proc. ACM Conf. Computer andComm. Security (CCS), pp. 150-158, 1998.

[11] P. Angelov and X. Zhou, “Evolving Fuzzy Rule-Based Classifiersfrom Data Streams,” IEEE Trans. Fuzzy Systems: Special Issue onEvolving Fuzzy Systems, vol. 16, no. 6, pp. 1462-1475, Dec. 2008.

[12] M. Panda and M.R. Patra, “A Comparative Study of Data MiningAlgorithms for Network Intrusion Detection,” Proc. Int’l Conf.Emerging Trends in Eng. and Technology, pp. 504-507, 2008.

[13] A. Cufoglu, M. Lohi, and K. Madani, “A Comparative Study ofSelected Classifiers with Classification Accuracy in User Profiling,”Proc. WRI World Congress on Computer Science and InformationEng. (CSIE), pp. 708-712, 2009.

[14] R. Polikar, L. Upda, S.S. Upda, and V. Honavar, “Learn++: AnIncremental Learning Algorithm for Supervised Neural Networks,”IEEE Trans. Systems, Man and Cybernetics, Part C (Applications and Rev.), vol. 31, no. 4, pp. 497-508, http://dx.doi.org/10.1109/5326.983933, Nov. 2001.

[15] D. Kalles and T. Morris, "Efficient Incremental Induction ofDecision Trees," Machine Learning, vol. 24, no. 3, pp. 231-242, 1996.

[16] F.J. Ferrer-Troyano, J.S. Aguilar-Ruiz, and J.C.R. Santos, "DataStreams Classification by Incremental Rule Learning with ParameterizedGeneralization," Proc. ACM Symp. Applied Computing

[17] G.A. Carpenter, S. Grossberg, and D.B. Rosen, "Art2-a: AnAdaptive Resonance Algorithm for Rapid Category Learningand Recognition," Neural Networks, vol. 4, pp. 493-504, 1991.

[18] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds, andD.B. Rosen, "Fuzzy Artmap: A Neural Network Architecture forIncremental Supervised Learning of Analog MultidimensionalMaps," IEEE Trans. Neural Networks, vol. 3, no. 5, pp. 698-713, Sept.1992.

[19] N. Kasabov, "Evolving Fuzzy Neural Networks for Supervised/Unsupervised Online Knowledge-Based Learning," IEEE Trans.Systems, Man and Cybernetics—Part B: Cybernetics, vol. 31, no. 6,pp. 902-918, Dec. 2001.

[20] G. Widmer and M. Kubat, "Learning in the Presence of ConceptDrift and Hidden Contexts," Machine Learning, vol. 23, pp. 69-101, 1996.

[21] R. Xiao, J. Wang, and F. Zhang, "An Approach to IncrementalSVM Learning Algorithm," Proc. IEEE Int'l Conf. Tools withArtificial Intelligence, pp. 268-278, 2000.

[22] Jose Antonio Iglesias, Plamen Angelov, Agapito Ledezma and Araceli Sanchis, "Creating Evolving User Behavior Profiles Automatically". IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 5, MAY 2012.

AUTHORS

Md. Ahemad Pasha,

Pursuing M.Tech (CSE) in SR Engineering College, Warangal, Andhra Pradesh, INDIA. Received a B.Tech Degree in Computer Science and Engineering. My main research interest includes Data Mining

Prof. R. Vijaya Prakash

Associate Professor in the Department of Computer Science & Engineering, SR Engineering College, Warangal, India. He has 15 Years of Teaching Experience. His areas of interest are Data Mining and Warehousing. He has published papers in International Journal, International Conference and National Conference.