International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 2 Issue 1, January- 2013

A Survey on Analysis of Evolving User Behavior Profile for Predicting Masquerades

Devendra B. Dandekar, Vinay S. Kapse
Department of Computer Science and Engineering
Tulsiramji Gaikwad-Patil
College of Engineering, RTMNU, Nagpur

Abstract

Knowledge about computer User is very essential. In this survey paper we approach for recognition of user behavior is very beneficial for assisting & predicting their future actions. By considering most existing techniques available as assuming handcrafted user profile which encode repertoire of the observed user. We find out from this survey actual causes however, the construction of effective user profile problematic to human behavior is often erratic and sometimes it is different for their change of goals. The novel approach applicable to any problem of dynamic evolving user behavior modeling where it can be represented as a sequence of actions or events. We will be able to see challenges from this survey with comparative study during creating evolve system approach & predict it.

Keywords - Evolving Fuzzy Systems, Fuzzy-Rule Based (FRB) Classifier, User Modeling.

1. Introduction

Knowledge about computer user is very beneficial to assist, to predict for creating & recognize behavior of profile. The recognition of other behavior profile in real time significant offers different tasks such as to predict their future action. Specifically, computer user modeling learned about ordinary observing user to promote a way of experience user profile. However, the construction of effective user profile problematic to human behavior is often erratic and sometimes it is different for their change of goals. There exists several definition for user profile [1]. It defined as description of user interests, characteristics, behaviors and preferences. In recent years, significant work has been carried out for profiling to the environment and new goals of the user. Example behind this profile which proposed in a previous work [2]. This paper proposed an adaptive approach for creating behavior and recognizing computer users. We approach (EVABCD) Evolving Agent behavior Classification based on Distribution of relevant events based on representing observed behavior user.

The UNIX operating system environment is used in this research for explaining and evaluating EVABCD. A user behavior is represented as a sequence of UNIX command in command-line interface. Previous research studies in this environment [3],[4] focus on detecting masquerades who individual impersonate other users on computer networks and system. However, EVABCD creates evolving user profiles and classifies new users into one of the previously created profiles. Thus the goal of EVABCD in UNIX environment can devided into two phases :

1.1 Creating and updating user profiles from the commands the users typed in UNIX shells.
1.2 Classifying a new sequence of commands into the predefined profiles.

Because we used an evolving classifier which constantly learned and adapting existing classifier structure. It accommodate the newly observed emerging behaviors, once a user is classified, relevant action can be done. But this task is not addressed to previous work. The techniques Macedo et al, [5] used to find out relevant information related to human behavior based on the captured history of navigation.

In Summary, our Contributions are:
• We discover the limitations and their root causes when creating user behavior profile in terms of classifying relevant sequence of events.
• We generalize proposed previous work regarding knowledge about computer user with increased complexity of thinking user behavior.
• We extend new algorithm to execute the environments in which segmentation of subsequent relevant events evaluated by using frequency based method.

• A comparative study to revise existing hypothesis than it is to generate hypothesis when each time new instance is observed.
• To detect Masquerades (Un-Authorized work) when it tends to knowledge of computer user.

2. Literature Review

Various approaches have been proposed as literature point of view that user profile usually changes to recognize behavior of others in real-time. To predict, to coordinate, to recognize human brain capacity for future actions. Different methods have been used to find out relevant information in computer user behavior in different computer areas :

2.1 Discovery of navigation patterns

Spiliopoulou and Faulstich [6] present the Web Utilization Miner (WUM), a mining system for discovering interesting navigation patterns in website.WUM prepares the web log data for mining and the language MINT mining the aggregated data according to the directives of the human expert [7].

2.2 Web recommender systems

Macedo et al. [8] propose a system (WebMemex) that provides recommended information based on the captured history of navigation from a list of known users. WebMemex captures information such as IP addresses, user Ids and URL accessed for future analysis.

2.3 Web page filtering

Gody and Amandi [9] present a technique to generate readable user profiles that accurately capture interests by observing their behavior on the Web. The proposed technique is built on the Web Document Conceptual Clustering algorithm, with which profiles without an a priori knowledge of user interest categories can be acquired.

2.4 Computer security

Pepyne et al [10] describe a method using queuing theory and logistic regression modeling methods for profiling computer users based on simple temporal aspects of their behavior.

3. Motivation & Preliminary

We show the limits of existing construction user behavior which based on evolving profile-library approaches with statistical classification method. "Figure 1. shows the high-level system architecture framework for Evolving Behavior Model Library Profile" to extract significant pieces of sequence of command. According to this aspect, it was used in [11] order to get representative set of subsequences from the acquired sequence. By using trie data structure [12] to learn a team behavior and [13] to classify the behavior patterns of a RiboCup soccer simulation team.

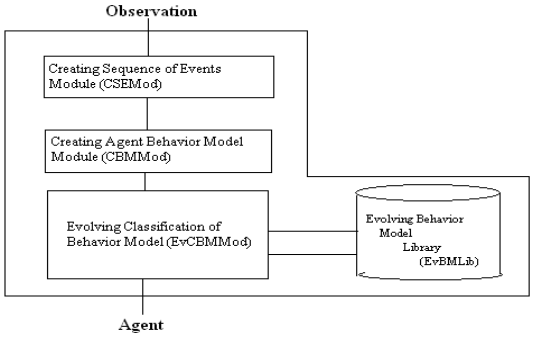

Figure 1. High-level system architectural framework for Evolving Behavior Model Library Profile

4. Existing Effective Classification Techniques In Observed Classifier

Following several incremental effective classifier in evolving fuzzy rule based system which work with automatically gain by observed behavior for adaptive distribution of relevant events. This classifier implemented using different framework.

4.1 Decision trees

The problem of processing data streaming has motivated with development of algorithm which designed to learn dicision trees incrementally [16], [17]. Some algorithm construct incremental decision trees as ID4 [18], ID5 [19], ID5R [20].

4.2 Artificial neural network (ANN)

In order to find out Adapdtive Resonance Theory (ART) networks [21] which proposed unsupervised ANN to carpenter that dynamically determine number of cluster based on a vigilance parameter [22]. In addition, Kasabov porposed other incremental learning neural network call as Evolving Fuzzy Neural Network (EFuNN) [23]. This architecture does not require to see previously data and accommodate new classes. New approach in incremental learning proposed by Seipone and bullinaria [24]. This is an evolutionary algorithm to evolve MLP parameters. This process produce network parameter with better incremental abilities.

4.3 Prototype-based supervised algorithm

Learning Vector Quantization (LVQ) is the nearest prototype learning algorithm [25]. LVQ considered to be a supervised clustering algorithm which each weight vector interpreted as a cluster center. Using this algorithm number of reference vectors has to be set by user. Poirier and Ferrieux proposed a method to generate new porotypes dynamically. LVQ [26] method lacks the generation of prototype for application with noisy data in Dynamic Vector Quantization (DVQ).

4.4 Bayesian classifier

This is an effective methodology for solving classification problem when features are considered simultaneously. However, features of Bayesian classifier forward selection method, huge computation is involved. Agrawal and Bala [27] proposed an incremental versions of Baysian classifier.

4.5 Support Vector Machine (SVM)

A machine which performs classificaition to constructing N-dimensional hyperplane that optimally configured into two categories. Training SVM "incrementally" discard all previous data except their support vectors which gives only approximate result. It is an exact online method to construct the solution. Xiao et al [14] propose which utilize the property of SV set and accumulates distribution of knowledge sample space through adjust table parameter.

However, this research focus command line interface, it is necessary to approachable process in real time streaming data. It is also cope with huge amount of several incremental classifier while structure of incremental classifier assumed to be fixed. They can not address problem of concept drift and shift [15]. By drift they refer to modification of the concept over time and shift. To capture sudden and abrupt changes in streaming data with necessary not only tuning parameter but also change in structure. Taking these aspect when proposed a paper to evolving fuzzy-rule-based system; However, approach has important advantage which makes it very useful in real environments :

- It can cope with huge amounts and data.
- Its evolving structure can capture sudden and abrupt changes in the streams of data.
- Its structure meaning is very clear, as we propose a rule-based classifier.
- It is monitoring in single pass computation with efficient and fast.
- Its classifier structure is simple and interpretable.

5. Investigational Outcome

From this literature survey we investigate how it is more efficient to analysis of evolving user behavior profile in one time solution proposed incremental classifier. It should able additional information from new data. It should not required access original data to train existing classifier. It should preserved previously acquired knowledge. It should accommodate new classes that may be introduced with new data.

6. Comparison of Existing Different Classifier Technique

Below table show the comparative study of different classifier. Difference made on bass of techniques that are used in respective algorithms, advantages and disadvantages.

Table 1. Comparisons of different incremental classifier algorithms

Algorithm/Technique	**Description**	**Advantage**	**Disadvantage**
Instance-Based Learning (IBL) [28]	The first rule find out adding extra relevant information to human behavior..	1.Reduce data storage requirement. 2.Represent as sequence of events.	1.Not interfacing with command-line environment.
Fuzzy Classifier [29]	It is effective for to characterize similarity between sequence of command.	1.Detection of task. 2.Easy to use pairwise sequence alignment.	1.Not Suitable for detecting Masquerades.
Sequential Minimal Optimization (SMO) [31]	The selection based on distribution of large user profile data.	1.Simple and easy to implement. 2.Response time is high. 3.Continuous stream of data.	1.It ignore the fact that user behavior can not change and evolve.
Evolving Profile Library Classifier (EPLib) [32]	The library contain Creating and evolving the classifier with different expected behavior.	1.Evolving learning observation.2. Utilization of evolving influence.	1.library contain different expected behaviors.
Evolving	The library	1.Evolving	1.library

Profile Library Classifier (EPLib) [32]	contain Creating and evolving the classifier with different expected behavior.	learning observation. 2.Utilization of evolving influence.	contain different expected behaviors.
Incremental learning Algorithm (ILA) [33]	It is more efficient to revise existing hypothesis than generate hypothesis.	1.Overcomes incremental classifier 2. Adaptive to dynamic environments	1.Used in complex knowledge streams.

7. Conclusion

This survey paper present approach for profiling and lassifying computer user from command-line interface. User profile is represented by a distribution of relevant subsequence and modification to proposed recongnition of user. Finally this paper shows in hundreds of users ABCD can easily modified with aspect implemented evolving systems [29] for future work.

8. References

[1] D. Godoy and A. Amandi, "User Profiling in Personal Information Agents: A Survey," Knowledge Eng. Rev., vol. 20, no. 4, pp. 329 361, 2005.
[2] J.A. Iglesias, A. Ledezma, and A. Sanchis, "Creating User Profiles from 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 Using Truncated Command Lines," Proc. Int'l Conf. Dependable Systems and Networks (DSN), pp. 219 228, 2002.
[5] A. Alaniz Macedo, K.N. Truong, J.A. Camacho Guerrero, and M. Graca Pimentel, "Automatically Sharing Web Experiences through a Hyperdocument Recommender System," Proc. ACM Conf. Hypertext and Hypermedia (HYPERTEXT '03), pp. 48 56, 2003.
[6] Spiliopoulou, M., Faulstich, L.C.: Wum: "A web utilization miner," In: Proceedings of EDBT Workshop WebDB 1998, pp. 109–115. Springer, Heidelberg (1998).
[7] Wexelblat, A.: "An environment for aiding information-browsing tasks." In: Proc. Of AAAI Spring Symposium on Acquisition, Learning and Demonstration: Automating Tasks for Users. AAAI Press, Menlo Park (1996).
[8] Macedo, A.A., Truong, K.N., Camacho-Guerrero, J.A., da GraC¸a Pimentel, M.: "Automatically sharing web experiences through a hyperdocument recommender system," In: HYPERTEXT 2003, pp. 48–56. ACM, New York (2003).
[9] Godoy, D., Amandi, A.: "User profiling for web page filtering," IEEE Internet Computing 9(4), 56–64 (2005).
[10] Pepyne, D.L., Hu, J., Gong, W.: "User profiling for computer security," In: Proceedings of the American Control Conference, pp. 982–987 (2004).
[11] Fredkin, E.: "Trie memory," Comm. ACM 3(9), 490–499 (1960).
[12] Kaminka, G.A., Fidanboylu, M., Chang, A., Veloso, M.M.: "Learning the sequential coordinated behavior of teams from observations," In: Kaminka, G.A., Lima, P.U., Rojas, R. (eds.) RoboCup 2002. LNCS, vol. 2752, pp. 111–125. Springer, Heidelberg (2003).
[13] Iglesias, J.A., Ledezma, A., Sanchis, A., Kaminka, G.A.: "Classifying efficiently the behavior of a soccer team". In: Burgard, W., et al. (eds.) Intelligent Autonomous Systems 10. IAS-10, pp. 316–323 (2008).
[14] R. Xiao, J. Wang, and F. Zhang, "An Approach to Incremental SVM Learning Algorithm," Proc. IEEE Int'l Conf. Tools with Artificial Intelligence, pp. 268 278, 2000.
[15] G. Widmer and M. Kubat, "Learning in the Presence of Concept Drift and Hidden Contexts," Machine Learning, vol. 23, pp. 69 101, 1996.
[16] D. Kalles and T. Morris, "Efficient Incremental Induction of Decision Trees," Machine Learning, vol. 24, no. 3, pp. 231 242, 1996.
[17] F.J. Ferrer Troyano, J.S. Aguilar Ruiz, and J.C.R. Santos, "Data Streams Classification by Incremental Rule Learning with Parameterized Generalization," Proc. ACM Symp. Applied Computing (SAC), pp. 657 661, 2006.
[18] J.C. Schlimmer and D.H. Fisher, "A Case Study of Incremental Concept Induction," Proc. Fifth Nat'l Conf. Artificial Intelligence (AAAI), pp. 496 501, 1986.
[19] P.E. Utgoff, "Id5: An Incremental Id3," Proc. Int'l Conf. Machine Learning, pp. 107 120, 1988.
[20] P.E. Utgoff, "Incremental Induction of Decision Trees," Machine Learning, vol. 4, no. 2, pp. 161 186, 1989.
[21] G.A. Carpenter, S. Grossberg, and D.B. Rosen, "Art2 a: An Adaptive Resonance Algorithm for Rapid Category Learning and Recognition," Neural Networks, vol. 4, pp. 493 504, 1991. [22]. Fredkin, E.: Trie memory. Comm. ACM 3(9), 490–499 (1960).
[22] G.A. Carpenter, S. Grossberg, N. Markuzon, J.H. Reynolds, and D.B. Rosen, "Fuzzy Artmap: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps," IEEE Trans. Neural Networks, vol. 3, no. 5, pp. 698 713, Sept. 1992.
[23] 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.
[24] T. Seipone and J.A. Bullinaria, "Evolving Improved Incremental Learning Schemes for Neural Network Systems,"

Proc. IEEE Congress on Evolutionary Computation, pp. 2002 2009, 2005.
[25] T. Kohonen, J. Kangas, J. Laaksonen, and K. Torkkola, "Lvq pak: A Program Package for the Correct Application of Learning Vector Quantization Algorithms," Proc. IEEE Int'l Conf. Neural Networks, pp. 725 730, 1992.
[26] F. Poirier and A. Ferrieux, "Dvq: Dynamic Vector Quantization An Incremental Lvq", Proc. Int'l Conf. Artificial Neural Networks, pp. 1333 1336, 1991.
[27] R. K. Agrawal and R. Bala, "Incremental Bayesian Classification for Multivariate Normal Distribution Data,"Pattern Recognition Letters, vol. 29, no. 13, pp. 1873 1876, http://dx.doi.org/10.1016/j.patrec.2008.06.010, 2008.
[28] T. Lane and C.E. Brodley, "Temporal Sequence Learning and Data Reduction for Anomaly Detection," Proc. ACM Conf. Computer and Comm. Security (CCS), pp. 150 158, 1998.
[29] P. Angelov and X. Zhou, "Evolving Fuzzy Rule Based Classifiers from Data Streams," IEEE Trans. Fuzzy Systems: Special Issue on Evolving Fuzzy Systems, vol. 16, no. 6, pp. 1462 1475, Dec. 2008.
[30] M. Panda and M.R. Patra, "A Comparative Study of Data Mining Algorithms for Network Intrusion Detection," Proc. Int'l Conf. Emerging Trends in Eng. and Technology, pp. 504 507, 2008.
[31] A. Cufoglu, M. Lohi, and K. Madani, "A Comparative Study of Selected Classifiers with Classification Accuracy in User Profil ing," Proc. WRI World Congress on Computer Science and Information Eng. (CSIE), pp. 708 712, 2009.
[32]. P. Riley and M.M. Veloso, "On Behavior Classification in Adversarial Environments," Proc. Int'l Symp. Distributed Autono mous Robotic Systems (DARS), pp. 371 380, 2000.
[33] R. Polikar, L. Upda, S.S. Upda, and V. Honavar, "Learn++: An Incremental Learning Algorithm for Supervised Neural Net works," 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.