ISSN No. 0976-5697

Volume 4, No. 4, March-April 2013

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Visual role mining to avoid text complexity using fuzzy logic decision system

M.K.Shanmathi
Computer Science Engineering
SRM Easwari Engineering College Chennai, India
shanmathimk@yahoo.co.in

G. Narmada*
Computer Science Engineering
SRM Easwari Engineering College Chennai, India
narmada92@yahoo.co.in

Ms.N.D.Thamarai selvi
Assistant Professor M.E SRM Easwari Engineering College
Chennai, India
thamarai.nd@gmail.com

Abstract: Access control is currently one of the most chief topics in information security. The exigent areas of research relating to access control are to recognize approaches and models to efficiently administer user privileges. With the ever-increasing number of users and IT systems, organizations have to administer large number of users and permissions in an efficient manner. This paper proposes a new approach for data visualisation which acts as an aid to Role Engineering .The key idea is to represent the roles within an organisation in a graphical manner so as to have better elicitation and understanding of the data. This is the primary step to Role Based Access Control (RBAC). The data is viewed in the form of a Bicluster using a tool named BicOverlapper. Further for best representation of roles we propose a fuzzy decision tree induction approach to role mining. It facilitates classification of the roles and reduces the problem complexity. The value of this visual analysis in business environments is demonstrated through examination on real life as well as constructed datasets

Keywords- Role engineering, Role Based Access Control (RBAC), Bicluster visualisation.

I. INTRODUCTION

Access control when in the case of information security is the process of restricting the necessary resources to the needed users. Most of the information handling organisations are having recourse to access control models that enables them to control the ability of a process and have securable access to objects for various system administration tasks. The best norm for this is by making use of ROLE BASED ACCESS CONTROL(RBAC). The success of RBAC models is mainly because they are simple: a role is assigned to every user within the organisation and a role identifies a set of permissions for itself.

Recently, there has been a growing interest as to how RBAC can be deployed in any industry by bagging a bearable cost. Textual approaches to role engineering envision a high cost for deployment and also they overlook the ambiguities that might arise while performing the task of role mining.

A visual approach to RBAC gives a quick analysis and better elicitation of the meaningful roles within an organisation. This is the new way to devise rbac approach and the rationale behind this is that a graphical representation can increase cognition, showing at a glance what it would take a lot of data to dilate. Our paper proposes a new idea to role based access control by resorting to the bicoverlapper tool that represents the data in the matrix format similar to access control matrices (ACM), possessing rows as users and columns as roles, which is then provided to a classification algorithm for choosing out the respective permissions that head under each role.Applications to visual approaches are that it allows role engineers to draw out conclusions and gain insight of the existing data sets.

II. RELATED WORK

IT security administration processes, access control and granting access rights to users are of prime importance in today's business scenario. Organizations have been unwilling to move to RBAC because of the conceived high cost for role engineering. First paper to introduce visual role mining was [1].

There are very few algorithms that are proposed for role engineering purposes.Kuhlmann et al. [2] first introduced the words "Role Mining" trying to elicit roles from access data. Visual representation of mining data have been considered mainly for binary data[3]. Leung and Carmichael [4] developed a visualization tool for frequent item sets. However, frequent item sets are not the only proper patterns for role mining.

Our approach to a certain degree is motivated by [5]. It introduced the BicOverlapper tool. It was typically used to represent gene data- rows and columns of an analyzed matrix, and was not suitable for role mining. Our approach greatly differs from [6]: primarily we represent the data in a form of a Bicluster. Secondly we use classification algorithm to classify the data for further investigation on them. According to that we use a fuzzy decision tree algorithm. Fuzzy logic can process uncertainty and imprecision and produce more interpretable models. Finally the classified data is again represented visually and demonstrated on both simple and very large datasets.

III. BACKGROUND

Before moving into the required formalism used in role engineering, we first review some concepts of the ANSI/INCITS RBAC standard required for the analysis. [1]There are a few entities like:

a. PERMS, USERS, and ROLES are the sets of all access permissions, users, and roles, respectively;
b. UA □ USERS × ROLES, set of all role-user relationships;
c. PA □ PERMS × ROLES, set of all role-permission relationships.

Following functions are provided:

a. ass_users: ROLES → 2USERS to identify users assigned to a role. We consider it as derived from UA, that is ass_users(r) = {u ∈ USERS | □u, *r*□∈ UA}.
b. ass_perms: ROLES → 2PERMS to identify permissions assigned to a role. We consider it as derived from PA, that is ass_perms(r) = {p ∈ PERMS | p,r∈ PA}.

It is now possible to formally define the main objective of role engineering: given USERS and their assigned ROLES, we are interested in determining the best setting PERMS, that is all combinations of permissions that can be assigned to the user.

IV. PROPOSED SYSTEM:

From the observations made in the previous sections, we now describe a viable, fast way to best represent user permission information in the visual form. To simplify the technique of visual representation, we use BicOverlapper tool.BicOverlapper[5] is a framework to support visual analysis by means of biclustering. A visual approach can complement statistical analysis and reduce the time spent by specialists interpreting the results of biclustering algorithms.

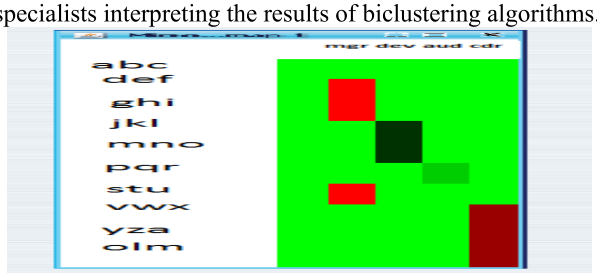

Figure 1: visual representation of roles

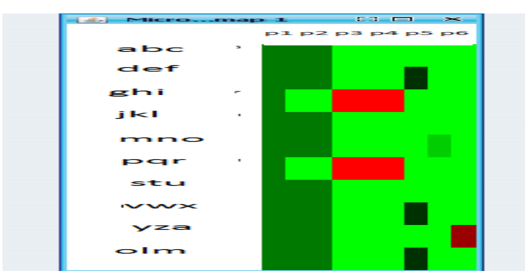

Figure 2: visual representation of user permission assignments.

Biclustering algorithms have become a spread technique to analyze microarray experiments. Since it is an unsupervised technique, biclustering is specially useful at discovering new knowledge from data. In order to improve the visualization of biclusters, it uses Overlapper, a visualization technique to simultaneously represent all biclusters from one or more biclustering algorithms, based on a force-directed layout. This visualization technique is integrated in BicOverlapper[5], along with several other visualization techniques and biclustering algorithms.

Initially user- role assignments are the minimal data set required. A natural representation for this information is the binary matrix, where rows and columns correspond to users and roles, and each cell is "on" when a certain user has that role. This is achieved by using BicOverlapper tool mentioned above. The output of this page looks like figure 1.

V. ROLE CLASSIFICATION

In this paper we propose a fuzzy decision tree induction approach [7] for data set classification. Decision trees are popular models in machine learning due to the fact that they produce graphical models and text rules that end users can easily understand.It is comparatively faster and requires minimum resources. Fuzzy systems, on the other hand, provide techniques to handle imprecision anduncertainty in data, based on the fuzzy logic and fuzzy sets theory. The combination of fuzzy systems and decision trees has produced fuzzy decision tree models, which benefit from both techniques to provide simple, accurate, and highly interpretable models.

A. Fuzzy classification system:

The classification task can be roughly described as in [7] as:

Given a set of objects E = {e_1, e_2, ..., e_n}, also named examples, which are described by m features, assign a class c_i from a set of classes C = {C_1,C_2,...,C_j} to an object e_p, e_p =(a_{p1},a_{p2},...,a_{pm}).

A typical fuzzy classification rule can be expressed by: R_k :IF X_1 is A_{11} AND…..AND $X_m$isA_{ml} THEN Class=C_i□where $R_k$is the rule identifier, X_1,…..,X_m are the features of the example considered in the problem, A_{11},...,A_{ml} are the linguistic values used to represent the feature values, and C_i∈ C is the class.

The first block of fig 3.illustrates a dataset with three attributes (At1, At2, and At3) and a class attribute. The fuzzyfied version of this dataset is presented in the second block. This fuzzyfied set of examples is used to induce the final DT, illustrated in the last block.

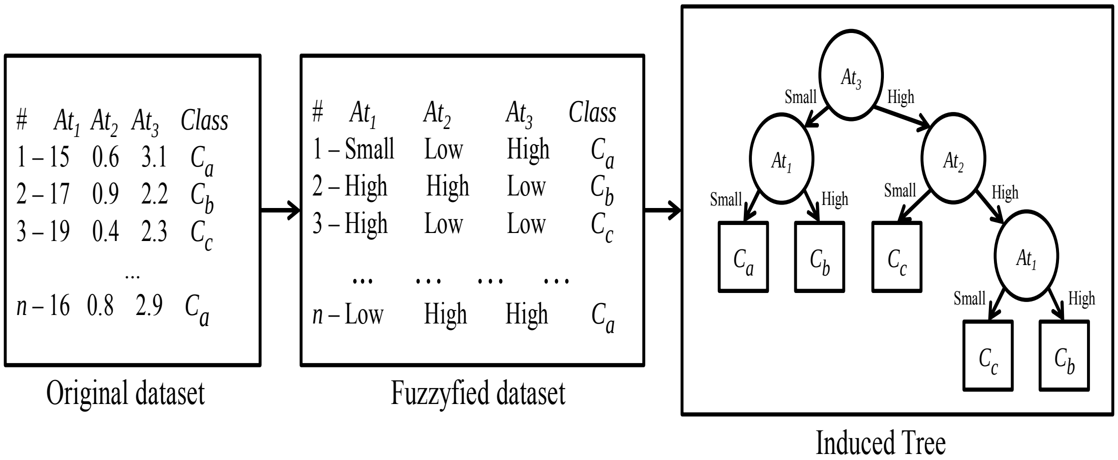

Figure 3: Fuzzy classification system[7]

Algorithm: The FuzzyDT algorithm from [7]:

a. Define the fuzzy data base, i.e., the fuzzy granulation for the domains of the continuous features;
b. Replace the continuous attributes of the training set using the linguistic labels of the fuzzy sets with highest compatibility with the input values;
c. Calculate the entropy and information gain of each feature to split the training set and define the test nodes of the tree until all features are used for all training examples are classified;
d. Apply a post-pruning process, similarly to C4.5, using 25% confidence limits as default.

Table 1: Attributes in a business environment

Attribute	Possible Values
Designation	Manager, developer, auditor, coordinator
Experience	Continuous
Qualification level	Continuous
Headquarters	True,false

Table 2: Training data

Designation	Experience	Qualification level	Headquarters	Permission
Manager	0.20	0.80	True	P1,P2,P3,P4
Manager	0.15	0.95	True	P1,P2,P3,P4
Manager	0.18	0.70	False	P1,P2,P4
Developer	0.12	0.65	True	P1,P2,P3
Developer	0.10	0.70	True	P1,P2,P5,P6
Auditor	0.8	0.73	False	P1,P2,P6
Auditor	0.10	0.64	True	P1,P7
Auditor	0.11	0.55	False	P1,P2,P6
Auditor	0.9	0.72	True	P1,P2,P5
Coordinator	0.5	0.88	False	P1,P7
Coordinator	0.6	0.77	False	P1,P7
Coordinator	0.2	0.60	True	P1

In general from previous studies on information gain and entropy, if we are given a probability distribution P = (p_1, p_2, ..,p_n) then the *Information conveyed by this distribution*, also called *the Entropy of P*, is:

$$I(P) = -(p_1*\log(p_1) + p_2*\log(p_2) + .. + p_n*\log(p_n)) \quad (1)$$

If a set T of records is partitioned into disjoint exhaustive classes C1, C2, .., Ck on the basis of the value of the categorical attribute, then the information needed to identify the class of an element of T is **Info(T)** = I(P), where P is the probability distribution of the partition (C1, C2, .., Ck):

$$P = (|C_1|/|T|, |C_2|/|T|, ..., |C_k|/|T|) \quad (2)$$

The weighted average of Info(Ti):

$$\text{Info}(X,T) = \text{Sum for i from 1 to n of } |Ti|/|T| * \text{Info}(Ti) \quad (3)$$

Quantity Gain(X,T) defined as

$$\text{Gain}(X,T) = \text{Info}(T) - \text{Info}(X,T) \quad (4)$$

The information gained from this approach is again passed on to the BicOverlapper tool. This produces the user permission matrix that is highly beneficial when used to gain an overview of the underlying data set. The output produced after this stage looks similar to that in figure 2.

VI. EXPERIMENTAL RESULTS

This section presents practical applications of our methodology. Fig. 4 shows the application of our methodology on the data set of Table 3. The situation hugely changes as the data set becomes larger. Even though large matrices cannot entirely be represented on a computer screen, their construction is beneficial. For instance, visualizing a small "sliding window" or implementing zooming still represent a valuable way of browsing data.

VII. CONCLUDING REMARKS

The proposed representation provides user-permission assignments in an intuitive graphical form simplify the role engineering process. This helps both in IT and business environment. In this paper we proposed how role engineering can be of great benefit by visual representation. As for further work these can be used in analysis of DNA fingerprints and related fields.

VIII. ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers who with their comments, helped increase the quality of the paper.

Table 3: Contents of synthetic data

NAME	E1 A	E1 B	E2 A	E2 B	E3 A	E3 B	E4 A	E4 B	E5 A	E5 B	E6 A	E6 B	E7 A	E7 B	E8 A	E8 B	E9 A	E9 B	E10 A	E10 B
ABC	0.5	0.9	0.3	0.5	0.5	0.3	0.9	0.3	0.5	0.6	0.8	0.8	0.6	0.4	0.3	0.9	0.8	0.6	0.5	0.4
DEF	0.3	0.7	0.5	0.3	0.4	0.5	0.7	0.7	0.4	0.9	0.3	0.2	0.2	0.6	0.7	0.4	0.2	0.3	0.3	0.7
HIJ	0.3	0.7	0.5	0.3	0.4	0.5	0.7	0.7	0.4	0.9	0.3	0.2	0.2	0.6	0.7	0.4	0.2	0.3	0.3	0.7
KLL	0.3	0.7	0.5	0.3	0.4	0.5	0.7	0.7	0.4	0.9	0.3	0.2	0.2	0.6	0.7	0.4	0.2	0.3	0.3	0.7
HUI	0.3	0.7	0.5	0.3	0.4	0.5	0.7	0.7	0.4	0.9	0.3	0.2	0.2	0.6	0.7	0.4	0.2	0.3	0.3	0.7
RPO	0.5	0.9	0.3	0.5	0.5	0.3	0.9	0.3	0.5	0.6	0.8	0.8	0.6	0.4	0.3	0.9	0.8	0.6	0.5	0.4
WER	0.6	0.5	0.4	0.6	0.6	0.2	0.9	0.7	0.3	0.3	0.2	0.9	0.8	0.5	0.4	0.3	0.7	0.1	0.6	0.5
FGD	0.4	0.3	0.5	0.4	0.4	0.7	0.6	0.7	0.2	0.9	0.8	0.6	0.6	0.1	0.1	0.5	0.3	0.2	0.5	0.7
REW	0.9	0.6	0.5	0.4	0.7	0.7	0.3	0.2	0.4	0.2	0.7	0.7	0.2	0.8	0.9	0.7	0.1	0.1	0.4	0.7
QEE	0.4	0.3	0.5	0.4	0.4	0.7	0.6	0.7	0.2	0.9	0.8	0.6	0.6	0.1	0.1	0.5	0.3	0.2	0.5	0.7
RTY	0.9	0.6	0.5	0.4	0.7	0.7	0.3	0.2	0.4	0.2	0.7	0.7	0.2	0.8	0.9	0.7	0.1	0.1	0.4	0.7
PLI	0.3	0.7	0.5	0.3	0.4	0.5	0.7	0.7	0.4	0.9	0.3	0.2	0.2	0.6	0.7	0.4	0.2	0.3	0.3	0.7
QSV	0.3	0.7	0.5	0.3	0.4	0.5	0.7	0.7	0.4	0.9	0.3	0.2	0.2	0.6	0.7	0.4	0.2	0.3	0.3	0.7
BNG	0.5	0.9	0.3	0.5	0.5	0.3	0.9	0.3	0.5	0.6	0.8	0.8	0.6	0.4	0.3	0.9	0.8	0.6	0.5	0.4
FAS	0.3	0.7	0.5	0.3	0.4	0.5	0.7	0.7	0.4	0.9	0.3	0.2	0.2	0.6	0.7	0.4	0.2	0.3	0.3	0.7
KTL	0.5	0.9	0.3	0.5	0.5	0.3	0.9	0.3	0.5	0.6	0.8	0.8	0.6	0.4	0.3	0.9	0.8	0.6	0.5	0.4
TRR	0.6	0.5	0.4	0.6	0.6	0.2	0.9	0.7	0.3	0.3	0.2	0.9	0.8	0.5	0.4	0.3	0.7	0.1	0.6	0.5
NCD	0.4	0.3	0.5	0.4	0.4	0.7	0.6	0.7	0.2	0.9	0.8	0.6	0.6	0.1	0.1	0.5	0.3	0.2	0.5	0.7
GSA	0.4	0.3	0.5	0.4	0.4	0.7	0.6	0.7	0.2	0.9	0.8	0.6	0.6	0.1	0.1	0.5	0.3	0.2	0.5	0.7
LWR	0.9	0.6	0.5	0.4	0.7	0.7	0.3	0.2	0.4	0.2	0.7	0.7	0.2	0.8	0.9	0.7	0.1	0.1	0.4	0.7

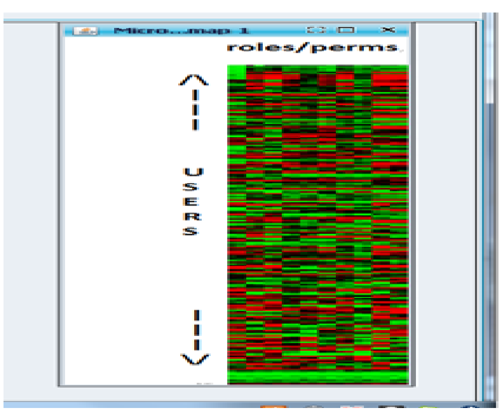

Figure 4: visual representation of the static data set

IX. REFERENCES

[1]. A. Colantonio, R. Di Pietro, A. Ocello, and N.V. Verde,"Visual Role Mining A Picture Is Worth a Thousand Roles", " Computers Security, vol.24, June 2012

[2]. M. Kuhlmann, D. Shohat, and G. Schimpf, "Role Mining—Revealing Business Roles for Security Administration Using Data Mining Technology," Proc. Eighth ACM Symp. Access Control Models and Technologies (SACMAT '03), pp. 179-186, 2003.

[3]. C. Chen, "Top 10 Unsolved Information Visualization Problems," IEEE Trans. Computer Graphics and Applications, vol. 25, no. 4, pp. 12-16, July 2005.

[4]. C.K.-S. Leung and C.L. Carmichael, "FpViz: A Visualizer for Frequent Pattern Mining," Proc. ACM SIGKDD Workshop Visual Analytics and Knowledge Discovery (VAKD '09), pp. 30-39, 2009.

[5]. R. Santamaria, R. Theron, and L. Quintales, "BicOverlapper: A Tool for Bicluster Visualization," Bioinformatics, vol. 24, no. 9, pp. 1212-1213, 2008.

[6]. R. Jin, Y. Xiang, D. Fuhry, and F.F. Dragan, "Overlapping Matrix Pattern Visualization: A Hypergraph Approach," Proc. IEEE Int'l Conf. Data Mining (ICDM '08), pp. 313-322, 2008.

[7]. FuzzyDT- A Fuzzy Decision Tree Algorithm Based on C4.5Marcos E. Cintra, Maria C. Monard, and Heloisa A. Camargo

