Mining Association Rules with Static and Dynamic Behavior of Learner in the Internet

Dr. R. Siva Rama Prasad
Research Director, Dept. of CSE,
Acharya Nagarjuna University, Guntur – 522 510, A.P., India

D. Bujji Babu
Associate Professor, Dept. of CSE,
Prakasam Engineering College, Kandukur.- 523105, A.P., India

ABSTRACT
Now a days several Number of users are depending on internet to do their routine tasks, because the world wide web providing several services required to the people. Here the main problem is the internet environment providing huge number of services so we need to find the behavior of the user in various dimensions. First we performed a study on static model of the learner. Second we performed a study on dynamic model of the learner. In general the Association rules are extracted from the market basket analysis problem with using the apriori algorithm. Here we concentrated mainly on the unification process and apriori algorithm was improved and we experimented the internet based learning and we present the experimental results.

Keywords: Association rule, Static learning, dynamic learning, unification process.

1. INTRODUCTION
The learning process is done generally via the formal class room environment . In this environment some advantages and disadvantages are there. As a part of advantages the leaner can directly interact with the teacher , hence the learning rate is very fast, the disadvantage is highly tough to find the subject export as well as skilful teachers in a large number. The virtual class room environment having so many advantages . virtual class room (internet environment) decreases the cost of establishment of a physical class room environment (formal class room environment) , at the same time the risk factors are also reduced a lot, and single skilful teacher can broadcast the lesson.

As a trend change several learners are utilizing the world wide web as resource of learning . i.e, Via the virtual class room environment . Hence leaning rate is increased drastically several universities, are conducting virtual class room environment and conducting more arrived 60-70% . In the field of education industry. In the Internet based learning Environment, Discoveries the static and Dynamic behavior of the of the learner is essential. The Research is on going on this issue by the psychologists and pedogogists for their application areas in learning process The Internet based learing is a Novel Approach in modern Education system. This concept is a wide are a of research in the fields of psychology, Information Technology and information services and so on . All these Areas are working to find the relationship between the static behavior and the dynamic behavior of a learner, We build the learner model based on various raw data complex, dynamic , distributed properties, then the selection of suitable method for discovery the patte relationship based on the behaviors .

1.1 Users learner model:- IEEE 1484.2 PAPI (Public and private & information) is the eight aspects [1] of the user learning process which is composed up with two models, the first one is personality model of the learner which has eight aspects as show below

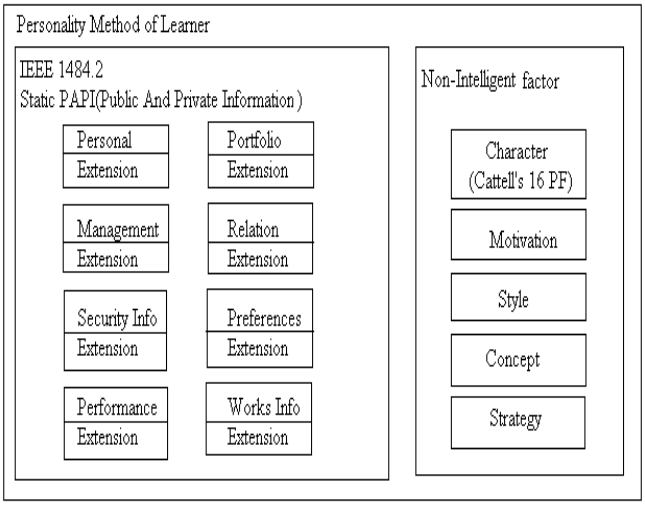

Fig: 1 Static Method

Even though the public and private sectors serving to the education industry well, but these sectors are not allowing the learners for personalization. The personalized education system is the demand of the learners.So,we are using the non-intelligent factors towards the personalized learning process, in the non-intelligent factors are personality, motivation, style, concept and strategy.

In the internet based learning environment the users prefers personalization process in the learning process. It is very clear that personalization means based on the learning interest he can choose the concept .Here the behavior of the learner are very important in learning process, In this concept we use the behavioral method which has six concepts. The following is the Dynamic method in learning process.

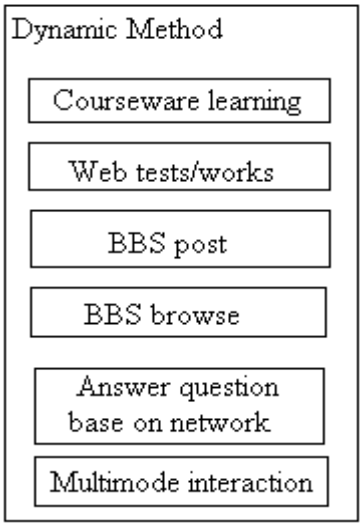

Fig:2 Dynamic Method

The User Learner Method(ULM) containing both Static Method and Dynamic Method , it will be look like as follows.

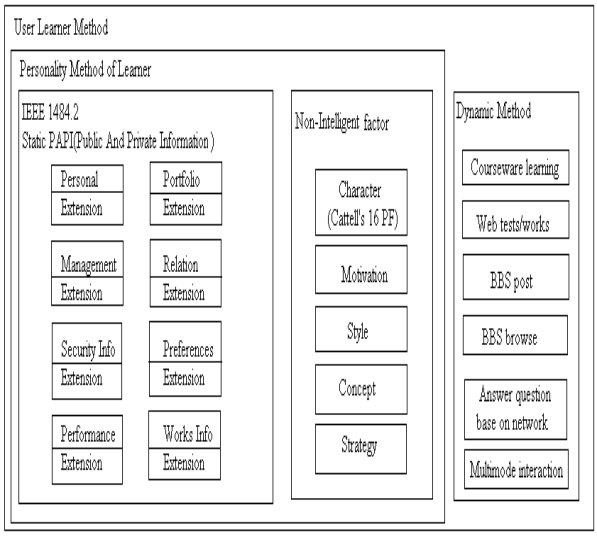

Fig: 3 The User Learner Method

The relationship between the static method and the dynamic method can be defined as

ULM=<SM,DM>

1.1.1 Static Method: During the learning process some elements are stable. These stable elements are represented in static method (SM) and it can be defined as

SM=<UL_P,UL_M,UL_S,UL_C,UL_T>

Non-Intelligent factor character was the sum of the static properties of the users[3].The learning process is affected by the style of dynamic learning[4]. According to Cattell's 16 Personality Factors the users learning was defined as following tuple .

UL_P::=<UID,A,B,C,E,F,G,H,I,L,M,N,O,Q1,Q2,Q3,Q4>

UID refers the unique identity number of the learner and the remaining are the Cattell's 16 Personality Factors.

Factor		Description	
A	Warmth	Reserved	Outgoing
B	Reasoning	Less Intelligent	More Intelligent
C	Emotional Stability	Affected by feeling	Emotionally stable
E	Dominance	Humble	Assertive
F	Liveliness	Sober	Happy-go-lucky
G	Rule Consciousness	Expedient	Conscientious
H	Social Boldness	Shy	Venturesome
I	Sensitivity	Tough-minded	Tender-minded
L	Vigilance	Trusting	Suspicious
M	Abstractedness	Practical	Imaginative
N	Privateness	Straightforward	Shrewd
O	Apprehension	Self-Assured	Apprehensive
Q1	Openness to Change	Conservative	Experimenting
Q2	Self-Reliance	Group-dependent	Self-sufficient
Q3	Perfectionism	Self-conflict	Self-control
Q4	Tension	Relaxed	Tense

Similarly, the other elements of PM can be defined as following formal expression,

Study Motivation Information of Learner:

L_M :: =< UID,M_C,M_I,M_R,M_D,M_S,M_E >

Where M_C is information of challenge, M_I is interest, M_R is curiosity, M_D is independence, M_S is Success and M_E is extrinsic motivation.

Study Style Information:

L_S :: =< UID,S1,S2,S3,S4,S5,S6,S7,S8 >

Study Concept Information:

L_C :: < UID,C_M,C_E,C_A >

Study strategy Information:

L_T :: =< UID,T_M,T_C,T_E,T_R,T_S,T_O >

1.1.2. Dynamic Method: In internet learning environment we collected the historical information from the browsers to identify the Dynamic learning. As per the user

learning dynamic model we describe 6 issues the following is tuple.

UDM = { D_C, D_T,D_B1,D_B2,D_A,D_I }

Where D_C – Dynamic Behavior on Courseware learning
D_T – Dynamic Behavior on Test/homework
D_B1- Dynamic BBS posting
D_B2- Dynamic BBS browsing
D_A- Dynamic Behavior of answer question
D_I- Dynamic Behavior of interaction.

The learner details can be quantifiable using statistics via the Dynamic Method. Both values of static method and dynamic method are mapped into a normalized manner i.e, uniform integer values and these values are ranked high, middle, low and 1,2,3 respectively.

2. ALGORITHM (Association Rule Mining)

Fist we define association rule using mathematical notation[2].

2.1 Definition: Association Rule

Let D be a set of transactions and transaction and the set of items $I=\{i_1,i_2,i_3\ldots\ldots\ldots,i_m\}$T is a set of items such that $T \subseteq L$ .An unique identifier,TID, is associated with each traction.T contains X,a set of some items in L, if .

Rule form:"Body $\Rightarrow$ Head[Support,confidence]".

Association rule

$X \Rightarrow Y \;\; X \subset T \;, Y \subset T, \; X \cap Y = \phi$

Support: X U Y (the percentage of records that contain both X and Y in the database, called

support of the rule)

confidence: X ∩ Y(the percentage of records containing X that also contain Y, called the confidence of the rule)

To reduce the cost of apriori we used the apriori decision domain algorithm for our problem.

Here we use the Dynamic behavior of the learner as input parameters,and we generate some association rules between the static method and the dynamic method as DM $\Rightarrow SM$. here we discover the attribute relationship between static method and the dynamic method.

2.2 Definition: Domain

The attributes set I in ULM can divided into two subsets: $I = I_S \cup I_D$, and $I_S \cap I_D = \phi$. We named I_S, I_D as a domain, and the domain I_S, I_D expressed as $I_S=\{s_1,s_2,s_3,\ldots\ldots..s_n\}$, $I_D=\{d_1,d_2,d_3,\ldots..d_m\}$.

2.3 Algorithm Analysis:

If $g_1,\ldots\ldots,g_i \Rightarrow h_1,\ldots\ldots,h_j$ existed, $g_1,\ldots\ldots\ldots,g_i \Rightarrow h_1$ . . . $g_1,\ldots\ldots,g_i \Rightarrow h_2$, ... $g_1,\ldots\ldots\ldots,g_i \Rightarrow h_j$ must exist. As per the apriori principle "any subset of a frequent item set must be frequent". So the problem can be translated as $I_G \Rightarrow I_H$ into the rule set as $\{ \wedge g_i \Rightarrow hj\}$.

2.4 Definition: Decision Domain (DD)

We need to find out the association rules as $h_i,h_j\ldots\ldots h_m \Rightarrow S$ where { $h_i,h_j\ldots\ldots h_m,S$ }is attribute set , $h_i,h_j\ldots\ldots h_m$, belong to domain $I_{D,S}$ belong to I_S . Here S was known and name as a Decision Domain (DD).

2.5 Monotonicity Property:

A Sequence { a_n } is said to be monotonic if $\{a_n\}$ is monotonically increasing or monotonically decreasing.

A Sequence { a_n } is said to be monotonic if $\{a_n\}$ is monotonically increasing if $a_{n+1}>=an$, $\forall n \;\; \varepsilon \; N$. That is

$a_1, <= a_2 <= a_3\ldots\ldots.<= a_n <= a_{n+1} <=\ldots\ldots\ldots\ldots$

A Sequence { a_n } is said to be monotonic if $\{a_n\}$ is monotonically decreasing if $a_{n+1}<=an$, $\forall n \;\; \varepsilon \; N$. That is

$a_1, >= a_2 >= a_3\ldots\ldots\ldots>= a_n >= a_{n+1} >=\ldots\ldots\ldots$

While using the the apriori with Decision domain algorithm ,if k-items $(h_1,h_2,h_3\ldots\ldots\ldots h_{k-1},s)$ is not a frequent item set,according to Monotonicity property.the $h_1,h_2,h_3\ldots\ldots\ldots h_{k-1}$ must been invalid frequent items to generate the rules as $h_1,h_2,h_3\ldots\ldots\ldots h_{k-1,}\ldots\ldots\ldots h_n \Rightarrow s$. Here invalid means even though $(h_1,h_2,h_3\ldots\ldots\ldots h_{k-1},s)$ can generate the frequent items such as $(h_1,h_2,h_3\ldots\ldots\ldots h_{k-1,}\ldots\ldots\ldots h_{n)}$, it can not generate frequent item set as $(h_1,h_2,h_3\ldots\ldots\ldots\ldots\ldots h_{k-1,}\ldots\ldots\ldots h_n,s)$.

3. ALGORITHM: APRIORI WITH DECISION DOMAIN

Step 1: divide L_k(k-frequent items) between L_{k1}which includes decision domain and L_{k2} which excludes decision domain,both L_{k1} and L_{k2} are k-frequent items.

Step 2: generate the k-candidate set$C_{(k\ +\ 1)1}$which includes decision domain from L_{k1},L_{k2}

Step 3: counting the items in $C_{(k+1)1}$,generate (k+1) frequent items L(k+1)1 which include

decision domain.

Step 4: supposing the item which included in $C_{(k+1)1}$ and excluded in $L_{(k+1)1}$ is $h_i,h_j,\ldots\ldots h_k,s$;

Step 5: delete all the items which include $h_i,h_j,\ldots\ldots h_k$ from L_{k2}

Step 6: generate k+1 candidate $C_{(k+1)2}$ which exclude decision domain from L_{k2}

Step 7: counting the items in $C_{(k+1)2}$,generate k+1 frequent items L(k+1)2 which exclude decision domain;

Step 8: repeat step 1 to step 7 till the largest set of frequent items is generated.

3.1 Comparative Analysis

(Time Complexity Evaluation)

Here we evaluate the dime complexity evaluation to the traditional apriori algorithm and also to the apriori with decision domain algorithm and we present the results.

The general apriori algorithm complexity is illustrates as

$$O(n*(C_{m}^{1} + C_{m-M-N}^{2} + \ldots\ldots\ldots\ldots\ldots + C_{m-M}^{m-M--1})) = O(n * 2^{m-M})$$

The time complexity of Apriori algorithm with decision domain is illustrated as

$$O(n*(C_{m}^{1} + C_{m-M-N}^{2} + \ldots\ldots\ldots\ldots\ldots + C_{m-M-N}^{m-M-N-1})) = O(n * 2^{m-M-N})$$

Then $O(n * 2^{m-M}) >> O(n * 2^{m-M-N})$ Hence, the algorithm with decision domain can reduce the number of k-candidates C_k efficiently.

4. EXPERIMENT

The experiment is conducted in a famous engineering college on students static and dynamic behavior with observing what they are browsing in the internet hour and we collected 420 students browsing history via band width management software. We collected 30 days browsing history of the students. Totally 12600 transactions were observed and we executed both apriori as well as apriori decision domain. Both algorithms were executed on the machines with Pentium IV 3.0 G.HZ processor and 512 MB RAM and we present the performance graph. Apriori has taken 2613m.s and Apriori DD has taken 1131m.s to process 12600 transactions.

5. RESULTS

Here we present the experimental results in the form of association rules and also we present a comparison graph of apriori and apriori with DD algorithms.We present a few rules.

R1:Any student who is rule conciousness he is using (browsing) courseware learning in the interner this associan is having 90 percent support.

R2:Any student who is not rule conciousness he is using (browsing) courseware learning in the interner this associan is having 60 percent support.

R3:Any student is social bold ness he is using (browsing) varity of websites for learning in the interner this associan is having 69 percent support.

R4:Any student is not social bold ness he is using (browsing) varity of websites for learning in the interner this associan is having 40 percent support

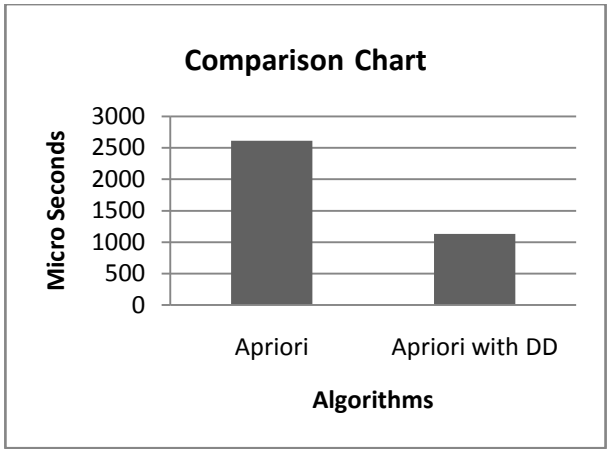

Fig:4 Comparison chart

6. CONCLUSION

In this work we collected the internet usage details of several students and we analysed the data based on their behaviour .Finally wecompared the Apriori algorithm and the apriori with DD and we presented the association rules aswell as the comparision graph .We are going to work on various types of data analysis in multiple dimensions to correlate the static and dynamic method in an effectively.

7. ACKNOWLEDGEMENTS

We would like to express sincere thanks to Mr.Dr. Kancharla Ramaiah Secretary and Correspondent of Prakasam Engineering College,Kandukur. for his continuous encouragement with providing the research environment.

8. REFFERENCES

[1] IEEE Learning Technology Standards Committee (LTSC), IEEE 1484.2 "PAPI Learner Model"

[2] E.V. Wilson, Students characteristics and computer-mediated communication, Computers & Education, 2003 (34), 67-76

[3] E.B.Kim, The role of personality in Web-based distance education courses Communications of the ACM,Volume 47, Issue 3 March 2004

[4] Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques, Morgan kaufman publishers. 2001.

[5] The Comparative of Boolean Algebra Compress and Apriori Rule Techniques for New Theoretic Association Rule Mining Model. Somboon Anekritmongkol, Kulthon Kasamsan International Journal of Advancements in Computing Technology, Volume 3, Number 1, February 2011.

[6] An Enhanced Scaling Apriori for Association Rule Mining Efficiency European Journal of Scientific Research ISSN 1450-216X Vol.39 No.2 (2010), pp.257-264

[7] Agrawal R, H. Mannila, R. Srikant, H. Toivonen, and A.Verkamo. "Fast discovery of association rules. *Advances in Knowledge Discovery and Data Mining, San Jose, CA*, pages 307-328, 1996.

[8] Agrawal R, T. Imielinski, and A. Swami. "Mining association rules between sets of items in large databases". *Proc. of the ACM SIGMOD Washington, D.C*, pages 207-216, May 1993.

[9] Alok Sharma, and Kuldip K. Paliwal, "Rotational Linear Discriminant Analysis Technique for Dimensionality Reduction*", IEEE Transactions on Knowledge and Data Engineering*, Vol. 20, No. 10, October 2008.

[10] Anthony K.H. Tung, Hongjun Lu, Jiawei Han, Member, and Ling Feng, "Efficient Mining of Inter transaction Association Rules" *IEEE Transactions on Knowledge and Data Engineering*, Vol. 15, No.1, January/February'03.

[11] Bodon. F, "A Survey on Frequent Itemset Mining", Technical report, Budapest Univ. Of Technology and Economics, 2006.

[12] Cheung D, V.T Ng, A. Fu, and Y.Fu. "Efficient mining of association rules in distributed databases". *IEEE Trans. Knowledge and Data Engineering*, pp 1-23, 1996.

[13] Elena Baralis, Tania Cerquitelli, and Silvia Chiusano, "IMine: Index Support for Item Set Mining", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 21, No. 4, April 2009.

[14] Ghosh, A. and S. Dehuri, "Evolutionary algorithms for multi-criterion optimization: A survey". *International Journal on Computers and Infromation Science*, 2: 38–57, 2004.

9. AUTHORS PROFILE

1. Dr. Siva Rama Prasad currently working as a head of the International Business Administration department at Acharya Nagarjuna University,India. He has published several research papers in various peer reviewed international journals . Authored seven books and also he is working as a research director in the department of CSE.

2. D. Bujji Babu currently working as an associate professor in the department of Computer Science and Engineering, at Prakasam Engineering College kandukur,A.P. India. He is having 4years of research and 10 years of teaching experience.He is a research scholar in the department of CSE at Acharya Nagarjuna University,India.