ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 5, May 2014

A Strategic approach in Knowledge Mining for Business intelligence using Temporal Data

Shruti Kirti Nandan[1], Chetan Mudgal[2], Dr. Ranjana Vyas[3]

Student, MBA-IT Division, Indian Institute of Information Technology, Allahabad, India[1,2]

Faculty, MBA-IT Division, Indian Institute of Information Technology, Allahabad, India[3]

Abstract: Though Business Intelligence has significantly benefitted with Data Mining which involves evaluating large set of data to find relevant patterns for better understanding and decision making in a particular Business context. In last decade data mining has though addressed many business issues but has also shown some limitations as well. One of the problematic areas of data mining is handling of temporal data, as it is established that transactional data has some seasonal behavior (As Supermarket Sales pattern changes during weekends) and thus need of temporal data mining. And another area of concern is of integrating existing domain knowledge in the mined results. Our paper proposes to address these issues. During past few years many approaches of temporal data mining were put forward with useful applications but they were largely incorporated either on Association Rule Mining or on Classification but We have proposed Temporal aspect integrated with Associative Classification. This integration of Temporal Associative Classification will make the mining process fast with better results, which were otherwise not revealed, which can further enhanced Organizations to have Effective decision making and time based Strategy Planning. Knowledge Management aspect is crucial in any organizational decision making and thus needs to be incorporated in Temporal Data mining process. Data mining acts as a tool for organisation, integration, extracting the data of relevance, error correction aids in a better revenue generation.

Keywords: Temporal Data Mining, Knowledge Management, Temporal predictive association rule

I. INTRODUCTION

1.1 Temporal Data Mining

Drastic increase of database in last decade has lead to the cause in discovery of hidden patterns in database, the emphasis is mainly on, data clustering, data classification and relationship discovery [Claudia][1][3]. Temporal data has great importance in the field of financial data forecasting, web usage monitoring, sales monitoring etc. Temporal data mining constitutes of mining vital information from data with time stamping, example of temporal data are the time series data.[**Temporal Data Mining** By Theophano Mitsa][4]

1.2 Temporal database

Temporal database constitutes of valid time and transaction time.

Valid time (exact time of data record), Transaction time (time of entry of record in database)

1.2.1 Pattern Discovery in temporal data

The main concept behind pattern discovery is to discover patterns occurring in the database like sequential pattern, associations, classification models etc. Pattern discoveries also concerned with areas such as neural network, clustering, sequential pattern mining, association mining.

Pattern discovery in event sequences are:

1) Sequence mining
2) Association rule discovery
3) Frequent Episode Discovery.

Sequence mining involves algorithms like Apriori and GSP which are Multiple Data Sequences Pattern Mining. Whereas Association rule discovery is single data sequence pattern mining. Frequent Episode Discovery is also a single data sequence pattern mining and it is a sequence of events appearing within a particular time window in a definite order.

Pattern discovery in time series comprises of

1) Motif and anomaly discovery
2) Streaming data pattern discovery.

Motif and anomaly discovery based pattern discovery is widely used in bioinformatics and computer networking monitoring. Whereas streaming data pattern discovery based pattern discovery finds its applications in Financial Data analysis or sensor data.

Table 1:shows Sequential pattern mining for multiple data sequences.

Table I

Sequence ID	Purchase data record
1	<bread, milk>
2	<(wheat, rice), bread, (berry, sausage)>
3	<(bread, butter, sausage)>
4	<bread, cheese, sausage>
5	<butter>

Table 2 shows Sequential pattern mining for a single data sequence

Table II

Data Sequence
< A1, A2,A3,A4,A5,A6,A7……...An>

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 5, May 2014

Prediction

In order to use a model to assess the class of an unlabelled sample is known as Prediction. Prediction is also used to estimate the value or value ranges of an attribute that a given sample can have. Two major types of prediction problems are Classification and Regression. In which Regression is used to predict continuous value and Classification is used to predict discrete or nominal values. There are the following three broad categories of predictions.

1) Event prediction
2) Event duration prediction
3) Time Series Forecasting:

Event prediction covers rare event prediction. Whereas Event duration prediction includes Regression and finally Time Series Forecasting which considers Moving average, Auto regression and ARMA models.

Classification

The two most crucial data mining techniques are Classification Rule Mining and Association Rule Mining. The objective of classification rule mining is identification of small set of rules present in the database to construct a errorless classifier. Whereas Association rule mining discovers all rules in database that meets minimum support and minimum confidence limitations (aggarwal et al). Both Classification and Association rule mining are crucial in empirical implementation. Thus integration of these two approaches leads in formulation of associative classification.

Associative Classification

It has been observed that Association Rules could also be used to model relationships between class labels and features from a training set. Therefore, association rules were used to efficiently build a classification model from very large training datasets. There are some Associative Classifiers proposed by data mining researchers mainly differing in the strategies.

- *Associative Classification* constitutes association mining and classification into one system [5], [6], [7], [8], [9],[10], [11].
- Main goal of Association Mining is to uncover the descriptive knowledge from databases.
- Whereas classification focal point is on constructing a classification model for categorizing new data.
- Association classification and Classification rule mining are crucial in empirical data mining applications. Integration of these two works can result in convenience of the end user.

Associative Classification system is formulated in two steps:

1) Discovering all the event associations.
2) Using association pattern to construct classification rules in order to build a classifier.

In the first step, the goal is to uncover the fundamental association patterns in a database.

In the second step, the task is to select a small set of relevant association patterns discovered to construct a classifier given the predicting attribute.

In second step, using Attribute prediction a classifier is constructed through selection of small set of applicable association patterns.

Data mining in the proposed associative classification framework involves three steps:

i. Continuous attributes is discrediting.
ii. Class association rules are generated.
iii. Building a classifier based on the generated CARs.

Following endowments made by this work are:

1. New way to construct accurate classifiers is offered. (Quinlan 1992)
2) It makes association rule mining techniques relevant to classification tasks.
3) By using existing classification systems solution to number of problems is archived.

Temporal Associative Classification Rule Approach

In Temporal data item the corresponding time related attribute is associated with temporal association rules that can be more useful and informative than basic association rules. For example

{Bread}⇒ {Butter} raises to 50% during 7 pm to 10 pm everyday

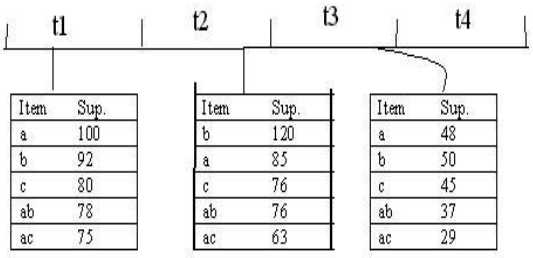


Fig.. 1 Division of Data in different Time-Interval

Temporal Associative Classifiers

As the data is dynamic in nature and changes with respect to time. Thus incorporating temporal features to this approach will give good results .The objective of temporal approach is to produce pattern among items with respect to time.

For example {bread} ⇨ {butter} mining from the temporal data we can get a more understanding fundamentals that the support of {bread} ⇨ {butter} rises to 50% during 7 pm to 10 pm everyday [3]. These fundamentals are more explanatory and useful in formulating strategic decision making in every field.

Why Temporal Associative Classification..?

Time plays a vital role in planning the strategy.Data can be accessed according to Time intervals and these time intervals are:

1) Hours/weeks
2) Quarter
3) Bi-monthly
4) months

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 5, May 2014

For example instead of the basic association rule of {bread}⇒ {butter} mining can be done from the temporal data. Like {bread}⇒ {butter} raises to 50% during 7 pm to 10 pm every day, marketing strategies can be formulated for weekends sale, effective decision making and other strategies for sales can be formulated using temporal association rule.

Comparative Study of Associative Classifier and Temporal Associative Classifier

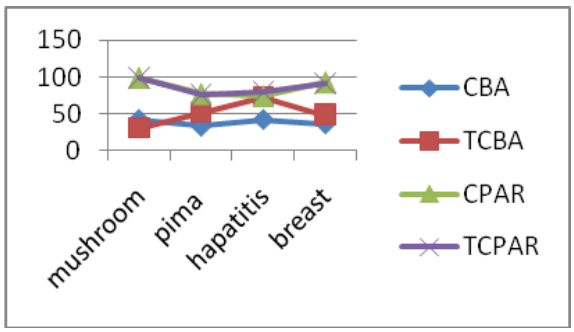
II. MINING ON MONTHLY BASIS USING CLASSIFIERS CBA,TCBA,CPAR,TCPAR

We have analyzed different dataset with the associative classification and same dataset was used with temporal Associative Classification. We can see that the Temporal Associative Classification accuracy is much higher then the normal associative classifiers. Comparative study was made between CBA TCBA CPAR and TCPAR. This classifier was again used with same dataset Quarterly basis. The result shows that the accuracy is higher than the what we used in the normal classifier.

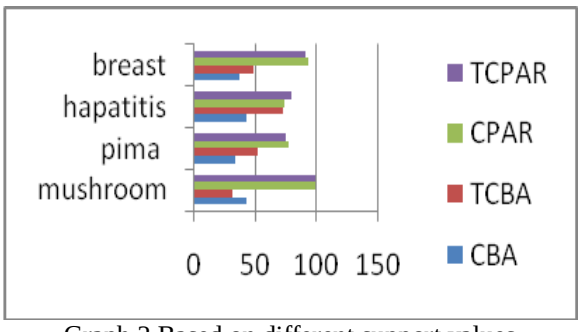
Table:III

DATASET NAME	CBA	TCBA	CPAR	TCPAR
mushroom	42.42	31.18	98.77	98.67
pima	33.22	52.14	77.4	75.5
hapatitis	42.52	72.85	74.34	80
breast	36.53	48.87	92.95	91.43

Table 3 Analysis of Associative Classifier & Temporal Associative Classfier



Graph 1 Based on different support values



Graph 2 Based on different support values

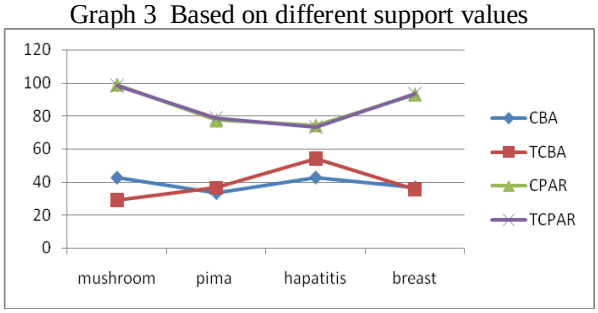
III. MINING ON QUARTERLY BASIS USING CLASSIFIERS CBA, TCBA, TCPAR, CMAR

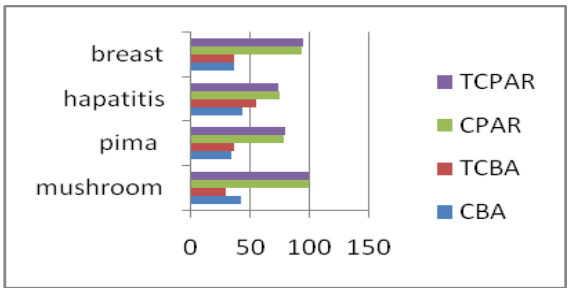
Table:IV

DATASET NAME	CBA	TCBA	CPAR	TCPAR
mushroom	42.42	29.17	98.77	98.51
pima	33.22	36.52	77.4	78.81
hepatitis	42.52	54.15	74.34	73.27
breast	36.53	35.88	92.93	93.66

Table:4 Analysis of Associative Classifier & Temporal Associative Classifier done on Quarterly bases

Graph 3 Based on different support values





Graph 04 Based on different support values

Knowledge Management Approach:

It was observed that in order to reach to a point where we get an useful technique for mining crucial knowledge from the Data Mining process, the appropriate integration of existing knowledge is required. With a reference to Market Basket Analysis, integrating DSS and DM the innovative concept of knowledge management approach which integrates Data Mining and knowledge inference is discussed. Conclusions were made that knowledge which was derived from data will definitely add value to a business through products, processes and if Association Rules generated by Data Mining process can be converted into Business rules then decision making will become more accurate and thus it would be easy to formulate business strategy if efficient knowledge management approach can be used. With a purpose to use and implement past experience, domain knowledge and specialized monitoring coupled with appropriate data mining techniques such as Multilevel Sequential mining and Temporal Associative Classifications.

Before understanding knowledge management it is provident to discuss various categories of knowledge. These are:

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 5, May 2014

1) Explicit knowledge.
2) Tacit knowledge

The knowledge which can be articulated in formal language is known as explicit knowledge. It includes grammatical statements (like numbers, words), mathematical expression, manual, specification etc. Explicit knowledge can be easily transmitted electronically, processed by a computer or can be easily stored in databases. Whereas the knowledge which is difficult to be transformed into useful form is known as Tacit Knowledge.

Temporal CPAR, a Temporal Associative Classifier which is discussed in above section id considered as an excellent Data Mining techniques in Supermarket Transaction database. In this, the items which are purchased by a customer will be stored in time stamping of items purchases in day and time. And with the help of this item sale pattern which are hidden will be obtained at this stage. Data Mining and Decision Support Application can be transformed into competitive Marketing Strategy. In these Knowledge Management Approach as discussed below. And thus by using product and processes a great value can be added to the business by Data Driven Knowledge which eventually leads to better decision making. Henceforth the processes can be named as Knowledge Based Marketing.[2]

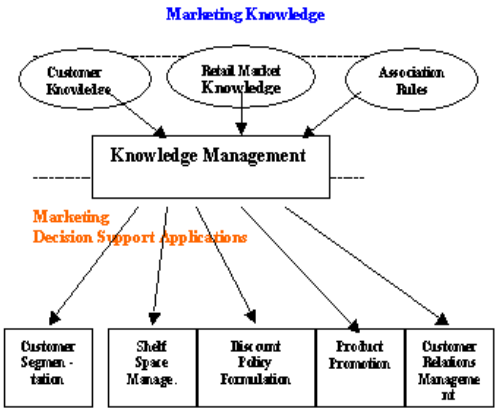


Fig.. 2 Knowledge Management approach for Marketing

Ranging from Operation to Finance the analysis of large data stream in order to identify hidden patterns is of great importance. Thus in this paper number of datasets are proposed for identifying the temporal pattern of data which is very crucial in formulating Business Strategy. Thus this Knowledge Management System approach is very important in various organizations to integrate Temporal Data Mining results while incorporating Domain Knowledge.

With reference to fig 05, by incorporating Domain Knowledge and experience, Association Rules can be converted into Business Rules. Likely customer knowledge, Retail Market knowledge and related experience in general may combine with Association Rules for Decision Making in the market basket analysis.

The proposed Knowledge Management approach can be extended to other application Domain of Data Mining.

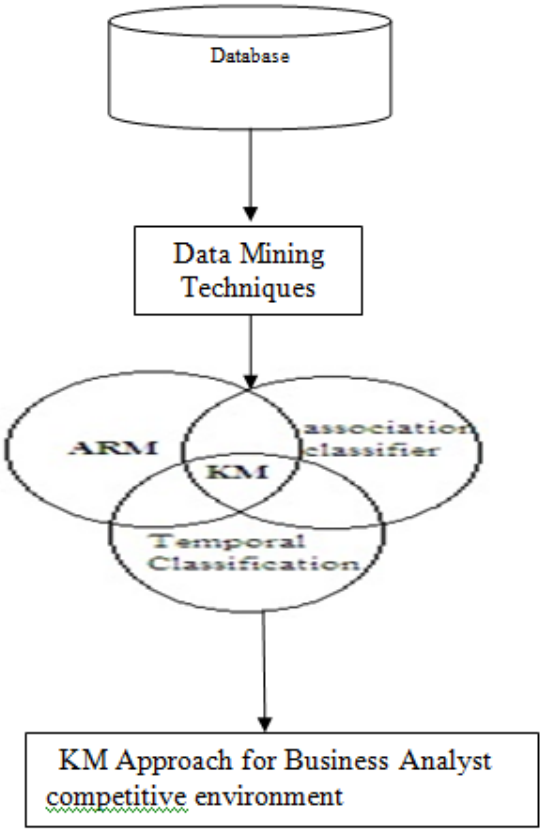


Fig. 3 Integration of Three methods for KM

IV. CONCLUSION

In this paper we have used data mining techniques through integration of Associative and Temporal classification, which attempts to provide a new approach of Knowledge Management. This information and knowledge can be Beneficial to corporate world and this information if given in right time in right format helps managers and analyst to take better decisions. As in our analysis, we have noticed that information, if bifurcated according to time then the analysis will not only be faster but the accuracy of the temporal classifier is much high, so helps organizations to plan there Marketing policy, Strategizing and bringing about new innovation trends and can be planned and handled them in more efficient and organized manner. In today's scenario, Knowledge management and Business intelligence can better utilize their intellect, through which Companies can acquire and utilize this information in much better ways to utilizing with Data mining Techniques.

REFERENCES

[1] Sunita Soni ,O.P.Vyas, "Using Associative Classifiers for Predictive Analysis in Health Care Data Mining"
[2] Jean-Louis ERMINE, "CHALLENGES AND APPROACHES FOR KNOWLEDGE MANAGEMENT"
[3] Cláudia Antunes , "Temporal Pattern Mining Using a Time Ontology"
[4] Theophano Mitsa, "Temporal Data Mining"
[5] G. Dong, X. Zhang, L.Wong, and J. Li. CAEP,"classification by aggregating emerging patterns". In Proceedings of The Second International Conference on Discovery Science (DS'99), pages 43–55, Japan, December 1999.
[6] J. Li, G. Dong, K. Ramamohanarao, and L. Wong. DeEPs: "a new instance-based lazy discovery and classification system". Machine Learning, 54(2):99–124, 2004.

[7] W. Li, J. Han, and J. Pei. CMAR: "Accurate and efficient classification based on multiple class-association rules". In Proceedings of The 2001 IEEE International Conference on Data Mining (ICDM'01), pages 369–376, San Jose, CA, November 2001.

[8] B. Liu, W. Hsu, and Y. Ma. " Integrating classification and association rule mining". In Proceedings of the Fourth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 80–86, New York, NY, August 1998.

[9] K. Wang, S. Zhou, and Y. He. "Growing decision tree on support-less association rules". In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining(KDD'00), pages 265–269, Boston, MA, August 2000

[10] Y. Wang and A. K. C. Wong. "association to classification: Inference using weight of evidence". IEEE Trans. On Knowledge and Data Engineering, 15(3):764–767, 2003.

[11] X. Yin and J. Han. CPAR: "Classification based on predictive association rules". In Proceedings 2003 SIAM International Conference on Data Mining(SDM'03), pages 331–335, San Francisco, CA, May 2003