A Survey of Cyclic Association Rule

Kaushal K. Patel
Master Of Engineering
Computer Science And Engineering
Hashmukh Goswami College Of Engineering,Ahmedabad,India

***Abstract* - Recently, the major problems of major companies how to mine effective data from their data sources, and they also want fast retrieval of datasets, therefore it is big area for researcher to researching new technique. so many researchers are try to find and develop new methods or new techniques for finding fast retrieval item sets or generate new datasets. In data mining there are many techniques for mine effective data from data sources, the techniques like classification, association, generalization, regression, clustering, etc. Among them I am try to focusing on cyclic association rule mining, which cyclic association rule is classified as a category of the temporal association rules, and temporal association rule classified as a category of association rule. Cyclic association rule have been introduced in order to discover rules from items that characterized by their regular variation over time or cyclic in nature. There are lots of algorithms for generating rules that the algorithm can yield significant performance benefits when compared to other algorithms and also provide predication that benefit to take decisions on bases of predication. So in this survey paper I am representing the comparison of different algorithms which are introduced by various authors during some past years.**

***Key Words*-Association Rule, Data set, Cyclic Association Rule, Generalization, Regression, Clustering**

I.INTRODUCTION DATA MINING

Nowadays in data collection and storage of data technology have made allowed for many companies to keep or store large amounts of data relating to their business. At the same time, cheap computing power has also made some automatic analysis of this data feasible. This activity is commonly referred to as “data mining”[1].

Data mining is technique of computer science which allowed us to the computational process of discovering patterns in large datasets which include methods at the intersection of artificial intelligence, machine learning, statistics and database system.

The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. The actual data mining task is the automatic or semi-automatic analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining). This usually involves using database techniques such as spatial indices. These patterns can then be seen as a kind of summary of the input data, and may be used in further analysis or, for example, in machine learning and predictive analytics. For example, the data mining step might identify multiple groups in the data, which can then be used to obtain more accurate prediction results by a decision support system. Neither the data collection, data preparation, nor result interpretation and reporting are part of the data mining step, but do belong to the overall KDD process as additional steps.

The Knowledge Discovery in Databases (KDD) process is commonly defined with the stages:

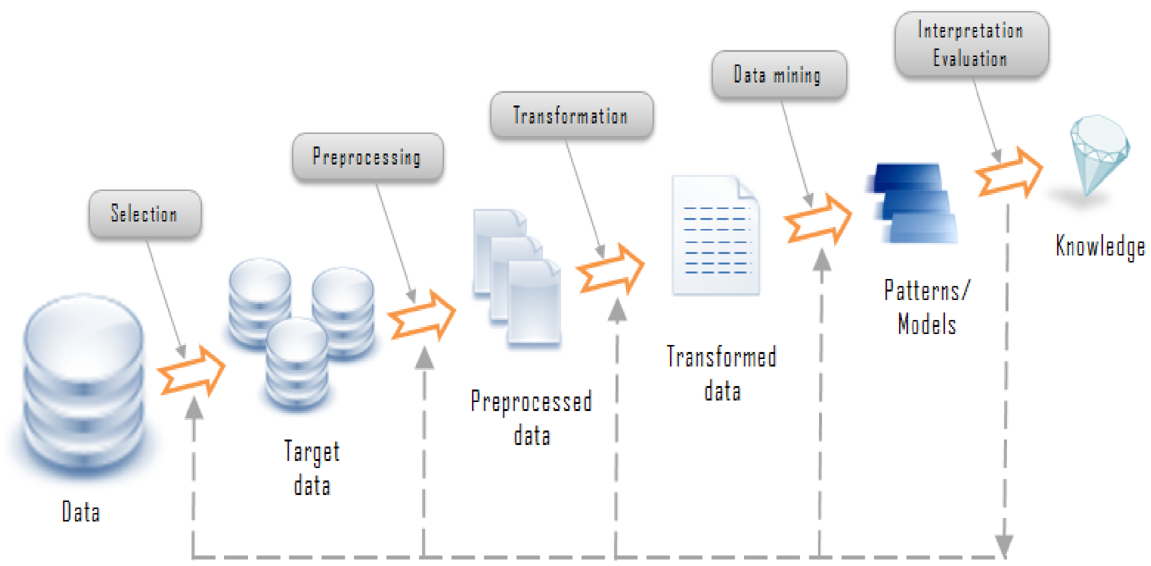

Fig.1 KDD Process

(1) Selection
(2) Pre-processing
(3) Transformation
(4) *Data Mining*
(5) Interpretation/Evaluation.

Selection
We have many data sources where our data are reside there.so first of all we have to select data source,on which data we are intresting.
Pre-processing
Before data mining algorithms can be used, a target data set must be assembled. As data mining can only uncover patterns actually present in the data, the target data set must be large enough to contain these patterns while remaining concise enough to be mined within an acceptable time limit. A common source for data is a data warehouse Pre-processing is essential to analyze the multivariate data sets before data mining. The target set is then cleaned. Data Cleaning removes the observations containing noise and those with missing data.

II. Data mining involves six common classes of tasks:

- **Anomaly detection (Outlier/change/deviation detection)** – The identification of unusual data records, that might be interesting or data errors that require further investigation.
- **Association rule learning (Dependency modeling)** – Searches for relationships between variables. For example a supermarket might gather data on customer purchasing habits. Using association rule learning, the supermarket can determine which products are frequently bought together and use this information for marketing purposes. This is sometimes referred to as market basket analysis.
- **Clustering** – is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data.
- **Classification** – is the task of generalizing known structure to apply to new data. For example, an e-mail program might attempt to classify an e-mail as "legitimate" or as "spam".
- **Regression** – attempts to find a function which models the data with the least error.
- **Summarization** – providing a more compact representation of the data set, including visualization and report generation.

III.ASSOCIATION RULE MINING:
Following the original definition by Agrawal et al. the problem of association rule mining is defined as: Let $I=\{i_1,i_2,\ldots,i_n\}$ be a set of n binary attributes called *items*. Let $D=\{t_1,t_2,\ldots,t_m\}$ be a set of transactions called the *database*. Each transaction in D has a unique transaction ID and contains a subset of the items in I. A *rule* is defined as an implication of the form $X \Rightarrow Y$ where $X,Y \subseteq I$ and $X \cap Y = \emptyset$. The sets of items (for short *itemsets*) X and Y are called *antecedent* (left-hand-side or LHS) and *consequent* (right-hand-side or RHS) of the rule respectively.
To illustrate the concepts, we use a small example from the supermarket domain. The set of items is $I=\{\mathrm{milk},\mathrm{bread},\mathrm{butter},\mathrm{beer}\}$ and a small database containing the items (1 codes presence and 0 absence of an item in a transaction) is shown in the table to the right. An example rule for the supermarket could be $\{\mathrm{butter},\mathrm{bread}\} \Rightarrow \{\mathrm{milk}\}$ meaning that if butter and bread are bought, customers also buy milk.
Note: this example is extremely small. In practical applications, a rule needs a support of several hundred transactions before it can be considered statistically significant, and datasets often contain thousands or millions of transactions.

Example database with 4 items and 5 transactions				
transaction ID	**milk**	**bread**	**butter**	**beer**
1	1	1	0	0
2	0	0	1	0
3	0	0	0	1
4	1	1	1	0
5	0	1	0	0

IV.ASSOCIATION RULE ALGORITHMS
1) Apriori algorithm
Apriori is the best-known algorithm to mine association rules. It uses a breadth-first search strategy to count the support of itemsets and uses a candidate generation function which exploits the downward closure property of support.
2) FP-growth algorithm
FP stands for frequent pattern.
In the first pass, the algorithm counts occurrence of items (attribute-value pairs) in the dataset, and stores them to 'header table'. In the second pass, it builds the FP-tree structure by inserting instances. Items in each instance have to be sorted by descending order of their frequency in the dataset, so that the tree can be processed quickly. Items in each instance that do not meet minimum

coverage threshold are discarded. If many instances share most frequent items, FP-tree provides high compression close to tree root.
Recursive processing of this compressed version of main dataset grows large item sets directly, instead of generating candidate items and testing them against the entire database. Growth starts from the bottom of the header table (having longest branches), by finding all instances matching given condition. New tree is created, with counts projected from the original tree corresponding to the set of instances that are conditional on the attribute, with each node getting sum of its children counts. Recursive growth ends when no individual items conditional on the attribute meet minimum support threshold, and processing continues on the remaining header items of the original FP-tree.
Once the recursive process has completed, all large item sets with minimum coverage have been found, and association rule creation begins

V.TYPES OF ASSOCIATION RULE

1) Context Based Association Rules

Context Based Association Rules is a form of association rule. Context Based Association Rules claims more accuracy in association rule mining by considering a hidden variable named context variable which changes the final set of association rules depending upon the value of context variables. For example the baskets orientation in market basket analysis reflects an odd pattern in the early days of month. This might be because of abnormal context i.e. salary is drawn at the start of the month.

2) Generalized Association Rules hierarchical taxonomy (concept hierarchy).

3) Quantitative Association Rules categorical and quantitative data.

4) Interval Data Association Rules e.g. partition the age into 5-year-increment ranged.

5) Maximal Association Rules

6) Sequential pattern mining

Discovers subsequences that are common to more than minsup sequences in a sequence database, where minsup is set by the user. A sequence is an ordered list of transactions.

7) Sequential Rules

Discovering relationships between items while considering the time ordering. It is generally applied on a sequence database. For example, a sequential rule found in database of sequences of customer transactions can be that customers who bought a computer and CD-Roms, later bought a webcam, with a given confidence and support.

8) Temporal association rule

However the algorithms used for association rule mining do not consider the temporal patterns in customer shopping behavior. For example consider association rules of the form {turkey ->pumkin pie} or {plum cake -> Christmas decorations} which apply during periods close to Thanksgiving and Christmas. These association rules might be infrequent for the rest of the year and therefore do not cross the minimum support threshold required for them to appear in the frequent items sets and henceforth in the association rules discovered by the mining system. Also, if they do appear in the association rules determined by our mining system, these association rules are misleading as they are not valid for all time periods. Therefore, taking temporality into consideration is important for deriving more robust association rules.

VI.TYPES OF TEMPORAL ASSOCIATION RULE:

To solve the temporal association rules issue, five streams approaches are identified:

1)Interval Association Rules : involve discovering of association rules during a time Interval 1 related to each item ;

2)Calendric Association Rules : Ramaswamy et al have introduced the calendric association rules. Then, Li et al generalized the already proposed model .The Calendric Association Rules offer a generic model of calendric association rules extraction based on calendar algebra expression. In fact, the latter is a collection of time intervals describing some real-life phenomenon.By integrating calendars and association rules, we can generate discovered patterns. For example, we buy cars equipped with air conditioning in summer : Cars => Air conditioning in Summer. And we deduce that there are more accidents and victims on weekends : Accidents => Victims on Weekends;

3)Temporal Predicate Association Rules : extend the association rule model by adding to the rules a conjunction of binary temporal predicates that specify the relationships between the time stamps of transactions. By finding associated items first and then looking for temporal relationships between them, it is possible to incorporate potentially valuable temporal semantics. The approach of temporal reasoning accommodates both point-based and interval-based models of time simultaneously. For example, we buy milk and butter in a given time-interval : Milk => Butter between 19 o'clock and 23 o'clock. And we buy bread and juice at a given time : Bread => Juice at nine o'clock;

4)Sequential Association Rules : dedicated to the extraction of association rules from data represented as sequences. Indeed, a sequence contains sorted sets of items ordered by transaction time . For example, association rule mining does not take the time stamp into account, the rule can be Buy A => Buy B. If we take time stamp into account then we can get more accurate and useful rules likewise : Buy A implies Buy B within a week, or usually people Buy A every week;

5)Cyclic association rules : aim to discover association rules from articles having regular cyclic variation over time.

1) CYCLIC ASSOCIATION RULE, Banu ozden,Sridhar Ramaswamy,Avi Silberscatz.

- They showed by exploring the relationship between cycles and large itemsets.
- They identified optimization techniques that allow us to minimize amount of wasted work perform during the data mining process.
- They introduced interleaved algorithm for cyclic large itemset detection.

- The study showed that performance benefits ranging from 5% to several hundred percent can be obtained through the use of the optimizations when compared to the more straightforward approach.

2) PRIVACY PRESERVATION FOR GLOBAL CYCLIC ASSOCIATIONS IN DISTRIBUTED DATABASES,Nirali Nanavati,Devesh Jinwala
- They extend the Interleaved algorithm to find global cycles in cyclic association rules privately.
- And they used Techniques for privacy preservation are segregated on the basic of homomorphic technique and secret sharing.
- They define for future work is that how global cycles can be detected in a co-opetative setup while maintaining the privacy of the individual parties. However there are a few open research challenges which include applying these privacy preserving theories to other temporal rule mining methods like calendric association rules and temporal predicate association rules.

3) TOWARDS AN INCREMENTAL MAINTAINCE OF CYCLIC ASSOCIATION RULES, Eya ben Ahmed,Mohmed salah Gouider
- They introduced IUPCAR algorithm for incremental mining of cyclic association rule.
- This algorithm provides the benefits of fast incremental mining and efficient cyclic association rules extraction.

4) TOWARDS A NEW MECHANISM OF EXTRACTING CYCLIC ASSOCIATION RULES BASED ON PARTITION ASPECT, Eya ben Ahmed,Med salah Gouider
- They addressing the issues of cyclic association rule conduct as to identify abnormalities affecting the pioneering approaches.
- They introduced a new model known as PCAR.
- In light of its experimental study, They stress on the considerable performance of their contribution and its outperformance vs. the INTERLEAVED and SEQUENTIAL algorithms.

5) INCREMENTAL UPDATE OF CYCLIC ASSOCIATION RULE, Eya ben ahmed
- They introduced the problem of incremental maintaince of cyclic association rule.
- Therefore,flying over the pioneering approaches handling the incremental update of association rule.
- That's why they introduced new proposed called IUPCAR algorithm dedicated particularly to update the cyclic association rule.
- Evaluate its efficiency.

6) MINING CYCLIC ASSOCIATION RULES FROM MULTIDIMENSIONAL KNOWLEDGE,Eya ben Ahmed,Ahlem nabil,Faize Gargouri
- They proposed a new methods to extract cyclic association rule from multidimensional context such as "A pharmaceutical company sells a product i.e. astradol,with a total turnover bracket ranging between 50,000 90,000 every month"
- Thus, they provide new algorithm, called "RACYM" to extract such pattern.

7) FINDING CYCLIC FREQUENT ITEMSETS,F.A Mazorbuiya,M.Shenify,A.Khan
- They take the output of algorithm proposed by mahanta et al. And takes the input as out of mahanta's algorithm supplies all frequent sets which are cyclic in nature.
- They may have some frequent itemsets where the time gaps are almost equal in length but the duration of the intervals of frequency are not equal even in the approximate sense.
- They may also have some frequent itemsets where the time-gaps are not equal in length but durations of the intervals are almost equal.
- And they modified algorithm according to find such frequent itemsets.
- In future, They may also modify their algorithm to get more accurate results. They would also like to find partially periodic patterns and other types of patterns which may exist in the datasets.

8) PERIODIC PATTERN MINING ALGORITHM & APPLICATION, G.N.V.G Sirosha,M.Shashi & G.V padma Rayu
- They said that Periodic patterns can be mined from datasets like biological sequences, continuous and discrete time series data, spatiotemporal data and social networks. Different criteria are used to classify the periodic patterns.
- They present an overview of different types of periodic patterns and Application along with a discussion of the algorithm that are used to mine these patterns.
- They are also discussed about efficiency user interaction needed & noise resilient nature of their proposed algorithms.
- They proposed ideas for future direction research that they mine a large number of patterns when min_sup is set very low. Like any frequent pattern mining algorithm, frequent periodic pattern mining algorithms also generate a large number of periodic pattern when min_sup or min_conf is set low. In order to reduce the redundancy of the generated

output and to improve the efficiency of mining process there is a need for the development of algorithms that mine closed and maximal periodic patterns.

9) AN EFFICIENT ALGORITHM FOR MINING LOCALLY FREQUENT ITEMSETS, F. A. Mazarbhuiya, M. Shenify, Md. Husamuddin

- They provide an algorithm which extract various types of frequent itemsets that may present in a dataset and cannot be extracted by J.M Ale and G.H.Rossi 's method.
- They discussed on tri-based implementation.
- The trie-structure is implemented in such a way that it can facilitate efficient candidate generation.
- They said that the implementation discussed in their paper is the most efficient one but it serves the purpose of justifying the claim made by A. K. Mahanta, F. A. Mazarbhuiya and H. K. Baruah; Finding Locally and Periodically Frequent Sets and Periodic Association Rules and there is an ample scope of improving the work. In future it will be looked for better implementation. In future comparative studies can be made with other known methods.

10) Mining Periodic Event Pattern from RDF Datasets, Anh Le and Michael Gertz

- They presented a framework to mine periodic patterns from RDF datasets describing events.
- The patterns represent topics of events that co-occur and repeat over time. Event topics are modelled based on event templates that are derived from events by generalizing subjects, predicates, objects, time and locations.
- As a separate input to their approach, hierarchies for concepts, time and locations are employed to obtain event templates and their generalizations.
- They demonstrated their framework using real datasets and found some interesting patterns.
- For future work they are currently extending their framework to support more expressive types of such constraints, e.g., stating that ETs of a pattern must be similar in terms of concepts, time or locations. They are also investigating how to apply the framework to predict future events and to detect outliers in RDF-based event datasets.

Table -1 Comparison Algorithm of Various Authors

YEAR	Proposed Algorithm	Function
1998	Interleaved Algorithm	Identified Optimization Technique which allow us to minimize the amount of wasted work. Studied show that performance increase from 5% to several hundred percent.
2010	IUPCAR Algo(Incremental Update of C.A.R)	Generate the quality of c.a.r Study deeply the significance of the extracted c.a.r from human experts. Use vertical database representation to improve iupcar algo.
2010	PCAR Algo(Partition C.A.R)	Its outperformance vs Interleaved and sequential algo. Use Data cubes to generate multidimensional C.A.R
2011	MIHYCAR(multi level hybrid C.A.R)	They extend the scope of the studying of mining C.A.R from single level to multiple concept levels. Methods extracts patterns in the respect to parallel concept.
2012	CBCAR(Constraint-Based C.A.R)	Reduced the generated rule To integrate the concept on C.A.R derivation in order to guide the expert and extraction process
2012	Extend the INTERLEAVED Algo	For Find global cycles
2013	PPatterns algo	Proposed Framework for finding patterns

References

[1] B. Ozden, S. Ramaswamy, A. Silberschatz" Cyclic Association Rules, 14th International Conference on Data Engineering(ICDE'98), 412, 1998.

[2] R. Agrawal, T. Imielinski and A. N. Swami, Mining association rules between sets of items in largedatabases, In Proc. of 1993 ACM SIGMOD Int'l Conf on Management of Data, Vol. 22(2) of SIGMOD Records, ACM Press (1993), pp 207-216.

[3] Tamas Horvath,Thomas Gartner,Stefan Wrobel,Cyclic patterns kernels for predictive graph mining published in KDD'04 processing of the tenth ACM SIGKDD international conference on knowledge discovery and data mining pages 158-167.

[4] Eya Ben Ahmed,Mohamed salah gouider, Towards an Incremental maintenance of cyclic association rules. published in International journal of database management system (IJDMS),vol.2,no.4,November 2010 pages 46-57.

[5] Eya Ben Ahmed,Mohamed salah gouider, Towards a new mechanism of extracting cyclic association rule based on partition aspect published in IEEE Research Challenges in Information Science (RCIS), 2010 Fourth International Conference on 19-21 May 2010,Pages-69-78.

[6] Eya Ben Ahmed, Incremental update of cyclic association rules published in springer Intelligent data Engineering and Automated Learning – IDEAL 2010 Lecture Notes in Computer science Volume 6283, 2010, pp 387-395

[7] Eya ben Ahmed,Ahlem nabil,Faiez Gargouri, Mining Cyclis association rule from multidimensional knowledge published in IEEE Digital Information Management (ICDIM), 2011 Sixth International Conference on pages 12 - 17

[8] Nirali Nanavati,Devesh Jinwala, Privacy preservation for global cyclic association in distributed databases,published in Elsevier sciverse sciencedirect in 2nd international conference on communication,computing & security[ICCCS-2012],pages 962-969.
[9] Wafa Tebourski,Wahiba ben Abdessalem karaa, Cyclic association rules mining under constraints,published in International Journal of computer applications(0975-8887) vol.49,no.20,July 2012.
[10] F.A.Mazarbhuiya,M.Shenify,A.Khan,A.Farooq, Finding Cyclic Frequent Itemsets,published in IJCSI International Journal of computer science issues,vol.9,issue 6,no.1,Nov 2012.
[11] Mining Periodic Event Patterns from RDF Datasets,Anh Le and Michael Gertz,published 17th East European Conference, ADBIS 2013 Genoa, Italy, September 2013.
[13] PERIODIC PATTERN MINING ALGORITHM & APPLICATION,G.N.V.G,M.Shash,G.V.Padma,published in global journal ,2013
[14] http://en.wikipedia.org/wiki/Data_mining