A SURVEY OF DIFFERENT TYPES OF ASSOCIATION MINING ALGORITHMS

[1]Kamal Choudhary, [2]Megha Singh

[1,2]CSE DEPARTMENT , CENTRAL INDIA INSTITUTE OF TECHNOLOGY, INDORE (MP),India

ABSTRACT

Data is essential property for everyone. Vast amount of data is accessible on the planet. There are different archives to store the data into data distribution centers, databases, data storehouse and so forth. This huge measure of data needs to process with the goal that we can get helpful data. Data mining is a system to process data, select it, incorporate it and recover some valuable data. Data mining is an explanatory device which enables clients to break down data, classifications it and synopses the connections among the data. It finds the valuable data from extensive measure of social databases. Data mining can play out these different exercises utilizing its system like grouping, characterization, forecast, affiliation learning and so forth. This paper introduces a diagram of association rule mining calculations. Calculations are talked about with legitimate illustration and analyzed in view of some execution factors like precision, data bolster, execution speed and so on.

Keywords: *Data mining, Association rule mining*

1. INTROUDUCTION

Data mining [8] is the way toward breaking down data from alternate points of view and outlining it into valuable data. Data mining is an expository apparatus for examining data. It enables clients to examine data, order it, and compress the connections among data. Actually, data mining is the way toward discovering relationships or examples in expansive social databases. It includes some normal assignments like anomaly detection, clustering, association run learning, regression, summarization, classification and so on. Anomaly detection is the scan for things or occasions which don't fit in with a normal example. These identified examples are called oddities and mean basic and significant data in different application areas. It is likewise alluded as anomalies. Association run the show learning scans for connections among factors. For illustration a general store may assemble data about how the client obtaining the different items. With the assistance of association administer, the store can distinguish which items are much of the time purchased together and this data can be utilized for advertising purposes. This is now and then known as market based investigation. Clustering finds the gatherings and structures in the data somehow or then again another comparable route, without utilizing known structures in the data. Grouping sums up known structure to apply to new data. Take an illustration; an email program may endeavor to characterize an email as "authentic" or as "spam" mail. Relapse endeavors to discover a capacity which models the data with the slightest blunder. Rundown gives a more minimized portrayal of the data set, which incorporates representation and report age. Figure 1 indicate Knowledge Discovery in Database forms where it takes data from different stores like data stockroom, database, data storehouses, social database and so on. It performs different activities like data cleaning, mix, change and so on and produces valuable data from that used to present the mined knowledge to the user)

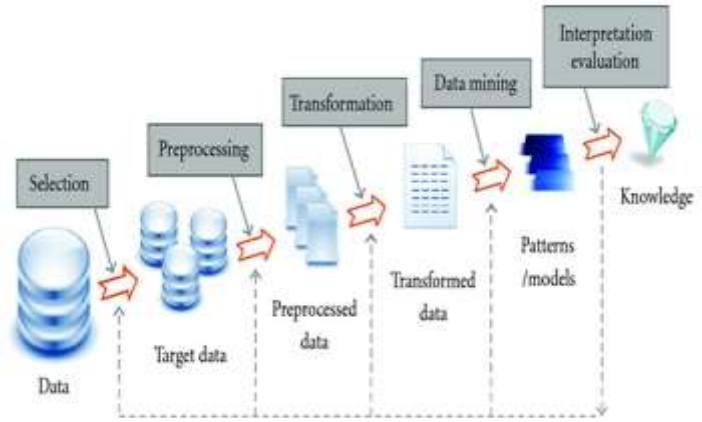

Figure 1.1: Data mining as a step in the process of knowledge discovery

Knowledge discovery as a process is depicted in Figure 1.1 and consists of an iterative sequence of the following steps:

- Data cleaning (to remove noise and inconsistent data)
- Data integration (where multiple data sources may be combined)
- Data selection (where data relevant to the analysis task are retrieved from the database)
- Data transformation (where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance)
- Data mining (an essential process where intelligent methods are applied in order to extract data patterns) Pattern evaluation (to identify the truly interesting patterns representing knowledge based on some interestingness measures

Knowledge presentation (where visualization and knowledge representation techniques are

2.Comparision of Different types of Algorithm

1) Apriori Algorithm

Apriori calculation is the inventiveness calculation of Boolean affiliation guidelines of mining incessant thing sets, raised by R. Agrawa and R. Srikan in 1994. The core principles of this theory are the subsets of frequent item sets are frequent item sets and the supersets of infrequent item sets are infrequent item sets. The calculation is utilized to discover all the regular thing sets.

In the main cycle, thing set A straightforwardly constitutes the first applicant thing set C1. Accept that A={a1,a2,… ,am}, at that point C1={{a1},{a2},… ,{am}}. In the Kth emphasis, right off the bat, the applicant thing set Ck of this emphasis rises as indicated by the regular thing set Lk-1 found in the last emphasis. (The candidate item set is the potential frequent item set and is the superset of the K-1th frequent item set. Item set with k candidate item sets is expressed as Ck, which was consisted by k frequent item sets Lk.) Then distribute a counter which has a initial value equals to zero to ever item set and scan affairs in database D in proper order. Make sure every affairs belongs to each item sets and the counter of these item sets will increase. When all the affairs have been scan, the support level can be gotten according to the actual value of |D| and the minimum support level of the certain Ck of the frequent item set. Repeat the process until no mew item occurs[10].
Algorithm consists mainly 2 steps first one is connecting step and pruning step.

Connecting step: in order to get Lk, connect Lk-1 with itself.
Set this candidate as Ck and assume L1 and L2 are the item sets of Lk-1. Li[j] is the jth item of Li.

Assume the affairs and items of the item set are in the dictionary order. Execute the connection Lk-1Lk-1, in which the elements of Lk-1, L1 and L1, are connectable, if they have the same first (k-2)th item.
Pruning step: Ck is the superset of Lk, that is that the members of it could be frequent or not, but all the k frequent item sets are all include in Ck. Scan the database, clear the counters of every candidate item sets of Ck to assure Lk. However, Ck might be very large, and then the amount of calculation will be huge. In order to decrease Ck, there are following method using the prosperities of Apriori: any infrequent item sets with k-1 items are not the subset of frequent item sets with k items.

structure and dynamically adjusts links in the mining process. A distinct feature of the proposed method is that it has a very limited and precisely predictable main memory cost and runs very quickly in memory-based settings. Moreover, it can be scaled up to very large databases using database partitioning.

2) AprioriTID

AprioriTID proposed by [14]. This algorithm has the additional property that the database is not used at all for counting the support of candidate itemset after the first pass. Rather, an encoding of the candidate itemsets used in the previous pass is employed for this purpose

3) FDM(Fast Distributed Mining)

FDM of affiliation rules has been proposed by [16], which has the following distinct highlights.
FDM (Fast Distributed Mining of association rules) has been proposed by [16], which has the following distinct features. 1. The generation of candidate sets is in the same spirit of Apriori. However, some relationships between locally large sets and globally large ones are explored to generate a smaller set of candidate sets at each iteration and thus reduce the number of messages to be passed.

4) GSP: Generalized Sequentiel Patterns (GSP)

Generalized Sequentiel Patterns (GSP) is representative Apriori-based sequential pattern mining algorithm proposed by Srikant & Agrawal in 1996 [17]. This algorithm uses the downward-closure property of sequential patterns and adopts a
multiple pass, candidate generate-and-test approach.

5) H-Mine:

H-Mine is an algorithm for discovering frequent itemsets from a transaction database developed by Peet al. [36] in 2007. They proposed a simple and novel data structure using hyper-links, H-struct, and a new mining algorithm, Hmine, which takes advantage of this data

structure and dynamically adjusts links in the mining process. A distinct feature of the proposed method is that it has a very limited and precisely predictable main memory cost and runs very quickly in memory-based settings. Moreover, it can be scaled up to very large databases using database partitioning.

6) SPADE:
SPADE is an algorithm for mining frequent sequential patterns from a sequence database proposed in 2001 by Zaki [25]. The author uses combinatorial properties to decompose the original problem into smaller sub-problems, that can be independently solved in main-memory using efficient lattice search techniques, and using simple join operations. All sequences are discovered in only three database scans.

3. ACKNOWLEDGEMENT

The most essential tasks of frequent pattern mining approaches are : itemset mining, sequential pattern mining, sequential rule mining and association rule mining. A decent number of data mining algorithms exist in the literature for mining frequent patterns. In this paper, we have introduced a concise diagram of the present status and future bearings of frequent pattern mining. Furthermore, we have played out a thorough investigation of a few calculations

and techniques that exists for the mining of successive examples. With over a time of broad research, a great number of research productions, advancement and application exercises in this space have been proposed. We give a short exchange of various calculations introduced along this decade with a similar investigation of a couple of critical ones in light of their execution. Be that as it may, we require to lead a profound research in view of a few basic issues with the goal that this space may have its genuine presence and profound effect in data mining applications.

4. REFERENCES

1. R. Agrawal, R. Srikant, "Fast Algorithms for Mining Association Rules", pp. 487-499.

2. X. Liu, P. He, "The Research of Improved Association Rules Mining Apriori ALgorithm", Proceedings of the Third International Conference on Machine Learning and Cybermetics, Shanghai, 26-29 August 2004, pp. 1577-1579.

3. J. Lei, B. Zhang, J. Li, "A new improvement on Apriori Algorithm", International Conference on Computational Intelligence and Security, Vol. 1, IEEE, 2006, pp. 840-844.

4. Y. Xie, Y. Li, C. Wang, M. Lu, "The Optimization and Improvement of the Apriori Algorithm", Education Technology and Training, International Workshop on Geoscience and Remote Sensing, ETT and GRS, Vol. 2, IEEE, 2008, pp. 663-665.

5. Z. Changsheng, L. Zhongyue, Z. Dongsong, "An Improved Algorithm for Apriori", First International Workshop on Education Technology and Computer Science, 2009, pp. 995-998.

6. L. Jing et. al, "An Improved Apriori Algorithm for Early Warning of Equipment Failure", 2009, pp. 450-452.

7. K. Shah, S. Mahajan, "Maximizing the Efficiency of Parallel Apriori Algorithm", International Conference on Advances in Recent Technologies in Communication and Computing, 2009, pp. 107-109.

8. H. Wu, Z. Lu, L. Pan, R. Xu, W. Jiang, "An Improved Apriori-based Algorithm for Association Rules Mining", Sixth International Conference on Fuzzy Systems and Knowledge Discovery, 2009, pp. 51-55.

9. Y. Liu, "Study on Application of Apriori Algorithm in Data Mining", Second International Conference on Computer Modelling and Simulation, 2010, pp. 111-114.