ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

Mining Human Activities in Smart Home for Health Care Application

Nithya R[1], Rajkumar S[2]

B.E Student, Department of Computer Science and Engineering, K.S. Rangasamy College of Technology, Tiruchengode, Tamilnadu, India[1]

Assistant Professor, Department of Computer Science and Engineering, K.S. Rangasamy College of Technology, Tiruchengode, Tamilnadu, India[2]

ABSTRACT: Frequent itemset mining may be a widely exploratory technique that focuses on discovering recurrent correlations among data. The steadfast evolution of markets and business environments prompts the need of data mining algorithms to urge significant correlation changes. Change mining, within the context of frequent itemsets, focuses on detecting and reporting significant changes within the set of mined itemsets from just one occasion period to a different. The generalized itemsets discovery that occur within the source data, and supply a high-level abstraction issues and new challenges within the analysis of itemsets that become rare, and thus are not any longer extracted. This project proposes a novel kind of dynamic pattern, namely the An Incremental FP- Growth Frequent Pattern Analysis, that represents the evolution of an itemset in consecutive time periods, by reporting the knowledge about its frequent generalizations characterized by minimal redundancy in case it becomes infrequent during particular time. To address Frequent Pattern Growth mining, it proposes Frequent Pattern Growth, an algorithm that focuses on avoiding itemset mining followed by post processing by exploiting a support-driven itemset generalization.

KEYWORDS : KDD, FP-Growth, Group Count techniques

I. INTRODUCTION

KNOWLEDGE DISCOVERY IN DATABASES

Knowledge discovery in databases (KDD) is that the process of discovering useful knowledge from a set of knowledge . Major application include marketing, fraud detection, telecommunication and manufacturing. Traditionally, data processing and knowledge discovery was performed manually. As time passed, the quantity of knowledge in many systems grew to larger than terabyte size, and will not be maintained manually. Moreover, for the successful existence of any business, discovering underlying patterns in data is taken into account essential. As a result, several software tools were developed to get hidden data and make assumptions, which formed a neighborhood of AI .

The KDD process has reached its peak within the last 10 years. The goal is to find high-level knowledge from low-level data.

The desire and wish for information has led to the event of systems and equipment which will generate and collect massive amounts of knowledge . Many fields, especially those involved in decision making, are participants in the information acquisition game. Application include: finance, banking, retail sales, manufacturing, monitoring and diagnosis, health care, marketing and science data acquisition. Advances in storage capacity and digital data gathering equipment have made it possible to seek out massive datasets, sometimes called data warehouses that measure in terabytes. NASA's Earth Observing System is expected to return data at rates of several gigabytes per hour by the end of the century.

(1) Modern scanning equipment record many transactions from common daily activities like supermarket or emporium checkout-register sales. The explosion within the number of resources available on the planet Wide Web is another challenge for indexing and rummaging through a continually changing and growing "database." Our ability to wade

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

through the data and turn it into meaningful information is hampered by the size and complexity of the stored information base. In fact, the shear size of the data makes human analysis untenable in many instances, negating the effort spent in collecting the data.. There are several viable options currently being used to assist in weeding out usable information. The information retrieval process using these various tools is mentioned as Knowledge Discovery in Databases (KDD)."The basic task of KDD is to extract knowledge (or information) from lower level data (databases)."
(2) There are several formal definitions of KDD, all agree that the intent is to reap information by recognizing patterns in data . Let us examine definition proposed by Fayyad, Piatetsky- Shapiro and Smyth, "Knowledge Discovery in Databases is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data."
(3) The goal is to distinguish from unprocessed data, something that may not be obvious but is valuable or enlightening in its discovery. Extraction of knowledge from raw data is accomplished by applying Data Mining methods. KDD has a much broader scope, of which data mining is one step in a multidimensional process.

CLASSIFICATION OF DISCOVERED KNOWLEDGE FREQUENT PATTERN

The problem of frequent pattern mining has been widely studied within the literature because of its numerous applications to a variety of knowledge mining problems like clustering and classification. In addition, frequent pattern mining also has numerous applications in diverse domains like spatiotemporal data, software bug detection, and biological data. The algorithmic aspects of frequent pattern mining are explored very widely. Frequent pattern mining is a rather broad area of research, and it relates to a wide variety of topics at least from an application specific-perspective. Broadly speaking, the in the area falls in one of four different categories:
Technique-centered: This area relates to the determination of more efficient algorithms for frequent pattern mining. A wide variety of algorithms have been proposed in this context that use different enumeration tree exploration strategies, and different data representation methods. In addition, numerous variations like the determination of compressed patterns of great interest to researchers in data processing .
Scalability issues: The scalability issues in frequent pattern mining are very significant. When the data arrives in the form of a stream, multi-pass methods can no longer be used. When the data is distributed or very large, then parallel or big-data frameworks must be used. These scenarios necessitate different types of algorithms.
Advanced data types: Numerous variations of frequent pattern mining are proposed for advanced data types. These variations have been utilized in a wide variety of tasks. In addition, different data domains such as graph data, tree structured data, and streaming data often require specialized algorithms for frequent pattern mining. Issues of interestingness of the patterns are also quite relevant in this context.
Applications: Frequent pattern mining have numerous applications to other major data processing problems, Web applications, software bug analysis, and chemical and biological applications. A significant amount of has been dedicated to applications because these are particularly important within the context of frequent pattern mining.

FREQUENT PATTERN MINING IN DATA STREAMS

In recent years, data stream became very fashionable due to the advances in hardware and software technology which will collect and transmit data continuously over time. In such cases, the main constraint on data processing algorithms is to execute the algorithms during a single pass. This is challenging because frequent and sequential pattern mining methods.

Frequent Items or Heavy Hitters: Frequent 1-itemsets need to be determined from a data stream in a single pass. Such an approach is usually needed when the entire number of distinct items is just too large to be held in main memory. Typically, sketch-based methods are used in order to create a compress data structure in order to maintain approximate counts.

Frequent itemsets: In this case, it is not assumed that the number of distinct items are too large. Therefore, the main challenge in this case is computational, because the typical frequent pattern mining methods are multi-pass methods. Multiple passes are clearly not possible in the context of data streams.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

FREQUENT PATTERN MINING WITH ADVANCED DATA TYPES

Although the frequent pattern mining problem is of course defined on sets, it are often extended to varied advanced data types. This was one among the earliest proposed extensions and is mentioned as sequential pattern mining. Algorithms are basic component for the frequent pattern mining problem. In general, the essential frequent pattern mining algorithms got to be modified carefully to deal with the variations required by the advanced data types.

ASSOCIATION RULE IN FREQUENT PATTERN

Association rule mining is used to find the frequent patterns found in various kinds of databases. Given a group of transactions, association rule mining aims to seek out the principles which enable us to predict the occurrence of a selected item supported the occurrences of the other items in the transaction.

Association rule mining is that the data processing process of finding the principles which will govern associations and causal objects between sets of things . So during a given transaction with multiple items, it tries to seek out the principles that govern how or why such items are often bought together. For example, spread and jelly are often bought together because tons of individuals wish to make PB&J sandwiches.

UTILITY PATTERN MINING

Utility pattern mining finds patterns from a database that have their utility value no less than a given minimum utility threshold. The utility of a pattern defines its importance and makes mined patterns more relevant surely applications. Primarily, the interest in utility patterns arises because it allows to associate relative importance to different items, and accounts for multiplicity of things . On the other hand, frequent-pattern mining can’t be used to find high utility patterns, due to its limitation of treating every item with equal importance with no use of item-quantity information. Applications like retail stores, where each item has different profit values and a transaction can have multiple copies of an item, will have a direct role of high utility pattern mining.

SEQUENTIAL PATTERN

Sequential pattern mining is the subject of data mining for finding statistically relevant patterns between itemsets. It is usually assumed that the values are discrete and the statistic mining is closely related.

There are several key computational problems occured within this field. It may includes building efficient databases and indexes for sequence information extracting the frequently occurring patterns, comparing sequences. Sequence mining problems are often classified as string mining which is usually supported string processing algorithms and itemset mining. Local process models extend sequential pattern mining to more complex patterns which can include choices, loops, and concurrency constructs additionally to the sequential ordering construct.

HIGH UTILITY MINING, CATEGORIZE OF UTILITY MINING PROBLEMS

The objective of utility mining is to discover the itemsets with highest utilities by considering user preferences. In utility mining, the utility of an itemset u(i) is defined as the sum of the utilities of itemset i in all the transactions containing i. An itemset i is named a high utility itemset if and as long as u(i)>= min_utility, where min_utility may be a user defined minimum utility threshold. Measuring the discovered patterns is important and the various criteria such as conciseness, coverage, reliability, peculiarity, diversity, novelty are used.

CATEGORIES OF UTILITY MINING

Interesting measures for mining high utility patterns are classified as objective measures, subjective measures and semantic based measures. An objective measure is predicated only on the data . Most objective measures are supported theories in probability, statistics, or scientific theory . Conciseness, generality, reliability, peculiarity, and variety depend only on the info and patterns, and thus are often considered objective. Objective measures such as support or confidence are based only on data. A subjective measure takes under consideration both the info and therefore the user of those data.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

II. EXISTING SYSTEM

In existing system a comprehensive survey of traditional data mining problems such as frequent pattern mining in the context of uncertain data can be found. Some concepts and issues arising from traditional sequential pattern mining and the mining of uncertain data.

The problem of sequential pattern mining has been well studied in the context of deterministic data. It can only examine a combinatorial explosive number of intermediate subsequences. Most of the previously developed sequential pattern mining methods, such as promoting data, explore a candidate generation-and-test approach to reduce the number of candidates to be examined.

However, this approach may not be efficient in mining large sequence databases having numerous patterns and/or long patterns. The low performance and support of the pattern-growth approach may lead to its further extension toward less accuracy mining of other kinds of frequent patterns, such as frequent substructures.

DRAWBACKS OF EXISTING SYSTEM

- The user prospects on the invention method of the mining patterns and also need the background of the user have not been
  thought-about then this lead to high price and very exhausting to affect the mining method.
- Slower performance in speed and space by suggests that of those approaches.
- High in memory use.
- Complex data handling in sequence graphs to manage the temporal constraints while large data mining.

III. PROPOSED SYSTEM

The process starts by cleaning and preparing the data and then applying frequent pattern mining for discovering appliance-to-appliance association., determining which appliances are operating together. Then, it uses cluster analysis to determine appliance to time associations. With these two processes, the system is able to extract the pattern of appliance usage which is then used as input to the Bayesian network for short-term and long-term activities prediction. The output of the system can be used by specific health care applications. The health care provider might only curious about knowing activities associated with cognitive impairment where tracking the sequence of daily activities is crucial for reminding the patient when abnormal behavior is detected.

ADVANTAGES OF PROPOSED SYSTEM

• On single-level projection, since the advantage of bi-level projection may not be significant when the pseudo-projected database is stored in main memory.
• Low in memory usage.
• High in performance and data retrieval latency time.
• It can measure the efficiency of the uncertain stream clustering method.
• The running time of all the algorithms increases almost linear.

IV. MODULES

BASE INFORMATION ANALYSIS:

In the base information analysis module represents We can mine the complete set of frequent itemsets, based on the completeness of patterns to be mined: it can distinguish the following types of frequent itemset mining, given a minimum support threshold the co-efficient , which refers to the variety of items, including first or most significant itemset .the combitorial represents the itemset ‘j’ represents the length of an itemset. If the length of an itemset is 2(j=2) means, it contains 1-itemset and 2-itemset (i=1,2) ‘m’ represents the target itemset length. m=k+1. Here ‘m’ denotes the itemset length that we are going to find the approximate count. (eg., if k=2, m=3) ‘k’ represents the bottom

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

information size. In the base information, if k=2 means, it denotes that, it contains 1-itemset and 2-itemset. represents the ith itemset of jth itemset to use for finding approximation count. As show in the fig 4.1

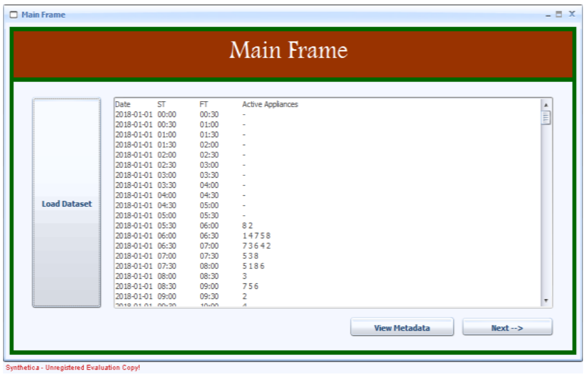

Figure 4.1 Main frame of the database

APPROMIZATION COUNT CALCULATION:

This module is to generate the maximal frequent itemsets with minimum effort. Instead of generating candidates for determining maximal frequent itemsets, this module uses the concept of partitioning the information source into segments then mining the segments for maximal frequent itemsets. Additionally, it reduces the number of scans over the transactional data source to only two. Thus, the time spent for candidate generation is eliminated.

This algorithm involves the following steps to determine from a data source:

1. Segmentation of the transactional data source.
2. Prioritization of the segments.
3. Mining of segments.

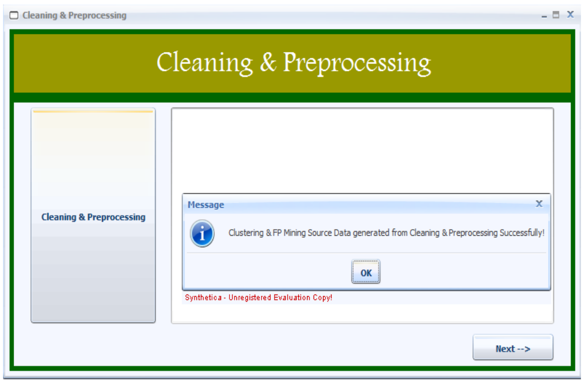

Figure 4.2 Cleaning and preprocessing

FREQUENT ITEMSET LIST GENERATION

In this module the window model is employed . The window should be divided into two sub-windows. The entire window is denoted as 'w' and the sub-windows are 'w0' and 'w1'. The sub-windows should be partitioned dynamically based on the inputs.it can derive all frequent induced subgraphs from both directed and undirected graph structured data having loops (including self-loops) with labeled or unlabeled nodes and links. Its performance is evaluated through the applications to Web browsing pattern analysis and chemical carcinogenesis analysis to avoid the problem of numerous database scans and candidate generate –and-test process.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

The corresponding algorithm is called FP Growth Algorithm. To obtain the information about the database, it requires two scans only. Frequent patterns are mined from the tree structure, since contents of the database are captured in a tree structure. Specifically, Incremental FP-Growthstarts by scanning the database once to find all frequent 1-itemsets. Afterwards, the algorithm makes a ranking table, during which items appear in descending frequency order. As show in the fig 4.3.

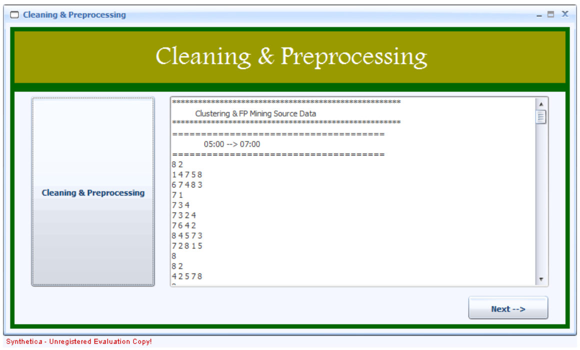

Figure 4.3 Frequent itemset list generation

SKIP AND COMPLETE TECHNIQUE

In this module is to get skip count by dividing the database during a number of non-overlapping segments. After the primary database scan, item set that are frequent locally in each segment are often found. For an item set to be globally frequent within the database, it must be locally frequent item set in a minimum of one partition (or segment). So, after gathering all local frequent item set, the Partition algorithm scans the database for the second and last time to see which of these local frequent item set are actually frequent globally in the whole database.

As a result, this system reduces drastically the amount of scans needed by Apriori-based algorithms to only two. So, Partition algorithm always depends on the info distribution and therefore the number of segments. As the database is scanned, this counter is updated by subtracting the corresponding “over-estimate” for every item within the pattern. If the counter gets below the minimum support, any pattern containing that item can't be frequent and hence are often pruned.

DP with its two improvements is a very effective technique and it improves both runtime and memory requirements of Fp-Growth algorithm. Even though it is still bounded by the generate and test approach limitations, the application of the decremental technique (known as Fp-Growth algorithm) is a reasonable Apriori-based adaptation for uncertain data. As show in the fig 4.4.

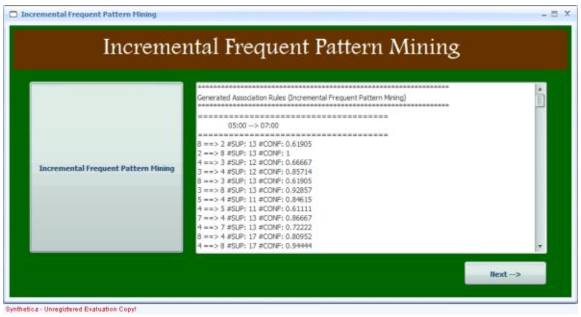

Figure 4.4 Skip and complete technique

GROUP COUNT TECHNIQUE

In this module to generate the data report as Tree Structure. By using this structure, the algorithm tries to enhance the mining time. Once the H-struct (Fp-Growth tree Structure) is constructed, the Incremental FP-Growth

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

algorithm just needs to maintain and update the numerous links that point from one transaction to the next that contains the same set of items.

Since Fp-Growth keeps all transactions that contain frequent items in memory, there is no need to read the database more than once. From that time on, all information is extracted from the H- struct. Incremental FP-Growth out performed Apriori by finding frequent patterns quicker and requiring less memory than Fp-Growth , especially with small minimum support threshold. As show in the fig 4.5.

Figure 4.5 ActivityRecognition

V. CONCLUSION

The prediction model utilizes appliance-to-appliance and appliance-time associations to predict multiple concurrent operating appliances. Based on the above results, it is easy to see the strong relationship between appliance usage inside the smart houses and human activity recognition. Learning the appliance-to-appliance and appliance-to-time associations extracted from the frequent pattern mining and cluster analysis are key processes to trace patients/people's routines and possibly provide them with health services when needed. The model is presented for recognizing human activities patterns from low resolution smart meters data. Occupants habits and behavior follow a pattern that could of time. Thus it is demonstrated that applicability of the proposed model to properly detect multiple appliance usage and make short and future prediction at high accuracy. For future work, it is going to refine the model and introduce distributed learning of massive data processing from multiple houses during a near real-time manner. This will help health applications to promptly take actions like sending aware of patients or care providers. Furthermore, we are getting to build a health ontology model to automatically map discovered appliances to potential activities. This means we will efficiently train the system and increase the accuracy of detecting human activities.

REFERENCES

1. Ahmed C. F., Tanbeer S. K., Jeong B.-S., and Lee Y. -K., "Efficient tree structures for high utility pattern mining in 01incremental databases," IEEE Transactions on Knowledge and Data Engineering, Vol. 21, No. 12, pp. 1708– 1721, 2009.
2. Agrawal R., Imielinski T., and Swami A., "Mining association rules between sets of items in large databases," In Special Interest Group on Knowledge Discovery in Data. Association for Computing Machinery, pp. 207–216, 1993.
3. Anusmitha A., Renjana Ramachandran M., "Utility pattern mining: a concise and lossless representation using up growth", International Journal of Advanced in Computer and Communication Engineering, Vol. 4, No. 7, pp. 451– 457, 2015.
4. Chun-Wei Lin J., Wensheng Gan., Fournier-Viger P., and Yang L., Liu Q., Frnda J., Sevcik L., Voznak M., " High utility itemset-mining and privacy-preserving utility mining," Vol. 7, No. 11, pp. 74–80, 2016.
5. Dawar S., Goya V. l., "UP - Hist tree: An efficient data structure for mining high utility patterns from transaction databases," In Proceedings of the 19th International Database Engineering & Applications Symposium. Association for Computing Machinery, pp. 56–61, 2015.
6. De Bie T., "Maximum entropy models and subjective interestingness: an application to tiles in binary databases," Data Mining and Knowledge Discovery, Vol. 23, No. 3, pp. 407–446, 2011.
7. Erwin A., Gopalan R. P and. Achuthan N. R., "Efficient mining of high utility itemsets from large datasets," In Proceeding of the Pacific-Asia Conference on Knowledge Discovery and Data Mining, pp. 554–561, 2008.
8. Fournier-Viger P., Wu C.-W., Zida S., and Tseng V.S., "Fhm: Faster high-utility itemset mining using estimated utility Co-occurrence pruning," In Proceedings of the 21th International Symposium on Methodologies for Intelligent Systems. Springer, pp.83-92, 2014.
9. Geng L., Hamilton H.J, "Interestingness measures for data mining: A survey," Association for Computing Machinery. Vol. 38, No. 3, pp.1–9, 2006.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 8, Issue 3, March 2020

10. Han J., Pei J., Yin Y., Mao R., "Mining frequent patterns without candidate generation: a frequent-pattern tree approach," Data Mining Knowledge Discovery in Data.Vol. 8, No. 1, pp. 53–87, 2004.
11. Junqiang Liu., Ke Wang., Benjamin., Fung C.M.,"Mining High Utility Patterns in One Phase without Generating Candidates", IEEE Transactions on Knowledge and Data Engineering, Vol. 28, No. 5, pp.1–14, 2016.
12. Jyothi Pillai., Vyas O.P., "Overview of itemset utility mining and its applications," International Journal of Computer Applications, Vol. 5, No. 11, pp. 9 –13, 2010.
13. Liu J., Wang K., and Fung B., "Direct discovery of high utility itemsets without candidate generation," In Proceedings of the 12th International Conference. IEEE, pp. 984–989, 2012.
14. Liu M., Qu J., "Mining high utility itemsets without Candidate generation," Conference on Information and Knowledge Management. Association for Computing Machinery, pp. 55–64, 2012.
15. Liu J., Pan Y., Wang K., and Han J., "Mining frequent item sets by opportunistic projection," In Special Interest Group on Knowledge Discovery and Data Mining. Association for Computing Machinery, pp.229–238, 2002.
16. Liu V., Liao W., and Choudhary A., "A fast high utility itemsets mining algorithm," in utility – Based Data Mining Workshop in Special Interest Group on Knowledge Discovery in Data. Association for Computing Machinery, pp. 253 – 262, 2005.
17. Li Y.-C., Yeh J.-S., and Chang C.-C., "Isolated items discarding Strategy for discovering high utility itemsets," Data & Knowledge Engineering, Vol. 64, No. 1, pp. 198–217, 2008.
18. Sarode, Nutan, and Devendra Gadekar, " A review on efficient algorithms for mining high utility itemsets, "International Journal of Science and Research, Vol. 3, No. 12, pp.708 –710, 2014.
19. Shankar S., Purusothoman T.P, Jayanthi S., Babu N., "A fast algorithm for mining high utility itemsets" ,In Proceedings of IEEE International Advance Computing Conference (IACC), Patiala, India, pp.1459-1464, 2009.
20. Tan P.N., Kumar V., and Srivastava J., "Selecting the right objective measure for association analysis," Information Systems, Vol. 29, No. 4, pp. 293–313, 2004.
21. Tseng V. S., Shie B.-E., Wu C.-Wand Yu P. S., "Efficient algorithms for mining high utility itemsets from transactional databases," IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 8, pp. 1772–1786, 2013. .,
22. Yao H., Hamilton H. J., Butz C.J., "A foundational approach to mining itemset utilities from databases," ICDM 2004, pp. 482-486.
23. Yao H., Hamilton H. J., Geng L., "A unified framework for utility-based measures for mining itemsets," in Utility-Based Data Mining," In Special Interest Group on Knowledge Discovery in Data. Association for Computing Machinery, pp. 28–37, 2006.
24. Zaki M.J., "Scalable algorithms for association mining," IEEE Transactions on Knowledge and Data Engineering, Vol. 12, No. 3, pp. 372–390, 2000.