IJITKM Volume 6 • Number 2 • June 2013 pp. 175-180 (ISSN 0973-4414)

Discovery and Ranking of Outliers using Weighted Association Rule Mining with Clustering

Deepak Sinwar

Department of Computer Science& Engineering
BRCM College of Engineering & Technology, Bahal, Bhiwani, Haryana
sdeepak.sinwar@gmail.com

Abstract—*Data mining is an important issue in the development of knowledge-based systems. To ensure the correctness of data mining results, outlier mining is absolutely necessary. Discovery and Ranking of Outliers using Weighted Association Rule Mining with clustering aims to discover hidden as well as interesting patterns which are different from others, we firstly detect them and then a ranking will be provided to them on the basis of some criterion. This paper introduces a new approach towards outlier mining as a two step process. The former selects patterns on the basis of weighted rules, whereas the later one mines these patterns using clustering. The process will then be continued towards providing a rank to the outliers mined using the above two step process. The benefit of this approach is that it will be easier to optimize the profit of an organization by taking effective decisions based on these outliers. As we know that the decisions based on outliers will reflect the optimized performance definitely.*

I. INTRODUCTION

Outlier is defined as an observation that deviates too much from other observations. The identification of outliers can lead to the discovery of useful and meaningful knowledge and has a number of practical applications in different areas. Now days, it seems very important to detect as well as to rank outliers from different types of data for finding out their severities/ influences/ benefits. Because the handling of anomalous or outlying observations in a data set is one of the most important tasks in data pre-processing. Once outliers have been detected it may be either retained or rejected. In order to successfully distinguish between noisy outlying data and noise free outliers, different kinds of information are normally needed. These should not only include various data characteristics and the context in which the outliers occur, but also includes relevant domain knowledge. Consider a supermarket scenario, containing a variety of items which fall into different categories like food, home appliances, cloths etc. We will use the terms items and attributes interchangeably throughout this paper. Obviously every supermarket manager will be interested in enhancing their profit. The profit may be enhanced by increasing sales, reduction in wastages or from any means. Normally a market manager tries to increase their sales of all available items, for gaining higher profit, but most of them are unaware of the significance or weights of different products which are available in the store. They may take effective business decisions to gain high profit, from increasing sales of some selected items based on their weights/significance. In other words, we can say that, beyond the traditional fashion of increasing profit by increasing sales of all items, one another kind of approach exists called, 'selective profit maximizing approach' which mainly consists sale of those items which are highly profitable and some of them whose selling frequency is so higher than others. We can say in this approach we try of maximize the profit using this approach. For example sale of 10g gold gives us more profit than 10g copper or silver. So that, we can classify this approach into two different types as follows: (i) Lesser sale frequencies with high profit; (ii) More frequently sale with low profit. Typical business decisions that the management of supermarket makes includes are: what to put together, what to put on front, how to design store layout, and what promotional schemes to consider etc. So to make all these decisions in addition of analysis of past as well as current data a profit maximizing strategy is also necessary. For example, in addition with monthly, quarterly or yearly sales analysis, a profit maximizing strategy is necessary. It is an association rule discovering mechanism which makes organizations to take informal business decisions, which are helpful to maximize the profit.

The proposed research work use Weighted Association Rule Mining as a building block and also combines a clustering technique towards discovery and ranking of outliers. The output of clustering scheme will be refined by selecting only those outliers which matches some specified criterion. Finally the selected outliers will be provided a rank to signify the severity of the discovered outliers in terms of benefits or influences.

Let us take a brief review of the techniques that will be used in the proposed work i.e. ARM, WARM and Outliers as follows:

II. BACKGROUND WORK

A In the field of Data Mining, there have been many studies on Frequent Pattern Mining and Association Rule Mining to discover new and interesting patterns and correlations. As we know that Association Rule Mining (ARM) is a well established data mining technique used for the extraction of hidden and interesting patterns called Association Rules. ARM works well by considering uniform importance of items. But in real life scenario different items plays different role in our daily life. So now days, it is seems necessary to introduce such ARM algorithms which understands the importance (in terms of weights) of items while mining day to day market-basket transaction records. This type of ARM is called Weighted Association Rule Mining (WARM). Plenty of work has been done for mining simple as well as weighted association rule mining, but the most of work of traditional ARM focuses on the well established ARM method i.e., Apriori developed by Agrawal et al [1]. On the other hand different researchers adopt different scenarios for Weighted Association Rule mining.

A. *Traditional Association Rule Mining*

Generally ARM was primarily proposed for market basket analysis to understand consumer purchasing patterns in retailing industry [1]. Most of the research in this area during that decade moves around the so called Apriori algorithm developed by Agrawal and Srikant in 1994 at 'IBM Almaden research center'. But due to some limitations of Apriori algorithm like needs several iterations over the data, use of uniform minimum support threshold etc, there was a need to generate such efficient ARM method which saves time by reducing the number of passes over the database as well as without the help of candidate generation. So in this way many researchers have proposed different variations of the Apriori method. Based on the downward closure property (if an itemset is not frequent, its subsets cannot be frequent; and vice versa), if one nonempty subset of an itemset in not frequent, the itemset is not frequent, there is no mean to take it as a new candidate. A new IApriori method of this kind was also developed, that can reduce the times of scanning database, optimize the join procedure in order to reduce the size of the candidate itemsets have been discussed in [3]. In order to improve the ARM without candidate generation, a new method of Frequent Pattern growth has been developed by Han et al., in [4]. There are many reasons behind the popularity of this method like there is no need for candidate generation; only two passes over the database are required, and it constructs FP-tree which is compressed highly to reduce the original database size etc. Li-Juan Huang [5] developed a combined approach called "FP-Grwoth Apriori" which that will help resolve two neck-bottle problems (e.g. heavy I/O loading and huge candidate sets) of traditional Apriori algorithm and has more efficiency than original one.

A combination of both Apriori algorithm and FP-growth has been proposed in many research publications. One of them is a new method Apriori-Growth has been presented in very effective manner by Wu et al., 2008 in [6]. Using computational results shown by them, we can easily see the performance of Apriori-Growth is much faster than Apriori algorithm, and it is almost as fast as FP-Growth [4], but it needs smaller memory. Because ordinary association rule mining algorithms considers same function of each attribute, they cannot effectively discover the association rules which can reveal cause and effects relationships between attributes. Secondly another combination of Apriori and FP-growth is merged in APFT algorithm [8]. The advantage of APFT (Apriori FP-Tree) is that it does not need to generate conditional pattern bases and sub-conditional pattern tree recursively. They consider using the Apriori method to mining the frequent itemsets basing on the FP-tree, the divide-and-conquer strategy is still adopted by mining process. It means that the compressed FP-tree is partitioned off a set of conditional subtree, each of the conditional subtree associated with a frequent item. If there are n 1-frequent items Ii(i=1,2,….n), then the FP-tree can be divided into n conditional subtree FPTi (i=1,2,….n) , and FPTi is the conditional subtree associating with frequent item Ii .Then use the Apriori algorithm to mine each conditional subtree, and gain all the frequent itemsets with the first prefix item Ii. The AFPT algorithm includes two steps, the first step is to construct the FPtree as FP-growth does, the second step is to use of the Apriori algorithm to mine the FP-tree.

B. *Weighted Association Rule Mining*

The principle of weighted association rule mining was first given by Cai et al., in [9]. Their aim is to produce such association rule mining algorithm which does not take all items of a basket database uniformly. They generalize this case where items are given weights to reflect their importance to user. The weights may correspond to special promotions on some products, or the profitability of different items They propose two new algorithms to mine weighted binary association rules namely MINWAL(O) and MINWAL(W) which uses a metric called the k-support bound. Usually MINWAL(O) and MINWAL(W) algorithms are based on Apriori algorithm. With the help of experiment they shows that these algorithms have reasonable performance for large databases and MINWAL(O) performs better than MINWAL(W) in most cases. So the problem of invalidation of the "downward closure property" in the weighted setting is solved by and improved model of weighted support measurements and exploiting a "weighted downward closure property" by Tao et al. [10]. The concepts of 'weighting attributes', 'Item weight', 'Itemset weight' and 'Transaction weights' are used in an improved model i.e. WARM (Weighed Association Rule Mining); where

- ***Weighting attributes*** are variables selected to calculate weights
- ***Item weight*** is a value attached to an items representing its significance
- I***temset weight*** can be derived from the weights of its enclosing items, one simple way is to calculate the average value of the item weights
- ***Transaction weight*** is a type of itemset weight in which a value is attached to each of the transactions
- W***eighting space(WS)*** is the context within which the weights are evaluated, it comprises of:
  Inner-transaction space WSt: this space refers to the host transaction that an item is weighted in.

Item space WSI : this space refers to the space of the item collection that covers all the items appears in the transaction.

Transaction space WST: this space is defined for transactions rather than for items.

- ***Weighted support (WSP):*** A set of transactions T respects a rule R in the form of A B, where A and B are non-empty sub-itemsets of the item space I and they share no item in common. Its weighted support is the fraction of the weight of the transactions that contains both A and B relative to the weight of all transactions. This can be formulated as:

$$wsp(AB) = \frac{\sum_{k=1}^{|WS_T|\&(A\cup B)\subset t_k} weight(t_k)}{\sum_{k=1}^{|WS_T|} weight(t_k)}$$

Where,

$$weight(t_k) = \frac{\sum_{i=1}^{|WS_t(t_k)|} weight(item(i))}{|WS_t(t_k)|}$$

By this means, weighted support is modelled to quantify the actual quota of an itemset in the transaction space in weighted association rule mining scenario. We know that it is very difficult to retain downward closure property while mining weighted association rules. For the first time ever, weighted downward closure property has been retained in this method with the concept of weighted support defined above.

One another approach of this kind has been proposed by Yun and Leggett [11] named WFIM: Weighted Frequent Itemset Mining using global and conditional FP-tree, which generates concise and important weighted frequent itemsets in large databases, particularly dense databases with low minimum support, by adjusting a minimum weight and a weight range. The main focus of weighted frequent itemset mining is concerned with satisfying the downward closure property.

C. Outlier Detection

Outlier is defined as an observation that deviates too much from other observations. The identification of outliers can lead to the discovery of useful and meaningful knowledge and has a number of practical applications in areas such as transportation, ecology, public safety, public health, climatology, and location based services [12]. Jingke Xi has presented outliers into two categories viz. classic outlier approach and spatial outlier approach. The classic outlier approach analyzes outlier based on transaction dataset, which can be grouped into statistical-based approach, distance-based approach, deviation-based approach, density-based approach. The spatial outlier approach analyzes outlier based on spatial dataset, which can be grouped into space based approach, graph-based approach. Thirdly, they conclude some advances in outlier detection recently. But our aim is also to provide a rank to the detected outliers, in this regard Muller et. Al. 2008 proposed OutRank approach for ranking outliers in heterogeneous high dimensional data [13]. They introduce a consistent model for different attribute types. Novel scoring functions transform the analyzed structure of the data to a meaningful ranking. Promising results in preliminary experiments show the potential for successful outlier ranking in high dimensional data. As opposed to existing approaches, OutRank approach handles heterogeneous data, i.e. both continuous and categorical attributes, of high dimensionality. As we know that the clustering is one of the famous techniques towards outliers' discovery. Jiang et al. 2009 they generalize local outlier factor of object and propose a framework of clustering based outlier detection. Another approach of this kind was also developed by Jiang [7], based on outlier factor of cluster, a clustering-based outlier detection method, named CBOD. The method consists of two stages, the first stage cluster dataset by one-pass clustering algorithm and second stage determine outlier cluster by outlier factor. The time complexity of CBOD is nearly linear with the size of dataset and the number of attributes, which results in good scalability and adapts to large dataset. The theoretic analysis and the experimental results show that the detection method is effective and practicable. Zhang [14, 15] also proposed a novel approach to detect outliers based on clustering. After analyzing current detection technologies, a detection method of outlier based on clustering analysis is proposed, in which an effective sample is screened out from original data. According to agglomerative of hierarchical clustering, credible sample set is found. Then mathematical expectation and standard deviation are obtained by credible sample. Finally, global data will detected by the definition of outlier which is proposed in this paper. The data disposed by this method can be irrelative to the time scales. And it needs not to presuppose the number of outlier. The experiment result on IRIS shows that this method can detect outliers effectively. Yoon and Bae [16] proposed a new kind of approach i.e. a pattern-based outlier detection method that identifies abnormal attributes in software project data: after discovering the reliable frequent patterns that reflect the typical characteristics of the software project data, outliers and their abnormal attributes are detected by matching the software project data with those patterns. Empirical studies were performed on three industrial data sets and 48 artificial data sets with injected outliers. The results demonstrate that our approach outperforms five other approaches by an average of 35.27% and 107.5% in detecting the outliers and abnormal attributes, respectively, on the industrial data sets, and an average of 35.44% and 46.57%, respectively on the artificial data sets. But James and Dimitrijev [17] considered that the pixels associated with large distances obtained by inter-image pixel-by-pixels comparisons should be considered as inter-image outliers and should be removed from the similarity calculation used for the image classification.

Whereas a density-similarity-neighbor-based outlier mining algorithm [18] proposed by Cao et. al. 2010. First, the concept of k-density of an object is presented and the similar density series (SDS) of the object is established based on the changes of the k-density and the neighbors k-densities of the object. Second, the average series cost (ASC) of the object is obtained based on the weighted sum of the distance between the two adjacent objects in SDS of the object. Finally, the density-similarity neighbor- based outlier factor (DSNOF) of the object is calculated by using both the ASC of the object and the ASC of k-distance neighbors of the object, and the degree of the object being an outlier is indicated by the DSNOF. The experiments are performed on synthetic and real datasets to evaluate the effectiveness and the performance of the proposed algorithm. The experiments results verify that the proposed algorithm has higher quality of outlier mining and do not increase the algorithm complexity.

Outlier detection in streaming data is a very challenging problem. This is because of the fact that data streams cannot be scanned multiple times. Also new concepts may keep evolving. Irrelevant attributes can be termed as noisy attributes and such attributes further magnify the challenge of working with data streams [19]. Yogitaa and Toshniwala 2012 proposed a clustering based framework for outlier detection in evolving data streams that assigns weights to attributes depending upon their respective relevance. Weighted attributes are helpful to reduce or remove the effect of noisy attributes in mining tasks. Keeping in view the challenges of data stream mining, the proposed framework is incremental and adaptive to concept evolution. Experimental results on synthetic and real world data sets show that our proposed approach outperforms other existing approaches in terms of outlier detection rate, false alarm rate, running time and with increasing percentages of outliers. Other approaches to detect outliers can be found in [2, 20, 21, 22, 23, 24, 25 and 26], which are not based on rule mining approaches, but provides significant amount of outliers in the data mining industry.

III. Proposed Work

The main objectives of the proposed research work include:

1) Detection of outlying observations using WARM and Clustering.

2) Ranking of outlying observations on the basis of implications/ benefits.

Plenty of work has been done in the area of outlier detection and ranking of outliers, but most of the work has been done for the large databases like spatial databases, and these methods does not includes: Concept of weighting attributes and Combined approach (Association Rule Mining with Clustering). But this proposed work includes the concept of special weighting attribute (s), which can be any attribute specified; and a combined approach which will be outlined in Methodology section.

Methodology

The proposed methodology will combine two approaches firstly weighted frequent item set mining and secondly clustering these weighted frequent item sets for outlier detection. Lets us outline the main steps of the proposed methodology as follows:

Step1: Input a transaction dataset

Step2: Apply weights to the input dataset by weighting attributes

Step3: Perform a Weighted Association Rule Mining approach (preferably Weighted Frequent Pattern Growth) for finding out the weighted frequent item sets among the input dataset.

Step4: Apply clustering to the item sets discovered in step 3 for finding out the outliers among the data if number of weighted frequent item sets is two or more, otherwise declare the item (if any) as outlier and go to step6.

Step5: Apply clustering to those items found eligible in step4 for outlier detections.

Step6: Refine the discovered outliers upon some characteristics so that they can be ranked.

Step7: Test the hypothesis about to select or reject the refined outliers.

Step8. End with a list of selected outliers (if any) with their respective rank(s).

IV. Proposed Methodology As Two Step Outlier Detection Work

The proposed outlier detection algorithm will be divided into two steps as a whole; the former generates frequent patterns using Weighted FP-Growth [22] and the later clusters weighted frequent patterns generated by former one. Let us discuss these two steps in brief:

A. *Weighted FP-Growth*

A new type of Weighted Association Rule Mining algorithm [22] has been designed to produce weighted association rules by making some minimal changes to the Frequent Pattern Growth algorithm [4] using WEKA data mining tool (freely available at [23]) . The main objective is to extract such items which are sometimes infrequent in nature but are more significant from others.

Algorithm: WFP (Weighted FP-Growth) [22]

Input: A transaction dataset D, min_wt (minimum weight of an item), min_sup (the minimum support), min_conf (the minimum confidence threshold),

Output: A list of weighted frequent item sets

Method:

1. Firstly, scan the data set once to find out the support count (m) of each item a of transaction dataset.

2. Build the frequent pattern tree by inserting only those items of a transaction which satisfies both conditions:
   (a) If weight of that item is greater than minimum weight threshold i.e.,
   Weight(Item i) > min_wt
   (b) Support count of that item is greater than minimum support threshold i. e.,
   Supp_count(Item i) > min_sup
3. The FP-Tree is mined by calling FP-growth(FP-tree, null) [11].
4. End with a list of weighted frequent item sets.

The significance is based on the percentage of profit earned per unit of sale of that item or any other criterion as specified. With the help of weight vector, it is possible for customers/users to supply weights according to his/him choice, thus making WFP more effective than other WARM algorithms. WFP works on the principle of building Frequent-Pattern tree by firstly pruning/eliminating those items from the transaction database D whose weight is below the minimum weight threshold and secondly inserting these items as well as those having more support count than the minimum support threshold to the FP-tree.

Objective of WFP: The objective to introduce this proposed algorithm is to reveal the importance or significance of items. This work will use the concept of special weighting attributes (for example, Cost/Price, Profit, Selling Frequency, General Importance etc.). WFP will generate weighted frequent item sets; so that the discovered outliers are already refined. The main issues that will be considered are:

1. How weights are effective then support counts
2. Conciseness of results
3. Detection and Ranking of outliers

Distribution of Weights: Weights to attributes are distributed in a variety of ways to produce weighted association rules. The proposed work will provide functionality to the customer to provide item weights on his/ her own behalf. Weighting attributes can be many ranged from 1 to n. In addition to this scheme, we can also provide transaction weights in case of absence of weighting attributes. Let us study some sample cases to provide weights to the items as well as transactions:

1) According to support counts (flat weights)
2) General importance of the item towards customer satisfaction
3) All flat weighted except for one attribute
4) On the basis of real item prices and unique profit percentage
5) Credibility of the product being sold out.
6) Highly profitable transaction etc.
7) Any other method.

Architecture of Building Weighted FP-Tree

The WFP algorithm works on the principle of building the frequent pattern tree by inserting only those items to the tree which satisfies both the following conditions:

a) If weight of that item is greater than minimum weight threshold i.e.,
Weight(Item i) > min_wt and
b) Support count of that item is also greater than minimum support threshold i. e.,
Supp_count(Item i) > min_sup

B. Clustering

Clustering is the task of assigning a set of objects into groups (called clusters) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. A cluster is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters.

Outliers Detection: In statistics, an outlier is an observation that is numerically distant from the rest of the data. Grubbs defined an outlier as: An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs. Outliers can occur by chance in any distribution, but they are often indicative either of measurement error or that the population has a heavy-tailed distribution. In the former case one wishes to discard them or use statistics that are robust to outliers, while in the latter case they indicate that the distribution has high kurtosis and that one should be very cautious in using tools or intuitions that assume a normal distribution.

Causes: Outliers can have many anomalous causes. A physical apparatus for taking measurements may have suffered a transient malfunction. There may have been an error in data transmission or transcription. Outliers arise due to changes in system behaviour, fraudulent behaviour, human error, instrument error or simply through natural deviations in populations. A sample may have been contaminated with elements from outside the population being examined. Alternatively, an outlier could be the result of a flaw in the assumed theory, calling for further investigation by the researcher. Additionally, the pathological appearance of outliers of a certain form appears in a variety of datasets, indicating that the causative mechanism for the data might differ at the extreme end (King effect).

Identifying Outliers: There is no rigid mathematical definition of what constitutes an outlier; determining whether or not an observation is an outlier is ultimately a subjective exercise. Outlier detection has been used for centuries to detect and, where appropriate, remove anomalous observations from data. Outlier detection can identify system faults and fraud before they escalate with potentially catastrophic consequences. The original outlier detection methods were arbitrary but now, principled and systematic techniques are used, drawn from the full gamut of computer science and statistics.

There are three fundamental approaches to the problem of outlier detection:
Type 1 - Determine the outliers with no prior knowledge of the data. This is essentially a learning approach analogous to unsupervised clustering. The approach processes the data as a static distribution, pinpoints the most remote points, and flags them as potential outliers.

Type 2 - Model both normality and abnormality. This approach is analogous to supervised classification and requires pre-labelled data, tagged as normal or abnormal.

Type 3 - Model only normality (or in a few cases model abnormality). This is analogous to a semi-supervised recognition or detection task. It may be considered semi-supervised as the normal class is taught but the algorithm learns to recognize abnormality.

Ranking Outliers: This proposed work includes ranking of outliers as one of the final step, in this step we will provide ranks to the different outliers discovered so far in the previous steps. The final ranking will decide the severity of the outliers. For validating this new proposed method, the output of this method will be compared against already developed algorithms of outlier detection and ranking.

V. CONCLUSION

This paper presents a new approach to discover as well as rank outliers using two methodologies viz. Association Rule Mining and Clustering. Plenty of work has been done in the area of mining association rules and outliers; but we found no such technique which will combine the both methods. This new approach will definitely get benefit of both methodologies. The work may be extended to simulate the proposed work into real environment for validation of the new approach.

REFERENCES

[1] R. Agrawal, T. Imielinski and A. Swami, "Mining Association Rules betweek Sets of Items in Large Databases", in: Proceedings of the ACM SIGMOD Conference, Washington DC, USA, 1993.
[2] [1] O. Alan and C. Catal, "Thresholds based outlier detection approach for mining class outliers: An empirical case study on software measurement datasets", Expert Systems with Applications, Vol. 38, 2011, 3440–3445
[3] Y. Xie, Y. Li, C. Wang and M. Lu, "The Optimization and Improvement of the Apriori Algorithm", in: Proceedings of International Smposium on Intelligent Technology Application Workshops", Shanghai, 2008, pp. 1101-1103.
[4] J Han , J. Pei and Y. Yin ," Mining frequent patterns without candidate generation", in proc. Intl. conference on Management of Data, 2000, pp. 1-12
[5] L. Huang, "FP-growth Apriori Algorithm's Applicatoins in The Design For Individualized Virtual Shop On The Internet", in: Proceedings of sixth International Conference on Machine Learning and Cybernatics, Hong Kong, 2007, pp. 3800-3804.
[6] B. Wu, D. Zhang, Q. Lan and J. Zheng, "An Efficient Frequent Patterns Mining Algorithm based on Apriori Algorithm and the FP-tree Structure", in: Proceedings of Third IEEE International Conference on Convergence and Hybrid Information Technology, Busan, Vol. 1, 2008, pp. 1099-1102.
[7] S. Y. Jiang and Q. B. An, "Clustering-Based Outlier Detection Method", Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2009, pp. 429-433
[8] Q. Lan, D. Zhang and B. Wu, "A New Algorithm For Frequent Itemsets Mining Based on Apriori and FP-Tree", proc: WRI Global Congress on Intelligent Systems, Xiamen, 2009, pp. 360-364.
[9] C. H. Cai, A. W. C. Fu, C. H. Cheng and W. W. Kwong, "Mining Association Rules with Weighted Items", in: Proceedings of International Database Engineering and Applications Symposium, Cardiff, Wales, 1998, pp. 68-77.
[10] F. Tao, F. Murtagh and M. Farid, "Weighted Association Rule Mining using weighted support and signigificance framework", in: Proceedings of 9th ACM-SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining, Washington, D.C., 2003, pp. 661-666.
[11] U. Yun, J. J. Leggett, "WFIM: Weighted Frequent Itemset Mining with a weight range and a minimum weight", in: Proceedings of International conf. on Data Mining, SIAM, 2005, pp. 636-640.
[12] J. Xi, "Outlier Detection Algorithms in Data Mining", IEEE IITA, 2008, PP. 94-97
[13] E. Muller, I. Assent, U. Steinhausen, T. Seidl, "OutRank: ranking outliers in high dimensional data", IEEE ICDE Workshop, 2008, pp. 600-603
[14] Y. Zhang, J. Liu, H. Li, "An Outlier Detection Algorithm based on Clustering Analysis" First International Conference on Pervasive Computing, Signal Processing and Applications, 2010, PP. 1126-1128
[15] W. Zhang, H. Liao and N. Zhao, "Research on the FP-Growth Algorithm about Association Rule Mining", in: Proceedings of 2008 International Seminar on Business and Information Management", Wuhan, 2008, pp. 315-318.
[16] K.A Yoon, D.H. Bae, "A pattern-based outlier detection method identifying abnormal attributes in software project data", Information and Software Technology, Vol. 52, 2010, pp. 137–151
[17] A. P. James and S. Dimitrijev, "Inter image outliers and their application to image classification", Pattern Recognition, Vol. 43, 2010, pp. 4101–4112
[18] H. Cao, G. Si, Y. Zhang and L. Jia, "Enhancing effectiveness of density-based outlier mining scheme with density-similarity-neighbor-based outlier factor", Expert Systems with Applications, Vol. 37, 2010, pp. 8090–8101
[19] Yogitaa and D. Toshniwala, "A Framework for Outlier Detection in Evolving Data Streams by Weighting Attributes in Clustering", Procedia Technology, Vol. 6, 2012, pp. 214 – 222
[20] G. Bruno and P. Garza, "TOD: Temporal outlier detection by using quasi-functional temporal dependencies" Data & Knowledge Engineering, Vol. 69, 2010, pp. 619–639
[21] B. Huang and P. Yang, "Finding key knowledge attribute subspace of outliers in high-dimensional dataset", Expert Systems with Applications, Vol. 38, 2011, pp. 10147–10152
[22] D. Sinwar, Saroj, K. Kumar, "Discovering Concise Weighted Association Rules", In proceedings of "International Conference on Emerging Trends in Engineering & Technology (IETET2010)" Kurukshetra- Haryana (INDIA), Oct 14-16 2010
[23] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); the WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.
[24] F. Jiang, Y. Sui and C. Cao, "A hybrid approach to outlier detection based on boundary region", Pattern Recognition Letters, Vol. 32, 2011, pp. 1860–1870.
[25] S. Y. Jiang and A. M. Yang, "Framework of Clustering-Based Outlier Detection", Sixth International Conference on Fuzzy Systems and Knowledge Discovery, 2009, pp. 475-479S. Kim a, N. Cho, B. Kang and S. Kang, "Fast outlier detection for very large log data", Expert Systems with Applications, Vol. 38, 2011, pp. 9587–9596.