Volume 1, Issue 10, pp. 69-73, 2017.
International Journal of Scientific Engineering and Science
ISSN (Online): 2456-7361

An Algorithm for Interesting Negated Itemsets for Negative Association Rules from XML Stream Data

Juryon Paik
Department of Digital Information & Statistics, Pyeongtaek University, Pyeongtaek-si, S.Korea

***Abstract*—*This Upcoming IoT revolution changes completely around the world transforming everyday physical objects into innovation technologies. With the advances of sensor network, billions of objects communicate wirelessly, which generates gigantic volumes of stream data. It highly demands knowledge discovery methods to extract valuable information. Association rules mining is one of well-known methods to analyze, manage, and leverage huge volumes of data. Researchers are devoted to elaborating solutions for the stream data. However, they are faced with several problems, such as real-time processing, complex data structure, heterogeneous environments, and so on. In this paper, we consider discovery of meaningful but negated tree-structured itemsets for negative association rules from the famous tree-structured data type, XML, which is not just xml but streaming xml data. We provide a frame of algorithm with several measuring factors.*

***Keywords*—** *Negated tree-structured item; negative association rule:xml stream data.*

I. Introduction

Over the past few years, IoT has been widely paid attention and greatly improved by many researchers and experts. Equipped with cutting-edge technologies such as embedded sensor devices and big data management, sensor-rich smart environments inevitably produce tremendous volume of data. This brings numerous challenges with sensor data, because the deployment also continuously generates gigantic amounts of priceless data. If the raw data from sensors is analyzed properly and evaluated effectively through data mining approach, it can be definitely converted to predictive insights. When the researchers tried to apply traditional mining technologies to the data transferred from sensors, they were faced with serious problems such as gigantic size, lossy information, limited resources, and high failure rate. It was too expensive and centralized to use directly the traditional methods.

The mining streaming data clearly focuses on the data analytic side. Various methods to extract actionable insights have been devised and applied to the stream data such as frequent pattern mining, classification, clustering, prediction, and outlier detection [1]. Association rules mining is one of the important techniques based on the frequent pattern mining, which is a popular and well researched method for discovering interesting relations between variables in large databases. Over the last years, the association rules mining becomes indispensable to exploit streaming data and support many real applications. However, association rules mining on xml is another challenging issue. The flexibility, heterogeneity, and exchangeability of xml come from its structure. It has tree structure which provides sufficient expressive power to describe any types of data. Because of characteristics of xml tree such as hierarchy and labels, traditional association rule mining approaches cannot be applied directly to xml data. Managing continuously arriving xml data is expensive and complicated task, and many of the problems it presents have yet to be adequately solved. In this paper, we consider and describe obtaining negated tree-structured itemsets for negative association rules mining from xml stream data. Discovery of negative association rules are considered importantly nowadays due to its effective usage and impact.

II. Related Work

The approach of association rules from stream data did not exist before 2000's. With the rapidly increasing sensor network deployments and the ability to generate large volumes of data in current Internet of Things (IoT) infrastructure, researchers have burst into finding association rules from the stream data. Among the early studies, Loo et al. [2] proposed a framework for discovery of association rules from sensor networks. In their approach, values of sensors are treated importantly in mining association rules and the time is divided into intervals. Then, the values with the intervals are formed as transactions which are stored in the database. Afterwards, the authors adapt the lossy counting algorithm [3] for one-pass analysis of the data. The lossy counting algorithm is used to identify such elements that their frequency-counts exceed a user-given threshold in a data stream.

The researchers Boukerche and Samarah proposed the positional lexicographic tree (PLT) structure to discover association rules from wireless ad-hoc sensor data [4]. PLT follows the famous pattern growth technique, FP-growth approach [5]. The target objects of PLT algorithm are those of event detecting sensors, and the sensors are within same time interval no matter what their values are. Based on the epochs, PLT structure is separated into a set of partitions storing positional vectors. In the corresponding epoch's pattern, the positional vector has the same sensor as the last sensor. Each partition is then indexed by the rank of the last sensor in its vector. Mining PLT is begun with the sensor whose rank is maximum by preparing the conditional PLT in a recursive way. Like other pattern growth based algorithms, PLT can also load required partitions into the main memory. However,

http://ijses.com/

International Journal of Scientific Engineering and Science
Volume 1, Issue 10, pp. 69-73, 2017. ISSN (Online): 2456-7361

its computational cost limits the efficient use in managing stream data, because PLT requires two database scans as well as additional PLT update operations.

Corpinar and Güundem suggest a rule mining system that provides solution to positive and negative association rules computation [6]. However, their type of data stream is different to that of other approaches. The data is XML data stream. Due to its powerful flexibility and simple handling, XML is the standard language for describing and delivering data on the Web. Discovering association rules from XML data is, however, another challenging issue, because of its complicated tree structure. Actually, extracting information from XML world still requires further researches and studies in diverse areas. Under such circumstance, they first adapted the original FP-Growth method to support stream data mining and negative rules. In order to decrease the search space for negative association rules they devised new pruning thresholds along with adding correlation coefficient parameter into their methodology.

The recently published paper [7] presents several new definitions and scheme related to association rule mining over xml data streams in wireless sensor networks. The authors' proposed scheme is the first approach to mining association rules from xml stream data in the sense that it generates frequent tree items without any redundancy. The overall methodology can be applied to any individual block, as well as the whole stream. In this paper, we focus on the problem of extracting informative tree-structured itemsets for negative association rules from stream xml data. This feature enables their scheme to discover frequent tree items better than previously published other schemes.

In this article we focus the problem of extracting informative negative association rules over xml stream data. Particularly, we consider how to efficiently prune negated tree items and obtain useful negated tree itemsets because negative association rules are built from a huge number of infrequent and negated tree items. We mainly discuss two major methods, *Interestingness* vs. *Correlation Coefficient*, for the pruning phase.

III. NEGATIVE ASSOCIATION RULES FROM XML STREAM DATA

A. Negative Association Rule Definition

In the form of a negative association rule, the positive association rule *bread* ⇒ *milk* can be expressed by one of the followings: 1) *bread* ⇒ ¬*milk* implies the customers who buy bread usually do not purchase milk. 2) ¬*bread* ⇒ *milk*, the customers who do not buy bread usually purchase milk. 3) ¬*bread* ⇒ ¬*milk*, the customers who do not buy bread usually do not purchase milk, either. This paper considers the form *bread* ⇒ ¬*milk*, for convenience. The definition of a negative association rule is almost like that of a positive association rule, except that the negated itemset implies the occurrence of some itemsets characterized by the absence of others [8]. Because negative association rules encapsulate the relationship between the occurrences of one set of items with the absence of the other set of items, the support and confidence of the rule $X \Rightarrow Y$ must count non-existing items in transactions. However, it is difficult to count the absent items. Based on the paper [7], the barometers of strength and reliability, support and confidence, of a $X \Rightarrow \neg Y$ are defined as:

$$freq(\neg X, D) = 1 - freq(X, D), \quad (1)$$

$$support(X \Rightarrow \neg Y) = freq(X \cup \neg Y, D) = freq(X, D) - freq(X \cup Y, D), \quad (2)$$

$$confidence(X \Rightarrow \neg Y) = \frac{freq(X \cup \neg Y, D)}{freq(X, D)} = \frac{freq(X, D) - freq(X \cup Y, D)}{freq(X, D)} = 1 - confidence(X \Rightarrow Y), \quad (3)$$

where D is a transaction database, X and Y are itemsets, and the function *freq* computes the ratio of transactions containing a given itemset in D. While the barometer *support* is used to compute a ratio of transactions containing all items of head and body, *confidence* is a ratio of the transactions containing not only items of head but also items of body. As expressed in the equations, support and confidence of the negative association rules also consider absence of items not only the existence.

B. Barometers for XML Stream Data

An XML document is modelled as a tree, with nodes corresponding to elements and attributes [9]. Streaming XML data is a series of trees because XML stream data is a continuous sequence of XML documents. According to the paper [7], the counterparts of record and item are *fraction* and *titem*, respectively. For detailed explanation of some definitions, please refer the paper. The following is an equation of the length of xml stream.

$$|XDS| = \sum_{i=1}^{n} |XB_i| = |XB_1| + |XB_2| + \cdots + |XB_n|$$
$$= \left|\sum_{j=1}^{k_1} T_j\right| + \left|\sum_{j=1}^{k_2} T_j\right| + \cdots + \left|\sum_{j=1}^{k_n} T_j\right| = \left|\sum_{i=1}^{n}\sum_{j=1}^{k_i} T_{ij}\right|, \quad (4)$$

where XDS is a given stream-data arriving as n blocks and every T_{ij} is a transaction or document as tree structure. When F is a set of fractions collected from the all blocks, the entire fractions for the given xml stream data can be expressed as $F = \{F_{j,k}^i \mid F_{j,k}^i \preccurlyeq T_{i,j}\}$, where $1 \le i \le n$, $1 \le j \le |T_i|$ and $1 \le k$. Once fractions are collected from the original data, each one of fractions is eligible to be titems. The function *freq* must be applied the length of xml stream, the equation (4), because a frequency of titem set X is a total number of xml documents satisfying the following equation:

$$freq(X, XDS) = |\{T_{k_1} | (X \subseteq T_{k_1}) \wedge (T_{k_1} \in XB_1)\}| + |\{T_{k_2} | (X \subseteq T_{k_2}) \wedge (T_{k_2} \in XB_2)\}| + \cdots + |\{T_{k_n} | (X \subseteq T_{k_n}) \wedge (T_{k_n} \in XB_n)\}|. \quad (5)$$

The given X is frequently occurred and mainly considered for finding association rules if $freq(X, XDS) \ge |XDS| \times \delta$, where δ is an user-defined threshold which range is $0 \le \delta \le 1$. Otherwise, X is pruned from the consideration. However, in negative association rules can be mainly used those X because it is possible **the negated *X*** can be frequent. To supplement frequency check, we adopt two different measures, *interestingness* and *correlation coefficient* values. The former was used by Wu *et al*. [10] to identify potential interesting

International Journal of Scientific Engineering and Science
Volume 1, Issue 10, pp. 69-73, 2017. ISSN (Online): 2456-7361

itemsets, the latter which has been a famous factor in data mining was applied in the paper [11]. To judge a possibility of rule's interestingness Wu *et al*. defined a function interest with a threshold *min_int*. We slightly modify the function and apply to the rule $X \Rightarrow \neg Y$ as shown in the equation (6).

$$\begin{aligned} interest(X, \neg Y) &= |freq(X \cup \neg Y, XSD) \\ &\quad - freq(X, XSD) freq(\neg Y, XSD)| \\ &= |freq(X, XSD) - freq(X \cup Y, XSD) \\ &\quad - freq(X, XSD)(1 - freq(Y, XSD)| \\ &= |freq(X \cup Y, XSD + freq(X, XSD) freq(Y, XSD)| \end{aligned} \quad (6)$$

The rule $X \Rightarrow \neg Y$ is worth to discover if the computational output of *interest(X, ¬ Y)* is greater than or equal to a user defined threshold minimum interest *min_int* even though its support and confidence are low. Otherwise, it is judged to be not-interested and pruned. However, two titemsets can have strong co-relation which implies they are dependent each even its interest value is not satisfied a given condition.

Correlation Coefficient is another measurement to prune uninteresting items. It measures a strength of association between two variables [6]. The correlation coefficient value between random variables X and Y is the degree of linear dependency, which is known as the covariance of the two variables, divided by their standard deviations. By Karl Pearson ϕ coefficient was introduced to evaluate associations between two itemsets X and Y. It measures the association for two binary values, 1 or 0. Therefore, it is adequate to consider the existence of an itemset in transactions; if an itemset exists it is regarded as 1, otherwise 0. When X and Y are two binary variables, the associations of them can be summarized in a 2 × 2 contingency table given in Table I.

TABLE I. 2×2 contingency table for two binary variables.

	Y = 1	**Y = 0**	**TOTAL**
X =1	n_{11}	n_{10}	n_{1+}
X = 0	n_{01}	n_{00}	n_{0+}
TOTAL	n_{+1}	n_{+0}	n

In the table, n_{11}, n_{10}, n_{01}, n_{00} are positive counts of numbers satisfying both X and Y, and n is a total number of a data set. With the counts, the association is evaluated from the ϕ correlation coefficient;

$$\emptyset_{XY} = \frac{n_{11}n_{00} - n_{10}n_{01}}{\sqrt{n_{1+}n_{0+}n_{+1}n_{+0}}}. \quad (7)$$

The equation (7) is re-written to be composed of only the itemsets which binary values are 1 because 1 means existence.

$$\begin{aligned} \emptyset_{XY} &= \frac{n_{11}(n - n_{11} - n_{10} - n_{01}) - n_{10}n_{01}}{\sqrt{n_{1+}n_{0+}n_{+1}n_{+0}}} \\ &= \frac{n_{11}n - n_{11}n_{11} - n_{11}n_{10} - n_{11}n_{01} - n_{10}n_{01}}{\sqrt{n_{1+}n_{0+}n_{+1}n_{+0}}} \\ &= \frac{nn_{11} - (n_{11}^2 + n_{11}n_{10} + n_{11}n_{01} + n_{10}n_{01})}{\sqrt{n_{1+}n_{0+}n_{+1}n_{+0}}} \\ &= \frac{nn_{11} - (n_{11}+n_{01})(n_{11}+n_{01})}{\sqrt{n_{1+}n_{0+}n_{+1}n_{+0}}} = \frac{nn_{11} - n_{1+}n_{+1}}{\sqrt{n_{1+}n_{0+}n_{+1}n_{+0}}} = \\ &\frac{nn_{11} - n_{1+}n_{+1}}{\sqrt{n_{1+}(n - n_{1+})n_{+1}(n - n_{+1})}} \end{aligned} \quad (8)$$

The strength of correlation coefficient was described in the book [12]. Based on the arguments we define the strength of ϕ as; a correlation of ±0.5 is large, ±03. Is moderate, and ±0.1 is small. Table 2 is a contingency table for possible combination of titemsets x and Y and the eqaution (9) is for $\emptyset$ (X, Y), where N is a total number of transactions, actually $|XDS|$.

TABLE II. 2×2 contingency table for $X \Rightarrow Y$.

	Y	**¬Y**	***total***
X	*freq(X ∪ Y)*	*freq(X ∪ ¬Y)*	*freq(X)*
¬X	*freq(¬X ∪ Y)*	*freq(¬X ∪ ¬Y)*	*freq(¬X)*
total	*freq(Y)*	*freq(¬Y)*	*N*

$$\emptyset_{XY} = \frac{N \cdot freq(X \cup Y) - freq(X) \cdot freq(Y)}{\sqrt{freq(X) \cdot (N - freq(X)) \cdot freq(Y) \cdot (N - freq(Y))}} \quad (9)$$

C. Algorithm for Negated Treeitemsets

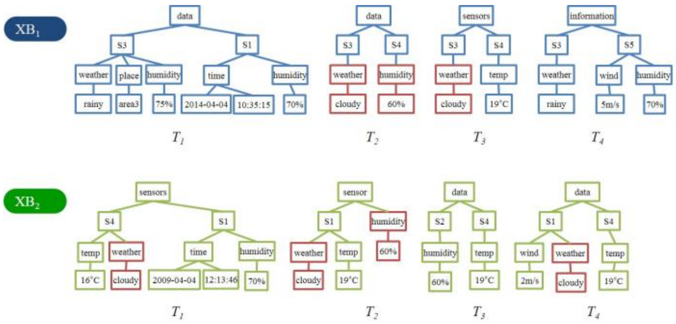

Fig. 1. XML stream data with two blocks.

Let us consider the data on Fig. 1. When we assume minimum support is 20%, the condition implies that only the fractions which occurrence counts in the data are more than 1.6 times are chosen for the next step. However, some unchosen fractions can have interesting information even though their supports are less than 20%, or on the contrary, some chosen fractions can have less interesting or useless information. Four possible candidate association rules are extracted and they are on Fig. 2. The support of each titemset configuring the rules satisfies the minimum support. However, when titemsets are correlated each other, the support values of their unions show different results. According to the equations (1) to (3) and previously known association rules equations, we evaluate the support and confidence values of each rule. We assume a minimum confidence is 50%.

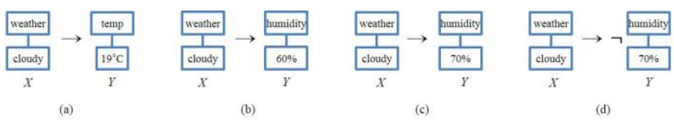

Fig. 2. Candidate association rules.

(a) $support(X \Rightarrow Y) = \frac{3}{8} \cdot 100 = 37.5\%$,

$confidence(X \Rightarrow Y) = \frac{3}{5} \cdot 100 = 60\%$,

(b) $support(X \Rightarrow Y) = \frac{2}{8} \cdot 100 = 25.0\%$,

$confidence(X \Rightarrow Y) = \frac{2}{5} \cdot 100 = 40\%$,

International Journal of Scientific Engineering and Science
Volume 1, Issue 10, pp. 69-73, 2017. ISSN (Online): 2456-7361

(c) $support(X \Rightarrow Y) = \frac{1}{8} \cdot 100 = 12.5\%$,

$confidence(X \Rightarrow Y) = \frac{1}{5} \cdot 100 = 20\%$,

(d) $support(X \Rightarrow \neg Y) = \frac{4}{8} \cdot 100 = 50.0\%$,

$confidence(X \Rightarrow \neg Y) = \frac{4}{5} \cdot 100 = 80\%$.

In the condition of reliability, confidence, (b) does not pass the condition, which means the rule is not a good association rule because the relationship between X and Y is weak; the higher the confidence is, the more likely it is for Y to be present in transactions that contain X. The third rule (c) cannot be generated because its support does not meet 20%. The last one (d) has a negated consequent, ¬Y, which is the negative version of (c). To evaluate its support and confidence, (c)' support and confidence values are used instead. The rule (d) shows a totally different result from that of (c); the support and confidence have the highest values among 4 candidate rules. There is nothing different except that the titemset Y is absent. By negating the consequent part, we obtain the most useful and important information with excellent strength and reliability.

With the equation (6) we compute the interestingness of each candidate association rule on Fig. 2 for the given *min_int* = 30%. For more clear comparison, we consider also ¬Y for all rules and measure their interestingness, except the rule (d). Following are the list showing the calculations:

(a) $interest(X,Y) = \left|\frac{3}{8} - \frac{5}{8} \cdot \frac{4}{8}\right| \cdot 100\% = 6.2\%$,

(b) $interest(X,Y) = \left|\frac{2}{8} - \frac{5}{8} \cdot \frac{3}{8}\right| \cdot 100\% = 1.56\%$,

(c) $interest(X,Y) = \left|\frac{1}{8} - \frac{5}{8} \cdot \frac{3}{8}\right| \cdot 100\% = 10.9\%$,

(d) $interest(X, \neg Y) = \left|\frac{1}{8} + \frac{5}{8} \cdot \frac{3}{8}\right| \cdot 100\% \approx 35.9\%$,

(a') $interest(X, \neg Y) = \left|\frac{3}{8} + \frac{5}{8} \cdot \frac{4}{8}\right| \cdot 100\% \approx 68.75\%$,

(b') $interest(X, \neg Y) = \left|\frac{2}{8} + \frac{5}{8} \cdot \frac{3}{8}\right| \cdot 100\% \approx 48.4\%$,

(c') $interest(X, \neg Y) = \left|\frac{1}{8} + \frac{5}{8} \cdot \frac{3}{8}\right| \cdot 100\% \approx 35.9\%$.

The interest values of (a), (b), and (c) are too low to satisfy *min_int*. In cases of (a) and (b) especially, those rules were considered to have enough occurrences to be association rules by the support factor. Even more the rule (a) has been evaluated to have strong reliability between titemsets by the confidence measure. However, neither (a) nor (b) are decided to be valuable by the method of interestingness. On the contrary, the negated versions of (a), (b), and (c) produce the opposite results; all of them have sufficient values of interestingness. In case of (c'), the original rule (c) was determined to be pruned from the rule generation because both support and confidence values did not satisfy the conditions. But, the negated version is quite interesting and is actually the rule (d). Consequently, only the negative association rule (d) has the condition-satisfying titemsets. From the obtained values we can observe the followings: 1) some rules can have high interest values even though they have not sufficient support nor confidence, and 2) it depends on how properly set up mi to find suitable titemsets for useful and usable association rules.

With the equation (9), we estimate the correlation coefficient value of each candidate rule. Note that for the rule (d), we use Y instead of ¬Y because the measure of correlation coefficient takes positive titemsets as input. The computed results are shown in the following list:

(a) $\emptyset_a = \frac{1 \cdot \frac{3}{8} - \frac{5}{8} \cdot \frac{4}{8}}{\sqrt{\frac{5}{8} \cdot \left(1 - \frac{5}{8}\right) \cdot \frac{4}{8} \cdot \left(1 - \frac{4}{8}\right)}} = \frac{1}{\sqrt{15}} \approx 0.258$,

(b) $\emptyset_b = \frac{1 \cdot \frac{2}{8} - \frac{5}{8} \cdot \frac{3}{8}}{\sqrt{\frac{5}{8} \cdot \left(1 - \frac{5}{8}\right) \cdot \frac{3}{8} \cdot \left(1 - \frac{3}{8}\right)}} = \frac{1}{15} \approx 0.07$,

(c) $\emptyset_c = \frac{1 \cdot \frac{1}{8} - \frac{5}{8} \cdot \frac{3}{8}}{\sqrt{\frac{5}{8} \cdot \left(1 - \frac{5}{8}\right) \cdot \frac{3}{8} \cdot \left(1 - \frac{3}{8}\right)}} = -\frac{7}{15} \approx -0.47$,

(d) $\emptyset_d = \frac{1 \cdot \frac{1}{8} - \frac{5}{8} \cdot \frac{3}{8}}{\sqrt{\frac{5}{8} \cdot \left(1 - \frac{5}{8}\right) \cdot \frac{3}{8} \cdot \left(1 - \frac{3}{8}\right)}} = -\frac{7}{15} \approx -0.47$.

For the case (a), its coefficient determination shows a modest value of 0.258 approximately, which means two titemsets move in tandem but their association is not strong enough to conclude that it will give many benefits. The interestingness value for (a) also determined that it would be few interest to be discovered, while the support and confidence judged that it would highly occur with strong reliability. Next, it can be informed that the case (b) is not worth considering because its value is less than +0.1, which means two titemsets are nearly independent each other and the association between them is seldom made. Then, it is quite interesting to look at the next two cases (c) and (d) - negative version of (c). The coefficients of both rules indicate that 1) the rules have negative relationship between titemsets, 2) the strengths of their coefficients are quite strong enough to give valuable information, and 3) the generated negative association rule will provide many opportunities for further mining, even though their support values are less than minimum support and the rules are not attracted in the ordinary association rules mining. With the frame of correlation coefficient, the hidden association of (c) provides benefits when it is mined for a negative association rule, which is not caught by support/confidence or even interest

For reliable and trustworthy pruning, we adapt the two measuring factors, interestingness and correlation coefficient. The figure 3 is a framed algorithm to discover interesting negated titemsets. Due to the lack of the time, details of the design and explanation of the algorithm will be given in an extended version of the paper.

```
Input : XDS, δ, min_sup, min_conf, min_int
Output: a set of interesting negated titemsets NT

derive every titem satisfying min_sup from XDS
for each titemset X, Y configure the form X ⇒ Y
      Y ← ¬Y
      if sup(X ⇒ Y) < min_sup or conf(X ⇒ Y) < min_conf
            compute interest(X, ¬Y)
            if interest(X, ¬Y) < min_int
                  then compute ϕ(X,¬Y)
                  if ϕ(X,¬Y) ≤ -0.3 or ϕ(X,¬Y) ≥ +0.3
                        then NT ← NT + {X ⇒ ¬Y}
            else
                  then NT ← NT + {X ⇒ ¬Y}
```

Fig. 3. Frame algorithm for negated titemsets.

IV. CONCLUSION

In this paper, we have presented a framed algorithm and some measuring factors to derive meaningful negated titemsets for the configuration of negative association rules. In the extended version of the article, the detailed explanation of the algorithm and experimental result will be presented for better understanding.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning(NRF-2017R1A2B1007015).

REFERENCES

[1] N. Jian and L. Grunewald, "Research issues in data stream association rule mining," *ACM SIGMOD Record*, vol.35, no. 1, pp. 14–19, 2006.
[2] K. K. Loo, I. Tong, B. Kao, and D. Chenung, "Online algorithms for mining inter-stream associations from large sensor networks," in *Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining, LNAI*, vol. 3518, pp. 143-149, 2005.
[3] R. Motwani and G. S. Manku, "Approximate frequency counts over data streams," in *Proceedings of the 28th International Conference on Very Large Data Base*, Hong Kong, pp. 346-357, 2002.
[4] A. Boukerche and S. Samarah, "A novel algorithm for mining association rules in wireless ad hoc sensor networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 19, pp. 865-877, 2008.
[5] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, USA, pp. 1-12, 2000.
[6] S. Corpinar and T. Í. Gündem, "Positive and negative association rule mining on xml data streams in database as a service concept," *Expert Systems with Applications*, vol. 39, no. 8, pp. 7503-7511, 2012.
[7] J. Paik, J. Nam, U. Kim, and D. Won, "Association rule extraction from xml stream data for wireless sensor networks," *Sensors*, vol. 14, pp. 12937-12957, 2014.
[8] X. Yuan, B. P. Buckles, Z. Yuan, and J. Zhang, "Mining negative association rules," in *Proceedings of the Seventh International Symposium on Computers and Communications*, Italy, pp. 623-628, 2002.
[9] S. Abiteboul, P. Buneman, and D. Suciu, *Data on the web – from relations to semistructured data and xml*. Morgan Kaufmann Publishers, 2000.
[10] X. Wu, C. Zhang, and S. Zhang, "Efficient mining both positive and negative association rules," *ACM Transaction on Information Systems*, vol. 22, no. 3, pp. 381-405, 2004.
[11] M. L. Antonie and O. R. Zaïane, "Mining positive and negative association rules: an approach for confined rules," in *Proceedings European Conference on Principles and Practice of Knowledge Discovery in Databases*, pp. 27-38, 2004.
[12] J. Cohen, *Statistical power analysis for the behavioral sciences*, Lawrence Erlbaum, New Jersey, 1988