EXTRACTION OF POSITIVE AND NEGATIVE ASSOCIATION RULES FROM TEXT: A TEMPORAL APPROACH

S. Mahmood*, M. Shahbaz and Z. U. Rehman

Department Of Computer Science & Engineering, UET, Lahore
Corresponding Author E-Mail: *sajid.mahmood@kics.edu.pk

ABSTRACT: The focus of the association rule mining community, to a major extent, has been towards mining positive association rules. The negative association rules, counterparts of the positive association rules, have attained much less attention of the researchers. Finding the association rules from temporal perspective is quite new to the data mining community. The datasets used for mining temporal associations have generally been either data streams or market basket transactions. Extraction of temporal associations has not been too popular with the association rule mining researchers, especially the negative associations in temporal domain. The temporal association rules enable answering vital questions for the application of association rules, thus enhancing the trustfulness of the generated rules. Associations among diseases and symptoms, both positive and negative, are time/season dependent e.g. flu can have different symptoms in summer and in winter. The rule should positively represent some stable and reliable relationships in the domain. We propose a technique for mining temporal positive and negative associations from medical blogs. Experiments prove the efficacy of the proposed technique.

Key words: Text Mining, Temporal Association Rules, Negative Association Rules, Frequent patterns, Infrequent Patterns.

INTRODUCTION

Data mining aims at the discovery of the helpful and interesting patterns from data. Generated knowledge patterns, generally represented as association rules, provide important insights for the users. There have been many researches focusing on algorithms for efficiently mining association rules. The problem faced by most of these techniques is the generation of abundant and redundant association rules (Liu et al., 1999; Piatetsky-Shapiro et al., 1994; Silberschatz and Tuzhilin, 1996). Interestingness measures, in recent years, have been proposed for extracting only the interesting and useful rules (Bayardo Jr and Agrawal, 1999; Geng and Hamilton, 2006; Gharib et al., 2010).

Traditional algorithms use the whole dataset, in general, for finding associations, before filtering out the unimportant and uninteresting rules. Hence they ignore the lively behavior of the people's dynamically changing thoughts over time. People tend to attribute their feelings and illnesses to the changes in the environment and the weather. We may attribute the environment change to the locality of a person expressing his/her thoughts; however, the weather change is a time-baud factor. We examine the association rules from the temporal perspective for their interestingness. The technique of finding temporal associations from the textual datasets is also bound to be more beneficial because most of the datasets are congregated over time. Therefore, the temporal associations represent the true and reliable behavior.

Additionally, many important association rules may be ignored when mining association rules without keeping in view the temporal context. $\boldsymbol{Pollens \rightarrow Flu}$, for example, hold a strong association during the spring in Islamabad, Pakistan. However, this may not hold when we mine associations over the whole year. Temporal rules thus, can provide the truly important association rules to the users.

Association rules have been extensively studied since their introduction in 1993 . An association rule is an insinuation of the form $x \rightarrow y$ where x and y both belong to *I* where $I = \{I_1, I_2, \dots, I_n\}$ is a set of items from D, a database of transactions (documents) *T*, each transaction *T* is a set of items such that $T \subseteq I$ and each transaction has an identifier T_{ID}. Both the sub-items x and y belonging to an itemset A hold the properties: $x \subseteq I$ and $y \subseteq I$, $x \cup y = A$ and $x \cap y = \emptyset$. The association rule $x \rightarrow y$ holds in database D with the following measures: *support supp,* such that s% of transactions in D contain $x \cup y$; *confidence conf,* such that c% of transactions in D contain *y* that also contain *x*. An association rule is considered interesting if it has support greater than the user defined *supp* threshold, and has a confidence greater the user defined threshold *conf*. Other rule interestingness measures have also been

reported in the literature for pruning the extracted association rules (Geng and Hamilton, 2006).

Wang et al introduced the temporal association rules (TARs) in (Wang et al., 2001). The temporal association rules help in extracting the associations in periodic datasets. They capture the temporal expressions in the form of association rules. Temporal association rules mainly differ from traditional association rules in that they attempt to characterize the temporal connections among data items. Literature reports different types of temporal association rules like inter-transactional, episodic, trend dependent, sequential, and calenderic associations.

Majority of the algorithms proposed for extracting temporal association rules did not consider the exhibition period of an item, the time window in which the item appears in the dataset (Huang et al., 2007). This handicapped the temporal association rule mining algorithms from capturing the true and effective temporal relationships. Algorithm for extracting general temporal association rules is proposed in (Lee et al., 2003). This algorithm calculates the support of the items according to their exhibition periods. The interestingness measures, support and confidence, are reformulated accordingly.

Ale J. M. et al introduced the notion of temporal support (Ale and Rossi, 2000). They extended the non-temporal model by integrating the time factor to the association rule mining model. ITARM algorithm proposed by Tarek et al (Gharib et al., 2010) discovers temporal frequent itemsets from the data. It depends on the previously generated 2-itemsets.

Lee C. H. et al proposed the progressive partition miner (PPM) algorithm (Lee et al., 2001). The algorithm proceeds by partitioning the dataset into time slices. It prunes infrequent itemsets for each partition using thresholds. Chang C. Y. et al proposed the segmented progressive filtering (SPF) algorithm (Chang et al., 2002). It divides the datasets into subsets using the common starting and end times. On each subset, it finds the candidate 2-itemsets using cumulative filtering threshold.

Chen M. et al employ temporal associations for event detection in a video dataset (Chen et al., 2007). They base their approach on feature extraction and hierarchical temporal association rule mining. They capture the temporal characteristics of the patterns of interest.

Most of the temporal association rule (TAR) mining techniques do not consider the individual item occurrence period. Miao R. et al present MPTAR algorithm, a periodic temporal association rule mining technique (Miao and Shen, 2010). It is a two-step process that mines the trend of continuous items in first phase and then calculates the period of the items in the second phase.

Research on association rule mining has produced numerous algorithms like Apriori and its variants, sampling, partitioning, and frequent pattern growth algorithms (Grahne and Zhu, 2005; Han et al., 2000). Additionally, some other variants like multiple supports, multiple confidence, correlational associations, associations from infrequent items, and temporal association rule mining algorithms are presented in the literature (Huang et al., 2007; Li and Chen, 2009; Wu and Chen, 2009). Generation of temporal association rules from the chronic data has gained popularity in recent research. Temporal associations add an interesting dimension to the traditional association rule mining techniques. Many domains including the likes of communication, blog texts, chronic disease spreads, finance, weather, economics, etc. present extensive temporal data.

Temporal association rules are prone to become useless with the addition of new data to the database. Few algorithms for incremental mining of association rules from the temporal database have been proposed to overcome this issue (Ng et al., 2007). These algorithms include Fast Update (FUP), Update Large Itemsets (ULI), Negative Border with Partitioning (NBP), Update With Early Pruning (UWEP), New Fast Update (NFUP), Fast Incremental Mining (FIM) and Pre-FUFP (Emam, 2009; Gharib et al., 2008; Lin et al., 2009; Yuping et al., 2009).

We, in this paper, present the statistical techniques for the analysis of temporal associations from the medical blogs. We begin by first dividing the data into temporal chunk of the blog texts (these chunks could be weeks, months, year, etc. according to the dates they are posted on the blog website). We extract associations among these temporal chunks using threshold support and confidence measures. We were able to identify three types of associations among diseases and symptoms from this temporal association rule mining:

1. **Seasonal/Periodic Associations:** associations among diseases and symptoms observed over a specific period e.g. during rainy season, and extreme cold weather etc.

2. **Event-based Associations:** associations discovered during specific events like flood, bird flu, etc.

3. **Established Associations:** associations among diseases and symptoms those remain almost constant during all seasons/times.

The experimental results demonstrate the promising efficacy of the associations extracted. Large availability of the data in temporal chunks, because many people are sharing their feelings and expressions, assures that data division does not degrade the importance of the extracted rules. Division of the discovered rules into *seasonal, event-base, and established* association rules can greatly help the medical domain practitioners.

1.1. **Terminology and Background:** Let $\boldsymbol{I} = \{i_1, i_2 \ldots i_N\}$ be a set of N distinct literals/terms called ***items***. Let ***D*** be a database of transactions (documents/blogs etc) where each transaction ***T*** is a set of items/terms such that **T** is a subset of '**I**'. Each transaction is associated with a unique identifier, called $\boldsymbol{T_{ID}}$. Let A, B be sets of items; an association rule is a derivation of the form, $\boldsymbol{A \Rightarrow B}$, where $\mathbf{A} \subset \mathbf{I}$, $\mathbf{B} \subset \mathbf{I}$, and $\mathbf{A} \cap \mathbf{B} = \boldsymbol{\emptyset}$. '***A***' is called the *antecedent* of the rule, and '***B***' is called the *consequent* of the rule. An association rule $\boldsymbol{A \Rightarrow B}$, can have different measures denoting its significance and quality. In our approach, we have employed i) **support**, we denote it as ***supp*** which is the percentage of transactions in database D containing both A and B ii) **confidence**, we denote it as ***conf*** which is representing the percentage of transactions in ***D*** containing ***A*** that also contain ***B*** which can be denoted in probability terms as $\boldsymbol{P(B|A)}$, iii) **Interest**, we denote it as **Interest** characterizing the direction of correlation between the antecedent and consequent of the association rule. Rules having the support value greater than user defined minimum support ***minsupp*** *i.e.* the item set needs to be present in minimum threshold number of transactions; and confidence greater than user defined minimum confidence ***minconf***, are called valid association rules. The Interest symbolizes the association whether positive or negative. A value of Interest greater than 1 indicates a positive relationship between the item sets; value of Interest less than 1 indicates a negative relationship; and where the value of Interest equals 1, the item sets are independent and there exists no relationship between the item sets.

We can represent a few of the above and other derived definitions in equations as follows:

$supp(A)$ /* Number/Percentage of transaction(s) containing A */ (1)

$supp(A \rightarrow B) = P(AB)$ /* Number/Percentage of transactions where A and B co-exist*/ (2)

$conf(A \rightarrow B) = \frac{P(AB)}{P(A)}$ /* Confidence measure of the rule that whenever A occurs (3)

B also occurs in transaction(s)*/

$interest(A \rightarrow B) = \frac{P(AB)}{P(A)P(B)}$ /* The strength of relationship between A and B */ (4)

$= 1 + \frac{P(AB) - P(A)P(B)}{P(A)P(B)}$

$Supp(\neg A) = 1 - Supp(A)$ (5)

$Supp(A \cup \neg B) = Supp(A) - Supp(A \cup B)$ (6)

$Conf(A \Rightarrow \neg B) = 1 - Conf(A \Rightarrow B) = \frac{P(A \neg B)}{P(A)}$ (7)

$Supp(\neg A \cup B) = Supp(B) - Supp(B \cup A)$ (8)

$Conf(\neg A \Rightarrow B) = \frac{(Supp(B)(1 - Conf(B \Rightarrow A))}{1 - P(A)} = \frac{Supp(\neg A \cup B)}{Supp(\neg A)}$ (9)

$Supp(\neg A \cup \neg B) = 1 - Supp(A) - Supp(B) + Supp(A \cup B)$ (10)

$Conf(\neg A \Rightarrow \neg B) = \frac{1 - Supp(A) - Supp(B) + Supp(A \cup B)}{1 - P(A)} = \frac{Supp(\neg A \cup \neg B)}{Supp(\neg A)}$ (11)

MATERIALS AND METHODS

We present a 3-step technique for the extraction of temporal associations from the medical blogs. We begin by first dividing the data into temporal subsets of the blog texts data (these subsets could span weeks, months, year, etc. and seasonal and event specific also).

I. Dataset Partitioning: We first divide original dataset D is into a number of sub-datasets D_1, $D_2 \ldots D_m$, according to the time-periods, T_1, $T_2 \ldots T_m$

II. Rule Mining from Partitioned Dataset: In second step, we extract association rules R_i from each sub-dataset D_i.

a. The set R of extracted rules can be defined as follows:

$$R = \{r \mid r \in (R_1 \cup R_2 \ldots \cup R_m)\}$$

This implies that a rule *r* is of potential interest if it appears in any rule set R_i. A rule r, appearing in any of the temporal rule sets R_i may not appear in another temporal rule set R_j. The reason could be that r may not have the threshold support and confidence in the temporal subset D_j, meaning that it is a seasonal/periodic association rule.

III. Temporal Analysis of Rules: We produce rules satisfying user given support and confidence values over time.

3.1 Proposed Algorithm: Frequent & Infrequent Itemsets Generation using FP-Tree

Input: FP-Tree; minimal support ***minsupp***.

Output: Frequent Itemsets, **FIS**; Infrequent Itemsets of interest, **InFIS**.

FIS ← Ø, InFIS ← Ø

For some node N_x in the header table of FP-Tree

Traverse Node Path $NP_x = \{N_1, \ldots, N_k\}$ ($N_x \notin NP_x$) in FP-Tree, where, N1,…, N_k, are node items.

A ← N_x.items, B ← Ø

For each Node $N_x \in NP_x$,

if supp(N_x.items) $\geq$ minsupp

Then A ← N_x.items; FIS ← FIS $\cup$ A

else B ← N_x.items; InFIS ← FIS $\cup$ B

return FIS, InFIS.

3.2 Positive and Negative Association Rule Generation

Input: database *D*; Frequent Itemsets *FIS*; Infrequent Itemsets *InFIS*; Minimum Support *minsupp;* Minimum Confidence *minConf*: threshold values;

Output: Positive Association Rules (PARs); Negative Association Rules (NARs);
/*generating all association rules from **FIS (Frequent itemsets)**.*/
For each itemset $I \in FIS$
do begin
For each itemset $X \cup Y = I$, $X \cap Y = \Phi$
do begin
/*generate rules of the form $X \Rightarrow Y$. */
If
$conf(X \Rightarrow Y) \geq min\_conf \;\&\&\; lift(X \Rightarrow Y) > 1$
then output the rule $X \Rightarrow Y$; $PAR \cup (X \Rightarrow Y)$
else
/*generate rules of the form $X \Rightarrow \neg Y$ and $\neg X \Rightarrow Y$. */
if
$\mathbf{conf(X \Rightarrow \neg Y) \geq min\_conf} \;\&\&\; lift(X \Rightarrow \neg \mathbf{Y}) > 1$
output the rule $\mathbf{X \Rightarrow \neg Y}$; $\mathbf{NAR \cup (X \Rightarrow \neg Y)}$
if
$\mathbf{conf(\neg X \Rightarrow Y) \geq min\_conf} \;\&\&\; lift(\neg X \Rightarrow \mathbf{Y}) > 1$
output the rule $\mathbf{\neg X \Rightarrow Y}$; $\mathbf{NAR \cup (\neg X \Rightarrow Y)}$
end for;
end for;
/*generating all negative association rules from **inFIS (Infrequent itemsets).***/
For each item $I \in InFIS$
do begin
For each itemset $X \cup Y = I$, $X \cap Y = \Phi$, $supp(X) \geq FIS$ **and** $supp(Y) \geq FIS$
do begin
/*generate rules of the form $X \Rightarrow Y$. */
If
$conf(X \Rightarrow Y) \geq min\_conf \;\&\&\; lift(X \Rightarrow Y) > 1$
then output the rule $X \Rightarrow Y$; $PAR \cup (X \Rightarrow Y)$
else
/*generate rules of the form $X \Rightarrow \neg Y$ and $\neg X \Rightarrow Y$. */
if
$conf(X \Rightarrow \neg Y) \geq \min\_conf \;\&\&\; lift(X \Rightarrow \neg Y) > 1$
output the rule $X \Rightarrow \neg Y$; $NAR \cup (X \Rightarrow \neg Y)$
if
$conf(\neg X \Rightarrow Y) \geq \min\_conf \;\&\&\; lift(\neg X \Rightarrow Y) > 1$
output the rule $\neg X \Rightarrow Y$; $NAR \cup (\neg X \Rightarrow Y)$
end for;
end for;
return PAR, NAR;

2. Analysis of Extracted Rules over Time: We analyze the extracted association rules with respect to support and confidence in the given time period. The support and confidence are not the only statistical measures that can be used for rule interestingness analysis. However, these are the most common used interestingness measure and are able enough of serving our purpose.

We categorized the identified associations, among diseases and symptoms, into three types:

a. **Seasonal/Periodic Associations:** associations among diseases and symptoms observed over a specific period e.g. during rainy season, and extreme cold weather etc.

b. **Event-based Associations:** associations discovered during specific events like flood, bird flu, etc.

c. **Established Associations:** associations among diseases and symptoms those remain almost constant during all seasons/times.

2.1. **Seasonal/Periodic Associations:** The association rules can have varying support and confidence measures over time i.e. a rule can have different interestingness at different times. We call such association rules the seasonal association rules. Alternately, we can call them the periodic association rules. We define a rule to be seasonal/periodic if its support and confidence meet threshold minsupp and minconf only in a specific season or a time period. The season could be for example, rainy season, winter season, etc. We can write a seasonal/periodic association rule mathematically as:

- Threshold minimum support *minsupp* *(1)*
- Threshold minimum confidence *minconf* *(2)*
- Support over the whole dataset $supp_D$ *(3)*
- Confidence over the whole dataset $conf_D$ *(4)*
- Support over a time period $supp_i$ *(5)*
- Confidence over a time period $conf_i$ *(6)*

An association rule is considered a seasonal/periodic if it satisfies the following conditions:

1. $supp_D < minsup$ and $conf_D < minconf$
2. $supp_i \geq minsup$ and $conf_i \geq minconf$

Example 1: Assume in a mining session we set minconf to 60%. There is a time period T_i in which the confidence

$conf_i$ of a rule, $A \rightarrow B$, is 65%. There are 1000 records in the dataset. We divide the data set into four subsets of 400, 350, 150, and 100 records each. The fourth subset containing 100 data records/transactions, all of them containing itemset *A*. Out of these 100 transactions 65 also contain the consequent *B* (i.e., $conf_i$ =65/100 = 65%). The question is whether $conf_i$ is statistically above 60%. The answer in above example is yes, both the support and confidence of $A \rightarrow B$ in the fourth subset satisfy the thresholds. Hence $A \rightarrow B$ is valid association rule in the fourth subset. This meets our definition of seasonal/periodic association rule.

2.2. **Event-based Associations:** A seasonal/periodic association rule requires its interestingness measures i.e. support and confidence, to satisfy threshold values for a specific time of year or spanning over parts of consecutive years. Event based associations, on the other hand, are the associations that capture the associations among diseases and symptoms on the time line of a particular event. The event could be a calamity of any kind e.g. flood, tsunami, hurricane, earthquake, etc. The users want to know the spread of disease during the outbreak of a particular type of calamity and the associated symptoms with a predictable behavior. We term such type of rules as event-based association rules. Intuitively speaking, an event-based association rule is a periodic association rule. The interestingness calculation of event-based rules is similar to that of periodic/seasonal association rules.

Example 2: Suppose we have a data set spanning multiple years and containing 3000 tuples. We partition the data into three sub-datasets according to the calamities occurring during the time frame. We want to analyze rules of the form, $A \rightarrow B$, $\neg A \rightarrow B$ etc. found in the dataset. The overall confidence values of the extracted association rules in different time periods are:

- T_1 (tsunami): $confidence_1$ = 64% (= 573/895)
- T_2 (Katrina): $confidence_2$ = 75% (= 400/1160)
- T_3 (Chili earthquake): $confidence_3$ = 83% (= 784/945)

We compute the confidence of a rule r_i in D_i using the support counts of the constituent itemsets {*A*} and $\{A \cap B\}$ in D_i.

2.3. **Established/Stable Rules:** An association rule that has its interestingness measures i.e. support and confidence above the *minsupp* and *minconf* thresholds respectively over time, is called an established/stable association rule. The users desire established/stable associations because the seasonal/periodic associations can be volatile. We consider an association rule to be established/stable if it satisfies the following conditions:

1. $supp_D \geq minsup$ and $conf_D \geq minconf$ eq. (3), (1), (4), (2)
2. $supp_i \geq minsup$ and $conf_i \geq minconf$ eq. (5), (1), (6), (2)

Example 3: Suppose we have a data set spanning multiple years and containing (1,00,000) tuples. We partition the data into four sub-datasets arbitrarily over time. We want to analyze rules of the form, $A \rightarrow B$, $\neg A \rightarrow B$ etc. found in the dataset. We compute the confidence of a rule r_i in D_i using the support counts of the constituent itemsets {*A*} and $\{A \cap B\}$ in D_i. The confidences of rules in different time periods are:

$support_1$ = 53% and $confidence_1$ = 70% over T_1 (sub-dataset-1)
$support_2$ = 47% and $confidence_2$ = 80% over T_2 (sub-dataset-2)
$support_3$ = 58% and $confidence_3$ = 63% over T_3 (sub-dataset-3)
$support_4$ = 55% and $confidence_4$ = 61% over T_4 (sub-dataset-4)

RESULTS AND DISCUSSION

The experimental results express the effectiveness of our proposed technique. We performed experimentation on real data collected from blog sites. We did not employ data sets from available online repositories such as UCI machine learning repository (Bache and Lichman, 2013) because the majority of datasets available there are not temporal datasets. We collected datasets from the following:

- Katrina Aftermath [http://katrina05.blogspot.com/]
- Cancer Survivors Network [http://csn.cancer.org/]
- Care Pages [http://www.carepages.com/forums/cancer]
- Chile Earthquake [http://news.yahoo.com/topics/chile-earthquake/]

Table 1 shows the generated association rules. The results include the seasonal/periodic, event-based, and the stable/established associations. We extracted these associations from the temporal sub-sets of the datasets we collected from the above-mentioned sources. The seasonal association rules give an insight into the associations among diseases and symptoms during a particular season. For example, from the results table we can see the associations that **Flu** has in winter and summer seasons. Because we are mining the seasonal

associations here, therefore we categorized the data into temporal subsets according to seasons i.e. data from April-September is considered to be summer season data and so on for other seasons. We can see from the results table that the association rules have high confidence values, this is because we incorporated the season as a mandatory sub-itemset to each of the itemset for association rule generation.

Table 1. Generated Association Rules, Seasonal/Periodic, Event-based, and Stable/Established

Rule	Support	Confidence
Seasonal/Periodic Association Rules		
{fever, muscle-ache, summer}->{flu}	0.23008	0.87501
{runny-nose, sore-throat, winter}->{ flu}	0.20651	0.92185
{runny-nose, fever, winter}->{ flu}	0.24625	0.83371
{headache, sore-throat, summer}->{ fever}	0.19478	0.68976
{fever, muscle-ache, flu}->{headache, winter}	0.24012	0.80174
{flu, runny-nose, muscle-ache}->{winter}	0.19285	0.63726
Event-based Associations Rules		
{¬road, ~communication, hurricane}->{damage}	0.24698	0.61428
{loot, weapon}->{storm}	0.28106	0.88034
{¬medicine, ¬sun, flood}->{allergy}	0.31735	0.71043
{water, rain, flood}->{infection}	0.18952	0.62538
{¬medicine, ¬sun, hurricane}->{allergy}	0.22865	0.64722
Established/Stable Associations Rules		
{runny-nose, sore-throat, fever }->{ flu}	0.10706	0.70324
{loot, weapon}->{storm}	0.12001	0.62857
{water, rain, flood}->{infection}	0.11096	0.59999
{fever, headache, cold}->{pneumonia}	0.10015	0.53691

Table 2. Support and Confidence threshold values used for rule generation

Rule Type	Threshold Support Value	Threshold Confidence Value
Seasonal/Periodic	0.15	6479
Event-based	0.2	5397
Established/Stable	0.25	4467

The event-based associations are less oriented towards the disease-symptom analogy. They reflect upon the behavior of people during a particular type of calamity. The event-based association rules demonstrate less confidence values as compared to the seasonal associations. One reason could be that we extract association rules for all types of calamities together from the dataset. The confidence values may increase if we extract the association rules for a specific calamity/event.

The last category in the results describes the established/stable rules i.e. rules that do not vary over time. The support and confidence measures of such rules are always greater than the threshold values. We can see from the results table that the established rules contain almost the same rules as the seasonal association rules.

3. Concluding Remarks and Future Work: In this research work, we contribute to the positive and negative association rule mining research from the unstructured textual data. We propose an algorithm for extracting positive and negative association rules from this data. Identification of temporal associations among text documents holds huge potential. The generation of negative associations alongside positive associations greatly adds to this potential. The discovery of temporal relationships among the diseases and their cause/symptoms can greatly help the medical practitioners. Temporal associations among diseases, symptoms and laboratory test results, whether positive or negative, can help a medical practitioner quickly in making a quick and appropriate decision about the presence or absence of a possible disease. Positive association rules such as $Headache, Runny-nose \Rightarrow Flu$ can tell us that a person who is suffering Flu experiences **Headache** and **Running-nose.** On the contrary, Negative association rules such as $\neg Throbbing - Headache \Rightarrow \neg Migraine$ tells us that if **Headache** experienced by a person is not **Throbbing**, then he may not have **Migraine** with a certain degree of confidence. The applications of this work include medical decision support systems among

others; finding disease-symptom temporal associations. The current work does not consider the context and semantics of the terms/items in the textual data. In future, we plan to assimilate the context of the features in our work, in order to enhance the eminence and practicality of the engendered association rules.

REFERENCES

Ale, J.M., Rossi, G.H.. An approach to discovering temporal association rules, in: Proceedings of the 2000 ACM Symposium on Applied computing-Volume 1. pp. 294–300 (2000).

Bache, K., Lichman, M. UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences (2013).

Bayardo Jr, R.J., Agrawal, R., Mining the most interesting rules, in: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 145–154 (1999).

Chang, C.Y., Chen, M.S., Lee, C.H. Mining general temporal association rules for items with different exhibition periods, in: Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference On. pp. 59–66 (2002).

Chen, M., Chen, S.C., Shyu, M.L. Hierarchical temporal association mining for video event detection in video databases, in: Data Engineering Workshop, 2007 IEEE 23rd International Conference On. pp. 137–145 (2007).

Emam, A. Future direction of incremental association rules mining, in: Proceedings of the 47th Annual Southeast Regional Conference. p. 74 (2009).

Geng, L., Hamilton, H.J. Interestingness measures for data mining: A survey. ACM Comput. Surv. 38 (2006).

Gharib, T.F., Nassar, H., Taha, M., Abraham, A. An efficient algorithm for incremental mining of temporal association rules. Data & Knowledge Engineering 69, 800–815 (2010).

Gharib, T.F., Taha, M., Nassar, H. An efficient technique for incremental updating of association rules. International Journal of Hybrid Intelligent Systems 5, 45–53 (2008).

Grahne, G., Zhu, J. Fast algorithms for frequent itemset mining using fp-trees. Knowledge and Data Engineering, IEEE Transactions on 17, 1347–1362 (2005).

Han, J., Pei, J., Yin, Y. Mining frequent patterns without candidate generation, in: ACM SIGMOD Record. pp. 1–12 (2000).

Huang, J.W., Dai, B.R., Chen, M.S. Twain: Two-end association miner with precise frequent exhibition periods. ACM Transactions on Knowledge Discovery from Data (TKDD) 1, 8 (2007).

Lee, C.H., Chen, M.S., Lin, C.R. Progressive partition miner: an efficient algorithm for mining general temporal association rules. Knowledge and Data Engineering, IEEE Transactions on 15, 1004–1017 (2003).

Lee, C.H., Lin, C.R., Chen, M.S. On mining general temporal association rules in a publication database, in: Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference On. pp. 337–344 (2001).

Li, H., Chen, H. Mining non-derivable frequent itemsets over data stream. Data & Knowledge Engineering 68, 481–498 (2009).

Lin, C.W., Hong, T.P., Lu, W.H. The Pre-FUFP algorithm for incremental mining. Expert Systems with Applications 36, 9498–9505 (2009).

Liu, B., Hsu, W., Mun, L.F., Lee, H.Y. Finding interesting patterns using user expectations. Knowledge and Data Engineering, IEEE Transactions on 11, 817–832 (1999).

Miao, R., Shen, X.J. Construction of periodic temporal association rules in data mining, in: Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference On. pp. 2133–2137 (2010).

Ng, V., Chan, S., Lau, D., Ying, C.M. Incremental mining for temporal association rules for crime pattern discoveries, in: Proceedings of the Eighteenth Conference on Australasian database-Volume 63. pp. 123–132 (2007).

Piatetsky-Shapiro, G., Matheus, C.J., others. The interestingness of deviations, in: Proceedings of the AAAI-94 Workshop on Knowledge Discovery in Databases. pp. 25–36 (1994).

Silberschatz, A., Tuzhilin, A. What makes patterns interesting in knowledge discovery systems. Knowledge and Data Engineering, IEEE Transactions on 8, 970–974 (1996).

Wang, W., Yang, J., Muntz, R. TAR: Temporal association rules on evolving numerical attributes, in: Data Engineering, 2001. Proceedings. 17th International Conference On. pp. 283–292 (2001).

Wu, S.Y., Chen, Y.L. Discovering hybrid temporal patterns from sequences consisting of point-and interval-based events. Data & Knowledge Engineering 68, 1309–1330 (2009).

Yuping, W., Nanping, D., Guanling, Z. Incremental Updating Algorithm Based on Partial Support Tree for Mining Association Rules, in: Control, Automation and Systems Engineering, 2009. CASE 2009. IITA International Conference On. pp. 17–20 (2009).