

## Research Article

### Discovering Analytic Associate Rule Filtering on Multi-Dimensional Data Streams

<sup>1</sup>P. Velvadivu and <sup>2</sup>C. Duraisamy

<sup>1</sup>Department of Computing, Coimbatore Institute of Technology Coimbatore-641014, Tamil Nadu, India

<sup>2</sup>Dean, School of Science and Humanities, Kongu Engineering College, Perundurai-638052,  
Tamil Nadu, India

**Abstract:** Multidimensional data points discover structural and chronological association inside the data streams. The PaDSkyline framework in a distributed environment utilized intra group optimization and multi filtering technique for skyline query processing. Skyline query processes within each group of distributed data sets, but dynamic filtering point selection was not performed with cost effective system. Similarity- Profiled temporal Association MINing mEthod (SPAMINE) used reference time sequences and threshold value to filter the information from real world data. Different similarity models for filtering temporal patterns were not very effective for performing the phase shift in time. To attain minimal phase shift based cost effective filtering on multidimensional data stream, Analytic Associate Rule Filtering (AARF) mechanism is proposed in this study. The main objective of AARF is to identify the relationship between attributes on multidimensional data and to filter out the independent attributes from the data streams. Initially, analytic association rule uses the weight computing factor to identify relationship and make inferences while testing multidimensional samples. Secondly, with the help of the analyzed relationship, the AARF mechanism uses the attribute independent criterion to discard negligible weight from association rule. Finally, to filter the analytic association rule with specified phase shift time, the 'if-then' strategy is used in AARF mechanism. AARF mechanism has an ability to make an analytic filtering with minimal phase shift time on multidimensional test dataset. The minimal phase shift time reduces the execution time factor and attains cost effective filtering system. Experiment is conducted using Japanese vowel multidimensional data set extracted from UCI repository for measuring the factors such as the average precision level, execution time, filtering query traffic efficiency and true positive rate.

**Keywords:** Analytic associate rule filtering, attributes, if-then strategy, multi-dimensional data streams, phase shift time, weight computing factor

## INTRODUCTION

Many types of data stream are being generated and processed from varied sources including forecasting of weather conditions, information about a specific location, log file monitoring and so on. Processing of queries evaluates it and performs the job of data processing in data stream management systems. Certain amount of data from similar sources is relevant. A skyline query given with a multidimensional point set retrieves those point of interest that are in no way dominated by certain other points.

In Lijiang *et al.* (2011) proposed, PaDSkyline framework evaluated constrained skyline queries from distributed environment that was present in unstructured manner, where data having more relevant items was geographically distributed between scattered sites. Though efficiency and effectiveness of the method was proved but dynamic filtering point

selection was not performed with cost effective system. Biomedical databases like PubMed retrieve several results of which certain items were relevant to the user. Abhijith *et al.* (2011) proposed a BioNav system, a novel search interface mechanism enabled the user to navigate higher amount of results of query by organizing them using MeSH hierarchy that minimized the navigation cost.

The objective of similarity based association mining with temporal data that change over time using transaction database is to identify associated item sets whose changes are observed on the basis of time are of similar nature using a threshold. Mining of associated patterns based on similarity-profiled was presented using SPAMINE algorithm developed by Jin and Shashi (2009) to derive the patterns and observed the sequences. The application of SPAMINE algorithm for similarity-profile transaction database minimized the search space using lower bounding distance and pruned candidate item sets with the help of monotonicity

**Corresponding Author:** P. Velvadivu, Department of Computing, Coimbatore Institute of Technology Coimbatore-641014, Tamil Nadu, India, Tel.: +91-94866-41259

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

property. Though meaningful results were obtained from real data different similarity models for filtering temporal patterns were not very effective for performing the phase shift in time.

The set of least items can be obtained using association rules which forms as a basis to identify the events that are infrequent and exceptional. Critical Relative Support (CRS) was presented by Ying *et al.* (2011) in “Effectively Indexing the Multi-Dimensional Uncertain Objects for Range Searching” to mine critical least association rules and also revealed that most of the uninterested rules were reduced. Though unwanted rules were significantly minimized, but it cannot be applied to real world scenarios. A parametric rough set model was introduced by Xu *et al.* (2013) for real world datasets using effective rule mining method with respect to threshold of rules and was proved to be fast in obtaining the desired set of rules. In” Finding association rules in semantic web data”, Victoria and Rafael (2011) presented a novel technique to mine association rules from semantic data repositories and proved to be efficient in extracting the patterns in lesser amount of time.

One of the widely adopted methods for efficient retrieval of information is the recommender systems that play a pivotal role by providing quality products and services to the end user. A novel recommendation method, Quantitative Association Rule based Filtering by Shweta and Kamal (2012) was designed in a way that the extracted relationships between the information available in prior and their ranking values that resulted in the optimization of computation cost. The rules that were identified formed as a basis for the user that enhanced the accuracy of the system. But trust based system was not addressed. A recommendation engine was designed by Ozgur and Murat (2012) to personalize an e-commerce website by integrating collaborative filtering and association rule mining. Multilevel association rules were discussed in “Multilevel Association Rule Mining for Bridge Resource Management Based on Immune Genetic Algorithm” by Yang *et al.* (2014) to address the issues related to association rule items and obtain patterns accordingly. Multi objective problem was addressed in “Attribute Index and Uniform Design Based Multi objective Association Rule Mining with Evolutionary Algorithm” a paper written by Jie *et al.* (2013) rather than single objective using Pareto frontier that significantly reduced the consumption of time.

Based on the aforementioned techniques and method, discovering Analytic Association Rule Filtering (AARF) on multidimensional data streams is presented where focus is made on identifying the relationship between attributes of multidimensional data in nature. This in further filters independent attributes from data streams maintaining average precision level. To reduce the execution time, AARF uses the if-then strategy with minimal phase shift time on multidimensional test dataset. AARF also applies weight computing factor to derive the relationship and

obtain inferences using multidimensional samples. With the application of attribute independent criterion in AARF, filtering query traffic efficiency is enhanced. As a result, AARF reduces the execution time factor and attains the cost effective filtering system.

## ANALYTIC ASSOCIATE RULE FILTERING ON MULTIDIMENSIONAL DATA STREAMS

In this section, we present an analytic association rule filtering on multidimensional data streams and shortly summarize if-then strategy based filtering algorithm. This algorithm reduces the phase shift time and efficiently filters the attributes and we show how time based cost effective ratio can be obtained.

**Filtering operation on multidimensional data:** The ultimate goal of Analytic Associate Rule Filtering mechanism is to construct a time based cost effective filtering system. The filtering mechanism in the association rule mining administers the analytic independence criterion on multidimensional data. With this, AARF follows non-arbitrary analytic capacity to filter out the data streams with minimal phase shift time. The objective of multidimensional data stream using AARF mechanism is to reach the target class with higher filtering efficiency result. The filtering operation on the multi-dimensional data is represented in Fig. 1.

AARF mechanism uses multi dimensional data for filtering out different types of multi dimensional data streams. A dataset that consists of different multi dimensional time series data from the Japanese vowel multidimensional data set is fetched to perform the filtering operation using AARF mechanism. The dataset consists of number of training and test samples for experimental evaluation of AARF mechanism. Analytic Association rule represent discovered knowledge and describe a close relationship between frequent time series items in a database. Association rule mining discovers the interesting information with the help of if-then strategy using AARF mechanism.

**Design considerations of AARF mechanism:** Let us assume that ‘S’ contains support rule for filtering multidimensional data using AARF mechanism with two chosen attributes, ‘x’ and ‘y’. If the attribute ‘x’ is satisfied with the rule and the verification made on the ‘y’ attributes is not matched, then the data stream is filtered out using the AARF mechanism. Followed by this, the Confidence ‘C’ in AARF mechanism is formulated to identify the probability satisfying ‘x’ and ‘y’ attributes:

$$Confidence (C (x \rightarrow y)) = \frac{S(xy)}{S(x)} \quad (1)$$

The confidence value  $C$  identifies the dependency level  $x \rightarrow y$  on multidimensional data which is obtained by dividing  $S(x)$ , the support rule of the attribute ‘x’

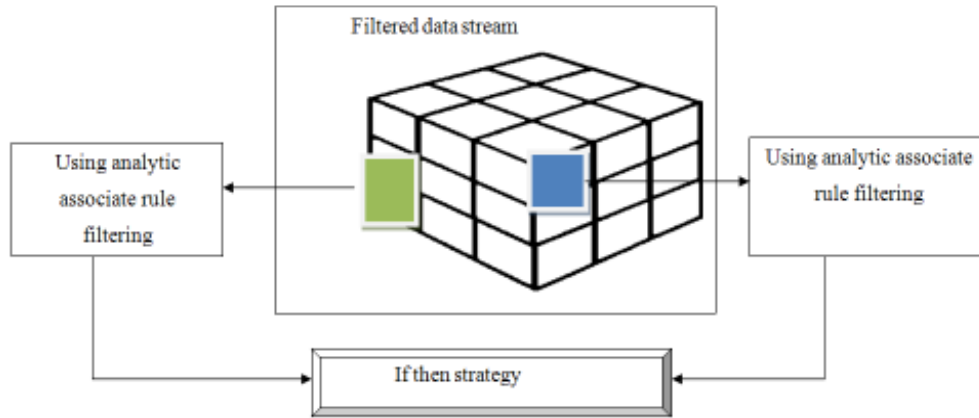


Fig. 1: Filtering operation on multi-dimensional data

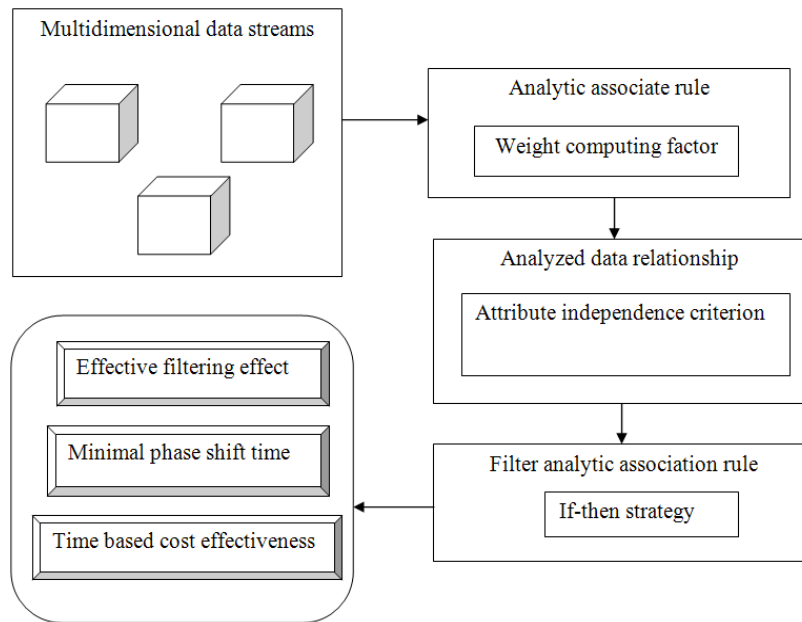


Fig. 2: Architecture diagram of AARF mechanism

and  $S(xy)$ , the relative form of attribute ‘x’ and ‘y’. In the context of the analytic association rule, the confidence value coincides with the analytic precision rule and hence AARF attains the higher precision ratio.

In AARF mechanism, filtering solution quantifies the result with effective analytical result. An association rule in AARF ( $x \rightarrow y$ ), the confidence and support rule is employed to attain minimal phase shift time based cost effective filtering system. The analytical  $\pi(C)$  and  $\pi(S)$  provide accuracy filtering result for all association rule mining on the given multidimensional data items. AARF gives the exact probability result over ‘i’ items with an association rule which is drawn at non-arbitrary distribution of data. Analytic Association Rule mining based Filtering algorithms guarantees lesser phase shift time for filtering multidimensional data. The architecture diagram of AARF mechanism is described in Fig. 2.

The overall design consideration of AARF mechanism is explained with the help of an architecture diagram as shown in Fig. 2, where multidimensional data streams are taken for association rule mining. Initially, analytic association rule is performed to identify the relationship between data attributes using the weight computing factor in AARF mechanism to measure the weight for each data attributes. The measured weight factor is used on the second part of AARF processing.

In order to remove the independent attributes from multidimensional data stream, the AARF mechanism analyze data relationship and employs attribute independence criterion which removes the negligible weight factor as they are independent to the data stream. Finally, to filter out the irrelevant data attributes with minimal phase shift time, the if-then strategy is designed using AARF mechanism. The if-then strategy

reduces the phase shift time on multidimensional data stream so that the cost factor is reduced based on the phase shift time in AARF mechanism.

**Design of analytic associate rule:** Analytic association rule procedure initially measures the support factor of the frequent time series items 'T'. The support factor of the analytic association rule is defined as given below:

$$S(T) = \frac{c(T)}{N} \tag{2}$$

The support factor is obtained using the confidence value  $C(T)$  of the frequent time series items 'T' over 'N' dimensional data points. The confidence value of identified data relationship uses polynomial distribution function in AARF mechanism. The polynomial distribution function for 'N' data points is computed as:

$$\text{Polynomial Distribution Function (PD)} = p[x + (1 - p)y]^N \tag{3}$$

The polynomial coefficient 'p' on data attribute 'x' and 'y' respectively is given in Eq. (3). The polynomial distribution is effectively computed using multidimensional data streams. The weighted computing factor in analytic association rule is multiplied with the polynomial distribution factor to measure the relationships. The analytic association rule form with weight factor is computed as:

$$\text{Analytical Rule Weight Factor} = \frac{\arg \max \sum_N \text{weight} (d_1, d_2, d_3 \dots d_n) * \max \text{value [PD]}}{\dots} \tag{4}$$

The association rule with weight factor is computed and then the relationship between the data is measured with higher precision rate with  $d_1, d_2, \dots, d_n$  data streams using multidimensional data streams. The max value of the polynomial distribution [PD] is obtained to attain the weight factor value using AARF mechanism.

**Attribute independence criterion:** The attribute independence criterion is developed in AARF mechanism to remove the irrelevant attributes from the data streams  $\{d_1, d_2, \dots, d_n\}$  as illustrated in the Fig. 3 with an association rule where the attribute 'K' is independent of the attributes 'x' and 'y' respectively. The independent attribute criterion removes the irrelevant attributes in multidimensional data streams.

The attribute independency and associative attributes are represented through Fig. 3. The data points are scattered in multidimensional data and support count is measured using the analytic associative attributes. The independent criterion is removed (i.e.,) filtered from multidimensional data stream:

$$\left. \begin{aligned} S &= x \cup y \\ S(xyK) &= S(xK)S(yK) \\ S(xyK) &= K[S(x)S(y)] \end{aligned} \right\} \tag{5}$$

Attribute independent is separated from multidimensional data stream in ARFF mechanism. 'S' denotes the support count on the attributes 'x' and 'y'. The 'K' is the independent attribute from the 'x' and 'y' attributes of the data stream. The Support vector between the attributed are computed and analyzed with the help of analytic association rule relationship. The small range of weight value on associating different data points denotes that the particular attribute is independent from all other attributes. The ARFF mechanism removes that independent attribute criterion from multidimensional attributes to improve the true positive rate.

**Design of If-then strategy in AARF:** AARF mechanism uses the if-then strategy to reduce the phase shift time. The if-then strategy is a powerful technique to filter the attributes on the multidimensional data streams. The if-then strategy based filtering algorithm is described as:

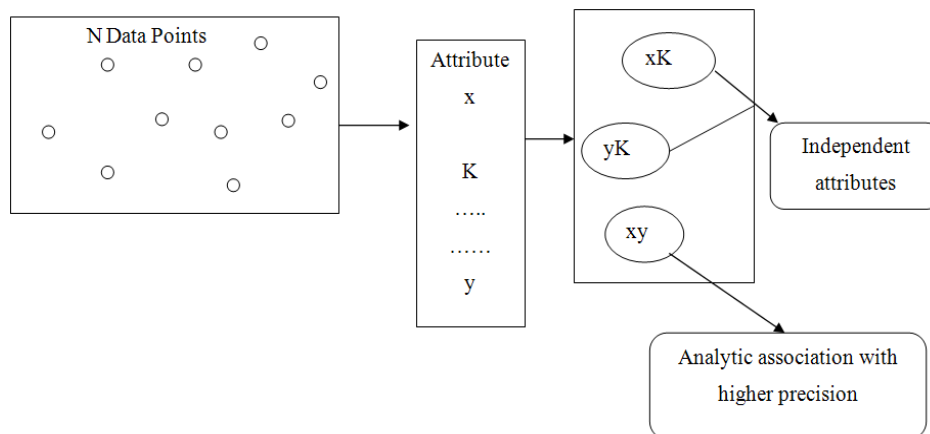


Fig. 3: Representation of attribute independence criterion

Table 1: Comparison of average precision level

| Dimensions | Average precision level (%) |                      |                   |
|------------|-----------------------------|----------------------|-------------------|
|            | AARF mechanism              | PaDSkyline framework | SPAMINE algorithm |
| 1          | 65                          | 61                   | 58                |
| 2          | 72                          | 68                   | 62                |
| 3          | 75                          | 70                   | 65                |
| 4          | 71                          | 69                   | 67                |
| 5          | 76                          | 72                   | 68                |
| 6          | 78                          | 74                   | 70                |
| 7          | 82                          | 75                   | 71                |

// If-then Strategy Procedure

Input: 'N' multi-dimensional data points with set of attributes (x, y, K)

Output: Filter irrelevant attributes from the data set

**Step 1** : Initial iteration = 0

**Step 2** : Analytic Association rule with (x and y) attributes

**Step 3** : For each attributes

**Step 4** : Calculate weight computing factor

**Step 5** : If (weight computing factor <negligible point value)

**Step 6** : Then, Remove the Independent attribute factor 'K'

**Step 7** : Else

**Step 8** : Associated attributes on multidimensional points

**Step 9** : End if else

**Step 10** : End for

The above if-then strategy based filtering algorithm removes the independence attributes from multiple phase data. The multiple phase shift data points are checked using the if-then strategy iterations. The iteration continues till all the phase shift information is executed using the if-then strategy. The reduced phase shift time in the AARF mechanism, reduces the time based cost effectiveness on the multi-dimensional data.

## EXPERIMENTAL EVALUATION

Analytic Associate Rule Filtering (AARF) mechanism developed on the multi dimensional data streams using JAVA platform. The Java platform uses the Weka tool for the effective rule mining and filters the irrelevant attributes from the data streams. AARF mechanism uses the Japanese Vowels multi dimensional Data Set from UCI repository to perform the experimental work. Japanese Vowels multi dimensional dataset records 640 time series of 12 Linear Predictive Coding (LPC) cepstrum coefficients taken from nine male speakers.

The collected multidimensional data are used for filtering the irrelevant Japanese vowels attributes from data streams. For each utterance, the analysis constraint explains below contain 12-degree linear prediction analysis to get hold of discrete-time series with 12 LPC cepstrum coefficients. Each utterance by a speaker forms a time series whose period is in the range 7-29

and each position of a time series is of 12 features (i.e.,) coefficients. The total number of the time series is about 640. We used one set of 270 time series for training samples and the other set of 370 time series for testing samples.

The frame length of the speaker's speech is about 25.6 ms and shift length of about 6.4 ms in Japanese Vowels multi dimensional Data Set. AARF mechanism is compared against the existing PaDSkyline framework and Similarity- Profiled temporal Association MINING mEthod (SPAMINE) algorithm. The experiment is conducted on the factors such as average precision level, execution time, filtering query traffic efficiency and true positive rate.

The average precision level *APL* in AARF indicates how much attributes returned from multidimensional data streams are of significant purpose given below:

$$APL = \left( \frac{Relevant\ Datastreams + Retrieved\ Datastreams}{Retrieved\ Datastreams} * 100 \right) \quad (6)$$

Execution time *ET* in AARF records the overall query processing time, from the moment when a multidimensional data stream is given as input to the moment when the final result is obtained is given as below:

$$ET = Time\ to\ process\ (d_1, d_2, \dots, d_n) \quad (7)$$

Filtering query traffic efficiency measures the number of multiple phase shift data points sent used during the extraction of interesting information. True positive rate in AARF measures the proportion of actual positives where the relevant data streams are identified correctly:

$$TPR = \left( \frac{Relevant\ Datastreams\ Identified + Relevant\ Datastreams\ Not\ Identified}{Relevant\ Datastreams} * 100 \right) \quad (8)$$

## RESULTS ANALYSIS OF AARF

The Analytic Associate Rule Filtering (AARF) mechanism is compared against the existing

PaDSkyline framework proposed by Lijiang *et al.* (2011) and Similarity- Profiled temporal Association MINing mEthod (SPAMINE) algorithm (POM-WIG) written by Jin and Shashi (2009). The experimental results using JAVA are compared and analyzed with the help of table and graph given below. Table 1 shows the simulation results of average precision level. Comparison is made with existing two methods, PaDSkyline and SPAMINE algorithm.

In Fig. 4, we observe that our mechanism AARF perform better than the existing works PaDSkyline framework developed by Lijiang *et al.* (2011) and Similarity- Profiled temporal Association MINing mEthod (SPAMINE) algorithm written by Jin and Shashi (2009) for different dimensions in the range of 1, 2, 3, ....7 respectively. This is because our mechanism considers the weight computing factor to observe the relationship and arriving at inferences while testing the multidimensional samples by improving the average precision level by 2-8% when compared to PaDSkyline framework introduced by Lijiang *et al.* (2011) Furthermore, AARF mechanism is guaranteed to find an optimal average precision level for each dimension in terms of cost effective filtering system. Our strategy based on analytic association rule can increase the average precision level significantly for AARFs by improving analytic precision rule by increasing the average precision level by 18-35% when compared to SPAMINE algorithm proposed by Jin and Shashi (2009). While average precision level increases slowly when dimension increases, it decreases sharply at dimension 4. When the dimension observed is of 7 almost all relevant streams are identified successfully and all the methods exhibit similar performance.

Table 2 and Fig. 5 depict the results of execution time. It can be observed that the time taken to execute using the AARF mechanism is comparatively lesser when compared to two other methods PaDSkyline framework introduced by Lijiang *et al.* (2011) and Similarity-Profiled temporal Association MINing mEthod (SPAMINE) algorithm introduced by Jin and Shashi (2009). Though with the increasing number of items, the execution time also gets increased in all the three methods, comparatively execution time is lesser using AARF mechanism. This is because of the application of analytic filtering with minimal phase

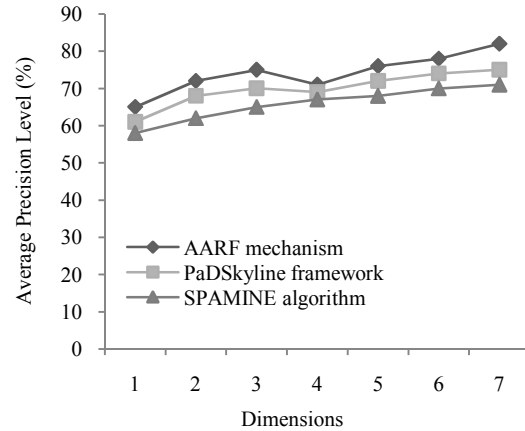


Fig. 4: Average precision level for different dimensions

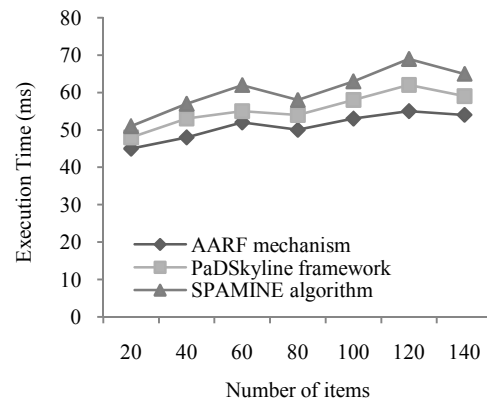


Fig. 5: Number of items versus execution time

shift time on multidimensional test dataset. The minimal phase shift time in turn further reduces the execution time factor by a rate of 6-12% when compared to PaDSkyline framework introduced by Lijiang *et al.* (2011). Moreover, with the introduction of if-then strategy iterations, multiple phase shift data points are checked that reduces the time based cost effectiveness on the multidimensional data by 46-56% than SPAMINE algorithm proposed by Jin and Shashi (2009).

Table 3 demonstrates the filtering query traffic efficiency of AARF mechanism, PaDSkyline framework and SPAMINE algorithm in MANET.

Table 2: Comparison of execution time

| Number of items | Execution time (ms) |                      |                   |
|-----------------|---------------------|----------------------|-------------------|
|                 | AARF mechanism      | PaDSkyline framework | SPAMINE algorithm |
| 20              | 45                  | 48                   | 51                |
| 40              | 48                  | 53                   | 57                |
| 60              | 52                  | 55                   | 62                |
| 80              | 50                  | 54                   | 58                |
| 100             | 53                  | 58                   | 63                |
| 120             | 55                  | 62                   | 69                |
| 140             | 54                  | 59                   | 65                |

Table 3: Comparison of filtering query traffic efficiency

| Dimensions | Filtering query traffic efficiency (data points) |                      |                   |
|------------|--|----------------------|-------------------|
|            | AARF mechanism                                   | PaDSkyline framework | SPAMINE algorithm |
| 1          | 640  | 655                  | 685               |
| 2          | 660  | 682                  | 690               |
| 3          | 675  | 695                  | 710               |
| 4          | 650  | 665                  | 695               |
| 5          | 690  | 710                  | 712               |
| 6          | 705  | 720                  | 735               |
| 7          | 712  | 725                  | 754               |

Table 4: Comparison of true positive rate

| Number of items | True positive rate (%) |                      |                   |
|-----------------|------------------------|----------------------|-------------------|
|                 | AARF mechanism         | PaDSkyline framework | SPAMINE algorithm |
| 20              | 60                     | 56                   | 53                |
| 40              | 65                     | 63                   | 57                |
| 60              | 70                     | 65                   | 60                |
| 80              | 66                     | 64                   | 62                |
| 100             | 71                     | 67                   | 63                |
| 120             | 73                     | 69                   | 65                |
| 140             | 75                     | 70                   | 66                |

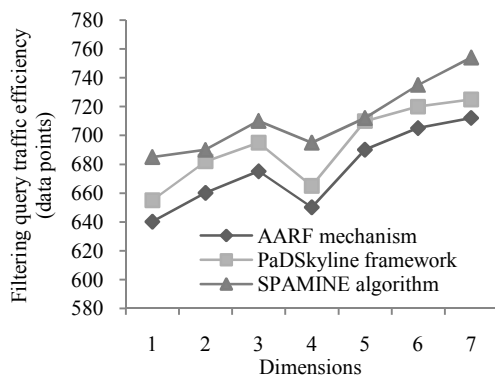


Fig. 6: Dimensions versus filtering query traffic efficiency

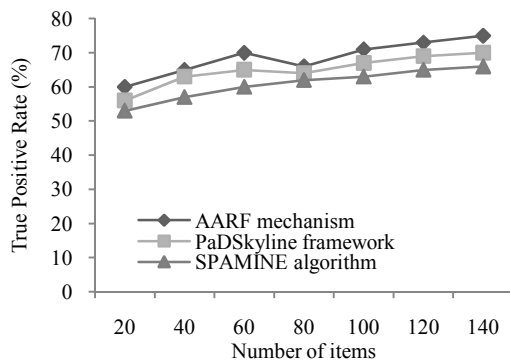


Fig. 7: Number of items versus true positive rate

In Fig. 6, the performance of filtering query traffic efficiency is considered with different number dimensions in the range of 1 to 7. Algorithm if-then strategy is used to filter the analytic association rule with specified phase shift time. It can be observed that when graph connectivity is high with respect to the dimensions, the performance increases in terms of

filtering query traffic efficiency and attains cost effective filtering system. This is because when graph connectivity is high in terms of dimension, the AARF mechanism uses attribute independent criterion to discard negligible weight from association rule by improving the filtering query traffic efficiency by 1-3% than the PaDSkyline framework introduced by Lijiang *et al.* (2011). Moreover, with the measure of multiple phase shift data points, the query traffic efficiency is improved by 40-60% when compared to SPAMINE algorithm by Jin and Shashi (2009).

A comparative measure of true positive rate is made in Table 4 on the basis of the number of items in the multidimensional data stream. The true positive rate is measured in terms of %.

Figure 7 depicts the true positive rate with respect to the number of items in multidimensional data stream. From the Fig. 7 it is illustrative that the true positive rate using AARF mechanism is higher than the other two methods. This increase in rate is due to the fact that the analytical  $\pi(C)$  and  $\pi(S)$  results in accurate filtering result for given multidimensional data items with all association rule mining increasing the true positive rate by 3 to 7% when compared to PaDSkyline framework introduced by Lijiang *et al.* (2011). In addition, the minimum weight value with different associated data points proves that the specific attribute is independent from all other attributes. This in turn makes the independent attribute to be removed from multidimensional attributes by improving the true positive rate by 25- 40% than SPAMINE algorithm proposed by Jin and Shashi (2009).

## LITERATURE REVIEW

The database society has observed highest contribution in research due to its importance in

different applications including data cleaning, evaluation of radio frequency, services based on location analyzing sensor data and soon. In “Effectively Indexing the Multi-Dimensional Uncertain Objects for Range Searching” Ying *et al.* (2011) proposed a novel type of indexing method using U-Quadtree was presented in multidimensional space to identify the uncertain objects in such a way that can be answered in an efficient manner using filtering techniques. Though effectiveness and efficiency was improved using indexing techniques, correlation related to multidimensional uncertain objects remains unaddressed. Lei and Lei (2011) proposed, analytic cost model that included monotone aggregate function based on the integration of top-k problem and dominant relationships. The indexing structure designed proved to be a cost efficient model for skylines record set. However dominant relationship analysis was unaddressed.

In many applications, multivariate dynamic data system was presented that was highly complex in nature. Multivariate Reconstructed Phase Space (MRPS) was introduced by Wenjing and Xin (2011) identified multivariate temporal patterns that were used to predict anomalies in dynamic range of system. This worked as an enhancement to univariate framework was designed on the basis of fuzzy unsupervised clustering method whereas MRPS used a novel method of categorization of data based on event definition.

Dominik *et al.* (2011) proposed a SwiftRule, in which a new model was designed for mining temporal types of data based on the classification rules understood by human experts using polynomial models. Followed by this, the classifiers then assessed short sequences with their rule premises for effective segmentation. But the model was not unaddressed for wider domain area. To improve the effectiveness of association rule mining techniques, an extended and generalized fuzzy technique was presented by Vivek *et al.* (2014) using fuzzy, rough, soft and vague set theories. In “An efficient algorithm for incremental mining of temporal association rules” written by Tarek *et al.* (2010) temporal association rules was introduced for providing solutions to temporal data involving time series data with the association rules. An incremental algorithm was also used to reduce to time taken to produce candidate item sets.

In many web scenarios, uncertainty is one of the important factors to be addressed. In web applications where uncertain database are involved many researchers have contributed on threshold queries based on probabilistic where all the results that satisfies the queries with the probability higher or equal to the threshold values were considered. Probabilistic Threshold Keyword Queries (PrTKQ) using XML data, was first introduced with the consideration of possible world semantics by Jianxin *et al.* (2013). Followed by

this, a Probabilistic Inverted (PI) index was then used to return the answers that were more suitable and also filtered out the unrelated ones using the two bounds either lower or upper bounds. But the time involved was too high and security was unaddressed.

A protocol called as the multi-party protocol was presented in “Secure Mining of Association Rules in Horizontally Distributed Databases” written by Tamir (2014) in order to mine the association rules in a secure manner present in horizontally distributed databases which was based on the Fast Distributed Mining (FDM) algorithm. The main portions included in multi-party protocol was that it evaluated the union of private subsets and integrated with the inclusion of an element that was held by a player in a sub-portion by another player that enhanced the level of security but was not efficient in terms of record matching. To address the problem related to record matching in the Web scenario, an unsupervised, an online record matching method was introduced by Weifeng *et al.* (2010) in “Record Matching over Query Results from Multiple Web Databases” that identified the duplicates present in the query in an efficient manner from multiple Web databases by improving the precision and recall.

Classification has been widely applied in many types of multidimensional data sets. But classification of road networks using multidimensional data sets has received lesser amount of research efforts. Jae-Gil *et al.* (2011) proposed classification of road networks and studied the same by observing the behavior of trajectories, along with the places where the nodes were visited in order to improve the classification accuracy. Classifications based on pattern for more sophisticated type involving multidimensional datasets remained unaddressed.

## CONCLUSION

In this study, we addressed the problem of obtaining minimal phase shift based cost effective filtering on the multidimensional data stream. We analyzed the impact of relationship of attributes on multidimensional data and filtering of independent attributes from data streams and filtering efficiency significantly. We have proposed and analyzed the algorithm if-then strategy based filtering algorithm for each attributes to reduce the phase shift time and to improve the overall performance. The time complexity of our algorithms is shown to reduce the time based cost effectiveness of multidimensional data and can filter the attributes in a more efficient manner. While if-then strategy based filtering algorithm is guaranteed to filter the analytic association rule with specified phase shift time, it was shown that AARF mechanism can reduce the execution time factor and results in cost effective filtering system. Through extensive simulation experiments, we observed that our mechanism AARF



performs better than PaDSkyline framework and SPAMINE algorithm for multidimensional data in terms of true positive rate and filtering efficiency. Furthermore, the execution time using multidimensional data is improved significantly in AARF.

## REFERENCES

- Abhijith, K., H. Vagelis, P. Michalis and T. Sotiria, 2011. Effective navigation of query results based on concept hierarchies. *IEEE Trans. Knowl. Data Eng.*, 23(4): 540-553
- Dominik, F., G. Thiemo and S. Bernhard and R. Swift, 2011. Mining comprehensible classification rules for time series analysis. *IEEE Trans. Knowl. Data Eng.*, 23(5): 774-787.
- Jae-Gil, L., L. Xiaolei and C. Hong, 2011. Mining discriminative patterns for classifying trajectories on road networks. *IEEE Trans. Knowl. Data Eng.*, 23(5): 713-726.
- Jianxin, L., L. Chengfei, Z. Rui and X.Y. Jeffrey, 2013. Quasi-SLCA based keyword query processing over probabilistic XML data. *IEEE Trans. Knowl. Data Eng.*, 26(4): 957-969.
- Jie, Z., W. Yuping and F. Junhong, 2013. Attribute index and uniform design based multiobjective association rule mining with evolutionary algorithm. *Hindawi Publishing Corporation, Scient. World J.*, Article ID 259347.
- Jin, S.Y. and S. Shashi, 2009. Similarity-profiled temporal association mining. *IEEE Trans. Knowl. Data Eng.*, 21(8): 1147-1161.
- Lei, Z. and C. Lei, 2011. Pareto-based dominant graph: An efficient indexing structure to answer top-K queries. *IEEE Trans. Knowl. Data Eng.*, 23(5): 727-741.
- Lijiang, C., C. Bin and L. Hua, 2011. Constrained skyline query processing against distributed data sites. *IEEE Trans. Knowl. Data Eng.*, 23(2): 204-217.
- Ozgur, C. and E.A. Murat, 2012. A recommendation engine by using association rules. *Proceedings of the Social and Behavioral Sciences, Elsevier*, pp: 452-456.
- Shweta, T. and K.B. Kamal, 2012. Enhanced new user recommendations based on quantitative association rule mining. *Proceedings of the Computer Science, Elsevier*, October, pp: 102-109.
- Tamir, T., 2014. Secure mining of association rules in horizontally distributed databases. *IEEE Trans. Knowl. Data Eng.*, 26(4): 970-983.
- Tarek, F.G., N. Hamed, T. Mohamed and A.T. Ajith, 2010. An efficient algorithm for incremental mining of temporal association rules. *Data Knowl. Eng.*, Elsevier, 69(8): 800-815.
- Victoria, N. and B. Rafael, 2011. Finding association rules in semantic web data. *Proceedings of the Knowledge-Based Systems, Elsevier*, May, pp: 51-62.
- Vivek, B., R.S. Thakur and G.S. Thakur, 2014. A review on dealing uncertainty, imprecision and vagueness in association rule mining using extended and generalized fuzzy. *Int. J. Eng. Comp. Sci.*, 03(07): 6963-6966.
- Weifeng, S., W. Jiying and H.L. Frederick, 2010. Record matching over query results from multiple web databases. *IEEE Trans. Knowl. Data Eng.*, 22(4): 578-589.
- Wenjing, Z. and F. Xin, 2011. Event characterization and prediction based on temporal patterns in dynamic data system. *IEEE Trans. Knowl. Data Eng.*, 26(1): 144-156.
- Xu, H., M. Fan and Z. William, 2013. Parametric rough sets with application to granular association rule mining. *Hindawi Publish. Corporat., Math. Problems Eng.*, 2013, Article ID: 461363.
- Yang, O., J.L. Zheng, R.K. Hamid and T. Ying, 2014. Multilevel association rule mining for bridge resource management based on immune genetic algorithm. *Hindawi Publish. Corporat. Abstract Appl. Anal.* 2014, Article ID: 278694.
- Ying, Z., Z. Wenjie, L. Qianlu and L. Xuemin, 2011. Effectively indexing the multi-dimensional uncertain objects for range searching. *Proceedings of the 15th International Conference on Extending Database Technology, New York*, pp: 504-515.