



Cyclic Repeated Patterns in Sequential Pattern Mining Based on the Fuzzy C-Means Clustering and Association Rule Mining Technique

Ramani Selvanambi*

Jaisankar Natarajan¹

¹*Vellore Institute of Technology University, Vellore, India*

* Corresponding author's Email: ramani.s@vit.ac.in

Abstract: The main aim of the proposed method is to remove cyclic repeated patterns in sequential pattern mining. Initially the input dataset is fed to the clustering process, in which fuzzy c means clustering algorithm is used to cluster the available data based on the similar sequential pattern. This approach is able to mine the patterns with the help of association rule mining, here two major tasks are present one is frequent item set generation and rule generation. In frequent item set generation, support and confidence value is evaluated for each pattern. Based on that, the rules are generated in rule generation. After the mining process, the threshold is fixed and based on that repeated cyclic patterns are removed and stored to the database. The performance of the suggested method is evaluated by means of execution time, memory and database difference ratio. The implementation is done with the help of JAVA platform.

Keywords: Sequential pattern mining, Fuzzy C-means clustering, Association rule mining, Frequent item set generation, Rule generation.

1. Introduction

Determining sequential patterns from an enormous database of categorization is an important problem in the field of knowledge discovery and data mining [1]. Sequential pattern mining is a data mining skill utilized to recognize patterns of well-ordered events [2]. There have been many fresh research works on data mining arena on determining interesting but unknown knowledge from a large amount of data. The data mining systems comprise association rules mining, classification, clustering, mining time sequences, and sequential pattern mining, to name a limited [3]. Sequential pattern mining that discovers recurrent subseries as designs in a series database is an important data mining problem with broad applications, including the analysis of user purchase patterns or web access patterns, the investigation of sequencing or time related processes such as natural disasters, scientific experiments, and disease treatments, the analysis of DNA sequences, etc [4]. Many sequential pattern mining approaches have been studied, such as overall sequential pattern mining, closed sequential

pattern (Clops) mining, maximal sequential pattern mining and interesting sequential pattern mining [5].

This domain is now mature and offers efficient techniques for abstracting such patterns. A typical base of sequences is a set of sequences of discrete events, in that each system has a sole sequence identifier [6]. In particular, closed sequential pattern mining has become adynamic topic in data mining community, as it is a compressed yet lossless compression of sequential patterns [5]. Frequent Sequential Patterns (FSPs) from deterministic databases has attracted a lot of attention in the research community because of its wide spectrum of real life implementations [7]. Sequential patterns are characterized by their support or frequency, which indicates the proportion of sequences in a database that comprises this pattern [8].

Sequential pattern mining is a critical data mining technique for decisive time-related behavior in series databases. In sequential pattern mining, a specific set of sequences, called data series, are utilized as input data. All data systems are a gradient of transactions, and each transaction encompasses a group of literals called items [9]. In business applications, whereas customers have long-term

relationships with a retailer, it becomes valuable to comprehend customers' repeated purchase behavior [10]. Original sequential pattern mining model only deliberates occurrence frequentness of sequential patterns, but disrespects their incidence periodicity [11]. In numerous implementations, some items appear recurrently in the data, although others appear infrequently. When the frequencies of items fluctuate a great deal, we will encounter the dilemma known as the rare item problem. If minimum support is fixed too high, we will not find those sequential patterns that comprise rare items in the information [12]. The overall objective of the suggested technique is removing the cyclic repeated patterns in sequential pattern mining. In order to attain the objective effective clustering and rule mining technique is examined in our proposed technique. The main advantage of the proposed technique is removing repeated pattern in an effective manner when compared with existing technique.

The remaining of this paper is organized as follows. Section 2 gives some brief background of researches related to the proposed technique. The background of the research is explained in section 3. Section 4 describes the proposed sequential pattern mining. The experimental results and discussions of the proposed approach are presented in Section 5. Finally, conclusions are summarized in Section 6.

2. Literature Survey

Ya-hanhu, et al. [13], has proposed cyclically repeated patterns (CRP) mining, in which a new parameter called repetition support, is considered in the mining process. In the paper, they have first redefined CRPs based on the MIR and original form of the sequence minimum support. A new algorithm, rep-prefix span, was developed for discovering a complete set of CRPs in sequence databases.

In X. Leslie R. Foulds, et al. [14], they have offered an alteration that permits any illustration of the irregular to be resolved as a kNN issue. Though that permits the implementation of any kNN algorithm, the alteration was too time overpriced to be practical for examples of practical dimensions. For that aim, they presented a condensation algorithm for the effective eradication of unfavorable training points and an operative parallel programming method on the basis of the discrete Fourier transform.

In this article, Zhou Zhao, et al. [15], they have projected to portion pattern frequentness on the basis of the imaginable world semantics. They have recognized two indeterminate sequence information

models abstracted from numerous real-life implementations linking uncertain series data, and express the issue of mining probabilistically frequent sequential patterns from information that follow to our models. U-Prefix Span efficiently evades the issue of "possible world's explosion", and if associated with their four pruning and authenticating approaches, attains even better performance. They have also offer a fast validating technique to additional speed up our U-Prefix Span algorithm.

In this article, Ozgur Kirmemis Alkan, et al. [16], have anticipated effective informational structures and pruning method that was on the basis of Cumulated Rest of Match (CRoM) based upper bound. CRoM, by describing a tighter upper bound on the usefulness of the candidates, permits more conservative pruning before candidate pattern generation in association with the available methods. Additionally, they have industrialized an effectual algorithm, HuspExt that computes the utilities of the child patterns on the basis of that of the parents.

Jerry Chun-Wei Lin¹, et al. [17], has proposed among the discovered knowledge, sequential-pattern mining was used to discover the recurrent subsequences from a sequence database. In the paper, the pre-large concept was adopted to handle the discovered sequential patterns with sequence deletion. An FUSP tree was RST built to keep only the frequent 1-sequences from the original database. The pre-large 1-sequences were also kept in a set for later maintenance approach.

Cluster-based Temporal Mobile Sequential Pattern Mine was projected by Eric Hsueh-Chan Lu et al [18]. In CTMSP-Mine, user clusters were fabricated with the help of a new algorithm called Cluster-Object-based Smart Cluster Affinity Search Technique (CO-Smart-CAST) and resemblances within consumers were assessed by the projected measure, Location-Based Service Alignment. In the meantime, a time segmentation method was offered to detect segmenting time intervals where alike mobile features exist. To their best knowledge, that was the principal work on mining and prophecy of mobile behaviors with deliberations of user relations and temporal property concurrently. Through experimental assessment under numerous simulated circumstances, the projected approaches were publicized to offer exceptional presentation.

In Jiaqi Zhu et al [19], they have projected Sequential Topic Patterns and express the issue of mining User-aware Rare Sequential Topic Patterns in document streams on the Internet. They were infrequent on the entire but comparatively frequent for precise consumers, so could be implemented in

numerous real-life situations, namely real-time monitoring on irregular consumer behaviors.

P. F. Viger et al. [20] have presented two algorithms. Rule Growth is a novel algorithm for mining sequential rules common to multiple sequences. The second algorithm (TRule Growth) allows the user to specify a sliding-window constraint on rules to be mined. Furthermore experiment publicized that the execution time and the number of valid rules found could be abridged by numerous orders of magnitude if the window size constraint was utilized.

From the literature survey, they mainly focused the sequential pattern mining with the help of prefix span algorithm. The main drawbacks of this algorithm is, it scans only the prefix subsequences and projects their corresponding postfix subsequences into the databases. And the main problem is, it is divided progressively into a series of sub-problems which are not intersecting. So that the suggested technique is using the effective clustering and rule mining technique for sequential pattern mining process.

3. Motivation

In this section we briefly discussed the problems identified in existing methods.

- In the existing method, contributed the main idea of Prefix span algorithm is based on the divide-and-conquer strategy.
- The problem of asynchronous periodic sequential patterns mining is divided progressively into a series of sub-problems which are not intersecting. So we recommend an association rule mining system in our anticipated study.
- The foremost idea of the projected technique is protracted from the association rule mining

algorithm a dissimilar issue is lectured in the arena of high utility sequential pattern abstraction.

- Association rule mining is a key strategy. However, because big amounts of association rules which is often from time to time simple.

In order to overcome these drawbacks, we have used an associative rule mining technique in our proposed study.

4. Proposed Approach

In sequential pattern mining, the support of the sequential pattern for the real life application database is defined only by the fraction of the supporting this sequence. Periodic patterns and cyclic patterns in sequence databases and the cyclic pattern mining, which considers a new parameter, named repetition support, into the mining process. The chief goal of the projected method is to project a technique to filtering the repetition of the cyclic periodic patterns. The past studies mainly focused on the sequential pattern mining with the help of prefix span algorithm. The main drawbacks of prefix span algorithm are the construction cost is high so that the suggested technique uses fuzzy c means clustering algorithm for dividing or grouping the sequential pattern. The main advantage of fuzzy c means clustering algorithm is easy tackling of noisy data and also it efficiently deals with various type of variables. At first, the input datasets are cluster with the aid of Fuzzy C-Means clustering technique. Then find the cyclic patterns from the dataset subsequently those cyclic patterns are going to the association rule mining technique to find the rules. Then calculate the support value for each rules, if the support value is greater than the threshold value the patterns are removed otherwise store the database. The overall block diagram of the proposed work is shown in fig.1 and the detail process of our proposed technique is illustrated in further section.

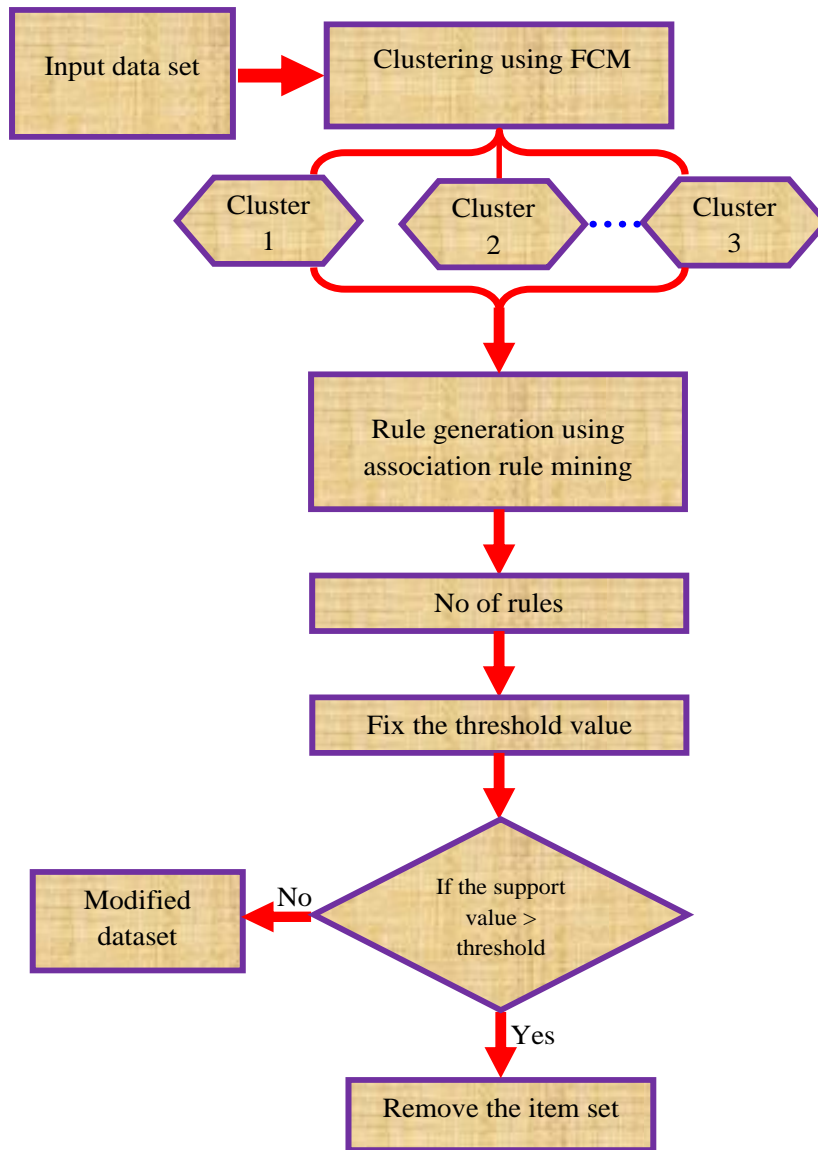


Figure.1 The overall block diagram of proposed method

Initially input data set is given to the clustering process. In our proposed method fuzzy c means clustering algorithm is used for clustering the input data set. The detail explanation of the fuzzy c means clustering algorithm is given below,

4.1 Fuzzy C means Clustering Algorithm (FCM)

Fuzzy C-means (FCM) is a data clustering technique wherein each data point belongs to a cluster to some degree that is specified by a membership grade. It provides a method of how to group data points that populate some multidimensional space into a specific number of different clusters. The main advantage of fuzzy c means clustering is that it allows gradual memberships of data points to clusters measured as degrees in [0,1]. This gives the flexibility to express

that data points can belong to more than one cluster. The proposed method uses fuzzy c means for clustering the input data. The objective function of proposed fuzzy c means algorithm is effectively explained as follows.

$$OF = \sum_{i=1}^n \sum_{j=1}^a M_{ij}^m \|d_i - a_j\|^2 \quad (1)$$

Where,
 “ M_{ij} ” is the membership of j^{th} data in the i^{th} cluster a_j
 “ a ” is the cluster center
 “ d_i ” is the input data
 “ m ” is the any real number greater than 1
 “ $\|*\|$ ” is the similarity between any measured data and the center

The pseudo code for fuzzy c means clustering algorithm is given below,

Pseudo code for fuzzy c means clustering algorithm

Input: Set of data points

Output: Clustered data set

Objective function: Obtain the clusters of the given data sets using equation (1)

Begin

Initialize the membership function,

$$M = [M_{ij}], M^{(0)}$$

Calculate the cluster center,

$$a^{(K)} = a_j, \text{ with } M^{(K)}$$

$$a_j = \frac{\sum_{i=1}^n M_{ij}^m d_i}{\sum_{i=1}^n M_{ij}}$$

Update the membership function,

$$M^{(K)} \text{ and } M^{(K+1)}$$

$$M_{ij} = \frac{1}{\sum_{K=1}^c \left(\frac{\|d_i - a_j\|}{\|d_i - a_K\|} \right)^{\frac{2}{m-1}}}$$

If $\|M^{(K+1)} - M^{(K)}\| < \epsilon$ then stop, “ ϵ ” is a termination criterion between 0 and 1.

otherwise calculate the cluster center,

end

Based on the above procedure of FCM the input data is clustered. After the FCM process, we obtain the number of cluster set such as $C_1, C_2, C_3, \dots, C_n$. Total number of data considered in our proposed work is 345. Here, we have fixed the cluster size as two; cluster one contains 114 datasets and second cluster contains 231 input data. This clustered dataset is used for the further processing. After the clustering process we have to generate the rule for each cluster. For generating the rules, the proposed technique employed the association rule mining. The clear explanation of association rule mining is described beneath,

4.2 Association Rule Mining

Association rule mining is an important component of data mining. Association rules are an important class of methods of finding regularities or

patterns in data. Two major tasks of association rule mining is recurrent item set generation and rule generation. **Frequent item set generation**, whose objective is to detect all the item sets that gratify the minimum support threshold; these item sets are called frequent item sets. **Rule generation**, whose objective is to abstract all the high confidence principles from the frequent item sets establish in the previous stage; these rules are known as strong rules.

In association rule mining, every rule must satisfy two users' specified constraints: one is measure of statistical significance called support and other is measure of goodness called confidence.

Support is a measure that defines the percentage or fraction of records or entries in the dataset contain $X \cup Y$ to the total number of records or entries.

$$Support(X \rightarrow Y) = P(X \cup Y) \tag{2}$$

Where X and Y are called as frequent item sets.

The support values are estimated for our separate product packages. An item packages along with support value below this minimum support value is usually taken away. This left product packages are usually picked. Further this picked product packages had been merged while using exact same product packages.

Confidence is a measure that defines the percentage or fraction of records or entries in the dataset that contain $X \cup Y$ to the total number of records or entries that contain just X. Using the formula, this confidence value can be determined.

$$Confidence(X \rightarrow Y) = \frac{P(X \cup Y)}{P(X)} \tag{3}$$

Where X and Y are called as frequent item sets.

To produce association guidelines, our typical item set made finally as well as the lowest confidence value is actually utilized. The complete item sets made have been in use. Applying solution (3), our confidence values for that decided on item sets have been computed. The product sets with confidence value greater than our lowest confidence value allotted usually are decided on. The actual left item sets usually are diminished. The actual decided on item sets will be the association guidelines. In association rule mining each cluster values are given to the input for rule generation, in which 145 rules and 171 rules are generated for each clusters. Finally we have to fix the threshold value for our proposed technique. If the desired value is greater than the threshold value we remove the rules from the input dataset and remaining item set or rules are stored in the modified dataset. The sample rules from association rule mining is shown in below table 1.

Here after we use the modified dataset for further process, in the modified dataset cyclic repeated patterns are removed. So that our proposed technique achieves the better result and also it saves the running time. The general performance of the projected system is appraised in section.5.

Table.1 Sample rules from association rule mining

Rule 1: 459 → 675 → 701 → 921 → 2: support 0.1
Rule 2: 675 → 701 → 459 → 921 → 2: support 0.1
Rule 3: 701 → 459 → 675 → 921 → 2: support 0.1
Rule 4: 459 → 675 → 921 → 701 → 2: support 0.1
Rule 5: 675 → 921 → 459 → 701 → 2: support 0.1
Rule 6: 921 → 459 → 675 → 701 → 2: support 0.1
Rule 7: 675 → 459 → 701 → 921 → 2: support 0.1
Rule 8: 459 → 701 → 675 → 921 → 2: support 0.1
Rule 9: 701 → 675 → 459 → 921 → 2: support 0.1
Rule 10: 675 → 459 → 921 → 701 → 2: support 0.1

5. Results and Discussion

The innovative cyclic repeated patterns in sequential pattern mining based on the fuzzy c-means clustering and association rule mining technique is performed and it is implemented using java.

Here we are considering mushroom dataset for our proposed work. The input dataset is taken from Frequent Item set Mining Dataset Repository. The mushroom dataset is available at <http://fimi.ua.ac.be/data/>. The experimental result of the proposed technique is described below.

5.1 Performance Evaluation Measures

The performance of our proposed method is evaluated based on execution time, cost and memory taken for execution process by varying the number of task or jobs in proposed approaches in detection process.

➤ Running Time

The running time in the Java program is squarely rooted on based on the memory values. The time complication is habitually represented in such a way as to ignore the coefficients and lower order terms and in most cases, the time duration is computed in millisecond (ms).

➤ Memory Usage

The Java program effectively organizes the memory for use. New objects are generated and positioned in the stack.

➤ Database Difference Ratio

The database difference ratio can be calculated by taking the difference ratio between modified database and the initial database.

5.2 Analysis Process

By varying the cluster size the proposed technique is evaluated the running time, memory value as well as the database difference ratio. The table 2 appearing below illustrates the running time, memory value and database difference ratio for the cluster size two. Here we are varying the support value and then find the corresponding run time, memory value and database difference ratio (DDR) of our proposed technique.

From analyzing the table.2, the proposed method achieves the run time by varying the support value. In our proposed technique, initially we fix the

support value is 0.1 in that time our suggested technique takes 2345841ms to complete the overall process. By varying the support value, the proposed run time also changes, to complete cyclic repeated patterns in sequential pattern mining by varying the support value is 0.2, 0.3 and 0.4 our suggested method takes 2644154ms, 3269954ms and 3841251ms respectively. The pictorial depiction of the run time for the innovative technique is furnished in Figure 2.

Table 2 represents the memory value by varying the support value in the innovative technique. The proposed method uses the memory value of 4569544 bits to complete the process with the support value of 0.1. The memory usage for completing the process with 0.2 support value is 4798961 bits. 4988874 bits and 5236934 bits memory space is needed to complete the support value of 0.3 and 0.4 respectively in the proposed cyclic repeated patterns in sequential pattern mining method. The pictorial depiction of memory value is given below. Here x axis represents the support value and y axis represents the memory usage value.

Table.2 Run time, memory value and DDR for cluster size two

Support value	Run time	Memory value	database difference ratio (DDR)
0.1	2345841	4569544	74.26
0.2	2644154	4798961	75.37
0.3	3269954	4988874	78.49
0.4	3841251	5236934	80.07

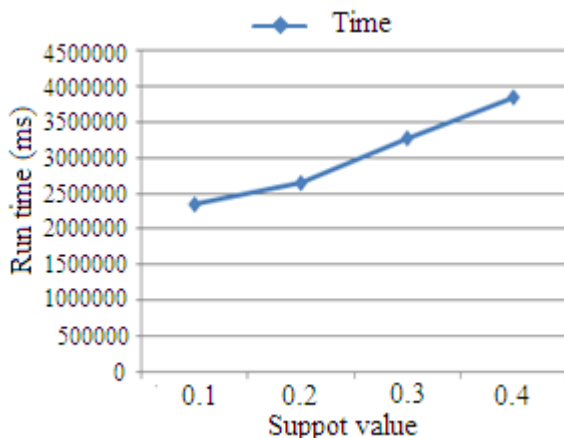


Figure.2 Run time for our proposed method by varying the support value (c=2)

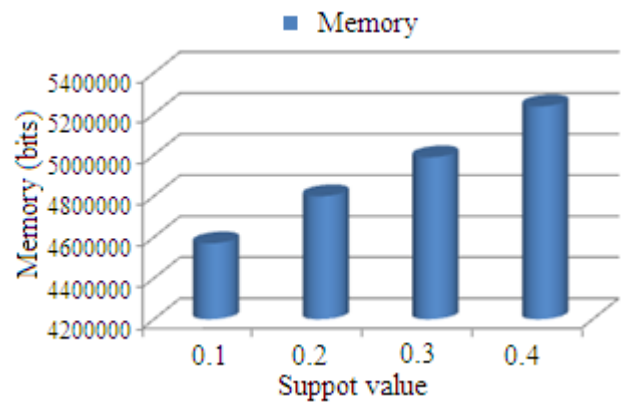


Figure.3 Memory value of our proposed method by varying the support value (c=2)

Table.3 Run time, memory value and DDR for cluster size three

Support value	Run time	Memory value	database difference ratio (DDR)
0.1	2136994	4136984	73.41
0.2	2546946	4469873	74.56
0.3	3026941	4698742	75.61
0.4	3654871	5136946	77.42

Table 3 shown above illustrates the running time, memory value and database difference ratio for the cluster size three.

When analyzing table 3, the proposed method achieves the run time by varying the support value. In our proposed technique, initially we fix the support value is 0.1 in that time our suggested technique takes 2136994ms to complete the overall process. By varying the support value, the proposed run time also changes, to complete cyclic repeated patterns in sequential pattern mining by varying the support value is 0.2, 0.3 and 0.4 our suggested method takes 2546946ms, 3026941ms and 3654871ms respectively. The pictorial depiction of the run time for the innovative technique is furnished below.

Table.3 represents the memory value by varying the support value in the innovative technique. The proposed method uses the memory value of 4136984 bits to complete the process with the support value of 0.1. The memory usage for completing the process with 0.2 support value is 4469873 bits. 4698742 bits and 5136946 bits memory space is needed to complete the support value of 0.3 and 0.4 respectively in the proposed cyclic repeated patterns in sequential pattern mining method. The pictorial depiction of memory value is given below.

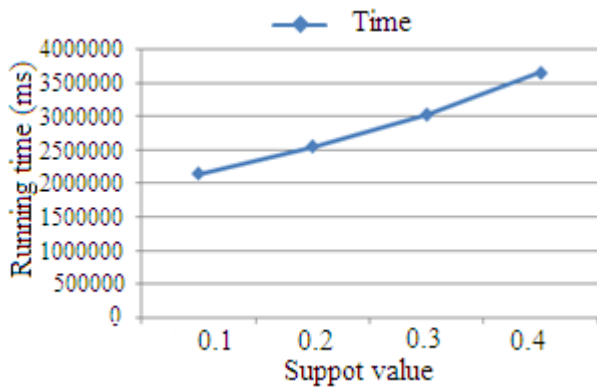


Figure.4 Run time for our proposed method by varying the support value (c=3)

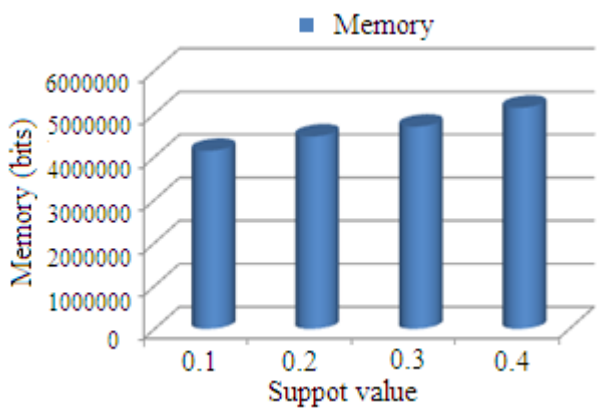


Figure.5 Memory value of our proposed method by varying the support value (c=3)

For cluster size two, we fix the support value is 0.1 our proposed method attains the database difference ratio is 74.26%. From the result 25.74% of repeated cyclic pattern is removed from the initial database. If the support value is vary from the 0.2, 0.3 and 0.4 our proposed technique achieves the database difference ratio is 75.37%, 78.49% and 80.07%. For cluster size three, we fix the support value is 0.1 our proposed method attains the

database difference ratio is 73.41%.

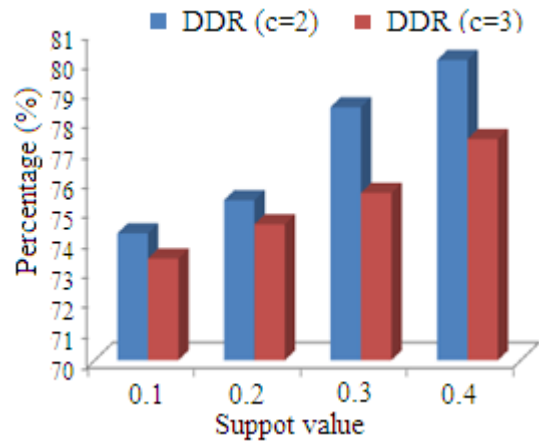


Figure.6 Database Difference ratio of proposed method by varying the support value

From the result 26.59% of repeated cyclic pattern is removed from the initial database. If the support value is vary from the 0.2, 0.3 and 0.4 our proposed technique achieves the database difference ratio is 74.56%, 75.61% and 77.42%. The pictorial representation of the DDR value is shown in fig.6.

5.3 Comparative analysis

In comparative analysis the proposed method is compared with existing technique, in order to prove that the proposed work is better. Here the overall execution time, memory and database difference ratio of the fuzzy c means clustering algorithm is compared with k means clustering algorithm and prefix span algorithm [14]. Here the overall execution time is represents in milliseconds and memory value is represents bits. The comparison result is tabulated in table.4; here the clustering size for comparison is two.

Table.4 Comparison result for proposed vs. existing method

Support value	Time (ms)			Memory (bits)			DDR (%)		
	FCM	k means	Prefix span [14]	FCM	k means	Prefix span [14]	FCM	k means	Prefix span [14]
0.1	2345841	2344641	2344721	4569544	4568339	4568258	74.26	74.06	74.12
0.2	2644154	2642654	2642865	4798961	4797309	4797584	75.37	75.12	75.22
0.3	3269954	3267934	3267525	4988874	4986909	4986563	78.49	78.13	78.23
0.4	3841251	3838243	3839581	5236934	5234466	5236656	80.07	79.29	79.56

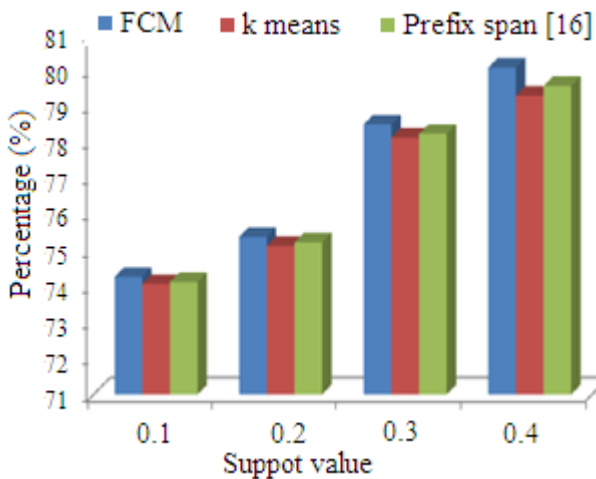


Figure.7 Database difference ratio for proposed vs. existing method

When analyzing the table.4, we evaluate the overall execution time, memory value and the database difference ratio for the fuzzy c means and existing k means clustering algorithm and prefix span algorithm [14]. From the result, the proposed technique achieves the execution time for the support value 0.1 is 2345841ms but the exiting k means technique takes 2344641ms and prefix span [14] algorithm takes 2344721ms for grouping or dividing the available pattern. And also the memory value taken for our FCM takes 4569544 bits memory and existing K means and prefix span [14] takes 4568339bits and 4568258 bits respectively. When compared to the FCM algorithm existing K means and prefix span [14] algorithm takes minimum time and minimum memory value. Even though the exiting technique achieves minimum execution time and memory space the proposed technique removes the repeated cyclic pattern in an effective manner when compared to the K means and prefix span [14]. Because of the effective clustering process using FCM, the suggested technique removes the cyclic repeated pattern with maximum percentage.

The graphical representation of the database difference ratio for the proposed fuzzy c means and existing k means clustering algorithm is shown in fig.7.

6. Conclusion

An effective method for cyclic repeated patterns in sequential pattern mining based on the fuzzy c-means clustering and association rule mining technique is proposed in this paper. Initially the mushroom dataset is fed to the fuzzy c means clustering algorithm. In fuzzy c means clustering algorithm the available input dataset is clustered

based on their similar sequential pattern. After the clustering process, the rules for each cluster are generated with the help of association rule mining algorithm. Finally, we fix the threshold value based on that the repetition of the cyclic pattern is removed and stored to the database. The performance of the proposed technique is evaluated by means of execution time, memory and database difference ratio. From the result our proposed technique removes the cyclic repeated pattern with effective manner. In future work, the cyclic repeated pattern could be effectively removed by various clustering and rule mining techniques.

References

- [1] M. Garofalakis, R. Rastogi, "Mining Sequential Patterns With Regular Expression Constraints", *IEEE Transaction on Knowledge and Data Engineering*, Vol. 14, No. 3, pp. 530-552, 2002.
- [2] A. P. Wright, A. T. Wright, "The use of sequential pattern mining to predict next prescribed medications", *Journal of Biomedical Informatics*, Vol. 53, pp. 73-80, 2015.
- [3] J. W. Huang, C. Y. Tseng, "A General Model for Sequential Pattern Mining with a Progressive Database", *IEEE Transaction on Knowledge and Data Engineering*, Vol. 20, No. 9, pp. 1153-1167, 2008.
- [4] J. Pei, J. Han, "Mining Sequential Patterns by Pattern-Growth: The Prefix Span Approach", *IEEE Transaction on Knowledge and Data Engineering*, Vol. 16, No. 11, pp. 1424-1440, 2004.
- [5] J. Zhang a, Y. Wang, "CCSpan: Mining closed contiguous sequential patterns", *Knowledge-Based Systems*, Vol. 89, pp. 1-13, 2015.
- [6] A. Julea, N. Méger, "Unsupervised Spatiotemporal Mining of Satellite Image Time Series Using Grouped Frequent Sequential Patterns", *IEEE Transaction on Geosciences and Remote Sensing*, Vol. 49, No. 4, pp. 1417-1430, 2001.
- [7] Z. Zhao, D. Yan, "Mining Probabilistically Frequent Sequential Patterns in Large Uncertain Databases", *IEEE Transaction on Knowledge and Data Engineering*, Vol. 26, No. 5, pp. 1171-1184, 2014.
- [8] C. Fiot, A. Laurent, "From Crispness to Fuzziness: Three Algorithms for Soft Sequential Pattern Mining", *IEEE Transaction on Fuzzy System*, Vol. 15, No. 6, pp. 1263-1277, 2007.
- [9] A. A. Shaw, N. P. Gopalan, "Finding frequent trajectories by clustering and sequential pattern

- mining", *Journal of Traffic and Transportation Engineering*, Vol. 1, No. 6, pp. 393–403, 2014.
- [10] Y. H. Hu, I. C. Chiang, "Mining Cyclic Patterns with Multiple Minimum Repetition Supports", *In Proceedings of IEEE International Conference on Fuzzy System and Knowledge Discovery (FSKD)*, pp. 1545-1549, 2011.
- [11] X. Yu, H. Yu, "An Asynchronous Periodic Sequential Patterns Mining Algorithm with Multiple Minimum item Supports", *In Proceedings of IEEE International Conference on p2p, parallel, Grid, Cloud and Internet Computing*, pp. 274-281, 2014.
- [12] W. Ouyang, Q. Huang, "Mining Positive and Negative Sequential Patterns with Multiple Minimum Supports in Large Transaction Databases", *IEEE 2010 Second WRI Global Congress on Intelligent Systems*, Vol. 2, pp. 190-193, 2010.
- [13] Y. H. Hu, C. F. Tsai, "A novel approach for mining cyclically repeated patterns with multiple minimum support", *Applied Soft Computing*, Vol. 28, pp. 90-99, 2015.
- [14] L. R. Foulds, J. P. de Morais Neto, "A variant of k-nearest neighbors search with cyclically permuted query points for rotation-invariant image processing", *Discrete Applied Mathematics*, Vol. 197, pp. 123–144, 2015.
- [15] Z. Zhao, D. Yan, "Mining Probabilistically Frequent Sequential Patterns in Large Uncertain Databases", *IEEE Transaction on Knowledge and data Engineering*, Vol. 26, No. 5, pp. 1171-1184, 2013.
- [16] O. K. Alkan, P. Karagoz, "CRoM and HuspExt: Improving Efficiency of High Utility Sequential Pattern Extraction", *IEEE Transaction on Knowledge and Data Engineering*, Vol. 27, No. 10, pp. 2645-2657, 2015.
- [17] J. C. W. Lin¹, "Efficiently Maintaining the Fast Updated Sequential Pattern Trees With Sequence Deletion", *IEEE Access*, Vol. 2, pp. 1374-1383, 2014.
- [18] Lu, E. H. Chan, V. S. Tseng, and S. Y. Philip, "Mining cluster-based temporal mobile sequential patterns in location-based service environments", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 23, No. 6, pp. 914-927, 2011.
- [19] Zhu, Jiaqi, K. Wang, Y. Wu, Z. Hu, and H. Wang, "Mining User-Aware Rare Sequential Topic Patterns in Document Streams", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 28, No. 7, pp.1790-1804, 2016.
- [20] F. Viger, Philippe, C. W. Wu, V. S. Tseng, L. Cao, and R. Nkambou, "Mining Partially-Ordered Sequential Rules Common to Multiple Sequences", *IEEE Transactions on Knowledge and Data Engineering*, Vol.27, No.8, pp. 2203-2216, 2015.