Webology (ISSN: 1735-188X)
Volume 19, Number 2, 2022

A Novel Matx-V Algorithm For College Course Demand And Admission Rate Prediction Using Sequential Pattern Mining

Mrs. VANI .N[1] , Dr. VEERAGANGADHAR SWAMY .T.M[2]

Ph.D Research Scholar, Dept of CSE, RYM Engineering College, Visvesvaraya Technological University(VTU), Karnataka.

[2]Professor & Head , Dept of CSE (AI &ML) , RYM Engineering College, Visvesvaraya Technological University(VTU) , Karnataka.

Abstract: The Sequential pattern mining does quintessence of patterns from quest of information dumped from large number of social media companies such as Google, Yahoo, Amazon, Flipcart and soon. Sequential pattern algorithms are very amelioration in extrication of knowledgeable patterns in various applications such as DNA analysis, Stock market, Intrusion detection etc. the extant algorithms are GSP, Prefix Span, SPADE, FAST and Lapin having numerous of pitfalls such as enormous of Time depletion, memory exhaustion and Complexity in investigating large amount of candidate sequence for macro databases. In this paper, a novel SPM technique for college course demand prediction and admission rate prediction is proposed using the MATX-V model. In the proposed work, the college log dataset is taken as input to predict the course demand and admission rates in college. Initially, preprocessing of the dataset is carried out. After that, user identification is performed to predict user behavior for future requests and then session identification takes place. Next, the proposed Mathematical model for Matrix Manipulation (MATX-V) is used to identify the frequent patterns in the dataset. Here, four different types of sequential database scanning are done by computing the centroid points and parsing them on the dataset. The proposed MATX-V algorithm uses Rabin-Karp Algorithm (RKA) for frequent pattern searching. At last, the frequent patterns obtained during scanning are stored by means of queuing technique. The outcomes of the MATX_V model demonstrate better results in terms of time and space efficiency and also the course demand and admission rates of college are predicted efficiently compared to the state-of-art methods.

Keywords: Sequential pattern mining (SPM), Rabin-Karp Algorithm (RKA), Demand prediction, Sequential analysis, Admission rate prediction, Preprocessing, Mathematical model for Matrix Manipulation (MATX-V).

1. INTRODUCTION

A sequential pattern mining having list of item-sets represents in terms of sequence in sequence-order. The application of this mining is customer-shopping, Amazon, flipcharts and online business transaction identifies most frequent sequences [7]. Sequential pattern mining is very complex in identifying DNA Analysis and stock market exchange for future predictions. The utilization of algorithms in various practices such as Decision support in online business transaction, disease prediction suggestion to doctor, Fraud detection in online bank transactions, Learning Status Analysis in online education exams, Intrusion detection, Stock Market Analysis for best shares for future prediction and Customer shopping Analysis in Amazon online purchases.

The sequential pattern mining has to follow certain rules:

i. Estimating or mines full set of patterns from sequential pattern mining.
ii. Minimum support threshold should be satisfied.
iii. Choosing high efficient algorithm for sequential mining.
iv. The selected algorithm should specify suitable number of scan based on user specified requirements.
v. Selection of algorithm which is scalable that suits for all types of requirement and demands.

Most of the algorithm may differ in various categories in sequential pattern mining:

- Every algorithm generates candidate sequences in different ways and different techniques.
- The goal of the algorithm need to reduce candidate sequences generation which reduces input and output cost.
- Estimation of minimum support value is required. The candidate sequence value which is less than minimum support threshold is eliminated so that memory consumption is minimized. The proposed algorithm has new ideas in sequential pattern mining.

The recommendation system accepts input from log file. There is need for course recommendation for students instead of seeking advice from academic expertise [15]. An identification of demanded courses from college website based on student logging-in into website is essential for students to select college. The proposed model develops recommendation system [15] for automated software for better prediction.

Moreover, the existing algorithms are time-consuming and require more memory for processing. Also, they have the complexity in investigating a large number of candidate sequences for macro databases. This paper proposes a novel SPM technique [7] for college course demand prediction and admission rate prediction using the MATX-V model. The proposed model fulfils the above demands on extricating patterns from a large database without the need for candidate generation.

2. LITERATURE SURVEY

Sequential pattern algorithm is accessible in DNA analysis and Medical prescription for patents and predicting behaviours in advance is helpful and also implemented in stock market for predicting shares of companies either profit or loss.

The linear technique of GSP [6] is first algorithm in sequential mining pattern. The complexity of the algorithm increases by increase in size of the database. The join and prune operation needs lot of time consumption process. The candidate sequence engendering of GSP is huge sizeable and generates lot of small patterns for all iteration. . The existing algorithms are SPADE (An efficient Algorithm for mining Frequent Sequences) and Prefix Span (Prefix-projected Sequential Pattern Mining). GSP uses Horizontal method, SPADE is vertical method and Prefix-SPAN is Projection-based pattern growth method. GSP was first algorithm in sequential pattern mining and uses Apriori approach and take huge number of database scans and lots of short patterns are generated in GSP algorithm so it is considered as less efficient when compared with other algorithms. SPADE [11] algorithm was invented by ZAKI. It uses vertical–id list and temporal joins of candidate sequences. SPADE outperforms better way in depth first search and breadth first search. SPADE has expensive I/O minimization problem. Prefix Span does better performance on candidate sequence generation. Bi-level projection is performed in Prefix Span. It takes lots memory space during processing of candidate generation.

The projection database of Prefix Span algorithm [4] produces projected patterns. Size of the database drastically reduces when correlated with GSP. The prefix span increases complexity when it uses large databases.

Title	Author(s)	Methodology	Review
PREFIXSPAN: Mining Sequential Patterns Efficiently By Prefix-projected Pattern Growth	Jian Pei, Jiawei Han	Prefix Span Algorithm Pseudo-projection	• Improves mining efficiency. • Efficient Mining technique for large database • Best suitable for DNA Database
SPADE: An Efficient Algorithm For Mining Frequent Sequences	Mohamed J.Zaki	SPADE Algorithm Vertical database format	• Fast Mining Algorithm • Multiple database scans • Uses simple temporal join operation
GSP:GENERALISED SEQUENTIAL PATTERN	Srikant and Agarwal	GSP Algorthm Horizontal database	• Iterative approach • Multiple database scans • Not compatible for

		format	large database scans
SEQUENTIAL PATTERN MINING: A Comparision Between GSP, spade And Prefixspan	Manika verma, Dr. Devarshi Mehta	Describes about GSP, SPADE, Prefix Span Algorithm	Comparative Analysis on GSP, Prefix Span, SPADE algorithm
SEQUENTIAL MINING: Patterns And Algorithms Analysis	Thabet slimini and Amor lazzer	Sequential pattern algorithms	• DFS based Algorithms • BFS based Algorithms • Apriori-Like algorithm
A TAXONOMY OF SEQUENTIAL PATTERN MINING	Nizar R.Mabroukh	Pattern Growth Algorithm	• Tree Projection, • Candidate sequence pruning • Search space efficiency

Table: 1 Literature survey of GSP, Prefix span and SPADE algorithm

The Spade algorithm is faster when correlated to GSP and Prefix Span algorithm [4]. By adopting equivalence classes such as dividing original task into smaller sub task. Spade uses Hash tree structure with objective of reducing I/O problem and limits number of scans when compared with GSP and Prefix Span. The Time and space complexity is slightly improved in Spade algorithm [11]. The Fast algorithm is faster in execution of algorithm with parameters of input is minimum threshold and maximum threshold of constraint in range of pattern searching algorithm. The Fast algorithm [11] is virtuous at word detection and text detection. It greatly avoids multiple database scans and improved performance is observed in sparse dataset other than dense dataset.

The LAPIN algorithm [10] is analogous to SPAM algorithm in terms of replacing Join operation of SPADE algorithm. The investigation of last position is performed. LAPIN algorithm executes backward tracking and forward tracking and calculates frequent itemset. Depth First Search strategy is implemented using Lexicographic Trees. LAPIN algorithm is acceptable to dense dataset than sparse dataset.

3. PROPOSED METHODOLOGY

Sequential pattern mining (SPM) is a technique that finds out frequent patterns from the data set with a time attribute. It is often applied to discover patterns in sequence databases which are widely used to model shopping sequences, medical syndromes and treatments, natural disasters, stock markets, and so on. A sequential pattern refers to a hierarchical pattern consisting of a sequence of frequent transactions with a particular ordering among these item sets. Many SPM methods have been proposed to meet a range of requirements, such as contrast SPM, tri-partition pattern mining, high utility pattern mining, and gap constraint SPM. The disadvantages of traditional SPM are enormous Time depletion, memory exhaustion, and Complexity in investigating a large amount of candidate sequence. Moreover, the traditional methods neglect the repetition in the sequence. Therefore, in order to overcome the limitations of traditional methods, this paper proposes a novel SPM technique for college course demand prediction and admission rate prediction using the MATX-V model. The goal of the proposed algorithm is to reduce candidate sequences generation and reduces input and output cost. In addition, the candidate sequence value which is less than the minimum support threshold is eliminated to minimize energy consumption. The proposed work consists of the following phases: Dataset collection, Preprocessing, User identification, Session identification, Four different types of database scan, Centroid position calculation, Database parsing (Top-Down Parsing, Bottom-Up Parsing, Left - Right Parsing, Right - Left Parsing, and Diagonal Parsing), Frequent pattern searching using RKA and Queuing. The working procedure of the proposed college course demand prediction and admission rate prediction system using SPM is displayed in figure 1.

3.1 Dataset Collection

At first, the college log dataset used for the proposed work is collected from publically available sources. This dataset is used for predicting user behaviour on the college admission database. It is also used to predict the demand of course and admission rates in college by analyzing the total number of visitors viewing the college website. After collecting the dataset, it is preprocessed for making it more suitable for analysis.

3.2 Preprocessing

Preprocessing is defined as the process of converting the raw data in the dataset into the structured format by using several operations on the dataset. This conversion will be more efficient for further processes, which reduces the time and memory required to process and store the data. The proposed work uses preprocessing steps like noisy data removal, superfluous information removal, and log file cleaning. These steps are described as follows,

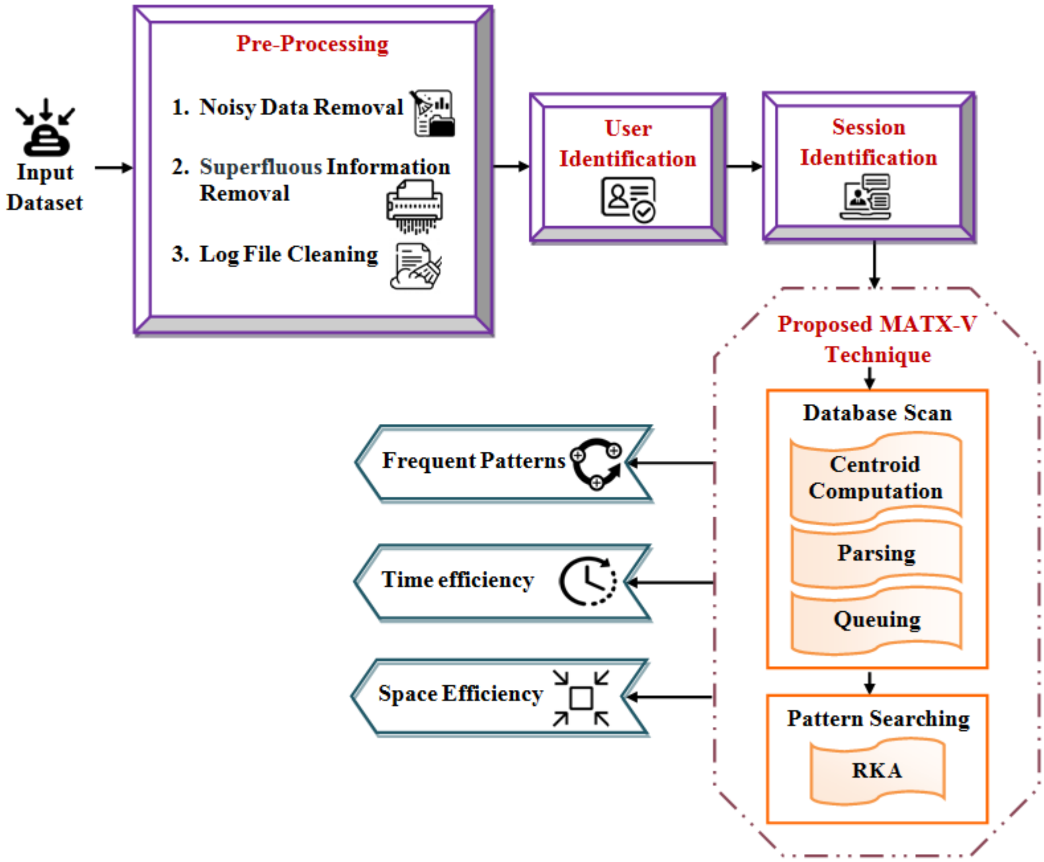

Figure 1: General structure of the proposed framework

- **Noisy data removal** – Noisy data determines the meaningless information present in the dataset, which is complex to understand and process. Such information can be generated due to faulty data collection, data entry errors, etc. Removing these types of noisy data works by dividing the entire log data into several stripes of equal size and loading it into a new file for processing.

- **Superfluous information removal** – It is the process of removing duplicate or irrelevant records from the dataset. Information replica occurs during data collection from multiple sites. So, deletion of such information may improve the efficiency of the overall process.

- **Log file cleaning** – This can be done through a data cleaning process. Data cleaning is defined as the process of identifying and eliminating corrupted or inaccurate records from the database. Then, the valid Uniform Resource Locator (URL), valid GET method, and valid STATUS are identified for pattern analysis.

After performing these steps, the preprocessed dataset is obtained which saves server log memory and improves the performance of the proposed pattern mining, college course demand, and admission rate prediction system. The step next to preprocessing is user identification.

3.3 User Identification

Here, the users are identified to predict their interest in a particular college and course for the future. This helps in identifying the demand for the course and admission rates of a college. Moreover, it will be valuable for search engines to load future requests in a system cache and improve system performance. User identification is a very difficult task because a single user may use different types of browsers on a single system to view only one website and also they use multiple browsers in different systems. Therefore, the identification of users is an important task in the prediction model. The proposed user identification process is given below.

- Initially, the log data is loaded and the users' Internet Protocol address (IP address), User ID, date/time, browser in which they are viewing the college website are examined.
- After that, the user ID and IP address values are sorted for further scrutiny.
- Then, the common IP addresses are identified based on the unique IDs' to compute the page access frequencies.
- Finally, IP address, browser, Date/Time, User ID are mapped to compute the frequency of searching patterns.

3.4 Session Identification

After users are recognized, session identification takes place. Session identification mentions the time duration taken by each user to send a request and response, to complete a task on a server. The time duration of each user is initially divided into multiple sessions and if the time duration of a user exceeds above 30 minutes, then it is considered as a second session for that user. Based on this condition, the number of sessions for each user is computed, and then the mapping is performed based on IP address, User ID, date/time (year, month, date, hour, minute, second), and a number of sessions. Then, the user identified and session identified data are used for predicting course demand and admission rates in college.

3.5 Proposed MATX-V Model for Course Demand and Admission Rate Prediction

In this phase, the proposed Mathematical Model using Matrix Manipulation (MATX-V) is used to envisage the college course demand and admission rates. This model uses Rabin-Karp Algorithm for searching frequent patterns in the dataset. In the proposed MATX-V technique, firstly, centroid position is identified and then four types of database scan such as Top-Down Parsing, Bottom-Up Parsing, Left - Right Parsing, Right - Left Parsing and Diagonal Parsing takes place for identifying frequent patterns via the Rabin Karp Algorithm (RKA) and then obtained sequential patterns are stored using the queuing technique. Let, $S^{(i)}$ be the sequential dataset used to identify the patterns, the $n-$number of elements present in the dataset is represented as,

$$S^{(i)} = S^{(1)}, S^{(2)}, \ldots\ldots, S^{(n)} \tag{1}$$

From the set of elements in the dataset, initially, database scanning is performed.

3.5.1 First database scan

At first, the sequence database is scanned and the centroid position is calculated. After centroid calculation, each element in the database is compared with the centroid value, if a frequent data element is found, then it is stored in the queue. If the first element is frequent, then it is stored in the first queue, if both the first and second elements are frequent, then it is stored in the second queue. If the first three elements are frequent, then it is stored in the third queue and the process continues by comparing each element in the database and storing it in the queue.

3.5.2 Centroid evaluation

The centroid position is used to estimate the frequent patterns from the dataset by comparing each element in the dataset with its corresponding centroid points. And the frequent patterns obtained are then stored in the corresponding queue. The centroid calculation for rows and columns in the dataset is mathematically notated below,

$$C_p^{row} = round\left({T_{row}}/{2} \right) \tag{2}$$

$$C_p^{col} = round\left({T_{col}}/{2} \right) \tag{3}$$

Where, C_p^{row}, C_p^{col} specifies the centroid point of rows and columns respectively, T_{row} is the total number of rows in the dataset, and T_{col} defines the total number of columns in the dataset. The Centroid computation of the proposed MATX-V technique is shown below,

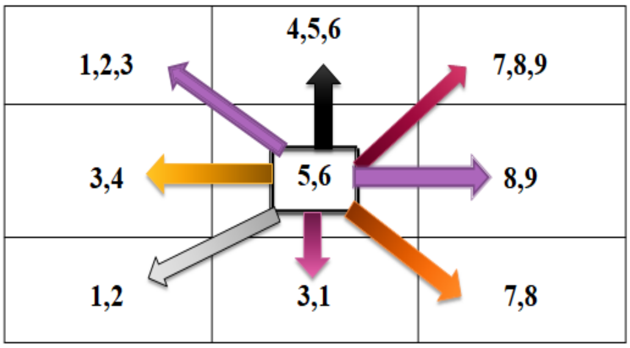

Figure 2: Centroid computation of proposed MATX-V technique

The mathematical model of the first database scan is expressed in equation (4)

$$\left(S^{(i)}\right)_{j+1} = \begin{cases} 1 & C_p \subseteq S^{(i)} \\ 0 & else \end{cases} \tag{4}$$

Here, $\left(S^{(i)}\right)_{j+1}$ mentions the first database scan. Then, a second database scan is carried out via top-down parsing.

3.5.3 Second database scan

Here, the sequential dataset is scanned to identify the frequent elements in the dataset. Parsing defines the process of analyzing a string of symbols to identify the frequent patterns between them. First of all, Top-Down parsing is performed to scan the database for frequent patterns. In this approach, the top element is compared with the bottom element and if a frequent pattern occurs then it is recorded in the queue and the process continues for the entire database. The second database $\left(\left(S^{(i)}\right)_{j+2}\right)$ scan is expressed as,

$$\left(S^{(i)}\right)_{j+2} = \begin{cases} Q^{(k)} & if\left(S^{(top)} \equiv S^{n-1}\right) \\ S^{(i)} & else \end{cases} \tag{5}$$

In equation (5), $S^{(top)}$ defines the top element in the database, $S^{(n-1)}$ signifies the bottom elements in the database and $Q^{(k)}$ determines the queue where, $k = 1,2,\ldots\ldots K$ indicates the number of queues to store the patterns. Then, Bottom-Up parsing is done in a similar way as that of Top-Down parsing. In Bottom-Up parsing, the bottom element is compared with the top element, and the frequent elements are stored in the queue.

$$\left(S^{(i)}\right)_{j+2} = \begin{cases} Q^{(k)} & if\left(S^{(bott)} \equiv S^{i+1}\right) \\ S^{(i)} & else \end{cases} \tag{6}$$

In the above equation, $S^{(bot)}$ illustrates the bottom element in the database and $S^{(i+1)}$ distinguishes the top element in the database.

```
Pseudocode of proposed MATX-V algorithm
Input: Sequence dataset S^(i)
Output: f_p , t_eff , ψ_eff

Begin
Generate sequential dataset S^(i)
While (i ≤ n) do
        Compute first database scan
        For all T_row & T_col
                Calculate C_p^row = round(T_row / 2)
                Evaluate column centroid point C_p^col
                If C_p is frequent with S^(i)
                        A. Store the frequent elements in queue
                        B. First frequent element is stored in first queue
                        C. First two frequent elements are stored in second queue
                        D. First three frequent elements are stored in third queue
                        E. First four frequent elements are stored in fourth queue
                        F. First five frequent elements are stored in fifth queue
                End
        End for
        Perform second database scan
        Estimate Top-Down parsing
        Repeat A to F
        Evaluate Bottom-Up parsing
        Repeat A to F
        Perform third database scan
        Carry out Right-Left parsing
        Repeat A to F
        Accomplish Left-Right parsing
        Repeat A to F
        Execute diagonal parsing
        Repeat A to F
End while
Attain frequent patterns (f_p), time efficiency (t_eff) & space efficiency (ψ_eff)
End
```

Figure 3: Pseudocode of proposed MATX-V algorithm

3.5.4 Third database scan

Next, the third database scan is performed in which Left-Right parsing and Right-Left parsing is carried out. Similar to the second database scan, the Left-Right parsing scans the database from left to right and looks for similar data. Here, the left element is taken as the comparison element. If frequent data is present in the database then it is stored in the queue. And the process continues for the entire dataset. Alike, Right-Left parsing is also done by taking the Right element as the main comparison element. The Left-Right and Right-Left parsing techniques are formulated below in equations (7) & (8),

$$\left(S^{(i)}\right)_{j+3} = \begin{cases} Q^{(k)} & if\left(S^{(left)} \equiv S^{right(i)}\right) \\ S^{(i)} & else \end{cases} \qquad (7)$$

$$\left(S^{(i)}\right)_{j+3} = \begin{cases} Q^{(k)} & if\left(S^{(right)} \equiv S^{left(i)}\right) \\ S^{(i)} & else \end{cases} \quad (8)$$

In the aforementioned equations, $\left(\left(S^{(i)}\right)_{j+3}\right)$ represents the third database scan, $S^{(left)}$ shows the left element in the database, and $S^{(right)}$ determines the right element in the dataset.

3.5.5 Fourth database scan

In this scanning, diagonal parsing is done in which the diagonal elements are compared in the database. If similar sequences are found then they are stored in the queue and the process continues. The diagrammatic representation of diagonal parsing technique is shown below,

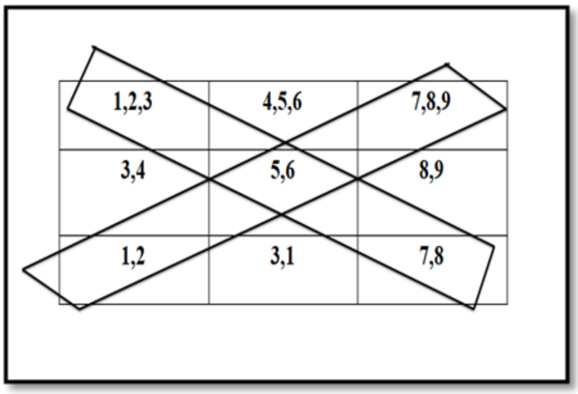

Figure 4: Diagonal parsing in proposed MATX-V algorithm

After performing four types of database scans, pattern searching using RKA is done. This algorithm finds frequent sequences in the database.

3.5.6 Pattern Searching with RKA

Rabin Karp Algorithm is a pattern recognition model, which uses a hash function to find patterns in the database. This technique transforms the normal string value in the dataset into a hash function and then compares the hash values within the dataset. Once strings matches, then the entire pattern is verified based on this approach. If both don't match, then the matching takes place to the next string in the database. Once, the hash value and patterns are exactly matched, then the patterns are identified and stored in the queue. This hashing procedure increases the searching speed of the proposed MATX-V algorithm and reduces the processing time complexity. The RKA hashing technique for pattern searching is detailed further.

Step 1: In the beginning, the elements in the dataset $\left(S^{(i)} = S^{(1)}, S^{(2)}, \ldots\ldots, S^{(n)}\right)$ are converted into a hash function such that,

$$h\left(S^{(i)} = S^{(1)}, S^{(2)}, \ldots\ldots, S^{(n)}\right) = \left(a^{i-1}S^{(1)} + a^{i-1}S^{(2)} + \ldots\ldots + a^{0}S^{(i-1)}\right) \cdot \bmod b \quad (9)$$

Where, $h(\bullet)$ shows the hashing function, a,b is the appropriately selected prime numbers.

Step 2: After the hash conversion of input data, string matching takes place to compare the hash value of the input string to the hash value of substrings in the database.

Step 3: Then, a similarity score is computed to identify the patterns in the document. The similarity score is calculated using the following equation,

$$\delta = \frac{2 * \beta}{x + y} \quad (10)$$

In the aforementioned equation, δ represents the database similarity score, β mentions the same hash value of strings and x, y indicates the number of hash schemes in the database. The output of the proposed MATX-V algorithm provides the frequent patterns (f_p) that occur in the sequential college log database and also it shows the time efficiency (t_{eff}) and space efficiency (ψ_{eff}) of the proposed model. Space efficiency defines the amount of memory consumed by the proposed technique to execute the pattern searching technique. Time efficiency mentions the time taken by the proposed algorithm to predict the course demands and admission rates in the college. The pseudocode of the proposed MATX-V algorithm is displayed in figure 4.

4. RESULTS AND DISCUSSION

The effectiveness of the proposed MATX-V technique for college course demand prediction and admission rate prediction is evaluated by comparing the outcomes of the proposed model with the existing Generalized Sequential Pattern (GSP) algorithm and Prefix-projected Sequential Pattern Mining (Prefixspan) algorithm. The two sets of college log datasets are used in the proposed work and are taken from publically available sources. The proposed course demand and admission rate prediction model is implemented in the working platform of PYTHON SPYIDER 3.7. The performance evaluation is described below,

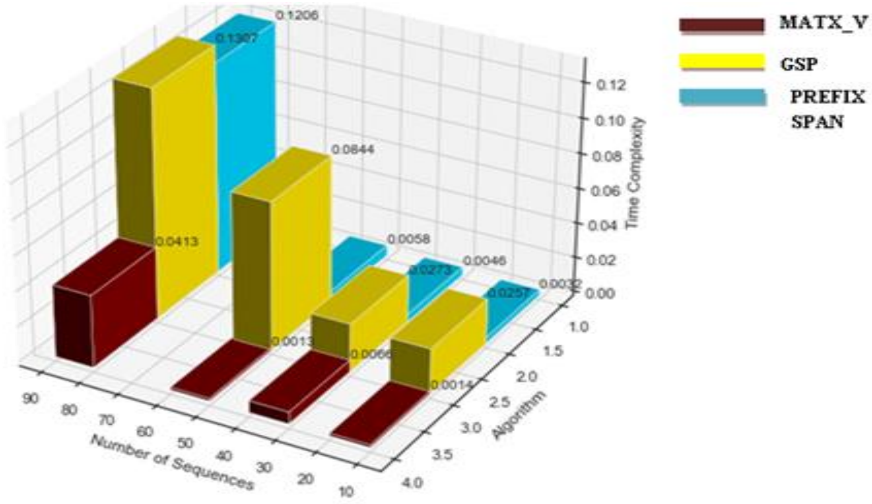

Figure 5: Time complexity assessment of dataset 1

Figure 5 compares the time efficiency of the proposed and existing techniques. From the graphical assessment, for the number of sequences ranging from 80-90, the time taken by the proposed MATX-V model to predict the course demand and admission rates is 0.0413sec. On the other hand, the existing GSP technique requires 0.1307sec for pattern identification. This is 0.0849sec greater than the proposed methodology. Likewise, the existing Prefixspan takes 0.1206sec to identify the patterns in the dataset. By comparing it with the proposed methodology, the proposed model predicts the patterns in a lower time than the other two existing techniques. Similarly, for the number of sequences between 50-60, the time required by the proposed technique to predict the pattern is 0.0013sec, whereas the other two existing GSP and Prefixspan models predict the patterns in 0.0058sec and 0.0844sec respectively. In this, the time complexity of the proposed system is low. In a similar manner, for the number of sequences below 50, the time complexity of the proposed method is lower than the existing ones. Therefore, it is clear that the proposed model predicts the pattern with less time compared to that of the existing techniques. The time complexity for dataset 2 is given by,

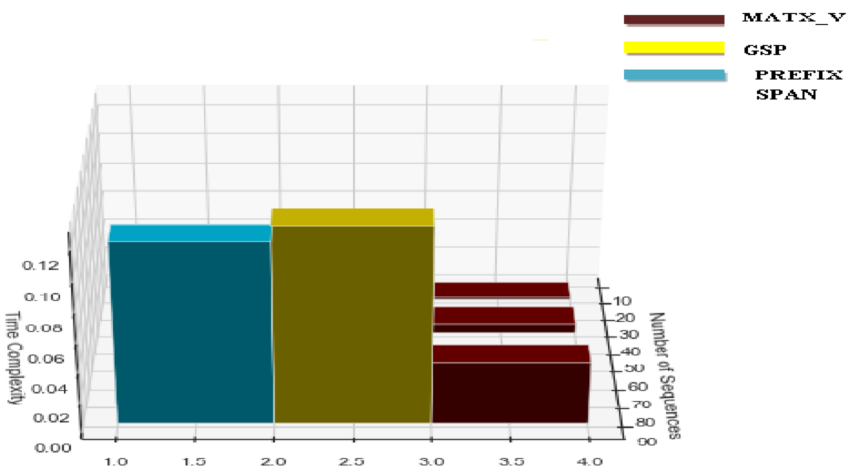

Figure 6: Time complexity of dataset 2

The above figure 6 illustrates the time complexities of the proposed and existing sequential pattern mining techniques. In the graphical assessment, the blue line represents the time complexity of the existing Prefixspan technique, the yellow part mentions the time complexity of the existing GSP model and the black part indicates the time taken by the proposed model to identify the patterns. Hence, from the graph, the time complexities of the proposed MATX-V model are low when compared to the existing methods. For the increasing number of sequences, the time complexities are increasing. For the increasing number of sequences also the proposed work requires less time to identify the patterns and to predict the demands. Therefore, for different datasets, the proposed work achieves best results than the other state-of-art models. The space complexity analysis of dataset 1 is shown below.

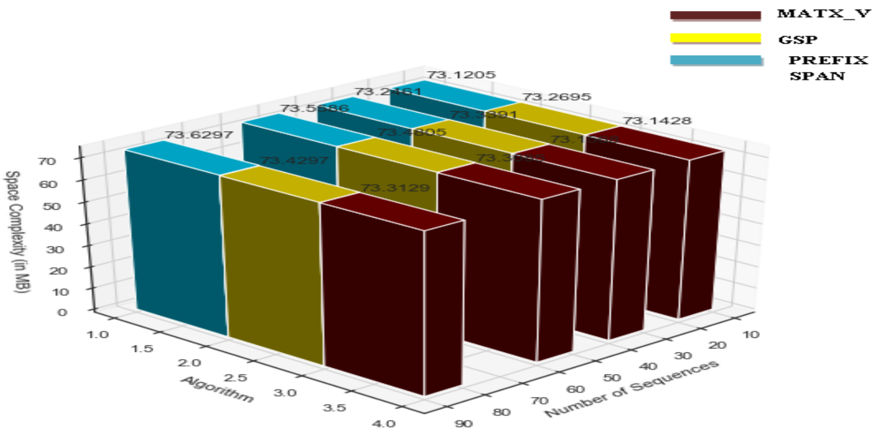

Figure 7: Space complexity analysis of dataset 1 using the proposed model

Space efficiency defines the amount of memory consumed by the proposed algorithm to execute the pattern searching technique. Lower memory consumption indicates the best performance. Figure 7 depicts the memory consumption of the proposed and existing sequential pattern mining algorithms. From the graphical representation, the space occupied by the proposed MATX-V algorithm to identify the frequent patterns is low when compared to the existing GSP and Prefixspan models. When total number of sequences is rangng from 10-90, the memory consumed by the proposed system increases gradually in terms of space complexity. But, the space complexity values are low compared to other models. For the lower number of sequence (10), the space complexity of the proposed technique is 73.1428 Mb. This is very low compared to other methods. Hence, it is sure that the proposed model is better than the existing one. The space complexity analysis of dataset 2 is graphically publicized in figure 8.

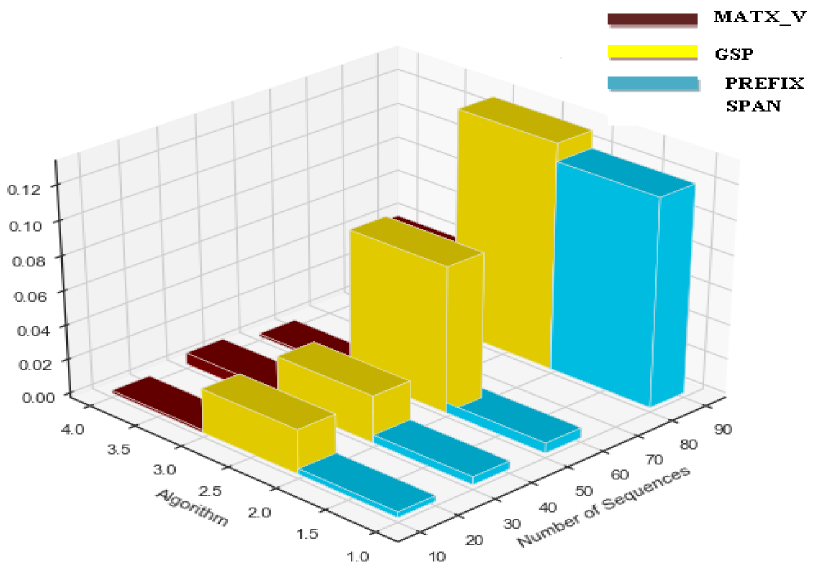

Figure 8: graphical representation of space complexity of dataset 2.

Figure 8 represents the space complexity of different sequential pattern mining algorithms with respect to the number of sequences. The graph shows that the space complexity of the proposed MATX-V model for the maximum number of sequences (90) is 58.9804Mb. But, for the existing Prefixspan model, the space complexity is increased to 124.72265Mb. Meanwhile, for the existing GSP, the space complexity is 94.5742Mb. From this discussion, the proposed one has a lower space complexity than that of the other two models. Therefore, the proposed system is better than other baseline techniques. The overall examination of the time complexity and space complexity of the proposed MATX-V model is low for the two different datasets while comparing it with the existing GSP and Prefixspan techniques. Also, the proposed system predicts the course demand and admission rates of college more effectively with less time and low memory consumption. As a result, it is evident that the proposed one is better than all other existing SPM systems.

5. CONCLUSION

In this paper, a novel SPM technique for college course demand prediction and admission rate prediction is proposed using the MATX-V model. The proposed work uses a college log dataset as input for predicting the course demand and admission rates in college. The proposed work goes through the following phases: Dataset collection, Preprocessing, User identification, Session identification, Four different types of database scan, Centroid position calculation, Database parsing (Top-Down Parsing, Bottom-Up Parsing, Left - Right Parsing, Right - Left Parsing, and Diagonal Parsing), Frequent pattern searching using RKA and Queuing. Finally, experimental analysis is carried out by comparing the outcomes obtained by using the proposed MATX-V model and the existing Prefixspan and GSP algorithms. In the proposed algorithm, no candidate sequences are used and No joining and pruning concept in the proposed algorithm. There are four database scans, which improve memory efficiency and time efficiency, and performance when compared with GSP, PrefixSpan, and Spade algorithm. The lower time efficiency achieved by the proposed model is 0.0413sec for database 1, similarly, the memory efficiency of the proposed model is 73.1428 Mb, which is very low. overall, the proposed approach outperforms the existing state-of-art methods and remains to be more reliable and robust. In the future, this model can be further extended by using advanced techniques for SPM.

REFERENCES

1. Maral kolahkaj, Ali Harounabadi, Alireza Nikravanshalmani and Rahim Chinipardaz, “A hybrid context aware approach for e-tourism package recommendation based on asymmetric similarity measurement and sequential pattern mining”, Electronic Commerce Research and Applications, vol. 42, pp. 1-50, 2020.
2. Taushif Anwar and Umaa V, “CD-SPM cross-domain book recommendation using sequential pattern mining and rule mining”, Journal of King Saud University - Computer and Information Sciences, vol. 32, no. 10, pp. 1-13, 2020.
3. Shah Mohammed Nuruddin, Didarul Islam, Shafiqul Alam, Jesan Ahammed Ovi and Ashraful Islam, “An efficient approach for sequential pattern mining on GPU using CUDA platform”, 4th International Symposium on Multidisciplinary Studies and Innovative Technologies, 22-24 October, Istanbul, Turkey, 2020.
4. Ji-Soo Kang, Ji-Won Baek and Kyungyong Chung, “Prefix span based pattern mining using time sliding weight from streaming data”, IEEE Access , vol. 8, pp. 124833 - 124844, 2020.
5. Hai Duong, TinTruong, Anh Tran, “ Fast Generation of Sequential Pattern with Item Constraints from concise Representation”, Springer Open, Journal of Big data,Springer, DOI:10.1186/s40537-019-0200-9,2019.
6. Jaewoong Choi, Byeongki Jeong and Janghyeok Yoon, “Technology opportunity discovery under the dynamic change of focus technology fields application of sequential pattern mining to patent classifications”, Technological Forecasting & Social Change, vol. 148, pp. 1-12, 2019.
7. Lamine Diop, Cheikh Talibouya Diop, Arnaud Giacometti, Dominique Li and Arnaud Soulet, “Sequential pattern sampling with norm-based utility”, Knowledge and Information Systems, vol. 62, pp. 2029-2065, 2019.
8. Maylawati D. S, Aulawi H and Ramdhani M. A, “The concept of sequential pattern mining for text”, IOP Conference Series Materials Science and Engineering, vol. 434, no. 1, pp. 1-8, 2018.
9. Chris Wong, “Sequence based course recommender for personalized curriculum planning”, Artificial Intelligence in Education, Springer, Cham, 1st Edition, ISBN no: 978-3-319-93845-5, 2018.
10. SPMF: A Sequential Pattern Mining FrameWork http://www.philippe-fournier-viger.com/spmf
11. Mohammed J Zaki, “SPADE: An efficient algorithm for Mining Frequent Sequences In Journal Machine Learning”, Volume 42 Issue 1-2, Pages 31-60.
12. John K Tarus, Zhendong Niu and Abdallah Yousif, “A hybrid knowledge-based recommender system for e-learning based on ontology and sequential pattern mining”, Future Generation Computer Systems, vol. 72, pp. 37-48, 2017.

13. Pilsun Choi and Buhyun Hwang, “Dynamic weighted sequential pattern mining for USN system”, 11th International Conference on Ubiquitous Information Management and Communication, 5-7 January, New York NY United States, 2017.
14. Mohammad Karim Sohrabi and Vahid Ghods, “CUSE: a novel cube-based approach for sequential pattern mining”, 4th International Symposium on Computational and Business Intelligence, 5-7 September, Olten, Switzerland, 2016.
15. Jie Xu, Tianwei xing and mihaela van der schaar, “personalized course sequence recommendations”, IEEE Transactions on Signal Processing, vol. 64, no. 20, pp. 5340-5352, 2016.
16. Ke Wang, Elaheh Sadredini and Kevin Skadron, “Sequential pattern mining with the micron automata processor”, ACM International Conference on Computing Frontiers, May 16-19, New York, United States, 2016.
17. Amina Kemmar, Yahia Lebbah, Samir Loudni, Patrice Boizumault and Thierry Charnois, “Prefix projection global constraint and top-k approach for sequential pattern mining”, constraints, vol. 22, no. 2, pp. 1-42, 2016.
18. Doddegowda B J, G T Raju, Sunil Kumar S Manvi,”Extraction of Behavioral Patterns from Preprocessed Web Usage Data for Web Personalization”, IEEE International Conference On Recent Trends In Electronics Information Communication Technology, May 20-21, 2016, India
19. Fabio Fumarola, Pasqua Fabiana Lanotte, Michelangelo Ceci and Donato Malerba, “CloFAST closed sequential pattern mining using sparse and vertical id-lists”, Knowledge of Information system, vol. 48, pp. 429-463, 2015.
20. Zailani Abdullah, Omer Adam, Tutut Herawan and Mustafa Mat Deris, “A review on sequential pattern mining algorithms based on apriori and patterns growth”, International Conference on Data Engineering, Springer, Singapore, 1st Edition, ISBN No: 978-981-13-1797-2, 2015.
21. Geethapriya Uvaraja, “Recommendation generation by integrating sequential pattern mining and semantics”, International Journal of Research in Engineering and Technology, vol. 3, no. 1, pp. 201-205, 2014.
22. Alpa Reshamwala and Neha Mishra, “Analysis of sequential pattern mining algorithms”, International Journal of Scientific & Engineering Research, vol. 5, no. 2, pp. 1034-1038, 2014.
23. Veeragangadhara swamy T M, G T Raju (2015) “ A Novel Prefetching Technique through Frequent Sequential Patterns from Web Usage Data”, COMPUSOFT, An international journal of advanced computer technology, 4 (6), Volume-IV, Issue-4,June-2015
24. Shinjee Pyo, Eunhui Kim and Munchurl Kim, “Automatic and personalized recommendation of TV program contents using sequential pattern mining for smart TV user interaction”, Multimedia Systems, vol. 19, pp. 527-542, 2013.
25. Thabet Slimani and Amor Lazzez, “Sequential mining patterns and algorithms analysis”, International Journal of Computer and Electronics Research, vol. 2, no. 5, pp. 639-647, 2013.
26. Ghim-Eng Yap, Xiao-Li Liand and Philip S Yu, “Effective next items recommendation via personalized sequential pattern mining”, 17th International

Conference on Database Systems for Advanced Applications, 15-19 April, Busan, Korea, 2012.

27. Madan Kumar K. M. V, Srinivas P. V. S and Raghavendra Rao C, “Sequential pattern mining with multiple minimum supports by MS-SPADE”, International Journal of Computer Science Issues, vol. 9, no. 5, pp. 285-292, 2012.
28. Chetna Chand, Amit Thakkar, Amit Ganatra. : Sequential Pattern Mining: Survey and Current Research Challenges. International Journal of Soft Computing and Engineering (IJSCE) ISSN: 22312307, Volume-2, Issue-1,2012.