Frequent Pattern Data Mining based clustering and classification using Fuzzy back propagation ResNet Convolutional networks

K.H Niralgikar*[1] , Dr.Mukesh A.Bulsara[2]

Abstract: Numerous scalable approaches are developed for mining various types of patterns, with frequent pattern mining being a key theme in data mining research. The internet is getting considerably more complicated, which makes it both more crucial and more difficult to process different data mining challenges across diverse areas. This research propose novel technique in frequent pattern data mining based on sequence using clustering and DL architecture. Frequent pattern based data is clustered and classified using fuzzy clustering with back propagation ResNet Convolutional networks. The experimental analysis has been carried out in terms of accuracy, precision, recall, F-1 score and RMSE. Proposed technique attained accuracy of 95%, precision of 71%, recall of 65%, F-1 score of 77% and RMSE of 62%.

Keywords: *Frequent pattern mining, data mining, sequence, clustering, deep learning*

1. Introduction

Finding hidden patterns in a data set that are interesting or valuable and using them as explicit knowledge is the goal of data mining. This method is quite computationally intensive. In general, knowledge that can explain a wide range of situations is regarded as being valuable, and frequency and confidence in the patterns that have been found are frequently employed as indicators of the quality of knowledge [1]. Along these lines, the co-occurrence of items in transactions under these two measures has been intensively explored using the Basket Analysis Apriori technique. The Apriori algorithm is effective in extracting co-occurrence of items, however the data structure it can handle is only a collection of items. Recent works [2] have reported on the extensions to address structured data in the form of sequences and/or taxonomies. One of the key data mining techniques is association rule mining. The process of using association rules to find potential connections between data points in large datasets. It makes it possible to discover regular patterns and connections between patterns with little risk to humans, bringing the majority of common knowledge to the surface for usage. Association rule has been determined to be a successful method for extracting important information from large datasets. Many of the various methods that were built have been included in numerous application fields that enable the telecommunications field [3]. The quality of the rules frequently prevents the use of the retrieved rules to dealing with complexity in the real world. When multi-level rules are involved, calculating the quality of association rules can be challenging, and current modes often seem to be at odds with one another. Massive amounts of databases and massive amounts of data in various disciplines have started to be created in order to advance information innovation. The study areas of database and information innovation have provided a mechanism to handle, store, and regulate this important data to support fundamental leadership. Data mining is a method used to extract valuable information for analysis from vast amounts of data. Since this method is so straightforward, patterns have been found. It is the same as what is referred to as information projection processes on the last day [4].

Contribution of this research is as follows:

1. To propose novel method in frequent pattern data mining based on sequence using clustering and deep learning architecture.
2. the frequent pattern based data has been clustered and classified using fuzzy clustering with back propagation ResNet Convolutional networks.

2. Literature Review:

Machine learning and deep learning are both very recent fields in artificial intelligence. To create more congested decision boundaries, it transforms the training inputs nonlinearly in different ways. Model training essentially uses supervised and unsupervised techniques. Multiple strategies have emerged as a result of the sequential pattern mining field's fast growing number of models. A new DL method for extracting pertinent characteristics from audio input was shown in Work [5]. Tzanetakis and Majorminer sets can be found in the benchmarking data. Both supervised and unsupervised learning processes were used to train the Deep Belief Networks. The features that were extracted are the inputs for the autotagging and genre classification processes. The categorization task is carried out using or SVM, classifier based on RBF kernel. In the

PHD,Research Scholar, CSPIT, CHARUSAT, CHANGA, Gujarat.
Mechanical Department, CHARUSAT.CHANGA.
kirtihn@yahoo.co.in
Professor Mechanical Department GCETV.V Nagar,Gujarat
mukeshbulsara@yahoo.com

extraction of pertinent information from audio spectra, this method outperformed Mel-Frequency Cepstral Coefficients (MFCCs), according to the evaluation step. Using Deep Belief Networks, author [6] proposed a new DL method for automatic speech recognition (DBNs). On TIMIT corpus, Associative Memory DBN (AM-DBN) and Back-Propagation DBN (BP-DBN) were applied. Both NN had strong generalisation capabilities in speech recognition, according to experimental investigations. A new DL technique based on BPTT for Sequential Click Prediction was introduced in Work [7]. The dynamic sequential states are memorised using the hidden layer. Ad features, User features (user ID and query), and Sequential features were chosen as the most pertinent features for precise Click-Through Rate (CTR) prediction. Recurrent Neural Networks outperformed Logistic Regression (LR) and NN models, according to an evaluation based on click-through records from a commercial search engine. Using auto-associative memory, author [8] presented a new DL method for mining frequent patterns. An input layer and an output layer are both present in the architecture. The Correlation Matrix Memory (CMM), a square matrix that represents the pattern occurrences, was employed by the neural network to identify the frequent patterns. The suggested model outperforms Apriori, Frequent Pattern Growth (FP-growth), Compressed FP-tree based method (CT-PRO), and Linear Time Closed Itemset Miner (LCM) according to experimental data [9]. The majority of DL methods are susceptible to learning instability, bias-variance problem, and parameter tuning. Additionally, the recognition accuracy frequently fell short of what was required for real-world engineering applications. We applied a new DL method based on ensemble learning as well as model selection to get around these restrictions [10].

3. Proposed model:

This section discuss novel technique in frequent pattern data mining based on sequence using clustering and DL architecture. Frequent pattern based data is clustered and classified utilizing fuzzy clustering with back propagation ResNet Convolutional networks. The proposed frequent pattern data mining based on sequence is shown in figure-1.

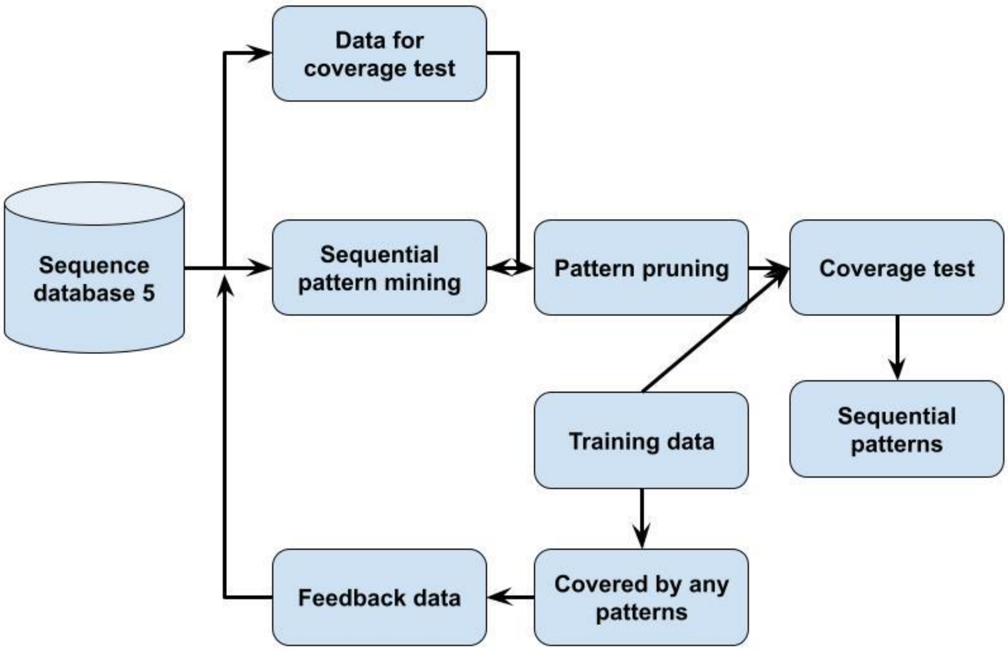

Figure-1 proposed frequent pattern data mining based on sequence

The classification model first has statistical significance and generalises effectively to the test data because it leverages frequent characteristics for induction. The model cannot generalise successfully to test data if an uncommon feature is employed since it was created from statistically insignificant observations. Overfitting is the term for this situation. Second, a pattern's support has a direct impact on how discriminatory it is. Using information gain as an illustration The information gain for a pattern represented by a random variable X is given by eq (1)

$$IG(C|X) = H(C) - H(C|X) \quad (1)$$

$$IG_{ub}(C|X) = H(C) - H_{lb}(C|X) \quad (2)$$

$$H(C|X) = -\sum_{x\in\{0,1\}} P(x) \sum_{c\in\{0,1\}} P(C|x) log P(c|x) \quad (3)$$

A function of p, q, and is H(C|X). P has a fixed value when given a dataset. Since H(C|X) is a concave function, the following criteria determine where it hits its lower bound with respect to q for fixed p and. When q = 0 or 1, H(C|X) hits its lower bound if p. When q = p/ or 1 (1 p)/, H(C|X) hits its lower bound if > p. The cases of θ ≤ p and p are identical. The analysis for the other cases is similar, but due to space constraints, we only cover the scenario for θ ≤ p. We only consider the scenario in which q = 1 because q = 0 and q = 1 are symmetric for case p. In that situation, eq (4) provides the lower bound Hlb(C|X)

$$H_{lb}(C|X) = (\theta - 1)(\frac{p-\theta}{1-\theta} log \frac{p-\theta}{1-\theta} + \frac{1-p}{1-\theta} log \frac{1-p}{1-\theta}) \quad (4)$$

fuzzy clustering with back propagation ResNet Convolutional networks:

For any R ϕ 2 (x)dx < ∞ and ϕ(x) 6= 0, so R R K(x, y)ϕ(x)ϕ(y)dxdy > 0. If the optimal classification's inner product takes the place of the dot product, the original eigenspace will have been transformed into a new feature space via eq (5),

$$\max Q(a) = \sum_{i=1}^{n} a_i - 0.5\sum_{i,j=1}^{n} a_i, a_j, y_i, y_j K(x_i, x_j) \quad (5)$$

$$\int (x) = \operatorname{sgn}[\sum_{i=1}^{n} a_i^* y_i K(x_i + x) + b^* \quad (6)$$

Here, an is the best possible solution, and b is the classification threshold, which can be determined by any pair of support vectors in two classes or by a support vector alone. We employ a kernel function R(x,y) satisfying the Mercer condition to prevent high dimensional operation. Simple operations in low-dimensional space can be used to achieve the inner product operation of high-dimensional space. The shape of the nonlinear mapping by eq(7) need not be taken into account.

$$[\ \Re(x, y) = \langle \zeta(x), \zeta(y) >, \quad (7)$$

Gaussian kernel is defined as eq. (8):

$$\Re(x_i, y_i) = \exp\left(-\frac{\|x_i - v_i\|^2}{\sigma^2}\right) \quad (8)$$

We first define object function of FCM by eq. (9).

$$\frac{\partial J}{\partial u_{ij}} = 0 => \left(1 - \Re(x_j, \nabla_i)\right) - \lambda_i = 0 \quad (9)$$

Where $\sum_{i=1}^{c} u_{ij} = 1$

Let partial derivative of uij for J(U, V) be zero by eq. (10)

$$\frac{\partial J}{\partial u_{ij}} = 0 => \left(1 - \Re(x_j, \nabla_i)\right) - \lambda_i = 0 \quad (10)$$

$$\sigma = \sum_{i=1}^{n} (d_i - d)^2/(n-1) \quad (12)$$

The degree of polymerization surrounding the grouping is represented by the distance variance between the sample points, which is also the degree of cluster structure compaction. The distance variance between the sample sites can be roughly estimated. As a result, we divide each feature map independently to perform fuzzy inference using the formula: $\phi_q^p = \prod_{i=1}^{9} \mu_{F_i^q}^p (x_i)$ where eq (13)

$$\mu_{F_i^p}^p(\zeta_{p,i}^5) = \begin{bmatrix} N(\zeta_{p,1}^5) \times \cdots N(\zeta_{p,i}^5) \times \cdots N(\zeta_{p,9}^5) \\ Z(\zeta_{p,1}^5) \times N(\zeta_{p,i}^5) \times \cdots N(\zeta_{p,9}^5) \\ P(\zeta_{p,1}^5) \times \cdots P(\zeta_{p,i}^5) \times \cdots P(\zeta_{p,9}^5) \end{bmatrix} \quad (13)$$

$$\psi = \begin{bmatrix} \phi_1^1 \\ \phi_2^1 \\ + \\ \dot{\phi}_1 \\ \phi_{19683}^{sn} \end{bmatrix}, z^8 = \begin{bmatrix} z_1^8 \\ z_2^8 \\ z_3^8 \end{bmatrix}, w^8 = \begin{bmatrix} w_{1,1}^8 & \cdots & w_{1,19683}^8 \\ w_{2,1}^8 & \cdots & w_{2,19683}^8 \\ w_{3,1}^8 & \cdots & w_{3,19833}^8 \end{bmatrix} \quad (14)$$

To derive $\partial y_i / \partial z_j^8$, we divide condition into two cases as Then, we can obtain by eq. (15)

$$\frac{\partial L}{\partial z_j^8} = -\sum_{i=1, i\neq j}^{3} \frac{d_i}{y_i}(-y_i y_j) - \frac{d_j}{y_j}\left(y_j(1-y_j)\right)$$

$$= \sum_{i=1, i\neq j}^{3} d_i y_j - d_j + d_j y_j = -d_j + y_j \sum_{i=1}^{3} d_i \quad (15)$$

The layers are introduced with the input image. The first layer consists of convolution plus a rectifier-using unit called a RELU (rectified linear unit). Max pooling layer comes after the following layer. The pooling procedure is carried out by choosing the most element possible from the region of filter-covered layer depths. Each layer output in the residual block is passed on to the following layer, and hops also take place across the identity connections. the output of a map that contains and has the most protuberant qualities of the previous map after max-pooling. The chunks that are utilised to indicate the use of earlier layers are shown by the arc blue outlines. Three identical blocks make up the first layer, four identical blocks make up the second layer, and four identical blocks make up the third layer.

4. Experimental analysis:

The computer platform is set up with an Intel Core I7 processor running at 4.0 GHz, 16GB of memory, and an NVIDIA GTX 780 GPU. MATLAB 2017 compiles the algorithm (a). In the experiment, the image size is 512 x 512 pixels. We downsized all of the downloaded photographs to 200x200 because our architecture only accepts input in the form of RGB 200x200 images. Generally, we validated the CNN model using k-fold validation. If there is enough data, prediction accuracy often reaches 99% precision rate. Our experiment had only processed 2100 photos, thus we utilised k-fold* to gauge the effectiveness of our strategy.

Data set description: The Biomedical Informatics and eHealth laboratory provided the dataset that we used in our experiments1. Total number of patterns in each category is represented by the slice in the frequency distribution. With an average category length of 164.70 and a standard deviation of 132.42, there are a significant number of patterns that are distributed unevenly between categories. The training and test sets are randomly selected from the data collection. Remaining 2305 patterns are utilized as training samples, and 989 of 3294 patterns are kept as test data for performance evaluation.

Table-1 Comparative analysis between proposed and existing technique based on clustering and classification for frequent data mining

Parameters	AM_DBM	CMM	FPDM_ResNet
Accuracy	85	88	95
Precision	71	73	77
Recall	61	63	65
F1_Score	72	75	77
RMSE	56	58	62

Table 1 shows Comparative analysis between proposed and existing technique based on clustering and classification for frequent data mining

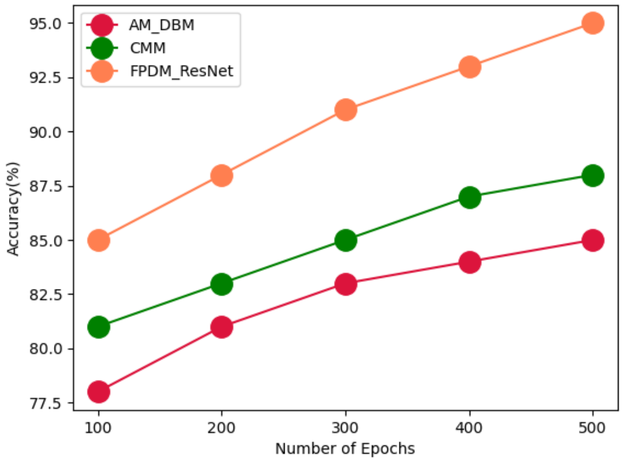

Figure-3 Comparison of accuracy

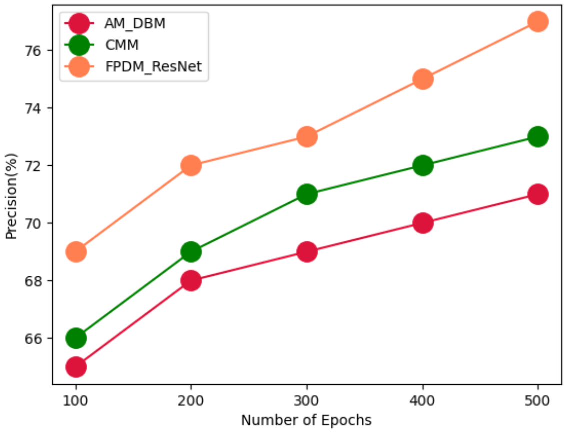

Figure-4 Comparison of precision

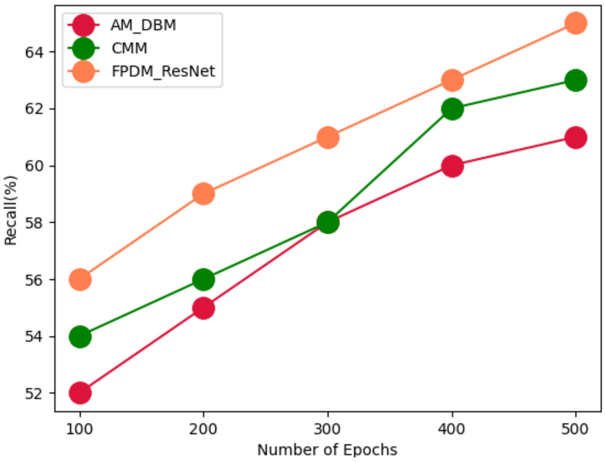

Figure-5 Comparison of Recall

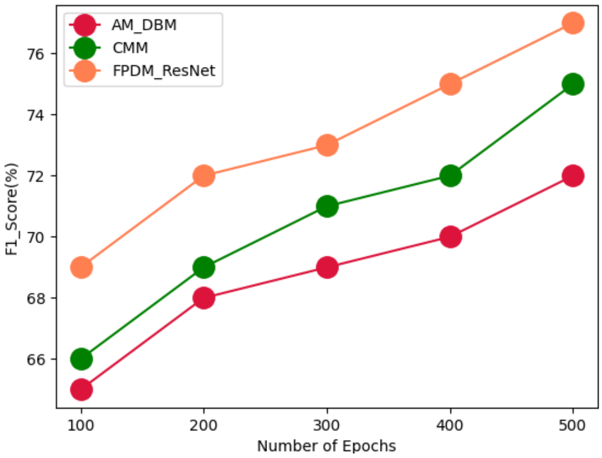

Figure-6 Comparison of F1- score

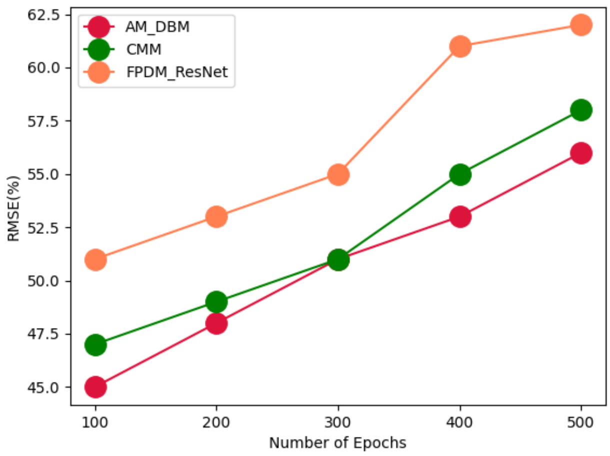

Figure-7 Comparison of RMSE

The above table-1 shows comparative analysis between proposed and existing technique based on clustering and classification for frequent data mining. Here the proposed technique has been analysed based on accuracy, precision, recall, F-1 score and RMSE. Existing technique compared are AM_DBM and CMM based on number of epochs. The proposed technique attained accuracy of 95% from figure-3, precision of 77% as shown in figure-4, recall of 65% as shown in figure- 5, F-1 score of 77% from figure 6 and RMSE of 62% by figure 7; while existing AM_DBM attained accuracy of 85%, precision of 71%, recall of 61%, F-1 score of 72% and RMSE of 56%; CMM accuracy of 88%, precision of 73%, recall of 63%, F-1 score of 75% and RMSE of 58%.

5. Conclusion:

The proposed framework designed in this research based on frequent pattern data mining based on sequence using clustering and deep learning architecture. the data has been clustered and classified using fuzzy clustering with back propagation ResNet Convolutional networks. One of the most well-liked difficulties in data mining with several applications is frequent pattern mining. To best of our knowledge, there is no adaptable framework for frequent pattern mining that can decompose relationships between data elements in sequence databases. In this study, we suggest a flexible and comprehensive framework for extracting common patterns from a set of sequences. The experimental analysis has been carried out in terms of accuracy, precision, recall, F-1 score and RMSE. Proposed technique attained accuracy of 95%, precision of 71%, recall of 65%, F-1 score of 77% and RMSE of 62%.

References

[1] Cheng, H., Yan, X., Han, J., & Hsu, C. W. (2006, April). Discriminative frequent pattern analysis for effective classification. In *2007 IEEE 23rd international conference on data engineering* (pp. 716-725). IEEE.

[2] Hsu, M. J., Chien, Y. H., Wang, W. Y., & Hsu, C. C. (2020). A convolutional fuzzy neural network architecture for object classification with small training database. *International Journal of Fuzzy Systems*, *22*(1), 1-10.