

Geneology Pattern Classifier Technique to Analysis Frequently Changing Uncertain Attribute Mining in Foreign Exchange Trading

^{I,III}Dr R.S. Vetrivel, ^{II}R. Geetha, ^{III}P.Jeyanthirani

^{I,III}Professor, Computer Science, Subramanya College of Arts & Science, Palani.

^{II}Research Scholar, Computer Science, Subramanya College of Arts & Science, Palani

Abstract

Data mining is the process of detect and examine the reliable patterns from large data sets in data base. Data mining method is developed for discovering the useful information which determines significant pattern and rules. In data management, mining the uncertain data is a major concern due to the indefinite nature of the data.

Existing system was presented a technique as Multi-viewpoint-based similarity measure for efficient clustering. In this method, two clustering criterion functions are described for multi-viewpoint based similarity measure. The existing method utilizes many different viewpoints, when the objects are not in the same cluster with the two objects being measured. The similarity measurement is effectively done while using of multiple view points based clustering. However, the existing cluster model properties are fixed.

In order to overcome these problems, we propose a technique for analyzing frequently changing uncertain attribute mining in foreign exchange trading based on geneology pattern classifier technique. In this method, document categorization is developed to classify the documents into groups based on the criterion. The geneology pattern classifier technique is combined with probabilistic database to select the frequently changing attributes. Finally, the uncertain attributes are extracted from the groups of documents which are related to the query from user.

Key Terms

Data Mining, Extraction, Uncertain Data, Attributes Probability, Geneology Pattern Classifier Technique.

I. Introduction

1. Data mining

Data Mining is defined as mining information (knowledge) from huge sets of data. Data mining is used to analyze the data from various perspectives and for summarizing it into valuable information. In data mining, unusual data records are discovered by identification of patterns and trends in data collected using classification, association, and clustering rules.

The aim of the data mining is the discovery of knowledge in Databases (KDD).

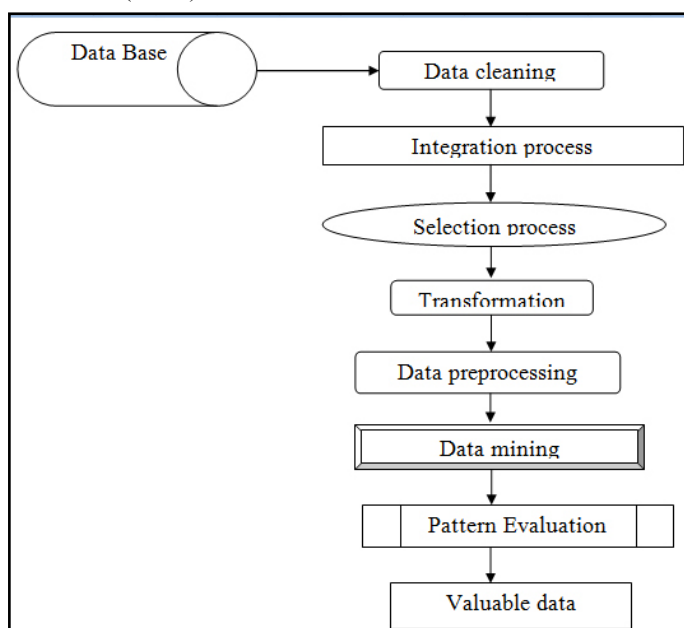


Fig. 1.1: Structure of Data mining process

From figure 1.1 we analysis the process of data mining to extract

valuable data information from the database. Initially, removing of noise and inconsistent data can be done by data cleaning process. Then data integration is developed to integrate (compile) multiple data sources form data base. The integrated data is carried out by data selection process to select relevant data. In transformation, summary normalization aggregation operations are performed to combine the data. Pre-processing is essential to analyze the multivariate data sets before data mining. Then data mining process is used to discover the knowledge of patterns. Finally the significant patterns of information are evaluated by pattern evaluation.

2. Uncertain Attribute Mining

The uncertain data is the data which contains noise that makes it deviate from the original values. Data uncertainty is the reason for various difficulties in the data base applications like data randomness and incompleteness, delay or loss of data updates. Discovering of frequent pattern in data base is technically challenging function because which contains huge amount of interesting information. Therefore mining of uncertain data is the most significant task in the data management. It can be achieved by using the statistics based model driven approaches.

The uncertainty of data must be handled in data management to reach the accurate results. The frequent patterns can be extracted by the probabilistic manner. Probabilistic database can have an exponential number of possible words. Probabilistic database describes the details about which query is issued by only once and which query is executed for an extensive amount of time.

3. Geneology Pattern classification

Geneology pattern classification is developed to analysis the frequently changing uncertain attribute by describing the attributes of genes. Pattern classification methods are used to support the research which related to gene products and functions. Geneology

pattern classification technique is employed to generate efficient document categorization.

Geneology pattern classification is established the patterns into groups of patterns which sharing the same set of properties. These properties are not fixed which contain criteria such as structure, intent, or applicability. Based on the chosen criteria, the patterns are easily grouped. In order to solve these kinds of norms, pattern classification is done by categorizing document in data management.

Initially, the user sent the query to database and related information of the query is extracted from text using the dictionary-based text mining. Once the text are extracted then preprocessing is done to remove the non-functional characters like stop words, comma, etc, from the extracted information. After completing of preprocessing, set of entities identified in the preprocessed text from text corpus and measure.

II. Literature Review

1. Mining significant association rules from uncertain data

In this paper [1], the author proposes a new technique for statistical significance test on association rules with uncertain data. The main aim of this method is to enhancing the reliability of association rule mining by recovering true resultant rules lost due to random error in data. Through computational procedures of the statistical test model, the author estimates a method to express the data error propagation. Based on error model, the combined analytic and simulative processes was described to attain an accurate result in statistical test for distortions and control the risk of spurious rules which caused by data error.

2. Efficient Mining of Frequent Item Sets on Large Uncertain Databases

The author was presented a technique in [2] to resolve the problem of discovering frequent item sets from the large uncertain database by constructing the possible world semantics (PWS). This technique is used to extract the threshold-based PFI results from large uncertain databases. In this work, the mining process can be done by Poisson binomial distribution to develop an approximate algorithm. The author introduced an incremental mining algorithm, for maintaining the mining result which enables to refresh the Probabilistic Frequent Item set (PFI) results. It reduces the need of re-executing the whole mining algorithm on the new database. Poisson binomial distribution and incremental mining algorithm are aids to pick out the frequent item sets in large database efficiently and accurately.

3. Mining Probabilistically Frequent Sequential Patterns in Large Uncertain Databases

In this paper [3], the author introduces a method to measure pattern frequentness, based on the possible world semantics. This technique proposes an algorithm as collectively called U-Prefix Span, for p-FSP mining. U-Prefix Span algorithm is used to resolve the issues of possible world's explosion effectively.

Four pruning and validating methods are combined to achieve better performance in mining frequent patterns These works estimated with two uncertain sequence data models which abstracted from many real-life applications involving uncertain sequence data, and formulate the problem of mining probabilistically frequent sequential patterns (or p-FSPs) from data.

4. EnBay: A Novel Pattern-Based Bayesian Classifier

The paper [4] was presented an effective approach to estimate the Bayesian probability. Based on Entropy-based Bayesian classifier, named as EnBay, the proposed method is introduced to select the minimal set of long and not overlapped patterns which are complies with a conditional-independence model. Long, frequent, and disjoint item sets are selected with help of entropy-based heuristics. This method is used to attain frequent item sets which the basis of joint Bayesian probability.

5. An Uncertainty-based Approach: Frequent Item set Mining from Uncertain Data with Different Item Importance

In this paper [5], the author was develops a technique to efficiently mining the frequent items from uncertain database. This method introduce a new uncertain item set mining algorithm which considering the important of items such as weight constrains. More meaningful item sets with high importance and existential probabilities can be obtained from uncertain database by this proposed method.

6. Probabilistic Convex Hull Queries over Uncertain Data

The author proposes a technique as Probabilistic Convex Hull (PCH) query in this paper [6]. Where location uncertainty is incurred by GPS devices or sensors, the proposed method can be exhibit applications such as Flickr landscape photo extraction and activity region visualization.

Based on geometric properties, the author develops an $O(N^3)$ algorithm, for solving the possible world explosion issue. $O(N^2 \log N)$ algorithm also developed to achieve better time complexity by maintaining a probability oracle in the form of a circular array with nice properties.

7. Clustering Uncertain Data Based on Probability Distribution Similarity

The author develops a model for uncertain objects in both continuous and discrete domains. The Kullback-Leibler divergence is introduced to measure the similarity between uncertain objects in both the continuous and discrete cases integrate it into partitioning.

Density-based clustering methods are used to cluster the uncertain objects. The proposed method is to estimate KL divergence in the continuous case by kernel density estimation to facing the cost and infeasible. KL divergence is helps to employ the fast Gauss transform technique to further speed up the computation.

8. FUDT: A Fuzzy Uncertain Decision Tree Algorithm for Classification of Uncertain Data

The proposed method as Fuzzy Uncertain Decision Tree Algorithm (FUDT) was developed to classify the uncertain data. In this paper [8], fuzzy entropy and similarity measure is introduced to characterize the uncertain data by using binary decision tree algorithm. For handle the uncertain data, Fuzzy entropy is used in proposed method to discover the best split point for the decision tree.

The mining of uncertain data can be improved with high accuracy by Similarity measure. Initially, select the best feature vector by the applying the fuzzy entropy for each feature vector. Then, from the selected feature vector, best split is selected. The binary tree starts to grow through the trained uncertain data. With the help of testing

phase of uncertain data, the constructed decision tree is evaluated by selecting the split point. The classified data can be obtained by subjecting the testing data into trained decision tree.

9. Mining Interesting Patterns from Uncertain Databases

In this paper, the author introduces a model as Weighted Uncertain Interesting Pattern Mining (WUIPM), in which a tree structure (WUIP-tree). Mining the correlated patterns from uncertain databases can be done by some new measures (e.g., uConf, wUConf).

Prefix proxy value and pProxy for WUIP-tree is used to enhance the mining performance. The proposed work is mainly developed to consider weight or importance of an individual item and correlation between items of patterns in uncertain databases.

10. A New Efficient Approach for Mining Uncertain Frequent Patterns using Minimum Data Structure without False Positives

Based on data structures and mining techniques, the author proposes an efficient algorithm in this paper [10] for mining the uncertain frequent patterns. This proposed method was developed to enhance the accuracy of the mining results without any false positives patterns.

The author was introduce a technique as list-based data structures and pruning techniques which permits to mining the set of uncertain frequent patterns efficiently without any pattern losses. The method was designed to perform the uncertain data mining operations efficiently with less runtime and memory resources.

11. Predicting the Pro-Longevity or Anti-Longevity Effect of Model Organism Genes with New Hierarchical Feature Selection Methods

In this work [11], the author introduces hierarchical feature selection methods for the classification task of data mining. These methods are applied to ageing-related data from four model organisms: *Caenorhabditis elegans* (worm), *Saccharomyces cerevisiae* (yeast), *Drosophila melanogaster* (fly), and *Mus musculus* (mouse).

The important part of the proposed feature selection method is to develop hierarchical relationships in the set of features (Gene Ontology terms). Transfer of classify model organisms genes into pro-longevity or anti-longevity genes exploited for improving the accuracy of the Naive Bayes and 1-Nearest Neighbour (1-NN) classifiers. The proposed method was developed to provide more accuracy than the standard (without feature selection) Naive Bayes and 1-NN classifiers.

12. Maximum Likelihood Estimation from Uncertain Data in the Belief Function Framework

In this paper [12] the author presents a technique as Maximum Likelihood Estimation for Uncertain Data in the Belief Function Framework. Based on the maximization of a generalized likelihood criterion the proposed method was developed to understand the degree of agreement between the statistical model and the uncertain observations.

In this method the author introduces a alternative of the EM algorithm to maximizes the criterion. For the categorical and continuous attributes, the proposed method is used to cluster the uncertain data with finite mixture models.

13. GOFAM: a hybrid neural network classifier combining fuzzy ARTMAP and genetic algorithm

In this paper [13] the author proposes a genetic algorithm to discover a better presentation order of training data for FAM. Genetic algorithm is combined with Fuzzy ARTMAP which is called Genetic Ordered Fuzzy ARTMAP (GOFAM). Some standard datasets from UCI repository of machine learning database are experimented for exhibiting the effectiveness of GOFAM. By evaluating and comparing the results from FAM, fixed order of training pattern presentation are estimated to FAM.

14. Modified AHP for Gene Selection and Cancer Classification using Type-2 Fuzzy Logic

The author presented a paper [14] to modify the analytic hierarchy process (AHP). This proposed technique was developed for cancer classification by selecting the most informative genes that serve as inputs to an interval type-2 fuzzy logic system (IT2FLS). The modified AHP was designed to permits the process for quantitative factors which are ranking results of individual gene selection methods that including t-test, entropy, receiver operating characteristic curve, Wilcoxon test and signal to noise ratio. The modified AHP enhances THE performance of classification not only for the IT2FLS but also for all other classifiers.

15. An Extensive Empirical Comparison of Probabilistic Hierarchical Classifiers in Datasets of Ageing-Related Genes

In this paper [15], the author develops a technique for predicting Gene Ontology (GO) terms related to ageing. This proposed method was estimated with 5 types of probabilistic hierarchical classification methods. The accuracy of the results can be improved by Local Hierarchical Classifier (LHC) and Predictive Clustering Tree algorithm (LHC-PCT). The hierarchical classification algorithm is used to test the impact of two types of variations (i) changing the base algorithm (we tested Naive Bayes and Support Vector Machines), and for pre-processing the impact of (ii) using or not the Correlation based Feature election (CFS) algorithm.

16. Adaptive multi-classifier fusion approach for gene expression dataset based on probabilistic theory

In this paper [16], we propose a fusion approach for gene expression dataset based on probabilistic theory. For improving performance and to optimize the parameters, the Naive Bayesian classifier is employed in this proposed work. The Naive Bayesian classifier performs like a supervisor. The author develops the technique which uses the base classifiers as Particle Swarm Optimization-Functional Link Neural Network (PSO-FLANN), Bat inspired-Functional Link Artificial Neural Network (BAT-FLANN) and Support Vector Machine (SVM). Through the objective functions like accuracy, parallel execution, time complexity and space complexity, the performance of proposed work was compared with Uniform Voting, Distribution Summation, Dempster-Shafer, Entropy Weighting and Density based Weighting on six benchmark gene expression datasets.

17. Informative gene selection and the direct classification of tumors based on relative simplicity

Based on measure of relative simplicity, the author proposes a technique in the paper [17] for the Informative gene selection and direct classification of tumors. In order to integrating vertical comparison with horizontal comparison, gene pairs are evaluated.

It constructs the RS-based direct classifier (RS-based DC) by the set of informative genes binary discrimination with paired votes strategy.

18. Gene Committee: a web-based tool for extensively testing the discriminatory power of biologically relevant gene sets in microarray data classification

The author develops a technique as geneCommittee for the following purposes a) To integrate the set of well-known classification methods b) enables the enrichment of original dataset variables with biologically relevant features c) based on the biological relevance and discriminative power, selecting the classifiers with the help of training classification solutions d) allow to apply the experts, a mix of classifier/variable, to diagnose new patients.

19. A DNA-based pattern classifier with in vitro learning and associative recall for genomic characterization and bio sensing without explicit sequence knowledge

In this paper [19], the author introduces an in vitro Biological Memory for confining and to store genomic information from biological samples as known and unknown. In vitro Biological Memory is also used to classify and compare their genomic patterns which have the capability to work with unknown organisms and sequence.

20. Molecular characterization of Gleason patterns 3 and 4 prostate cancer using Reverse Warburg effect-associated genes

In this paper [20], the author introduces two potential classifiers as stromal FOXO1 and stromal ATG5/GLUT1 which are used to determine the difference of aggressive and indolent forms of PCa. RWE-related with genes can be used to distinguish among GP3 and GP4 prostate cancers. The Reverse Warburg effect (RWE) is used to analyze the physiologic results to the epithelial phenotype which is independent to the epithelial genotype and of direct tumor sampling.

III. Methodology

Geneology pattern classifier technique is developed to effectively analyze the frequently changing uncertain attribute mining which improves the accuracy of performance. The main goal of clustering is to discover fundamental structures of data and to organize them into significant subgroups for future study and analysis purposes.

The existing method was developed to achieve effective clustering by multi view point based similarity measures. Multi view point clustering is used to measure the similarity between data objects in sparse and high dimensional domain. Most clustering methods are used to assume the cluster relationship between data objects. The dissimilarity/similarity measure is employed by single viewpoint which is the origin.

Similarity between a pair of objects was defined either explicitly or implicitly. The multi-view point similarity measurement uses the more than one point of reference. The author formulates the two clustering criterion functions for multi-view point clustering. In this method, the optimization algorithm is used for the document clustering.

The Clustering with Multi-view point based similarity measure technique have the following disadvantages

- Impractical to go through all the details of all the patterns.

- Existing clustering model unable to perceive main properties of patterns.
- Unable to structure between relationships pattern.
- Existing cluster modeled properties is fixed.

In Foreign exchange trading, to extract the frequently changing uncertain attributes, the proposed method is developed with technique as Geneology Pattern Classifier Technique.

Prediction of sequential data plays a major role in Foreign Exchange (FOREX) market data analysis. Geneology pattern classifier technique is used in this proposed method for discovering the user/object behavior information or the changing information in a multiple data set in data base.

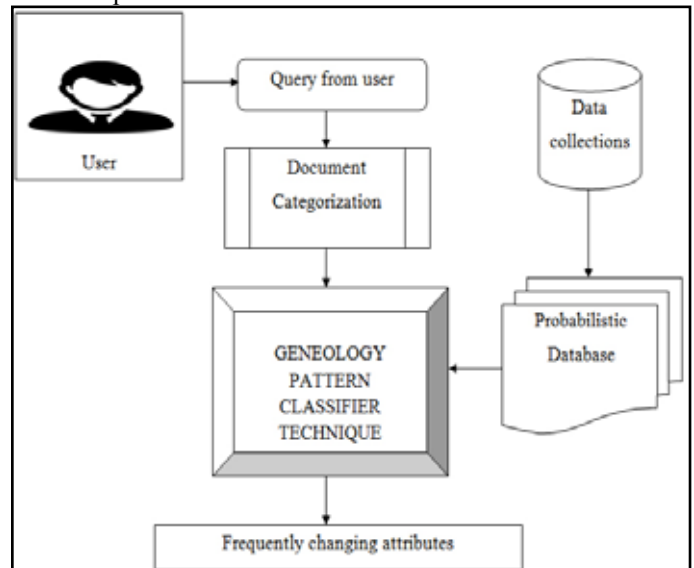


Fig. 3.1: Architecture of Geneology Pattern Classifier technique

From figure 3.1 we analyze the process for mining the uncertain attributes which are frequently changed in uncertain database. Initially, the query request from user is given to the document classification block. The document classification is used to group the documents based on the query. Then the geneology pattern classifier technique is developed to discover the attributes which are frequently changed with the help of Probabilistic database. Probabilistic database described the details about which query is performed for number of time in foreign exchange trading. It provides the frequently changing uncertain data from database efficiently.

Advantages

- The accuracy of the document classification can be enhanced for analyzing the FOREX values.
- The processing time to detect various changes of attribute is reduced by this proposed method using probabilistic based decision model.
- This proposed method is developed to minimize the processing time required for classification.

A. Modules

- Document Categorization
- Geneology pattern classifier technique for mining the frequently changing attributes
- Foreign exchange trading

B. Modules Description

1. Document categorization

Document classification or document categorization is the most significant task that assigns documents to one or more classes or categorizes. The documents can be classified such as texts, images, music, etc. Each type of document possesses its special classification overlapping problems.

The proposed method is developed with efficient document categorization to overcome these difficulties. Based on the chosen criteria or class, document categorization extracts the conceptual content of text items with knowledge which attributes are frequently changed in database. The request based approach is used in this method to discover required information.

An efficient method to classify patterns is to categorize them according to their purpose. This criterion is the most useful which describes the concrete situation. Documents may be classified according to their subjects or according to other attributes. The frequent patterns of attributes can be extracted in probabilistic manner from probabilistic database. Probabilistic database estimates the details about which query is issued by only once and which query is executed for an extensive amount of time.

2. Geneology Pattern Classifier technique for mining the frequently changing uncertain attributes

Geneology pattern classifier organizes patterns into groups based on the properties patterns which sharing same set of properties. Pattern classification technique is developed for document collection to use classification scheme. It helps to classify and selecting the patterns. Geneology pattern classifier scheme needs to be compact in terms of the number of criteria reduce complexity and uncertainty. Classification schema represents main properties of patterns to be classified.

Architectural patterns: specify the fundamental structure of applications. They specify structure of the system as subsystems, and their responsibilities, and define rules for relationships between them.

Design patterns: focus on subsystems and components, the middle level of a software system. Design patterns describe common problems arising in a given context of particular subsystem.

Coding patterns (idioms): use proven conventions and techniques during the implementation phase of applications.

3. Foreign Exchange Trading

This proposed work builds a strong consumer focus on retail, financial, communication, and marketing organizations in foreign exchange trading. It make easy for foreign exchange trades to determine relationships among internal features such as product positioning, cost of products and staff skills, and external features such as economic indicators, customer demographics and competition. And it also assists aids to assume the impact on corporate profits, sales, and customer satisfaction. According to, these types of details, we can identify the frequently changing uncertain attributes from data management with help of geneology pattern classifier technique.

For the following purposes, the proposed work is developed in Foreign exchange trading,

- Retailer to develop products and promotions for request to specific customer segments based on individual's purchase history to determine customer purchasing pattern.
- To analysis the details about what kind of products are bought by which kind of people.
- For identifying Customer Requirements- which helps to

discover the factors that may attract new customers.

- It also aids to Cross Market Analysis to perform correlations between product sales.
- To find out clusters of model customers who share the same characteristics such as interests, income.

Input: Database 'D', Number of queries 'Q'
Output: Identify the frequently changing uncertain attributes
Step 1: Begin Step 2: Enter the required text Step 3: Document categorization classify the documents into groups based on chosen criteria Step 4: Probabilistic database collects the documents from server which query is executed for more amount of time Step 5: Geneology Pattern Classifier provides results which are related documents to queries from server Step 6: The frequently changing uncertain attributes are discovered to the user Step 7: End

Fig. 3.2 : Algorithm for Geneology Pattern Classifier technique

IV. Experimental Evaluation

In this paper worked on performance evaluation in terms of Attributes probability, Accuracy, Processing time. The performance metric to evaluates and analyze the value in java environment simulations. This technique improves the performance in analysis of frequently changing uncertain attribute mining. The performance measures of the proposed work are analyzed with following metrics:

- Attributes probability
- Accuracy
- Processing time

1. Attributes Probability

Attributes probability is used for analyzing the document to gives the information about which query is executed for number of time.

They will be checking the folders to identify the frequently changed uncertain attributes.

Table 4.1: Tabulation for Attributes Probability

Document Size	Attributes probability	
	Clustering with multi view point	Geneology Pattern Classifier Technique
5	0.20	0.25
10	0.25	0.30
15	0.30	0.35
20	0.35	0.40
25	0.40	0.45

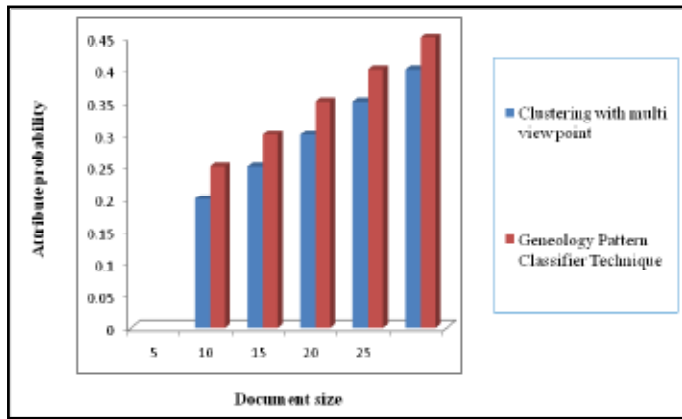


Fig. 4.1: Measurement of Attributes probability

Figure 4.1 demonstrates the Attributes probability. The documents taken for the experimental consideration is varied from 5 to 25. The X axis represent the number of documents as document size. The Y axis represents the attributes probability. From the figure it is clearly evident that the proposed Geneology Pattern Classifier Technique improves the attributes probability than the Existing Clustering with multi view point method. Hence, the attributes probability is improved up to 18% by the proposed Geneology Pattern Classifier Technique than the existing Clustering with multi view point method.

2. Accuracy

The accuracy of the results defined as ratio to the identification of frequently changing uncertain attributes in database with respect to the query request from user. Geneology Pattern Classifier Technique efficiently investigates to provide better search accuracy in foreign exchange trading. The accuracy of result is measured in terms of percentage (%).

Table 4.2: Tabulation for Accuracy of results

Document Size	Accuracy of results (%)	
	Clustering with multi view point	Geneology Pattern Classifier Technique
5	80.2	90.2
10	82.4	92.8
15	84.2	94.6
20	86.8	96.2
25	88.2	100.2

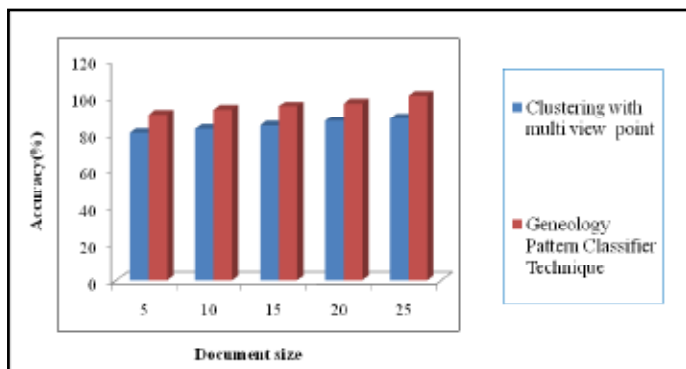


Fig. 4.2: Measurement of Accuracy of results

Figure 4.2 demonstrates the Accuracy of results. The documents taken for the experimental consideration is varied from 5 to 25. The

X axis represent the number of documents as document size. The Y axis represents the accuracy of results. From the figure it is clearly evident that the proposed Geneology Pattern Classifier Technique improves the accuracy of results than the Existing Clustering with multi view point method. Hence, the accuracy of results is improved up to 12% by the proposed Geneology Pattern Classifier Technique than the existing Clustering with multi view point method.

3. Processing Time

Processing time is defined as the amount of time taken for identifying the frequently changed attributes in FOREX trading. Processing time is measured in terms of milliseconds (ms).

Table 4.3 : Tabulation for Processing time

Document Size	Processing Time (ms)	
	Clustering with multi view point	Geneology Pattern Classifier Technique
20	12	14
40	16	20
60	20	26
80	26	32
100	30	40

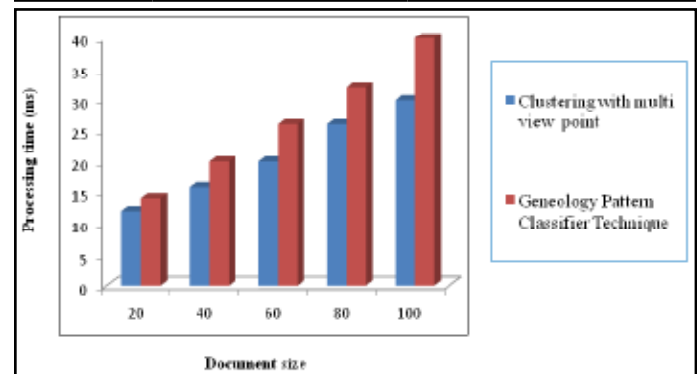


Fig. 4.3: Measurement of Processing time

Figure 4.3 demonstrates the Processing time. The documents taken for the experimental consideration is varied from 20 to 100. The X axis represent the number of documents as document size. The Y axis represents the Processing time. From the figure it is clearly evident that the proposed Geneology Pattern Classifier Technique improves the Processing time than the Existing Clustering with multi view point method. Hence, the accuracy of results is reduced to 26% by the proposed Geneology Pattern Classifier Technique than the existing Clustering with multi view point method.

V. Conclusion & Future Scope

A. Conclusion

Patterns can be classified depending on many criteria by the document categorization. In foreign exchange trading, the geneology pattern classifier is developed to discover the frequently changing uncertain attributes in an efficient manner. By this proposed method, accuracy of the document classification is improved for analyzing the FOREX values. The processing time to detect various changes of attribute is reduced by this proposed method using probabilistic based decision model. This proposed method is developed to minimize the processing time required for

classification.

Further analysis, the techniques are improved to attain frequently changing attributes as results with minimum execution time then the proposed method.

References

- [1] Anshu Zhang ,Wenzhong Shi-Geoffrey I,Webb “Mining significant association rules from uncertain data” *Data Mining and Knowledge Discovery*,Vol 30 Issue 4,July 2016,Pages 928-963.
- [2] Liang Wang, David Wai-Lok Cheung, Reynold Cheng, Member, IEEE, Sau Dan Lee, and Xuan S. Yang “Efficient Mining of Frequent Item Sets on Large Uncertain Databases” *IEEE transactions on knowledge and data engineering*, Vol. 24, no. 12, december 2012, Pages 2170-2183.
- [3] Zhou Zhao, Da Yan and Wilfred Ng “Mining Probabilistically Frequent Sequential Patterns in Large Uncertain Databases” *IEEE transactions on knowledge and data engineering*, Volume 26, Issue 5, May 2014, Pages 1171-1184.
- [4] Elena Baralis, Member, IEEE, Luca Cagliero, and Paolo Garza “EnBay: A Novel Pattern-Based Bayesian Classifier” *IEEE Transactions On Knowledge And Data Engineering*, Vol. 25, No. 12, December 2013, Pages 2780-2795.
- [5] Gangin Lee, Unil Yun and Heungmo Ryang “An Uncertainty-based Approach: Frequent Item set Mining from Uncertain Data with Different Item Importance”, *ELSEVIER, Knowledge-Based Systems*, Volume 90, December 2015, Pages 239-256.
- [6] Da Yan, Zhou Zhao, Wilfred Ng, and Steven Liu “Probabilistic Convex Hull Queries over Uncertain Data” *IEEE Transactions On Knowledge And Data Engineering*, Vol. 27, No. 3, March 2015, Pages 852- 865.
- [7] Bin Jiang, Jian Pei, , Yufei Tao, Xuemin Lin “Clustering Uncertain Data Based on Probability Distribution Similarity” *IEEE Transactions On Knowledge And Data Engineering*, Vol. 25, No. 4, April 2013, Pages 751-763.
- [8] S. Meenakshi, V. Venkatachalam “FUDT: A Fuzzy Uncertain Decision Tree Algorithm for Classification of Uncertain Data” *Arabian Journal for Science and Engineering*, November 2015, Volume 40, Issue 11, pp 3187–3196.
- [9] Akiz Uddin Ahmeda, Chowdhury Farhan Ahmedb, Md. Samiullah, Nahim Adnana, Carson Kai-Sang Leungc “Mining Interesting Patterns from Uncertain Databases” *Information Sciences*, ELSEVIER, Volume 354, 1 August 2016, Pages 60-85.
- [10] Gangin Lee and Unil Yun “A New Efficient Approach for Mining Uncertain Frequent Patterns using Minimum Data Structure without False Positives”, *Future Generation Computer Systems*, ELSEVIER, Volume 68, March 2017, Pages 89-110.
- [11] Cen Wan, Alex A. Freitas, and Jo~ao Pedro de Magalh~aes “Predicting the Pro-Longevity or Anti-Longevity Effect of Model Organism Genes with New Hierarchical Feature Selection Methods” *IEEE Transactions On Knowledge And Data Engineering*, Vol. 12, No. 2, March/April 2015, Pages 262-275.
- [12] Thierry Denoeux “Maximum Likelihood Estimation from Uncertain Data in the Belief Function Framework” *IEEE Transactions On Knowledge And Data Engineering*, Vol. 25, No. 1, January 2013, Pages 119-130.
- [13] Masoud Yaghini · Mohammad Ali Shadmani “GOFAM: a hybrid neural network classifier combining fuzzy ARTMAP and genetic algorithm”, *Artificial Intelligence Review*, March 2013, Volume 39, Issue 3, pp 183-193.
- [14] Thanh Nguyen and Saeid Nahavandi “Modified AHP for Gene Selection and Cancer Classification using Type-2 Fuzzy Logic” *IEEE Transactions on Fuzzy Systems*, 2016, Volume 24, Issue 2, Pages 273-287.
- [15] Fabio Fabris, Alex A. Freitas and Jennifer M. A. Tullet “An Extensive Empirical Comparison of Probabilistic Hierarchical Classifiers in Datasets of Ageing-Related Genes” *IEEE/ACM Transactions on Computational Biology and Bioinformatics* , Volume: 13, Issue: 6, Nov.-Dec. 1 2016, Pages 1045-1058.
- [16] Sashikala Mishra , Debahuti Mishra “Adaptive multi-classifier fusion approach for gene expression dataset based on probabilistic theory”, *ELSEVIER, Journal of the Korean Statistical Society* , Volume 44, Issue 2, June 2015, Pages 247–260.
- [17] Yuan Chen, Lifeng Wang, Lanzhi Li, Hongyan Zhang, Zheming Yuan “Informative gene selection and the direct classification of tumors based on relative simplicity” *BMC Bioinformatics*, Volume 17, Issue 44, 2016, Pages 1-16.
- [18] Miguel Reboiro-Jato1, Joel P Arrais, José Luis Oliveira , Florentino Fdez-Riverola1 “Gene Committee: a web-based tool for extensively testing the discriminatory power of biologically relevant gene sets in microarray data classification” *BMC Bioinformatics*, Volume 15, Issue 31, 2014, Pages 1-11.
- [19] Ju Seok Lee, Junghuei Chen, Russell Deaton, Jin-Woo Kim “A DNA-based pattern classifier with in vitro learning and associative recall for genomic characterization and bio sensing without explicit sequence knowledge” *Journal of Biological Engineering* Volume 8, Issue 25, 2014 Nov 6, Pages 1-12.
- [20] Ilinca Georgescu, Robert J. Gooding, R. Christopher Doiron, Andrew Day, Shamini Selvarajah, Chris Davidson, David M. Berman , Paul C. Park “Molecular characterization of Gleason patterns 3 and 4 prostate cancer using reverse Warburg effect-associated genes”, *Cancer & Metabolism* , Volume 4, Issue 8, 2016, Pages 1-12.