International Journal of Pure and Applied Mathematics
Volume 117 No. 15 2017, 1151-1162
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue
ijpam.eu

Fuzzy Set based Evolving user prediction and classification for Recommendation model

Mr. G.Silambarasan[1]
Research Scholar
CMJ University, Meghalaya, India
gssilambarasan@gmail.com

Mr. J.Anvar Shathik[2]
Assistant Professor
KGISL Institute of Technology, Coimbatore, India
anvarshathik@gmail.com

Abstract

Data Recommendation has stepped up to significant range in recent years for providing the user with choices based on their intension. User profiling is the practice of gathering, organizing, and interpreting the user profile information. Robust and Accurate recommendation is an important aspect in e commerce and marketing based applications. Many prediction based algorithms has been employed in literature for assisting the user with their future actions. Despite of more advantages, those techniques fails to predict the accurate recommendation for evolving user characteristics. In this paper, we propose a novel fuzzy set based user prediction and classification technique. The proposed Approach for characterising and recognising the user evolving patterns and hidden patterns is presented. It allows learning which characteristics are evolving and also predict the evolution trends. Fuzzy based classifier is employed to learn the user characteristics to obtain the class in order to cluster the data. Iterative model is embedding in this technique in order to generate the potential clusters to the data under streaming. The proposed model can be applicable to any problem of dynamic/evolving user behaviour modeling. The experimental results indicates that proposed system outperforms the state of art approaches in terms of accuracy and efficiency

Keywords: Fuzzy Classifier, Fuzzy Set, Fuzzy Rules, User Classification, Evolving Data

1. Introduction

Recommendation systems have been widely used to provide users with high-quality personalized recommendations about the contents from a large volume of choices [1]. The recognition of users can be very beneficial for assisting them or predicting their future actions. Most existing techniques for user recognition assume the supervised classification of the user profiles. User profiles can be expressed as the description of the user interests, characteristic and preferences[2].

In this paper, adaptive approach for predict the user behaviour and behaviour is classified based on the characteristic evolution[3]. The learning of the user data is considered as adaptive distributions. Fuzzy rules present an evolving method for updating and evolving the user profiles and classifying an observed user. It is constantly learning and adapting the existing classifier structure to accommodate the newly observed emerging behaviours. This aspect motivates the idea of automated sequence learning for data classification.

The descriptions of a particular behaviour itself may also evolve, so we assume that each behaviour is described by one or more fuzzy rules[4]. A conventional system does not capture the new patterns (behaviours) that could appear in the data stream once the classifier is built. In addition, the information produced by a user is often quite large. Therefore, we need to cope with large amounts of data and process this information in real time, because storing the complete data set and analyzing it in an offline (batch) mode would be impractical. In order to take into account these aspects, we use an evolving fuzzy-rule-based system that allows for the user behaviours to be dynamic and to evolve.

The remainder of the paper is organized as follows: Section 2 discusses the related works in instance integration and its impacts against the performing user data evolving, Section 3 briefly discusses the proposed technique in terms fuzzy rules for prediction and classification of the user data and Section 4 presents the experimental results on a number of data sets. Section 5 discusses conclusions and future work.

2. Related Works

There exist many techniques to user prediction are designed and implemented efficiently. Each of these techniques follows some sort of user mining, user prediction on the evolving data among few performs nearly equivalent to the proposed framework, which is described as follows

2.1.Sequential Minimal optimization

Sequential Minimal Optimization is used for training support vector machines by breaking a large quadratic programming (QP) optimization problem into a series of smallest possible QP problems [5].It present a learner with unlabeled sequential data that discover meaningful patterns of sequential behaviour from example streams. It can work with large user profile data.

2.2. Evolving user behaviour classification

It is Incremental algorithms build and refines the model at different points in time, in contrast to the traditional algorithms which perform the model in a batch manner. It is more efficient to revise an existing hypothesis than it is to generate hypothesis each time a new instance is observed. Therefore, one of the solutions to the proposed scenario is the incremental classifiers [6].

3. Proposed System

In this section, we describe the hybrid data reliability linkage mechanism using hybrid classifier such as Support Vector Machine and Expectation Maximization Technique on event matching. This process is represented as follows

3.1.Design Goals

The learning algorithm can be defined in terms of following utilizes in order to produce the effective results.

1. It should be able to learn additional information from new data.
2. It should not require access to the original data, used to train the existing classifier.
3. It should preserve previously acquired knowledge.
4. It should be able to accommodate new classes that may be introduced with new data.

3.2. Fuzzy Based Rules for Prediction of evolving user profiles

The Fuzzy based rules has important advantages which make it very useful in real environments for predicting the evolution of user profiles.

1. It can cope with huge amounts and data.
2. Its evolving structure can capture sudden and abrupt changes in the stream of data.
3. Its structure meaning is very clear, as we propose a rule-based classifier.
4. It is noniterative and single pass; therefore, it is computationally very efficient and fast.

5. Its classifier structure is simple and interpretable.

The fuzzy based rules are used for automatic clustering of the user data evolutions. The proposed system takes into account the fact that the behaviour of any user is not fixed, but is rather changing. Although the proposed approach can be applied to any behaviour represented by a sequence of characteristics changes.

3.2.Feature Extraction

Feature can be considered as behaviour of the user. The Feature extraction detects the features of the evolving data. The Feature is represented in form vector space model. In Feature extraction model we employ wordnet to determine its concept and semantic of the data to be predicted. Frequency based method is employed to detect the number of similar features for an aspect.

3.3. Classifier modelling based Fuzzy Rules

A classifier is a mapping from the feature space to the class label space. In the proposed classifier, the feature space is defined by distributions of subsequence’s of events. On the other hand, the class label space is represented by the most representative distributions. Thus, a distribution in the class label space represents a specific behaviour which is one of the prototypes of the user. The prototypes are not fixed and evolve taking into account the new information collected online from the data stream The number of these prototypes is not prefixed but it depends on the homogeneity of the observed behaviours. The following sections describe how a user behaviour is represented by the proposed classifier.

3.3.1. Class Representation

Features distributions must be represented in a Training phase in order to generate a class. Each feature distribution is considered as a data vector that defines a point that can be represented in the Feature set. The data space in which we can represent these points should consist of n dimensions, where n is the number of the different subsequence’s that could be observed. It means that we should know all the different subsequence’s of the data a priori. The dimension of the data space also evolves; it is incrementally growing according to the different subsequence’s that are represented in it

A prototype is a feature set (a behaviour represented by a distribution of subsequence's of commands) that represents several samples which represent a certain class. Then, each data sample is classified to one of the prototypes (classes) defined in the classifier. Finally, based on the potential of the new data sample to become a prototype. it could form a new prototype or replace an existing one.

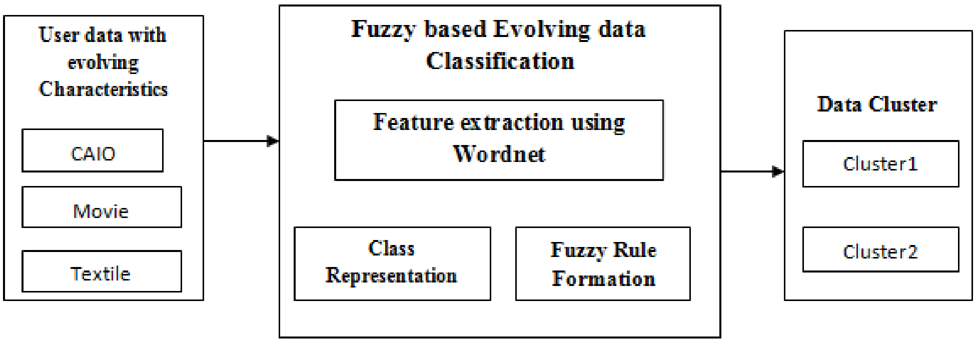

Figure 1: Architecture Diagram of the proposed model

The potential of the data sample which represents a function of the accumulated distance between a sample and all the other k -1 samples in the data space is represented in figure 1.

The data point difference is calculated using the Euclidean Distance. It is calculated using the cosine distance. Cosine distance has the advantage that it tolerates different samples to have different number of attributes; in this case, an attribute is the support value of a subsequence data.

The proposed fuzzy rule based classifier has been explained taking into account that the observed data samples do not have labels; thus, using unsupervised learning. In this case, the classes are created based on the existing prototypes and, thus, any prototype represents a different class. the observed data samples can have a label assigned to them a priori

Specific class is represented by several prototypes, because the previous process is done for each different class. Thus, each class has a number of prototypes that depends on how heterogeneous are the samples of the same class. In this case, a new data sample is classified in the class which belongs the closest prototype.

4. Experimental Analysis

In section, we describe the experimental results of the proposed framework against the existing illustrates that proposed mechanism outperforms the existing approaches in terms of Precision , Recall ,F1 score and Computation Time.

4.1.Dataset Description

We have done extensive experiments on 3 real datasets for evolving data streams is as follows

4.1.1. Movie Dataset

Movie Rating is widely used movie recommendation dataset. It contains 100,000 movies with ratings scale of 1–5 for different varieties of genre. The data set after which transformed into matrix format for further analysis.

4.1.2. Textile Dataset

It is a dataset composed of the rating to the different product with quality factor like size, price, and color. It contains 100 rating values which is given by the 10 different product like Allen solely, Louis Philip, Levis etc with rating scale of 1 to 5

4.1.3. Ciao dataset

It is likely to be the publicly available data set that contain both item ratings and social relationships specified by active users

4.2.Evaluation

The proposed Framework is evaluated against the following measures Pairwise Completeness (PC), Precision (P); Recall(R), F1-scorecs (F) and computation time.

4.2.1. Precision

Positive predictive value is the fraction of relevant instances among the retrieved instances. Precision is the number of correct feature divided by the number of all returned feature space.

$$\text{Precision} = \frac{\text{True positive}}{True\ positive + False\ Positive}$$

True positive is a number of real positive cases in the data and false negative is number of real negative cases in the data. The precision is evaluated against different dataset is depicted in the figure 2

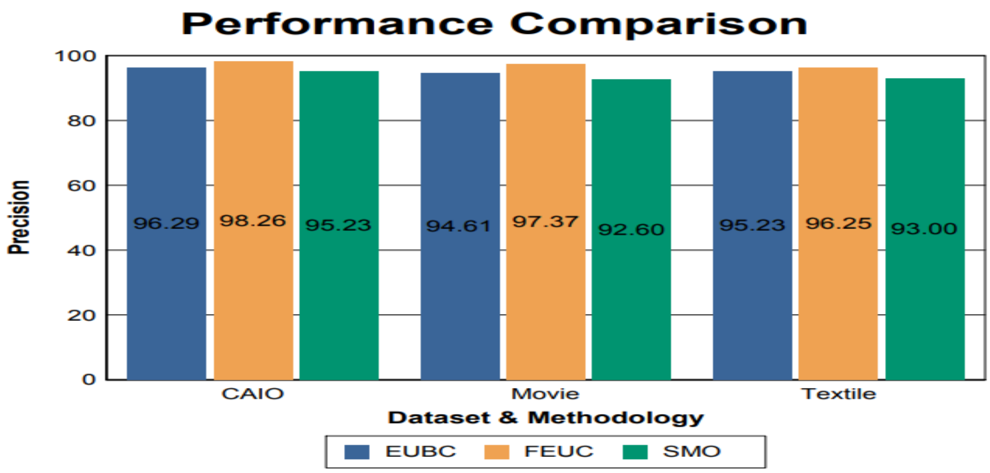

Figure 2: Performance Evaluation of the Methodologies on Precision against the different datasets

4.2.2. Recall

It is the fraction of relevant instances that have been retrieved over the total amount of relevant instances. The recall is the part of the relevant documents that are successfully classified into the exact classes

$$\text{Recall} = \frac{\text{True positive}}{\textit{True positive} + \textit{False negative}}$$

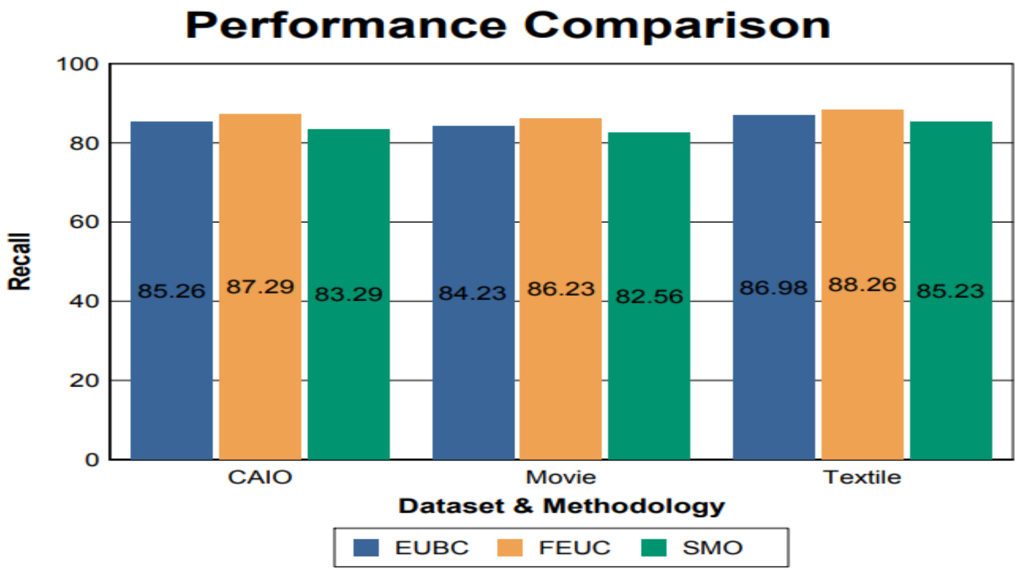

Figure 3: Performance Evaluation of the Methodologies on Recall against the different datasets

True positive is a number of real positive cases in the data and false negative is number of real negative cases in the data. The recall is evaluated against different dataset is depicted in the figure 3

4.2.3. F Measure

It is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test. The performance chart is described in figure 5 and its value indicated in Table 1.

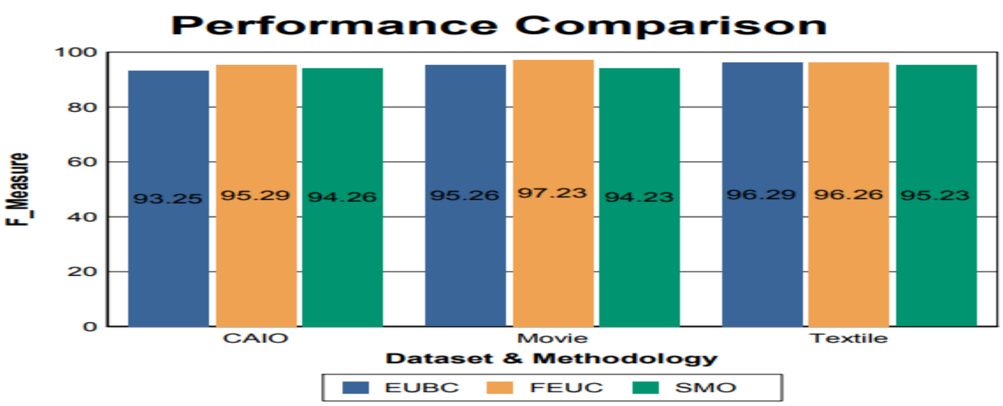

Figure 5: Performance Evaluation of the Methodologies on F Measure against the different datasets

4.2.4. Computation Time

It is defined as no of time taken to establish the instance matching for the different lexical words between the two heterogeneous sources.

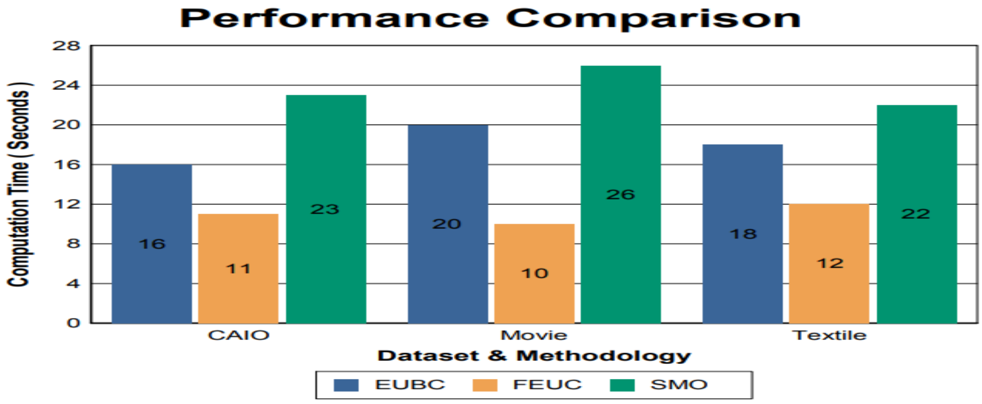

Figure 6: Performance Evaluation of the Methodologies on Computation Time against the different datasets

The performance evaluation chart and its values is described in figure 6 and Table 1.

Table 1: Performance Comparison of Methodology against measures for various dataset

Dataset	Technique	Precision %	Recall %	Fmeasure %	Computation Time (s)
Movie Dataset	FEUC SMO EUBC	97.37 92.6 94 .61	86.23 82.56 84.23	97.23 94.23 95.26	10 26 20
Textile Dataset	FEUC SMO EUBC	96.25 93 95.23	88.26 85.23 86.98	96.26 95.23 96.29	12 22 18
CAIO Dataset	FEUC SMO EUBC	98.26 95.23 96.29	87.29 83.29 85.26	95.29 94.26 93.25	11 23 16

As shown in Table 1, FEUC techniques enable the entire process to run 2-3 orders of magnitude faster than SMO and EUBC

In order to examine the impact of the different components of our proposed technique greatly increased the runtime for both candidate selection and the prediction process also it is been proved in terms of the fuzzy rules.

Conclusion

We designed and implemented the fuzzy based evolving data classification and prediction. The Model classifies user behaviours from a sequence of events. The underlying assumption in this approach is that the data collected from the dataset can be transformed into a sequence of events. The proposed evolving classifier is evaluated in terms of precision, Recall and f measure and reports its accuracy compared with existing techniques. Classifier can cope with huge amounts of data and the proposed approach is the most suitable alternative for prediction of user evolution.

References

[1] J.A. Iglesias, A. Ledezma, and A. Sanchis, “Creating User Profiles from a Command-Line Interface: A Statistical Approach,” Proc. Int’l Conf. User Modeling, Adaptation, and Personalization (UMAP), pp. 90-101, 2009.

[2] A. Cufoglu, M. Lohi, and K. Madani, “A Comparative Study of Selected Classifiers with Classification Accuracy in User Profiling,” Proc. WRI World Congress on Computer Science and InformationEng. (CSIE), pp. 708-712, 2009.

[3] F.J. Ferrer-Troyano, J.S. Aguilar-Ruiz, and J.C.R. Santos, “Data Streams Classification by Incremental Rule Learning with Parameterized Generalization,” Proc. ACM Symp. Applied Computing (SAC), pp. 657-661, 2006.

[4] R.K. Agrawal and R. Bala, “Incremental Bayesian Classification for Multivariate Normal Distribution Data,” Pattern Recognition Letters, vol. 29, no. 13, pp. 1873-1876, http://dx.doi.org/10.1016/ j.patrec.2008.06.010, 2008.

[5] J.A. Iglesias, A. Ledezma, and A. Sanchis, “Sequence Classification Using Statistical Pattern Recognition,” Proc. Int’l Conf. Intelligent Data Analysis (IDA), pp. 207-218, 2007.

[6]. Bo Tang, Steven Kay, Haibo He "Toward Optimal Feature Selection in Naive Bayes for Text Categorization “in IEEE Transactions on Knowledge and Data Engineering in Volume: 28, Issue: 9, Sept. 2016

[7] H. Ma, D. Zhou, C. Liu, M. Lyu, and I. King, “Recommender systems with social regularization,” in Proceedings of the 4th ACM International Conference on Web Search and Data Mining (WSDM), 2011, pp. 287–296.

[8] Y. Koren, “Factor in the neighbors: Scalable and accurate collaborative filtering,” ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 4, no. 1, pp. 1:1–1:24, 2010.

[9] P. Wang, H. Wang, X. Wu, W. Wang, and B. Shi, “A Low- Granularity Classifier for Data Streams with Concept Drifts and Biased Class Distribution,” IEEE Trans. Knowledge and Data Eng., vol. 19, no. 9, pp. 1202-1213, Sept. 2007.

[10] I. Katakis, G. Tsoumakas, and I. Vlahavas, “Dynamic Feature Space and Incremental Feature Selection for the Classification of Textual Data Streams,” Proc. Int’l Workshop Knowledge Discovery from Data Streams (ECML/PKDD), pp. 102-116, 2006.

Rajesh, M., and J. M. Gnanasekar. "Annoyed Realm Outlook Taxonomy Using Twin Transfer Learning." International Journal of Pure and Applied Mathematics 116 (2017): 547-558.

Rajesh, M. & Gnanasekar, J.M. Wireless Pers Commun (2017),https://doi.org/10.1007/s11277-017-4565-9

Rajesh, M., and J. M. Gnanasekar. "GCCover Heterogeneous Wireless Adhoc Networks." Journal of Chemical and Pharmaceutical Sciences (2015): 195-200.

Rajesh, M., and J. M. Gnanasekar. "CONGESTION CONTROL IN HETEROGENEOUS WANET USING FRCC." Journal of Chemical and Pharmaceutical Sciences ISSN 974: 2115.

Rajesh, M., and J. M. Gnanasekar. "GCCover Heterogeneous Wireless Ad hoc Networks." Journal of Chemical and Pharmaceutical Sciences (2015): 195-200.

Rajesh, M., and J. M. Gnanasekar. "CONGESTION CONTROL USING AODV PROTOCOL SCHEME FOR WIRELESS AD-HOC NETWORK." Advances in Computer Science and Engineering 16.1/2 (2016): 19.

RAJESH, M. "TRADITIONAL COURSES INTO ONLINE MOVING STRATEGY." The Online Journal of Distance Education and e-Learning 4.4 (2016).

Rajesh, M. "Object-Oriented Programming and Parallelism."