ISSN No. 0976-5697

Volume 4, No. 11, Nov-Dec 2013

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

A Fuzzy Based Approach for Student Admission Recommendation into Seminary

S. Kalyani[1], R V S Ratna Kumar[2] and S. Ram Prasad Reddy[3]
Asst. Professor, IT Dept.,
Vignan's Institute of Engineering for women,
Visakhapatnam, India
vishal_mohi@yahoo.co.in[1], mailmeto.satya@gmail.com[2], reddysadi@gmail.com[3]

Abstract Education is a complex systematic engineering, which is the guarantee of training high-quality talent, helping society make full use of educational outcomes and promote the healthy development of education. To step-in to that high quality education students have a chaos in evaluating the best among the several institutions and select the one. In this paper, we suggest an automatic student admission recommendation system that selects a set of institutions by considering a set of student prerequisites and semantically match them with their parameters. Fuzzy clustering technique is applied on categorized data for suggesting better suited colleges for a particular student based on his/her course option. Since the same student can opt for more than one college for a particular course, depending upon multiple parameters, fuzzy clustering acts as the best suited method for seminary recommendation. The relative fuzzy score called "degree of membership" calculated for each college indicates the membership of a particular student to different institutions. Subjective evaluation of the algorithm is tested on synthetic dataset and the experiments produce promising results.

Keywords: Synthetic dataset, Fuzzy sets, Clustering, Text Categorization, Smart selection

I. INTRODUCTION

The use of the Internet now has a specific purpose: to find information. Unfortunately, the amount of data available on the Internet is growing exponentially, creating what can be considered a nearly infinite and ever-evolving network with no discernable structure. This rapid growth has raised the question of how to find the most relevant information. Many different techniques have been introduced to address the information overload, including search engines, semantic web, and recommender systems, among others.

The internet has become the most liked media for collecting information, and communication for different age groups. As there are a wide number of colleges, it has become difficult for the students and their parents to grade a college and choose the best suitable college basing on their rank, reservation, locality and students' interest. With the emergence of web applications it has become easy for the users to use any kind of service at their door step. The college recommendation for students in web counseling contains the information regarding the students rank, reservation category, gender and locality collected from their registration process.

They are also provided with the exercising options to select Colleges and courses they wish to join and arrange them in the order of priority. The list of colleges along with their codes, courses offered and course codes are congregated. The apportionment will be processed based on merit and category and will be placed in the web by proposing the students to choose the best seminary among the various academies displayed.

Recommender systems are computer-based techniques that are used to reduce information overload and recommend colleges likely to interest a user when given some information about the student's profile. This technique is mainly used to suggest seminaries that fit a student's desired tendencies.

The use of recommender systems for seminary selection is intended to remove the chaos among public, students and parents by providing the needed information .A few algorithms have been proposed in recent years on clustering. They belong to different types like Zengyou's [1] method called squeezer picks the data and hierarchically clusters it with any one of the existing ones.

In this paper, architecture of a recommender system that uses fuzzy clustering methods [2, 3] for seminary selection is introduced. In addition, a comparison with the smart-selection system, a Web-based seminary Assistance Application used to aid students in finding the institution that is most in line with their preferences, is presented.

A recommender system for seminary selection specifies two basic entities, which include the user (i.e., student) and the seminary (i.e., Institution). The main goal of this type of recommender system that is used in seminary selection is to basically aid students in finding the institution that is most in line with their preferences. The main problem of recommender system includes the following:

a. Quality of Recommendations: The information received from a recommender system must be reliable; for that reason, recommender systems should minimize the number of false positive results

b. Sparsity: A recommendation system is related to the number of recommendations made for the students. The sparsity problem of recommender systems emerges when the number of rated colleges is small compared to the total number of colleges, which leads to weak recommendations since the recommender systems are based on similarities between individuals.

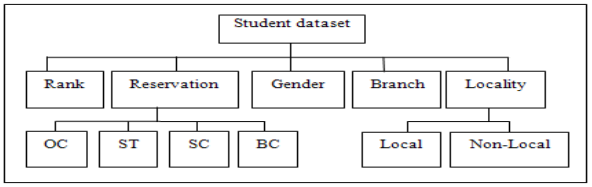

Figure 1. Clustering of Colleges

II. CLUSTERING

Clustering is a technique which groups data points such that points within a single group/cluster have similar characteristics (or are close to each other) while points in different groups are dissimilar. Most previous works on clustering focus on numerical data whose attributes are represented by numerical values and exploit the typical dissimilarity measures like Euclidean distance measure. A robust clustering algorithm (ROCK) [4] works based on a link similarity measure for merging similar clusters. Huang [5] proposed a k-modes categorical data clustering by extending the k-means algorithm.

Hence, our main objective is to "cluster the un-ordered categorical student data into predefined categories, while each student can belong to more than one college cluster with different features". So, we have used generalization hard partitioning method called fuzzy k-modes clustering algorithm. The fuzzy score called "degree of membership" calculated for each student indicates the likeliness of a particular student to different college clusters.

A. Clustering Of Colleges:

The pictorial representation of automatic college recommendation framework is given in Fig. 1. Our objective is to suitably match students to colleges, taking into account of various parameters/features like student rank, reservation category, gender, locality and branch interested in.

Fuzzy K-Modes clustering is applied on the pre-processed college data and clustered into three main college clusters – Category-A, Category-B and Category-C. The fuzzy score associated with each student gives the membership of the particular Category of the colleges recommended to the students. The various steps are explained in detail in the following sub sections.

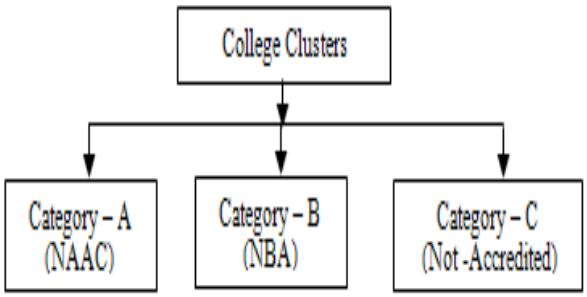

Figure 2. Example- Student Dataset Features

B. Student Annotation:

The features of the student data in the database are represented using textual keywords called annotations. These annotations carry comments, explanations or external remarks describing the information of an entity (student). A relevant set of annotation keywords, which sufficiently describe the student data features like rank, reservation category, gender, branch and locality detail are selected with the help of experts. All the keywords collected under these features act as a training data and they are categorized semantically using a categorization technique.

C. Experimental Database:

The details of the students data being collected is stored in the database and mapped with the different colleges applicable to them basing on their rank, gender, caste, branch and region using two different approaches of smarts-election and fuzzy k-modes clustering

Table: 1 xample Student Dataset

Rank	*Gender*	*Caste*	*Branch*	*Region*
500	Female	OC	CSE	Local
20000	Male	BC-A	EEE	Non-Local
3400	Female	ST	ECE	Local
50312	Male	SC	ECE	Non-Local
4213	Male	OC	CIVIL	Local
871	Female	BC-D	Mech	Local
1200	Female	BC-B	EEE	Non-Local
65700	Male	ST	ECE	Local
32678	Male	OC	IT	Local

As more number of colleges only increases the database size and have nothing to do with increasing the accuracy of the student selection, only data of certain colleges are collected and annotated by counselors. All the colleges and their corresponding annotations are stored in the database. So, we have classified all the available College data set into only three different types of categories: Category-A, Category-B and Category-C.

Table: 2 Example College Dataset

College Name	*Code*	*Accreditation*	*Rank*
Vignan Institute of engineering for women	NM	Not Accredited	7
Vignan Institute of Information Technology	L3	NAAC	2
Viswanatha Institute of Technology & Management	56	NBA	4

III. ARCHITECURE

The recommendation process is divided into three steps. In the first step the students must create their profiles by using a fuzzy interface described in greater detail in the following section. In the second step, once all necessary profiles have been created the user selects the recommendation target and the type of output. In the final step, once the recommendation engine has computed all information, the user receives a recommendation in a degree of member ship format.

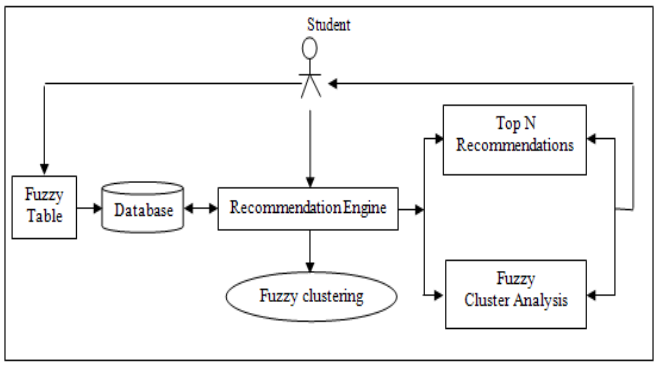

Figure 3. Fuzzy based Recommender System Architecture

Here we use two approaches i.e., smart-selection and fuzzy k-modes clustering methods for the computation of the recommendation process, compare the results and show

that fuzzy clustering method is better than smart-selection. The methods are described in detailed in the following sections

IV. SMARTSELECTION: A SELECTION ADVICE APPLICATION

A. About smart-selection clustering:

The smart-selection Web site is made accessible to the students and leads them through three steps in order to arrive at their individual selection. First, students must specify their profile. They are asked to answer the questionnaire about their details and interests, but they can choose between a "deluxe version" that consists of all questions and a "rapid version" that only includes certain questions.

The students have also a "no answer" option if they wish to leave out a number of questions, and they can weigh the answers according to the level of importance that the issues hold for them. The Web site provides students with additional background information, including pros and cons for each question. Second, students have to select the locality for which they want to receive a seminary recommendation.

B. Smart-selection Match Points Computation:

To generate its recommendations, the smart-selection system uses a statistical method to compute the "match points" by using (1).

$$MP_i(v,c,w) = 100 - |a_{iv} - a_{ic}| + b \qquad (1)$$

Where $MP_i(v,c)$ represents the number of points of agreement (i.e., match points), a_{ic} and a_{iv} represent the numerical answers given by student v and college c to questions i and b. The next step in the matching calculations is to take into account the relevance that each student gives to each question.

All the questions also have a weight, which consist of "+", "=", and "-." Depending on the weight assigned by the student, the corresponding match points are multiplied by the factors 2, 1, or 0.5 (weighting value "+" corresponds to factor value 2, weighting value "=" corresponds to factor value 1,and weighting value "-" corresponds to factor value 0.5), as shown in (2).

$$MP_i(v,c) = (100 - |a_{iv} - a_{ic}| + b\ ) \times w_i \qquad (2)$$

Where w_i is the weighting value that counsel admin c gives to a given questions i. A matching value, which is the percentage between student v and college c, is calculated using (3) and (4).

$$MP(v)_{max} = \sum_{i=1}^{n} a_{iv} \times w_i \qquad (3)$$

Where is the theoretical maximum possible match score, which depends only on the answers and weights of student v.

$$MP(v,c,w) = \frac{MP(v,c,w)}{MP(v)_{max}} \times 100 \qquad (4)$$

Where represents the matching value as the percentage between student v and college c. smart-selection can also be used in order to generate recommendations by full lists; in this case, the matching values are computed by using the mean average of all colleges on the list.

C. Symmetry Problem:

There is a drawback in using smart-selection matching point computation. Known as "the symmetry problem," this challenge can be illustrated with the following example:

Two individuals, p1 and p2, both answer the smart-selection questionnaire as both student's v1 and v2 and colleges c1 and c2, respectively. The responses to all of their questions as college and student are the same for p1 and p2. Assume that the answer to a specific question of p1 is "Yes" (score = 100), and p2 is "Probably Yes" (score = 75). The relation between pair's v1 – c2 and v2 – c1 are

$$\mathrm{MP}(v_1, c_2) = \frac{100 - 1100 - 751}{150} = 0.5$$

$$\mathrm{MP}(v_2, c_1) = \frac{100 - 1100 - 751}{100} = 0.75$$

V. FUZZY K-MODES CLUSTERING

The fuzzy clustering algorithms are well known for clustering the patterns to all the clusters with different degrees of membership. The k-means based fuzzy clustering technique is proven as one of the best methods of clustering. As the college data set is un-ordered, the Fuzzy K-Modes clustering [6] is the best suited method to recommend students to select one of the various college clusters with a different membership.

Let X = {X_1, X_2, .., X_n, .., X_N} be a set of N categorized data stored in the database. Each data point X_n, for all n is defined by the set of features {A_1, A_2,...A_m,...A_M} where each A_m has set of root attributes {a_{m1}, a_{m2}, . . .a_{mk}. . .,a_{mK}} and their corresponding child attributes. Let X_n be denoted as [$\mu^{(1)}x_{n1}$,$\mu^{(2)}x_{n2}$, . . ,$\mu^{(m)}x_{nm}$, . . ,$\mu^{(M)}x_{nM}$], where $\mu^{(m)}$, for all m are the weights given to the student data features with the constraint These weights are decided according to the degree of importance of the features of the data. The objective of the fuzzy k-modes is to minimize the loss,

$$J_f = \sum_{n=1}^{N} \sum_{k=1}^{K} \omega_{nk}^{\gamma} d(V_k, X_n)\ 0 \le \omega \le 1, 1 \le k \le K, 1 \le n \le N \qquad (5)$$

Subject to the constraints,

$$\sum_{k=1}^{K} \omega_{nk} = 1, \forall n\ , \quad 0 < \sum_{k=1}^{N} \omega_{nk} < N, \forall k \qquad (6)$$

In order to minimize the loss, we use the k-modes algorithm with a simple distance measure between the centers and datum, and updating the data cluster centers. The hamming dissimilarity measure d(V_k, X_n) between the centroid V_k and a data X_n is defined as,

$$d(V_k, X_n) = \sum_{m=1}^{M} \varphi(V_{km}, X_{nm}) \qquad (7)$$

Where

$$\varphi(V_{km}, X_{nm}) = \begin{cases} 0 & V_{km} = X_{nm} \\ 1 & V_{km} \neq X_{nm} \end{cases}$$

Algorithm

a. Let iteration index $\tau = 1$ and $V^{(\tau)} = \{V_1, V_2, .., V_k, ..V_K\}$ be a set of clusters. Select K initial modes, one for each cluster.

b. Calculate the membership values $W^{(\tau)} = \{\omega_{kn}\}$ for all $1 \le k \le K, 1 \le n \le N$, N such that J_f is minimized, calculated membership value $W^{(\tau)}$. ω_{kn} is the fuzzy membership value of n^{th} student belonging to k^{th} cluster.

c. Set iteration index $\tau: \tau + 1$. Update all K modes of the clusters using the membership value of each data point. i.e., $V^{\tau} = V^{\tau+1}$. Update $W^{(\tau)}$ using$V^{(\gamma+1)}$.

d. Continue iterating the above step until the difference between $W^{(\tau)}$and $W^{(\tau+1)}$ is less than the threshold $\in$.

The modes are updated using the procedure as in . The broad idea of updating the modes is as follows. The cluster center V_k is defined as $[v_{k1}, v_{k2}, \ldots, v_{km}, \ldots, v_{kM}]$, $\forall m$. The cost function Jf is minimized if and only if$v_{km} \in \{A_m\}$. i.e.,

$$\sum_{n, x_{nm} = a_{mr}} \omega_{kn}^{\gamma} \ge \sum_{n, x_{nm} = a_{mt}} \omega_{kn}^{\gamma},$$

$$1 \le n \le N, 1 \le m \le M \qquad (8)$$

Refer [3] for the proof of fuzzy k-modes update method. Refer Table 1 for some more examples. In the table, the student features are shown in the order of, Rank, Reservation, Gender, Branch, and so on.

VI. EXPERIMENTAL RESULTS

The Seminary selection framework involves different experimental phases like student annotation, smart-selection, fuzzy k-modes clustering for selecting a set of colleges for different types of students. The algorithm clustered a high number of students into its correctly relevant seminary and only few students into wrong seminary cluster.

Table: 3 Performance of Fuzzy k-modes clustering for College Selection

Clustering Results			
	Degree of Membership[0,1]		
Keywords [rank/gender/reserve/ Branch/locality]	***Category A***	***Category B***	***Category-C***
500/Female/OC/CSE/Local	0.55	0.25	0.20
20000/Male/BC-A/EEE/Non-Local	0.20	0.65	0.15
3400/Female/ST/ECE/Local	0.35	0.25	0.40

Table: 4 Comparison Results of Clustering

Effect of categorization on clustering						
	Without clustering			***With clustering***		
Keywords [rank/gender /reserve/ Branch/locality]	****Cat -A***	***Cat -B***	***Cat -C***	***Cat -A***	***Cat -B***	***Cat -C***
500/Female/OC /CSE/Local	0.55	0.25	0.20	0.85	0.15	0.00
3400/Female/S T/ECE/Local	0.35	0.25	0.40	0.25	0.05	0.70
Note:- *Cat - Category						

The fuzzy k-modes clustering of categorized student data are compared with the clustering results without categorization of data. It was found that the clustering accuracy was improved in case of clustering the categorized data. The overall system accuracy is directly dependent on the robustness of the categorization and clustering. The results are shown in graphically and proved that the clustering method performs better than other techniques

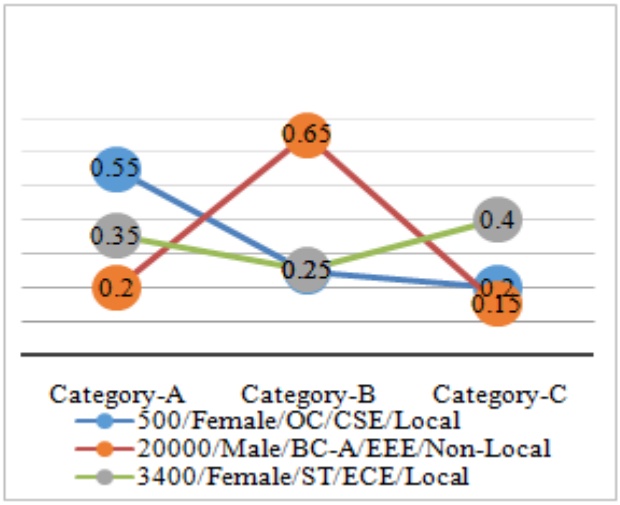

Figure 4. Graphical representation of the performance of Recommendation Engine

VII. CONCLUSIONS

We have proposed an automatic seminary recommendation framework which efficiently clusters the colleges into three different Categories – Category A, Category B, and Category C. It takes into consideration the various parameters of the student data for matching them with the college context. Our procedure involves adaptive categorization, fuzzy clustering technique for grouping the relevant colleges resulted in good performance.

VIII. REFERENCES

[1] He.Z, Xu.X and Deng.S, "Squeezer: An Efficient Clustering Algorithm for Categorical Data", Jl.of Computer Science and Technology, 2002.J.

[2] Sudha Velusamy, Lakshmi Gopal, Sridhar.V and Shalabh Bhatnagar, "Fuzzy Clustering Based Ad Recommendation for TV Programs".

[3] Stefani Gerber , Andreas Lander, Luis Terá ,"Using a Fuzzy-Based Cluster Algorithm for Recommending Candidates in eElections".

[4] Guha.S et al, "ROCK: A Robust Clustering Algorithm for Categorical Attributes", Proc.of Internation Conf.on Data Engineering,

[5] Huang.Z, "Extensions to the K-Means Algorithm for Clustering Large Datasets with Categorical Values", Data Mining Knowledge Discovery,

[6] Frank.H, Frank.K, Rudolf.K and Thomas.R, "Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition", Wiley Publications, 1999.