Volume 5, Issue 7, July 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Dream Graduation School Recommendation Using Cocktail Approach

Pruthvi U. V. S. S*
Computer Science,
GITAM University, India

Raga Ravali Pothireddy
Computer Science,
S.R.M University, India

Abstract*— Present years have witnessed students increased interest in doing higher studies in their dream universities. Despite many consultancies in this field, students find difficulty in getting into dream school depending on their capabilities like GRE and Toefl, financial situations etc. The judgemental challenge in this field is to consider every factor associated with the student and address the unique characteristics of each university. We first analyse the universities and develop a student-field-location-semester rank (SFLS) model. Then, on this, cocktail approach is applied to this model to get individualized. Further, we extend the recommendation to student-relation-field-location-semester rank (SRFLS) model for inert relationship among the students having the similar profile. Indeed, in this paper the moderate individualized recommendation of universities for each student and also the safe and ambitious list.

Keywords*— recommender systems, cocktail, safe, ambitious, moderate, rank model.

I. INTRODUCTION

According to recent reports, there has been a large stream of 75% in the number of bachelor students taking up GRE in 2014. Another report from an organization that conducts the GRE and TOEFL, the Educational Testing Service reveals that this increase is expected in the following years too. Getting through these tests with good grades eventually opens the gate for the aspiring candidates to apply to masters in dream schools. Most of the students appearing for these tests are at graduate level intending to join an MBA or a STEM (Science, Technology, Engineering and Mathematics) related program in the best universities. Globalization has made the world so small that with the emergence of different ways of support, more and more students are now fulfilling their dreams. The reason for desiring about Graduation is simple. Due to the high credentials and prominence associated with professional studies, a large number of students try for it every year. Also, the availability of proficient faculties, well equipped labs, interactive teaching methodologies, scholarships and good career prospects in best universities and also the availability of educational loans.

A. Cocktail Approach

To expeditiously mine sporadic associations, it is essential to develop a new data mining framework[10].The main concern about this paper is after taking all these tests, how can a student select a university from best universities of this world that is suitable for him. Universities consider the student's grades, research work, intelligence, scores of these tests, ability to get with universities works, etc. whereas students check for ranking of the university, scholarships, coursework, etc. Considering all these factors they have to select the best university. This is called as cocktail approach. It is a general approach followed in any selection. It is the method examining every detail of the problem and formulating a particular ratio according to their credit in the problem. This ratio is used to derive results.

For this, an intelligent recommender system is needed. Since recommender systems have been successfully applied to enhance the quality of service in a number of fields [3], [5], it can be feasible to many students to find the universities.

Despite all this, there are many intrinsic practical and domain issues in developing and implementing an effective recommendation system for individualised dream university recommendation system. There are many reasons to hinder this. Firstly, all the universities do not consider the similar criteria for selecting the students. Secondly, the students interested may not like the recommended university due to several factors like cost, area, etc. Thirdly, the reviews about the universities differ from their actual rankings and recommendations cannot completely rely on their reviews. Also, the rankings of universities depend on the semester and the course.

To solve these issues, in our preliminary work [3], cocktail approach was used. Particularly, analysing the important considerations of each university and divide their reachability based on their location, field of interest and their semester intake of students. Based on this, we develop a student-field-location-semester (SFLS) ranking model. In this model, the rank extraction is based on both the students and also the universities. Student factors like university location i.e. in which country it is located and to which field the student is applying i.e., to which subject he is interested in that university and also to which semester he is applying and some of the minor considerations of universities like scores of student and his interest. Upon, SFLS rank model, the cocktail approach can constitute all the interest of students, as well as universities. If this data is static then, the problem is solved now, but the cold start problem starts with the day to day

rankings of the universities and their particulars. In this paper, not only with this individualised recommendation, we group the students according to their profiles and provide relationship among them. This makes splitting the universities accordingly and suggest the students, not only their profiled universities, but also safe and ambitious universities.

Safe universities of a student are those which are below his capability, so that he can get admissions through them very easily. Their ranking is very less. Students may be provided with better scholarships and other alliances by the university.

Ambitious universities of a student are those which are much above his capability, so that he can just try for university and if university likes his profile he may get through that. But, student needs to spend much than what he expected it would be.

B. Review Extraction

We can extract the details and log of each university admits and rejects of every year through net. From these the required information is collected in logs using data mining techniques.

II. SFLS MODEL

In this section, the representation of universities and the students is shown by a rank model using Bayesian networks and the similarity between them is formulated like methods in [1] and [4] .The lists of mathematical notations used in this paper are shown in Table 1.

TABLE 1 MATHEMATICAL NOTATIONS

NOTATION	DESCRIPTION
$U = \{U_1,U_2,\ldots,U_N\}$	The set of universities. N
$B = \{B_1,B_2,\ldots B_A\}$	The set of bachelor students list. A
$S = \{spring, fall\}$	The set of semesters.2
$M = \{M_1,M_2,\ldots,M_Y\}$	The set of scores. Y
$L = \{L_1,L_2\ldots,L_X\}$	The set of locations. X
$Q = \{Q_1,Q_2,\ldots,Q_Z\}$	The set of QS world rankings of the universities. Z
$F = \{F_1,F_2,\ldots,F_S\}$	The set of student undergraduate fields. R
$C = \{C_1,C_2,\ldots,C_T\}$	The set of cost packages of each university. T

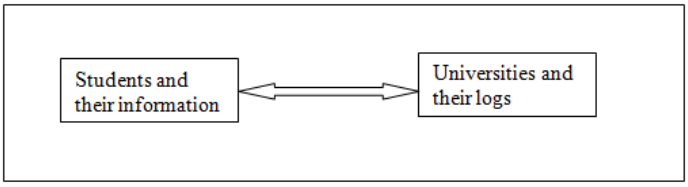

Fig. 1 SFLS model

To find the M set of Table 1 , we need to consider many details like scores of the student in GRE , TOEFL (say), grades and his previous research work .As the universities do not consider only the scores or percentages, we need to calculate all the minor details also in the particular ratio in which universities consider. In this equation, M represents marks in the tests and tests include GRE and Toefl. The second factor is his under graduation performance using grades, which include the field in bachelors and his final percentage. The final factor is Level of research, which include his previous projects and the journals he published.

Scores α [M(tests) * P(grades) * L(research)]

Scores α [M_i(GRE) . M_i(Toefl)] * [P_i(field). P_i(percentage)] * [L_i(projects).L_i(journals)]

Scores = R_a [M_i(GRE) . M_i(Toefl)] * R_b[P_i(field). P_i(percentage)] * R_c[L_i(projects). L_i(journals)]

Generally in a cockatail approach, according to definition 1, each details need to be considered in the particular ratio. R_a,R_b,R_c represents the ratio in which these details need to be considered. These ratios depend on the universities.

Next, is the location in which the university is located. There are several countries where the students wish to go. In this paper, present a countries only where many universities are found and these are formulated in the table. These are required to get the cocktail ratio of the scores of universities depending on location. Table 2 represents the area segmentation taken from [6].

TABLE II AREA AND SEGMENTATION TABLE

Area	Major segmentation
United States	Texas, Chicago, Newyork, Pennsylvania state
Canada	British Columbia, Toronto, Montréal, Vancouver and Quebec

Africa	South Africa, Kenya, Ghana, Tanzania, Uganda, Morocco, Egypt
Europe	Spain, Germany, England, France, Italy, Spain, Czech Republic, Italy.
Asia	India, China, Pakisthan, Nepal, Bangladesh, Indonesia, Iraq, Philippines, Sri Lanka ,Vietnam
Australia	Sydney, Melbourne, Brisbane, Wollongong
Mexico	Monterrey, Mexico city, Brazil
Russia	Moscow, Tomsk, Saint Petersburg

The next factor is the world ranking of the universities. In this paper, we consider the standard ratings given by QS rankings. The final issue is the cost factor and the scholarships to be provided. Students will consider the total tuition fee, accommodation, scholarships provided by the universities and the cost of living in that place. Here, there are several details to consider. If the university is the best and the salaries offered after graduation is higher, then students do not think about the cost much and if the student's financial status is low and if banks provide scholarships then this issue need not be considered. Hence, taking general strategy of students into consideration that is how much can be paid back in two years to the total expenditure to study as a cost package for a university. It is a general consideration; it varies according to student performance. In this paper, for general idea of expenditure this is to be considered.

Initially, the process of generating the university details in a rank model style, i.e., considering all the previous log details and reviews about the university. Then, the locations of the universities are drawn from location set. To choose the university, first we choose the university from the given ranks, i.e., external source called observed variable and rank extracted from rank modelling using the logs of the university is called as latent rank. In this model, there are two steps in which first step corresponds to inverse method in which, we get the university rank from the external world and next step is the generative process in which the rank is extracted using every detail and students log of that university like other details like cost, scholarships, location, etc. besides rank. This rank is the latent rank.

The SFLS rank model follows the similar Dirichlet distribution assumptions as [1], [7] and here universities are the "tokens" for rank modelling. Here, d, stand for ID of student and n for ID of university. L is the locations set and l is the particular location derived. Each recommendation can be of the vector form <University, Semester, Cost factor, Field > whereas, cost factor can be well described in vector form as <QS rank, cost, score>. This can be given in abbreviated form as, <U, S, C, F> and <Q, C, S> respectively. As, this is the student recommendation it does not mean that universities requirement is not considered. University, u is chosen from one universities list , U of that location set L and r is the latent rank for both student and the university whereas, z is the actual QS ranking of university.(L,Q,C) is the specific location-QS rank –cost triplet and (M,F,S) is the score-field-semester triplet. Φ, θ are say student consideration of university ranking and the university consideration of its ranking respectively and α, β are their corresponding hyper parameters for cocktail ratio, say.

$$\prod_{n=1}^{X}\pi \prod_{s=1}^{Z}\pi \prod_{c=1}^{C}\pi\, p\left(\frac{\emptyset_{nsc}}{\alpha}\right)$$
$$\prod_{y=1}^{Y}\pi \prod_{r=1}^{R}\pi \prod_{s=1}^{Z}\pi\, p\left(\frac{\theta_{yrs}}{\beta}\right)$$
$$\iint \prod_{d=1}^{D}\pi \prod_{i=1}^{N}\pi \prod_{l=1}^{A}\pi\, p\left(\frac{r_A}{\emptyset_{<L,Q,C>}}\right)$$
$$\sum_{r_u=1}^{N} p\left(\frac{r_N}{\emptyset_{<M,F,S>}}\right) d\theta\, d\emptyset$$

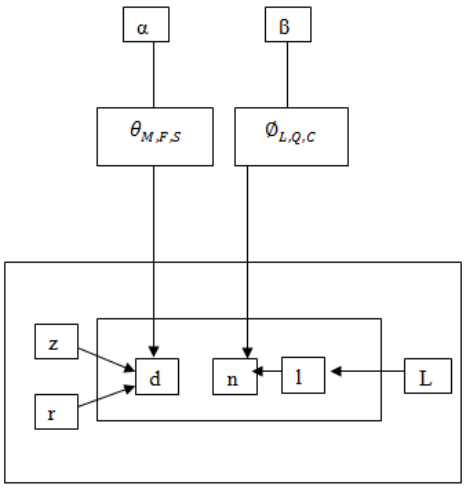

Fig. 2. SFLS graphical model

III. MODEL INFERENCE

In this paper, Gibbs sampling method [8],an easy implemented way to derive topics from the large sets of universities previous admit logs is used. When sampling, the generation of each university token for a given admits log depends on the corresponding student consideration QS rank-cost-score triplet and also university consideration of the specific location-QS rank –cost triplet. Finally, Φ and θ can be calculated by,

$$\emptyset_{nsc} = \frac{\beta + b_{slqc}}{\sum_{p=1}^{r}(\beta_p + b_{plqc})}$$

$$\theta_{yrs} = \frac{\propto + a_{rmfs}}{\sum_{q=1}^{z}(\propto_q + a_{qmfs})}$$

Here, a_{rmfs} is the number of logs of the universities for a particular profile based on their QS rankings and $b_{slqc}$is the logs of students research for a particular university of a specific location.

In this, for each iteration, θ derives the students profile requirement for the university along with near profiles and Φ derives the university requirements necessary for the students i.e., based on their factors as discussed above.

fig.3 gives the complete structure of how the SFLS model works using the previous data from the logs and the students details using cocktail approach.

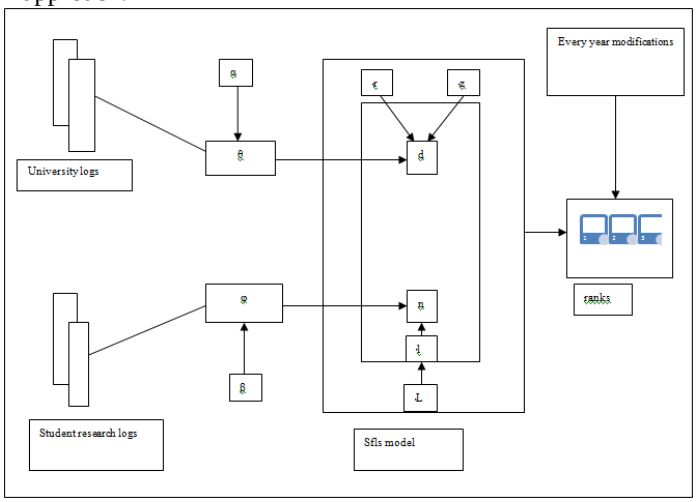

Fig 3. Complete Understanding Model of the SFLS model using cocktail approach

IV. CONCLUSION

Significant progress in recommender systems is being taken since last two decades. These play major role in all fields. In this paper mainly university ranking, cost factors and student profile is considered. There are many advances and enhancements to be taken. Here almost every detail for the university selection is considered and given in proportion according to their importance.

There are 2 major enhancements can be made. First, the proportion in which all factors are considered may be main point in the future work. Second, It is not compulsory that every year the same rankings are given to universities or universities do not select other students with different profiles. So, the variables are changed and to be taken with new values.

Furthermore, cocktail approach has certain disadvantages compared to recent intelligent systems. In intelligent systems [9], the preferences of each student are traced. Unlike that, here only the logs of universities and student research are considered. Although logs provide the basic pattern of thinking, this cannot work exactly for a single student who has his own interest in choosing universities. For such exceptions, intelligent systems can be used.

REFERENCES

[1] I.J. Michael ,D.M. Blei and Y.N. Andrew,"Latent Dirichlet Allocation," J. Machine Learning Research, vol. 3, pp. 993-1022, 2003.

[2] Y. Ge, Z. Li, H. Xiong Q. Liu, and E. Chen, "Personalized Travel Package Recommendation," Proc. IEEE 11th Int'l Conf. Data Mining (ICDM '11), pp. 407-416, 2011.

[3] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.

[4] M. Rosen-Zvi et al., "The Author-Topic Model for Authors and Documents," Proc. 20th Conf. Uncertainty in Artificial Intelligence (UAI '04), pp. 487-494, 2004.

[5] Y. Ge et al., "An Energy-Efficient Mobile Recommender System," Proc. 16th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '10), pp. 899-908, 2010.

[6] Qi Liu, Enhong, Hui Xiong, Yong Ge, Zhongmou Li, and Xiang Wu, "Cocktail approach for travel package recommendation", IEEE transactions on knowledge and data engineering , vol.26, NO. 2, February 2014.

[7] A. Mccallum, X. Wang, and A. Corrada-Emmanuel, "Topic and Role Discovery in Social Networks with Experiments on Enron and Academic Email," J. Artificial Intelligence Research, vol. 30, pp. 249-272, 2007.

[8] T.L. Griffiths and M. Steyvers, "Finding Scientific Topics," Proc. Nat'l Academy of Sciences USA, vol. 101, pp. 5228-5235, 2004.

[9] S. Staab et al., "Intelligent Systems for Tourism," IEEE Intelligent Systems, vol. 17, no. 6, pp. 53-66, Nov. 2002.

[10] Yanqing Ji, Hao Ying, Fellow, IEEE, John Tran, Peter Dews,Ayman Mansour, and R. Michael Massanari, "A Method for Mining Infrequent Causal Associations and Its Application in Finding Adverse Drug Reaction Signal Pairs," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 4, APRIL 2013

[11] Data Mining:Concepts and Techniques, Third Edition, Jiawei Han and Micheline Kamber, ISBN·l3,978·0123814791

