Course Recommender System in a Liberal Arts Context

anonymous for review

ABSTRACT

This paper describes a direct application of topic modelling and sequential rule mining to provide transparent course recommendations to students of the Liberal Arts and Sciences Bachelor from University College XXXX, based on their academic interests and performance in previous courses. The system is developed to complement academic advising and help students make well informed decisions. We find that course recommendations based on a topic modeling of course descriptions are useful and that sequence mining provides a rough method to control for prerequisites.

Keywords

Course recommendation, open curriculum, topic modeling, pre-requisite discovery.

1. INTRODUCTION

The Bachelor in Liberal Arts and Sciences offered at the University College XXXX, the Netherlands, is an honors program characterized by an open curriculum. The program offers over 150 courses ranging from artificial intelligence, to conflict resolution and to pop songs, and students can design their own curriculum in a fairly free fashion: 140 out of the 180 ECTS are free. Students can also enroll in courses offered at one of the other 12 departments of XXXX University. This allows the students to tailor their curriculum to their own interests, but also make the selection of courses overwhelming: given the large number of courses available, it is often difficult for students to make an informed decision. Firstly, students and their personal academic advisor only know the content of a limited number of courses. Secondly, since each student has a unique curriculum, it is difficult to determine whether a course's level is appropriate for a student or whether it is too advanced. Given the limited number of hours that a person can spend browsing course catalogues and reading course descriptions, it would be beneficial to develop a course recommender system for the students of the liberal arts program which took into account the student's interests and her/his performance in past courses in order to suggest potentially interesting courses and warn students in case they do not have the necessary prerequisites for a course.

To accomplish this, we have developed a system that offers courses recommendations based on a topic model of course content, and that gives red flags based on sequential association rules. We are currently working on an alternative approach to red flags that does not consider courses as black boxes (sequential association rules), but takes into account the skills developed by student during their curriculum. XXXX University and the University College XXXX place high value in the decision making process of their students, fostering them to take their education in their own hands. In this context, such course recommender system would enhance self academic advising. For this reason, the system that we have developed not only offers course recommendation and red flags, but also gives the reasons behind these. We see this transparency as crucial to allow students to make well informed decisions with regard to their course selections.

2. METHODOLOGY

2.1 Available Data

We have had access to two types of data: student data and course data. The student data consists of the transcripts of all previous and current students of the liberal arts program (n = 2,526). It contains the time at which a student took a course and the grade (s)he obtained. The course data consists of the five most recent course catalogues (2014-2019) which offer a one-page description of each course offered at the college, as well as the individual course manuals, which provide detailed information about the course's objective and content (usually 20 pages long), for the year 2017. We are currently incorporating the course catalogues from the other department of XXXX University to the analysis. In the future, we hope to have access to the literature (academic articles, book chapters and other material read by the students) of each course in order to have more detailed information about their content.

2.2 Approach

The course recommender system is composed of two pillars: course recommendations and red flags. This division is motivated by the two challenges that students face when choosing which courses to enroll in for the following semester: among the thousands of courses offered at the university, which ones match my academic interests and which ones have a level that suits my profile. We consider course recommendations in the traditional sense of selecting the top-n courses based on some predefined ranking metric. Red flags are issued when the system identifies course that are too advanced for the student given her/his performance in past courses. All computations are conducted on R[1].

2.2.1 Course Recommendation

To provide Course Recommendations, we identify the courses whose content best match the academic interests of the students. First, we identify the topics covered by a course with the help of a topic model fit on the course data with a Latent Dirichlet Allocation (LDA) algorithm. A topic model allows the probabilistic modeling of term frequency occurrences in documents by providing a probability distributions of topics per document (i.e. per course), and of words per topic. We fit topic models on the most recent course catalogue and on the course manuals. In the future, we hope to be able to fit a topic model on the academic literature of each course to obtain a more detailed representation of course content. We vary three parameters of the algorithm – number of topics, alpha and beta – and select the model with the best log-likelihood (Figure 1). At the time of this writing, we are still evaluating the optimal value of the latter two parameters; yet, the current topic model already gives very good

[1] https://github.com/XXXX

recommendations. In the future, we may also consider two alternative metrics: perplexity and coherence.

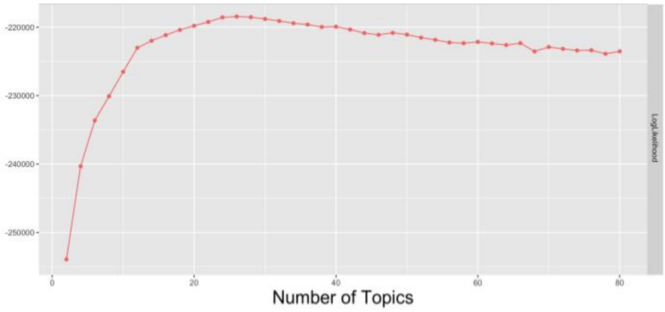

Figure 1. Log Likelihood of Topic Models on Course Catalogues for different numbers of topics.

Second, we map the key words introduced by the students into the system to the topics identified by the topic model. This way, we establish a topic profile of the student which consists of a numeric vector indicating the importance of each topic for the student. At the moment, student topic profiles are established only via the key words they enter in the system, but we are working on building a more complete profile by allow students to indicate courses and academic literature whose content particularly match their interests. Finally, we rank the courses based on their match with the topic profile of the student and recommend the top *n* courses.

To make these recommendations transparent, we include the key words that led to a course being recommended.

2.2.2 Pit Fall Avoidance - Red Flags

To provide Red Flags we mine the student data using the CSPADE algorithm. Given a set of sequences, the CSPADE algorithm identifies all subsequences with a support superior to a given threshold. Since we are working with a relatively small number of observations (around 2,526 student, each taking around 40 courses), we restrict the search to two item sequences, i.e. rules with a single antecedent and a single consequent. Each item is a dyad composed of a course and grade (<course, grade>). We also identify sequences whose antecedent is a course that has not been taken (see Figure 3, second red flag). The obtained subsequences can be conceptualized as association rules whose left-hand side is anterior to the right-hand side. We only consider sequences whose right-hand side corresponds to a low performance i.e. a grade of 7 or lower (using a stricter threshold i.e. a failing grade of 5.4 produces rules with extremely low support) and have a count superior to 20 students, a confidence superior to 0.4 and lift superior to 1.1.

To provide red flags, we ask the student to enter the courses s(he) intends to take the following semester. We then, select all rules that have one of the prospective courses as a consequent and check whether the student's transcript contains any of the <course, grade> dyads that appear as antecedents in the selected rules. If this is the case, we issue a red flag for the course that matches the consequent; otherwise, no warning is given. In order to make the red flags transparent, we provide a numerical summary of the subsequence on which the red flag is based.

We are currently working on an alternative approach to red flags based on the competences acquired by students throughout their curriculum. We are also working on providing a list of courses that a student faced with a red flag could follow as a preparation to the course flagged by the system.

3. RESULTS

This is work in progress, yet the results are promising. Course recommendations were particularly satisfactory. The recommended courses follow logically from the selected key words; an opinion that was corroborated by faculty members, academic advisors and current students of the liberal arts program. An example of the outcome can be seen in Figure 2.

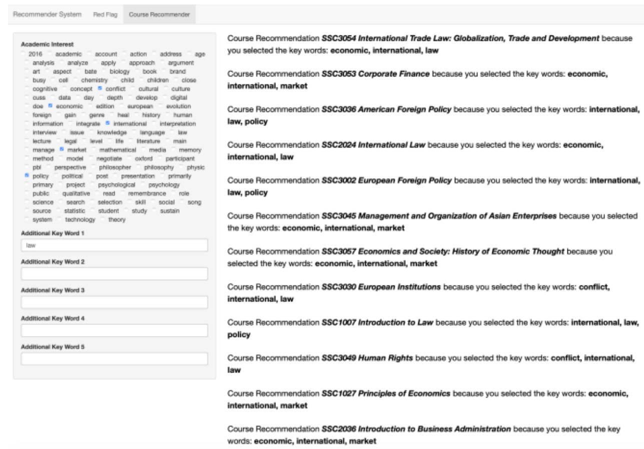

Figure 2.Example of Course Recommendations

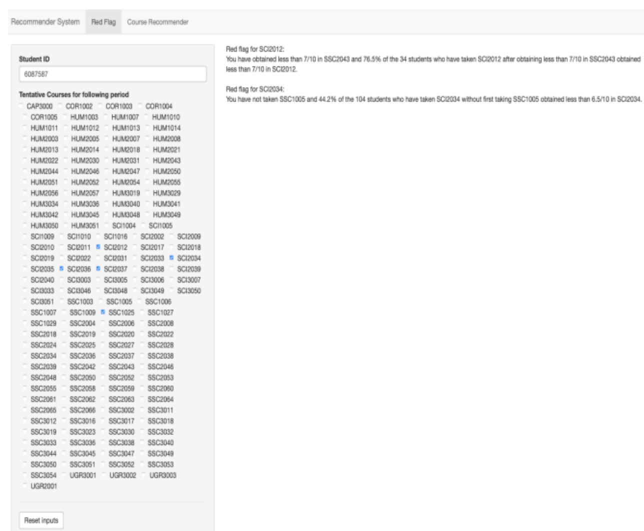

Figure 3. Example of Red Flags

Results for red flags were less satisfactory. Given the relatively small number of students, the large number of course and the flexibility of the program, sequences tends to have a low support and to be particularly sensitive to variance. Moreover, in its current form, the system does not permit to over-ride red flags; that is, even if there is ample evidence that a student is able to perform well in a course, if (s)he matches a single sequence, then a red flag is issued. Figure 3 provides an example of red flags.

4. FURTHER RESEARCH

In the light of these results, we would like to continue improving both pillars of the course recommender system. First, we would like to include course information from the other departments. This way, the system could recommend courses to the students which they had never considered because they ignored they existence and did not know that their content suit their academic interest. We also want to elaborate a more complete topic profile of student by taking into account courses or journal articles that the student found particularly interesting. Second, given the limited success of red flags based only on performance in past courses, we would like take into account the competences that the students acquire during their curriculum. One such approach is to consider the topics that the student has covered in past courses (obtained from a topic model).

5. ACKNOWLEDGMENTS

Our thanks to the University College XXXX, the Institute of Data Science, XXXX University, and the Department of Data Science and Knowledge Engineering, XXXX University, in particular to XXXX, and XXXX for supporting this project with their expertise.

6. REFERENCES

[1] Bettina Grün and Kurt Hornik. 2011. topicmodels: AnRPackage for Fitting Topic Models. *Journal of Statistical Software* 40, 13 (2011). DOI:http://dx.doi.org/10.18637/jss.v040.i13

[2] R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna,