Title	Module Advisor: Guiding Students with Recommendations
Authors(s)	Hagemann, Nina, O'Mahony, Michael P., Smyth, Barry
Publication date	2018-05-17
Publication information	Hagemann, Nina, Michael P. O’Mahony, and Barry Smyth. “Module Advisor: Guiding Students with Recommendations.” Springer, May 17, 2018. https://doi.org/10.1007/978-3-319-91464-0_34.
Conference details	The 14th International Conference (ITS 2018), Montreal, Canada, 11-15 June 2018
Series	Lecture Notes in Computer Science
Publisher	Springer
Item record/more information	http://hdl.handle.net/10197/10164
Publisher's statement	The final publication is available at Springer via http://dx.doi.org/10.1007/978-3-319-91464-0_34.
Publisher's version (DOI)	10.1007/978-3-319-91464-0_34

Module Advisor: Guiding Students with Recommendations

Nina Hagemann, Michael P. O'Mahony, and Barry Smyth

Insight Centre for Data Analytics, School of Computer Science
University College Dublin, Ireland
http://www.insight-centre.org

Abstract. Personalised recommendations feature prominently in many aspects of our lives, from the movies we watch, to the news we read, and even the people we date. However, one area that is still relatively underdeveloped is the educational sector where recommender systems have the potential to help students to make informed choices about their learning pathways. We aim to improve the way students discover elective modules by using a hybrid recommender system that is specifically designed to help students to better explore available options. By combining notions of content-based similarity and diversity, based on structural information about the space of modules, we can improve the discoverability of long-tail options that may uniquely suit students' preferences and aspirations.

Keywords: Recommender Systems; Content-based Filtering; Diversity; Collaborative Filtering; Module Recommendations; Elective Modules

1 Introduction

Today's students enjoy a wide variety of options regarding the availability of courses and modules, encouraging students to broaden their horizons, explore their interest and strengths, and develop new skills. One such opportunity offered in many universities is the possibility to freely choose elective modules from outside a student's main area of study. The taking of such elective modules is often a mandatory requirement of programmes of study, and can have a significant impact on students' academic experience and overall performance.

Unfortunately, in practice, student choices are often limited by discoverability challenges and overcrowded modules as students flock to popular options. As a result many students follow the crowd or their peers' recommendations when selecting electives. This trend was confirmed to exist in a preliminary exploratory data analysis of the Computer Science undergraduate students in our institution. An analysis of historical student data revealed an imbalance in elective module allocations. The percentage of students choosing modules outside of Computer Science decreased rapidly over time; instead, students selected from a limited set of popular modules. This lead to many unsuccessful allocations (given constraints on enrolment numbers), obliging students to settle for their second or third elective module choices. We hypothesise that one of the reasons for these trends is the low discoverability of elective modules, especially those outside of a student's core area. Therefore, our main objective is to support students in discovering elective modules outside of their main field of study.

The need for a recommender system for academic guidance has been established over ten years ago. Previous research has shown the possibilities and requirements for such systems [2, 4, 11]. More recently, the interest in recommender systems for the educational sector has grown and studies agree on the benefit of recommender systems for module exploration [1, 5]. However, the majority of this research focuses on grade prediction or the use of grades as an indirect way of measuring students' ratings of modules [3, 6].

Although we agree that success in a module is an important factor for students to choose their modules, in this work we focus on the content of a module and its relevance to students' interests. We focus on supporting students in finding modules that are related to, but outside, their main area of study. We developed a prototype application that includes a personalised recommender system which helps students to discover lesser known elective modules by introducing diversity into the recommendation process.

In this paper we briefly present the current prototype and the underlying recommender system techniques and discuss the results of a preliminary offline study.

2 User Interface Prototype

To help students discover suitable elective modules we developed a prototype web application as shown in Figure 1. The application includes a personalised recommender system where students can choose modules from their module history and receive elective module recommendations based on their choice. A slider allows students to control the degree of *discovery* in the recommendations that are made. Moving the slider introduces diversity into the recommender system algorithm and acts as a natural explanation for the recommended modules. Thus, students are facilitated to gradually explore modules outside of their field of study and to broaden their knowledge about available modules in different areas.

3 Module Recommendation Approaches

In this section, we describe the proposed hybrid approach to elective module recommendation used in the application. We further briefly describe a collaborative filtering approach that is used to evaluate our results. The following notation is introduced. Let S and M denote the set of students and modules in the system, respectively. Each student, $s_i \in S$, is profiled by a subset of the modules which they have previously taken. Let P_i denote the profile of student s_i, where $P_i = \{m_1, m_2, \ldots, m_l\}$ and $m_j \in M$ denotes a particular module. Based on the modules in the profile, candidate elective modules (i.e. all elective modules offered by the university) are ranked and a top-N list of recommendations is returned for student s_i.

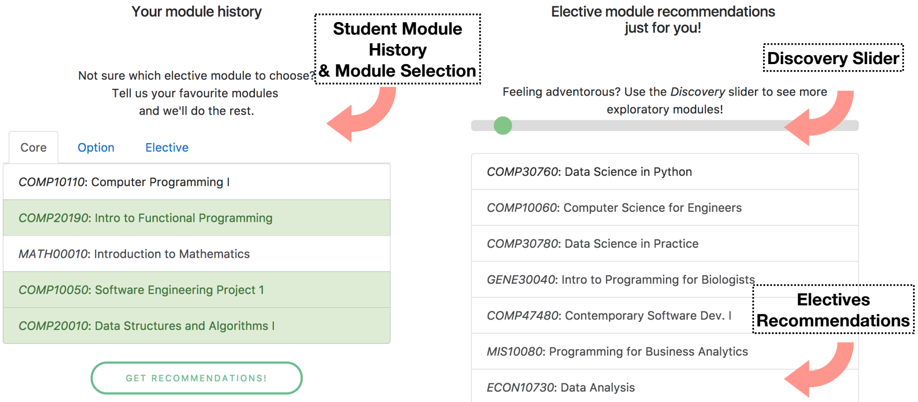

Fig. 1. Screenshot of the recommender system part of the prototype

The proposed hybrid recommender consists of two components to produce recommendations. The first component prioritises candidates that are similar in content to those in the student's profile; for this purpose, a traditional content-based (CB) recommender is used. The second component prioritises candidates from outside the student's programme area; in this case, a hierarchical taxonomy of the available programmes of study and associated modules is created, and candidates which are furthest from those in the student's profile are recommended.

Content-based Recommender. Each module has a descriptor which provides a textual description of its content, aims and learning outcomes. Thus, modules can be viewed as documents made up of the set of terms contained in their descriptors. Using the Vector Space Model (VSM) [9], each module is represented by a vector in an n-dimensional space, where each dimension corresponds to a term from the overall set of terms in the collection. Standard preprocessing of documents is performed, such as tokenisation, stop-words removal, and stemming [7]. Each module is represented as a vector of term weights, where each weight indicates the degree of association between the module and the corresponding term. For term weighting, we employ TF-IDF (Term Frequency-Inverse Document Frequency) [8], a commonly used scheme in information retrieval. The intuition behind TF-IDF is that a term which occurs frequently in a given document (TF), but rarely in the rest of the collection (IDF), is more likely to be representative of that document. Given the vector space representation the similarity between two modules is computed using cosine similarity [8].

The rank score of a candidate elective module, m_c, for student s_i is calculated as the mean cosine similarity between m_c and each of the modules in the student's profile, P_i, as follows: $\text{score}_{CB}(s_i, m_c) = \frac{1}{|P_i|} \sum_{m_j \in P_i} \text{sim}(m_c, m_j)$. Candidates are ranked in descending order of score.

Taxonomy-based Recommender. In order to recommend modules to students from outside their programme of study, an approach based on a hierarchical taxonomy of the academic structure of our university is used. Briefly, there are six Colleges, each with a number of constituent Schools. Each School offers a number of programmes of study, and each module is associated with one or more of these programmes.

While more sophisticated approaches are possible, here we make the general assumption that modules from the same programme are more closely related than those from different programmes. The following approach to used to calculate the rank score of a candidate elective module m_c for a given student s_i with profile P_i: $\text{score}_{TB}(s_i, m_c) = \frac{1}{|P_i|} \sum_{m_j \in P_i} \text{rel}(m_c, m_j)$, where $\text{rel}(m_c, m_j)$ is 0 if both modules belong to the same programme; 0.33 if the modules are from different programmes offered by the same School; 0.66 if the modules are offered by different Schools in the same College; and 1 if the modules are from programmes offered in different Colleges. Using this approach, higher scores are assigned to candidate modules which are further from those in the student's profile, thereby facilitating the student to to broaden their learning experience.

Hybrid Recommendation Ranking. The above provides two alternatives to elective module recommendation. The former prioritises candidates which are similar to a student's profile, while the latter prioritises candidates which are furthest from a student's core programme of study. These approaches can be combined to allow students to better explore the wide range of elective module choices available from across the university. An overall score for a candidate elective module m_c is calculated for student s_i as follows: $\text{score}(s_i, m_c) = \alpha \ \text{score}_{CB}(s_i, m_c) + (1 - \alpha) \ \text{score}_{TB}(s_i, m_c)$, where the parameter α can be varied to influence the diversity of elective modules recommended.

Collaborative Recommender. We also consider a neighbourhood-based *collaborative filtering* (CF) approach [12]. As before, each student s_i is profiled by a subset of previously taken modules, P_i. The neighbourhood for a given student s_i is determined based on profile similarity, where the similarity between two profiles, P_i and P_j, is calculated using the overlap coefficient [13]. Once the k most similar students to student s_i are identified, a top-N list of elective module recommendations, ranked by their frequency of occurrence in neighbour profiles, is then returned to the student. Using this approach, the elective modules which are popular among students with similar profiles are recommended.

4 Evaluation

We randomly selected 100 Computer Science students from the historical data set. Each student is represented by an average of 20 core modules, from which we randomly select three as the input to the recommender system, mimicking a student's input into the web application.

We conduct a leave-one-out test [10] and generate a top-10 recommendation set for each student for each recommendation approach: a pure content-based approach ($\alpha = 1$), three hybrid approaches ($\alpha = [0.25, 0.5, 0.75]$), and the collaborative filtering method (CF). To evaluate the offline results we are not using a classic accuracy score as we hypothesise that our ground truth, that is the set of elective modules actually taken by students, is skewed due to the reasons explained above (i.e. students largely following peer recommendations or simply choosing popular modules). One of our main objectives is to broaden the range of modules that students are aware of. Hence, we evaluate our results comparing the number of distinct modules recommended over all users, and the number of distinct subjects covered by these recommendations. To evaluate relevance, we use *sim-to-core* (StC), a metric that determines the average similarity of the most similar module in the student's profile, P_i, to each module in the recommendation set, R_i, as shown in Equation 1:

$$StC(P_i, R_i) = \frac{\sum_{m_k \in R_i} sim_{max}(m_k, P_i)}{|R_i|} \ , \tag{1}$$

where $R_i = \{m_1, ..., m_r\}$ is the set of elective module recommendations and $sim_{max}(m_k, P_i)$ returns the maximum similarity between the recommended elective module m_k and the modules in the students profile.

4.1 Results & Discussion

Firstly, we consider the overlap coefficient [13] of the recommendation sets produced by the various approaches (Table 1). As expected, as more diversity is introduced into the recommendation process (i.e. as α is decreased), a decrease in overlap between the recommended sets is observed. For example, an overlap of 76.5% in recommended sets is seen between the pure content-based recommender ($\alpha = 1$) and the hybrid approach with $\alpha = 0.5$. Comparing the recommendations made by the collaborative filtering approach, we see approximately only 3% of the same modules being recommended; since this approach operates over the limited set of largely popular modules actually selected by students, this result is also to be expected.

α	1	0.75	0.5	0.25	CF
1	1.000	0.901	0.765	0.626	0.031
0.75		1.000	0.862	0.724	0.032
0.5			1.000	0.860	0.033
0.25				1.000	0.034
CF					1.000

Table 1. Overlap of the recommended module sets by the different approaches.

α	D. Mod.	D. Sub.	% IPE	*StC*
1	149	31	24.1	0.012
0.75	156	34	21.1	0.011
0.5	157	37	17.9	0.009
0.25	154	44	12.3	0.008
CF	60	28	26.7	0.002

Table 2. Evaluation results for the different approaches.

Table 2 shows that there is a gradual increase in both the number of distinct modules (D. Mod.) recommended and the number of distinct subjects covered (D. Sub.) as diversity is introduced (i.e. as α decreases). Moreover, the percentage of in-programme (Computer Science) modules (% IPE) recommended also reduces, while the reduction in the *sim-to-core* (StC) metric is less pronounced. The results also show that the collaborative filtering approach produces recommendations with the lowest number of distinct modules and subjects covered, while the percentage of IPE modules recommended is the highest. Thus, it can be seen that the hybrid approach can successfully improve recommendation diversity, without significantly compromising relevance, while the collaborative filtering approach recommends from a relatively small set of modules.

5 Conclusion and Future work

The application of recommender systems seems opportune given the increasing tendency of our university's students to select from among a limited number of popular elective modules. We have shown that module descriptions can be used to make meaningful recommendations. While collaborative filtering approaches will give accurate results in a traditional sense, it will not help the problem of discoverability of modules as it promotes primarily already popular modules. We have shown that the proposed hybrid recommender system can add diversity to the set of recommendations. While the taxonomy-based recommender represents a first step, nonetheless it is capable of facilitating the discoverability of modules outside of the students' core area of study. In future work we plan on further developing our approach as well as conducting a live user study to understand how students will interact with the system and whether it leads to students choosing from among a more diverse range of elective module options.

References

1. Ajanovski, V.V.: Guided exploration of the domain space of study programs. In: 4 th Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS) 2017. p. 43 (2017)
2. Bendakir, N., Aïmeur, E.: Using association rules for course recommendation. In: Proceedings of the AAAI Workshop on Educational Data Mining. vol. 3 (2006)
3. Bydžovská, H.: Are collaborative filtering methods suitable for student performance prediction? In: Portuguese Conference on Artificial Intelligence. pp. 425–430. Springer (2015)
4. O'Mahony, M.P., Smyth, B.: A recommender system for on-line course enrolment: An initial study. In: Proceedings of the 2007 ACM Conference on Recommender Systems. pp. 133–136. RecSys '07, ACM, New York, NY, USA (2007)
5. Park, Y.: A recommender system for personalized exploration of majors, minors, and concentrations
6. Polyzou, A., Karypis, G.: Grade prediction with course and student specific models. In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. pp. 89–101. Springer (2016)
7. Porter, M.F.: Readings in information retrieval. chap. An Algorithm for Suffix Stripping, pp. 313–316. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1997)
8. Salton, G., McGill, J.: Introduction to Modern Information Retrieval. McGraw-Hill, Inc. (1986)
9. Salton, G., Wong, A., Yang, C.: A vector space model for automatic indexing. Communications of the ACM 18(11), 613–620 (1975)
10. Sammut, C., Webb, G.I. (eds.): Leave-One-Out Cross-Validation, pp. 600–601. Springer US, Boston, MA (2010)
11. Sandvig, J., Burke, R.: Aacorn: A cbr recommender for academic advising. Technical Report, TR05-015, DePaul University (2005)
12. Shardanand, U., Maes, P.: Social information filtering: Algorithms for automating "word of mouth". In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 210–217. CHI '95, ACM Press/Addison-Wesley Publishing Co., New York, NY, USA (1995)
13. Vijaymeena, M., Kavitha, K.: A survey on similarity measures in text mining. Machine Learning and Applications: An International Journal 3(2), 19–28 (2016)