Visualising Module Dependencies in Academic Recommendations

Nina Hagemann
Michael P. O'Mahony
Barry Smyth
firstname.lastname@insight-centre.org
Insight Centre for Data Analytics, University College Dublin
Dublin, Ireland

ABSTRACT

Starting their academic career can be overwhelming for many young people. Students are often presented with a variety of options within their programmes of study and making appropriate and informed decisions can be a challenge. Compared to many other areas in our every day life, recommender systems remain under used in the academic setting. In this part of our research we use non-negative matrix factorisation to identify dependencies between modules, visualise sequential recommendations, and bring structure and clarity into the academic module space.

CCS CONCEPTS

• **Information systems** → **Recommender systems**.

KEYWORDS

Recommender System, Non-Negative Matrix Factorisation, Academic Advising

ACM Reference Format:
Nina Hagemann, Michael P. O'Mahony, and Barry Smyth. 2019. Visualising Module Dependencies in Academic Recommendations. In *24th International Conference on Intelligent User Interfaces (IUI '19 Companion), March 17–20, 2019, Marina del Ray, CA, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3308557.3308701

1 INTRODUCTION

When starting university, students are faced with vast amounts of information and choosing the right modules that suit their strengths, interests and potential career goal can be a challenge. Recommender systems are omni-present in many aspects of our lives, helping us make informed decisions when choosing movies, books and even romantic connections. However, the academic area is still relatively underdeveloped when it comes to creating personalised recommendations. The educational sector has the potential to benefit from recommender systems technology and in recent years an increase in research interest in this area has been seen [1, 4, 5].

IUI '19 Companion, March 17–20, 2019, Marina del Ray, CA, USA

ACM ISBN 978-1-4503-6673-1/19/03…$15.00
https://doi.org/10.1145/3308557.3308701

The work presented here is part of an ongoing research project aiming to help students to make informed decisions in their academic career. This part of the research focuses on the often sequential nature of undergraduate degrees and the relations and dependencies between modules. Undergraduate programmes are often structured into levels (or years) with modules depending on preceding modules. While explicit module pre- and co-requisites are typically defined for university programmes, these formal requirements do not capture the informal dependencies that often exist between modules in a given programme.

Using non-negative matrix factorisation, dependencies between modules based on latent factors can be found. This information can be used in multiple use cases, from guiding students to select their most suitable personal curriculum, to helping students better understand the importance of modules in the wider context of their area of study.

2 APPROACH

In most universities undergraduate programmes are completed in three or four years, where modules are grouped by level or years. Modules are often presented to the student as a textual description in an online module catalogue.

Therefore, modules can be viewed as documents made up of the set of terms contained in their descriptors. We use the Vector Space Model (VSM) [7] to represent each module as a vector in an n-dimensional space, where each dimension corresponds to a term from the overall set of terms. We apply standard text preprocessing, such as tokenisation, stop-words removal, and stemming, before applying TF-IDF (Term Frequency-Inverse Document Frequency) [6] for term weighting. TF-IDF highlights terms that appear frequently in a given document, but rarely in the overall set, as these are more likely to be representative of the given document.

Non-negative matrix factorisation (NMF) [3] is applied to the created term-document matrix. NMF is a soft clustering algorithm, in which each item can be assigned to several clusters rather than definitively assigned to just one. NMF allows us to identify clusters of items that share latent features. The calculated coefficient matrix presents the membership weights for every module relative to each topic.

We apply Cosine similarity [6] on the coefficient matrix to calculate the similarity between modules. We use these similarity scores to present dependencies between modules in subsequent levels.

3 USE CASES

We can leverage the output of the above approach in multiple ways, serving different purposes and use cases. One such use case is the presentation of a personalised curriculum plan. In this scenario, a student might be interested in a specific area of their programme, and dependencies between modules that are important steps on the way to this goal can be highlighted. For example, Figure 1 shows the personal path for a first year student interested in the area of machine learning. For each year of the student's undergraduate programme, modules with the highest dependencies to this area are shown. Modules in this visualisation are clustered by their highest topic affiliation (colours of nodes) and year of study. The sizes of the nodes (modules) and arrows depict the importance of modules and relationships for the specified area of interest. In the example

Topic	Top Words
"0 - Computer Architecture"	parallel, processor, architecture, performance, cycle
"1 - Programming"	programming, language, java, object oriented
"2 - Theoretical Fundaments"	proof, discrete mathematics, sets, principles
"3 - Linear Algebra"	linear, matrix, algebra, equation, vector
"4 - Software Engineering"	software, development, design, tools, testing
"5 - Machine Learning & AI"	machine learning, game, intelligence, learning
"6 - Databases"	information, database, data, information systems
"7 - Technical Fundaments"	circuits, logic, architecture, digital, systems
"8 - Advanced Programming"	mobile, current, networks, technologies, applications

Table 1: Top words for each of the nine topics

shown, modules belonging to four out of nine different clusters (the number of clusters was determined by experiment; see Table 1) are associated with the learning pathway. In the first year, the diversity of topics is the highest, covering important foundational topics in Computer Science, while topics in later years are more specific to machine learning. Topic 5 (depicted in blue) clusters modules related to artificial intelligence and data science. Unsurprisingly, these make up the majority of the key dependencies, but also modules from topic 2 (theoretical foundations, depicted in red), show strong dependencies to machine learning. While these connections seem obvious to machine learning professionals, we note that none of the modules presented in this example are official pre-requisites.

This example shows how the proposed NMF approach can be used to highlight the thematic relationships between modules to students, as well as which modules taken in the early stages of programmes provide the learning outcomes required for more specialised modules offered in future years. As such, the approach helps students to better understand the module space and to make more informed decisions when planning their module choices. Moreover, it also provides students with a natural explanation as to why certain modules are important to their chosen area of study, thereby

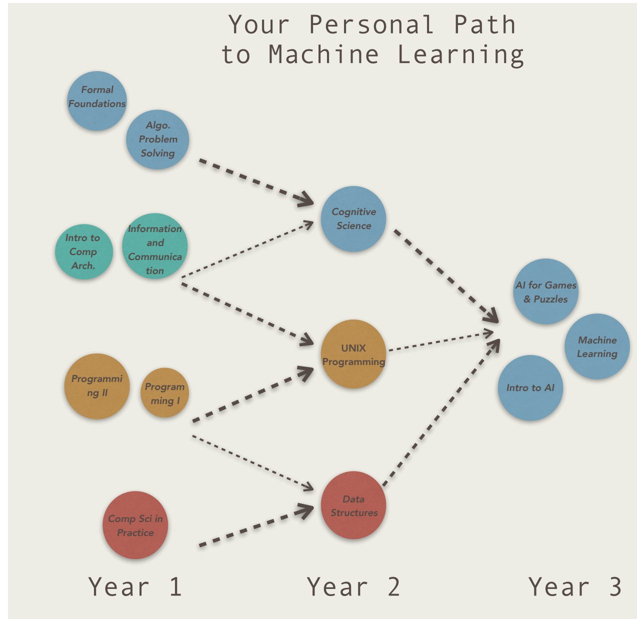

Figure 1: Use Case Visualisation Prototype

providing an added motivation to focus on and excel in modules which may appear tangential to their ultimate goals.

Ultimately, the output of the dependency analysis will allow us to improve our current recommender system [2] by adding valuable information to the user modelling process.

4 CONCLUSION & FUTURE WORK

Using NMF to detect dependencies allows us to find latent factors within the academic module space. Visualising these dependencies can serve multiple use cases and will ultimately allow students to better explore the module space, make informed decisions, and plan their academic career.

We conducted an offline evaluation using a ground truth build by university alumni. We found that our approach detects the majority of explicit dependencies and furthermore identifies additional implicit dependencies between modules. Our next steps involve improving our system [2] to include additional use cases, enhance the module recommender component and to undertake a comprehensive user evaluation of the system.

REFERENCES

[1] Vangel V Ajanovski. 2017. 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*. 43.

[2] Nina Hagemann, Michael P. O'Mahony, and Barry Smyth. 2018. Module Advisor: Guiding Students with Recommendations. In *International Conference on Intelligent Tutoring Systems*. Springer, 319–325.

[3] Daniel D Lee and H Sebastian Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *Nature* 401, 6755 (1999), 788.

[4] Sara Morsy and George Karypis. 2018. Learning Course Sequencing for Course Recommendation. (2018).

[5] Young Park. 2017. A Recommender System for Personalized Exploration of Majors, Minors, and Concentrations. (2017).

[6] G. Salton and J.M. McGill. 1986. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc.

[7] G. Salton, A. Wong, and C.S. Yang. 1975. A Vector Space Model for Automatic Indexing. *Commun. ACM* 18, 11 (1975), 613–620.