

Investigating Influence of Demographic Factors on Study Recommenders

Michal Huptych^{1,2}, Martin Hlosta¹, Zdenek Zdrahal^{1,2}, and Jakub Kocvara¹

¹ Knowledge Media Institute, The Open University, Milton Keynes, UK,
{michal.huptych,martin.hlosta,zdenek.zdrahal,jakub.kocvara}@open.ac.uk

² CIIRC, Czech Technical University in Prague, CZ

Abstract. Recommender systems in e-learning platforms can utilise various data about learners in order to provide them with the next best material to study. We build on our previous work, which defines the recommendations in terms of two measures (i.e. *relevance* and *effort*) calculated from data of successful students in the previous runs of the courses. In this paper we investigate the impact of students' socio-demographic factors and analyse how these factors improved the recommendation. It has been shown that education and age were found to have a significant impact on engagement with materials.

Keywords: personalised learning, educational recommender systems

1 Introduction

In distance education, most of the learning takes place in Virtual Learning Environments (VLE) allowing learners flexible way of studying. E-learning systems minimise the costs of education and enable to scale up the number of students, which would be impossible or very difficult to achieve in a traditional face to face learning environment. On the other hand, the students lose frequent contact with teachers, their supervision and possible valuable feedback, which can help them with guidance and organisation of their study. As a result, many students drop out before completing the course.

Education domain brings many challenges caused by its characteristic features. For example, if the intention is to increase students performance, a recommendation of study materials used by students with similar learning behaviour might not direct student towards success. The other specific feature includes the existence of limited time of learners they have to reserve to achieve a task, possible boundaries of the duration of the course, dependencies and prerequisites among study materials, taking into account the existing skills and knowledge of students before students' studying had started [2]. These specifics pose a necessity to develop new methods tailored to education.

According to review [2], most of the educational recommenders used Collaborative Filtering (CF), with growing interest of incorporating characteristics of the educational domain. For example, prioritising features and ratings of good performing students was used in [3]. Personalised recommender systems take

into account the past activities of each individual learner. CF achieves this by behavioural or rating pattern. However, it suffers from a cold-start problem, needing initial information about the users, which is not available at the start of the course [7]. The cold-start problem hinders also content-based methods, which operates on the similarity of the items that users liked in the past rather than similarity to users.

Existing research in RS mentions using demographics to mitigate the cold-start. However, [6] mentions student educational background only marginally and [8] is targeted on recommending courses. [1] found socio-demographic the best features for improving the student engagement with MOOCs, however for recommending peers. Students were more engaged when they were recommended to connect with their peers based on demographic factors, such as age and country. [4] found gender and achievement-goal orientation as valuable predictors of student engagement with their online learning system in some cases. Other than that, we were not able to found any other work utilising demographic data.

2 Personalisation by Socio-Demographic factors

The goal of this paper is to investigate methods of personalisation as described in our previous work in [5]. We design the personalised study recommender by constructing a click-based behavioural pattern from the activities of successful students. We define two new measures: *relevance* and *effort*. The *relevance* measures the average activity spent by the selected cohort (the excellent group) on each educational material in the previous presentation, while the *effort* indicates the engagement of each individual student with the educational material in the current presentation. *Relevance* is the 'content to be learned' while the *effort* shows what has the individual 'already learned'. The recommendation strategy says that for each educational resource and each current student his/her *effort* should balance the resource *relevance*.

At the start of the course no student has engaged with any study material, thus each student's *effort* is zero. *Relevance* calculated from activities of the excellent group in the previous presentation is the same for all students.

However, demographic and other static data about each student are available. Could we enrich the VLE activity data by this additional information to further refine and personalise the recommendations for groups of students with similar VLE characteristics? In another word, do students with different demographic data and equal VLE behaviour manifest different responses in the educational process? The available features included 1) **Highest education**; from 'No formal' to 'Postgrad', 2) flag if student **is new** to the university, 3) **Motivation**, 4) **Occupation**, 5) **Age**, 6) **Gender**, 7) **IMD** – Index of Multiple Deprivation.

3 Experiments and Impact of Socio-Demographic Factors

The experiments were performed on 4 presentations of a technology-related course. The goal was to evaluate the impact of each individual socio-demographic

factor on students' effort. The values of investigated factors were split into two groups. For categorical-valued factors, the groups were defined by the selection of categories, for ordinal-valued factors the groups were defined by selecting the threshold. The contingency table was constructed for each analysed factor and the groups of excellent students ($\geq 75\%$) and others. The dependency was evaluated by Mann-Whitney test at the significance level $p < 0.1$. The following results have been achieved:

The comparison of students with A level or higher education and the rest of the cohort revealed differences in activities depending on their **educational backgrounds**. For excellent students group, the mean cumulative *effort* was 11% higher for students with lower than A level education. These students were more active on relevant materials, although these differences were less noticeable among students with scores lower than 75%. These findings proved to be consistent across all presentations that were tested.

The most influential factor affecting students' behaviour was their **age**. The biggest gap in activities was found for the split at the age of 30 years. Therefore two groups were created: up to 29 years and 30+ years. Although the mean cumulative *effort* of these groups has similar values at the end of the presentation, important differences have been found at individual blocks and materials, e.g. 18 % higher activity for students 30+ in Block 3 Part 2 (see Figure 1). Older students systematically showed higher engagement with materials and also started interacting with these materials earlier, often even before they were assigned by the tutor. This was the case for both performance groups and all presentations.

Influence of factors new/continuing, occupation, motivation and IMD did not prove to be statistically significant.

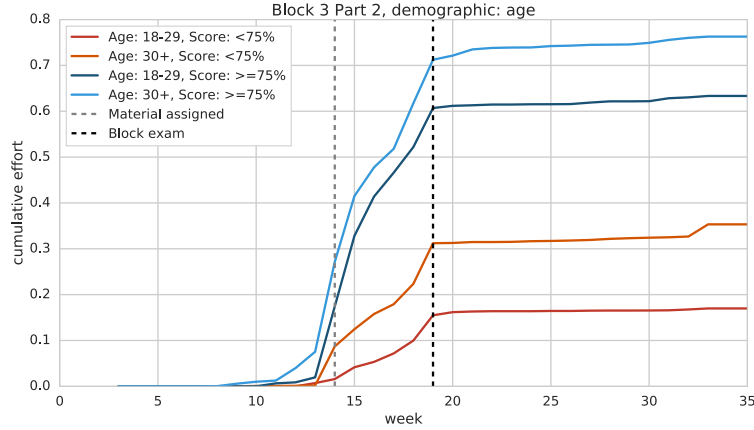


Fig. 1. Change of cumulative *effort* over time on Block 3 Part 2 in 2015. After a block exam in week 19, *effort* becomes saturated.

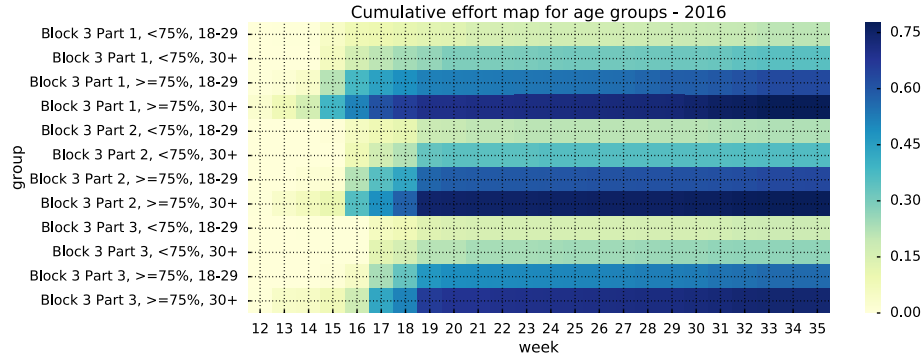


Fig. 2. Heat-map of cumulative *effort* over time for age and performance groups in 2016. Darker squares indicate higher engagement.

The cumulative *effort* has a similar pattern for the corresponding performance and age groups across study materials. The 30+ groups have always been more active and started with each Block/Part earlier than the 18-29 group, see Figure 1 and in Figure 2.

4 Discussion and Conclusions

We presented the analysis of demographic factors with the aim of investigation their impact on recommendations. Recommendation are defined using two measures: *relevance* and *effort*, which reflect the importance of the study material and engagement of the individual student with this material. We selected 4 presentations of one technology-related course and performed the examination of the impact of 5 factors' categories on the *effort* of students.

As expected, there are significant and consistent differences in *effort* between groups of students divided according to their performance. This fact is prevalent among all demographic factors and it confirms our presumption that correlation between *effort* and performance is always evident.

Furthermore, we analysed the *effort* within the excellent group and its relationship with students' demographic factors. We found two factors that contribute to the differences in *effort*: education and age. Based on our investigation, the students from the 30+ group had considerably higher engagement with the study materials. They started to interact with them earlier and reached higher values of *effort* at the moment of the Block exam, especially in the Parts of the Block 3. One of the reasons might be that younger students are more familiar with learning from online materials in general. But a closer investigation revealed that the Block 3 is focused on modern technologies such as Cloud computing and GPS.

The results also show that older students need more time to study such topics. This applies both to excellent students and their lower performing colleagues.

References

1. Bouchet, F., Labarthe, H., Yacef, K., Bachelet, R.: Comparing peer recommendation strategies in a mooc. In: Adjunct Publication of the 25th Conf. on User Modeling, Adaptation and Personalization. pp. 129–134. ACM (2017)
2. Drachsler, H., Verbert, K., Santos, O.C., Manouselis, N.: Panorama of Recommender Systems to Support Learning, In Recommender Systems Handbook (eds: F .Ricci and L. Rokach and S. Bracha). Springer US, Boston, MA (2015)
3. Ghauth, K.I., Abdullah, N.A.: The effect of incorporating good learners’ ratings in e-learning content-based recommender system. *Journal of Educational Technology & Society* 14(2), 248 (2011)
4. Guerra, J., Hosseini, R., Somyurek, S., Brusilovsky, P.: An intelligent interface for learning content: Combining an open learner model and social comparison to support self-regulated learning and engagement. In: Proc. of the 21st Intl. Conf. on Intelligent User Interfaces. pp. 152–163. IUI ’16, ACM, New York, NY, USA (2016), <http://doi.acm.org/10.1145/2856767.2856784>
5. Huptych, M., Bohuslavsek, M., Hlosta, M., Zdrahal, Z.: Measures for recommendations based on past students’ activity. In: LAK ’17 Proc. of the 7th Intl. Learning Analytics & Knowledge Conf. on - LAK ’17. pp. 404–408 (2017)
6. Kerkiri, T., Manitsaris, A., Mavridis, I.: How e-learning systems may benefit from ontologies and recommendation methods to efficiently personalise resources. *Intl. Journal of Knowledge and Learning* 5(3-4), 347–370 (2009)
7. Nabizadeh, A.H., Mário Jorge, A., Paulo Leal, J.: Rutico: Recommending successful learning paths under time constraints. In: Adjunct Publication of the 25th Conf. on User Modeling, Adaptation and Personalization. pp. 153–158. UMAP ’17, ACM, New York, NY, USA (2017)
8. Wen-Shung Tai, D., Wu, H.J., Li, P.H.: Effective e-learning recommendation system based on self-organizing maps and association mining. *The Electronic Library* 26(3), 329–344 (2008)