Measures for recommendations based on past students' activity

Michal Huptych[1,2] Michal Bohuslavek[1, 3] Martin Hlosta[1] Zdenek Zdrahal[1, 2]

Knowledge Media Institute [1]
The Open University, Walton Hall
Milton Keynes, MK7 6AA, UK
{michal.huptych; martin.hlosta; z.zdrahal}@open.ac.uk

CIIRC, [2]
Czech Technical University
Zikova street 1903/4
Prague, 166 36
Czech Republic

Faculty of Mechatronics, Informatics[3]
and Interdisciplinary Studies
Technical University of Liberec
Studentska 1402/2, 461 17 Liberec 1
Czech Republic
michal.bohuslavek@tul.cz

ABSTRACT

This paper introduces two measures for the recommendation of study materials based on students' past study activity. We use records from the Virtual Learning Environment (VLE) and analyse the activity of previous students. We assume that the activity of past students represents patterns, which can be used as a basis for recommendations to current students.

The measures we define are *Relevance*, for description of a supposed VLE activity derived from previous students of the course, and *Effort*, that represents the actual effort of individual current students. Based on these measures, we propose a composite measure, which we call *Importance*.

We use data from the previous course presentations to evaluate of the consistency of students' behaviour. We use correlation of the defined measures *Relevance* and *Average Effort* to evaluate the behaviour of two different student cohorts and the Root Mean Square Error to measure the deviation of *Average Effort* and individual student *Effort*.

CCS Concepts

•**Applied computing** → **Education;** *E-learning;* Distance learning;

Keywords

Learning strategy; Recommendation; Student Retention; Learning Analytics; Relevance; Effort

1. INTRODUCTION

Data and metadata generated by e-learning systems can be fed back to various education-related tasks, such as the evaluation of learning materials and the design of new materials [4], predictions of student performance [13][16], recommendation of learning materials and the creation of personalised study plans ([8][2][4][11]).

LAK '17, March 13 - 17, 2017, Vancouver, BC, Canada

ISBN 978-1-4503-4870-6/17/03…$15.00

DOI: http://dx.doi.org/10.1145/3027385.3027426

In order to improve student learning, it is necessary to know which learning activities lead the students towards success. In the case of an online environment with large amounts of materials, this might be difficult to obtain manually. However, there are several techniques that allows to process the data an automated way. The important is to specify strategy which might by use for description and representation of data.

Recommender systems provide information, items of interest or services to the user according to the users' activities and preferences. This paper presents a new approach to recommender design. Recommenders evaluate user behaviour and preferences and offer the user the most appropriate learning resource. There are different recommender techniques [1] [12] implemented in a number of recommender systems [7][15]. According to [12] these techniques can be divided into four categories:

- **Collaborative techniques** construct recommendations from the behaviour and results of similar learners. Similarity is usually calculated from the VLE activities of the recommendation recipient and other learners in the present or past courses. A detailed description of collaborative recommenders can be found in [12][3][5][2].
- **Content-based techniques** use for recommendation only information about the users and their histories [12]. Typical problem solving methods are Case Based Reasoning and Attribute-based techniques, which derive the recommendations from the learner profile [1][6] [18][12][14].
- **Matrix/tensor factorization techniques** consist of decomposition of a tensor to factors. The recommendation calculates factorization of known tensor values, and use the product of factors to obtain the vector of unknown values. For details see [16].
- **Association rules** are machine learning techniques for discovering dependence patterns in data. The recommender mines rules from activities of learners in the past to recommend activities to the current learner. Examples of association rules used for recommendation are in [10][17].

The approach presented in this paper draws from the collaborative techniques and association rules. We evaluate the VLE activities of successful students in the previous presentation, compare it with the currently supported learner and recommend activities that should decrease the differences between the two. Some topics introduced earlier in the study plan are prerequisites for ones presented later, e.g. knowing the HTML language is necessary to understand the design of web applications. These dependencies are reflected in the learner behaviour and could be discovered from the measures introduced in this paper.

2. PROBLEM DESCRIPTION

At the Open University (OU), courses (modules) take usually about 40 weeks and are offered to students in a number of consecutive years. Each module has a study plan which breaks the course content down into Blocks. Each Block presents a different topic taught in the course and is further divided into Parts (1 week long). Thus, Block 1 Part 1 refers to week 1, Block 1 part 2 to week 2, etc. The study plan usually does not significantly change between presentations. Study materials are provided in the Virtual Learning Environment (VLE) and therefore student clicks can be recorded. Each click has a 'semantic label' called activity type, which indicates the kind of interaction with the VLE. Examples of activity types are forum, resource, ou-content or quiz. Clicks on different activity types have different information content; resource is a page with text in pdf and therefore one click provides access to all the underlying content. On the other hand, ou-content refers to the study materials represented in usually highly structured HTML and the number of VLE accesses pretty well represents student effort. Key study materials in modules are represented as ou-content and for this reason we analyse clicks labeled as ou-content, both in the previous, already completed presentation and in the current one.

Each block in the study plan has associated expected study time. However, since student VLE interactions in the previous presentations are recorded, the real effort required for understanding each topic can be measured in terms of the average number of clicks of successful students on the corresponding web pages. We assume that the performance of students who passed the previous presentation well approximates the effort required at present.

The problem addressed in this paper is how to use ou-content VLE activities of the previous presentation and VLE data collected from current students to design a personalised study recommender that navigates students through the study plan.

Measuring of the time-on-task is not simple [9]. In our case the approximation by number of clicks is sufficient.

3. RECOMMENDATION STRATEGY

The recommendation strategy is constructed from relevance of the study material and learners' activity. These concept are formally defined in the following sections.

3.1 Capturing study materials' relevance

Relevance is defined as a normalized difference of the average cumulative students activity a, measured by the cumulative number of clicks on a specific study activity, between two consecutive weeks *i-1* and i:

$$R\left(w,a\right)=\frac{\sum_{i=1}^{w}c_p\left(i,a\right)-\sum_{i=1}^{w-1}c_p\left(i,a\right)}{\sum_{i=1}^{N}c_p\left(i,a\right)}, \tag{1}$$

where $c_p\left(i,a\right)$ is the number of clicks for the activity a in week i, $\sum_{i=1}^{w}c_p\left(i,a\right)$, $\sum_{i=1}^{w-1}c_p\left(i,a\right)$ are cumulative clicks from the beginning (week 1) to week w and $w-1$, respectively. $\sum_{i=1}^{N}c_p\left(i,a\right)$ is the cumulative sum until the last week N of the previous presentation. Henceforth:

- *Relevance* is always non-negative, $\forall w \forall a, R\left(w,a\right)\geq 0$,
- sum of the *Relevance* for each activity over all weeks is 1, $\forall a, \sum_w R\left(w,a\right)=1$,
- *Relevance* of each activity is the same for all students.

An example of the cumulative clicks for 5 selected activities is shown in Figure 1, the corresponding relevance is shown in Figure 2.

3.2 Capturing learners' activity

Further, we need a measure that can capture the activity of the learner in the VLE, that can be related to relevance. Therefore, we create a measure *Effort* and define it as:

$$E\left(w,a\right)=\frac{\sum_{i=1}^{w}c_c\left(i,a\right)-\sum_{i=1}^{w-1}c_c\left(i,a\right)}{\sum_{i=1}^{N}c_p\left(i,a\right)}, \tag{2}$$

where $c_c\left(i,a\right)$ is number of clicks for activity a in week i from current student,$c_p\left(i,a\right)$ is number of clicks for activity a in week i from previous presentation, $\sum_{i=1}^{w-1}c_c$, $\sum_{i=1}^{w}c_c\left(i,a\right)$ are the numbers of cumulative clicks for given activity from the beginning of the current presentation to week $w-1$ and w, respectively, and $\sum_{i=1}^{N}c_p\left(i,a\right)$ is the number of cumulative clicks until the last week of the previous presentation. Henceforth:

- sum of the *Effort* for each activity over all weeks can reach one of the following eventuality:

$$\forall a, \sum_w E\left(w,a\right) is \begin{cases} <1, & \text{if } \sum_{i=1}^{N}c_p\left(i,a\right) \\ & >\sum_{i=1}^{w}c_c\left(i,a\right) \\ =1, & \text{if } \sum_{i=1}^{N}c_p\left(i,a\right) \\ & =\sum_{i=1}^{w}c_c\left(i,a\right) \\ >1, & \text{if } \sum_{i=1}^{N}c_p\left(i,a\right) \\ & <\sum_{i=1}^{w}c_c\left(i,a\right) \end{cases} \tag{3}$$

- *Effort* is given for each student individually.
- *Average Effort* is given as average of *Effort* over all students

Thus, the effort represents an approximation of the progress for the given activity for an individual student. An example of effort is shown in Figure 3.

Relevance and Effort, formalized by (1) and (2), capture our intuition of a transferring of the past experience (*Relevance*) to the behaviour of current student (*Effort*).

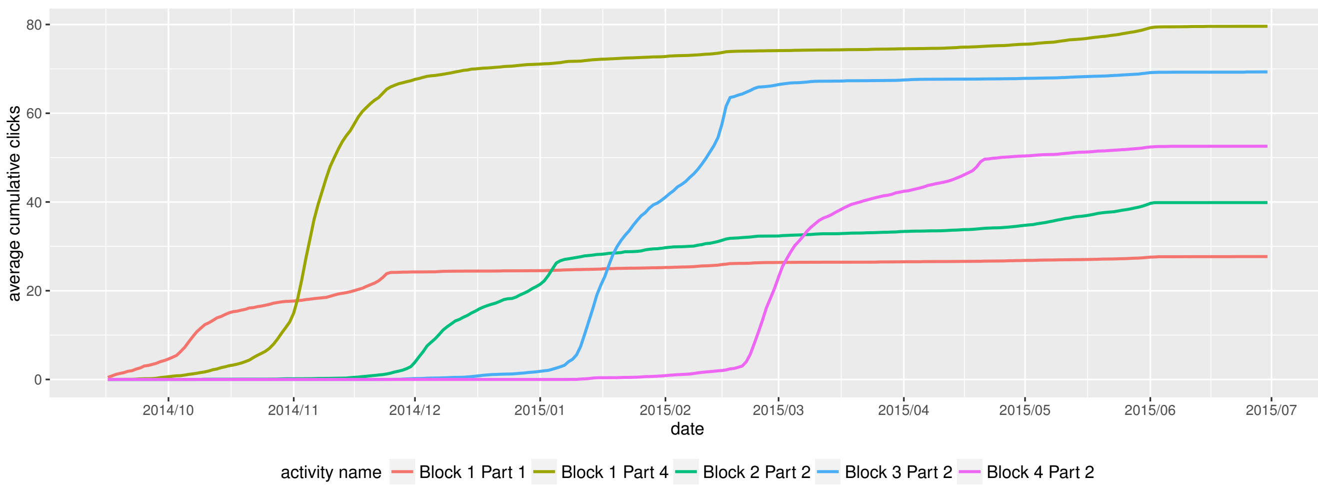

Figure 1: Average number of cumulative clicks in time

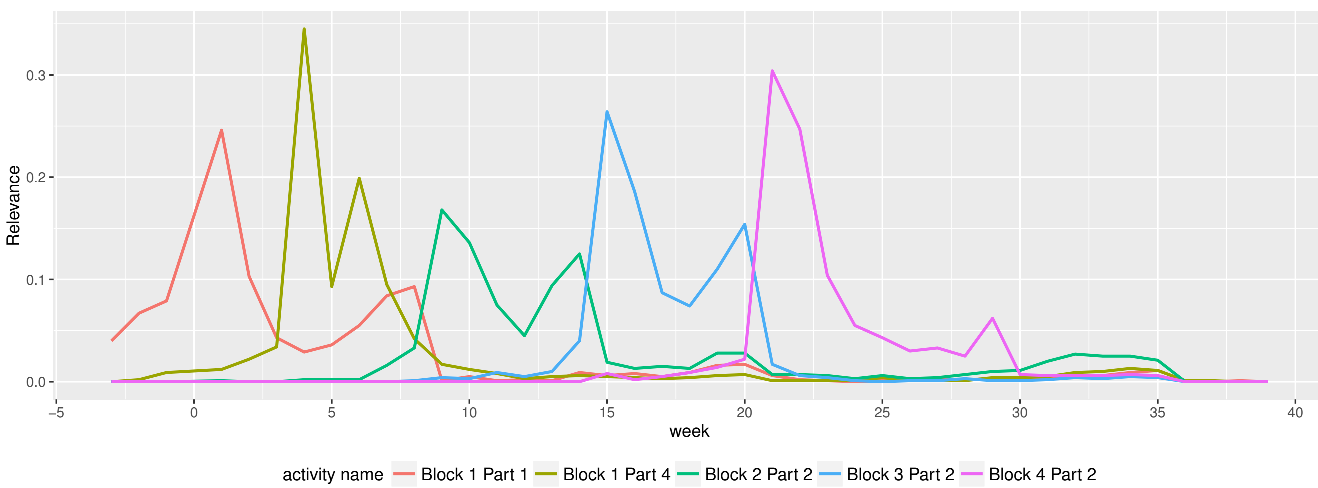

Figure 2: *Relevance* derived from the cumulative clicks

3.3 Recommendation

Thus, we propose a recommender strategy to output for each activity a in week i it's *Importance* as:

$$I(w, a) = R(w-1, a) - E(w-1, a), \qquad (4)$$

where $R(w-1, a)$ and $E(w-1, a)$ are appropriate *Relevance* and *Effort* for given activity in a previous week, respectively. Thus, the *Importance* represents a combination of information of the *Relevance* of some activity in the previous week and *Effort* of the student for the given activity.

4. EVALUATION

We can empirically evaluate similarity between students behaviour for the current and previous presentation. We use 2014 presentation for computing the *Relevance* and 2015 as the presentation for retrieving the learners *Effort*.

In both presentations, we select only successful students. We disregard the failed/withdrawn students because the previous research [13] shows that VLE behaviour is the discriminative factor between successful and unsuccessful students. From the previous presentation we selected 1,062 students and from the current one 922 students. We focus only on the activity types for which we know that the repeated clicking is relevant, i.e. ou-content.

The *Relevance* and the *Effort* are both positive for all activities and weeks. If we use an *Average Effort* (over all students) in particular weeks, we can postulate that the *Relevance* and the *Average Effort* should be correlated. To measure the similarity, we use Pearson's correlation.

Figure 4 shows that the *Relevance* of the educational activities in the previous presentation is similar with the *Effort* in the current presentation across all the weeks for successful students. This means that a) the behaviour of the successful students does not change from the previous to the current presentation and b) the use of *Effort* value will recommend the activity which should allow the learner to achieve similar results as the successful students in the topics where they

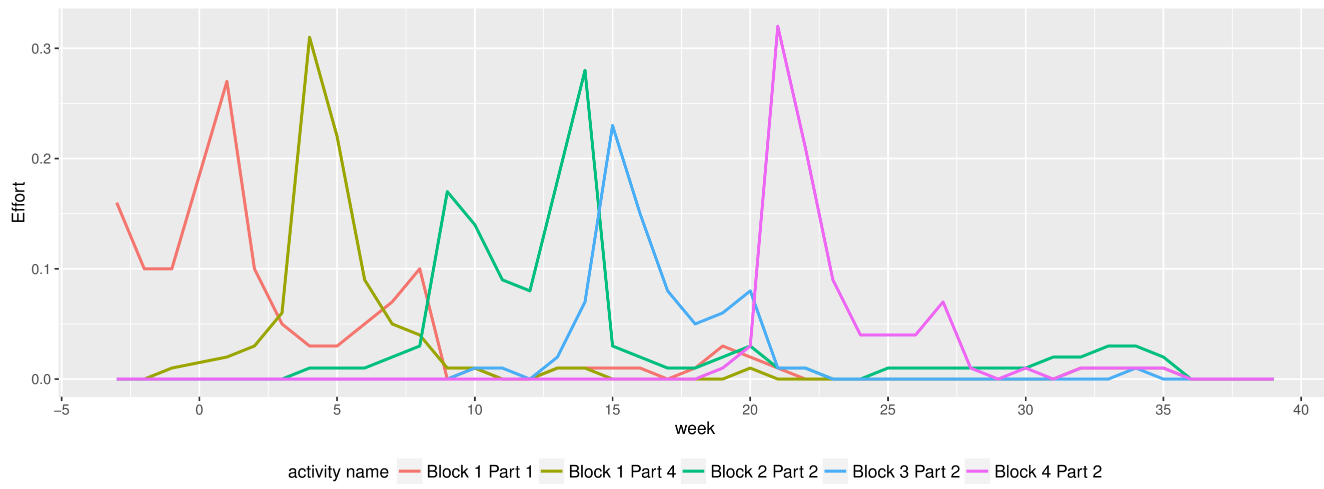

Figure 3: Example of the *Effort*

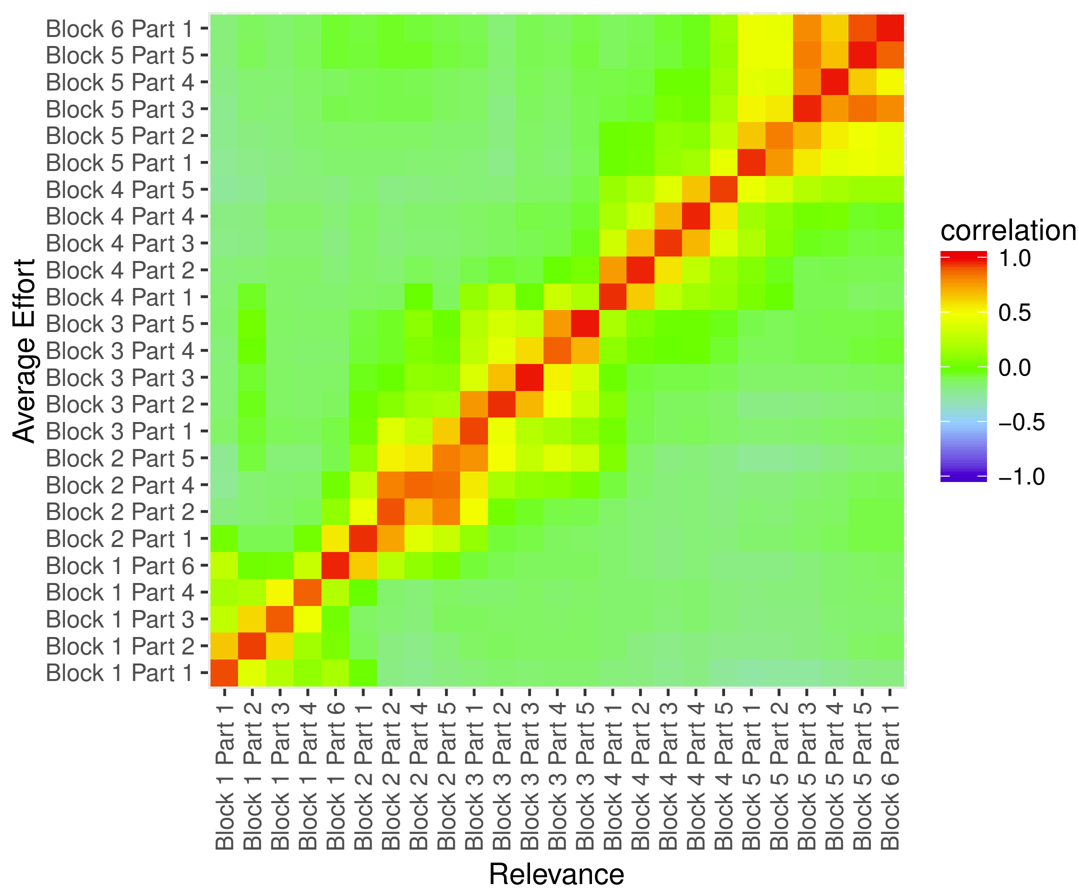

Figure 4: Correlation matrix for *Relevance* of previous presentation and *Average Effort* for current presentation

are lagging behind.

To show a deviation of the *Average Effort* and individual *Efforts* we use the Root Mean Square Deviation (RMSD) (definition in [16]). The RMSD for the selected particular activities is shown in Table 1.

Dependencies between topics are shown in Figure 2. For example, though the highest relevance of Block 1 Part 1 is in about week 1 of the presentation, the topic is obviously

Table 1: RMSD of average *Effort* and particular individual *Efforts*

Activity name	RMSD	Activity name	RMSD
Block 1 Part 1	0.14	Block 3 Part 5	0.15
Block 1 Part 2	0.13	Block 4 Part 1	0.10
Block 1 Part 3	0.11	Block 4 Part 2	0.11
Block 1 Part 4	0.12	Block 4 Part 3	0.12
Block 1 Part 6	0.11	Block 4 Part 4	0.14
Block 2 Part 1	0.17	Block 4 Part 5	0.14
Block 2 Part 2	0.16	Block 5 Part 1	0.12
Block 2 Part 4	0.17	Block 5 Part 2	0.12
Block 2 Part 5	0.18	Block 5 Part 3	0.14
Block 3 Part 1	0.14	Block 5 Part 4	0.16
Block 3 Part 2	0.11	Block 5 Part 5	0.12
Block 3 Part 3	0.14	Block 6 Part 1	0.14
Block 3 Part 4	0.12	–	–

also relevant in week 7 and 8. Similar dependencies exist between other topics.

5. CONCLUSIONS AND FUTURE WORK

In this work, we propose a novel strategy for personalized study recommendation that utilises the information from the successful students in the previous presentation. We define two measures, *Relevance* and *Effort*, which describe a past students' behaviour and current student's effort, respectively. Further, we define the theoretical principle of the recommendation based on these two measures, which we call *Importance*.

We use the historical VLE activity for evaluation of our concept by correlating *Relevance* and *Effort*, which represents consistency of students behaviour between both presentations. The result shows a correlation ($means \pm std = 0.94 \pm 0.05$) between the activities of previous and current students. We interpret this finding as confirmation that the successful students have an important and significant pattern of learning.

Currently, we are enriching the OUAnalyse system with the proposed recommender and we are planning to evaluate it's impact on students behaviour.

6. REFERENCES

[1] 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. on Knowl. and Data Eng.*, 17(6):734–749, April 2005.

[2] A. R. Anaya, M. Luque, and M. Peinado. A visual recommender tool in a collaborative learning experience. *Expert Systems with Applications*, 45:248–259, March 2016.

[3] J. Bobadilla, F. Serradilla, and A. Hernando. Collaborative filtering adapted to recommender systems of e-learning. *Knowledge-Based Systems*, 22(4):261–265, May 2009.

[4] M.-I. Dascalu, C.-N. Bodea, M. N. Mihailescu, E. A. Tanase, and P. O. de Pablos. Educational recommender systems and their application in lifelong learning. *Behaviour & Information Technology*, 35(4):290–297, January 2016.

[5] M.-I. Dascalu, C.-N. Bodea, A. Moldoveanu, A. Mohora, M. Lytras, and P. O. de Pablos. A recommender agent based on learning styles for better virtual collaborative learning experiences. *Computers in Human Behavior*, 45:243–253, 2015.

[6] H. Drachsler, H. G. Hummel, and R. Koper. Personal recommender systems for learners in lifelong learning networks: the requirements, techniques and model. *International Journal of Learning Technology*, 3(4):404–423, July 2008.

[7] H. Drachsler, K. Verbert, O. C. Santos, and N. Manouselis. *Panorama of Recommender Systems to Support Learning, In Recommender Systems Handbook (eds: F .Ricci and L. Rokach and and S. Bracha).* Springer US, Boston, MA, 2015.

[8] G. Durand, N. Belacel, and F. LaPlante. Graph theory based model for learning path recommendation. *Information Sciences*, 251:10–21, December 2013.

[9] V. K. et al. Does time-on-task estimation matter? implications on validity of learning analytics findings. *Journal of Learning Analytics*, 2(3):81–101, February 2016.

[10] E. García, C. Romero, S. Ventura, and C. de Castro. An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering. *User Modeling and User-Adapted Interaction*, 19(1):99–132, February 2009.

[11] A. Garrido, L. Morales, and I. Serina. On the use of case-based planning for e-learning personalization. *Expert Systems with Applications*, 60:1–15, October 2016.

[12] A. Klašnja-Milićević, M. Ivanović, and A. Nanopoulos. Recommender systems in e-learning environments: a survey of the state-of-the-art and possible extensions. *Artificial Intelligence Review*, 44(4):571–604, December 2015.

[13] J. Kuzilek, M. Hlosta, D. Herrmannova, Z. Zdrahal, and A. Wolff. Ou analyse: analysing at-risk students at the open university. *Learning Analytics Review*, LAK15-1:1–16, March 2015.

[14] J. Liu, P. Dolan, and E. R. Pedersen. Personalized news recommendation based on click behavior. In *Proceedings of the 15th International Conference on Intelligent User Interfaces*, pages 31–40. ACM, February 2010.

[15] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang. Recommender system application developments: A survey. *Decision Support Systems*, 74:12–32, June 2015.

[16] N. Thai-Nghe, L. Drumond, A. Krohn-Grimberghe, and L. Schmidt-Thieme. Recommender system for predicting student performance. *Procedia Computer Science*, 1(2):2811–2819, 2010.

[17] F.-H. Wang and H.-M. Shao. Effective personalized recommendation based on time-framed navigation clustering and association mining. *Expert Systems with Applications*, 27(3):365–377, October 2004.

[18] Y. J. Yang and C. Wu. An attribute-based ant colony system for adaptive learning object recommendation. *Expert Systems with Applications*, 36(2):3034–3047, March 2009.