Ogata, H. et al. (Eds.) (2015). Proceedings of the 23rd International Conference on Computers in Education. China: Asia-Pacific Society for Computers in Education

Context-aware Personalized Courses Search based on Hybrid Learner Profile

Haoran XIE[a], Di ZOU[b*], Fu Lee WANG[a], Tak-Lam WONG[c] & Ksenia TROSHINA[b]

[a]*Caritas Institute of Higher Education, Hong Kong SAR*

[b]*English Language Centre, The Hong Kong Polytechnic University, Hong Kong SAR*

[c]*Department of Mathematics and Information Technology,*
The Hong Kong Institute of Education, Hong Kong SAR

* daisyzou@polyu.edu.hk

Abstract: With the rapid growth of massive online open courses (MOOCs) on the Web, it is essential to provide learners with appropriate assistance in courses and learning materials. The extant approaches of personalized course search mainly consider historical learnt and enrolled courses of learners. That is, those courses which are contently similar to previous courses in learner profiles will be highlighted in the ranking results of the personalized course search. However, these approaches mainly neglect two distinguished characteristics in this domain, which are (i) context-dependent: course search which is highly correlated with learner contexts, e.g., a learner may have the individual learning schedule of the courses to be retrieved depending on the temporal contexts; and (ii) knowledge-constrained: learners are more willing to search and enroll in the courses that they have sufficient pre-knowledge about. To incorporate these two domain characteristics of the personalized course search, we therefore present a novel approach based on hybrid learner profile in this paper. Furthermore, we conduct the experiments which compare the performance of different methods on a dataset to verify the effectiveness of the proposed method for the personalized course search.

Keywords: context-aware, personalized course search, learner profile, e-learning, MOOCs

1. Introduction

With the rapid growth of massive online open courses (MOOCs) such as Coursera, Edx and Udacity on the Web, users have more choices and opportunities to take online high-quality open courses to acquire new knowledge and skills. On the other hand, they also face the problem of information overload when confronting such large volumes of learning resources. In other words, it is quite challenging for learners to find a suitable course which would match their interests. To address this issue, the personalized course search approaches which incorporate the individual preferences and the learning logs of learners into the process of finding relevant courses are very important and indispensible. The extant approaches of personalized course search mainly consider historical learnt and enrolled courses of learners. That is, those courses which are contently similar to previous courses in learner profiles will be highlighted in the ranking results of the personalized course search. However, these approaches mainly neglect two distinguished characteristics in this domain as follows.

Context-dependent: course search are often highly correlated with learner contexts. For example, a learner may have an individual learning schedule so the retrieved courses often depend on the temporal contexts of this learner (i.e., whether today is weekdays or not).

Knowledge-constrained: Individual knowledge background seems to constraint the course search results. More specifically, students are more willing to search and enroll in the courses that they have sufficient pre-knowledge about.

To incorporate two domain characteristics of the personalized course search, contextual information and pre-knowledge of users should be captured and modeled in the search framework. Therefore, we employ a hybrid learner profile which integrates various sources such as learning and browsing logs, pre-knowledge levels, demographic data as well as contextual information to facilitate the personalized course search. Particularly, we also propose a novel approach for personalized course search based on the hybrid learner profile in this paper. To evaluate the effectiveness of the proposed method, we further conduct the experiment which compares its performance to the search accuracy and efficiency of state-of-the-art baselines.

The rest of this paper is organized as follows. In Section 2, we review the related research on personalized approaches in MOOCs (or other e-learning systems) and learner profiling techniques as well as the corresponding application. Section 3 discusses our methodology including the hybrid learner profile and context-aware personalized search methods. Section 4 reports the processes, metrics and results of experiments. Finally, we briefly summarize this research and outline the potential directions for future research in Section 5.

2. Related Work

With the development of personalized techniques in information retrieval (IR) and data explosion in online e-learning systems such as MOOCs, it is quite natural and prominent to adopt personalized approaches to facilitating e-learning users' access and retrieving a large volume of learning resources. An intelligent agent, named e-Teacher, was presented to offer the personalized assistances to e-learning students by taking student behaviors such as learning style, learning performance into consideration (Schiaffino et al., 2008). Limongelli et al. (2012) proposed a comprehensive framework supporting the tasks of defining, retrieving, and importing learning objects for personalized courses in Moodle platform so that the teachers can retrieve and manage the learning materials according to their personalized contexts. More recently, web 3.0 approaches (Kurilovas et al., 2014), and generalized metrics (Essalmi et al., 2015) are exploited for personalization in e-learning systems.

The learner profiling techniques are generally based on the user modeling techniques in the area of human computer interaction and several domain characteristics of e-learning. Chen et al. (2007) exploited the association rule methods to mine the learner profile for identifying common learning misconceptions during learning processes. Özpolat and Akar (2009) addressed the problem of how to automatically extract the learner profile based on Felder–Silverman learning style model. Feng et al. (2011) investigated how to make use of roles to define the learner profile for promoting collaborative learning. Recently, a learner profile based on information flow approach was proposed to support personalized learning (Yang, 2013). Zou et al. (2014) adopted the involvement load to estimate the learner efforts in word acquisition for recommending suitable word learning tasks.

3. Methodology

In this section, we introduce the overall framework of the proposed methodology, which mainly tackles two research questions: (i) how to build the learner profiles from multiple data sources and (ii) how to facilitate the personalized course search based on the built learner profiles. The hybrid learner profiling and the personalized course search are detailed in the following subsections to address the above two questions, respectively.

3.1 Hybrid Learning Profiling

In our research, we also employ the vector form in construction of learner profiles (Yang, 2013; Zou et al., 2014). As there are multiple data sources available for building learner profiles, we therefore mainly address the following two issues: (i) what information should be extracted from multiple sources; and (ii) how to convert the various data sources containing diverse information into the vector presentation.

To answer the first question, we mainly extract two kinds of information from multiple sources which are pre-knowledge levels and learner preferences. Pre-knowledge levels are denoted by a knowledge distribution on all knowledge units of courses in an e-learning system. The details of how to establish the relationship between knowledge units and courses are discussed by Leung and Li (2007). In the context of this study, we consider the knowledge units as the basic elements of a course, which seems to be sufficient in terms of understanding the purpose of this research. Formally, we define a hybrid learner profile containing two following elements.

Definition 1: Let U_i=$\{u_1 : \varepsilon_1^i,\ u_2 : \varepsilon_2^i, \ldots u_n : \varepsilon_n^i\}$ be knowledge units of all courses and the corresponding degree of pre-knowledge for each unit by learner l_i, and T_i=$\{t_1 : \sigma_1^i,\ t_2 : \sigma_2^i, \ldots t_m : \sigma_m^i\}$

are topics (categories) for all courses and the corresponding degree of preference for each topic by learner, the learner profile of l_i is denoted by a vector $\vec{L}_i$ as:

$$\vec{L}_i = (U_i, T_i)$$

Note that we do not employ the courses in either U_i (knowledge units) or T_i (topics) for construction of the learner profiles. Specifically, the degree of pre-knowledge mastered by the learner for each knowledge unit can be obtained from learning historical documents (e.g., the learnt courses or completed assignments). In the most MOOCs, the learnt courses and the grades are available information sources for each learner, so we use the learnt courses as the source to obtain pre-knowledge levels of each learner. The degree to which the knowledge unit has been mastered by the learner can be inferred by the average scores of all courses containing the unit:

$$\varepsilon_x^i = \frac{1}{N_x} \sum\nolimits_{\forall c \in C} S(c)$$

where C is a set of all courses containing unit u_x (i.e., $C = \{c \mid u_x \in c\}$), N_x is the total number of courses containing the knowledge unit u_x (i.e., $N_x=|C|$), and $S(c)$ is a converting function which scales the score in different rating systems into a value between interval [0,1]. For example, it converses the score of 90 in 100 scoring system into the value of 0.9.

For the topics of courses, we mainly adopt those from the existing categorization for courses in Coursera (2015). In addition, users are required to specify their preferences on the categories in a 5-scale rating system (at least 5 categories in the experiment). The degree of preferences for course categories is therefore converted from 5-scale ratings.

As we mentioned, learning contexts seem to have a direct impact on the choice of courses to be accessed. To incorporate the context effects, the learning contexts are explicitly defined by a set of pre-defined contextual attributes and values as follows.

Definition 2: Let a_y is a pre-defined contextual attribute and v_y^b is the value of this contextual attribute under the learning context L_b, which is defined as a vector of attribute-value pairs as follows.

$$\vec{L}_b = (a_1 : v_1^b, a_2 : v_2^b, \ldots a_m : v_m^b)$$

Note that values for a contextual attribute are also pre-defined, e.g., a_y is "the week days" and v_y^b can be a specific element in the set of {"Monday", "Tuesday", ..., "Friday"}.

To mine contextual association rules, we adopt the threshold-based approach by setting a support value for mining. A contextual association rule is a mapping between a learning context and a course topic (i.e., $\vec{L}_b \to t_a$). If support value of a rule is greater than the pre-set threshold, we consider the rule to be a frequent contextual association rule. The threshold is set as 0.2 in this article. Finally, we estimate the factor of context effect. Specifically, we quantify the probability of a specific contextual association rule to a course topic in the set of frequent contextual association rules.

3.2 Personalized Course Search

The personalized course search is essential in calculating ranking scores for all courses by giving an issued query and a learner profile under a specific learning context. The learner profiles and contexts are presented in Definitions 1 and 2. For the representation of queries and courses, we also adopt a typical vector space model (VSM) to model both courses and queries, which are in the form of a bag-of-words. More specifically, queries are denoted as a vector of query terms assumed to have equal weight (Cai et al., 2010) as $\vec{q} = (w_1 : 1, w_2 : 1, \ldots w_s : 1)$, where w_1, w_2 ,…w_s are the query terms. Different from the query, the term features in courses have different weights. The definition of courses is given as follows.

Definition 3: Let $(w_1', w_2', \ldots w_r')$ be the terms relevant to course contents (e.g., in course titles and descriptions) and $(\mu_1^j : \mu_2^j, \ldots \mu_r^j)$ be the degree of relevance of each term to this course j, which is defined as a vector $\vec{C}_j$ of term-value pairs as follows.

$$\vec{C}_j = (w_1' : \mu_1^j, w_2' : \mu_2^j, \ldots w_r' : \mu_r^j)$$

where μ_r^j is the term frequency and inverse course frequency (TF-ICF) of the course, which is adapted from the term frequency and inverse document frequency (TF-IDF) (Baeza-Yates and Ribeiro-Neto, 1999). To achieve the goal of the personalized course search, we have taken both the learner preferences (learner profiles) and the learning contexts into consideration during the ranking processes. There are mainly three steps for the personalized course search.

(1) ***Learner Profile Contextualization***. According to previous research in context-aware information retrieval based on user profiles (Xie et al., 2012), the effects of a learning context can be interpreted as the rearrangement of the learning preferences in the learner profile. In other words, a learner may shift his preferences to courses in various contexts. Given a learning context $\vec{L}_b$ for learner *i*, the topic preferences T_i in the learner profile will be re-calculated in the two cases: (a) if the rule ($\vec{L}_b \rightarrow \iota_x$) is a frequent contextual rule, weights of categories in the learner profile which are re-calculated based on the probability of the rule ($r(\sigma_x^i)$=$p(\vec{L}_b \rightarrow \iota_x) \wedge \sigma_x^i$); (b) σ_x^i will be zero if there is no rule to be frequent to the course category. We notate the component of course topic preferences in the contextualized learner profile as T_i^*.

(2) ***Course Relevance Measurement.*** The relevance of each course is measured by two components, which are the course-to-query relevance and the course-to-profile relevance. The first component indicates how relevant a course content is to the issued query, and the second component refers to how relevant a course topic interest is to the learner. We employ the cosine similarity to measure the components course-to-query relevance as follows.

$$s_1(\vec{q},\vec{c}_j) = \cos(\vec{q},\vec{c}_j) = \frac{\vec{q}\cdot\vec{c}_j}{\|\vec{q}\|\|\vec{c}_j\|}$$

where $\vec{q}$ and $\vec{c}_j$ are the issued query and course *j*, which are in the form of vectors, the function $s_1(\ )$ denoting the ranking score of the first component. The second component (i.e., the course-to-profile relevance), indicating the degree of course topic interested by the learner, can be calculated by the mean of degrees of interest for those categories of the course.

$$s_2(c_j,T_i^*) = \sum\nolimits_{\forall t_k, c_j \in t_k} \frac{r(\sigma_k^i)}{K}$$

where T_i^* is the contextualized learner profile obtained in step (1), $r(\sigma_k^i)$ is the re-calculated course topic preference by the learner, and *K* is the total number of course topics as a course may incorporate multiple topics, e.g., the course "Introduction to Bioinformatics" can belong to both categories of "biology" and "computer science". Furthermore, we aggregated the ranking scores of two components and obtained the ranking scores as follows.

$$s = e^{s_1(\vec{q},\vec{c}_j)} \cdot e^{s_2(c_j,T_i^*)}$$

where *e* is the natural logarithm to smoothen and aggregate ranking scores of $s_1(\vec{q},\vec{c}_j)$ and $s_2(c_j,T_i^*)$. Courses are ranked in the search result lists based on the ranking scores.

(3) ***Knowledge-based Filtering.*** The objective of third stage is to eliminate those courses which can hardly be learnt by the learner due to his/her insufficient pre-knowledge. In this step, the component U_i in the learner profile can assist us to identify these courses. By comparing the knowledge level a threshold ε_x^* for the degree of required knowledge unit *x* to learn a course, we can remove those courses which are hardly learnt by the learner from the search result list.

4. Experiment

4.1 Experimental Setting

To evaluate the performance of the proposed approach for the personalized course search, a prototype system is developed for the experiment. We have adopted the similar user interface as Coursera (2015) and crawled 135 courses in 8 categories (including computer science, biology and life science, education, etc) from Coursera (2015). We have extracted course titles, introductions and brief descriptions to build the course feature vectors. Generally, the number of courses in each category is from 14 to 27.For the participants, 19 undergraduates from diverse programmes in a university were involved in the experiment. There are 12, 6 and 1 subjects are aged from 17-19, 19-21 and 21-23, respectively. Most subjects (15) are come from Hong Kong, and 11 of them are female.

4.2 Metric & Baseline

Precision@N (*P@N*) is used in our experiments for evaluation. *P@N* mainly reflects the accuracy of the proposed personalized search results. Formally, it can be calculated as follows.

$$P@N = \sum\nolimits_{q\in Q} \frac{p(q)}{|Q|} = \sum\nolimits_{q\in Q} \frac{n_q}{N\cdot|Q|}$$

where q is a query from the set of all queries (Q), and $p(q)$ is the accurate rate for each query, which is based on the portion of relevant courses (n_q) in top-N results. The larger value of *P@N* indicates more accurate of the personalized search approach.

To validate the soundness and effectiveness of the proposed methods, we compare it with several state-of-the-art baseline methods as follows.

- ***BASIC.*** The basic method is a non-personalized approach, which only uses the basic vector space between query vectors and course vectors (Gudivada et al., 1997).
- ***COS.*** The cosine method (notated as *COS*) employs the vector space model and cosine similarity among the user (learner) profiles, queries and items (courses) as proposed by Nolland Meinel (2007). Although the method is a personalized search method by exploiting user profiles, the search contexts and user knowledge are not taken into account.
- ***BM25.*** The best match 25 (BM 25) in personalized search is proposed by Vallet et al. (2010). The idea is similar to *COS*, which measures the relevance among the user profiles, query and items. The difference is that the best match 25 paradigm is adopted for profile construction.

4.3 Experimental Results

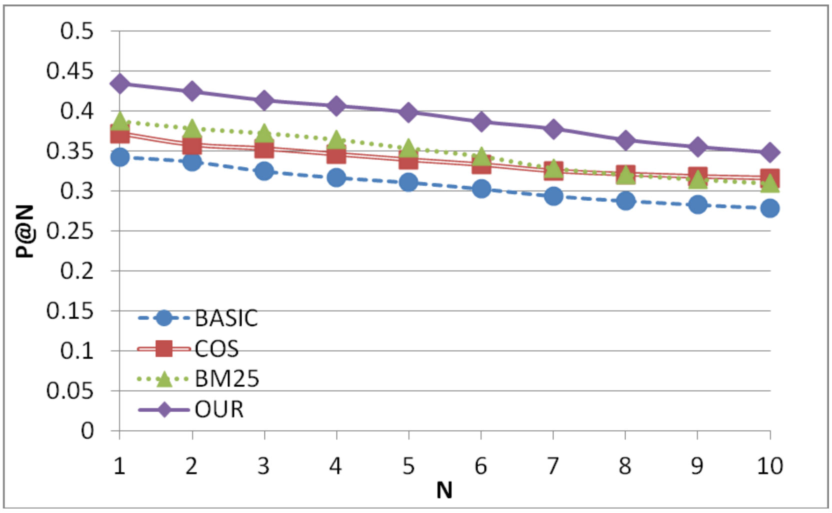

Figure 1. The metric P@N by all five approaches.

In the experiments, we explicitly extract data from historical transcripts from students. As participants are in their first or second year of study, there are only limited historical courses which can be easily converted to knowledge unit levels manually. For the query and relevance judgment, we ask each participant to generate more than 10 queries and marked the relevant courses on the result list merged from top-10 results by all four approaches (ours and three baselines). The performances of *P@N* of all five approaches are shown in Figure 1. We can observe that *P@N* of all approaches decreases with the increase of *N*. This is mainly because the relevant courses judged by participants are quite limited for each query. When *N* becomes larger, *P@N* normally will be dropped down. Furthermore, we can find that our method achieves the best performance in *P@N* (*P@1*=0.435 *and P@10*=0.348), BM25, COS and BASIC has the second, third and fourth *P@N*, respectively. The result is consistent with the finding in the previous research (Cai et al. 2010). We have also performed student t-test between each pair of methods and verify that all improvements are significantly ($p<0.05$). In addition, the experimental result reveals that integration of learner profiles (including pre-knowledge levels and learning preferences) and learning contexts is a reasonable and effective framework than those frameworks (baselines) without them. More specifically, both BM25 and COS neglect either knowledge levels or learning contexts, while BASIC not only ignores knowledge and contexts but also the learner profiles.

5. Conclusion

In this paper, we focus on the issue of how to assist learner to search their interested courses efficiently and accurately. We explicitly propose the models of learning preferences, pre-knowledge levels and learning contexts, so that a context-aware personalized course search method based on learner profile is facilitated. By conducted experiments on real 19 learners who compare the search result of the proposed method to those of four baselines, the accuracy and efficiency are verified. In the future, we plan to investigate the problem of how to automatically identify the required knowledge for a course.

Acknowledgements

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (UGC/FDS11/E06/14).

References

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). Modern information retrieval (Vol. 463). New York: ACM press.

Cai, Y., Li, Q., Xie, H., & Yu, L. (2010). Personalized resource search by tag-based user profile and resource profile. In Web Information Systems Engineering–WISE 2010 (pp. 510-523). Springer Berlin Heidelberg.

Chen, C. M., Hsieh, Y. L., & Hsu, S. H. (2007). Mining learner profile utilizing association rule for web-based learning diagnosis. Expert Systems with Applications, 33(1), 6-22.

Coursera Categorization List. (2015, April 15). Retrieved from http://www.coursera.org/courses

Essalmi, F., Ayed, L. J. B., Jemni, M., & Graf, S. (2015). Generalized metrics for the analysis of E-learning personalization strategies. Computers in Human Behavior, 48, 310-322.

Feng, X., Peng, Y., Xie, H., & Yan, Z. (2011, July). Role-Based Learning Path Discovery for Collaborative Business Environment. In 2011 International Conference on Control, Automation and Systems Engineering (CASE), (pp. 1-4). IEEE.

Gudivada, V. N., Raghavan, V. V., Grosky, W. I., & Kasanagottu, R. (1997). Information retrieval on the world wide web. IEEE Internet Computing, 1(5), 58-68.

Kurilovas, E., Kubilinskiene, S., & Dagiene, V. (2014). Web 3.0–Based personalisation of learning objects in virtual learning environments. Computers in Human Behavior, 30, 654-662.

Leung, E. W. C., & Li, Q. (2007). An experimental study of a personalized learning environment through open-source software tools. Education, IEEE Transactions on, 50(4), 331-337.

Limongelli, C., Miola, A., Sciarrone, F., & Temperini, M. (2012, July). Supporting teachers to retrieve and select learning objects for personalized courses in the Moodle_LS environment. In Advanced Learning Technologies (ICALT), 2012 IEEE 12th International Conference on (pp. 518-520). IEEE.

Noll, M. G., & Meinel, C. (2007, November). Web search personalization via social bookmarking and tagging. In Proceedings of the 6th international The semantic web and 2nd Asian conference on Asian semantic web conference (pp. 367-380). Springer-Verlag.

Özpolat, E., & Akar, G. B. (2009). Automatic detection of learning styles for an e-learning system. Computers & Education, 53(2), 355-367.

Schiaffino, S., Garcia, P., & Amandi, A. (2008). eTeacher: Providing personalized assistance to e-learning students. Computers & Education, 51(4), 1744-1754.

Vallet, D., Cantador, I., & Jose, J. M. (2010). Personalizing web search with folksonomy-based user and document profiles. In Advances in Information Retrieval (pp. 420-431). Springer Berlin Heidelberg.

Xie, H., Li, Q., & Mao, X. (2012). Context-aware personalized search based on user and resource profiles in folksonomies. In Web Technologies and Applications (pp. 97-108). Springer Berlin Heidelberg.

Yang, J. (2013). A study on online learner profile for supporting personalized learning. Knowledge Management & E-Learning: An International Journal (KM&EL), 5(3), 315-322.

Zou, D., Xie, H., Li, Q., Wang, F. L., & Chen, W. (2014). The Load-Based Learner Profile for Incidental Word Learning Task Generation. In Advances in Web-Based Learning–ICWL 2014 (pp. 190-200). Springer International Publishing.