A Heuristic Approach for New-item Cold Start Problem in Recommendation of Micro Open Education Resources

Geng Sun[1], Tingru Cui[1], Dongming Xu[2], Jun Shen[1], Shiping Chen[3]

[1] School of Computing and Information Technology, University of Wollongong, Australia
[2] UQ Business School, The University of Queensland, Australia
[3] CSIRO Data61, Australia
gsun@uow.edu.au, tingru@uow.edu.au, D.Xu@business.uq.edu.au, jshen@uow.edu.au, Shiping.Chen@data61.csiro.au

Abstract. The recommendation of micro Open Education Resources (OERs) suffers from the new-item cold start problem because little is known about the continuously published micro OERs. This paper provides a heuristic approach to inserting newly published micro OERs into established learning paths, to enhance the possibilities of new items to be discovered and appear in the recommendation lists. It considers the accumulation and attenuation of user interests and conform with the demand of fast response in online computation. Performance of this approach has been proved by empirical studies.

Keywords: Cold Start, Open Education Resources, Adaptive Micro learning, Heuristic Recommendation, Learning Path

1   Introduction

Along with various leading universities opening up access to their courses, open education resources (OERs) are becoming increasingly available. The rise of OERs gains large popularity in the entire higher and adult education sector, and these emerging learning paradigms have attracted many researchers' attention, from educational, social, and computational views [1]. However, because of the newness of the phenomenal education trend, its popularity has been limited due to the lack of personalized services so that current OER delivery often fails to meet comparatively diverse demands from both OER providers and learners. Moreover, these OERs are suggested to be consumed in a micro learning mode, which conforms to the characteristic of the modern e-society where mobile and pervasive computing becomes dominant.

A service-oriented system, Micro Learning as a Service (MLaaS), aims to deliver personalized OER with micro learning to satisfy learners' personal demands in real time. It customizes adaptive micro learning contents as well as provides learning path identifications tailored for each individual learner. MLaaS consists of an offline computation and an online computation domain to provide recommendations jointly to improve computation performance and respond in the granularity of seconds. In this paper we will introduce a heuristic approach to overcome the new-item cold start problem in micro OER recommendation, which will be realized by the online computation of MLaaS.

2   Background

2.1   Micro Learning

Compared to traditional learning modes, now learners' overall efforts to go through an entire concept (or learning objective) will proceed in an intermittent way rather than a consecutive way. Hence, micro learning through OER (i.e. micro open learning) is becoming mainstream for next generation learners, who learn on the move, with easy access to the 'cloud' or Internet of Things [2]. Generally micro learning refers to short-term learning processes, which contains knowledge or learning content in small units. Typically, a micro learning activity is carried out through mobile devices within a time frame of 15 minutes [3]. As an emerging educational phenomenon, micro learning is more user-centric and it requires different learning schedules from on-campus learning or even standard e-learning and m-learning [4].

2.2   Micro Learning as a Service

Basically, the offline computation of MLaaS runs on a basis of compound transactions. When a learning activity launches, a compound transaction is generated associated with it, which can be therefore represented as: *{Learner Profile, Micro*

OER profile, Association}, where a learner profile and an OER profile are involved, linked by an association showing the learner's properties against the micro OER.

The offline computation is responsible to transform implicit user behaviors into explicit information [5]. A knowledge base is in charge of the semantic construction and storage of the learner profile and OER representation [6]. The semantic construction of learner profiles enables MLaaS to consider individual learning styles, learner's context, application capabilities, and teaching materials structure, leading to a customization of the type and delivery format of learning information in response to the user. Similarly, an augmented micro OER ontology is also built [7].

The authors of [7] have proposed a comprehensive learner model which involves features that can impact and constrain the micro learning experiences and outcomes, and is enclosed in an ontological representation [8]. By taking advantage of the comprehensive learner model, the LearnerProfile can be broken down to: *{InternalFactors, ExternalFactors} = {IntellectualFactors & NonIntellectualFactors, ExternalFactors}*, where the internal factors can be classified into personal intellectual and non-intellectual factors, differentiated by whether a factor is related to a learner's cognitive and intelligence level or not. External factors come from the environmental and social-economical contexts.

Given the OER delivery has a 'big data' context, ideally there could be sufficient data sets to be used in data modeling and machine learning, and rule mining is compulsory to impute unknown values for online-prediction. The rules can be represented as:*{LearnerProfile, MicroOERProfile}→{Association}.*Technically rule mining and learner clustering operation runs throughout all the offline computation process.

As the core of MLaaS, the Adaptive Engine processes the results from all other services and transmits its output to the user interfaces straightforward.

However, the initial MLaaS system has insufficient information about the learners as well as new OERs without existing ratings, which leads to infeasible profile construction. This places the cold start problem as the central challenge of micro OER delivery. Consequently, MLaaS's online computation is mainly in charge of the cold start problem, and also making up limitation in timeliness and renewal of the offline computation by retrieving real-time usage and keeping the comprehensive learner model and leaner-micro OER profiles up-to-date.

In [9] a solution to deal with the new-user cold start problem has been proposed. A top-N recommendation is adopted to provide learners a range of options to kick off their micro open learning journey. It takes a rule-based heuristic approach to generate candidate learning paths, and then the first micro OER in each learning paths is picked up to form the recommendation list. In this paper we will carry on the efforts put in the new-user cold start problem and follow the rules to come up with a new heuristic for the new-item cold start problem, namely inserting newly published micro OERS into well-established learning paths.
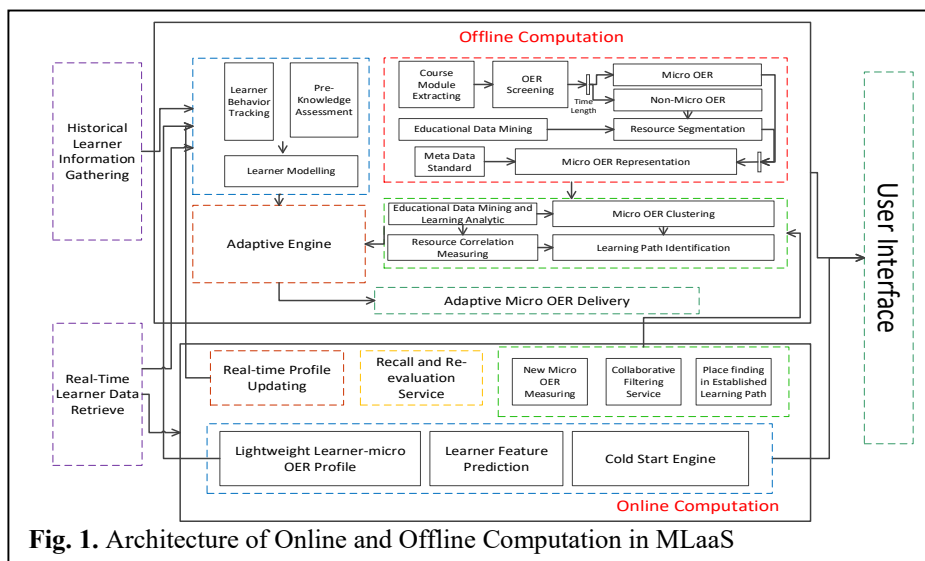
Fig. 1. Architecture of Online and Offline Computation in MLaaS

2.3 Research Problem

The cold start problem is generally triggered by three factors: new community, new item and new user. Both open learning and micro learning are comparatively novel to many people in the e-society. The followers of this novel trend, no matter new education pursuers or regular learners migrated from other online learning modes, newly join into this emerging community of e-society. Meanwhile, MLaaS faces the new-item cold start which is of great practical importance. This is because new OERs are kept publishing day after day, and effectively recommending them is essential for keeping the users continuously engaged[10]. For newly published micro OERs, it is central to their acceptance and popularity that it can be discovered and appear in recommendation list as soon as being released.

The learning demands and expectations of learners engaged in open learning are much more practical than conventional university students. They are mostly self-regulated so that it is totally flexible for them to decide when to join or quit the online course at their own willingness, and switch among courses frequently [11].

In addition, in micro open learning, the accumulation and attenuation of user interests and demands can be periodical and may vary in different patterns than other online activities [12]. It is very likely to see that a learner accesses OERs covering similar knowledge are-as again (and again) offered by different educational institutions. This cross-learning phenomenon can be attributed to purposes of reviewing or mutual supplementation, by comparing the ways of knowledge imparting as well as learning two or more micro OERs simultaneously [11-12].

3 Semantic Representation of Micro OERs and Learners

3.1 Investigation of Micro OER Correlation

From the item-based perspective, the general ontology of OER is augmented to adapt the needs of micro learning, in which an annotation of a micro OER is self-describing with metadata exploring its educational parameters, such as typology, type of interaction, didactic model, and non-functional attributes. Each node in the augmented OER ontology indicates a micro OER chunk[7]. There is no completely independent chunk and each of them is part of a relational web rather than merely a conceptual object. This ontology explicitly classifies the OERs to recommend among a pedagogically defined set of distinctive main concepts, fed as the raw material in the reasoning process of MLaaS [13].

The investigation over 'big' open learning data comes up to the OER side. Among the massive OERs, three types of relations are mainly targeted:

- ConsistsOf is an inclusion relation. This relation can be generally found between two OERs or one OER and one micro OER. Two items with this relation are located in different hierarchies of the augmented micro OER ontology, the ancestor at the top and the descendant at the bottom. Let R denote an *OER* and MR denote a micro OER, and ConsistsOf(R_a, MR_b), or $MR_b \in R_a$ indicates that the original OER R_b （ancestor） is segmented into several micro OERs (descendants), and MR_b is one of them. Certainly, as a micro OER can be further subdivided, there can be ConsistsOf(MR_b, MR_c), or $MR_c \in MR_b$ provided that MR_c is a tinier micro OER derived from MR_b.
- RequiredSequence is a strong order between two items (OER or micro OER), where the former micro OER is necessary to be learnt before the latter one, due to course setting and educational consideration.
- RecommendedSequence is a weak order relation between two items (OER and micro OER), where the former micro OER is suggestive to be learnt before the latter one, according to the instructors' guidance, but it is not mandatory.
- Certainly, two items (OER or micro OER) can have no relation at all.

Both relations regarding sequence can be inherited by entities' descendants. For example, when there is a RecommendedSequence (R_1, R_2) indicating an OER R_1 is preferably learnt prior to R_2, then, for $MR_1 \in R_1$ and $MR_2 \in R_2$, there is a RecommendedSequence (MR_1, MR_2). The inheritance is also valid if the ConsistsOf relation is between two micro OERs.

3.2 Lightweight Learner-Micro OER profile

Motivated by the cold start condition, the comprehensive learner model was simplified to a lightweight learner-micro OER profile [7]. It merely deals with necessary information for decision making in order to act on the initialization agilely.

The lightweight learner profile is managed by MLaaS with a static part and a dynamic part. The static part can be represented by a vector, which contains the demographic and educational information. By matching these two augmented

4

ontologies, for item and user respectively, the dynamic part of a learner node is denoted as a pair, $L_j = \{MR_u, ML_j\}$, $L_j \in L$. Herein, the element MR_u denotes the uth micro OER, which is a particular version of the micro OER ontology, as introduced in the previous subsection, and a three-dimensional element $ML_j \{P_{u,j}, TA_j, D_j\}$ is exclusive to j^{th} learner during the micro learning process. Herein, the element $P_{u,j}$ indicates the learner's preference, TA_j indicates the jth learner's instant time availability, and D_j denotes the level of distraction in terms of the given learning environment and surroundings. Each of these three features proposed in the lightweight profile is associated with a confidence degree to reflect its subjective relevance. Whenever MLaaS gathers any information from the learner's learning process over OER, the learner profile will be updated regarding ML_j.

4   Insertion of New Micro OERs into Established Learning Path

4.1   Micro OER Screening and Rules

For each existing micro OER, once MLaaS has acquired its final preference value and confidence degree, those nodes, which do not meet the minimum requirement of confidence degree, is rejected by the system. When generating a list of recommended micro OERs, the ones with higher learners' interests are placed at the top. For two micro OERs MR_u and MR_w, their sequence is determined according to some heuristic rules which are defined in accordance with the extraction of three types of relations discussed in the Section 3.1. These rules are executed sequentially with priority. Herein, the first rule is deemed as a hard rule which should be strictly obeyed and the rest rules are soft rules which can be violated with educational consideration, from case to case.

1. If there is a RequiredSequence relation between these two micro OERs, the prerequisite one is placed above (refer to the Section 3).
2. If the preference regarding these two OERs, $P_{u,j}$, $P_{w,j}$, the former one is higher than the latter one, then the MR_u is above MR_w
3. If, in the absolute terms, the confidence degree $CD(P_{u,j})$ is high and the $CD(P_{w,j})$ is low, then the MR_u is above MR_w.
4. If there is a RecommendedSequence relation between these two micro OERs, the one which is suggested to be accessed first is placed above (refer to the Section 4).
5. The micro OER, which is more related to the learners' education background, or falls in the relevant disciplines or interdisciplines is placed with priority if the disciplinary difference between this two candidate micro OERs is obvious.
6. Otherwise the recommended micro OER list is randomly ordered if none of the above rules applies.

4.2   Learning Path Establishment

Candidate learning path solutions (chromosomes) are randomly generated where each of them is a learning path with a series of micro OERs, rather than an individual micro OER. For a chromosome, its violation degree is investigated by examining the relations between each contiguously prior/posterior micro OER pair against the first 5 rules listed in the previous Section 4.1, and then summing up. For such pair in a chromosome, its violation degree, $VD(MR^t, MR^{t+1})$, is calculated by the weighted sum of its violations against rule 2 to rule 5, respectively, where MR^t is the tth micro OER in k and MR^{t+1} is the $t+1^{th}$. The higher the violation degree is, the more serious the candidate's learning path violates the rules. The violation degree of a candidate learning path, k, is calculated using the following equation (1)

$$VD_k = \sum VD(MR^t, MR^{t+1}) \quad (1)$$

A given micro learning resource MR_u's real-time suitability for micro learning, $RT_{u,j}$, is calculated in a previous work [9]. We borrow the definition of this variable. Hence, for the candidate learning path, k, $RT_{k,j}$ denotes the sum of the real-time suitability of micro OERs it contains. Similarly, $P_{k,j}$ sums up all the predicted preferences from the learner L_j versus micro OERs that k contains.

$$\eta = \min(\alpha VD_k + \beta RT_{k,j} + \gamma / P^1_{k,j} + \delta / P_{k,j}) \quad (2)$$

where α，β，γ and δ serve as weight for each variable and suggestively $\alpha > \beta > \gamma > \delta$, $P^j_{k,j}$ denotes the L_k's preference value of the first micro OER in k. A heuristic algorithm in [9] infers a few optimized learning paths, by minimizing the fitness of each. A recommendation list is generated with a size of N, namely N learning paths with lowest finesses are selected and

their first micro OERs are placed in the recommendation list. Once the learner makes his/her first option micro open learning, the following items in the same learning path will be adapted to him/her sequentially.

4.3 Optimization of Micro OER Similarity Calculation

Hereby, the similarity calculation among micro OERs is crucial to the quality of item-based collaborative filtering approach. Using the equation (3), this calculation is not only based on their Euclidean distance on educational settings, but also added a time decay factor, which considers accumulation and attenuation of interest, and a penalty term, which tackles the filter bubbles. These two operators are shown as the latter multipliers in equation (3).

$$sim^M(n,g) = \sum (|n,g|) * N_0 e^{-\lambda(t_1-t_2)} * \frac{1}{\log_a(O_j+c)} \tag{3}$$

where the t_1 is the current time and t_2 is the time when the existing micro OER, MR_g, was accessed. O_j refers to the times of a specific learner, L_j's operation, retrieved from the real-time MLaaS usage (as stated in Section 4). The constant c keeps the denominator unequal to zero.

The L_j's preference values are selected and evaluated to obtain their mean. A K nearest neighbor (KNN) algorithm is able to cluster items with higher similarities with the new micro OER, MR_n. Its neighbors form as a set, G. Consequently, the prediction for the L_j's preference against a new micro OER MR_n, is calculated by the equation (4):

$$P_{n,j} = \overline{P_j} + \frac{\sum_{g\in G} sim^M(MR_n,MR_g)*(P_{g,j}-\overline{P_j})}{\sum_{g\in G} sim^M(MR_n,MR_g)} \tag{4}$$

4.4 Inserting New Micro OERs into Established Learning Path

The new items will be inserted into established learning paths according to the Algorithm 1:

Algorithm 1: Insert New OER into Established Learning Path

Input: MR_n (a newly published micro OER), the set of existing micro OERs, Established learning paths, Semantic Relationships of OERs (as defined in the Section 5)
Output: Optimized learning paths which contain the MR_n inserted at the suitable position

---

begin: Investigate the MR_n in terms of measurements in the Section 3.
    Measure its similarities with existing micro OERs using eq(3)
    Use KNN algorithm to cluster MR_n's neighbors
    //Let **G** denote the set of neighbors of MR_n
    MR_n's semantic relationships among neighbors in G are firstly examined according to the standard provided in Section 3.
    Invoke established learning paths or use the Algorithm in [9] to produce candidate learning paths

    for each MR_g ∈G **do**
      select learning paths that contain MR_g
      cut a segment that contain MR_g and few micro OERs prior/posterior to it in each learning path
    //find rough positions where the MR_n might be located at
        for each segment **do**
      MR_n's semantic relations among micro OERs in the few places are examined again
          if there is a 'RequiredSequence'
          locate the place for MR_n
            interposition it between the micro OERs according to this strong order (as in the Section 3)
          end if
          if there is a 'RecommendedSequence'
          put MR_n among the micro OERs according to this weak order
            or
          put MR_n in parallel with one of existing micro OERs alternatively

```
                    compare the predicted preference values of MRₙ and all ex-
                    isting micro OERs' in this segment
                    apply the rule 2 in the Section 4.1
                measure the fitness of new learning paths using eq(2)
                use Hill Climbing algorithm to compare fitness
                replace an existing OER in established learning path with
                worse fitness
                or
                insert MRₙ between two existing micro OERs, keep MRₙ added on
                as extra if the fitness is satisfactory //in this case the
                overall quantity of items in the new learning path is in-
                creased
                end if
            end for
        end for
    generate new learning paths containing MRₙ
    end
```

A Hill Climbing algorithm effectively abandons the learning paths with poor fitness by searching locally and reduces the times of iterations [14]. This approach does not examine throughout all elements in all matrices, hence its computing complexities and running time are acceptable for online computation which is in demand of fast response.


5   Evaluation

In this section we evaluate the qualities of generated learning path with newly published micro OERs inserted. We borrow the concept of 10-cross validation, by dividing the micro OERs in the relevant fields into two portions, in a ratio of 1:9. Learning paths were generated among the nine-tenth micro OERs. For each learning path, one micro OER from the rest is selected and treated as a newly published micro OER in the experiment. In total, 3674 micro OERs in the information technology field, 4479 micro OERs in the business field and 3254 in the social science field are picked up as candidates; and 366, 448 and 325 of them were selected out as test items, respectively. The Algorithm 1 is executed to find a place for each new member to the majority.

The Fig. 2 gives the violation degree (i.e. *VD* as defined in the Section 4.2) for the learning paths with new micro OERs inserted generated by using MLaaS approach against those generated by using the shortest-path approach [15] and competency-based approach [16]. In the information technology field 2044 new learning paths come out, while in the business and social science fields the numbers are 3746 and 2329. This is because one or more places are found for a new micro OERs; or according to the Algorithm 1, two new learning paths are generated when there is a 'RecommendedSequence' relation. The shortest-path (SP) approach and competency-based (C-based) approach are executed as well to put newly published micro OERs into places among or in parallel with items in established learning paths. Actually, the working principles of SP and C-based are not finding a place for the newly published micro OER within the established learning path, while rebuilding a new learning path thoroughly. According to the Fig. 2, the average violation degrees of the learning paths generated separately by the three approaches are compared in terms of the three disciplines. It can be found that MLaaS approach outperforms SP and C-based approach overall, as average *VD*s of MLaaS-generated learning paths in each discipline are far less than the others. Also, the SP approach is difficult to identify a reasonably learning path provided that there are many micro OERs loosely correlated (i.e. with the weak order RecommendedSequence).

Afterwards, we evaluate the best solutions that are produced by MLaaS, SP and C-based. Therein, one test micro OER is only allowed to be involved in one learning path; and for each test object, the new learning path with lowest *VD* value is selected. This is to eliminate the potential influences brought by the loose coupling of prior/posterior micro OER pairs. In this case, the amount of the nominated learning paths exactly equals to the amount of test micro OERs, namely 366, 448 and 325, respectively.
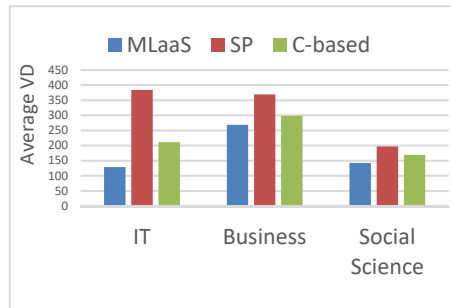
Fig. 2. Performance Comparison – Average VDs of All Generated Learning Paths
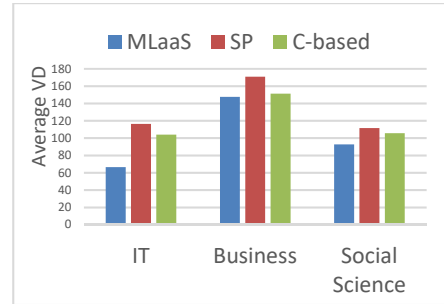
Fig. 3. Performance Comparison – Average VDs of the Best Learning Paths for Each Test Micro OER

The results in the Fig. 3 show same observation that the MLaaS approach surpasses the other two approaches, by finding better places to insert newly published micro OERs meanwhile breaking the rules (as described in Section 3) less times. It is worth noting that the average violation difference is considerably larger in the information technology discipline. This is probably because the learning paths in this field are relatively longer. In other words, an individual learning path that goes through a complete information technology knowledge area consists of more micro OERs than those of business and social science. In addition, all of the three approaches produce learning paths with higher violation degrees in the business discipline. It can be potentially attributed to that micro OERs having other OERs as knowledge pre-requirement, and one micro OER being closely related to more than one of the others, are more often found in the business discipline.

6   Conclusion

In this paper we have introduced a heuristic approach to overcome the new-item cold start problem in micro OER recommendation. The newly published micro OERs are evaluated using an optimized similarity, by considering the accumulation and attenuation of interest and filter bubbles. Experiment results have proved that the proposed approach can generate learning paths with higher conformity with heuristic rules, hence finding more appropriate places in established learning paths for new items.

References
1. J. Hylen, D.V. Damme, F. Mulder and S. D' Antoni, 'Open Educational Resources: Analysis of Responses to the OECD Country Questionnaire', OECD Education Working Papers No. 76, June 2012.

8

2. M. I. Souza and S. F. D. Amaral, 'Educational Micro Content for Mobile Learning Virtual Environments', Creative Education, vol. 5, pp. 672-681, 2014

3. D. Kovachev, Y. Cao, R. Klamma and M. Jarke, 'Learn-as-you-go, New Ways of Cloud Based Micro-learning for the Mobile Web', 10th International Conference on Web-based Learning, Hongkong, December 2011.

4. P. A. Bruck, L. Motiwalla and F. Foerster, 'Mobile Learning with Micro-content: A Framework and Evaluation', 25th Bled eConference, Bled, Slovenia, 2012, pp.527-542.

5. Z. Sun, G. Guo and J. Zhang, 'Exploiting Implicit Item Relationships for Recommender Systems', 23rd International Conference on User Modeling, Adaptation and Personalization, Dublin, Ireland, June 2015.

6. G. Sun, T. Cui, G. Beydoun, S. Chen, F. Dong, D. Xu and J. Shen, 'Towards Massive Data and Sparse Data in Adaptive Micro Open Educational Resource Recommendation: A Study on Semantic Knowledge Base Construction and Cold Start Problem', Sustainability, vol.9, no.6, pp. 898.1-898.21, 2017.

7. G. Sun, T. Cui, W. Guo, G. Beydoun, D. Xu and J. Shen, 'Micro Learning Adaptation in MOOC: a Software as a Service and a Personalized Learner Model', the 14th International Conference on Web-based Learning (ICWL), Guangzhou, China, November 2015, pp.174-184.

8. G. Sun, T. Cui, J. Shen, D. Xu, G. Beydoun and S. Chen, 'Ontological Learner Profile Identification for Cold Start Problem in Micro Learning Resources Delivery', the 17th IEEE International Conference on Advanced Learning Technologies (ICALT), Timisoara, Romania, 2017, pp.16-20

9. G. Sun, T. Cui, D. Xu, H. Chen, S. Chen and J. Shen, 'Assisting Open Education Resource Providers and Instructors to Deal with Cold Start Problem in Adaptive Micro Learning: a Service Oriented Solution', the 14th IEEE International Conference on Services Computing (SCC), Hawaii, USA, June 2017, pp. 196-203.

10. B. Lika, K. Kolomvatsos, S. Hadjiefthymiades, 'Facing the Cold Start Problem in Recommender Systems', Expert Systems with Applications, vol. 41, no. 4, pp. 2065-2073, 2014.

11. S. Miranda, G. R. Mangione, F. Orciuoli, M. Gaeta and V. Loia, 'Automatic Generation of Assessment Objects and Remedial Works for MOOCs', 12th International Conference on Information Technology Based Higher Education and Training, Antalya, Turkey, October 2013.

12. I. Nawrot and A. Doucet, 'Building Engagement for MOOC Students', Introducing Support for Time Management on Online Learning Platforms', Proceeding of WWW'14 Companion, 2014.

13. A. Moreno, A. Valls, D. Isern, L. Marin and J. Borras, 'SigTur/E-Destination: Ontology-Based Personalized Recommendation of Tourism and Leisure Activities', Engineering Applications of Artificial Intelligence, vol.26, no.1, pp.633-651, 2013.

14. V. Tam, E. Y. Lam and S.T. Fung, 'a New Framework of Concept Clustering and Learning Path Optimization to Develop the Next-Generation e-Learning Systems', Journal of Computers in Education, vol.1, no.4, pp.335-352, 2014.

15. M. Alian. M and R. Jabri, 'a Shortest Adaptive Learning Path in eLearning Systems: Mathematical View', Journal of American Science, vol. 5, no. 6, pp. 32-42, 2009.

16. W-C. Hsu and C-H. Li, 'a Competency-based Guided-Learning Algorithm Applied on Adaptively Guiding e-Learning, Interactive Learning Environments', vol. 23, no. 1, pp. 106-125, 2015.