Lin, J., Sun, G., Beydoun, G., & Li, L. (2022). Applying Machine Translation and Language Modelling Strategies for the Recommendation Task of Micro Learning Service. *Educational Technology & Society*, *25* (1), 205-212.

Applying Machine Translation and Language Modelling Strategies for the Recommendation Task of Micro Learning Service

Jiayin Lin[1*], Geng Sun[1], Ghassan Beydoun[2] and Li Li[3]

[1]University of Wollongong, Wollongong, NSW, Australia // [2]University of Technology Sydney, Sydney, NSW, Australia // [3]Southwest University, Chongqing, China // jl461@uowmail.edu.au // gsun@uow.edu.au // ghassan.beydoun@uts.edu.au // lily@swu.edu.cn

[*]Corresponding author

ABSTRACT: A newly emerged micro learning service offers a flexible formal, informal, or non-formal online learning opportunity to worldwide users with different backgrounds in real-time. With the assist of big data technology and cloud computing service, online learners can access tremendous fine-grained learning resources through micro learning service. However, big data also causes serious information overload during online learning activities. Hence, an intelligent recommender system is required to filter out not-suitable learning resources and pick the one that matches the learner's learning requirement and academic background. From the perspective of natural language processing (NLP), this study proposed a novel recommender system that utilises machine translation and language modelling. The proposed model aims to overcome the defects of conventional recommender systems and further enhance distinguish ability of the recommender system for different learning resources.

Keywords: Information filtering, Recommender system, Micro learning, Big data, Natural language processing

1. Introduction

The achievements in 5G Internet and mobile devices boost the real-time multi-media interaction in various applications such as commercial, entertainment, and online education. In the meantime, the fast-paced modern life and booming of knowledge in the big data era drive people to seek a more flexible way to acquire knowledge or carry out personalised learning activities. All the above factors gave birth to micro learning service (Sun et al., 2015b), aiming to utilize user's daily fragmented spare time and assist the learner in conducting self-regulated personalized learning activities. The term "micro" used in this study refers to the small (micro) personalised chunks of learning materials containing a small volume of knowledge. As pointed out in the study (Syeda-Mahmood & Ponceleon, 2001), users are less likely to leave out the knowledge points for a short learning session, such as a short video. And the engagement of an online learning activity plunges quickly after 7 minutes (Anderson et al., 2014; Guo et al., 2014). With the advantage of Internet technology, massive learning materials are uploaded to the Internet every day in various disciplines, format, and difficulty levels. Hence, a serious information overload problem challenges the learner experience of the micro learning service. Hence, filtering out irrelevant information and picking the one that matches the learner's learning requirement is the key to such a personalized online learning service.

Especially for the online service/application that deploys in the context of big data, a sophisticated recommender system is a key factor to guarantee efficiency and personalization. Even information filtering and retrieval were classified into two different research disciplines, the boundary between information filtering (i.e., the main function of recommender system) and information retrieval (i.e., the main function of search engine) is relatively vague. The former one aims to find irrelevant resources and filter them out, and the recommenders can be further classified into three categories (Wasid & Ali, 2017): content-based filtering (CB), collaborative filtering (CF), and hybrid recommending strategy. The latter one aims to find relevant resources and rank them based on their relevant degrees. In General, both of them try to distinguish relevant and irrelevant information (Belkin & Croft, 1992). Due to pedagogical issues (Lin et al., 2020), ranking the recommended learning materials is significant to delivering a suitable learning resource to the learner. However, as discussed in (Valcarce, 2015), there is little research about applying information retrieval (IR) ideas to boost the performance of a recommender system (RS). From the perspective of IR, the probabilistic model has a solid statistical foundation. Hence, as discussed in (Belkin & Croft, 1992) that the probabilistic model can make significant improvements to the field of recommender system as it did in IR.

Based on one previous work (Lin, 2020), in this paper, we further refine the proposed recommender system, which can precisely rank the recommended learning materials. This system makes use of the mathematical concepts of machine translation and language modelling to model the learning materials and reflect the mapping relationships between historical learning records and new learning materials. The remainder of the paper is organised as follows. Section 2 will discuss challenges in the recommendation task of the micro learning service.

The related work in this research will be discussed in Section 3. The proposed model will be introduced in Section 4. We introduce and explain the evaluation of the proposed recommendation strategy in Section 5. The conclusions and future work of this study is discussed in Section 5.

2. Challenges in recommendation for micro learning

2.1. The drawbacks of conventional recommendation models

As discussed before, the algebraic-based recommender system such as collaborative filtering and matrix factorization has been demonstrated to be very effective in filtering out irrelevant online resources, but such a model lacks the ability to precisely distinguish the difference between the remained resources. Most of the system used algebraic-based strategy can only predict the rating value of the resources but cannot provide any detail information of the resource with the same rating value. One study proved that the algebraic-based collaborative filtering cannot generally provide good result in the top-k recommendation task (Valcarce, 2015). Hence, the authors of this study argued that the probabilistic method could be a more effective and formal way for generating personalized rankings of recommendations (Valcarce, 2015). And in the study of (Koren et al., 2009), the researchers demonstrated that matrix factorization based recommender systems is guided by the rating value and does not involve any explicit features, which could not represent the ranking information among items.

2.2. Micro learning service and recommendation in e-learning

Most studies on recommenders found in the e-learning field were focused on the suitability of learning materials against learners' personalization. Formally, a micro learning activity is carried out within a time span of 15 minutes through a mobile device (typically, though). One pilot work investigated the possibility of customising open educational resources (OERs) to meet the demand of microlearning (Sun et al., 2015b). And another work provided a comprehensive learner model oriented to micro learning through OERs in (Sun et al., 2015a). In (Sun et al., 2018) the researchers discussed the mainstream typology (video, audio, text), type of interaction (expositive, active, mixed, two-way), didactic model (e.g., inductive, deductive, learning by doing) of the online learning materials, in particular, for micro learning.

A content-based convolutional neural network (CBCNN) recommender system was proposed in a prior study (Shu et al., 2018), which shows fairly satisfying ability in mining new or unpopular learning materials for a target learner. Another study proposed a new way to calculate similarities between online learning materials for recommendation tasks (Niemann & Wolpers, 2013). And the authors in that paper argued that the usage context-based model has the potential to outperform the content-based model, if the usage data is sufficiently fine-grained. And a system for recommending OERs in MOOC was proposed in (Hajri et al., 2017), which emphasized the significance of modelling users and learning materials. However, none of these studies mentioned the significance of the ranking for the success of an online learning service.

2.3. The significance of ranking ability of the recommender system for online learning

Unlike the personalized service in other areas (e.g., e-commerce and entertainment), complex pedagogical issues (Sikka et al., 2012; Wu et al., 2015) or requirements influence the learning outcome to a great extent. For example, the description of a learning material might contain vague information and pre-requested knowledge are required for some courses. Letting learner know what he/she should learn first what he/she needs to learn next is vital for an informal or non-formal online learning. Hence, for the online learning service like micro learning, a recommender system should be able to precisely distinguish the importance differences of the recommended resources (ranking).

3. Related work

3.1. Conventional recommendation strategies

Collaborative filtering and content-based filtering are two typical conventional recommendations strategies, which have been proven as effective and been commonly used in many studies or real applications. The recommendation results for a target user given by collaborative filtering are based on his/her correlation among other users of the system. As indicated in one previous study (Pazzani, 1999), collaborative filtering presents a uniform approach to finding items of potential interest and predicting the rating that the target user would give to the item.

Content-based filtering strategy generates recommendations by comparing and analysing the description of the items that have been rated by the target user and the descriptions of the items to be recommended (Pazzani, 1999). However, as the user's profile is constructed based on the user's historical activities, such recommendation strategy lacks the ability to explore and recommend the new items, which might vary greatly from the historical ones.

3.2. Language model and translation model for information retrieval and information filtering

As discussed in Berger and Lafferty (2017), applying the strategy of machine translation to solve the recommendation problem is not a fanciful idea but a feasible one. In this study, the researchers demonstrated constructing using a statistical machine translation model to handle the information retrieval task. Similarly, in another research (Lavrenko & Croft, 2017), researchers used a language model to reflect the mapping relationships between a query and a document. Statistical language models were explored and analysed for handling the recommendation task (Valcarce, 2015). However, applying a language model a translation model to solve a recommendation problem is still less touched.

4. Computation model

In this section, we roll out a novel recommendation strategy based on the combination of the concept of language and translation model. This strategy is realised in the recommender module of the proposed system in one early work (Lin et al., 2019a).

4.1. Translation model and language model

4.1.1. Statistical machine translation model

The translation is a probabilistic mapping procedure that a string e in one language can be translated to a string f in another language with the probability of $P(f|e)$. In the natural language processing (NLP) area the probability distribution of $P(f|e)$ can be modelled in different ways. For example, Bayes Theorem is used in one previous study to represent this distribution (Brown et al., 1992):

$$P(f|e) = \frac{P(e)P(f|e)}{P(f)} \qquad (1)$$

Since the denominator only correlates with source language f and we only consider the result of target language e, we can simplify this distribution as Equation (2):

$$P(e|f) \propto P(e)P(f|e) \qquad (2)$$

Finding the best translation result $\hat{e}$ is realised by finding the one that gives the highest probability:

$$\hat{e} = \underset{e}{argmax}\, P(e)P(f|e) \qquad (3)$$

4.1.2. Language model

Generally, the expression of a language is composed of sentences and phrases, and the representation of sentences or phrases is a sequence of words. The language model is a probability distribution of a sequence of words. In Brown et al. (1992), the authors assumed that the production of a piece of English text could be characterized by a set of conditional probabilities. Given an English sentence or phrase e which contains k words, its probability can be formulated as Equation (4):

$$P(e) = P(w_1, \ldots, w_k) = P(w_1)P(w_1|w_2)P(w_3|w_2, w_1) \ldots P(w_k|w_{k-1}, w_{k-2}, \ldots, w_1) \qquad (4)$$

N-gram model is one of the most representative language models applied in many NLP tasks, such as speech recognition, spelling correction, and translation. Given a sequence of *n-1* words, the n-gram model predicts the probability of the next word after this sequence. As the n-gram model keeps the continuous combination of n words, it is capable to preserve and represent some semantic information. When using the n-gram model, the probability of producing a sequence of words can be formulated as Equation (5):

$$P(w_1, \ldots, w_k) = P(w_1)P(w_1|w_2)\ldots P(w_k|w_{k-1}, w_{k-2}, \ldots, w_1) \approx \prod_{i=1}^{k} P(w_i|w_{i-(n-1)}, \ldots, w_{i-1}) \quad (5)$$

The probabilities used in these models can be simply calculated by using maximum likelihood estimation (MLE), for n-gram models and the translation procedures the probability can be formulated as Equation (6):

$$P\left(w_i \middle| w_{i-(n-1)}, \ldots, w_{i-1}\right) = \frac{C(w_{i-(n-1)}, \ldots, w_i)}{C(w_{i-(n-1)}, \ldots, w_{i-1})} \quad (6)$$

Here, $C(w_1, \ldots, w_i)$ represents the frequency of the word sequence $w_1, \ldots, w_i$ in the training sample.

4.2. The combination of the language model and the translation model

As discussed in the early pilot study (Lin, 2020), the language model can be used to model the online learning materials and historical learning records, and the machine translation model can be used to model the mapping relationship between historical learning records and the new learning materials. More specifically, in NLP, a language model is used to reflect the combination between words, while the proposed system utilizes a language model to reflect the combination between features of a certain learning material. For a translation task, a machine translation model is used to reflect the mapping relationship between two different languages, while the proposed system uses a machine translation model to reflect the mapping relationship between historical learning records and new learning materials.

The visualization of the process of using a language model to reflect the combination between words and feature combination of a learning material is shown in Figure 1. We can see that a sentence is composed of a sequence of words ($w_1, w_2, w_3, \ldots, w_n$), and similarly, a learning material is represented by several features ($f_1, f_2, f_3, \ldots, f_n$) such as type, language, discipline etc. For the recommendation process that tackled in this study, it is represented through a mapping procedure between the historical learning records and the new learning material, such procedure is similar to the process of language translation. The visualization of the translation procedure is shown in Figure 2. In the left part of Figure 2 is the translation procedure between one sentence in the original language and a set of translation results in the target language. Different translation results have different probabilities. The result with a higher probability means the result is the more reasonable one Similarly, in our proposed solution, the recommendation procedure between the historical learning activities and the recommended new learning materials is represented in the right part of Figure 2. Given the historical learning activities of a user, the system will recommend several different new learning materials with different probabilities. The higher probability of the learning material means it is more suitable to this user.

Figure 1. The modelling process of a sentence and a piece of learning material

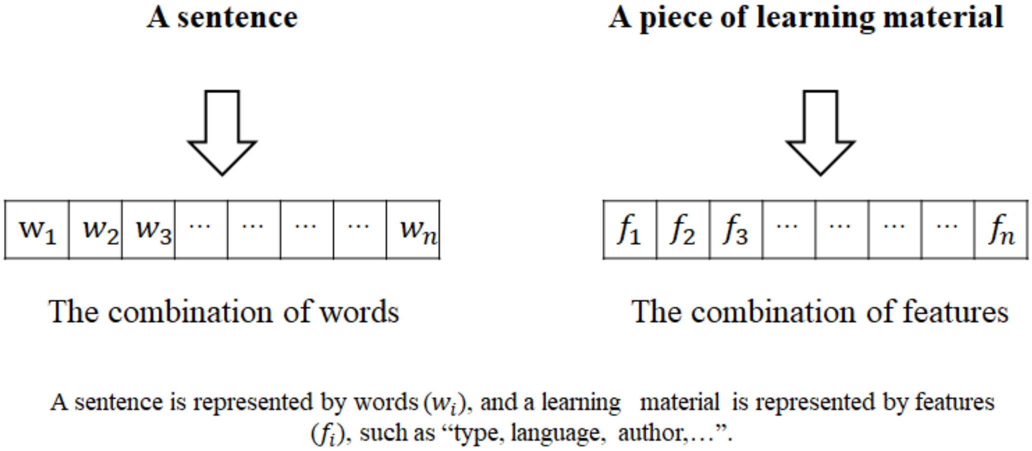
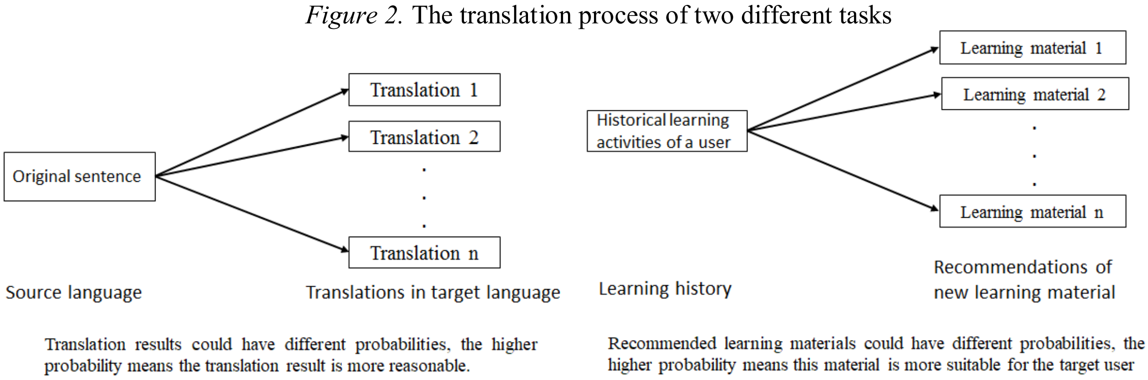


Figure 2. The translation process of two different tasks



Hence, the proposed recommender system in this study is composed of a language model and a machine translation model. The recommendation process is formulated as Equation (7) below:

$$P(l|h) \propto \prod_{i=1}^{k} P(f_i|f_{i-(n-1)}, \dots, f_{i-1})P(h|l) \qquad (7)$$

Herein based on the concepts defined in Lin (2020), l represents the learning material, h represents the historical learning records of the target user, f_i is the i-th feature that is used to represent the leaning material. The probability $P(l|h)$ represents the degree of correlation between a user's historical learning activities and the new learning material. Finding the best recommendation result $\hat{l}$ is realised by finding the one that gives the highest probability which is formulated as Equation (8):

$$\hat{l} = \underset{l}{argmax} \prod_{i=1}^{k} P\left(f_i\middle|f_{i-(n-1)}, \dots, f_{i-1}\right) P(h|l) \qquad (8)$$

4.3. Sub-translation for different types of features

As the different types of features contain different amount of information, it is more reasonable to interpret different types of feature/metadata separately. For example, some descriptive features, such as a subject title and the introduction of a course, are more important than the metadata (e.g., the resolution degree of the lecture video). However, such less-descriptive metadata can also reflect some latent information of the target user (Al-Hmouz et al., 2011), to some extent. During the recommendation process, different types of feature/metadata should be treated in different manners. Hence, it is more reasonable that the translation procedure proposed above is separated into several sub translation tasks, then all the translation results are assembled with different weighting values. For linearly assembling the several translation results, the final relevance score s between the recommended learning material and the user's historical learning records can be estimated by the following equation:

$$s = \sum_i \alpha_i p_i \text{ , and } \sum_i \alpha_i = 1 \qquad (9)$$

Herein, α_i are the weight for the i-th translation result. As described in Equation (7), p_i represents the correlation degree between the recommended learning material and the user's historical learning records produced by the i-th translation task. The top k recommended items are generated by ranking the s scores and picking the k items with the highest values. The workflow of our proposed translation set is shown in Figure 3.

5. System evaluation

In this section, we will discuss the feasibility analysis of our proposed model, the dataset used in the experiments, and the relevant evaluation metrics.

5.1. Feasibility analysis

The analysis of the feasibility of the proposed model and relevant experiments stems from three perspectives, model design, experimental dataset, and the evaluation metrics.

The proposed recommendation strategy does not involve any less-explainable "black-box" structure, such as the neural network. All the model designs are based on the Naïve Bayes rule and its variants, and all the formula deductions discussed in the previous sections are based on the solid mathematical foundation and probability theory.

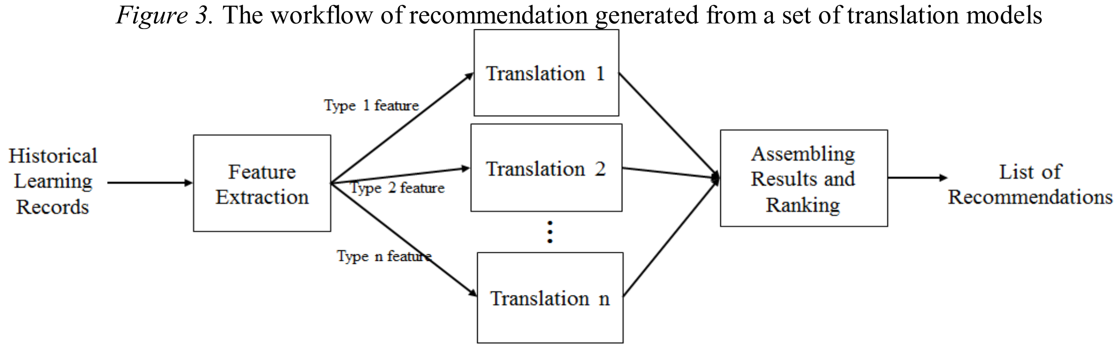
The dataset used in the relevant experiments of this study is a well-acknowledged public dataset, which is widely used in various recommender system related studies. The details of such dataset will be discussed in the next subsection.

In our work, we do not involve any novel evaluation metrics, all the evaluation metrics used in the experiments are also well-acknowledged in the research area of the recommender system.

5.2. Dataset

In one earlier work, researchers investigated the readiness of public and academic data sources that were adopted in e-learning literature (Lin et al., 2019b). By comparing the pros and cons of available public data sets, the dataset used in the experiments of this study is Book-Crossing (Ziegler et al., 2005). For the experiment, we also crawled some other descriptive information like description and comments for each book. The reasons of choosing this dataset can be summarized as follow:

Figure 3. The workflow of recommendation generated from a set of translation models



Compared to other open-source datasets like MovieLens (Harper & Konstan, 2015) and Jester (Goldberg et al., 2001), Book-Crossing dataset is more closer to the educational domain.

This work aims to demonstrate and prove that the translation and language models have the potential in mining latent useful information for boosting the recommending results. To demonstrate the effectiveness of applying this recommending strategy to the different formats of online learning material is beyond the scope of this paper.

5.3. Evaluation metrics

5.3.1. Normalized Discounted Cumulative Gain (NDCG)

As the proposed system aims to precisely rank the recommended learning materials, comparing the proposed system's ranking ability with the baseline is necessary. Hence, Normalized Discounted Cumulative Gain (NDCG) (Wang et al., 2013) will be used for measuring the ranking performance of the baselines and the proposed system.

5.3.2. Precision and recall at top K

Moreover, for a comprehensive comparison, we will also use Precision@K and Recall@K to evaluate the models for top K recommended results. These two metrics reflect the ability of a model to find relevant learning resources from the online repository.

6. Conclusions and future work

In this paper, we proposed a novel recommender system which applies the idea of language modelling and machine translation. From mathematic derivation, we can see that the proposed system can distinguish the importance differences of the recommended results. In the future, we will conduct the experiment of the proposed model based on the dataset from the educational domain. In this study, for assembling the set of translators, we proposed a simple linear assembling strategy. However, as demonstrated in one relevant study (Sagi & Rokach, 2018), some non-linear ensemble strategies also have strong ability to discover the complex structure and learn high-level concepts in large datasets. Hence, investigating how to better integrate sub-translators effectively is another area for our future work.

References

Al-Hmouz, A., Shen, J., Al-Hmouz, R., & Yan, J. (2011). Modeling and simulation of an adaptive neuro-fuzzy inference system (ANFIS) for mobile learning. *IEEE Transactions on Learning Technologies, 5*(3), 226-237.

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. In *23rd International Conference on World wide Web* (pp. 687–698). https://doi.org/10.1145/2566486.2568042

Belkin, N. J., & Croft, W. B. (1992). Information filtering and information retrieval: Two sides of the same coin? *Communications of the ACM, 35*(12), 29-38. https://doi.org/10.1145/138859.138861

Berger, A., & Lafferty, J. (2017). Information retrieval as statistical translation. *ACM SIGIR Forum, 51*(2), 219–226. https://doi.org/10.1145/3130348.3130371

Brown, P. F., Della Pietra, V. J., Desouza, P. V., Lai, J. C., & Mercer, R. L. (1992). Class-based n-gram models of natural language. *Computational Linguistics, 18*(4), 467-480.

Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001). Eigentaste: A Constant time collaborative filtering algorithm. *Information Retrieval, 4*(2), 133-151.

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An Empirical study of MOOC videos. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 41–50). https://doi.org/10.1145/2556325.2566239

Hajri, H., Bourda, Y., & Popineau, F. (2017). MORS: A System for recommending OERs in a MOOC. In *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)* (pp. 50-52). IEEE. https://doi.org/10.1109/ICALT.2017.89

Harper, F. M., & Konstan, J. A. (2015). The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TIIS), 5*(4), 1-19. https://doi.org/10.1145/2827872

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer, 42*(8), 30-37. http://dx.doi.org/10.1109/MC.2009.263

Lavrenko, V., & Croft, W. B. (2017). Relevance-based language models. *ACM SIGIR Forum, 51*(2), 260–267. https://doi.org/10.1145/3130348.3130376

Lin, J. (2020). Hybrid translation and language model for micro learning material recommendation. In *2020 IEEE 20th International Conference on Advanced Learning Technologies (ICALT)* (pp. 384-386). IEEE. https://doi.org/10.1109/ICALT49669.2020.00121

Lin, J., Sun, G., Cui, T., Shen, J., Xu, D., Beydoun, G., Yu, P., Pritchard, D., Li, L., & Chen, S. (2020). From ideal to reality: Segmentation, annotation, and recommendation, the vital trajectory of intelligent micro learning. *World Wide Web, 23*(3), 1747-1767. http://dx.doi.org/10.1007/s11280-019-00730-9

Lin, J., Sun, G., Shen, J., Cui, T., Yu, P., Xu, D., & Li, L. (2019a). A Survey of segmentation, annotation, and recommendation techniques in micro learning for next generation of OER. In *2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD)* (pp. 152-157). IEEE. https://doi.org/10.1109/CSCWD.2019.8791891

Lin, J., Sun, G., Shen, J., Cui, T., Yu, P., Xu, D., Li, L., & Beydoun, G. (2019b). Towards the readiness of learning analytics data for micro learning. In *International Conference on Services Computing, SCC 2019. SCC 2019. Lecture Notes in Computer Science* (Vol 11515, pp. 66-76). Springer. https://doi.org/10.1007/978-3-030-23554-3_5

Niemann, K., & Wolpers, M. (2013). Usage context-boosted filtering for recommender systems in TEL. In *European Conference on Technology Enhanced Learning, Scaling up Learning for Sustained Impact. EC-TEL 2013. Lecture Notes in Computer Science* (Vol. 8095, pp. 246-259). Springer. https://doi.org/10.1007/978-3-642-40814-4_20

Pazzani, M. J. (1999). A Framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review, 13*(5), 393-408.

Sagi, O., & Rokach, L. (2018). Ensemble learning: A Survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(4), e1249. https://doi.org/10.1002/widm.1249

Shu, J., Shen, X., Liu, H., Yi, B., & Zhang, Z. (2018). A Content-based recommendation algorithm for learning resources. *Multimedia Systems, 24*(2), 163-173. http://dx.doi.org/10.1007/s00530-017-0539-8

Sikka, R., Dhankhar, A., & Rana, C. (2012). A Survey paper on e-learning recommender system. *International Journal of Computer Applications, 47*(9), 27-30.

Sun, G., Cui, T., Guo, W., Beydoun, G., Xu, D., & Shen, J. (2015a). Micro learning adaptation in MOOC: A Software as a service and a personalized learner model. In *International conference on web-based learning, Lecture Notes in Computer Science* (Vol. 9412, pp. 174-184). Springer. https://doi.org/10.1007/978-3-319-25515-6_16

Sun, G., Cui, T., Xu, D., Shen, J., & Chen, S. (2018). A Heuristic approach for new-item cold start problem in recommendation of micro open education resources. In *International conference on intelligent tutoring systems 2018. Lecture Notes in Computer Science* (Vol. 10858, pp. 212-222). Springer. https://doi.org/10.1007/978-3-319-91464-0_21

Sun, G., Cui, T., Yong, J., Shen, J., & Chen, S. (2015b). MLaaS: A Cloud-based system for delivering adaptive micro learning in mobile MOOC learning. *IEEE Transactions on Services Computing, 11*(2), 292-305.

Syeda-Mahmood, T., & Ponceleon, D. (2001). Learning video browsing behavior and its application in the generation of video previews. In *Proceedings of the ninth ACM international conference on Multimedia* (pp. 119–128). https://doi.org/10.1145/500141.500161

Valcarce, D. (2015). Exploring statistical language models for recommender systems. In *Proceedings of the 9th ACM Conference on Recommender Systems* (pp. 375–378). https://doi.org/10.1145/2792838.2796547

Wang, Y., Wang, L., Li, Y., He, D., Chen, W., & Liu, T.-Y. (2013). A Theoretical analysis of NDCG ranking measures. In *Proceedings of the 26th annual conference on learning theory* (COLT 2013), (Vol. 8, pp. 6).

Wasid, M., & Ali, R. (2017). Use of soft computing techniques for recommender systems: An Overview. In *Applications of soft computing for the web* (pp. 61-80). Springer. https://doi.org/10.1007/978-981-10-7098-3_5

Wu, D., Lu, J., & Zhang, G. (2015). A Fuzzy tree matching-based personalized e-learning recommender system. *IEEE transactions on fuzzy systems, 23*(6), 2412-2426. http://dx.doi.org/10.1109/TFUZZ.2015.2426201

Ziegler, C.-N., McNee, S. M., Konstan, J. A., & Lausen, G. (2005). Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web* (pp. 22–32). https://doi.org/10.1145/1060745.1060754