British Journal of Educational Technology *Vol 51 No 5 2020* 1734–1747
doi:10.1111/bjet.12951

Course video recommendation with multimodal information in online learning platforms: A deep learning framework

Wei Xu and Yuhan Zhou

Wei Xu is an associate professor at School of Information, Renmin University of China. His research interests include big data analytics, business intelligence and decision support systems. He has published over 150 research papers in international journals and conferences, such as Production and Operations Management, Decision Support Systems, European Journal of Operational Research, IEEE Trans. and International Journal of Production Economics. Yuhan Zhou is a graduate student at School of Information, Renmin University of China. Her research interests include big data analytics, business intelligence and decision support systems. Address for correspondence: Wei Xu, School of Information, Renmin University of China, Beijing 100872, P.R. China. Email: weixu@ruc.edu.cn

Abstract
With the rapid development of online learning platforms, learners have more access to various kinds of courses. However, they may find it difficult to make choices due to the massive number of courses. The main contribution of our research is the design of a course recommendation framework which extracts multimodal course features based on deep learning models. In this framework, different kinds of information of course, such as course title, and course audio and course comments, are used to make proper recommendation in online learning platforms. Moreover, we utilize both explicit and implicit feedback to infer learner's preference. Based on real-world datasets, our empirical results show that the proposed framework performs well in course recommendation, achieving an AUC score of 79.03%. This framework can provide technical support for course video recommendation, thus helping online learning platforms to manage course resources and optimize user learning experience.

Introduction

In the age of Web 2.0, teaching methods have shifted from offline to online. Online education can improve the effectiveness of the learning process (Kekkonen-Moneta & Moneta, 2002), thus becoming an increasingly more important way for users to learn courses in a convenient and inexpensive way. In the education field, when combing e-learning with traditional teaching methods, students' academic performance can improve a lot (Condie & Livingston, 2007). Online learning platforms such as MOOC and Coursera allow learners to have quick access to various kinds of courses, with over thousands of courses and over millions of enrolled students. However, the massive number of courses makes it difficult for learners to choose a proper one. Each learner has his/her own interests and capabilities, thus personalized recommendation is important. A recent study shows that the completion rate of MOOCs is extremely low (De Freitas, Morgan, & Gibson, 2015). The mismatch between courses' difficulty and students' skills becomes a main hindrance of leveraging online learning to provide knowledge for learners. In this paper, we focus on making proper course recommendation using multimodal data based on deep learning methods.

Practitioner Notes

What is already known about this topic

- Online learning has become an increasingly more important way for users to learn courses in a convenient and inexpensive way.
- With the rapid development of online learning platforms, several studies have explored the recommendation problem in online learning platforms.
- Some factors have been used in online learning recommendation; however, previous studies have only focused on textual information like title and introduction.

What this paper adds

- The main contribution of our research is the design of a course recommendation framework which extracts multimodal course features based on deep learning models.
- In this framework, different kinds of information of course videos, such as course title, and course audio and course comments, are used to make proper recommendation in online learning platforms.
- We utilize both explicit and implicit feedback to infer learner's preference.
- Our empirical results show that the proposed deep learning framework outperforms baselines in course video recommendation.

Implications for practice and/or policy

- This paper can not only extract features to facilitate the recommendation of courses, but also help explaining what factors attract learners in online learning platforms.
- The combination of learner's basic information, explicit feedback and implicit feedback can fully infer his/her preferences.
- This framework can provide technical support for course video recommendation, thus helping online learning platforms to manage course resources and optimize user learning experience.

Previous research has explored the recommendation problem in online learning platforms (Liu, Fan, Chou, & Chen, 2010). However, these studies only considered some numerical features (eg, course relevance score) when recommending courses; a few studies extracted other hidden features from courses (eg, textual features extracted from course titles, and visual features extracted from course videos) using deep learning methods. Moreover, though some studies tried to make recommendation by integrating social information into existing methods (Intayoad, Becker, & Temdee, 2017), a few of the previous studies inferred learner's preference from his/her play record or view record. This study can fill the existing research gaps.

When learners want to start a new course, they make choices according to course's title to see whether the course will satisfy their demands or not. Also, the number of comments represents the popularity and quality of the course. Accordingly, both numerical and textual information can be leveraged to recommend courses. Moreover, both acoustics information and visual information of course videos have simultaneous impact on learners when they are learning courses. One novelty of our research is to extract acoustics and visual features using deep learning methods with common numerical and textual features to improve the effect of course recommendation. As for learners' behavior on online learning platforms, playing course video can be treated as explicit feedback, as it directly reflects the learner's preference. Meanwhile, just viewing course's title, comments, etc but not playing the video can indirectly reflect the learner's preference. Learners

may be interested in the course and watch the course video next time. This implicit feedback should also be considered when recommending courses.

In brief, our research work provides the following main contributions. First, we take multimodal data, such as textual information, video and audio recordings of courses into consideration during the recommendation process. Second, we propose the recommendation model which integrates learners' basic information as well as explicit and implicit feedback data. Finally, we design a novel deep learning framework which extracts hidden information of learners and courses in online learning platforms. To the best of our knowledge, in the online learning field, this is the first successful research of integrating different types of learners' preferences and courses' multimodal features in a deep learning recommendation model. The practical implication of our research is that the proposed deep learning framework can be applied in online learning platforms to recommend proper courses based on learners' preference, and hence they can improve learner click rate and reduce learner attrition rate.

The remainder of this paper is organized as follows. The Related Work section summarizes existing literature related to recommendation in online learning platforms and deep learning methods. The Proposed Method section describes the proposed framework with details for recommending course videos. Discussions of our experimental results are presented in the Empirical Analysis section. The Conclusions and Future Work section gives conclusions and future work of our research.

Related work

The overview of online learning

Online learning has received growing research attention over recent years. Research problems mainly focus on the role of online learning in the education field (Condie & Livingston, 2007; Kekkonen-Moneta & Moneta, 2002; Zhang, Perris, & Yeung, 2005), the use of new education tools (Griff & Matter, 2013; Wang, Woo, Quek, Yang, & Liu, 2012; Warburton, 2009), and learners' behavior (Choi, Lee, & Kang, 2009; Shin & Chan, 2004; Wang, 2007).

To study the role of online learning in the education field, most students hold positive perceptions toward distance education due to the convenience of the Internet according to the results of questionnaire survey in Hong Kong (Zhang *et al.*, 2005). Findings suggest that online students perform better considering the applied conceptual learning, and the use of interactive e-learning can improve the effectiveness of the learning process (Kekkonen-Moneta & Moneta, 2002). Condie and Livingston (2007) find that combining information and communication technology with traditional methods will have more positive impacts on students' academic performance.

To study the use of new education tools, Wang *et al.* (2012) use Facebook as a learning management system (LMS) to explore its potential for teaching and learning. Second Life is a popular virtual world platform used in the education field, and its effectiveness is examined from three aspects including technical, immersive and social aspect (Warburton, 2009). The adaptive online learning system, LearnSmart, can perform the best when it has a close link with course goals and texts (Griff & Matter, 2013).

To study learners' behavior, Choi *et al.* (2009) explore the impact of four learning styles on students' learning experience and outcomes in e-learning environment, and find the slight influence of active–reflective learning style during the early stage. The level and the requirement of the course are two important factors that influence students' engagement in online learning (Shin & Chan, 2004). Wang (2007) explores the effect of power distance index (PDI) on students' perceptions of e-learning experience.

Recommendation in online learning platforms

Online learning recommendation objects can be e-learning resources (Manouselis, Vuorikari, & Van Assche, 2010) and cooperative partners (Zheng & Yano, 2007). Generally, there are three types of filtering-based approaches in recommendation: collaborative filtering (CF), content-based filtering (CB) and hybrid approaches (Adomavicius & Tuzhilin, 2005). Based on learner's study history, the collaborative filtering (CF) method has been widely used in online learning recommendation. The content-based filtering method recommends courses similar to what users have learned in the past. Hybrid approaches are the combination of the CF and CB methods. Based on learning records, several studies cluster learners with similar preferences and interests into groups. As a result, relevant courses can be recommended. Liu *et al.* (2010) use concept association techniques to find the relevance between courses, thus making proper recommendation.

With online learning becoming increasingly popular, learners' interests, preferences and education backgrounds are more complicated. Considering the difference in students' learning capabilities and knowledge backgrounds, Aher and Lobo (2013) propose an improved algorithm using the combination of clustering technique and association rule algorithm, which performs well compared to other methods. A social context-aware personalized recommendation system is developed by taking learner's learning style and knowledge background into consideration (Intayoad *et al.*, 2017).

Although the course is usually presented in the form of a video, no attention has been paid to the impact of acoustics information and visual information on the effectiveness of recommendation. Therefore, our research proposes a new deep learning framework using multimodal data. In addition, we consider not only basic, but also explicit and implicit information of learners, which are meaningful in making proper course recommendation in online learning platforms.

Deep learning-based recommendation

Traditional recommending methods have many drawbacks, such as sparsity of data and cold-start problem. Deep learning methods have been widely used to tackle these problems. A dynamic personalized recommendation algorithm is proposed to address the problem of sparse rating matrix. This algorithm catches the latent relations between ratings using information in profile contents and ratings, which performs well on public datasets (Tang & Zhou, 2012). For cold start items, Stacked Denoising Auto Encoder (SDAE) model is used to extract the content features of the items (Wei, He, Chen, Zhou, & Tang, 2017). Shi, Zhao, and Shen (2017) propose a novel interview-based model to cluster all users into different groups. Many recommendation methods integrate social information like social relations (Liu, Wu, & Liu, 2013) into existing models. Social contagion theory and social homophily theory are considered to fully infer user's preference (Li, Wang, & Liang, 2014). Besides, new mechanisms like Markov chain model (Fouss, Pirotte, Renders, & Saerens, 2007) and ontology-based semantic similarity (Al-Hassan, Lu, & Lu, 2015) have been used.

While traditional recommending methods mainly focus on numerical and textual information, deep learning methods attempt to extract features from images, audio and video recordings, like clothing pictures (Guan, Wei, & Chen, 2019) and music (Cheng & Shen, 2016). A multiview model is proposed to take multiple sources like product photos and comment texts into count, which achieves high accuracy (Guan *et al.*, 2019). To solve the problem of unavailability of specific video features, rich contents like text, motion and audio are used for video recommendation (Zhao *et al.*, 2013).

Diverging from previous research, our study proposes a recommendation method based on deep learning in the online learning field. The use of deep learning methods can help extract hidden information of learners and courses, thus making proper recommendation.

The proposed method

In this study, we design a new multimodal information framework for online course video recommendation. In this framework, different kinds of information of course are used to make proper recommendation in online learning platforms. Both play record and view record are utilized to infer learner's preference, which can be regarded as explicit and implicit feedback. The deep learning methods are employed to recommend online course video. The proposed multimodal information framework for online course video recommendation is presented in Figure 1. As can be seen from Figure 1, four steps, namely, data collection, feature extraction, profiling and recommendation, are consisted in the proposed framework.

Data collection

Course and learner data from an online learning platform are retrieved. The retrieved course data include numerical data (eg, comment volume), textual data (course title) and video data. Three numerical features shown in Table 1 are utilized by the proposed recommendation model. Course titles are used as textual features because learners form an initial impression of courses based on their titles. As for video features, both acoustics information and visual information are applied due to their simultaneous impact on learners when they play courses. The retrieved learner data include learner's basic information (ie, age and gender), play record and view record. Play record and view record are two lists of courses, which are regarded as learner's explicit and implicit feedback. Numerical, textual, acoustics and visual features of these records can be extracted in the same way that the previously mentioned course features are extracted.

Feature extraction

In this paper, we employ different modalities of data, including numeric indicators, titles and videos. In addition, we have explicit and implicit feedback data of learners. The numerical data are preprocessed to represent concrete features. Numerical course features (ie, duration, play volume and comment volume) and learner's basic information (ie, age and gender) are converted into 3-dimensional and 2-dimensional vectors respectively. As for titles and videos, we use different processing methods to extract hidden textual and video features.

Textual feature extraction

Text in our paper refers to course title. Textual data are preprocessed, ie, word segmentation and stop-word removal. Each Chinese sentence is divided into words in the word segmentation process. Afterward, stop words like "the" and "is" are removed from sentences. Considering that our dataset is retrieved from one of the biggest online learning platforms in China, we applied Chinese word embeddings pretrained with a large Chinese corpus (Li *et al.*, 2018) on titles to capture the semantic information among words. Each word in a title is converted into a 300-dimensional vector, and we use the average of these vectors to represent textual information of the title. In this way, we convert each title into a 300-dimensional feature vector.

Video feature extraction

In addition to textual features, we mine visual and acoustics features from video and audio streams of course videos. The visual information represents the content of the course, and the acoustics information affects learner's listening experience. We use different processing methods to extract

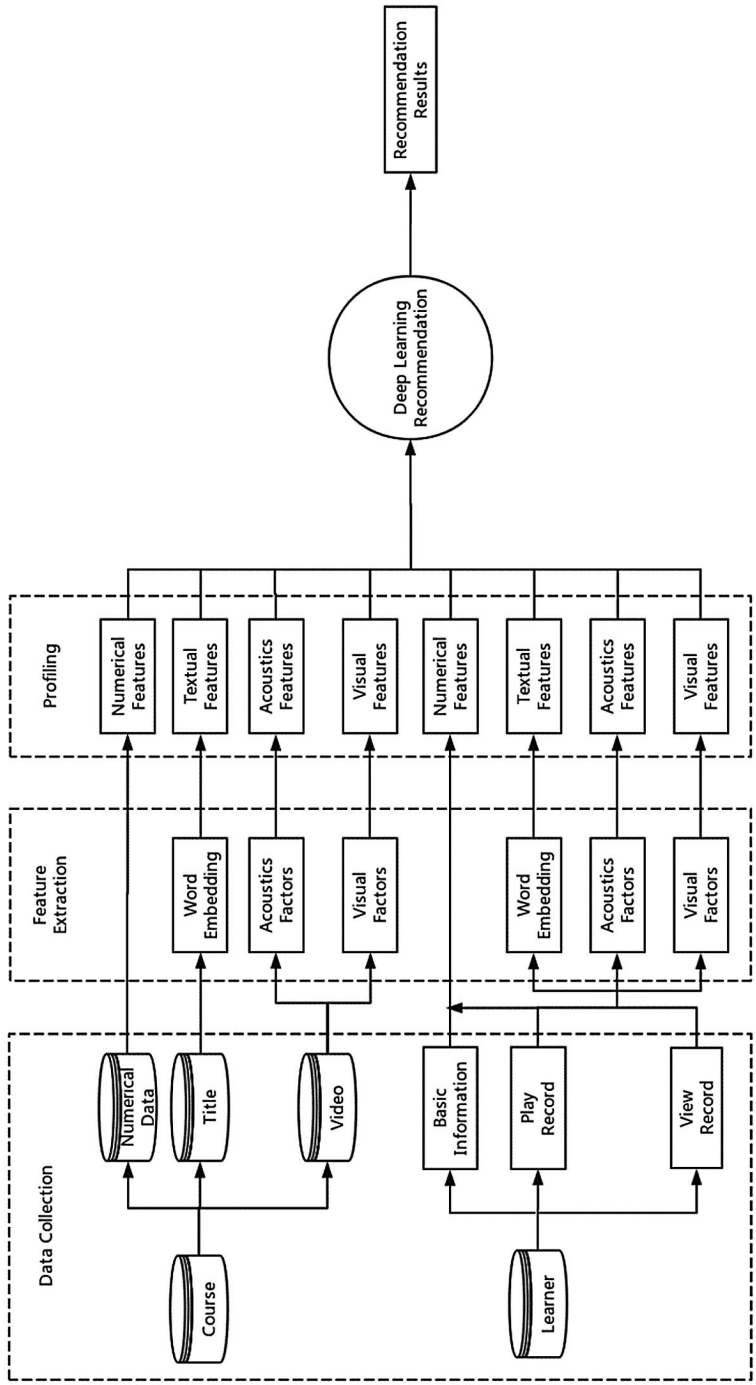

Figure 1: A multimodal information framework for online course recommendations

Table 1: Numerical course features

Feature	*Description*
Duration	Number of seconds of a course
Play volume	Number of times played by learners
Comment volume	Number of comments of a course

hidden features from videos while retaining the inherent correlation in the data. The impact of visual information and acoustics information will be discussed separately in the Empirical Analysis Section.

To extract visual features by frame, we treat a video stream as a series of images and extract one frame per second from a video. To maintain the effectiveness and integrity of each frame, we apply the pretrained ResNeXt-50 model (Xie *et al.*, 2017), which is released by Facebook Research and implemented on PyTorch. Each frame is scaled and regularized, and each image is prepared as three channels of RGB. By initializing weights pretrained on the ImageNet dataset, this pretrained model can convert the three channels into a 2048-dimensional vector. After extraction by the ResNeXt-50 model, the series of video frames is converted into a series of the 2048-dimensional frame features. We use the average of these feature vectors to represent visual information of the whole video. Thus, videos with different numbers of frames can be converted into feature vectors of the same 2048 dimensions.

The audio streams are processed by the python module, Librosa, which has been extensively applied in audio signal analysis. We extract five acoustics features including zero-crossing rate, spectral centroid, spectrum attenuation, Mayer frequency cepstrum coefficient and chrominance frequency for audio evaluation. The meanings of these five acoustics features are presented in Table 2. Finally, we obtain five features from each audio stream and prepare them for further modeling.

Using the different processing methods above, we convert each course video into a 3-dimensional vector of numerical features, a 300-dimensional vector of textual features, a 2048-dimensional vector of visual features and a 5-dimensional vector of acoustical features. For each learner, we get a 2-dimensional vector of basic information. As for courses listed in his/her play record and view record, the numerical, textual, visual and acoustics features can be extracted the same way. The dimensions of play record vector and view record vector are both 2356.

Table 2: The meaning of acoustics features

Acoustics feature	*Meaning*
Zero-crossing rate	The rate of sign-changes in a signal
Spectral centroid	A measure which indicates where the center of mass of the spectrum is, reflecting the impression of brightness of a sound
Spectrum attenuation	A measure of the shape of a signal
Mayer frequency cepstrum coefficient	A measure which describes the overall shape of the spectrum envelope, which simulates the characteristics of human voice
Chrominance frequency	The average of frequencies in 12 intervals, representing the 12 distinct semitones or shades of a musical octave

Profiling

After the multimodal feature extraction of courses, the raw data are processed into vectors of different dimensions. To represent the feature of each course, we vertically concatenate the numerical, textual, visual and acoustics feature vectors, and finally get a 2356-dimensional course vector. As for learner's explicit and implicit feedback, the numbers of played courses and viewed courses vary from learner to learner. To get vectors with a fixed dimension, we calculate the average of course feature vectors. For example, a learner has played courses A and B, and has viewed courses C and D. We use the average of two 2356-dimensional course vectors of A and B to represent the learner's explicit feedback. For his/her implicit feedback, the average of two 2356-dimensional course vectors of C and D is calculated. Therefore, for each learner, we get a 2-dimensional vector of basic information, a 2356-dimensional vector of explicit feedback and a 2356-dimensional vector of implicit feedback. For each learner, we vertically concatenate these vectors, and finally get a 4714-dimensional learner vector to represent his/her feature.

To sum up, for each learner and course pair, we convert the raw data into a 4714-dimensional learner feature and a 2356-dimensional course feature. We vertically concatenate these two features as the input of our recommendation model. The output target is 1 if the learner has played the course, otherwise 0.

Recommendation

Deep learning models are proposed to carry out the recommendation tasks. As a widely used machine learning method, deep learning has been used into the recommendation field. The deep learning usually outperforms well-known classifiers like NN and SVM for a variety of recommendation tasks. The modeling process is shown in Figure 2.

The model consists of two parts, which correspond to the processing of course and learner raw data. For each course, the numerical data are processing to represent concrete features. The textual title of course is transferred to vector by word embedding. The video stream is segmented into a series of video frames, and each frame is processed by the pretrained ResNeXt-50 model for feature extraction. The average of all frame vectors is the visual feature. Also, five acoustics features are extracted using Librosa. For each learner, his/her age and gender are converted into a 2-dimensional vector, and the average vectors of courses listed in his/her play and view records are used. We vertically concatenate the two groups of course and learner features as the input of our recommendation model. The aggregate feature is fed into a single-directional LSTM, and then a fully connected layer for the classification and recommendation. We employ the output of the fully connected layer as our prediction result. The computation of the accuracy between the true value and the prediction result helps us optimize our model parameters.

Empirical analysis

The datasets

Our learner and course datasets were retrieved from one of the biggest online learning platforms in China. For each learner, we collected his/her basic information (ie, age and gender), play record and view record. For each course video, we collected the video streams and audio streams, titles and numerical data (ie, duration, play volume and comment volume) as input multimodal materials for our recommendation model. To fully infer learner preferences from one's play record, we ensured each learner played at least two courses in November 2019. After omitting some learners, we selected a total of 1853 learners, 5479 courses, 21619 play records and 24258 view records from various categories of the platform.

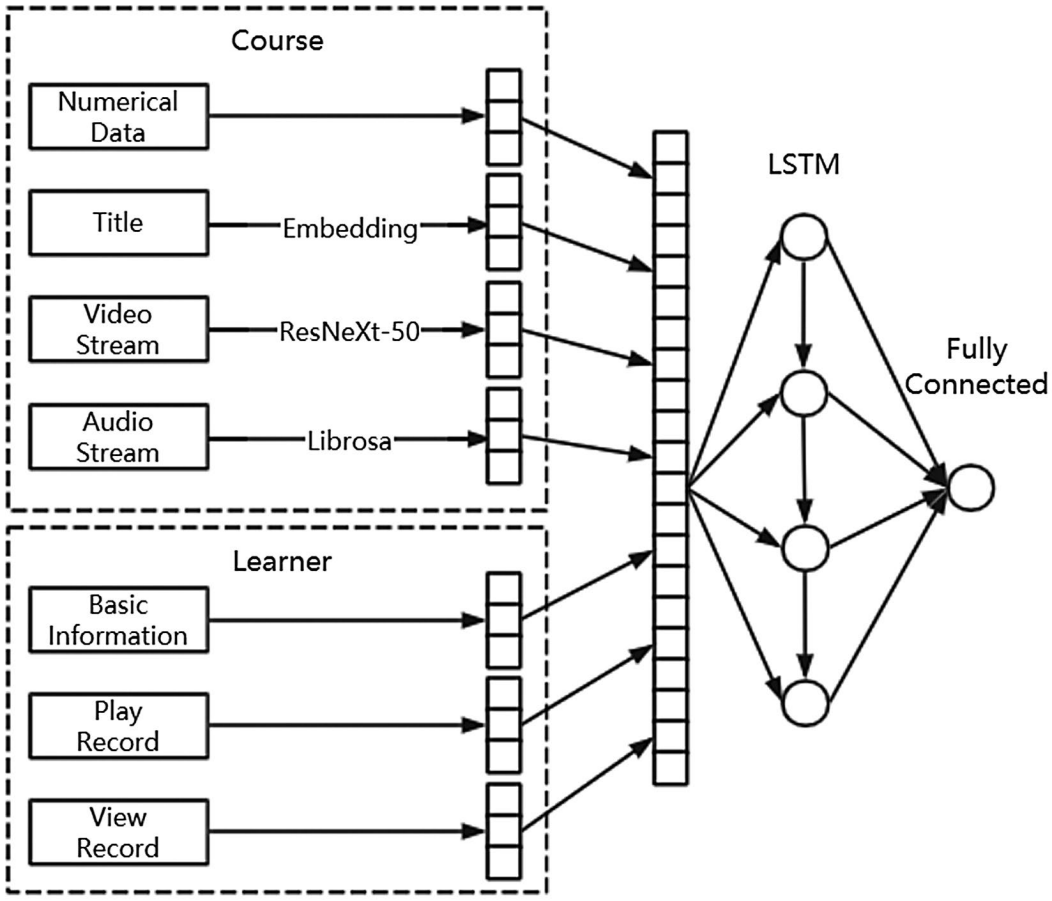

Figure 2: The deep learning modeling process

The performance measures

We applied two performance measures (Guan *et al.*, 2019) to evaluate the effectiveness of the proposed deep learning recommendation framework. The first measure is Area Under the ROC Curve (AUC). The AUC measures the ratio of correctly predicted learner and course pairs to all pairs. For each learner m and each course n, the correctly predicted pair is defined in (1):

$$X_{mn} = \begin{cases} 1 & (\text{Learner } m \text{ has played course } n.) \\ 0 & (\text{Learner } m \text{ has not played course } n.) \end{cases} \quad (1)$$

The second measure is Hit Ratio, which is defined as the proportion of learners who have one or more correctly recommended courses in his/her top-K recommendation course list. In the following sections, different values of K were chosen.

Table 3: Recommendation performance (%) of various models

	DNN	*LSTM*	*CNN*
AUC	74.67	79.03	67.72
HIT @50	5.21	7.61	4.43
HIT @100	13.67	15.76	11.34
HIT @150	21.52	24.78	20.99
HIT @200	29.34	31.09	25.63

Experimental settings

We implement our recommendation framework by Keras, and use Python to process the data. We employ a back-propagation and batch training method for the training step. Each layer in the LSTM uses the Relu activation function, while the fully connected layer uses the Sigmoid activation function because it is a binary classification problem. We use the RMSProp as the optimizer. As for parameter settings, the batch size is 16, the number of iterations is 100 and the learning rate is 0.001. The loss function is "binary_crossentropy" and the metrics is "accuracy." For the model evaluation, a half of the played courses of each learner are randomly chosen for the training set, and the other played courses are chosen for the test set.

Experimental results of recommendation models

We evaluated the recommendation performance of a traditional Deep Neutral Network (DNN) as well as other well-known deep learning methods such as Long Short Term Memory (LSTM) which is a special type of Recurrent Neural Network (RNN), and Convolutional Neural Network (CNN). The adopted feature set included all course features and learner information. Table 3 reports our experimental results. Based on AUC and four Hit Ratios, the proposed LSTM outperforms all the baseline models. The possible reason is that compared with LSTM, the traditional neutral network does not take the sequence information into consideration. CNN may lose some important information because of the pooling layer. Common pooling mechanisms like max-pooling and average-pooling can avoid overfitting but somehow miss important features when reducing dimensions. Therefore, for the rest of our experiments, we applied LSTM as the recommendation model.

Comparison for various kinds of learner information

We evaluated the effectiveness of different combinations of learner information. In this section, we used four information sets, namely, S1, S2, S3 and S4 which represented basic information (ie, age and gender), basic information and explicit feedback, basic information and implicit feedback, and using all available learner information. In these information sets, all course features are considered. Our experimental results are shown in Table 4.

The experiments of various kinds of learner information are enlightening. As we previously described, explicit feedback comes from learner's play record and implicit feedback comes from

Table 4: Recommendation performance (%) of various kinds of learner information

	S1	*S2*	*S3*	*S4*
AUC	74.05	78.48	75.76	79.03
HIT @50	3.75	5.82	7.28	7.61
HIT @100	8.91	15.16	14.24	15.76
HIT @150	16.90	19.40	20.71	24.78
HIT @200	21.36	26.47	27.01	31.09

Table 5: Recommendation performance (%) of various course feature sets

	P1	*P2*	*P3*	*P4*	*P5*
AUC	60.01	63.79	65.17	75.70	79.03
HIT @50	2.31	4.77	4.14	7.18	7.61
HIT @100	13.80	14.18	14.15	15.26	15.76
HIT @150	17.05	19.89	20.19	21.03	24.78
HIT @200	22.13	25.22	25.76	27.88	31.09

learner's view record, which can directly and indirectly reflect learner's preference. It is distinct that the information subset S4 (ie, basic information, explicit feedback and implicit feedback) leads to the best performance. As shown in Table 4, the deep learning model can achieve an AUC of 79.03%. Another metric, Hit Ratio, measures the ranking-based recommendation accuracy. When K varies from 50 to 200, S4 performs the best across all K levels. The AUC and four Hit Ratios of integrating all learner information are higher than the baseline by 6.73%, 102.93%, 76.88%, 46.63% and 45.55% respectively. The comparison of experimental results proves that by integrating different kinds of information of learners, the proposed method can improve the recommendation effectiveness in online learning platforms.

Comparison for various course feature sets

The effectiveness of different combinations of course features is also tested. For this series of experiments, we tried five feature sets, namely, P1 (ie, numerical features), P2 (ie, numerical features and textual features), P3 (ie, numerical features, textual features and acoustics features), P4 (ie, numerical features, textual features and visual features) and P5 (ie, all available course features). In these feature sets, all kinds of learner information are considered. Our experimental results are presented in Table 5.

On the whole, it is distinct that the feature subset P5 (ie, all available features) leads to the best performance. By using all available features, the percentage of AUC performance improvement over using only numerical features is 31.69%. These experimental results confirm the benefits of the multimodal course features when recommending courses. Also, it is noticeable that the feature subset P4 (ie, numerical features, textual features and visual features) achieves better performance than P3 (ie, numerical features, textual features and acoustics features). The results indicate that visual features matter more than acoustics features. When watching course videos, video streams contain more important information for course recommendation.

Discussions

The proposed deep learning framework for course recommendation is effective. First, when learners want to start a new course, they browse the information pages and then click to play the course video that they find useful. In this process, they make choices according to the course title to see whether the course will satisfy their demands or not. Then, the play volume and the comment volume represent the popularity and quality of the course, which can help learners make decisions. And the duration represents how much time a learner needs to spend on the course. Finally, when learners watch a course video, acoustics information and visual information have a simultaneous effect on learners, which are extracted as synergistic features. By applying the proposed recommendation model to mine multimodal information from titles and videos, rich hidden features are obtained. To enhance the proper recommendation of online courses, the rich hidden features are combined with numerical features to build a multimodal feature set. As for

learners, the basic information includes age and gender, which also affect the recommendation results. In addition, learner's preference can be fully inferred from play record and view record, which can be regarded as explicit feedback and implicit feedback. Combined with learner's basic information, these features can improve the performance of the proposed recommendation model. Based on real-world data collected from one online learning platform in China, our experimental results show that the proposed framework can achieve an AUC score of 79.03%. When K varies from 50 to 200, the Hit Ratios are 7.61%, 15.76%, 24.78% and 31.09%, which show high recommendation accuracy. On the whole, our empirical results show that course's multimodal information and learner's rich preferences are two important determinants of accurate course recommendation. These features can not only facilitate the recommendation of courses but also help explaining what factors attract learners in online learning platforms. By mining hidden features from learners and courses, the platform can recommend more suitable courses to learners, and hence they can better manage and distribute course resources. Also, the effective recommendation of courses can boost both traffic and revenues of the platform.

In addition, our experiments compare prediction performance of various kinds of learner information and course features respectively. The results indicate that visual features are more effective than acoustics features. In addition, our experimental results show that different deep learning methods do influence the final recommendation performance. By using a neural network which takes sequence information into consideration (eg, LSTM), it proves to be more effective than using a traditional neural network (eg, DNN).

Conclusions and future work

In the age of Web 2.0, online learning has become one of the most important sources for learners to have access to various courses. Due to the explosive growth in online learning, how to recommend proper courses to learners to improve the learning effectiveness in MOOCs becomes a key problem. Previous studies have introduced some numerical features of courses for course recommendation; however, textual features and video features have been little explored. Moreover, though learner's preference has been considered according to their play record, their implicit feedback is neglected. Our research provides the following contributions. First, a novel deep learning framework is designed for recommending courses in online learning platforms. Second, multimodal features extracted from courses (ie, numerical features, textual features, acoustics features and visual features) are fully fused. Third, learner's basic information, explicit feedback and implicit feedback are combined to infer his/her preferences. Finally, an empirical study is conducted to identify the effective features that influence recommendation performance.

Using real-world learner and course datasets, our experimental results show that multimodal features mined from courses are useful for recommendation. The LSTM recommendation model achieves an AUC score of 79.03% and Hit Ratios of 7.61%, 15.76%, 24.78% and 31.09%, which indicate high recommendation accuracy. Our experimental results also show that visual features are more important than acoustics features. Moreover, the model which uses learner's all information outperforms the model which only uses learner's basic information by 6.73% in terms of the AUC score. The practical implications of our research work are that online learning platforms can apply the proposed deep learning framework to recommend proper courses to learners. As for learners, they need not waste time choosing proper courses from long lists of courses. Recommendation can help learners have quick access to courses that they really want to study, thus improving the efficiency of the learning process. In addition, since our deep learning framework has taken course's multimodal information and learner's rich preferences into consideration, the contents of recommended courses can fully satisfy learner's needs. Learners with

different learning capabilities and knowledge backgrounds can find their own suitable courses. As a result, the completion rate and learner's learning quality can be highly improved.

More sophisticated deep learning models such as the Restricted Boltzmann Machine (RBM) model and the attention model will be explored to enhance the course recommendation process in our future work. In addition, our current research only examines textual features of course titles. To further improve the performance of the proposed recommendation model, more course attributes (eg, course descriptions and comments) will be incorporated. Finally, our current research only examines learner and course data collected from one online learning platform in China. Our future work will conduct a larger scale of experimentation to prove the effectiveness of our study. For instance, we will use data retrieved from more online learning platforms in the near future.

Statements on open data, ethics and conflict of interest

We can provide the research data by email to the corresponding author. Email address: weixu@ruc.edu.cn (W. Xu).

We have to read and follow the guidelines set out in Wiley's Best Practice Guidelines on Publishing Ethics. We have written an entirely original work, and others' work has been appropriately cited in our paper. We can provide the research data by email to the corresponding author, and we also anonymize the research data to protect them.

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work. There is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled.

References

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge & Data Engineering*, *6*, 734–749.

Aher, S. B., & Lobo, L. M. R. J. (2013). Combination of machine learning algorithms for recommendation of courses in e-learning system based on historical data. *Knowledge-Based Systems*, *51*, 1–14.

Al-Hassan, M., Lu, H., & Lu, J. (2015). A semantic enhanced hybrid recommendation approach: a case study of e-government tourism service recommendation system. *Decision Support Systems*, *72*, 97–109.

Cheng, Z., & Shen, J. (2016). On effective location-aware music recommendation. *ACM Transactions on Information Systems (TOIS)*, *34*(2), 13.

Choi, I., Lee, S. J., & Kang, J. (2009). Implementing a case-based e-learning environment in a lecture-oriented anaesthesiology class: do learning styles matter in complex problem solving over time? *British Journal of Educational Technology*, *40*(5), 933–947.

Condie, R., & Livingston, K. (2007). Blending online learning with traditional approaches: changing practices. *British Journal of Educational Technology*, *38*(2), 337–348.

De Freitas, S. I., Morgan, J., & Gibson, D. (2015). Will MOOCs transform learning and teaching in higher education? engagement and course retention in online learning provision. *British Journal of Educational Technology*, *46*(3), 455–471.

Fouss, F., Pirotte, A., Renders, J. M., & Saerens, M. (2007). Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Transactions on Knowledge and Data Engineering*, *19*(3), 355–369.

Griff, E. R., & Matter, S. F. (2013). Evaluation of an adaptive online learning system. *British Journal of Educational Technology*, *44*(1), 170–176.

Guan, Y., Wei, Q., & Chen, G. (2019). Deep learning based personalized recommendation with multi-view information integration. *Decision Support Systems*, *118*, 58–69.

Intayoad, W., Becker, T., & Temdee, P. (2017). Social context-aware recommendation for personalized online learning. *Wireless Personal Communications*, *97*(1), 163–179.

Kekkonen-Moneta, S., & Moneta, G. B. (2002). E-learning in Hong Kong: comparing learning outcomes in online multimedia and lecture versions of an introductory computing course. *British Journal of Educational Technology*, *33*(4), 423–433.

Li, S., Zhao, Z., Hu, R., Li, W., Liu, T., & Du, X. (2018). Analogical reasoning on Chinese morphological and semantic relations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (pp. 138–143). Melbourne, Australia: Association for Computational Linguistics.

Li, X., Wang, M., & Liang, T. P. (2014). A multi-theoretical kernel-based approach to social network-based recommendation. *Decision Support Systems*, *65*, 95–104.

Liu, C. C., Fan, C. S. H., Chou, C. Y., & Chen, S. Y. (2010). Knowledge exploration with concept association techniques. *Online Information Review*, *34*(5), 786–805.

Liu, J., Wu, C., & Liu, W. (2013). Bayesian probabilistic matrix factorization with social relations and item contents for recommendation. *Decision Support Systems*, *55*(3), 838–850.

Manouselis, N., Vuorikari, R., & Van Assche, F. (2010). Collaborative recommendation of e-learning resources: an experimental investigation. *Journal of Computer Assisted Learning*, *26*(4), 227–242.

Shi, L., Zhao, W. X., & Shen, Y. D. (2017). Local representative-based matrix factorization for cold-start recommendation. *ACM Transactions on Information Systems (TOIS)*, *36*(2), 1–28.

Shin, N., & Chan, J. K. (2004). Direct and indirect effects of online learning on distance education. *British Journal of Educational Technology*, *35*(3), 275–288.

Tang, X., & Zhou, J. (2012). Dynamic personalized recommendation on sparse data. *IEEE Transactions on Knowledge and Data Engineering*, *25*(12), 2895–2899.

Wang, M. (2007). Designing online courses that effectively engage learners from diverse cultural backgrounds. *British Journal of Educational Technology*, *38*(2), 294–311.

Wang, Q., Woo, H. L., Quek, C. L., Yang, Y., & Liu, M. (2012). Using the Facebook group as a learning management system: an exploratory study. *British Journal of Educational Technology*, *43*(3), 428–438.

Warburton, S. (2009). Second life in higher education: assessing the potential for and the barriers to deploying virtual worlds in learning and teaching. *British Journal of Educational Technology*, *40*(3), 414–426.

Wei, J., He, J., Chen, K., Zhou, Y., & Tang, Z. (2017). Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Systems with Applications*, *69*, 29–39.

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1492–1500). Honolulu, HI: IEEE.

Zhang, W. Y., Perris, K., & Yeung, L. (2005). Online tutorial support in open and distance learning: students' perceptions. *British Journal of Educational Technology*, *36*(5), 789–804.

Zhao, X., Yuan, J., Wang, M., Li, G., Hong, R., Li, Z., & Chua, T. S. (2013). Video recommendation over multiple information sources. *Multimedia Systems*, *19*(1), 3–15.

Zheng, Y., & Yano, Y. (2007). A framework of context-awareness support for peer recommendation in the e-learning context. *British Journal of Educational Technology*, *38*(2), 197–210.