

# Consideration of the Local Correlation of Learning Behaviors to Predict Dropouts from MOOCs

Yimin Wen\*, Ye Tian, Boxi Wen, Qing Zhou, Guoyong Cai, and Shaozhong Liu

**Abstract:** Recently, Massive Open Online Courses (MOOCs) have become a major online learning methodology for millions of people worldwide. However, the dropout rates from several current MOOCs are high. Usually, dropout prediction aims to predict whether a learner will exhibit learning behaviors during several consecutive days in the future. Therefore, the information related to the learning behaviors of a learner in several consecutive days should be considered. After in-depth analysis of the learning behavior patterns of the MOOC learners, this study reports that learners often exhibit similar learning behaviors on several consecutive days, i.e., the learning status of a learner for the subsequent day is likely to be similar to that for the previous day. Based on this characteristic of MOOC learning, this study proposes a new simple feature matrix for keeping information related to the local correlation of learning behaviors and a new Convolutional Neural Network (CNN) model for predicting the dropout. Extensive experimental validations illustrate that the local correlation of learning behaviors should not be neglected. The proposed CNN model considers this characteristic and improves the dropout prediction accuracy. Furthermore, the proposed model can be used to predict dropout temporally and early when sufficient data are collected.

**Key words:** Massive Open Online Courses (MOOCs); dropout prediction; local correlation of learning behaviors; Convolutional Neural Network (CNN); educational data mining

## 1 Introduction

Massive Open Online Courses (MOOCs) are widely regarded as a new revolution of education, and has

- Yimin Wen, Ye Tian, and Guoyong Cai are with Guangxi Key Laboratory of Trusted Software and the School of Computer Science and Information Security, Guilin University of Electronic Technology, Guilin 541004, China. E-mail: ymwen2004@aliyun.com; 976022567@qq.com; ccgycai@gmail.com.
- Boxi Wen is with the School of Business, Guilin University of Electronic Technology, Guilin 541004, China. E-mail: wenboxi@aliyun.com.
- Qing Zhou is with the College of Computer Science, Chongqing University, Chongqing 400044, China. E-mail: tzhou@cqu.edu.cn.
- Shaozhong Liu is with the College of Foreign Studies, Guilin University of Electronic Technology, Guilin 541004, China. E-mail: shaozhong@hotmail.com.

\* To whom correspondence should be addressed.

Manuscript received: 2019-03-31; accepted: 2019-04-04

become a hot cross research subject of education, psychology, information technology, and data science<sup>[1]</sup>. The joint efforts of academia and industry have led to the recent development of multiple MOOC platforms, such as Coursera, Udacity, Edx, and XuetangX, to adequately address diverse learning needs and cater to service learners by providing thousands of well-designed online courses. In a typical MOOC platform, learners can not only access speech videos, assignments, and examinations, but also use learning tools, such as online discussion forums and wiki, for participating in peer-to-peer interactions. MOOC has become the main choice of online learning for millions of people worldwide.

Unlike traditional education, learners with extensively different motivations exhibiting various participation styles enroll in MOOCs. Many of them have insufficient amount of incentives for completing their courses. Consequently, MOOCs are infamous for their high dropout rate with a completion rate of less than 10%

with respect to majority of the courses<sup>[2,3]</sup>. The low completion rate has become a major hurdle for further development of MOOCs. Therefore, imminent actions should be performed to accurately identify and predict dropouts. If a teacher detects the likelihood of a learner to drop out, he/she can resolve to timely measures during the teaching process. For instance, he/she can resort to email reminders or provide positive feedbacks to the learners.

“Big data” is becoming a popular topic in the field of education. Educational data mining<sup>[4]</sup> and learning analytics<sup>[5]</sup> aim to understand educational data for achieving enhanced teaching and learning experiences. From the viewpoints of educational data mining and learning analytics, dropout prediction can be usually achieved using supervised machine learning algorithms. However, this process is difficult because of several reasons. First, the MOOC learners may drop out for different reasons<sup>[6]</sup>. Some learners may only attempt to experience MOOC as a new type of learning. In this regard, dropping out is perceived as a natural occurrence after an initial attempt. Some learners may lose interest in the selected courses. When a learner believes that the course is boring, he/she just drops out. Some learners may lack fundamental knowledge for studying the selected courses, therefore, he/she will drop out when faced with difficulties during the learning process. Furthermore, some learners may have insufficient time for attending the selected courses. If he/she is busy with other things, he/she just stops attending the course. Thus, it is difficult to evaluate the reasons associated with such behaviors of the learners. Second, dropout prediction is not a simple classification problem. Usually, dropout prediction aims to predict whether a learner will exhibit a learning behavior during several consecutive days in the future. Learners who exhibit no learning behavior during a future time interval will be labeled as dropout. Otherwise, they will be labeled as retain. Hence, the dropout concept can be simple and convenient. However, the concept of retention may present several different situations. For example, a retained learner may only display learning behavior on the first day of an observation period, on the last day of a time range, or during several consecutive days in a future time interval. Therefore, all the aforementioned learners should be defined as retained even though they are under completely different situations.

Even though many studies have reported good

dropout prediction accuracy, two shortcomings have been observed in the related literature to date. (1) The objective of dropout prediction is to predict whether a learner exhibits a learning behavior during several consecutive days in the future. Therefore, the information associated with the learning behaviors of learners on several consecutive days should be considered. However, this information is often ignored by the current classification models. (2) After analyzing the learning behaviors of the learners, we can observe that their learning status remains the same for several consecutive days, i.e., he/she always learns or exhibits no learning behavior for several consecutive days. This phenomenon, which is referred to as the local correlation of learning behaviors in this study, was often overlooked in related previous discussions.

To overcome these shortcomings, this study aims to predict whether a learner will exhibit any learning behavior in the next 10 days based on his/her learning behaviors in the previous days. If the learner exhibits no learning behavior during the next 10 days, he will be labeled as dropout, otherwise, he will be labeled as retainer. To take advantage of the local correlation characteristics of the aforementioned learning behaviors, the features for each day of the learner’s learning process are extracted based on his/her clickstream logs. We further propose a Convolutional Neural Network (CNN) model to extract high-level features containing the local correlation information of learning behaviors for dropout prediction. To the best of our knowledge, this study is the first to consider the local correlation of learning behaviors for conducting dropout prediction. Further, this study offers the following three novel contributions:

(1) We propose the learning status concept to find the local correlation of the learning behavior.

(2) We propose a novel simple feature matrix and CNN model for dropout prediction. We also analyze the influence of the local correlation of learning behaviors on the prediction accuracy.

(3) We explore the feasibilities of using the proposed CNN model to implement temporally and early dropout prediction.

The remainder of this study can be organized as follows. In the next section, we summarize the previous studies conducted with respect to MOOC dropout prediction. In Section 3, we describe the dataset that has been utilized in this study. The proposed prediction method is presented in Section 4, and the experimental results obtained using the dataset are reported in Section

5. Finally, we conclude this study with a brief discussion of the observations and conclusions.

## 2 Related Work

In this section, we will briefly discuss the studies related to MOOC dropout prediction. Dalipi et al.<sup>[6]</sup> performed a survey of studies conducted for MOOC dropout prediction. They observed that different research groups tended to choose different data sources and extract different features. The clickstream logs, assignment grades, social networks, and even demographic information are frequently used in the literature as potential sources for feature extraction. Further, various classification models are employed for dropout prediction.

Related studies are identified using common classification algorithms to implement supervised learning for dropout prediction. In the study conducted by Kloft et al.<sup>[7]</sup>, a learner was labeled as a dropout when he/she exhibited no learning behavior in a week. Kloft et al. attempted to predict the dropout in the next week based on the data obtained from several previous weeks. The clickstream logs were used as the source data to extract 19 types of features of a learner, such as the number of active days and the number of video views, with respect to a selected course. The features of several weeks were subsequently concatenated. For example, a feature vector with a length of 19 obtained based on the first week was used to predict the dropout in the second week, the features of the first and second weeks were concatenated as a feature vector with a length of 38 to predict the dropout in the third week. The features were subsequently processed using Principal Component Analysis (PCA) and used to train support vector machine for predicting the weekly dropouts.

In the study conducted by Taylor et al.<sup>[8]</sup>, a learner was labeled as a dropout when he/she failed to submit any further assignments or exercise problems. For example, if a learner submitted his/her final assignment in the third week, he/she was considered to drop out in the fourth week. Thus, 27 interpretive features were extracted from the self-proposed, self-extracted, and crowd-proposed covariates, which included the learner's total time spent on all the resources, number of forum posts posted by a learner, number of the weekly homework problems accurately answered by a learner, total number of forum responses posted by all the learners, and learner's average number of submissions

as a percentage of the maximum average number of submissions in that particular week. After defining time slices as weekly units, the features of several weeks were concatenated to construct a feature vector for predicting the dropout. Logistic regression and multiple classifier methods were used in this study.

In the study conducted by Liang et al.<sup>[9]</sup>, a learner was labeled as a dropout when he/she exhibited no learning behavior over the last 10 days, with the data from the first 30 days used as input, and clickstream logs, course information, and enrollment information used as the source data for feature extraction. Overall, 112 features, including the number of total clicks, total times of access, number of registered learners, total engage time of learners, and enrolled time, are extracted. During this process, four common classification models, including logistic regression, support vector machine, random forest, and gradient boosted decision tree, were trained to predict the dropout.

Ramesh et al.<sup>[10]</sup> proposed that the learner activities in forums were good indicators of his/her engagement, which could be classified into active, passive, and disengaged. They leveraged the behavioral, linguistic, temporal, and structural features to train the framework of probabilistic soft logic, which represented the engagement of learners as a latent variable.

Semi-supervised learning algorithms are also used to perform dropout prediction. Similar to the study conducted by Liang et al.<sup>[9]</sup>, Li et al.<sup>[11]</sup> predicted whether a learner will exhibit learning behaviors with respect to a course in the next 10 days based on his/her learning behavior in the first 30 days. By considering the clickstream logs, six types of learning behaviors, including viewing assignments, viewing videos, accessing other objects except videos and assignments, closing page, accessing wiki, and accessing forum, were considered to be the features. The number of learning behavior records in a course was counted on a weekly basis for each learner. Different types of features were obtained from the first 30 days to construct different views. Therefore, a multi-view semi-supervised learning method was proposed to use a co-training method to predict the dropout. Logistic regression, naive Bayes, Support Vector Machine (SVM), and decision tree methods, which utilized the concatenated features, were used for performing comparison.

Several studies applied temporal models for performing dropout prediction. In the study conducted by Balakrishnan et al.<sup>[12]</sup>, a learner was labeled as a

dropout when he/she exhibited no learning behavior during a week. Four types of source data, including the cumulative percentage of the available lecture videos watched, number of threads viewed on the forum, number of posts posted on the forum, and number of times the course progress page was checked, were used to extract the input features. The four defined features and learner's label were used to train the Hidden Markov Model (HMM). In the study conducted by Fei and Yeung<sup>[13]</sup>, a learner was labeled as a dropout when he/she exhibited no learning behavior during a week. Further, the weekly learning behaviors of a learner up to the current week was used to predict dropout in the next week. Seven features of the Coursera courses and five features of the edX courses were extracted based on the browser- or server-side events. By considering the dropout prediction task as sequential labeling, a variant of HMM and two Recurrent Neural Network (RNN) models, including the vanilla RNN network and long short-term memory network, were trained to predict the dropout.

Few studies used deep neural network models for dropout prediction. In the study conducted by Whitehill et al.<sup>[14]</sup>, target label or proxy label was used to denote a dropout. Altogether, 37 features were extracted from the clickstream and personal information data. The values of each feature vector before week  $t$  were summed as a new feature vector to predict the dropout in the  $t$ -th week. Further, a logistic regression model and a fully connected feed-forward network with five hidden layers were trained to predict the dropout. Similar to Refs. [9] and [11], Wang et al.<sup>[15]</sup> predicted whether a learner will exhibit learning behavior in a course during the next 10 days based on his/her learning behaviors within the first 30 days. 186 features were extracted from the raw records and used to train baseline methods such as linear SVM, SVM with Radial Basis Function (RBF) kernel, logistic regression, decision tree, AdaBoost, gradient tree boosting, random forest, and Gaussian naive Bayes. A combination of CNN and RNN with 30 matrices as input, each with a size of  $24 \times 48$ , was trained using mini-batch stochastic gradient descent.

Thus, dropout can be predicted via machine learning methods, and the proposed dropout prediction models can achieve comparable prediction accuracy. The extraction of features is important for dropout prediction, and different methods are proposed for feature extraction. However, some shortcomings can still be identified. First, dropout prediction aims to

predict whether the learners will exhibit a learning behavior during several consecutive days in the future. Therefore, the learning behaviors should be considered with respect to several consecutive days while training the classification models. However, to reduce the length of the feature vector, majority of the existing studies sum the features of different days to obtain the features for a week, which lead to an inevitable overlook of several other details associated with the learner's learning behavior. Second, in the majority of the existing studies, the features of different weeks are concatenated to construct a one-dimensional feature vector in which the correlations between different features are ignored. Considering these shortcomings of the existing studies, this study proposes to extract features for each day during the learning process of learners based on his/her clickstream logs. The row feature vectors for consecutive days from the start to the current day are arranged one below one to construct a feature matrix. A new CNN model is subsequently proposed to extract high-level features containing the local correlation information of learning behaviors for dropout prediction. The proposed method is different from those developed by Whitehill et al.<sup>[14]</sup> and Wang et al.<sup>[15]</sup> The former did not construct a feature matrix for training the proposed deep neural network, whereas the latter only constructed feature matrix from one-hot vectors, which cannot capture the local correlation of learning behaviors.

### 3 Data Set

The experimental dataset was obtained from 39 courses on the XuetangX platform<sup>[16]</sup>. The dataset was used in KDDCup2015<sup>[17]</sup> and contained labeled and unlabeled data. The labeled data included the records for 30 days and a label representing whether a learner exhibited learning behavior in the fourth 10 days. If a learner exhibits no learning behavior during the fourth 10 days, the learner will be labeled as dropout, otherwise, he will be labeled as retainer. The unlabeled data only included the records for 30 days and were used to test the player's model. Only the labeled data are used in this study. As depicted in Fig. 1, the first column represents the enrollment ID that indicates the course selected by the learner, the second column represents the time point for an event, the third column represents the sources of events, the fourth column represents the types of event, and the final column represents the objects that the learners access or navigate. Each type of event

enrollment_id	time	source	event	object
1	2014-06-14T09:38:29	server	navigate	0J6eGzrddg8M1aCtaqlIkY6zruSrb71b
1	2014-06-14T09:38:39	server	access	3T6Xvo1MKgo157cm29Rjy8FXVFc1onx1
1	2014-06-14T09:38:39	server	access	qxv8NHTf1RkNcCvM0hcGw6hvhDQvnd4
1	2014-06-14T09:38:48	server	access	2cmZrZW2b6I1911t03e89FCaBLWhf3W
1	2014-06-14T09:41:49	browser	problem	RMtgC2bTAqEefTenUUYia504wsyzeZVf
1	2014-06-14T09:41:50	browser	problem	RMtgC2bTAqEefTenUUYia504wsyzeZVf
1	2014-06-14T09:42:28	browser	problem	RMtgC2bTAqEefTenUUYia504wsyzeZVf
1	2014-06-14T09:42:30	browser	problem	RMtgC2bTAqEefTenUUYia504wsyzeZVf
1	2014-06-14T09:43:20	browser	problem	RMtgC2bTAqEefTenUUYia504wsyzeZVf
1	2014-06-14T09:43:25	browser	problem	RMtgC2bTAqEefTenUUYia504wsyzeZVf
1	2014-06-14T09:43:25	server	problem	RMtgC2bTAqEefTenUUYia504wsyzeZVf
1	2014-06-14T09:43:40	server	problem	RMtgC2bTAqEefTenUUYia504wsyzeZVf
1	2014-06-14T09:44:29	browser	page_close	3T6Xvo1MKgo157cm29Rjy8FXVFc1onx1
1	2014-06-19T06:21:04	server	navigate	0J6eGzrddg8M1aCtaqlIkY6zruSrb71b
1	2014-06-19T06:21:16	server	access	3T6Xvo1MKgo157cm29Rjy8FXVFc1onx1
1	2014-06-19T06:21:16	server	access	8BopEkeW8JHRxR06g7IH70dTK1nJDJGg
1	2014-06-19T06:21:32	server	access	qxv8NHTf1RkNcCvM0hcGw6hvhDQvnd4
1	2014-06-19T06:21:32	browser	page_close	3T6Xvo1MKgo157cm29Rjy8FXVFc1onx1
1	2014-06-19T06:21:45	server	access	00kCvDvaJhsSkoN6yuhvmxMAJXu8tx6G
1	2014-06-19T06:21:46	browser	page_close	3T6Xvo1MKgo157cm29Rjy8FXVFc1onx1

Fig. 1 Clickstream logs.

indicates a type of learning behavior. The types of event are presented in Table 1. Line 2 in Fig. 1 denotes a learner with enrollment ID 1 who navigated to a course at 9:38:29 on June 14, 2014.

To understand the learning behavior of learners, we count the frequency of different types of event as presented in Table 1 in the experimental dataset, and the results are shown in Table 2. From Table 2, it can be observed that wiki is a resource that is not frequently accessed by the learners. The statistics for dropout and retained learners are presented in Table 3. Only approximately 20% of the learners will continue with their learning pattern after 30 days.

In this study, dropout is denoted by the lack of learning behaviors of the learners in the fourth 10 days. Meanwhile, retained learners include several different cases. Because of the lack of log data for learners in the fourth 10 days, we calculate the number of learners who exhibit learning behavior in the third 10 days as an alternative to reflect the case for the fourth 10 days. In

Table 1 Types of events in the clickstream log.

Event	Description
Access	Accessing other course objects except videos and assignments
Problem	Working on the course assignments
Page_close	Closing the web page
Navigate	Navigating to other part of the course
Video	Watching the course videos
Discussion	Accessing the course forum
Wiki	Accessing the course wiki

Table 2 Frequency of different types of learning behaviors.

Access	Problem	Page_close	Navigate	Video	Discussion	Wiki
3 112 191	1 261 170	1 237 883	1 009 309	796 958	649 259	90 507

Table 3 Statistics of dropout and retained learners.

	Amount	Proportion (%)
Dropout	95 581	79.29
Retain	24 961	20.71

Fig. 2, the horizontal axis indicates the number of days on which a learner exhibits a learning behavior, whereas the vertical axis indicates the number of learners. It can be observed from Fig. 2 that majority of the learners with retainer label only learn one day, whereas other retained learners exhibit learning behaviors on several days. Thus, we can consider that the case for retain is more complex when compared with that for dropout.

## 4 Proposed Method

In this section, we will present our dropout prediction method. The proposed method initially extracts a feature matrix for each pair of learner-course. A new CNN model is proposed for dropout prediction.

### 4.1 Local correlation of learning behaviors

In this section, we analyze the learning behavior pattern of learners during the first 30 days. To ensure convenience, the concept of learning status is proposed in this section. If a learner exhibits a learning behavior on one day, then his/her learning status for that day is set as 1, otherwise the learning status for that day is set as 0. A learner's learning status for a course in 30 days can be represented as a Boolean vector  $S = (s_1, \dots, s_i, \dots, s_{30})$ , where  $s_i$  indicates the learning status on the  $i$ -th day.

Each learner's learning status vector can be further divided into multiple learning status segments. For example, there is a learning status vector of [0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 1, 1], the learning status segments can be expressed as [[0, 0], [1], [0, 0, 0, 0], [1], [0], [1, 1, 1, 1], [0, 0], [1], [0, 0], [1, 1, 1], [0], [1, 1], [0, 0, 0, 0], [1, 1]]. We can observe that each learning

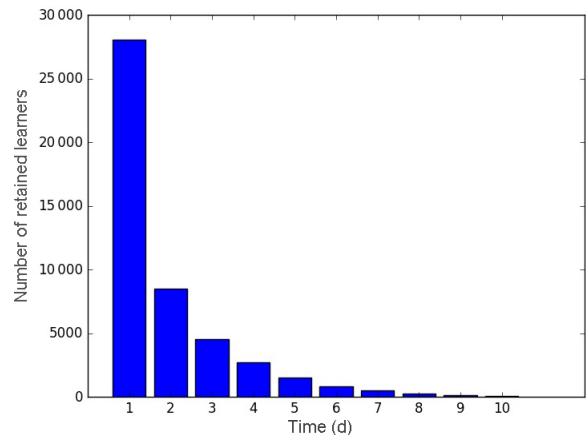
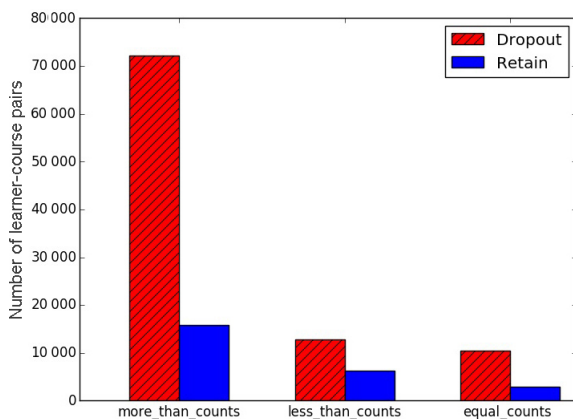


Fig. 2 Number of retained learners with different number of days of learning in the third 10 days.

status segment contains a learning status that is valid for one day or several consecutive days. The learning status segment with respect to multiple days can be referred to as a continuous learning status segment, whereas the learning status segment with respect to only one day can be referred to as a discontinuous learning status segment. For different pairs of learner-course, three cases of learning status segments can be given as follows: the number of continuous learning status segments is larger than the number of discontinuous learning status segments (namely *more\_than\_counts*), the number of continuous learning status segments is less than the number of discontinuous learning status segments (namely *less\_than\_counts*), and the number of continuous learning status segments is equal to the number of discontinuous learning status segments (namely *equal\_counts*). Further, we count the number of learner-course pairs in the three cases for two types of learners. As depicted in Fig. 3, the blue bar represents the dropout learners, whereas the red bar represents the retained learners. The horizontal axis represents three cases, whereas the vertical axis indicates the number of learner-course pairs. Based on Fig. 3, majority of the learners have more continuous learning status segments than discontinuous learning status segments. This phenomenon remains the same for dropout and retained learners. Hence, majority of the MOOC learners tend to maintain the same learning status for several consecutive days.

The transitions of two consecutive learning status are studied to explore the local correlation of learning behaviors. If the learning status of the next day is the same as the previous day, the case can be referred to as the same status transition, while the learning status



**Fig. 3** Statistics of the learning status segments for two types of learners.

of the next day is different with the previous day, the case will be referred to as the different status transition. The number of occurrences and proportions of these two cases for two types of learners are separately presented in Tables 4 and 5. Based on these statistical results, it can be observed that the learning status for the next day has a considerable likelihood of being in agreement with the learning status of the previous day. Hence, the learners tend to maintain the same learning status on adjacent days.

These two aforementioned analyses denote that learners' learning behaviors during several consecutive day exhibit strong correlations, indicating that the learner's learning behaviors are locally correlated. Actually, during learning process, learners are often affected by their own state or external conditions which always influence learners' learning behaviors for several consecutive days.

#### 4.2 Feature extraction

In the previous section, we observed certain local correlations of the learning behaviors. Therefore, an intuitive feature extraction methodology is to count the number of learning behavior records per day. We extracted 7 types (Table 1) of behavior features from the clickstream logs. Dropout prediction is defined as a binary classification problem, specifically, each  $\langle X, y \rangle$  represents a sample and its label. If a learner drops out from a course, the sample will be labeled as  $y = 1$ , otherwise  $y = 0$ .  $X$  represents a feature matrix denoted in Fig. 4. The  $i$ -th row represents the seven types of behavior features on the  $i$ -th day, each  $X_{ij}$  represents the count of the  $j$ -th learning behavior on the  $i$ -th day for a learner.

An ideal image feature preprocessing method is borrowed to normalize  $X$ . For a grayscale image, the maximum value of pixels is 255, and the value of each pixel is divided by 255 to obtain normalized data. However, in our problem, for majority of the learners, the count of certain type of learning behavior per day

**Table 4** Status transition for dropout learner.

	Occurrence	Proportion (%)
Same status transition	2 503 166	90
Different status transition	268 683	10

**Table 5** Status transition for retained learner.

	Occurrence	Proportion (%)
Same status transition	548 825	76
Different status transition	175 044	24

$X_{1,1}$	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	$X_{1,5}$	$X_{1,6}$	$X_{1,7}$
$X_{2,1}$	$X_{2,2}$	$X_{2,3}$	$X_{2,4}$	$X_{2,5}$	$X_{2,6}$	$X_{2,7}$
$X_{3,1}$	$X_{3,2}$	$X_{3,3}$	$X_{3,4}$	$X_{3,5}$	$X_{3,6}$	$X_{3,7}$
...						
$X_{30,1}$	$X_{30,2}$	$X_{30,3}$	$X_{30,4}$	$X_{30,5}$	$X_{30,6}$	$X_{30,7}$

Fig. 4 Feature matrix.

is relatively small, while some learners exhibit a very large number of learning activities on a single day. In this case, dividing each value of the feature matrix by the maximum value will cause majority of the values to become approximately zero. Therefore, we choose a normalization parameter to represent the maximum value of all the features. The values of all the features larger than this parameter are truncated to this parameter. By setting different parameters to try, we obtain a suitable normalization parameter, that is 240. The values of all the features larger than 240 are truncated to 240, and each of the remaining elements  $X_{ij}$  are set as  $X_{ij} = X_{ij}/240$ .

### 4.3 Construction of the CNN

To take advantage of the local correlation characteristic of the learning behaviors, we use CNN to extract high-level features containing the local correlation information and make dropout prediction. By considering that no related work conducted to use CNN model for predicting the dropout, we have obtained a network structure through continuous attempts. The proposed CNN model structure is depicted in Fig. 5.

The proposed CNN model comprises six layers, in addition to the layer for input, all layers contain trainable parameters (weights). We take the feature matrix proposed in the previous Section 4.2 as the input. The size of input is  $30 \times 7$ . Layer C1 is a

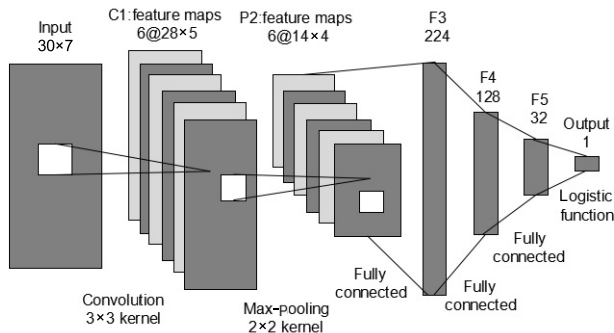


Fig. 5 Proposed CNN model.

convolutional layer with six feature maps having the size of  $28 \times 5$ . Each unit in each feature map is connected to a  $3 \times 3$  neighborhood in the layer of input. In a feature map, all the units share the same set of nine weights and the same bias. Therefore, the same feature is detected at all possible locations in the input. Different feature maps in the layer use different sets of weights and biases, thereby extracting different types of local features. Layer P2 is a max-polling layer with 6 feature maps with the size of  $14 \times 4$ . Each unit in each feature map is connected to a  $2 \times 2$  neighborhood in the corresponding feature map in C1. The step for row is set as 1 while the step for column is set as 2. The maximum of 4 values in C1 is input into the unit in P2. Layer F3 contains 224 units, layer F4 contains 128 units, and layer F5 contains 32 units, which are all fully connected to previous layer. Finally, the logistic function is used in the output layer to complete the classification task.

We consider the strategy of using the Root Mean Square prop (RMSprop) as optimizer to train the parameters in the proposed network. The Rectified linear unit (Relu)<sup>[18]</sup> function is used as the activation function in the C1, F3, F4, and F5 layers. To solve overfitting, we considered the dropout technique for F3 and F4, the key idea is to randomly drop units (along with their connections) from the neural network during training, this can prevent units from considerably co-adapting<sup>[19]</sup>.

## 5 Experiments and Discussions

### 5.1 Evaluation criterion

Variables such as the precision, recall, F-measure, and accuracy, are considered as metrics to evaluate the performance of the proposed method. They are calculated using Eqs. (1)–(4), respectively, based on confusion matrix.

$$precision = \frac{TP}{FP + TP} \quad (1)$$

$$recall = \frac{TP}{FP + FN} \quad (2)$$

$$F\_measure = \frac{2 \times precision \times recall}{precision + recall} \quad (3)$$

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (4)$$

Confusion matrix is always used to evaluate the classification accuracy of a method. In the binary classification problem, the confusion matrix is a matrix with four entries such as True Positive (TP), False Negative (FN), False Positive (FP), and True Negative

(TN), which are presented in Table 6. TP denotes the number of dropout learners who have been accurately predicted, FP denotes the number of retained learners who have been predicted as dropout, FN denotes the number of dropout learners who have been predicted as retainer, and TN denotes the number of retained learners who have been accurately predicted.

Based on the confusion matrix, precision denotes the proportion of true dropout learners from among the learners who have been predicted as dropout, whereas recall represents the proportion of true dropout learners from among all the dropout learners. In general, precision and recall suppress each other whereas F-measure represents the harmonic mean of precision and recall. Accuracy represents the proportion of learners who have been accurately predicted from among all the learners.

## 5.2 Baselines

The experimental results of seven common classification algorithms will be considered as baselines to verify the effectiveness of the proposed methods. These classification algorithms are commonly used in dropout prediction, including Classification And Regression Tree (CART), Naive Bayes (NB), Linear Discriminant Analysis (LDA), Logical Regression (LR), SVM, Random Forest (RF), and Gradient Boosted Decision Tree (GBDT). The parameters of some algorithms are presented in Table 7. The  $n\_estimator$  is the number of base classifiers in RF and GBDT.  $C$  and  $\gamma$  are the parameters of SVM and are set as the default values of the parameters in libsvm<sup>[20]</sup>, and  $num\_feature$  is the number of features of the samples. In this study,  $num\_feature$  is set as 210.

For the aforementioned common classification algorithms, each  $\langle X, y \rangle$  represents a sample and its label. Different with the aforementioned feature matrix,

$X$  denotes a one-dimensional feature vector of length 210, which all the rows in the feature matrix are sequentially concatenated.  $X$  is normalized by the z-score before training, it makes the mean input approximately 0 and the variance approximately 1, which will accelerate learning.

The main difference between the common classification algorithms and the proposed method is that the feature vector with one dimension is used in the former, whereas the feature matrix is used in the latter. Tenfold cross-validation is used to compute the prediction results.

## 5.3 Experiment results

In this section, we aim to answer the following questions based on the experimental results: how about the performance of dropout prediction based on the proposed CNN is? Does the local correlation of learning behaviors in the feature matrix play a role in improving predictive performance? And can the proposed method predict the dropout at any learning stage?

In Fig. 6, 40 days are divided into four stages, each stage includes 10 days. The learning behavior records of the first three stages are known, and the dropout in the fourth stage has to be predicted. Based on the experimental results presented in Table 8, the proposed CNN model achieves the best results in case of F-measure and accuracy, and the second best result in case of precision. RF obtains the best result for precision, however its recall is not good. LDA obtains the best result for recall, however its precision is not good. The average values of the baseline methods with respect to each evaluation metric are computed to compare with the results of the proposed CNN model by considering



Fig. 6 Description of the prediction task.

Table 8 Experiment results.

Method	Precision	Recall	F-measure	Accuracy
CART	0.8807	0.8868	0.8837	0.8150
NB	0.8813	0.9247	0.9024	0.8414
GBDT	0.8892	0.9626	0.9244	0.8751
LDA	0.8597	<b>0.9771</b>	0.9147	0.8555
LR	0.8736	0.9675	0.9181	0.8632
RF	<b>0.8985</b>	0.9446	0.9209	0.8714
SVM	0.8836	0.9595	0.9200	0.8676
Average	0.8809	0.9461	0.9120	0.8556
CNN	0.8938	0.9579	<b>0.9247</b>	<b>0.8764</b>

Table 6 Confusion matrix.

	Predicted as dropout	Predicted as retainer
Actually dropout	TP	FN
Actually retain	FP	TN

Table 7 Parameters setting.

Method	Parameter	Value
SVM	$C, \gamma$	$C = 1, \gamma = \frac{1}{num\_feature}$
RF	$n\_estimator$	500
GBDT	$n\_estimator$	500

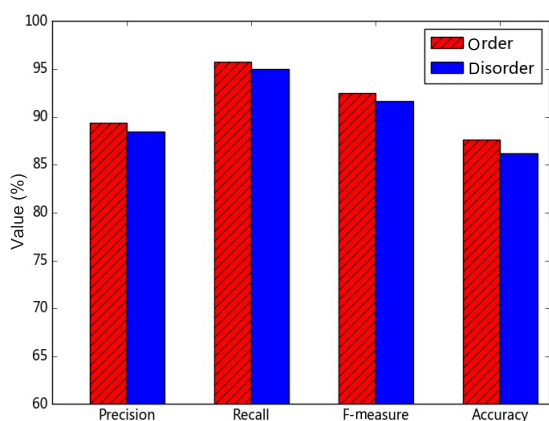


the selected parameters for the baseline methods may not be optimal. It can be observed that the proposed CNN model outperforms the average of all baseline methods with respect to all the evaluation metrics.

The superior performance of the proposed CNN model when compared with those of all the baseline methods could be attributed to the fact that the features are concatenated by the time order. The baseline methods treat all the dimensions of feature vector as independent individuals and fail to take advantage of the local correlation of learning behaviors.

In the previous analysis, we have observed that a learner's learning behaviors during the MOOC learning process exhibit local correlation, i.e., the learning behaviors on several consecutive days exhibit a strong correlation. In the feature matrix, each row feature vector represents the number of learning behaviors on a single day. The row feature vectors for different days are arranged one below one according to the time order. Thus, the local correlation of the learning behaviors is kept in the feature matrix. If the row feature vectors for different days in the feature matrix are randomly disordered, i.e., the local correlation of the learning behaviors may be destroyed, then whether the proposed CNN can still obtain best results for dropout prediction remains unknown.

To explore the aforementioned problem, we randomly disorder the row feature vectors for different days in the feature matrix with 100 times to obtain the average results of 100 experiments. The results are computed by using 10 cross-validation. In Fig. 7, the average results are compared with those of the proposed CNN model with the original ordered features. As depicted in Fig. 7, all the evaluation metrics decline after the destruction of the local correlation of the learning behaviors. Such



**Fig. 7** Comparison of the dropout predictions when ordered or disordered feature matrix is used as the input.

experimental results illustrate that the local correlation of the learning behaviors should not be ignored and the proposed CNN model can take advantage of the local correlation of the learning behaviors.

In actual applications, we may achieve temporal dropout prediction, i.e., predicting the dropout of the next stage according to the known stages. In order to verify the effectiveness of the proposed method in this scenario, experiments were performed in two situations. The descriptions of the prediction tasks are provided in Figs. 8 and 9. Figure 8 shows the dropout prediction for the second stage based on the learning records of the first stage, whereas Fig. 9 indicates the dropout prediction for the third stage based on the learning records of the first two stages. In another scenario, we may wish to predict dropout in the final stage as early as possible. Experiments are also conducted in two situations. The description of the prediction tasks are depicted in Figs. 10 and 11. Figure 10 indicates the dropout prediction for the fourth stage based on the learning records of the first stage, whereas Fig. 11 shows the dropout prediction for the fourth stage based on the learning records of the first two stages.

The experimental results of Situation 1 and Situation 2 are presented in Tables 9 and 10, respectively. In Table 9, the proposed CNN method achieves the best results for precision and accuracy. In this situation, the proposed CNN model exhibits a precision that is approximately 10% better than the average precision, but exhibits a recall that is 10% less than the average recall. These means that the proposed CNN model is inadequate for detecting the majority of dropout cases, and is not likely



**Fig. 8** Temporal dropout prediction in Situation 1.



**Fig. 9** Temporal dropout prediction in Situation 2.



**Fig. 10** Early dropout prediction in Situation 3.



**Fig. 11** Early dropout prediction in Situation 4.

**Table 9** Experimental results of temporal dropout prediction in Situation 1.

Method	Precision	Recall	F-measure	Accuracy
CART	0.7543	0.5397	0.6292	0.6452
NB	0.6112	0.8178	0.6995	0.6082
GBDT	0.7513	0.6470	0.6882	0.6755
LDA	0.6005	<b>0.8869</b>	<b>0.7161</b>	0.6077
LR	0.6016	0.8693	0.7111	0.6059
RF	0.7947	0.5813	0.6714	0.6826
SVM	0.7813	0.6138	0.6875	0.6887
Average	0.6992	0.7079	0.6861	0.6448
CNN	<b>0.7969</b>	0.5975	0.6828	<b>0.6903</b>

**Table 10** Experimental results of temporal dropout prediction in Situation 2.

Method	Precision	Recall	F-measure	Accuracy
CART	0.8371	0.8507	0.8438	0.8087
NB	0.6777	0.8383	0.7495	0.6596
GBDT	0.8372	0.9202	0.8767	0.8428
LDA	0.6598	0.9236	0.7697	0.6643
LR	0.6608	0.8922	0.7593	0.6563
RF	<b>0.8641</b>	0.9137	0.8882	0.8603
SVM	0.8529	0.9360	0.8925	0.8631
Average	0.7699	0.8963	0.8256	0.7650
CNN	0.8522	<b>0.9401</b>	<b>0.8937</b>	<b>0.8642</b>

to predict retained learners as dropout learners. The experimental results of Situation 2 are presented in Table 10. From Table 10, it can be observed that the proposed CNN method achieves the best results with respect to recall, F-measure, and accuracy. When compared with the average, the proposed CNN model achieves the best results with respect to all of the evaluation metrics.

The experimental results of Situation 3 and Situation 4 are presented in Tables 11 and 12, respectively. In Table 11, the proposed CNN model only achieves the best result for accuracy. However, when compared with the average, the proposed CNN model achieves the best results for precision, recall, F-measure, and accuracy. In Table 12, the proposed CNN model achieves the best

**Table 11** Experimental results of early dropout prediction in Situation 3.

Method	Precision	Recall	F-measure	Accuracy
CART	0.8285	0.9045	0.8648	0.7759
NB	<b>0.8611</b>	0.8347	0.8477	0.7622
GBDT	0.8293	0.9705	<b>0.8943</b>	0.8182
LDA	0.8290	0.9678	0.8930	0.8162
LR	0.8236	<b>0.9751</b>	0.893	0.8147
RF	0.8365	0.9506	0.8899	0.8136
SVM	0.8320	0.9624	0.8925	0.8161
Average	0.8342	0.9379	0.8821	0.8024
CNN	0.8358	0.9600	0.8936	<b>0.8188</b>

**Table 12** Experimental results of early dropout prediction in Situation 4.

Method	Precision	Recall	F-measure	Accuracy
CART	0.8511	0.8904	0.8703	0.7895
NB	<b>0.8884</b>	0.8599	0.8739	0.8033
GBDT	0.8579	0.9621	0.9070	0.8436
LDA	0.8529	0.9644	0.9052	0.8399
LR	0.8511	<b>0.9672</b>	0.9054	0.8398
RF	0.8677	0.9425	0.9036	0.8405
SVM	0.8602	0.9560	0.9055	0.8419
Average	0.8613	0.9346	0.8958	0.8283
CNN	0.8639	0.9558	<b>0.9075</b>	<b>0.8456</b>

results for F-measure and accuracy. When compared with the average, the proposed CNN model achieves the best results for precision, recall, F-measure, and accuracy.

From Tables 9 to 12, we can conclude that the amount of historical data will influence the dropout prediction results. In general, there are more historical data in Figs. 9 and 11, the local correlation of the learning behaviors can be accurately captured when large amount of data are used. We select the dropout prediction results of the proposed CNN model from Tables 8, 10, and 12 to construct Table 13. It is found very interesting from Table 13 that three stages maybe enough for dropout prediction. Thus, we can guess that this is why only three stages of historical data are provided in KDDCup2015.

## 6 Conclusion

In this study, we probed alternatives for explaining dropout prediction to improve completion rates of MOOCs. We conducted detailed analysis of the learning behavior patterns of MOOC learners. The results denote that learners often exhibit similar learning behaviors on several consecutive days, i.e., the learning status of a learner on the next day has a considerable likelihood of being similar to the learning status of the learner on the previous day. To the best of our knowledge, this study is an initial endeavor to address the local correlation of the learning behaviors. A new simple feature matrix is proposed for retaining the information of the local correlation of learning behaviors to maximize the usage of this characteristic of MOOC learning. A new CNN

**Table 13** Results of dropout prediction using the proposed CNN model in different scenarios.

Scenario	Precision	Recall	F-measure	Accuracy
CNN in Fig. 6	0.8938	0.9579	0.9247	0.8764
CNN in Fig. 9	0.8522	0.9401	0.8937	0.8642
CNN in Fig. 11	0.8639	0.9558	0.9075	0.8456

model is constructed for extracting high-level features pertaining to the local correlation of learning behaviors and dropout prediction.

Extensive experimental validations were conducted to denote that the local correlation of learning behaviors should not be ignored, the proposed CNN model can take reasonable advantage of this characteristic and improve dropout prediction. Moreover, the proposed CNN model can be used for temporal dropout prediction and early dropout prediction once sufficiently large amount of data are obtained. In our future study, we plan to construct more reasonable network structures to predict the dropout, and further explore the correlations between parallel MOOCs selected by learners.

### Acknowledgment

This work was partially supported by the National Natural Science Foundation of China (Nos. 61866007, 61363029, 61662014, 61763007, and U1811264), the Natural Science Foundation of Guangxi District (No. 2018GXNSFDA138006), Guangxi Key Laboratory of Trusted Software (No. KX201721), Humanities and Social Sciences Research Projects of the Ministry of Education (No. 17JDGC022), Chongqing Higher Education Reform Project (No. 183137), and we would like to gratefully acknowledge the organizers of KDDCup2015 as well as XuetangX for making the datasets available.

### References

- [1] R. Wescrues, G. M. Henricks, M. Perry, S. Bhat, C. J. Anderson, N. Shaik, and L. Angrave, How do gender, learning goals, and forum participation predict persistence in a computer science MOOC?, *ACM Transactions on Computing Education*, vol. 18, no. 4, pp. 1–18, 2018.
- [2] A. Ramesh, D. Goldwasser, B. Huang, H. Daumé III, and L. Getoor, Learning latent engagement patterns of students in online courses, in *Proc. the Twenty-Eighth AAAI Conference on Artificial Intelligence*, Québec, Canada, 2014, pp. 1–7.
- [3] J. He, J. Bailey, B. I. P. Rubinstein, and R. Zhang, Identifying at-risk students in massive open online courses, in *Proc. the Twenty-Ninth AAAI Conference on Artificial Intelligence*, Austin, TX, USA, 2015, pp. 1749–1755.
- [4] P. A. Alejandro, Educational data mining: A survey and a data mining-based analysis of recent works, *Expert Systems with Application*, vol. 41, no. 4, pp. 1432–1462, 2014.
- [5] M. Vahdat, A. Ghio, L. Oneto, D. Anguita, M. Funk, and M. Rauterberg, Advances in learning analytics and educational data mining, in *Proc. European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, Belgium, 2015, pp. 22–24.
- [6] F. Dalipi, A. S. Imran, and Z. Kastrati, MOOC dropout prediction using machine learning techniques: Review and research challenges, in *Proc. IEEE Global Engineering Education Conference*, Islas Canarias, Spain, 2018, pp. 1007–1014.
- [7] M. Kloft, F. Stiehler, Z. Zheng, and N. Pinkwart, Predicting MOOC dropout over weeks using machine learning methods, in *Proc. the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, Doha, Qatar, 2014, pp. 60–65.
- [8] C. Taylor, K. Veeramachaneni, and U. O'Reilly, Likely to stop? Predicting stop out in massive open online courses, <https://arxiv.org/abs/1408.3382?context=cs.CY>, 2014.
- [9] J. Liang, J. Yang, Y. Wu, C. Li, and L. Zheng, Big data application in education: Dropout prediction in edX MOOCs, in *Proc. IEEE Second International Conference on Multimedia Big Data*, Taipei, China, 2016, pp. 440–443.
- [10] A. Ramesh, D. Goldwasser, B. Huang, H. Daumé III, and L. Getoor, Modeling learner engagement in MOOCs using probabilistic soft logic, in *Proc. NIPS Workshop on Data Driven Education*, Lake Tahoe, NV, USA, 2013, pp. 21–62.
- [11] W. Li, M. Gao, H. Li, Q. Xiong, J. Wen, and Z. Wu, Dropout prediction in MOOCs using behavior features and multi-view semi-supervised learning, in *Proc. International Joint Conference on Neural Networks*, Vancouver, Canada, 2016, pp. 3130–3137.
- [12] G. Balakrishnan, G. K. Balakrishnan, and D. Coetzee, Predicting student retention in massive open online courses using hidden markov models, Technical Report No. UCB/EECS-2013-109, Electrical Engineering and Computer Sciences University of California, Berkeley, CA, USA, 2013.
- [13] M. Fei and D. Yeung, Temporal models for predicting student dropout in massive open online courses, in *Proc. IEEE International Conference on Data Mining Workshop*, Atlantic City, NJ, USA, 2015, pp. 256–263.
- [14] J. Whitehill, K. Mohan, D. Seaton, Y. Rosen, and D. Tingley, Delving deeper into MOOC student dropout prediction, <https://arxiv.org/pdf/1702.06404.pdf>, 2017.
- [15] W. Wang, H. Yu, and C. Miao, Deep model for dropout prediction in MOOCs, in *Proc. the 2nd International Conference on Crowd Science and Engineering*, Beijing, China, 2017, pp. 26–32.
- [16] XuetangX, <http://www.xuetangx.com/>, 2019.
- [17] Kddcup2015, <https://biendata.com/competition/kddcup2015/>, 2015.
- [18] A. Krizhevsky, I. Sutskever, and D. E. Hinton, ImageNet classification with deep convolutional neural networks, in *Proc. of the Neural Information Processing Systems Conference*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.
- [19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [20] C. C. Chang and C. J. Lin, LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems & Technology*, vol. 2, no. 3, pp. 1–27, 2011.



**Yimin Wen** received the PhD degree from Shanghai Jiao Tong University, Shanghai, China, in 2007. He is currently a professor at the School of Computer Science and Information Safety, Guilin University of Electronic Technology. His research interests include machine learning, pattern recognition, media and image mining, and educational data mining.



**Ye Tian** is a master student at Guilin University of Electronic Technology, Guilin, China, in 2018. His research interests include data mining and educational data analysis.



**Boxi Wen** is currently a master student in Business School of Guilin University of Electronic Technology, His research interests include data mining and human resources management.



**Qing Zhou** received the PhD degree from Chongqing University, Chongqing, China, in 2008. He is currently a professor in the College of Computer Science at Chongqing University. His research interests include machine learning and educational data mining.



**Guoyong Cai** received the PhD degree from Zhejiang University, Hangzhou, China, in 2008. He is currently a professor in School of Computer Science and Information Safety, Guilin University of Electronic Technology. His research interests include machine learning, data mining, and social network analysis.



**Shaozhong Liu** received the PhD degree in high education and teacher development from the University of North Carolina at Greensboro in 2011 and the PhD degree in linguistics and applied linguistics from Guangdong University of Foreign Studies, in 1997. He is currently a professor in the College of Foreign Studies, Guilin University of Electronic Technology. His research interest is linguistic issues in higher education settings.