Educational Data Mining: Dropout Prediction in XuetangX MOOCs

Chengjun Xu[1] · Guobin Zhu[2] · Jian Ye[3] · Jingqian Shu[4]

Abstract

With the rapid development of educational data mining and learning analytics, this study tries to make sense of education data and improve teachers' competence and teaching experience. In recent years, massive open online courses (MOOCs) have become the first choice of online learning for tens of millions of people around the world. However, the dropout rates for MOOCs are high. The goal of dropout prediction is to predict whether learners will exhibit learning behavior in several consecutive days in the future. Therefore, in this study, we consider the correlation information of learners' learning behaviors for several consecutive days. Through the in-depth statistical analysis of learners' learning behavior, it is found that learners' learning behavior on the next day is similar to that of the previous day. Based on this characteristic, we propose a Lie group region covariance matrix to represent the local correlation information of learning behavior and construct a convolutional neural network model with a multidilation pooling module to extract the local correlation high-level features of learning behavior for dropout prediction. In addition, extensive experiments show that the local correlation of learners' learning behavior cannot be ignored, which is fully considered in our model. Compared with the existing methods, our method achieves the best experimental results in accuracy, F-measure, precision, and recall, which is better than the current methods.

Keywords Convolutional neural network (CNN) model · Dropout prediction · Lie group · Local correlation of learning behaviors · Massive open online courses (MOOCs)

✉ Chengjun Xu
jjtu_001@163.com

1   School of Software, Jiangxi Normal University, Nanchang, China

2   School of Remote sensing and Information Engineering, Wuhan University, Wuhan, China

3   Jiangxi Teachers College, Yingtan, China

4   Nanchang University College of Science and Technology, Nanchang, China

1 Introduction

Massive open online courses (MOOCs) are widely recognized as a revolution in the field of education, along with interdisciplinary research topics in the field of education, information technology, psychology, pedagogy, and artificial intelligence [1]. With the rapid development of science, technology, academia, and industry, many MOOC platforms have emerged, such as Coursera, EdX, Udacity, and XuetangX. These platforms can provide a variety of well-designed courses to meet the different learning requirements of learners. On the MOOC platform, learners can not only access assignments, transcripts, and video lectures, but also communicate with other learners through online forums and wikis. Therefore, MOOC has become the first choice of online learning for tens of millions of people around the world.

Different from traditional education, MOOC learners have extensively different motivations, thus exhibiting various ways of participation. Most learners do not have enough perseverance to complete their courses. One of the major concerns of MOOC is the high dropout rate. According to statistics, the completion rate of most courses is less than 10% [2,3], which means that the vast majority of learners who have registered for the course from the beginning will eventually fail to complete their studies. Scholars are trying to find out the reasons why learners fail to complete the course and take timely actions to accurately identify and predict dropouts. If teachers find that learners are likely to drop out at an early stage, they can take timely measures during the teaching process, such as providing positive feedback to the learner's email or voice.

In the field of education, Big Data has become a hot topic. Educational Data Mining [4] and Learning Analytics [5] are trying to make sense of educational data and further improve teaching experience to achieve enhanced teaching and learning experience. From the above two perspectives, the dropout situation can be predicted by supervised learning method. However, due to the following practical reasons, this process becomes very difficult. Firstly, dropout prediction is not a simple classification problem. The goal of dropout prediction is to predict whether a learner will exhibit a learning behavior in the next few consecutive days, and label those who do not exhibit learning behavior as dropouts. Otherwise, it is labeled as retain. Therefore, from the above analysis, we find that the concept of dropout is simple and convenient, but the concept of retain is very complex. For example, the learners of retain may only have learning behavior in a certain period and certain days. Secondly, MOOC learners drop out for a variety of reasons [6]. Some learners are just trying to get a taste of the new MOOC learning style, and dropping out is a natural consequence of trying. Another part of learners may lack the fundamental knowledge of the chosen course and drop out when they faced with difficulties in the learning process. Some learners may find the course boring and lose interest, they will drop out. In addition, some learners do not have enough time to study due to work and other reasons, they just stop studying the course temporarily. Therefore, it is difficult to evaluate learners' behavior.

Although there are many good dropout prediction models in existing studies, the following deficiencies still exist:

(1) Most of the current studies based on static scenario design and implementation cannot flexibly handle dynamic changes and ignore the local relevance of learning behaviors.
(2) The current prediction model ignores the information related to learners' learning behaviors for several consecutive days. We note that learners' learning behavior usually remains the same for several days, for example, learners have no learning behavior for several days.

To address the above problems, this study focuses on predicting whether learners will exhibit learning behaviors in the next 10 days by learning behaviors in the previous days. If the learner does not exhibit any learning behavior within the next 10 days, it will be labeled as dropout, otherwise, it will be labeled as a retainer. In this study, the local correlation features and time changes of learning behavior are fully considered, and a convolutional neural network (CNN) model with a multidilation pooling module is further proposed to extract high-level features with learning behavior and time changes, which are specifically used for dropout prediction.

The main contributions of this study are as follows:

(1) A novel Lie Group regional covariance feature matrix and a CNN model with a multidilation pooling module are proposed, which are specially used for dropout prediction.
(2) We explored the possibility of using the CNN model to implement temporally and early dropout prediction.
(3) We carry out some extensive experiments on the data set. Experimental results show that our method is superior to competitive methods and improves the accuracy of prediction.

The remainder paper is organized as follows. In Sect. 2, we briefly review recent advances that work on previous studies. Section 3, we introduce our proposed prediction approaches. Section 4, we introduce the dataset. Section 5, we describe the experimental results. Section 6, finally conclude the paper.

2 Related Work

In this section, we briefly introduce the research related to the prediction of dropout by MOOCs. Fisnik Dalipi et al. [6] reviewed recent research on machine learning application toward predicting, explaining, and solving the problem of dropout in MOOCs, and found that different research groups often choose different data sources and extract different characteristics. This study used assignment grades, social networks, clickstream data, and even demographic information as features to extract, and combined with various classification models to predict dropout.

In the existing research, many commonly used classification algorithm models are used to implement supervised learning dropout prediction. Kloft et al. [7] believed that if a learner does not exhibit any learning behavior within 7 days, it will be labeled as a dropout. They used clickstream logs as the data source, they extracted 19 types of characteristics of each learner, such as the number of video views for a course, and attempted to predict the dropout rate in the coming week. Specifically, 19 types of features extracted in the first week were used to predict the dropout learners in the second week, 38 types of features extracted in the first 2 weeks were used to predict dropout learners in the third week, and so on. Then principal component analysis (PCA) is used to process the features, and the support vector machine (SVM) is trained to realize the weekly prediction.

Taylor et al. [8] believed that if a learner failed to submit any exercise or assignments questions, he will be labeled as a dropout. For example, learners submitted final assignments in the sixth week and were considered dropout in the seventh week. They defined the time interval as weeks, combined the features of several weeks, and constructed a feature vector to predict dropouts. Multiple classifications and Logistic regression methods were used in this study. Liang et al. [9] believed that a learner who had no learning behavior over the past 10 days was labeled as a dropout. They used data from the first 30 days as input to the model, which contained 112 features such as enrollment information, number of registered learners, total engage the time of learners and enrolled time. Four common classification

models, SVM, gradient boosted decision trees, logistic regression, and random forest (RF), were used to predict dropouts.

Ramesh et al. [10] proposed that learner's activities in the forum can be divided into disengaged, active, and passive as measurement indicators. They leveraged structural features, linguistic, behavior, and temporal to train and construct probabilistic logical models, taking learner involvement as a potential variable.

Scholars also used a semi-supervised learning method to predict dropout. Similar to the previous study by Liang et al. [9], Li et al. [11] utilized the learning behavior of learners in the first 30 days to predict whether learners would exhibit learning behavior for a certain course in the next 10 days. The features extracted in this method include accessing forums, viewing videos, and accessing wiki. Count the number of learning behavior records per week of each learner, capture different types of features in the first 30 days, and construct different views. Therefore, they proposed a multi-view semi-supervised learning algorithm, using the co-training method to predict dropouts, and compared with SVM, Naive Bayes (NB), decision tree, and logistic regression methods.

Several studies have used temporal model-based prediction of dropouts. Balakrishnan et al. [12] suggested that learners who did not exhibit learning behavior within 7 days were labeled as dropouts. The method in this study uses four types of data, such as the number of threads viewed on the forum, the number of times the course progress page was checked, the cumulative percentage of course videos watched and the number of posts posted on the forum. The Hidden Markov Model (HMM) is trained by using the learner's label and four defined features. Fei and Yeung [13] also considered that if the learner did not exhibit learning behavior within 7 days, he was labeled as a dropout. Furthermore, the method also uses learners' learning behavior in 1 week to predict the dropout in the next week. Five features of the EdX courses and seven features of Coursera are extracted based on the browser- or server-side events. Two new Recurrent Neural Network models (RNN), i.e., long short-term memory network and a variant of HMM are trained to predict dropout.

There are few kinds of research on the prediction of dropouts using deep learning network model. Whitehill et al. [14] extracted 37 features from the personal information and clickstream data and used proxy label or target label to represent dropouts. The values of the various vector before $(t-1)$ weeks are added to a new feature vector to predict the dropout of the tth week. Then, a logistic regression model and a fully connected feed-forward network model with five hidden layers are used to predict dropouts. Similar to the previous [9] and [11], Wang et al. [15] also used the learning behavior of the first 30 days to predict whether learning behavior would be performed in the next 10 days. 186 features were extracted from the raw data for training, such as Radial Basis Function (RBF), logistic regression, RF, and linear SVM. The mini-batch stochastic gradient descent algorithm is used to train the combination of RNN and CNN with 30 matrices as input. The size of each matrix is 24×48.

Marque-Vera et al. [16] used data mining methods to deal with imbalanced behavior and high-dimensional data, to predict dropouts. Kim et al. [17] used a large number of learners' watching video records to summarize various video viewing modes in online learning. Kizilcec et al. [18] combined different types of user data, such as learner demographic data, learner course enrollment data, and learner behavior data, and used an unsupervised machine learning algorithm to divide learners into a few stereotypes. Gerben et al. [19] extracted the behavior features of learners from the previous curriculum, combined with machine learning algorithms to predict dropout. In addition, Xu et al. proposed a variety of prediction and classification methods based on the CNN model [20–22].

According to the above references, machine learning can be used to predict dropout and can achieve better prediction accuracy. Feature extraction is very important for dropout pre-

diction. There are many different feature extraction methods for dropout prediction. However, there are still deficiencies in these approaches. Firstly, most of the existing methods connect the features at different times into a one-dimensional feature vector, which ignores the correlation between different features. Secondly, the goal of dropout prediction is to predict whether learners will exhibit learning behavior in consecutive days in the future. Therefore, continuous-time learning behavior should be considered in the training of the classification model. However, in the existing research approaches, to reduce the length of feature vectors and reduce the amount of calculation, the features in different days are superimposed into the features of a week, and other details related to learners' learning behaviors are directly ignored. To address the deficiency of existing methods, in this study, we put forward based on learners clickstream logs to extract the features of the learners' learning every day, will be the features of the construction of Lie Group region covariance matrix, and then, we put forward a novel kind of CNN model, the extract contains learning behavior of high-level features, dedicated to dropout prediction. Different from the above methods [14,15], we constructed the Lie Group regional covariance matrix and considered the correlation between features.

3 Proposed Approach

In this section, we will introduce our method of predicting dropouts. Firstly, the learning behavior patterns of learners in the first 30 days were statistically analyzed to verify the local correlation of learning behavior. Secondly, a Lie Group regional covariance feature matrix is extracted for each pair of learner-course. Finally, a CNN model with a multidilation pooling module is proposed to predict dropout.

3.1 Local Correlation Analysis

For a more intuitive understanding, we define the concept of the learner's learning status.

Definition 1 If a learner exhibits a learning behavior on a certain day, then his/her learning status is defined as True (T), otherwise, his/her learning status is defined as False (F).

We measured the learning behavior patterns of learners during the first 30 days. We use a vector to represent the learner's learning status in the first 30 days, $LS = (LS_1, LS_2, \cdots, LS_i, \cdots, LS_{30})$, where LS_i represents the learning status of the learner on the ith day.

The vector representation of each learner's learning status can be divided into several learning status subsegments. For example, if a learner's learning status of [T, T, T, F, F, T, T, F, F, F, F, T, T, T, T, F, T, T, F, T, F, F, T, F, T, T, T, F, T, F, T], can be divided into [T, T], [T, F, F, T, T, F], [F, F], [F, T, T, T, T, F], [T], [F, T], [F, F, T, F, T], [T, T], [F], [T, F, T]. Among them, only one day's learning status subsegment is recorded as a discontinuous learning status subsegment, and multiple days of learning status is recorded as continuous learning status. The segmented learning status is valid in one day or several consecutive days.

For different learner-courses, we define three cases of learning status segments:

Definition 2 If $A < B$, it is recorded as less than the count, if $A = B$, it is recorded as equal to the count, if $A > B$, it is recorded as more than the count, where A represents the number of continuous learning status segments, and B represents the number of discontinuous learning status segments.
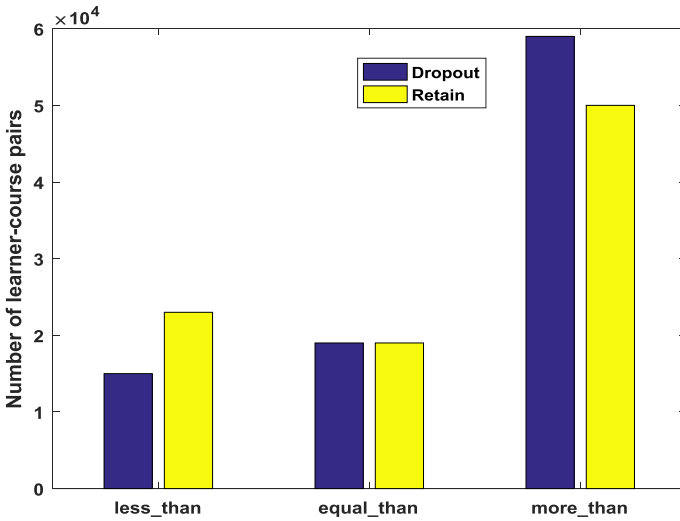
Fig. 1 Statistics of the learning status segments

Table 1 Statistics transition for dropout learner

Types	Occurrence	Proportion(%)
Status transfer	2327645	92
Status transformation	202404	8

We counted the cases in the above three cases, as shown in Fig. 1. Yellow represents the retained learners, blue represents the dropout learners. The horizontal axis represents the three cases, and the vertical axis represents the learner-course data statistics. It can be seen from Fig. 1 that the continuous learning status segment of most learners is larger than the discontinuous learning status segment, which also exists in the retained and dropout learners. Therefore, we believe that most MOOC learners prefer to keep continuous learning status.

The above content describes the two situations of learning state and analyzes the local correlation of learning behaviors. Next, we define the transformation of learning:

Definition 3 If the learning status of the second day is the same as that of the first day, it is recorded as the same status transfer. If the learning status of the second day is different from that of the first day, it is recorded as a different status transformation.

We have calculated the learning status transition, as shown in Tables 1 and 2. From Tables 1 and 2, we find that the probability that the learning status of the second day is the same as that of the first day is very large. Therefore, learners tend to maintain the same learning status and behavior in the adjacent time.

Through the above statistics and analysis, we find that learners' learning behaviors in consecutive days exhibit strong correlation, indicating that learners' behaviors are locally correlated. However, in the real world, learners are affected by various external factors, which often affect their learning behavior for several consecutive days.

Table 2 Statistics transition for retained learner	Types	Occurrence	Proportion (%)
	Status transfer	564618	78
	Status transformation	159251	22

3.2 Regional Covariance Characteristic Matrix of Lie Groups

In the previous section, we found the local correlations of learning behaviors. Therefore, we directly extract the daily learning behavior records of learners, which has the advantage of being more intuitive. We extracted seven types of behavior features from the clickstream logs, as shown in Table 3. Dropout prediction is a binary classification problem. Specifically, each D(L, T) represents the data sample and its label. If the learner drops out from the course, it is labeled as False (F), otherwise it is labeled as True (T).

Firstly, we map the data samples onto the Lie Group manifold space to obtain the Lie Group samples.

$$M_{ij} = log(L_{ij}) \tag{1}$$

where L_{ij} represents the statistical number of the jth learning behavior on the ith day for a learner, and M_{ij} represents the statistical number of the jth learning behavior of a learner on the manifold space of a Lie Group [23].

For the sample data set of the Lie Group, and the attribute of each data can be expressed as $I(x, y)$. In the manifold space of the Lie Group, let F denote the feature extracted from the sample of Lie Group:

$$F(x, y) = \phi(I, x, y) \tag{2}$$

where ϕ represents the mapping function that extracts an $d-dimensional$ feature. For the Lie Group sample set, given rectangular region $R \subset F$, the corresponding ϕ of all elements in the set can be expressed as $\{z_i\}_{i=1}^n$, z_i is a $d-dimensional$ real space vector, and n represents the number of elements in the set. The region R covariance within feature points:

$$C_R = \frac{1}{n-1} \sum_{i=1}^{n} (z_i - \mu)(z_i - \mu)^T \tag{3}$$

where μ represents the intrinsic mean of the data sample points [23]. For other detailed introductions, please refer to reference [24,25]. The calculation method of the intrinsic mean of the Lie Group covariance matrix can be referred to references [26,27].

3.3 Constructing CNN

To make full use of the local correlation characteristics and consider the characteristics of its high-level features to predict dropouts. We propose a novel deep learning model through a large number of experiments.

The general architecture is shown in Fig. 2. Our proposed model is mainly comprised of two parts: a lightweight CNN and a multidilation pooling module. We use a lightweight CNN as a base network to extract high-level features. We take the Lie Group region covariance matrix as the input, the size of the input is 30×7 (30 represents row data, representing the count of 30-day learning behavior, 7 represents 7 types (Table 3) of behavior features extracted from the clickstream logs data). Layer $C1$ is a convolutional layer with four feature maps having
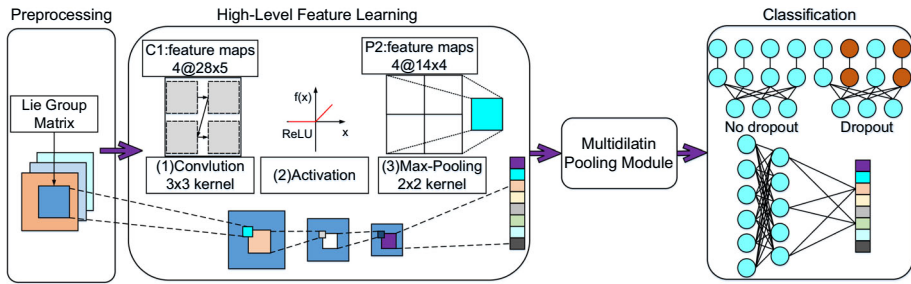
Fig. 2 The general architecture of the proposed network

the size of 28×5, in which each unit in each feature map is connected to a 3×3 neighborhood in the input layer. Use the Rectified linear unit (ReLU) as the activation function [28]. Layer $P2$ is the max-pooling layer and has four feature maps of size 14×4. Each unit in each feature map is connected to the 2×2 neighborhood in the corresponding feature map in Layer $C1$. According to the comparative analysis of the experiment, we set the steps of rows and columns of the Lie Group region covariance matrix to be 1. Then, the multidilation pooling module is applied to extract multidimensional features, and by connecting multidimensional features, the feature map is combined to learn multiscale association information. Finally, complete the classification task.

To achieve a multidimensional feature representation, we added a multidilation pooling module, as shown in Fig. 3. This module is a pyramidal pooling module, whose idea originates from SPPNet [29]. To eliminate the fixed-size constraint of the network model when introducing the SPP layer into SPPNet, since the input scale is very flexible, the SPP layer can pool features extracted with a variable scale. It can be seen from Fig. 3 that our module has multiple branches, which can not only complete the extraction of features of different dimensions but also has good robustness. One branch is used to directly use the global average pooling, and the other three branches are divided into global average pooling, dilated convolution, and SE layer. Finally, the features of multiple branches are joined together to capture the global multidimensional feature representation. Multiple dilation rates of 2, 5, and 6 are set in three dilated convolutions to extract features of different dimensions.

4 Data Preparation

We selected 39 categories of courses from the XuetangX platform [30], which are used in KDDCup2015 [31] and contain both labeled and unlabeled data. The labeled data contains 30 days of recorded information and a label representing whether the learner exhibited learning behavior during the last 10 days. If the learner does not exhibit any learning behavior in the last 10 days, it is labeled as a dropout, otherwise, it is labeled as a retainer.

For the convenience and intuitionistic display of learner behaviors in the follow-up research, we have statistically analyzed the frequency of different types of events in the data set, as shown in Tables 3 and 4. From Table 4, we find that wikis are a resource that learners do not frequently access. Dropout and retained learners statistics as shown in Table 5, we find that only 21% of the learners complete their studies.
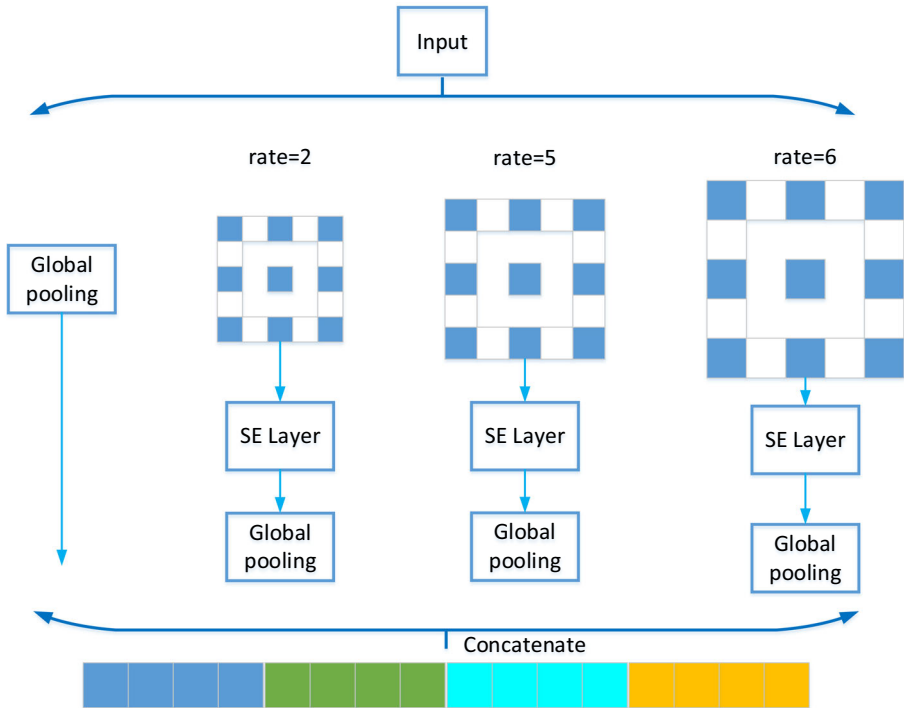
Fig. 3 Structure of the multidilation pooling module

Table 3 Types of events in the clickstream log

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

Table 4 Frequency of different types of learning behaviors

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

Table 5 Statistics of dropout and retained learners

Types	Amount	Proportion (%)
Dropout	95,581	79
Retain	24,961	21

Table 6 Confusion matrix

	Predicted negative class	Predicted positive class
Actual negative class	False negative (FN)	True positive (TP)
Actual positive class	True negative (TN)	False positive (FP)

Table 7 Experimental environment parameters

Item	Content
Processor	Inter Core i7-4700 CPU with 3.60 GHz
Memory	16 GB
Operating system	Windows 7 Pro
Hard disk	500 GB
Software	Matlab 2019a
GPU	Nvidia Titan-X $\times 2$
PyTorch	v1.1
Learning rate	10^{-8}
Momentum	0.9
Weight decay	5×10^{-4}

5 Experimental Analysis

5.1 Evaluation Metrics

There exist five widely-used, standard evaluation metrics in classification: Accuracy, F-Measure, Precision, Recall, and Confusion Matrix (as shown in Table 6). As follows:

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} \tag{4}$$

$$F\text{-}Measure = \frac{2 \times Precision \times Recall}{Recall + Precision} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{FP + FN} \tag{7}$$

5.2 Experimental Setup

The experiments and the contrast experiments were carried out in the same environment. The experimental environment and relevant parameters are shown in Table 7. The parameter values were set based on reference to [32,33].

The Classification algorithms used in this study include Classification And Regression Tree (CART), CNN, Gradient Boosted Decision Tree (GBDT), Linear Discriminant Analysis (LDA), Logical Regression (LR), NB, RF, SVM. Table 8 lists the parameters of some algorithms, where $n\_estimator$ represents the number of base classifiers in RF and GBDT, C and γ represent the parameters of SVM, default parameter values in libSVM [34] are

Table 8 Parameters setting

Method	Parameter	Value
GBDT	n_estimator	600
RF	n_estimator	600
SVM	C, γ	C=1, $\gamma = \frac{1}{num\_feature}$

adopted, and $num\_feature$ represents the number of features of data samples, which is set to 200 in this study.

5.3 Experiment Results

5.3.1 Local Correlation Experiment Results

In the previous discussion, we found that learners' learning behaviors during the MOOC learning process exhibited a strong local correlation over several consecutive days. In the feature matrix, each row represents the learner's learning behavior statistics for a day. By arranging the feature vectors of each day's learning behavior in time order, the obtained matrix maintains local relevance. If the vectors in the feature matrix are arranged randomly, the local relevance of learning behavior is destroyed. Can the proposed method achieve the best prediction results?

To explain the above problem, we randomly arrange the feature vectors of each row in the matrix 200 times and obtain the average results of 200 experiments. The experimental results were calculated by 20 cross-validations. At the same time, we compare the experimental results with the ordered experimental results. As shown in Fig. 4, after the local correlation of learning behavior is destroyed, all indicators have a certain degree of decline. The experimental results show that the local correlation of learning behavior can not be ignored, and our model can make good use of the local correlation of learning behavior.

5.3.2 Experimental Results Under Different Situations

In the actual scenario, we can predict the next stage of dropout based on the known stages. As shown in Fig. 5, the dropout of the third stage is predicted according to the learning behavior of the first two stages. As shown in Fig. 6, the dropout of the second stage is predicted based on the learning behavior of the first stage. As shown in Fig. 7, the dropout of the fourth stage is predicted according to the learning behavior of the first two stages, and that of the fourth stage is predicted according to the learning behavior of the first stage, as shown in Fig. 8.

Tables 9 and 10 show the experimental results in Situations 1 and 2, respectively. In Table 9, our approach achieved the best results in accuracy, F-measure, precision, and recall. Compared with the average, our approach achieves the best results in all evaluation metrics. In Table 10, our approach achieved the best results in accuracy, precision, and recall. Compared with the average, our approach has the best results in accuracy, precision, and recall.

Tables 11 and 12 show the experimental results in Situations 3 and 4, respectively. In Table 11, our approach achieved the best results in accuracy, F-measure, precision. Compared with the average, our approach achieved the best results in accuracy, F-measure, precision. In Table 12, our approach achieved the best results only in accuracy. But compared with the average, our approach achieved the best results in accuracy and recall.
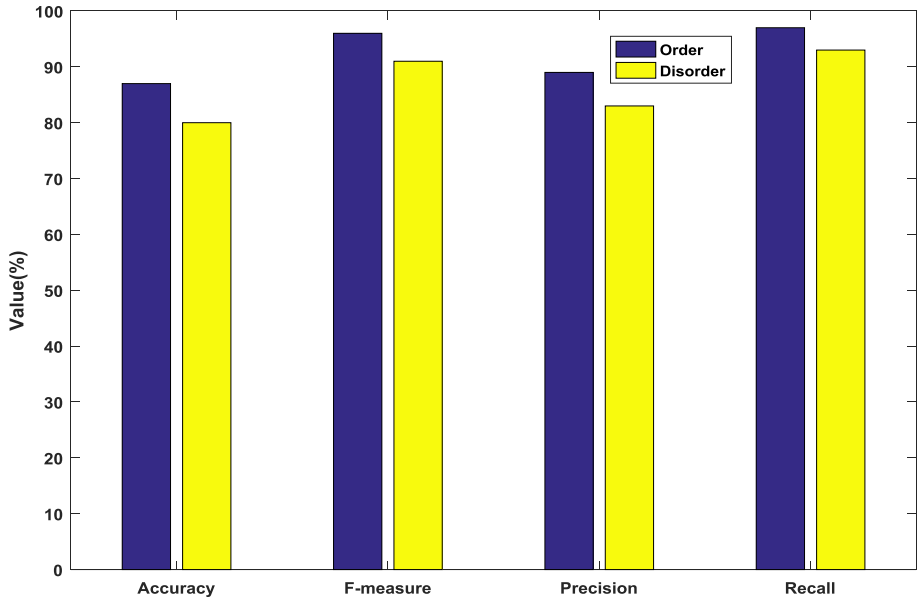
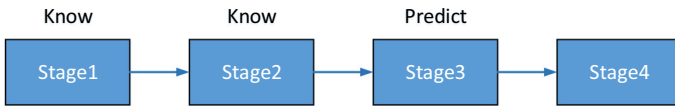**Fig. 4** Comparion of the dropout predictions when ordered or disordered
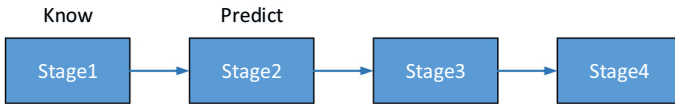

Fig. 5 Temporal dropout prediction in Situation1

Fig. 6 Temporal dropout prediction in Situation2
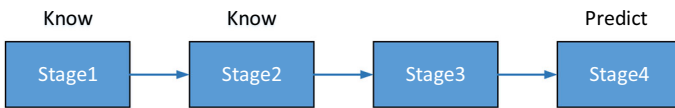

Fig. 7 Early dropout prediction in Situation3
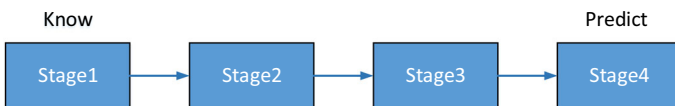

Fig. 8 Early dropout prediction in Situation4

Table 9 Experimental results of temporal dropout prediction in Situation1

Method	Accuracy	F-measure	Precision	Recall
CART [35]	0.8087	0.8438	0.8371	0.8507
CNN [35]	0.8642	0.8937	0.8522	0.9401
GBDT [35]	0.8428	0.8767	0.8372	0.9202
LDA [35]	0.6643	0.7697	0.6598	0.9236
LR [35]	0.6563	0.7593	0.6608	0.8922
NB [35]	0.6596	0.7495	0.6777	0.8383
RF [35]	0.8603	0.8882	0.8641	0.9137
SVM [35]	0.8631	0.8925	0.8529	0.9360
Average	0.7774	0.8341	0.7802	0.9019
Proposed	**0.9237**	**0.8967**	**0.9624**	**0.9153**

Table 10 Experimental results of temporal dropout prediction in Situation2

Method	Accuracy	F-measure	Precision	Recall
CART [35]	0.6452	0.6292	0.7543	0.5397
CNN [35]	0.6903	0.6828	0.7969	0.5975
GBDT [35]	0.6755	0.6882	0.7513	0.6470
LDA [35]	0.6077	**0.7161**	0.6005	0.8869
LR [35]	0.6059	0.7111	0.6016	0.8693
NB [35]	0.6082	0.6995	0.6112	0.8178
RF [35]	0.6826	0.6714	0.7947	0.5813
SVM [35]	0.6887	0.6875	0.7813	0.6138
Average	0.6505	0.6857	0.7115	0.6942
Proposed	**0.7361**	0.6937	**0.8352**	**0.8973**

Table 11 Experimental results of temporal dropout prediction in Situation3

Method	Accuracy	F-measure	Precision	Recall
CART [35]	0.7895	0.8703	0.8511	0.8904
CNN [35]	0.8456	0.9075	0.8639	0.9558
GBDT [35]	0.8436	0.9070	0.8579	0.9621
LDA [35]	0.8399	0.9052	0.8529	0.9644
LR [35]	0.8398	0.9054	0.8511	**0.9672**
NB [35]	0.8033	0.8739	0.8884	0.8599
RF [35]	0.8405	0.9036	0.8677	0.9425
SVM [35]	0.8419	0.9055	0.8602	0.9560
Average	0.8305	0.8973	0.8617	0.9373
Proposed	**0.8961**	**0.9164**	**0.9026**	0.9267

It can be seen from the above experimental results that the classification accuracy of traditional methods (such as CART and GBDT) is relatively low, mainly because these methods treat the dimensions of feature vectors as independent individuals and fail to fully consider the correlation between features and the local correlation of learning behaviors. The CNN [35] method achieves higher accuracy than the above methods because it extracts

Table 12 Experimental results of temporal dropout prediction in Situation4

Method	Accuracy	F-measure	Precision	Recall
CART [35]	0.7759	0.8648	0.8285	0.9045
CNN [35]	0.8188	0.8936	0.8358	0.9600
GBDT [35]	0.8182	**0.8943**	0.8293	0.9705
LDA [35]	0.8162	0.8930	0.8290	0.9678
LR [35]	0.8147	0.8930	0.8236	**0.9751**
NB [35]	0.7622	0.8477	**0.8611**	0.8347
RF [35]	0.8136	0.8899	0.8365	0.9506
SVM [35]	0.8161	0.8925	0.8320	0.9624
Average	0.8045	0.8836	0.8345	0.9407
Proposed	**0.8316**	0.8751	0.8295	0.9573

more abstract and higher-level features. Compared with the above methods, our proposed method achieves superior performance. The main reason is that the features in the model are concatenated by the time order, instead of treating all dimensions of the feature vector as independent individuals, taking full advantage of the local correlation of learning behaviors. In our method, the local correlation of learning behavior is retained in the Lie Group regional covariance matrix, and the row feature vectors of different dates are arranged according to the time order. The local correlation feature is retained while the feature dimension is reduced, which has good robustness and computational performance.

In the analysis in Sect. 5.3.1, we found that learners' learning behaviors in the MOOC learning process show local correlation, that is, the learning behaviors on several consecutive days showed very strong correlation characteristics. From Tables 9, 10, 11 and 12, different experimental results are obtained for different situations. Specifically, the prediction accuracy of situation 1 is the highest and that of situation 2 is the lowest. Further analysis shows that the amount of known historical data will influence the dropout prediction results. Situation 1 has more historical data and higher prediction accuracy, while situation 2 has less historical data and lower prediction accuracy. Situation 4 is similar to situation 2, although there is only less historical data, it still retains some information of other stages (such as stage 2 and stage 3), so the prediction accuracy is relatively higher. Therefore, we can conclude the following conclusions: (1) The amount of historical data will influence the dropout prediction results, (2) When a large amount of historical data are used, local correlation of learning behavior can be captured effectively. From the experimental results in Tables 9, 10, 11 and 12, we found that the first two stages are known may be enough for dropout prediction. Therefore, we can speculate that this is the reason why KDDCup2015 only provides the historical data of three stages.

6 Conclusion

In this study, we explored methods of predicting dropouts to improve the completion rate of MOOCs. Firstly, we conducted a detailed statistical analysis of MOOC learners' learning behaviors. The research results showed that learners exhibit similar learning behaviors on consecutive days of learning, and learners' learning status on the next day might be related to their learning status on the previous day. We proposed a Lie Group regional covariance matrix to represent the local correlation information of learning behavior, and construct a

CNN model with a multidilation pooling module to extract the local correlation high-level features of learning behavior for dropout prediction.

In the future, we will continue to study related CNN models to further improve the accuracy of dropout prediction.

References

1. Crues RW (2018) How do gender, learning goals, and forum participation predict persistence in a computer science MOOC? ACM Trans Comput Educ (TOCE) 18(4):1–14
2. Ramesh A et al (2014) Learning latent engagement patterns of students in online courses. In: Twenty-eighth AAAI conference on artificial intelligence, pp 1–7
3. He J et al (2015) Identifying at-risk students in massive open online courses. In: Twenty-ninth AAAI conference on artificial intelligence, pp 1749–1755
4. Peña-Ayala A (2014) Educational data mining: a survey and a data mining-based analysis of recent works. Expert Syst Appl 41(4):1432–1462
5. Vahdat M et al (2015) Advances in learning analytics and educational data mining. In: Proceedings of of ESANN2015, pp 297–306
6. Dalipi F, Imran AS, Kastrati Z (2018) MOOC dropout prediction using machine learning techniques: review and research challenges. In: 2018 IEEE global engineering education conference (EDUCON). IEEE, pp 1007–1014
7. Kloft M et al (2014) Predicting MOOC dropout over weeks using machine learning methods. In: Proceedings of the EMNLP 2014 workshop on analysis of large scale social interaction in MOOCs, pp 60–65
8. Taylor C, Veeramachaneni K, O'Reilly U-M (2014) Likely to stop? predicting stopout in massive open online courses. arXiv preprint arXiv:1408.3382
9. Liang J et al (2016) Big data application in education: dropout prediction in edx MOOCs. In: 2016 IEEE second international conference on multimedia big data (BigMM). IEEE, pp 440–443
10. Ramesh A et al (2013) Modeling learner engagement in MOOCs using probabilistic soft logic. In: NIPS workshop on data driven education, pp 21–62
11. Li W et al (2016) Dropout prediction in MOOCs using behavior features and multi-view semi-supervised learning. In: 2016 international joint conference on neural networks (IJCNN). IEEE, pp 3130–3137
12. Balakrishnan G, Coetzee D (2013) Predicting student retention in massive open online courses using hidden markov models. Electrical Engineering and Computer Sciences University of California at Berkeley, vol 53, pp 57–58
13. Fei M, Yeung D-Y (2015) Temporal models for predicting student dropout in massive open online courses. In: 2015 IEEE international conference on data mining workshop (ICDMW). IEEE, pp 256–263
14. Whitehill J et al (2017) Delving deeper into MOOC student dropout prediction. arXiv preprint arXiv:1702.06404
15. Wang W, Yu H, Miao C (2017) Deep model for dropout prediction in MOOCs. In: Proceedings of the 2nd international conference on crowd science and engineering, pp 26–32
16. Márquez-Vera C et al (2013) Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. Appl Intell 38(3):315–330
17. Moon S, Potdar S, Martin L (2014) Identifying student leaders from MOOC discussion forums through language influence. In: Proceedings of the EMNLP 2014 workshop on analysis of large scale social interaction in MOOCs, p 15
18. Kizilcec RF, Piech C, Schneider E (2013) Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In: Proceedings of the third international conference on learning analytics and knowledge, pp 170–179
19. Dekker GW, Pechenizkiy M, Vleeshouwers JM (2009) Predicting students drop out: a case study. International Working Group on educational data mining
20. Xu Y, Bo D, Zhang L (2020) Assessing the threat of adversarial examples on deep neural networks for remote sensing scene classification: attacks and defenses. IEEE Trans Geosci Remote Sens 59(2):1604–1617

21. Xu Y et al (2018) Hyperspectral image classification via a random patches network. ISPRS J Photogramm Remote Sens 142:344–357
22. Xu Y, Bo D, Zhang L (2019) Beyond the patchwise classification: spectral-spatial fully convolutional networks for hyperspectral image classification. IEEE Trans Big Data 6(3):492–506
23. Tuzel O, Porikli F, Meer P (2006) Region covariance: a fast descriptor for detection and classification. In: European conference on computer vision. Springer, Berlin, pp 589–600
24. Xu C, Zhu G (2021) Intelligent manufacturing Lie group machine learning: real-time and efficient inspection system based on fog computing. J Intell Manuf 32(1):237–249
25. Xu C, Zhu G, Shu J (2020) Robust joint representation of intrinsic mean and kernel function of Lie group for remote sensing scene classification. IEEE Geosci Remote Sens Lett 18(5):796–800
26. Xu C, Zhu G, Shu J (2021) A lightweight intrinsic mean for remote sensing classification with Lie group kernel function. IEEE Geosci Remote Sens Lett 18(10):1741–1745
27. Xu C, Zhu G, Shu J (2022) A lightweight and robust Lie group-convolutional neural networks joint representation for remote sensing scene classification. IEEE Trans Geosci Remote Sens 60:1–15
28. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
29. He K et al (2015) Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Trans Pattern Anal Mach Intell 37(9):1904–1916
30. XuetangX http://www.xuetangx.com/ (2019)
31. Kddcup2015 https://biendata.com/competition/kddcup2015/ (2015)
32. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
33. Sivic J, Zisserman A (2003) Video Google:a text retrieval approach to object matching in videos. In: IEEE international conference on computer vision, vol 3. IEEE Computer Society, pp 1470–1470
34. Chang C-C, Lin C-J (2011) LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol (TIST) 2(3):1–27
35. Wen Y et al (2019) Consideration of the local correlation of learning behaviors to predict dropouts from MOOCs. Tsinghua Sci Technol 25(3):336–347