# A Data Mining Method for Students' Behavior Understanding

https://doi.org/10.3991/ijet.v15i06.13175

Wei Na
Hunan University of Arts and Science, Hunan, China
276677761@qq.com

**Abstract**—A framework for predicting students' learning performance based on behavioral model is proposed to model students' behavior and describe their behavior characteristics accurately and comprehensively. The framework extracts features from multiple perspectives to describe behaviors more comprehensively, including statistical features and association features. In addition, a multi-task model is designed for fine-grained prediction of students' learning performance in the curriculum. A framework for predicting mastery based on online learning behavior is also put forward. Additional context information is added to the collaborative filtering algorithm including student-knowledge-point mastery, class-knowledge-point, and students' mastery is predicted according to the learning path excavated. Considering the time-varying of mastery, the approximate curve of students' mastery of knowledge points is fitted according to the Ebinhaus forgetting curve. The experiments show that the proposed framework has a high recall rate for the prediction of learning performance and also shows a certain practicability for early warning. Further, based on the model, the correlation between student behavior patterns and learning performance is discussed. The inclusion of additional information has improved the prediction efficiency, especially the operational efficiency. At the same time, the proposed framework can not only dynamically assess students' master of knowledge but also facilitate the system to review feedback or adjust the learning order and provide personalized learning services.

**Keywords**—Education data mining, student behavior pattern, sequential pattern mining, learning performance prediction

## 1 Introduction

In recent years, with the rapid development of data mining in education, the combination of data mining and machine learning methods to analyze students' behavior data has become a popular trend. It is mainly devoted to the prediction of students' learning performance, the discovery of future behavior & interest and the extraction of students' individual or group characteristics [1]. It is also necessary to study the correlation between communication behavior patterns and students' performance in online learning. Moreover, to explore the differences of behavior patterns among different performance groups, dedicated to find the most frequent sequence of activities and the impact of the sequence of activities on emotional states such as confusion and boredom [2]. The relationship between students' behavior, mental health and academic performance is studied by using the collected behaviors of students such as movement, sleep, dialogue and learning. A learning framework is constructed to model multiple heterogeneous behavior characteristics that automatically assess students with financial difficulties, solve the time-consuming shortcomings and the hidden dangers of fairness of traditional manual selection methods.

Students' behavior patterns have objective reflections on their learning performance, emotional state, mental health and even economic status. With the widespread attention to the research on students' behavior patterns, how to model students' behavior and describe their behavior characteristics accurately and comprehensively is an important issue to be considered.

## 2 Literature Review

In terms of the evolution and trend of behavior patterns, Cowan et al. (2017) focused on the similarity of students' behavior patterns and grouped students dynamically by improving evolutionary clustering algorithm [3]. Bian et al. (2015) considered the temporal variability in human behavior modeling, so they added temporal information and paid attention to the changing trend of behavior [4]. From the perspective of study habits, Feoktistov et al. (2015) realized automatic classification of learners' learning styles through online learning behavior and compared the predictive performance under several different classifiers. It was found that tree-based classifiers could achieve predictive accuracy up to 91% [5]. Paulus et al. (2015) studied the mining sequential patterns of user learning and discussed the topological relationship and relevance between knowledge points [6]. Wiemken et al. (2015) used neural networks to learn user's learning paths, improve undifferentiated learning strategies and mine the relevance between courses [7]. In view of assessment and prediction of learning performance, Henley et al. (2016) predicted the dropout crisis of middle school students by using Logistic Regression method based on the information of classroom participation behavior (such as absence of classes) and historical achievement data. They also applied work-improved regularization Logistic Regression algorithm to model online learning behavior periodically and predict students' performance of dropping out from online courses in advance [8]. As far as personalized learning services is concerned, Samuel et al. (2016) constructed an adaptive learning system and used back propagation neural network to recommend personalized learning content to learners [9]. Shi et al. (2017) combined clustering algorithm with Apriori algorithm, analyzed the association rules of the learning sequence patterns of various students in the course and recommended the appropriate course according to the learning paths of various students' preferences [10].

To sum up, the above findings mainly reflect the evaluation of students' behavior patterns, classification of learners' learning styles, students' performance and so on. Based on the above researches, student behavior recognition and behavioral modeling in behavior sequences are described. Then, the experiment is carried out to discuss the students' behaviors and learning performance based on campus behavior pattern is predicted. Moreover, prediction framework, student behavior characteristics and learning performance prediction models are analyzed. At last, the research results are analyzed in details. The advantage of the paper is that it well used data mining method to discuss students' behavior understanding which has great practical value. However, students' behaviors are complex and difficult to analyze at a deep level. So, only using data mining method to talk about students' behaviors is not enough. Therefore, it is supposed to combine data mining method with other technologies or algorithms to explore students' behaviors further.

# 3 Method

## 3.1 Student behavior recognition

Campus card, as an important tool for students' campus life has become an important medium as well as for automatically capturing students' daily consumption and identity data. With each card swipe, a record containing the corresponding trigger time, location & related attributes are generated and stored in a database or data warehouse. Table 1 shows an example of regular data format of campus card. For example, the original campus card records in Table 1 contain information such as UID (student ID), DEALTIME (behavior time), TRANSMONEY_ (transaction amount), ORGINID (behavior location), and DEALCODE (transaction type such as recharge, consumption, etc.). Through further processing and mining of campus card records, each user's behavior sequence is obtained as the basis of subsequent modeling. Firstly, a trigger record of the user (including the information of swipe time, location, consumption type, etc.) is defined as an event unit. The formal definition of events is:

Definition 1 (event): For a given user $n \in U$, an event is defined as $e = (t, p, c, m)$, $e \in Eu$, including event timestamp t, location $p \in P$, transaction type $c \in C$ and transaction amount $m \in R$. Ru represents the behavior set of user u.

For any specific application scenario, location set P, consumption type set C and student group should be limited sets. The granularity of location can be flexibly set according to actual application. For example, the first five records in Table 1 correspond to a recharge event, a store consumption event, two restaurant consumption events and a store consumption event respectively according to the corresponding location, time and consumption type.

**Table 1.** An example of regular campus card recording

| UID | Dealtime | Transmoney | Orginid | Dealcode |
|---|---|---|---|---|
| U1 | 02-03-2015 09:29 | 400 | 1 | 4 |
| U1 | 02-03-2015 09:47 | 2.8 | 1.02E+09 | 20 |
| U1 | 02-03-2015 11:56 | 2.5 | 1012002 | 20 |
| U1 | 2015 | 5 | 1012002 | 20 |
| U1 | 2015 | 2 | 1049 | 20 |
| … | … | … | … | … |

Further, for a group of consecutive events with the same location category, the same transaction type and within a certain time window, it can be merged into an activity. For instance, a series of consecutive consumption events in multiple windows of a restaurant can be merged into a specific meal activity (e.g. the third and fourth events in Table 1). The formalization of activities is defined as:

Define 2 (activity): a set of continuous event sequences (e1,e2,...,eG) that the user u has with the same location and transaction type in a certain period of time and they are merged into an activity with v=(tstart,tend,p,c,m). In this case, the start time is tstart=e1, t, and the end time is tend=eG.t. For any ei, v.p=ei.p in the event sequence, v.c=ei.c, m is the cumulative operation of ei.m. Vu is used to represent the activity set of user u. Among them, as the largest window of activity, $\delta$ can be set according to actual needs to limit the maximum time of an activity (e.g. an hour).

So far, according to the time sequence, the sequence of user u's activity in the specified period T is Seq(u,t0)=(v1,v2,...,vs, where vi, vj∈Vu for i<j, vitstart<vjtstart, t0≤v1tstart, vstend≤t0+T.

A diverse set of activities often implies the same behavior, such as eating that corresponds to a variety of activities because of the different distribution of time or place. For a specific application scenario, there exists a mapping function f, which defines a uniquely determined behavior tag (such as breakfast, shopping, etc.) for an activity according to the activity time, location category and transaction type. Therefore, the behavior of an activity can be formally defined as:

Definition 3 (behavior): Given the user u's activity v∈Vu, its corresponding behavior is h=f(v), h∈H, and the effective behavior set H is a finite behavior set. Function f is a predefined mapping for the actual application scenario, for example, for input vi,

If vi.p is campus store, then f(vi)=Shopping;

If vi.p is restaurant∧(6.am≤vi.tstart≤10.am), then f(vi)=Breakfast;

If vi.c is recharge type, then f(vi)=Recharge.

Similarly, for a set of activity sequences, the corresponding behavior sequence Seq (u, t0) is defined as Seq'(u,t0)=(h1,h2,…,hS), where hi∈H,hi=f{vi).

For a given user in a specified period T, the whole process of acquiring Seq's activity sequence and Seq's behavior sequence from the event set is simply described. $\delta$ is the largest window defined by an activity, such as an hour. Firstly, the activity sequence Seq and the behavior sequence Seq' are initialized as empty sets, an open queue q is set up to assist in storing the pending events and the attribute m is initially set to 0.

### 3.2 Behavioral modeling in behavior sequences

Markov model is one of the most widely used models due to the limitation of movement rule and speed in behavior modeling which assumed that people's behavior in the future always depends on the current behavior and accords with certain regularity. It is simply assumed that students' daily behavior conforms to Markov nature and hidden Markov model (HMM) is used to model students' behavior, in which the observed variables describe the user's diverse activities and the hidden state is used to describe the implicit behavior behind the user's diverse activities.

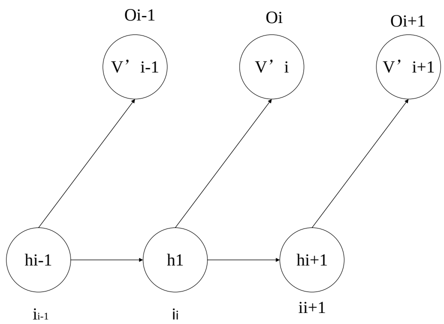


**Fig. 1.** Schematic diagram of hidden Markov models for behavior sequences

Firstly, it is necessary to construct discrete activity space V' and divide the activity time into hours, ignoring the transaction amount for the sake of simplicity.

By using the processed active sequence, the observation sequence O= (o1,o2,...,oS:) can be constructed. It is assumed that the size of the observation space V' is M, the corresponding behavior sequence is extracted as the hidden state sequence I= (i1,i2,…,is) and the size of the hidden state space H is N. Based on the sequence pairs of L cycles, hidden Markov model λ= (π,A,B) is constructed for {(O1,I1),(O2,I2),…,(OL,IL)}, as shown in Figure 1. π(π1, π2,…, πN) is the initial probability distribution vector, A=[aij]N*N is the state transition probability distribution matrix and B=[bjk]N*M is the observation probability distribution matrix.

For the learning of model parameter λ, if the training data contains the observation sequence and its corresponding hidden state sequence, it can be realized by supervised learning method and if only the observation sequence is included, it can be realized by unsupervised learning algorithm - EM (Expectation Maximization) algorithm. Because the hidden state space is the behavior label space of predefined behavior set, the hidden state sequence is known and given sequence pairs {(O1,I1),(O2,I2),…,(OL,IL)}, the maximum likelihood estimation method can be used to estimate the parameters of the model as follows:

Initial probability distribution π(π1, π2,…, πN): πi refers to the frequency of the initial action of the user for hi in the L implicit state sequence.

State transition probability distribution A=[aij]N*N: aij represents the probability of the user's behavior (hidden state) occurring at time t + 1 under the condition that the user's t behavior (hidden state) for hi at a certain time, and the estimated value is shown in Formula (1):

$$a_{ij} = P(i_t + 1 = h_j \mid i_t = h_i) = \frac{A_{ij}}{\sum_{j=1}^{N} A_{ij}}, i = 1,2,..., N;\ j = 1,2,..., N \tag{1}$$

In Formula (1), Aij indicates the corresponding frequency;

Observation probability distribution B=[bjk]N*M: bjk indicates the probability of showing a specific activity (observation) v'k when the user's behavior (hidden state) is hj, and its estimated value is shown in Formula (2):

$$b_{jk} = P(o_t = v'_k \mid i_t = h_j) = \frac{B_{jk}}{\sum_{k=1}^{M} B_{jk}}, j = 1,2,..., N; k = 1,2,..., M \tag{2}$$

In Formula (2), Bij denotes the corresponding frequency.

### 3.3 Experiment and analysis

In the experiment, the campus card records of a university student in the spring semester of 2015 are used as the data set and desensitized. In random sampling, 1000 students were active in campus activities (i.e. the record is more than 300 and no card-missing weeks are found). The extraction period of behavior sequence is set to one week. Table 2 is a detailed description of the data set.

**Table 2.** Data set description

| Description item | Statistic |
|---|---|
| Number of users | 1000 |
| Number of records | 813046 |
| Periodic number | 18 |
| Number of locations | 121 |
| Time interval number | 24 |
| Behavior (hidden state) category | 12 |
| Weekly average behavior sequence length | 45 |

Considering the diversity of locations, for example, a school may have more than one campus, a campus may have more than one canteen and a canteen will have more than one business or window. So, it is necessary to determine the particle size of the physical location to be analyzed according to a priori. After determining the particle size of the location, the number of locations in the campus record will be reduced to 121. The 12 behavior (hidden state) categories in the experiment are defined as shown in Table 3.

Weekly average distribution of different behaviors: Based on the behavior sequence in the data set, the frequency distribution of different types of behavior in each cycle is shown in Figure 2. The four types of behavior with the highest proportion are Shopping, Lunch, Dinner and Breakfast which are consistent with the normal campus life of students and also coincide with the use of campus card to record students' consumption behaviour

**Table 3.** A set of common behavior set examples

| Behavior | Description |
|---|---|
| Breakfast | Having breakfast in the school cafeteria |
| Lunch | Having lunch in the school cafeteria |
| Dinner | Having dinner in the school cafeteria |
| Shopping | Shopping at school stores |
| Exercise | Exercise at the school gym |
| Treatment | Visiting a school hospital |
| Buying materials | Buying learning materials |
| Library entrance | Registering in Library |
| Recharge | Campus card recharging |
| Card service | Having card service |
| School bus | Taking the school bus |
| Dorm access | Access control in dormitories |

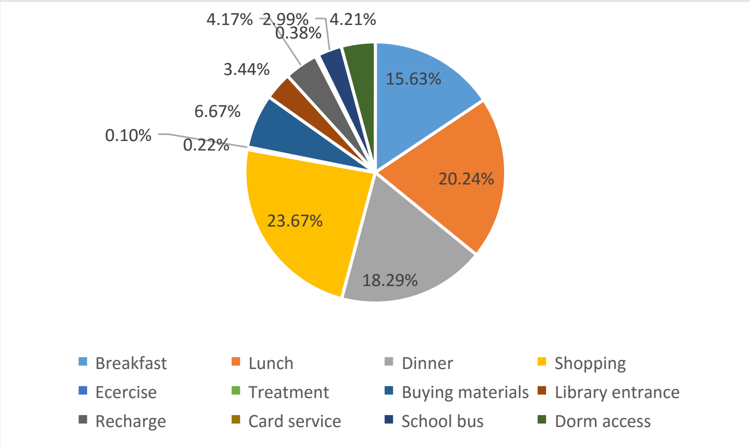


**Fig. 2.** Frequency distribution of average weekly behaviors

Model confusion: confusion is used to measure the performance of the model. The degree of confusion is used to measure the quality of a probability distribution or probability model in information theory. It is also often used to evaluate language models. The lower the confusion degree is, the better the effect of the model is. The definition of model confusion is shown in Formula (3):

$$Perplexity(M_i) = e\{-\frac{\sum_{j=1}^{m} \log p(I_j)}{\sum_{j=1}^{m} N_j}\} \quad (3)$$

The exponential part is the cross-entropy of the model. For each behavior sequence Ij, p(Ij) represents its corresponding likelihood function which is the length of the behavior sequence.

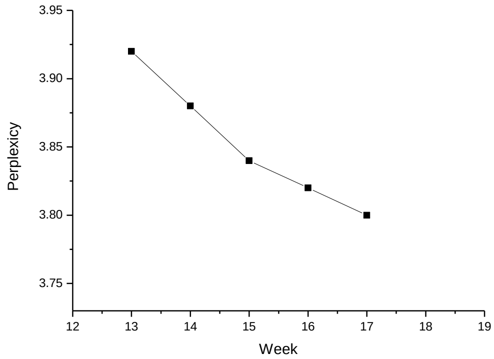


**Fig. 3.** Model confusion

The behavioral sequence set training models in different time ranges are used. The degree of confusion of the model varies with the time range of the training set as shown in Figure 3. The abscissa represents the time range of the training set from 13 to 18 weeks. Although the confusion degree of the model in Figure 3 does not vary much, it decreases with the increase of the time range of the training set. It indicates that the prediction effect of the model is better with the increase of the number of training sets.

### 3.4 Prediction of learning performance based on campus behavior pattern

At present, the related work of learning performance prediction mainly focuses on static prediction factors such as student demographic information, historical achievement data CGPA (Cumulative Grade Point Average), mid-term assessment, college background information, social information and verifies its effectiveness. However, the current work ignores the impact of dynamic behavior, specifically among college students; the dependence between the results of different courses is relatively small. In addition to the work related to online learning behavior and learning performance, few studies have analyzed the impact of students' behavior on school learning performance. For example, students' behavior data are collected based on student Life data set and mobile client and questionnaire but limited by the collection method, the collected behavior is limited and the accuracy is poor. With the gradual evolution of campus card into the main media of college life, its automatic perception method records the information of time, space and consumption generated by students' diverse behaviors which opens up a new and effective way for us to depict students' behavior.

The correlation between behavior and learning performance has also been studied in sociology. From 1950 to 2013, Ovid MEDLINE (The National Library of Medicine) published literatures and it was found that up to 36 articles have studied the effect of breakfast behavior on students' classroom behavior and learning performance. The results consistently show that habitual breakfast behavior as one of the important indicators of healthy life has a positive impact on students' learning and performance. Through multi-level Logistic Regression assessment of the correlation between healthy behavior and learning, the results show that students with unhealthy lifestyles such as low-quality diet or lack of physical activity are more likely to have poor learning performance. Despite the obtained result, more systematic quantitative analysis and regular learning are needed to identify the correlation between students' behavior and learning performance.

Traditional research methods for identifying learning performance are mostly based on single task learning, mainly including Support Vector Machine, Logistic Regression, Decision Tree, K-Nearest Neighbor, Neural Network and other mainstream algorithms. There are two main strategies to solve the single-task learning model: one is regarding each student's performance in all courses as a whole which ignores the differences between the courses and the selected groups; the other is to model each student's performance in each course separately which has data limitations and the trouble of repetitive and tedious modeling process. Compared with traditional single-task learning, multi-task learning often reflects more obvious advantages specifically when the amount of data per single task is small and the quality of data is poor. For retaining the uniqueness of a single task, it can simultaneously learn from multiple tasks by capturing the relationship between tasks and to improve the generalization performance and interpretability of the model while improving the prediction effect of a single task and the overall task. At present, multi-task learning is widely used in face recognition, multi label image classification and other applications.

### 3.5 Prediction framework

Figure 4 is a general picture of the framework. It mainly consists of three parts:

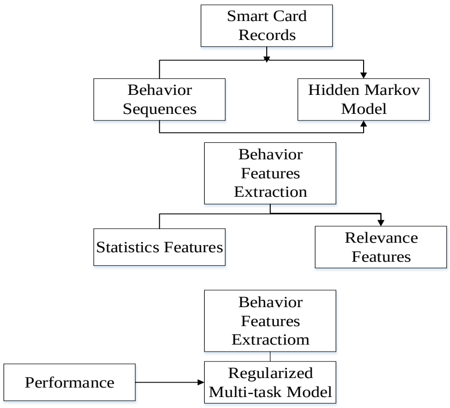


**Fig. 4.** Prediction framework

First of all, the behavior patterns of students are excavated. Students' behaviors and activities are extracted from campus card records and students' behavior patterns are extracted based on behavior sequence. Secondly, the behavior characteristics are extracted from two perspectives. Based on the behavior sequence, the statistical characteristics of behavior are extracted including behavior level, behavior trend and behavior law; based on behavior pattern, behavior correlation features are extracted including behavior transfer entropy and observation entropy. Finally, the abnormal learning performance prediction model is constructed. In order to predict the learning performance of each course at the same time, the framework design used a regularized multi-task model rather than a single-task solution.

### 3.6 Student behavior characteristics

Statistical method is the most widely used as intuitionistic description method. According to the average weekly distribution of different behaviors, Breakfast, Lunch and Dinner accounts for more than 50% of the average weekly meal behavior. So, in addition to these three behaviors, Meal at school and Meal on weekend’s two common aggregation behaviors are added. In particular, Dorm access is measured by leaving the dormitory before a specific time point and returning to the dormitory after a specific time point. This time threshold can be set to the peak value of the distribution, such as 8 a.m. and 10 p.m.

Each behavior $h \in H$ contains two types of measurable attributes: hfre is the cumulative frequency of the behavior h in a specified period T (such as a week).

For each specific behavior of any of the above attributes, the features are extracted from three aspects: behavior level, behavior change and behavior law. For the cumulative frequency attributes of behavior h, it is assumed that the mean value of hfre in the P-th unit cycle is hfre', and the mean sequence of continuous P cycles is (hfre1’, hfre2’,…,hfrep’), a linear fitting hfre’=at+b of the mean sequence with the fluctuation of the period t is made, then:

**Behavior level:** the average level of a specific behavior measured by mean. Behavior change: For linear fitting of mean series over successive periods, slope a is used to measure the change trend of a particular behavior. Behavior law: For linear fitting of mean series over successive periods, the regularity of a particular behavior is measured by sum of residual squares.

Although statistical features are relatively intuitive and easy to interpret, they ignore the correlation between different behaviors. The transfer of behavior implies the user's inherent behavior habits and the distribution of behavior in different activities also implies the user's certain or uncertain life state. In this section, the behavioral association features are extracted based on the previous model. The entropy is used to measure the transfer between behaviors as well as the uncertainty of the distribution between behavior and activity.

**Behavior transfer entropy:** The transfer of one behavior to another implies to some extent the laws and habits of daily life. From each row of the state transition probability distribution A in the hidden Markov model, the transfer entropy of a specific behavior can be obtained as shown in Formula (4). The smaller transfer entropy means that students behave more regularly

$$Entropy\_T(h_i) = -\sum_{h_j \in H} P(h_j \mid h_i) \log P(h_j \mid h_i) = -\sum_{j=1}^{N} a_{ij} \log(a_{ij}) \qquad (04)$$

**Behavior observation entropy:** The diversity of life is reflected by the distribution of different activities displayed by a behavior. Over-enriched or over-active lifestyle may be detrimental to learning status. From each row of the observed probability distribution B in the hidden Markov model, the observed entropy of a specific behavior can be obtained as shown in Formula (5). The greater the entropy of observation is, the more varied the students' daily activities are.

$$Entropy\_O(h_i) = -\sum_{v_j \in V'} P(v_i \mid h_i) \log P(v_j \mid h_i) = -\sum_{k=1}^{M} b_{jk} \log(b_{jk}) \qquad (05)$$

The d1 dimension statistical feature and d2 dimension correlation feature are extracted from the above process and the total dimension is d=d1+d2.

### 3.7 Learning performance prediction model

The single task learning model is widely used in related work. There are usually two solutions:

Students' overall learning performance is taken as a prediction goal. The input is a single matrix X∈Rn*d, and the learning expression label is a vector Y∈{0,1}n. If student u is not performing well in any course, Yu is set to 1. Obviously, this method neglects the differences of course selection groups among courses and the prediction is coarse-grained and lacks guiding significance.

A model is trained individually for each course. For each course i, the input is matrix X(i)∈Rni*d, and the corresponding learning performance label is Y(i)∈{0,1}ni. However, each group of sample data is relatively small; each model is learnt alone for each course and also accompanied by repeated cumbersome modeling process.

In view of the shortcomings of the two single task solving strategies, multi-task model is used to improve prediction granularity and generalization performance. Taking into account the different course groups of each course, naturally, a course is regarded as a task. The curriculum of the same college is related to some extent, so l2,1-norm is added to select the shared feature space.

For the new sample x input model of course i, its learning performance is labeled y=sign(W(i)Tx),W∈Rd*m. Each column of W(i) is the coefficient vector corresponding to course i, and its k-th item means the importance of the k-th feature of task i.

To solve the model, that is, the coefficient matrix W, the criterion function is minimized:

$$\min_{W} \sum_{i=1}^{m} \sum_{j=1}^{n_j} \log(1+\exp(-Y_j^{(i)}(W^{(i)T} X_j^{(i)}))) + p_1 \| W \| 2,1 + p_{L2} \| W \|_F^2 \tag{06}$$

In the above formula, the first one is logistic loss and the last two are regularization items including: l2,1-norm used to select shared feature space. Frobenius-norm ||W||F used to enhance the robustness of the model and pl & pl2 respectively control the sparsity and complexity of the model. Formula (6) can be divided into the following two parts:

$$\min_{W} g(x) + p_1 \| W \| 2,1 \tag{07}$$

$$g(x) = \sum_{i=1}^{m} \sum_{j=1}^{n_j} \log(1+\exp(-Y_j^{(i)}(W^{(i)^T} X_j^{(i)}))) + p_{L2} Tr(W^T W) \tag{08}$$

The first g(x) is derivable and its gradient in each item Wk(i) is:

$$\nabla g(W)_{k,i} = \frac{\partial g(x)}{\partial W_{k,i}} = \sum_{j=1}^{nj} (-Y_j^{(i)} X_{j,k}^{(i)} + \frac{Y_j^{(i)} X_{j,k}^{(i)}}{1+\exp(-Y_j^{(i)}(W^{(i)T} X_{j,k}^{(i)}))}) + 2 p_{L2} W_{k,i} \tag{09}$$

The above formula satisfies the Lipschitz continuity condition and uses the near-end gradient descent to get the solution under the minimum criterion function. At the same time, for the selection of the optimal parameters, the multi-fold cross-validation under the grid search is used.

## 4 Results and Discussion

All the experiments are completed in 8GB memory and single processor platform of 15-4460CPU@3.2GHz processor. The experiments are conducted on the same data set as before. Considering the homogeneity of the groups in each college, the learning performance of students in 12 courses is modeled by using the campus records of one college student in the spring semester of 2015. There are 302 students and 224,138 campus records in total. In order to ensure users' privacy information is not leaked; all data are anonymously collected before mining. Simplified in the process of extracting features for actions without transaction amount, such as 6ms, the relevant features on the attributes of cumulative amount can be neglected. 96 dimensions features are extracted totally including 72 dimensional statistical features and 24 dimensional association features.

Before starting the experiment, it is necessary to do a series of preprocessing of the data. Firstly, students with low activity are removed (activity can be defined as the length of the sequence of behavior within the unit cycle T). Non-discriminative courses are considered for students with unusual performance. In reality, the proportion of abnormal learning performance group is low and the data is biased. So, it is necessary to apply SMOTE (Synthetic Minority Oversampling Technique) algorithm independently for each course to balance two kinds of samples for each sample of abnormal class, a group of samples are randomly selected from the nearest neighbor sample with automatic sampling rate to make linear difference and the constructed new samples are added to the data set.

The following three main single task learning models are used as benchmarks.

**Logistic Regression:** Converts the result of linear regression to approximate posterior probability by logic function. The l1- paradigm and the l2- norms are selected as candidate regularization terms. Add C to control the intensity of punishment.

**SVM (Support Vector Machine):** Searches for the optimal hyper-plane based on structural risk minimization to divide different classes to the greatest extent. The experiment is carried out in linear kernel and radial basis function (RBF) kernel respectively and RBF kernel needs to set coefficient $\gamma$.

**Random Forests:** Multiple base classifiers (usually decision trees) are used for ensemble learning. Each base classifier selects different sample subsets and feature subsets for training and the average results of multiple base classifiers are used as the final output. The Gini co-efficient and entropy is chosen as the candidate criteria for the partition of the base classifier. For the single-task benchmark model, the overall learning performance of students in all courses is taken as the predictive target and if student u does not perform well in any course, then Yu is 1.

The evaluation indicators adopted are as follows: accuracy, precision, recall, FI-Score, ROC (Receiver Operating Characteristic) and AUC (Area under concentration - time curve).

**Comparison of different single task models:** After grid search, the optimal parameters of the single task model are as follows, Logistic Regression: using l1 paradigm as regularization term, C=1; SVM: based on RBF kernel, $\gamma$=0.01, C=10; Random Forests: using entropy as criterion, the number of base classifiers is 100.

From the point of view of ROC and AUC, all the three models show good generalization performance. For coarse-grained prediction, single task has a good performance, especially SVM. These results all indicate that the extracted behavioral characteristics have certain effect on the prediction of students' learning performance. It is speculated that, although there is no direct information about achievement, students' behavioral patterns can reflect their potential learning state to a certain extent and help to discover their future learning performance.

**Comparison between single task model and multi-task model:** According to grid search, the optimal parameters are pl=0.01 and pl2=0.001, and the MTL (Matrix Template Library) model performs well in all indicators. In practical applications, it is very important to accurately and comprehensively find students with abnormal learning performance which means that MTL model with higher recall rate has higher practicability.

**Comparison of the running time of each model:** LR (Load Runner) runs the shortest time followed by SVM and MTL. RF (Ratio Frequency) runs the longest time in the training process of ensemble learning which is about five times as long as LR. Even so, on the whole, the MTL model can get enough efficiency and accuracy.

**Early warning:** According to all the experiments mentioned above, the predictive framework can obtain better accuracy when given the behavior characteristics of students throughout the semester (one semester is divided into 18 weeks). However, this may not be practical enough for practical application, specifically for those who may have abnormal learning conditions. If screening can be made only through the behavior characteristics of students in the first half of the semester, it will be conducive to the early warning and effective guidance of abnormal performance.

## 5 Conclusion

A general framework is proposed to discover students' learning performance based on their behavior patterns. Based on the spatial and temporal data of card perception, students' behavior characteristics are extracted from two perspectives: statistics-based and model-based. In the classification work, the multi-task learning method is applied to model students' learning performance in multiple courses simultaneously. The experimental results show that the framework has a high recall rate for the prediction of learning performance and is flexible enough for early warning. In addition, according to the parameters of the model, the correlation between daily behavior and learning performance can be quantitatively measured. It is found that some behaviors (such as shopping, dining in restaurants and leaving the dormitory before 8 o'clock) are significantly correlated with learning performance; moderately active participation in campus life as well as maintaining healthy living habits in diet and early rising are conducive to the formation of good learning performance.

Therefore, this study is of great significance to guide students to improve their study habits and living habits in order to improve their performance. However, there are still some limitations in this study. For example, the subjects selected in this study are all from a school, the distribution of students is not balanced and the number of students is also small. In the latter study, students of different grades in different schools will be taken as subjects and the number of experiments will be increased as many as possible. It is hoped that the study can provide better guidance for students' life and learning, and train more talents for the society.

## 6 References

[1] Henley, A, J., DiGennaro, Reed, F, D., Reed, D, D., et al. A crowdsourced nickel-and-dime approach to analog OBM research: A behavioral economic framework for understanding workforce attrition. Journal of the Experimental Analysis of Behavior, 2016, vol. 106(2), pp. 134-144. https://doi.org/10.1002/jeab.220

[2] Middelkoop, D, V., Ballafkih, H., Meerman, M. Understanding diversity: a Dutch case study on teachers' attitudes towards their diverse student population. Empirical Research in Vocational Education & Training, 2017, vol. 9(1), pp. 1. https://doi.org/10.1186/s40461-016-0045-9

[3] Cowan, R, J., Abel, L., Candel, L. A Meta-Analysis of Single-Subject Research on Behavioral Momentum to Enhance Success in Students with Autism. Journal of Autism & Developmental Disorders, 2017, vol. 47(5), pp. 1-14. https://doi.org/10.1007/s10803-017-3076-6

[4] Bian, C., Xu, S., Wang, H., et al. A Study on the Application of the Information-Motivation-Behavioral Skills (IMB) Model on Rational Drug Use Behavior among Second-Level Hospital Outpatients in Anhui, China. Plos One, 2015, vol, 10(8), pp. e0137928. https://doi.org/10.1371/journal.pone.0137928

[5] Feoktistov, D., Sharapova, I., Orlova, E., et al. Experimental Research of Wetting and Drop Evaporation Processes on a Heated Substrate as a Method for Research Skills Development for Heat and Power Engineering Students. Procedia - Social and Behavioral Sciences, 2015, vol. 206, pp. 307-314. https://doi.org/10.1016/j.sbspro.2015.10.056

[6] Paulus, Markus. Licata, Maria., Kristen, Susanne., et al. Social Understanding and Self-Regulation Predict Pre-Schoolers' Sharing with Friends and Disliked Peers: A Longitudinal Study. International Journal of Behavioral Development, 2015, vol. 39(1), pp. 53-64. https://doi.org/10.1177/0165025414537923

[7] Wiemken, T, L., Carrico, R, M., Kelley, R, R., et al. Understanding why low-risk patients accept vaccines: a socio-behavioral approach. Bmc Research Notes, 2015, vol. 8(1), pp. 813. https://doi.org/10.1186/s13104-015-1816-2

[8] Henley, A, J., DiGennaro, Reed, F, D., Reed, D, D., et al. A crowdsourced nickel-and-dime approach to analog OBM research: A behavioral economic framework for understanding workforce attrition. Journal of the Experimental Analysis of Behavior, 2016, 1vol. 06(2), pp. 134-144. https://doi.org/10.1002/jeab.220

[9] Samuel, Greiff., Christoph, Niepel., Ronny, Scherer., et al. Understanding students' performance in a computer-based assessment of complex problem solving. Computers in Human Behavior, 2016, vol. 61(C), pp. 36-46. https://doi.org/10.1016/j.chb.2016.02.095

[10] Shi, W., Xue, K., Meshot, E,R., et al. The carbon nanotube formation parameter space: data mining and mechanistic understanding for efficient resource use. Green Chemistry, 2017, vol. 19(16). https://doi.org/10.1039/c7gc01421j

## 7 Author

**Wei Na** works for the Hunan University of Arts and Science in city of Hunan in China.

Article submitted 2020-01-13. Resubmitted 2020-02-24. Final acceptance 2020-02-25. Final version published as submitted by the authors.