

A Deep learning Grade Prediction Model of Online Learning Performance Based on knowledge learning representation

Shuaileng Yuan¹, Sukrit Leelaluk², Cheng Tang², Li Chen², Fumiya Okubo², Atsushi Shimada²

¹ Laboratory for Image and Media Understanding, Kyushu University, Japan

² Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan

Abstract

In recent years, due to the impact of Coronavirus disease (COVID-19), digital platforms have developed rapidly and accumulated a large amount of data. To better utilize the comprehensive and diverse data stored in online platforms for data mining, such as learning behavior analysis or performance prediction, and to provide guidance and valuable feedback for educator became more important. For the current analysis of learning behaviors by time series data with DNN method, the interpretability is not enough. This paper proposes a method based on the simultaneous use of learning behaviors and learning materials to obtain the representation of learned knowledge, and through multiple cross-validations, the effect of this knowledge representation has a certain improvement on the original data, and the interpretability can promote the feedback function.

Keywords

Performance prediction, Knowledge representation, TabNet, Behavior representation

1. Introduction

With the continuous development of school online digital platform construction, students have generated a lot of data of learning life, but the data accumulated in the system as an asset has not yet fully released its value. Making good use of educational data resources with the help of data mining technology could provide scientific guidance and valuable decision-making support for school teaching, student learning and other tasks[1]. It is unable to intervene in learners' bad learning behaviors in time and cannot provide help when some students have poor learning results. If we can predict that students are at risk of failing a class, and analyze learning method through data mining, better support can be used to guarantee learning result[2].

At present, many research methods focus on handling relational model of learning behavior by machine learning[3][3][4], finding key influencing factors through correlation analysis of behavioral statistical data, or using NLP model with the time series data of behavior type or text[6]. A major challenge is that correlation analysis is difficult to find complex relationships, and the DNN model that processes time series data has weak interpretability. Based on a relatively intuitive understanding, we can see that the process of students interacting with teaching materials during the learning process records a process of students acquiring knowledge. Here we proposed a method to predict the grade by learning behavior and corresponding knowledge, which can obtain richer information and interpretability at the same time.

LAK-WS 2024: Joint Proceedings of LAK 2024 Workshops, March 18–19, Kyoto, Japan



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

2. Related work

Study performance individual difference

The modern humanistic point of view pays more attention to the development of learners' learning attitude and learning motivation in the learning process. understanding learners' characteristics and learning willingness to meet learners' learning interests and learning resources[7]. According to humanistic behavioral theory, through the analysis of student behavior to judge the result, one of the key points is to obtain the development process of the student's learning motivation and learning attitude from the student behavior data[8]. The method we propose is to better reflect these differences in feature representation.

2.1. TabNet

Students' performance is recorded as tabular data, and there were some widely used DNN model for this kind of data. TabNet[9] is designed, which not only absorbs the advantages of the tree model, but also inherits the advantages of DNN. TabNet has been designed based on retaining the end-to-end and representation learning characteristics of DNN. It also has the advantages of tree model and sparse feature selection, which makes it suitable for interpretability. We choose TabNet DNN structure to process tabular data of learning behaviors, which can automatically mine learning method pattern information.

2.1.1. Feature selecting process

TabNet has a decision-tree function as the Figure 1 :

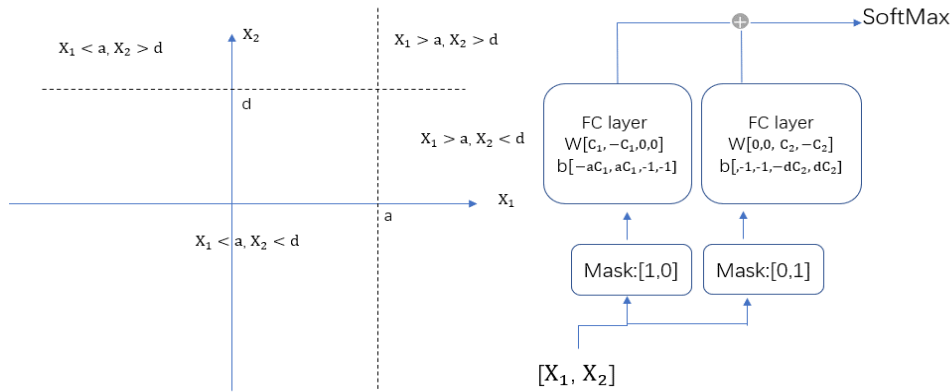


Figure 1: Embodying the decision-making process

TabNet also uses the idea of sequential attention to imitate the behavior of decision trees. Each step of TabNet's decision-making module will allocate attention through mask function and obtain feature information by selecting features that have a greater impact on the results. Its interpretability can be defined by the formula below. $M_{b,j}$ of each decision can be analyzed as pattern information, and the final $M_{agg-b,j}$ can be used as the overall attention attributes for correlation analysis and other functions.

$$M_{agg-b,j} = \frac{\sum_{i=1}^{N_{steps}} \eta_b[i] M_{b,j}[i]}{\sum_{j=1}^D \sum_{i=1}^{N_{steps}} \eta_b[i] M_{b,j}[i]} \quad (1)$$

3. Study aim

The purpose of this study is based on the data of the existing digital learning platform, to establish an effective system that combines course materials and corresponding learning records to predict students' learning results. Providing well interpretability at the same time. In this proposed system, our goal is to explore the relationship between knowledge of different topics and learning performance in vectorized learning materials and establish a connected module to help obtain the feature representation of learning results.

4. Proposed method

4.1. Structure of proposed model

The focus of this study is to combine learning materials and learning performance to establish an effective system to predict student learning outcomes, providing a reference for the analysis of teaching and student learning situations.

From the system overview in Figure 2, we can see that the system is divided into two parts: the knowledge representation part and the learning performance part. Each student's learning materials and online learning operation behavior records are used as input, and the final score is used as a label. The knowledge representation part is mainly responsible for converting learning materials into knowledge vectors. These knowledge vectors contain some connections and correlations between various knowledge points of learning. The learning performance part is mainly responsible for converting the operation records of the student's learning process into feature vectors of the learning performance of learning this part of knowledge. Then, by exploring the effect of the learning results on the knowledge vector, a knowledge vector representing the learning results is obtained, and performance predictions are made based on this. The output is the probability of each student achieving different grades. Because there are few direct connections between material from different courses, experiments can only be conducted for the same course. During training, the model records the data of some students in the same course as a training set and evaluates the effectiveness of the model by predicting the performance of the remaining students.

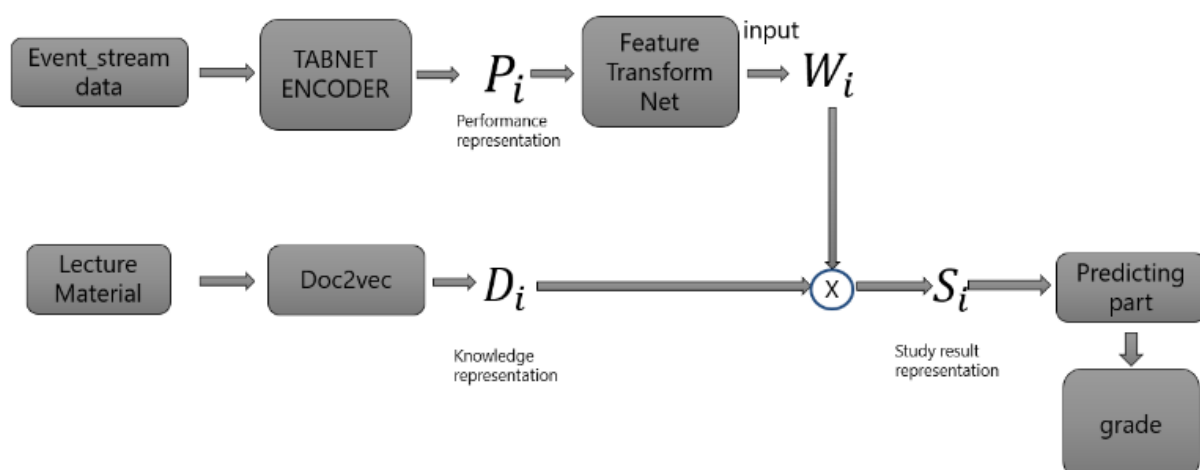


Figure 2: The structure of proposed method

4.2. Preprocessing slide material

A typical teaching slide, which briefly expresses the main knowledge points of the course content, such as the meaning of some concepts, described formulas and relationships, etc. Some parts will also describe knowledge points through some pictures. Here we use pure text parts to summarize the main knowledge points representing the teaching materials. Mainly from the perspective of specific nouns and the relationships between them, we explore the internal relationships and interrelationships between knowledge points on different topics. When making textbooks, the main knowledge points in different chapters are somewhat different, but at the same time they are also connected according to certain correlations. Therefore, under normal circumstances, the internal relationship and correlation of knowledge points in different chapters are reflected to a certain extent through the distribution of nouns related to the knowledge points. So here we use all the text data in the course slides, first perform certain basic processing, and using Doc2Vec to convert it into a vector expression of word embedding. These word embeddings express information about the structure and content of a certain degree of course knowledge. Therefore, it can reflect the differences in the distribution ratio and structure of knowledge points in different chapters, which could also be regarded as sets of knowledge vector.

The original data like Figure 3 mainly includes the specific course chapter information of each course, the start and end time of each chapter, the student ID of each course, the event stream records of all students operating on the online platform, the operation part mainly including events such as turning pages forward and backward, clicking on links, annotations, notes, etc[10][11].

userid	contentsid	operati	pageno	marker	device	eventtime	userid	contentsid	pageno	num_visits	CLOSE	OPEN	NEXT	PREV	...	
F-2021_U1	F-2021_C14	CLOSE	11		pc	2021/10/7 11:30	0	F-2021_U10	F-2021_C1	1	4	0	4	4	0	...
F-2021_U2	F-2021_C16	OPEN	1		pc	2022/1/19 4:52	1	F-2021_U10	F-2021_C1	3	1	0	0	0	1	...
F-2021_U2	F-2021_C16	NEXT	1		pc	2022/1/19 4:52	2	F-2021_U10	F-2021_C1	4	4	0	0	4	0	...
F-2021_U2	F-2021_C16	CLOSE	2		pc	2022/1/19 4:53	3	F-2021_U10	F-2021_C1	5	3	0	0	2	1	...
F-2021_U3	F-2021_C26	PREV	16		pc	2021/11/25 23:5	4	F-2021_U10	F-2021_C1	6	4	0	0	3	1	...
F-2021_U3	F-2021_C26	PREV	14		pc	2021/11/25 23:5	
F-2021_U3	F-2021_C26	PREV	12		pc	2021/11/25 23:5										
F-2021_U3	F-2021_C26	PREV	11		pc	2021/11/25 23:5										
F-2021_U3	F-2021_C26	PREV	10		pc	2021/11/25 23:5										
F-2021_U3	F-2021_C26	PREV	9		pc	2021/11/25 23:5										

Figure 3: the original data of event stream and the structured data by OpenLA

Here we structured data through the API OpenLA[12] which is developed by our research laboratory. The original data generated by the learning platform is the information generated by all operations saved based on time, including the username of the operation, time, operation type, etc., and using OpenLA, we can obtain the structured data we need according to specific needs. As you can see from the Figure 4, what we use here is the statistical record of the operational behaviors of all students during weekly study. Including forward, backward, add bookmarks, jump and so on.

4.3. The concept of knowledge vectors

As mentioned above that the meaningful words of each chapter of the course can represent information about the content and structure of a certain degree of knowledge. On the other hand, the learning process is the process of integrating new knowledge into one's own knowledge structure, which is the mastery of the content of knowledge itself and the correlation between knowledge. Therefore, based on these analyses, it is considered here that knowledge (descriptive words) embedding vector can be used as a standard to measure the effect of learning.

The materials studied by all students are consistent, that is, the knowledge content being learned is consistent. However, when different students study, their mastery of different knowledge points will be different, and the establishment of relationships between knowledge will also be different, so in the end there will be many differences in the learning results under the same learning materials. From this perspective, analyzing students' learning performance

based on learning results based on knowledge content has a certain theoretical basis and is also consistent with some qualitative analysis results.

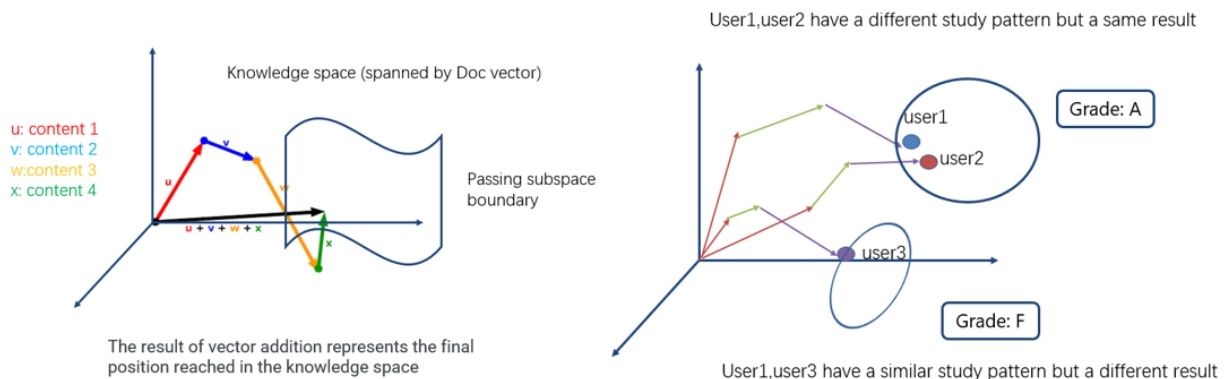


Figure 4: The study process from knowledge vector perspective

From the Figure 4, we can see that if we look at the entire learning process from the perspective of knowledge vectors, the learning process is like a process of moving forward in the knowledge space which is containing all the knowledge vector of the course. In the beginning, it is at the origin of this knowledge space, which means not mastering any knowledge, after each period of study, students will reach a new position in the space. The boundary represents that certain dimensions reach an appropriate range, and students can get a good understanding of the course and pass the course.

When embedding the knowledge vector, the dimension should be considered with the amount of information of the text. More dimension means more complex relationship of knowledge. For not enough words in the slide, we set the dimension to 100.

The materials in each chapter are consistent, that is, the knowledge vectors being learned are consistent. However, if the influence of learning ability and learning habits is mapped on the knowledge vector, based on the standard vector that is being learned, different proportions will be obtained in different dimensions. Specifically from the right part of Figure 4, for example, user 1 and user 2 here have very different learning modes, so the direction and distance of learning the same chapter in the knowledge space are quite different, but in the end, a similar total learning effect has been achieved; and user 1 and user 3 have similar learning patterns, so they have similar directions in each study, but due to different learning abilities, time spent and other reasons, the positions they end up are completely different, and the final results are naturally different.

4.4. Feature encoding for learning performance

In the past, data analysis of learning behavior was generally mainly combined with manual analysis. Selecting the part that we need from the various information existing in the learning operation data and combine it with other analysis models based on learning behavior for verification.

The main advantages of using TabNet to analyze learning behavior lie in two points. Firstly, no preprocessing is required, reducing manual operations in the process, and providing a more convenient way of use for the overall system design; Secondly, it can automatically display the data in tabular form the ability to automatically select features and combination relationships from the data. That means it could automatically search for student behavior patterns and behavioral rules in records, and automatically discover students' implicit learning motivations and learning habits and other information.

The specific process is to use students' operational data as input and student scores as labels for training. Finally, after training, the encoder structure of the trained TabNet is used to convert

the student's operation data into a dense vector that meets the usage requirements. We can see by Figure 5, TabNet model will take multiple steps. The process of each step is like the decision-making process of a decision tree. Different attentions are used to obtain learning performance information with different pattern information. The vectors obtained by the encoder in all steps are combined into a new vector. This new vector contains learning performance information based on different patterns, so it also contains richer pattern information of the learning process .

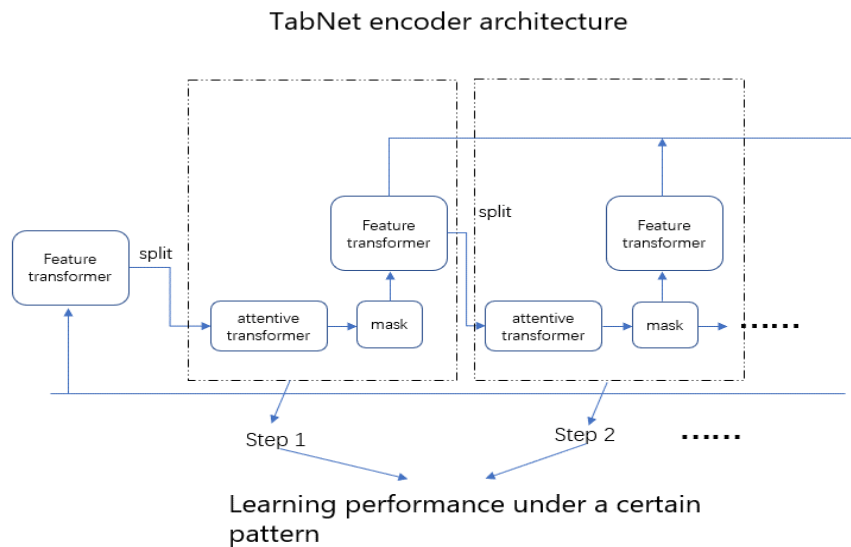


Figure 5: the multiple decision making step of TabNet

4.5. Reflecting learning process on knowledge vector space

As mentioned earlier, the knowledge vector represented by word embedding contains the structural relationship of the learned knowledge on different topics, and the knowledge vectors of different topics have different emphases in different dimensions like Figure 6. The knowledge vectors of some chapters will be larger in some dimensions. proportion. Therefore, we can speculate that if the knowledge of certain topics is better learned during the actual learning process, the results in these dimensions will also be better.

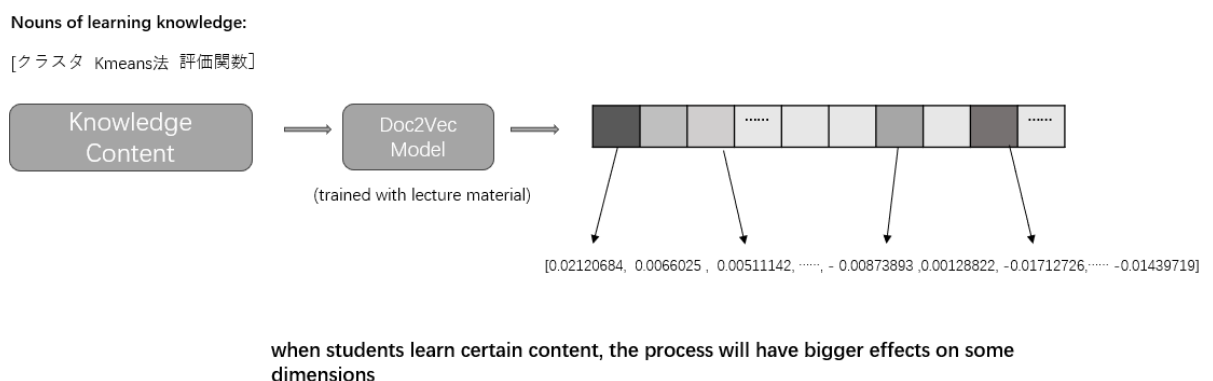


Figure 6: the learning contents have different effects among dimensions

Therefore, we can try to explore the impact of performance feature on specific parts of knowledge, and these different knowledges have different emphases in different dimensions. So here we can try to translate it into the impact of focusing on different dimensions of the knowledge space. The method we propose here is to convert the characteristics of learning performance into weights of different dimensions in the knowledge space, and then multiply the weights and the corresponding knowledge vectors in each dimension to obtain the final learned knowledge vector.

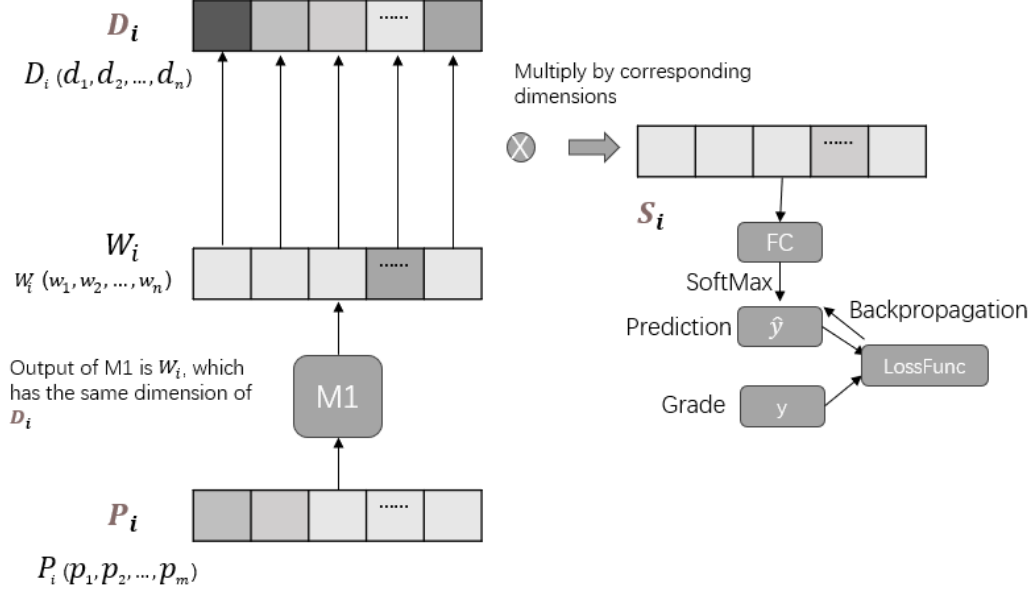


Figure 7: the training process of the network

Here we express this mapping by training a MLP model 'M1'. The training method is as Figure 7 Taking the learning performance feature P_i as input, and finally the output of this MLP has the same dimension as the knowledge vector. This output is the effect of learning performance in different dimensions on the knowledge vector, that is the weight value $W_i (w_1, w_2, \dots, w_n)$. The better the learning effect, the closer this effect weight value will be to 1, and the final knowledge vector will be closer to the standard value. By multiplying the output of MLP and the knowledge vector in one-to-one correspondence in each dimension, the vector expression S_i of the learned knowledge is obtained. This output would be transferred through another fully connected layer, and the probability of the final score of this knowledge vector is obtained through SoftMax. The better the training effect, the better the effect of M1 on reflecting the influence of learning feature vectors to knowledge vectors.

$$S_i = D_i \otimes W_i \quad (2)$$

$$Porbability = softmax(FC(S_i)) \quad (3)$$

5. Experiment

To evaluate the effectiveness of the designed model, here we conducted a series of experiments to verify our method. The experiment used data of BookRoll reading behavior datasets and all slide text in a CS course, recording the log data of all student operation events. These mainly include operational events such as turning pages forward and backward, clicking links, marking, and taking notes. Through 5-fold cross-validation, training sets and validation sets are randomly selected from groups with different scores in different proportions. At the same time, in a more practical way, according to the standard of whether they are in the danger zone,

students with different grades are finally classified into two categories to test the effect of the model. Because in the overall data, there is a gap of nearly 10 times between students in the danger zone and non-danger zone, so when setting the task indicators, students in the danger zone are set as label 1 to mainly examine the performance of predicting in risk crisis. It can more effectively identify students who need to provide early warning. The second experiment is using same prediction model, inputting knowledge representation of our method and behavior data separately, to verify whether our method extract more meaningful information for students with different behavior.

5.1. Result

To effectively evaluate the prediction quality of this knowledge vector model, here we selected traditional machine learning algorithms commonly used in other similar studies in the past as comparison baselines, including logistic regression (LR), extreme gradient boosting (XGBOOST), Lightgbm, random forest.

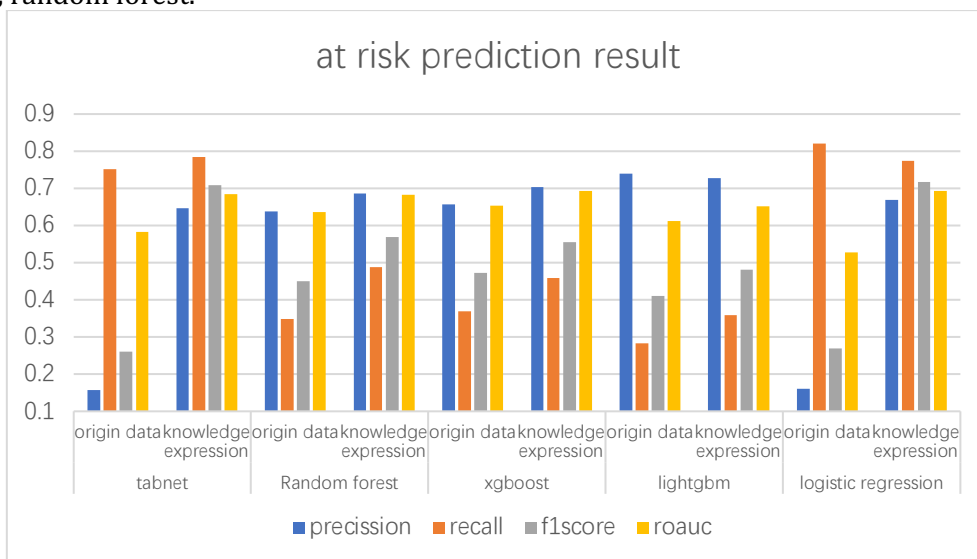


Figure 7: Comparison of at risk student prediction

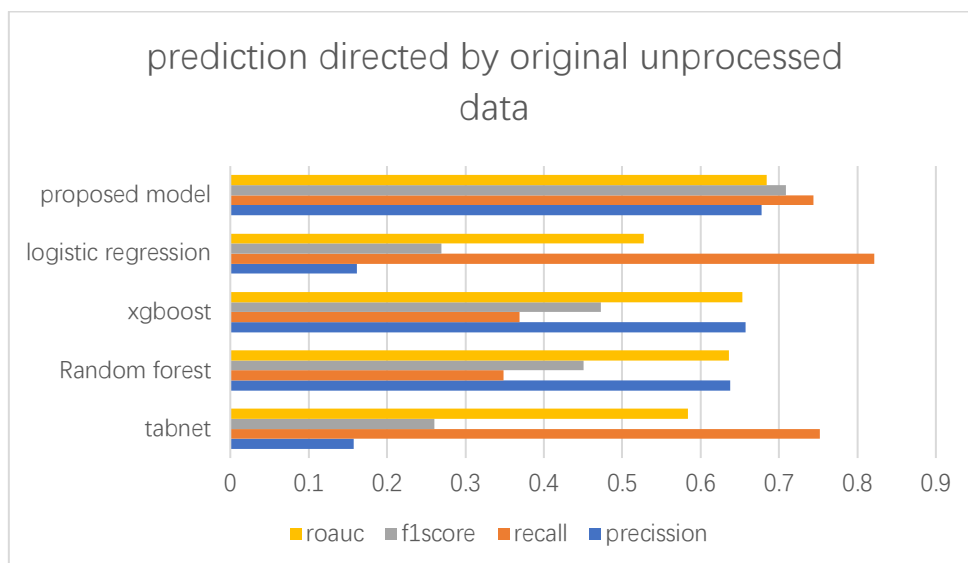


Figure 8: comparison from roauc aspective

The algorithm of knowledge vector proposed here have a certain improvement than the result predicted by behavior data alone. Performance improvements have been achieved in both F1

score and AUROC. Result shows our knowledge representation is more efficient than behavioral data only, which has ability to explore individual differences such as different grade of similar behaviors, or same grade of different behavior. It is worth mentioning that when the model we proposed predicts the learned knowledge vector, it only uses a simple fully connected layer to predict the results, which shows that the knowledge representation has a better classification boundary when predicting multiple grade categories. At risk students who were consistently mispredicted using behavioral data alone were more likely to be correctly identified by our method.

5.2. Pattern information of TabNet

From Figure 11 we can see that for the most important three features in step1 and step2 of TabNet’s decision steps, the coefficient of variation of students with good learning effects is larger, which means that the features discovered by TabNet are meaningful. If we conduct a qualitative analysis of "NOTGETIT, BOOKMARKJUMP, SEARCHJUMP", we can guess that the explanation of this pattern information is: For students with strong learning initiative, they will be more proactive in learning difficult-to-understand knowledge points after marking them. They may learn relevant content by searching. Therefore, this learning behavior pattern may correspond to the difficult knowledge in each chapter. So, we provide feedback on students' learning behavior and help improve learning results by counting the records of better performance in these patterns and higher performance records of their own.

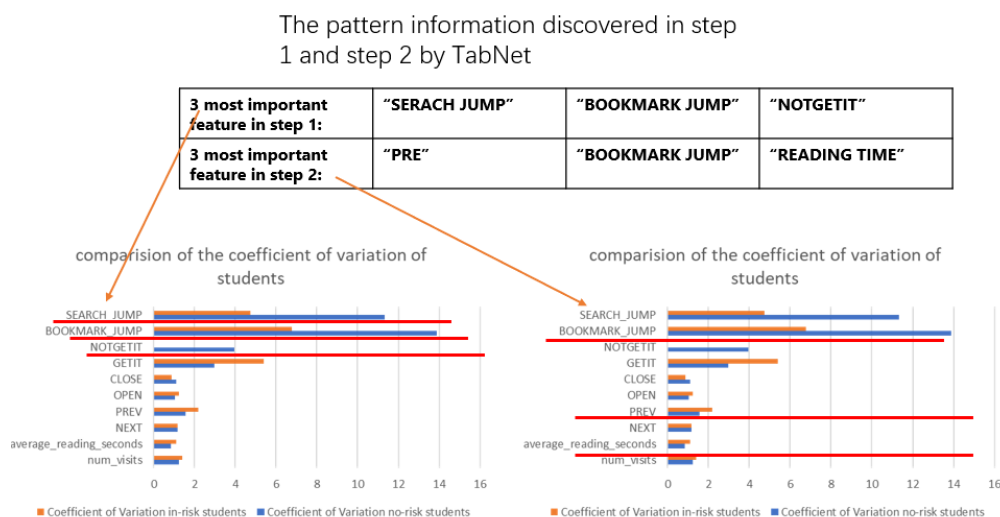


Figure 9: The possible pattern information discovered by TabNet

6. Conclusion

In this paper we hope to predict student learning result by combining the course material data of the digital education platform and the operation record data of student learning process. By introducing the idea of learning knowledge expression, we can dig out the result expression of learning knowledge from the materials to be learned and the learning process records. Combined with TabNet's advantages of not requiring preprocessing and interpretability on tabular data, it can provide better help for subsequent analysis and feedback work. It can be seen from the experimental results that the prediction model established here provides good prediction results.

As the current structure is not suitable for different courses’ grade predicting. For more different courses, it is necessary to establish a set of methods to effectively handle the knowledge expression of all courses together. Based on the current excellent performance of the BERT model in large-scale text expression, in the future we hope to explore further using the BERT model to adapt to prediction methods for different courses in terms of knowledge expression.

7. References

- [1] Cheng P, Ding R. The effect of online review exercises on student course engagement and learning performance: A case study of an introductory financial accounting course at an international joint venture university[J]. *Journal of Accounting Education*, 2021, 54: 100699.
- [2] Mueen A, Zafar B, Manzoor U. Modeling and predicting students' academic performance using data mining techniques[J]. *International Journal of Modern Education and Computer Science*, 2016, 8(11): 36.
- [3] Jia, Y. ., & Zhao, Q. (2022). The Learning Behavior Analysis of Online Vocational Education Students and Learning Resource Recommendation Based on Big Data. *International Journal of Emerging Technologies in Learning (IJET)*, 17(20), pp. 261–273.
- [4] Hu Q, Rangwala H. Reliable deep grade prediction with uncertainty estimation[C]//*Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. 2019: 76-85.
- [5] Hussain S, Muhsin Z, Salal Y, et al. Prediction model on student performance based on internal assessment using deep learning[J]. *International Journal of Emerging Technologies in Learning*, 2019, 14(8).
- [6] Minn S. BKT-LSTM: Efficient Student Modeling for knowledge tracing and student performance prediction[J]. *arXiv preprint arXiv:2012.12218*, 2020.
- [7] Bouton M E. Context, attention, and the switch between habit and goal-direction in behavior[J]. *Learning & behavior*, 2021, 49(4): 349-362.
- [8] Zulfa M N M, Setiawan D, Fardani M A. Analysis of Habit Patterns in Academic Behavior in Student Learning Discussions[J]. *International Journal of Elementary Education*, 2020, 4(3): 392-399.
- [9] Arik S Ö, Pfister T. TabNet: Attentive interpretable tabular learning[C]//*Proceedings of the AAAI conference on artificial intelligence*. 2021, 35(8): 6679-6687.
- [10] Flanagan, B., Ogata, H. (2018). *Learning Analytics Platform in Higher Education in Japan, Knowledge Management & E-Learning (KM&EL)*, Vol.10, No.4, pp.469-484.
- [11] Lu, O.H., Huang, A.Y., Flanagan, B., Ogata, H., & Yang, S.J. (2022). A Quality Data Set for Data Challenge: Featuring 160 Students' Learning Behaviors and Learning Strategies in a Programming Course. *30th International Conference on Computers in Education Conference Proceedings*, pp. 64-73.
- [12] Murata, R., Minematsu, T., Shimada, A. (2020). OpenLA: Library for Efficient E-book Log Analysis and Accelerating Learning Analytics. In *International Conference on Computer in Education (ICCE 2020)*, pp. 301-306.