SCAN ALGORITHM TO OPTIMIZE DYNAMIC INSTRUCTIONAL INFORMATION NETWORK FOR PREDICTING STUDENT BEHAVIOR

JUNTING TENG*

Abstract. With the integration of work, study and life on campus taking shape gradually in major universities across China, the concept of education from classroom to life is more widely accepted and the smart education model has become a trend in education informatization. However, most of the existing studies are not applicable to student behaviour data in the campus environment, and the temporal as well as cyclical characteristics in the data are not accessible in the student behaviour prediction problem. In this study, a hypergraph-based dynamic campus behaviour information network is designed to address the needs of the student behaviour prediction problem, and a student campus behaviour prediction algorithm is proposed in the dynamic campus behaviour information network-based student behaviour prediction problem. The effectiveness and rationality of the algorithm is verified through experiments with real campus data sets. The experimental results demonstrated that the periodic nature of the data acquired by the cycle gated cyclic unit module needs to be built on top of the snapshot gated cyclic unit module to help the algorithm achieve better results. The area under the curve of the algorithm proposed in the study achieves more advantageous results on both the 21-day and 35-day datasets. The cycle gated cyclic unit module of the student campus behaviour prediction algorithm proposed in this study can more effectively extract the cyclic features present in the dynamic campus behaviour information network and better accomplish the prediction of student campus behaviour.

Key words: Dynamic instructional information networks; Student behaviour prediction; Cycle gated cyclic units; Snapshot gated cyclic units

1. Introduction. With the release of the national standard "General Framework for Smart Campus", the construction of a smart campus that integrates campus work, learning and life has been gradually carried out and taken shape in major universities across China [1]. The concept of education from classroom to life has been accepted by the public, and traditional teaching concepts and methods can no longer meet the demand for personalized training of current innovative talents [2]. The smart education model, which takes advantage of information technology such as big data analysis and artificial intelligence, has become a trend for new types of education. By analyzing students' interests, hobbies, habits and other behaviours, it has a significant impact on their personal development and academic performance [3]. The key to predicting student behaviour is an understanding of behavioural patterns in students' habits. In today's smart campuses, students' daily behaviours are recorded through information technology [4]. These data records, which truly reflect the habits and behavioural patterns of students within the campus environment, contain information on the number of students, locations, times and events. Campus behavioural data is dense, multi-sourced, dynamic and cyclical in nature, and as such, a network structure is a widely used approach to data modelling today [5]. Based on the analysis of the characteristics of campus behavior data, this study constructed a multi-source campus behavior information network for student search behavior to more accurately predict student behavior. To solve the problem of predicting student behavior in dynamic campus behavior information networks, a SCAN algorithm based on dynamic behavior information prediction is proposed. Research and design a Similar Campus Lifestyle Student Miner (SCLSM) algorithm that integrates multi-source behavior information mining and student similarity analysis to address the problem of similar student search in multi-source campus behavior information networks. Processing high-dimensional data: Because in a smart campus environment, student behavior data usually contains multiple types and dimensions of information. These data are high dimensional and are characterized by high complexity, dynamics and periodicity. Therefore, SCAN algorithm is selected in

---

*Department of Economic Management, North China Electric Power University, Baoding, 071000, China (`tengjunting2023@163.com`)

this study because SCAN algorithm can effectively process high-dimensional and complex data. Secondly, it can capture this dynamic nature, and can effectively integrate these multi-source information, and carry out in-depth mining and analysis. Moreover, the cyclic neural network module in SCAN algorithm can effectively process time series data and capture the dynamic characteristics of students' behaviors over time. However, other machine learning algorithms are often only able to process data from a single source and cannot efficiently process time series data or capture these dynamic features. This study innovatively integrates multi-source behavioral data of students and conducts deep mining and analysis based on the SCAN algorithm to construct a graph of student behavior characteristics. This study provides more effective support for further improving the accuracy of student behavior prediction and school education management. The research is divided into three main parts, the first of which analyses the characteristics of campus behavioural data as they exist. A multi-source campus behaviour information network is designed for the student search problem, and a dynamic campus behaviour information network is designed for the student behaviour prediction problem. The second part is the implementation of the SCAN algorithm for student behaviour prediction in a dynamic campus behaviour information network. Each of the individual structures in the algorithm framework is explained, and two feasible SCAN algorithm framework designs are also designed. The third part is an experimental recording of student behaviour on campus to verify the feasibility of the SCAN algorithm and the effectiveness of the SCAN-S and SCAN-P framework structures on a real data set.

2. Related works. With the development of information technology such as big data analysis and artificial intelligence, the smart education model has become a trend in education informatics, which can better facilitate personal and social development by predicting student behaviour. Wang et al. equationted behavioural entries as a set of contextual items and proposed a multi-type item set embedding learning method. Finding complementary contexts for people to make their behaviour more likely to succeed. The proposed method was shown to be both efficient and applicable when compared to state-of-the-art behavioural prediction and contextual recommendation methods [6]. Yang X et al. proposed a method for recommending student learning resources based on a bidirectional long and short-term memory recurrent neural network. Based on the deep learning model, learning analysis techniques are used to construct students' specific learning patterns and provide appropriate learning resources. The experimental results show that the student adaptive learning system of the deep learning model has better stability and interpretability in terms of recommendation results [7]. Zhang M et al. proposed a learning pattern analysis method to predict academic performance by describing students' beliefs about learning and their motivation, and by analyzing individual students' learning patterns. The results of the study showed that the proposed method could effectively extract micro-level behavioural information and produced better prediction results [8]. Ji et al. proposed a Bayesian inference and learning scheme. The problem of target tracking under completely unknown maneuvers was investigated to cope with complex target manoeuvres. Tests showed that the proposed tracker can significantly reduce the peak tracking error, which is one of the most important performance indicators to prevent tracking failure [9]. Joshi A et al. proposed to integrate a model with robust machine learning and evaluated a prediction model. This model improves the accuracy of previous student performance predictions through identification. The results of the study showed that our integrated machine learning model Cat Boost outperformed standard machine learning models with an accuracy of 92.27%. This new model was able to prove itself reliable through the use of shocks and hyperparameter optimization, which proved to be a valuable method and approach [10]. Zhang Z et al. proposed an extensible semantic description structure. On this basis, a heterogeneous information embedding framework based on CMG is designed. In the fusion process, an attentional mechanism is used to automatically learn the weights of these potential vectors so that each final node representation focuses on the proximity of the most information-rich order. Experimental results show that the proposed method is superior to the most advanced isomorphic and heterogeneous network embedding methods in three network mining tasks: node classification, node clustering and node similarity search [11].

With the development of information technology, people's habits and behavioural patterns are recorded in the form of data, and it has become a trend to use network structures and intelligent algorithms to solve problems in large environments. Gardoni et al. tested the effect of different biological deacidifications on volatile and non-volatile components. Malolactic fermentation was carried out with brewer's yeast inoculated with Brettanomyces and Lactobacillus plantarum strains. Different fermentation treatments were carried out in 0.75 L

vessels. The results showed that the concentration of higher alcohols, fatty acids and acetic acid was reduced and that the simultaneous fermentation of alcohol and malolactic acid reduced malic acid by approximately 80% [12]. Huang Z et al. based on Convolutional Neural Networks (CNN) for a permeability inversion scheme. A single PET scan from a radiotracer pulse injection experiment was used as input. The results showed that the scheme provides an unprecedented method for effectively characterizing multi-scale permeability inhomogeneities in complex geological samples [13]. Prigioniero et al. explored the relationship between particulate matter, PAH leaf concentrations, uptake rates and foliar functional traits in four Mediterranean evergreen trees. uptake of PM10 was positively correlated with interspecies and upper cuticle thickness, and PM2.5 was positively correlated. The uptake of different components of PAHs was usually weakly correlated with different leaf functional traits. Experimental results indicated that both plant surface morphology and chemical leaf characteristics affect the retention of PM and PAHs [14]. Meng et al. proposed a new ironing technique to study a mechanism design problem with network agents and stochastic evolutionary private information. The network intervention problem is further discussed based on the optimal dynamic mechanism obtained to define some important nodes and edges in the network under different ranges of synergy parameters, i.e., subjects can intervene to change the ex-ante distribution of individual types [15]. Khan A. et al. proposed A least significant steganography method to hide secret information into the original image. A convolutional neural network was used to train the model to detect and extract patterns of hidden features in the image, and to classify the steganographic original image and cover image. The experimental results show that the proposed scheme can realize both information hiding and information revealing, with an accuracy of 95.1%, and the model is robust in terms of efficiency [16]. Qin X et al. developed a new structured Bayesian interactive analysis method to effectively integrate network information. Using Bayesian technique, an effective variable dB Bayesian expectation maximization algorithm is designed to explore the posterior distribution. Experimental studies show that the results are biologically sensitive, and the prediction accuracy and selection stability are satisfactory [17].

In summary, the prediction of people's behaviour within the wider environment in everyday life can contribute to the development of people as individuals as well as to the progress of society. However, existing research still suffers from semantic deficits and insufficient temporal and periodic features in the face of similar behaviour search problems, as well as behaviour prediction problems. Therefore, further research is still needed to face the student behaviour prediction problem to help the student behaviour prediction goal to be better accomplished.

3. Prediction of Student Behaviour Based on Dynamic Instructional Information Networks. For different groups of students, the network is flexibly used to represent their behavioural data, to design a multi-source campus behavioural information network suitable for student search, and a dynamic campus behavioural information network suitable for student behaviour prediction. This chapter will focus on analyzing the characteristics of students' behavioural data at school to provide a basis and scheme for this research, which will be validated by means of experiments.

3.1. Research on campus behavioral information network model. Campus behavior data is a record of students' behavior on campus, which is composed of people, time, place and events [18, 19]. However, students' behaviors on campus occupy a relatively large proportion of students' life, resulting in a large number of recorded behaviors, complex, and high intensity. Therefore, this study designed a campus behavior information network model. Among them, the single-source campus behavior information network is the campus behavior data built through a single source, and its information network is shown in Figure 3.1.

There are five types of binding for the single-source campus behavioural information network, such as student ( ), behaviour ( ), time ( ), place ( ) and event ( ). There are four types of articulation and weight types, relationships between students and instances, relationships between instances and time, relationships between behaviours and locations, and relationships between behaviours and events, and each relationship has a weight value of 1. However, student behavior changes with time, and the single source campus behavior information network cannot predict student behavior in time. Therefore, the definition of time node in the single-source campus behavior information network mode is changed, and the network model after the change is shown in Figure 3.2.

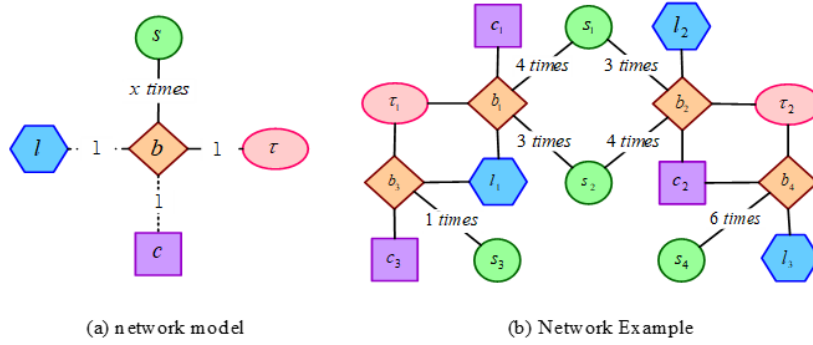Use hypergraphs to represent relationships between students, places, and events. The date indicates an

Fig. 3.1: A Single Source Campus Behavioral Information Network
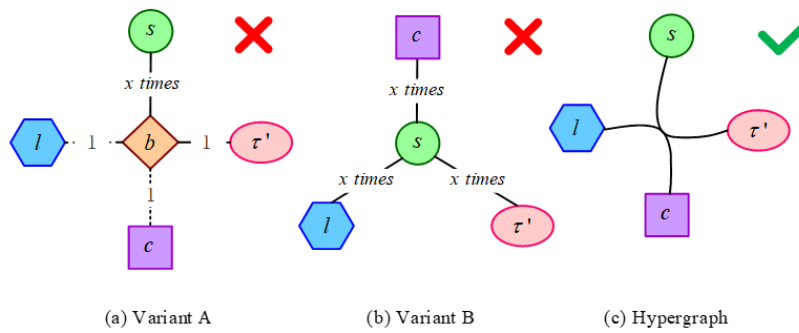

Fig. 3.2: Time node defines the single-source campus behavior information network model after the replacement

hour in a day. However, the data source of campus behavior data is highly variable, and the whole network structure needs to be updated constantly, which is not conducive to the expansion of the algorithm. Therefore, it is necessary to combine the dynamic properties of student behavior prediction on the basis of single source campus information network structure. While retaining the benefits of a behavioral information network, you can also retain additional data when the data source changes.

3.2. Research on SCAN algorithm based on dynamic behavior information prediction. Aiming at the problem of Student Behavior Prediction in dynamic Campus network, the SCAN algorithm of Student Campus Behavior Prediction based on Hypergraph Information Propagation (HIP) is studied [20, 21]. The algorithm combines the connection prediction of hypergraphs with dynamic features and the extraction of periodic features. Among them, It includes HIP module, Gated Recurrent Unit-Snapshot (GRUs), Gated Recurrent Unit-Period (GRUp), and Hyperlink interactive scoring module Interaction Scoring,INT). The flow framework of SCAN algorithm is shown in Figure 3.3.

HIP first transforms the super-edge into a similar ordinary graph structure during the computation against the super-edge to facilitate subsequent computations [22]. The SCAN algorithm, on the other hand, uses the graph convolution module to process the clusters formed by the hyperedges and in this way obtains the vector representation corresponding to each node in the hyperedges. Since the direct use of the regimented representation of student behaviour leads to ambiguity, the representation of each node in is obtained for a given hyperedge using information propagation. The equation for this is shown in equation 3.1.

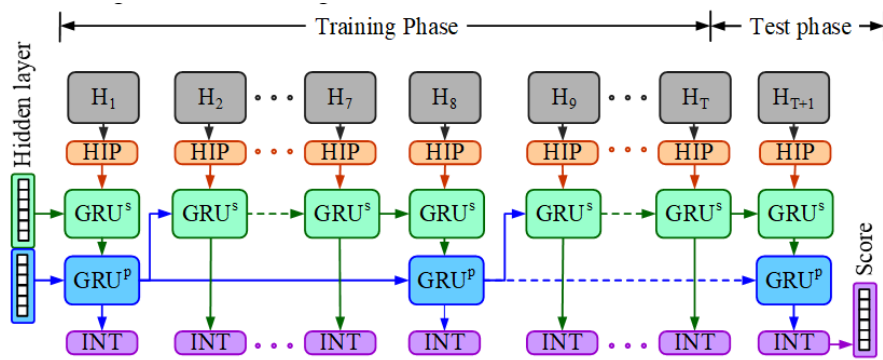$$g_v^\varepsilon = W_{\text{clique}} \sum_{u \in \varepsilon - \{v\}} r_u + b_{\text{clique}} \tag{3.1}$$

Fig. 3.3: SCAN algorithm process framework

In equation 3.1, u is denoted as the nodes in the hyperedge ϵ except v, r_u is denoted as the initial vector representation of node u, and BBB and b_{clique} are denoted as the parameters to be trained. Vector representations of nodes other than the current node are added as representations of nodes for information propagation. At the same time, the initial vector representation is also used for information propagation, so that more information can be retained during calculation. Therefore, a is added to the information dissemination part carried out by the initial characterization. Its calculation formula is shown in equation 3.2.

$$l_v = W_{\text{self}} r_v + b_{\text{self}} \tag{3.2}$$

In equation 3.2, r_u is represented as the initial vector representation of node AA, and W_{self}, b_{self} are represented as the parameters to be trained. The campus behavior information network has the characteristics of temporal relationship, so it is necessary to extract the structure of temporal characteristics to obtain the change of student behavior with time. In the SCAN algorithm framework, the cyclic units of recurrent neural network are used, including RNN units, LSTM units and GRU units. These units can help to better understand student behavior patterns. The mathematical expression of RNN unit structure is shown in equation 3.3.

$$h_t = \tanh\left(W \cdot [h_{t-1}, x_t] + b_n\right) \tag{3.3}$$

In equation 3.3, h_{t-1} is the hidden layer of the RNN at the previous moment, x_t is the input at the current moment, $tanh$ is the activation function, W and b_n are the parameters that the RNN needs to be trained. The Z_t mathematical expression of the unit structure in GRU is shown in equation 3.4.

$$Z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t] + b_z\right) \tag{3.4}$$

In equation 3.4, h_{t-1} is the previous moment hidden layer, x_t is the current moment input, σ is the activation function, W_z and b_z are the parameters to be trained by GRU, and the structural equation of Z_t unit in GRU is shown in equation 3.5.

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t] + b_r\right) \tag{3.5}$$

In equation 3.5, h_{t-1} is the hidden layer at the previous moment, x_t is the input at the current moment, σ is the activation function, and W_r, b_r are the parameters to be trained by GRU. The SCAN algorithm uses a clustering method based on shared neighbors. For any pair of nodes u and v in graph G whose similarity is higher than a given threshold, the two nodes are considered similar. The similarity here is calculated by the
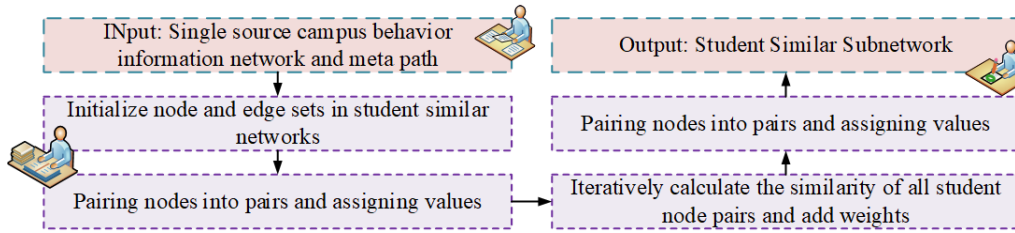
Fig. 3.4: The construction steps of a similar student network

intersection of the set of neighbor nodes of nodes u and v. In this process, Jaccard similarity is used as a metric, and its mathematical expression is shown in equation 3.6.

$$\text{Sim}(u, v) = \frac{|N(u) \cap N(v)|}{|N(u) \cup N(v)|} \tag{3.6}$$

In formula 3.6, $N_{(u)}$ and $N_{(v)}$ represent the set of neighbor nodes of nodes u and v, respectively. The SCAN algorithm divides the time into multiple time Windows, then calculates the similarity of nodes within each time window, and takes the average value of these similarities as the final similarity, the mathematical expression of which is shown in equation 3.7.

$$\text{Sim}(u, v) = \frac{\sum_t w_t \cdot |N(u) \cap N(v)|}{\sum_t w_t \cdot |N(u) \cup N(v)|} \tag{3.7}$$

In formula 3.7, $N_t(u)$ and $N_t(v)$ represent the set of neighbor nodes of nodes u and v in time window t, and w_t is the weight of time window t. Compared with the traditional RNN structure, LSTM increases a lot of parameters, but it occupies too much computing resources and space resources, and GRU does not need to calculate and save the cell state.

3.3. Construction of multi-source behavioral information similar student search model optimized by SCAN algorithm. In order to solve the problem of searching Similar students in multi-source Campus behavior information network, a Similar Campus Lifestyle Student Miner (SCLSM) algorithm combining multi-source behavior information mining and student similarity analysis was designed [7]. By integrating multi-source behavioral data of students, this algorithm carries out in-depth mining and analysis based on SCAN algorithm to build the behavioral feature map of students. At the same time, an efficient similarity search algorithm is introduced to achieve accurate search and recommendation for similar students. The characteristics of selection mainly include students' learning behavior, living habits and other multi-source behavioral information, which can fully reflect the characteristics of students' behavior. By integrating the multi-source behavior data of students, the SCAN algorithm is used for in-depth mining and analysis [24, 25, 26] to build the behavioral feature map of students. In addition, an efficient similarity search algorithm is used to achieve accurate search and recommendation for similar students. Among them, the construction steps of similar student network are shown in Figure 3.4.

Calculating the similarity between students is a key step in constructing the student similarity network, in which the similarity calculation between every two students is independent. However, there is a certain correlation between these characteristics, for example, students' learning behaviors and living habits may influence each other. By calculating the similarity between the features and applying the constrained meta-path similarity calculation method, the correlation between these features is considered and mined. Therefore, by introducing multi-process or multi-thread method, the execution efficiency of the program can be improved during the construction of the similar sub-network connection of a single student. In addition, the process of constructing similar subnetworks of different students is also non-interference, in which the similar networks do not affect each other. Based on the multi-source behavioral information similar student search model optimized by SCAN algorithm, the network model is optimized and calculated to improve the search accuracy and
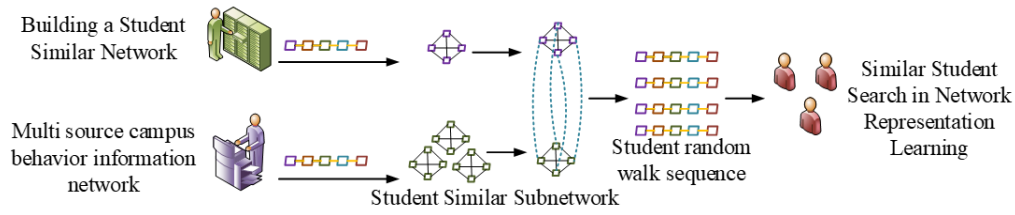
Fig. 3.5: A SCLSM algorithm framework that integrates student similar networks and network representations

meet the needs of campus education management. The SCLSM algorithm framework that integrates students' similar networks and network representations is shown in Figure 3.5.

Based on the correlation between features, a clustering method based on shared neighbors is used to calculate the similarity between students. Moreover, the biased random walk is used to generate the random walk sequence of students on the student similarity network to capture the contextual semantic information of students, and further improve the precision of the similar student search. The constrained meta-path similarity calculation method is adopted in the construction design of student similarity subnetwork. For each single campus behavior information network and each meta-path, the similarity between all students can be calculated. Students are the nodes and similarity is the weight. The biased random walk is performed on the student similar network to generate the student random walk sequence. In the similar student search process based on network representation learning, the random walk sequence of students is taken as the contextual semantics of students to obtain the vector representation of students. Then the similarity with other students is calculated by vector representation, and then sorted to get similar students. For the node vector features output by GRUs and GRUp, the possibility of constituting a super-edge needs to be evaluated using an evaluation function, which is shown in equation 3.8.

$$\text{Score}_\varepsilon = \sigma(W \cdot f(\varepsilon) + b) \tag{3.8}$$

In equation 3.8, Score_ε is denoted as the score of the possibility of the existence of the super-edge, σ is the activation function, and W, b are denoted as the training parameters. $f(\varepsilon)$ is denoted as the function to obtain the ε vector representation through the representation of node $v \in \varepsilon$ in the super-edge. The SCLSM loss calculation equation for each network snapshot H_t is shown in equation 3.9.

$$L_t = \frac{1}{|\varepsilon_t^+|} \sum_{\varepsilon^+ \in \varepsilon_t^+} + \wedge \left(\frac{1}{|\varepsilon_t^-|} \sum_{\varepsilon^- \in \varepsilon_t^-} (\text{Score}_{\varepsilon^-} - \text{Score}_{\varepsilon^+}) \right) \tag{3.9}$$

In equation 3.9, ε_t^+ denotes the hyper-edge (ε_t), H_t denotes the hyper-edge that exists in the momentary network snapshot H_t in the dynamic campus behavioural information network, $\wedge(x)$ denotes a monotonically non-decreasing function, and L_t denotes the loss obtained in the model from the t momentary network snapshot H_t, so the total loss of the dynamic campus behavioural information network is calculated as shown in equation 3.10.

$$L = \sum_{t=1}^{T} L_t \tag{3.10}$$

In equation 3.10, L is denoted as the loss function of the SCLSM algorithm to obtain the smallest L after the final optimization objective. the objective function of the optimization is to have as many super-edge ε_t^+ with ratings higher than the average rating of the super-edge ε_t^- as possible, in order to achieve the objective

Table 4.1: Scale of dynamic campus behavior information network

Number					
Time span	Node	Student	Train Hyperlink	7-days Increases	Test Hyperlink
7-days	2238	2147	45348	45349	6770
14-days	2272	2175	90808	45456	6541
21-days	2491	2395	136468	45654	7483
28-days	2526	2417	186511	50041	7471
35-days	2529	2426	23278	46287	7316

Table 4.2: Comparison between SCAN model and different ML models in AUC scores and Recall@k scores

Data Set	7-days		14-days		21-days		28-days	
Algorithm	AUC	R@k	AUC	R@k	AUC	R@k	AUC	R@k
SVM	0.523	0.23	0.561	0.288	0.622	0.301	0.487	0.234
Logistic Regression	0.527	0.267	0.584	0.295	0.583	0.276	0.556	0.274
Random Forest	0.835	0.423	0.826	0.431	0.797	0.41	0.814	0.406
SCAN-MEAN	0.849	0.417	0.894	0.448	0.877	0.434	0.853	0.397
SCAN	0.875	0.434	0.939	0.475	0.91	0.45	0.878	0.446

of having the ratings of the presence of super-edge ε_t^+ higher than all non-super-edge ε_t^- ratings. In the multi-source behavior information similar student search model optimized by SCAN algorithm, the similarity calculation between each two students is independent. The method takes students as nodes and similarity as weights, and obtains the most similar students through the representation calculation of student vectors, so as to realize campus education management.

4. Validation Test of SCAN Algorithm to Optimize Dynamic Instructional Information Network. This section tests the effectiveness of dynamic campus behavioural information networks in real campus behavioural data, as well as validating the feasibility of the SCAN algorithm for student behavioural prediction. The advantages and disadvantages of two SCAN framework structures, SCAN-P and SCAN-S, are explored in terms of their effectiveness. Data from selected time periods were taken for experimental recording, with data from 7, 14, 21, 28 and 35 days taken as training datasets. The experimentally recorded data sets are shown in Table 4.1.

In table 4.1, students' campus behavior continues to be cyclical. In the first cycle, there was a significant increase in the number of students, and students were more involved in activities. In the third cycle, the number of students' campus behaviors was more stable. In the fourth cycle, the number of students' campus behavior tended to increase again, but in the fifth cycle, it showed a decline. For the reliability of student behavior prediction, the SCAN model was compared with other successful ML models, SVM, Logistic Regression and Random Forest model. There are two evaluation indicators used in the experiment, namely AUC and Recall@k, to verify the validity of the SCAN algorithm. The comparison between SCAN model and different ML models in AUC scores and Recall@k scores is shown in Table 4.2.

As can be seen from Table 4.2, since the SCAN algorithm takes into account the dynamic and periodic characteristics of campus behavior data, the effect of student behavior prediction is significantly better than other ML models. The AUC score of SCAN model is 0.08 higher than that of SVM, 0.12 higher than Logistic Regression, and 0.10 higher than that of Random Forest. On Recall@k, the SCAN model also performs well, which is 0.15 higher than SVM, 0.20 higher than Logistic Regression, and 0.18 higher than Random Forest. It can be seen that the SCAN model shows significant advantages in predicting students' behavior, surpassing other successful ML models in both AUC scores and Recall@k scores, indicating the superiority of the SCAN model in processing such dynamic and periodic data. However, the accuracy of this prediction needs to be improved, and more factors, such as students' personal background and course load, will be considered in subsequent studies to improve the accuracy of the prediction. To this end, in order to verify the soundness of

(a) AUC Score

(b) Recall@ k score

Fig. 4.1: AUC scores of SCANS and its variants and Recall@k fraction

the framework structure of the SCAN algorithm, the results of the two variants, the SCAN-P algorithm and the SCAN-S algorithm, will be compared on the dataset as shown in Figure 4.1.

As can be seen from Figure 4.1, SCAN achieves better results on the main evaluation metric AUC scores, with data only slightly lower than SCAN-P at day 28. On the Recall@k evaluation metric, SCAN-P overtakes SCAN in its dataset at day 21, with a gradual trend towards higher SCAN-P. This shows that the characteristics of the data extracted by the GRUs module are particularly important here. SCAN-S has a more pronounced trend of decreasing model effectiveness when the vector representations of nodes are updated without the GRUp module. As the size of the dataset increases, the results are shown in Figure 4.2 in order to verify the effect of SCAN on efficiency improvement.

From Figure 4.2a, it can be concluded that the SCAN algorithm has good scalability and can still support the prediction of student behaviour on campus when the number of students is 20,000. From Figure 4.2b, it can be concluded that α has little effect on the efficiency of SCAN and only takes slightly less time at lower hours. From Figure 4.2c it can be concluded that when the parameter n increases in a multiplicative form, SCAN elapsed time also shows a non-linear increase, but the rate of increase does not increase in a multiplicative form. From Figure 4.2d it can be concluded that there is an overall trend of non-linear increase, with a decreasing trend in elapsed time and a significant decrease in efficiency as parameter d grows from 16 to 32. It can be seen that the periodic nature of the GRUp module for acquiring data needs to be built on top of the GRUs module in order to help the SCAN algorithm achieve better results. Its effectiveness results are shown in Figure 4.3.

From Figure 4.3, it can be concluded that the SCAN-GRUs algorithm outperformed the SCAN algorithm in both evaluation metrics in the overall 35-day dataset, instead the GRUp module made the overall model less effective. Therefore, the updated nodal vector representation using the GRUp module is not applicable to everyday student behaviour. Selective use of the GRUp module is performed in special scenarios. In SCAN, the hyperparameters that have an impact on task relevance and the period that affects the frequency with which the GRUp module updates the hidden layers of the GRUs module. Sensitivity analysis of the two parameters was performed on 7, 21-and 35-day data sets. A demonstration of the sensitivity validation results for the h and *period* parameters is shown in Figure 4.4.

From Figure 4.4, it can be concluded that SCAN achieves better results for the main evaluation metric AUC on both the 21- and 35-day datasets when *period* is 7. This shows that the GRUp module updates the frequency of the hidden layer with the periodic features present in the data, and achieves the best results when the fit is right. In addition, in the data set with longer time span, the experimental effect obtained at 7 is more advantageous. However, when the time span is short or even less than one cycle, it cannot achieve great advantages, and even the effect is completely inferior to the experimental effect obtained by updating the hidden layer and vector characterization with high frequency. In this regard, we will further optimize our algorithm model to make it better adapted to the prediction task.
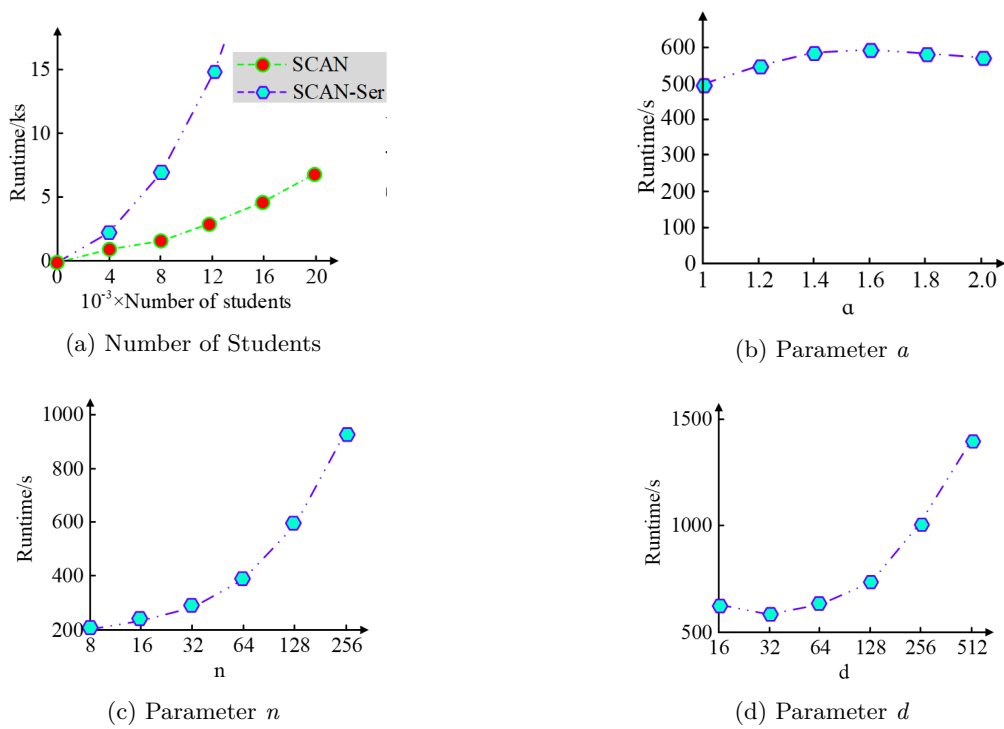
(a) Number of Students

(b) Parameter a

(c) Parameter n

(d) Parameter d

Fig. 4.2: Impact of Extensive Testing and Parameters on SCAN Efficiency
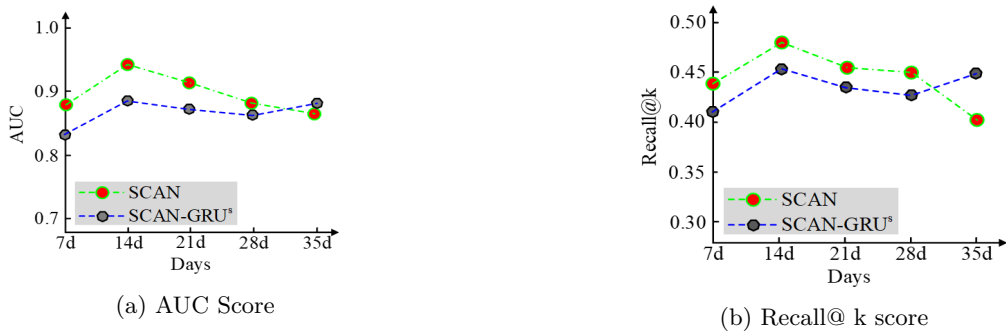

(a) AUC Score

(b) Recall@ k score

Fig. 4.3: Effectiveness of GRUp module in SCAN

5. Conclusion. Predicting student behaviour on campus has a positive effect on controlling footfall in campus areas, course check-in status, and resource allocation on campus. The problem of predicting student behaviour is transformed into a super-edge prediction problem in dynamic campus behavioural information networks by modelling campus behavioural data. This study proposes the SCAN algorithm to address the shortcomings of the temporal and periodic features in the acquired data. the SCAN algorithm has good scalability and can still support the prediction of campus student behaviour when the number of students is 20,000. In addition, the AUC scores and Recall@k scores of the SCAN algorithm were analyzed separately, and it was concluded that the SCAN algorithm obtained better results on the second cycle of the dataset. For the
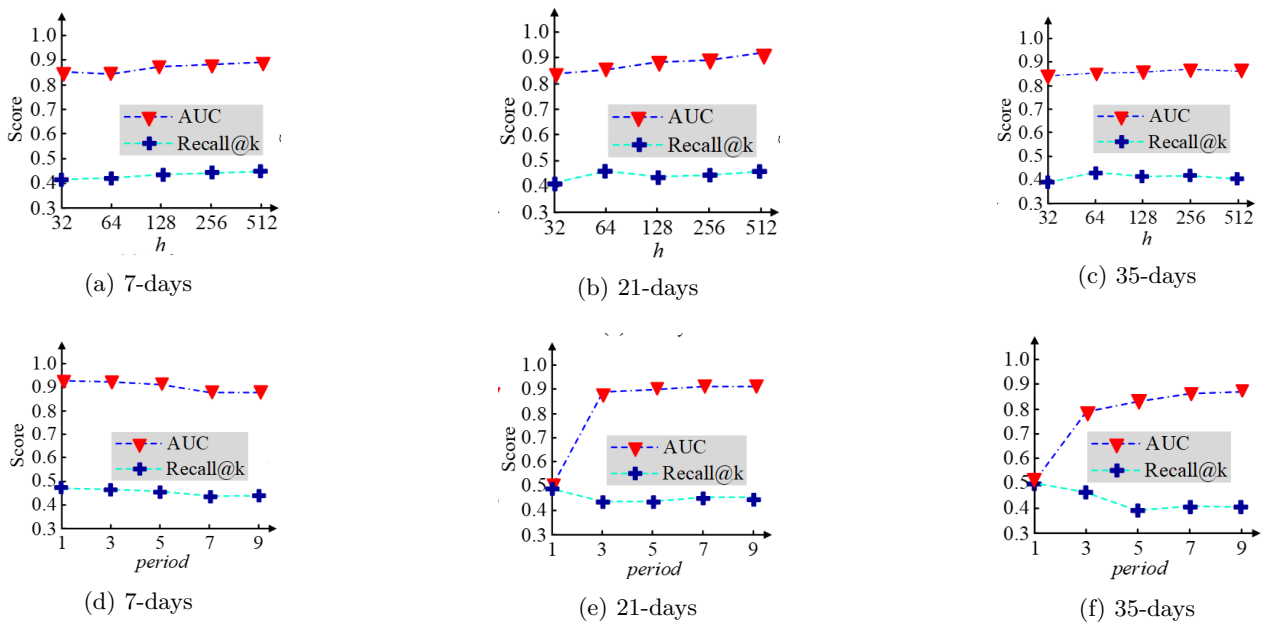
(a) 7-days

(b) 21-days

(c) 35-days

(d) 7-days

(e) 21-days

(f) 35-days

Fig. 4.4: Sensitivity analysis of SCAN parameters

Table 5.1: Abbreviated table

HIP	Hypergraph Information Propagation
SCAN	Student Campus Behavior Prediction
GRUs	Gated Recurrent Unit-Snapshot
GRUp	Gated Recurrent Unit-Period
INT	Hyperlink Interaction Scoring
SCLSM	Similar Campus Lifestyle Student Miner

dynamic and periodic features present in the campus behaviour data, the GRUs module was used to obtain the temporal features of each network snapshot in the dynamic campus behaviour information network, as well as the GRUp module to update the vector representations of the hidden layers and nodes in cycles as time phases. The results indicated that the updated node vector representations using the GRUp module are not applicable to everyday student behaviour. Selective use of the GRUp module is carried out in special scenarios. the periodic character of the data acquired by the GRUp module needs to be built on top of the GRUs module in order to help the SCAN algorithm to obtain better results. And with an parameter of 7, SCAN achieved better results for both the 21-day and 35-day datasets for the main evaluation indicator AUC. It can be seen that the GRUp module updates the frequency of the hidden layer with the periodic features present in the data, and is able to achieve the best results when it fits. However, the existing research in the face of similar behaviour search problems, and behaviour prediction problems still cannot be used in everyday life, but only in special situations. Therefore, further research is still needed to face the student behaviour prediction problem.

REFERENCES

[1] Tominaga, M., Shinkawa, S., Hyodo, T. & Yamaguchi, K. Dynamic behavior of macrocycle-based organic frameworks in single-crystal to single-crystal guest exchanges. *CrystEng Comm.* **23**, 7039-7043 (2021)

[2] Su, H., Yuan, X., Tian, R., Rong, L. & Yan, H. Study on the Wireless Body Area Network MAC Protocol Using Dynamic Super frame for e-Health IoT. *Journal Of Physics: Conference Series.* **1948**, 14-16 (2021)

[3] Jeslet, D., Komarasamy, D. & Hermina, J. Student Result Prediction in Covid-19 Lockdown using Machine Learning Techniques. *Journal Of Physics Conference Series.* **1911**, 8-12 (2021)

[4] Nsugbe, E. Toward a Self-Supervised Architecture for Semen Quality Prediction Using Environmental and Lifestyle Factors Artificial Intelligence and Applications. 2023. (0)

[5] Masood, F., Masood, J. & Zahir, H. Novel approach to evaluate classification algorithms and feature selection filter algorithms using medical data. *Journal Of Computational And Cognitive Engineering.* **2**, 57-67 (2023)

[6] Wang, D., Zeng, Q., Chawla, N. & Yinjuan, S. Modeling Complementarity in Behavior Data with Multi-Type Itemset Embedding. *ACM Transactions On Intelligent Systems And Technology.* **12**, 2-25 (2021)

[7] Yang, X., Zhou, Z. & Xiao, Y. Research on Students' Adaptive Learning System Based on Deep Learning Model. *Scientific Programming.* **13**, 2-13 (2021)

[8] Zhang, M., Du, X., Rice, K., Hung, J. & Li, H. Revealing at-risk learning patterns and corresponding self-regulated strategies via LSTM encoder and time-series clustering. *Information Discovery And Delivery.* **50**, 206-216 (2022)

[9] Ji, R., Liang, Y. & Xu, L. Recursive Bayesian inference and learning for target tracking with unknown maneuvers. *International Journal Of Adaptive Control And Signal Processing.* **36**, 1032-1044 (2022)

[10] Joshi, A., Saggar, P., Jain, R., Sharma, M., Gupta, D. & Khanna, A. CatBoost - An Ensemble Machine Learning Model for Prediction and Classification of Student Academic Performance. *Advances In Data Science And Adaptive Analysis: Theory And Applications.* **13**, 2-28 (2021)

[11] Zhang, Z., Huang, J., Tan, Q., Sun, H. & Zhou Y. CMG2Vec:, A. composite meta-graph based heterogeneous information network embedding approach. *Knowledge-Based Systems.* **216**, 2-14 (2021)

[12] Gardoni, E., Benito, S., Scansani, S., Brezina, S., Fritsch, S. & Rauhut, D. Biological Deacidification Strategies for White Wines. *South African Journal For Enology And Viticulture.* **42**, 114-122 (2021)

[13] Huang, Z., Kurotori, T., Pini, R., Benson, S. & Zahasky, C. Three-Dimensional Permeability Inversion Using Convolutional Neural Networks and Positron Emission Tomography. *Water Resources Research.* **58**, 2-21 (2022)

[14] Prigioniero, A., Zuzolo, D., Niinemets, U., Izzo, F., Postiglione, A., Sciarrillo, R. & And, G. and polycyclic aromatic hydrocarbon uptake in relation to leaf surface functional traits in Mediterranean evergreens: Potentials for air phytoremediation. *Journal Of Hazardous Materials.* **435**, 2-17 (2022)

[15] Cohen, Y., Grinshpun, A., Meir, K. & Pet/ct, G. uptake reveals a single metastasis inside vertebral hemangioma, catch me if you scan. *European Journal Of Nuclear Medicine And Molecular Imaging.* **48**, 2066-2067 (2021)

[16] Khan, A., Shaikh, A., Cheikhrouhou, O., Laghari, A., Rashid, M. & Shafiq, M. et al. *MG-forensics: Multimedia-enabled Information Hiding Investigation Using Convolutional Neural Network.* **16**, 2854-2862 (2022)

[17] Qin, X., Ma, S. & Wu, M. Gene-gene interaction analysis incorporating network information via a structured Bayesian approach. *Statistics In Medicine.* **40**, 6619-6633 (2021)

[18] Meng, D., Sun, L. & Tian, G. Dynamic mechanism design on social networks. *Games And Economic Behavior.* **131**, 84-120 (2022)

[19] Moura, B. Towards Precision Medicine in Psychosis: Benefits and Challenges of Multimodal Multicenter Studies-PSYSCAN: Translating Neuroimaging Findings From Research into Clinical Practice. *Schizophrenia Bulletin.* **48**, 722-723 (2022)

[20] Loone, C. Multispectral Line Scan Camera. *Vision Systems Design.* **27**, 2-46 (2022)

[21] Zhang, X., Lin, H., Xu, B., Li, C., Lin, Y., Lin, Y. & Ma, F. Dynamic intent-aware iterative denoising network for session-based recommendation. *Information Processing & Management: Libraries And Information Retrieval Systems And Communication Networks: An International Journal.* **59**, 10-14 (2022)

[22] Jaraut, P., Abdelhafiz, A., Chenini, H., Hu, X., Helaoui, M., Rawat, M., Chen, W., Boulejfen, N. & Ghannouchi, M. Augmented Convolutional Neural Network for Behavioral Modeling and Digital Predistortion of Concurrent Multiband Power Amplifiers. 2021. (0)

[23] Yan, B., Jian, L., Ren, R., Fulk, J., Sidnam-Mauch, E. & Monge, P. The Paradox of Interaction: Communication Network Centralization, Shared Task Experience, and the Wisdom of Crowds in Online Crowdsourcing Communities. *Communication Research.* **48**, 796-818 (2021)

[24] Ismail, S., Hamid, S., Ahmad, M., Alaboudi, A. & Jhanjhi, N. Exploring students engagement towards the learning management system (LMS) using learning analytics.. *Computer Systems Science & Engineering.* **37** (2021)

[25] Annadurai, C., Nelson, I., Devi, K., Manikandan, R., Jhanjhi, N., Masud, M. & Sheikh, A. Biometric Authentication-Based Intrusion Detection Using Artificial Intelligence Internet of Things in Smart City. *Energies.* **15**, 7430 (2022)

[26] Alex, S., Jhanjhi, N., Humayun, M., Ibrahim, A. & Abulfaraj, A. Deep LSTM Model for Diabetes Prediction with Class Balancing by SMOTE. *Electronics.* **11**, 2737 (2022)