International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-15

https://doi.org/10.70917/ijcisim-2026-0109
Article

Research on the Dynamic Prediction Model of Student Behavior in the Information Management System of Ideological and Political Education for Student Groups in Colleges and Universities

Danqiong Wang *

Information Office, University of Shanghai for Science and Technology, Shanghai, 200093, China; danqiongwang@163.com

Abstract: This study proposes a dynamic prediction model of student behavior based on multilayer fragment dynamic semantic spatio-temporal graph convolutional network (MF-STGCN) to meet the needs of information management of ideological and political education in colleges and universities. By integrating heterogeneous data from multiple sources, such as campus card consumption, WiFi track, course schedule, etc., a three-level data system of "activity-behavior-abnormal behavior" is constructed. First, activities are defined and clustered into five types of functional areas (teaching, dormitory, cafeteria, etc.). Secondly, we extracted the spatial and temporal characteristics, calculated the proportion of students' stay time in each area on a weekly basis, and formed the spatial and temporal vectors to characterize behavioral preferences. Finally, the MF-STGCN model is designed, which innovatively introduces a multi-fragment segmentation mechanism, combining dynamic semantic graph convolution, temporal convolution and fully connected layers to realize multi-scale feature fusion. The empirical study shows that the entropy value of WiFi usage in the teaching area is the lowest (0.18-0.41), reflecting the class schedule constraints. The dormitory has medium entropy value (0.47-0.72), which rises to 0.68 at the end of the period (schedule disruption). The cafeteria has the highest entropy value (0.65-0.82), which is consistent with fragmentation. MF-STGCN has an average relative error of only 4.23%-5.65% at the size of 100-10,000 people, which is 2.93% lower than K-nearest neighbor and GCN -4.53%, and the operational efficiency is improved by 5-9 times, with 10,000-person prediction taking 35.53 seconds. The model effectively improves the accuracy and timeliness of behavioral prediction, and promotes the transformation of civic education management from empirical decision-making to data-driven.

Keywords: civic education; information management; college students' behavioral characteristics; student behavior prediction; GCN

1. Introduction

The advent of the information age has transformed the way information is disseminated, driving industries across the board toward information-driven development. In the field of education, information-driven models have similarly brought new opportunities and challenges to ideological and political education in higher education institutions. Therefore, we can fully leverage the technological advantages of the information-driven model to innovate and advance the management model of ideological and political education, thereby enhancing the quality and standards of such education. By utilizing the massive data collection and dynamic data acquisition advantages of educational information management systems, we can achieve timely capture and multi-dimensional presentation of students' behavioral dynamics in daily ideological and political education for college students. Additionally, we can create a multi-dimensional, three-dimensional “holographic image” of individual students [1-4]. Based on this, educators can gain a more comprehensive, in-depth, precise, and rapid understanding of the behavioral characteristics and ideological traits of educational subjects [5]. Moreover, “everything is recorded, everything is analyzed” is the ultimate goal of educational information management systems. Many ambiguous issues in the daily ideological and political education of college students, such as students' ideological tendencies, value concepts, and cognitive abilities, can be measured, statistically analyzed, and visually presented using big data technology [6-7]. By utilizing data analysis technology to statistically analyze, intelligently process, and visually present data related to educational subjects and educational processes, educators can identify the underlying factors and relationships driving changes in college students' ideological and behavioral dynamics, as well as explore the patterns of daily ideological and political education [8-10].

Information technology provides technical support for educators to comprehensively and deeply understand educational subjects and the patterns of daily ideological and political education. Zhou, W., and Yang, T. indicate that modern information technology provides effective support for ideological and political education management in higher education institutions, and that information-based educational management methods are more efficient than traditional management methods [11]. Qiu, J. utilized data mining technology to conduct an in-depth analysis of students' ideological and political education performance, pointing out that educational management activities based on this analysis have certain guiding significance and theoretical foundation [12]. Han, W. elaborated on the application pathways of information technology in ideological and political education in higher education institutions, aiming to improve the efficiency of student management work and thereby promote the high-quality development of ideological and political education [13]. Huang, L. used teacher and student assignment assessment scales as the data foundation and applied clustering methods to analyze the effectiveness of ideological and political education in higher education institutions, the clustering results fully reflect the current state of ideological and political education management in higher education institutions, providing valuable insights for improving management standards [14]. It is evident that educational information management systems lay the foundation for the implementation and optimization of all aspects of daily ideological and political education for college students, while analyzing and predicting student behavior dynamics facilitates scientific decision-making and personalized education.

The article proposes a set of systematic methodological framework for dynamic prediction of students' behavior, which provides data support and decision-making basis for precise ideological governance, and becomes the key to enhance the effectiveness of the management of ideological education informationization. Firstly, behavioral elements (activities, behaviors, abnormal behaviors) are precisely defined to build the management and analysis basis of multi-source heterogeneous campus activity data. Secondly, digging deeply into behavioral patterns and extracting features reflecting students' interests and patterns from the spatial and temporal dimensions. Focus on extracting key features reflecting students' behavioral patterns from raw activity data. Functional categorization of locations to unify semantic information. Processing temporal features in terms of days and weeks, the core is to calculate the percentage of time students spend in different types of locations. And the spatio-temporal feature vector representing individual student's behavioral patterns is constructed, which calculates the distribution of students' average daily stay time in various types of functional areas, effectively capturing their behavioral preferences and regularities. Finally, a prediction model based on multilayer fragmented dynamic semantic spatio-temporal graph convolutional network (MF-STGCN) is proposed. The model structure contains a dynamic semantic graph convolutional layer (extracting behavioral semantic features, e.g., inferring study habits using class schedules), a temporal convolutional layer (capturing temporal features, e.g., inferring rest and relaxation patterns using WIFI trajectories), and a fully connected layer. The core improvement lies in the introduction of a multi-fragment segmentation and fusion mechanism, which extracts the features of three key cycles: day, week, and month, respectively, and then fuses them through dynamic graph structure learning and spectral space feature capturing. Finally, the complete prediction process is demonstrated, which forms a closed-loop prediction management system from multi-source data collection and cleaning, activity clustering to behavioral modeling, and then to multi-fragment feature extraction and prediction output of MF-STGCN.

2. Construction of a Dynamic Prediction Model for Student Behavior Based on Multi-Source Campus Data fusion

2.1. Definition of Mathematical Relationships Related to Campus Data

The construction of digital campus aggregates a huge amount of daily activity data of college students on campus, and these multi-source heterogeneous data provide a solid foundation for data-driven college students' behavior mining research based on data, which gets rid of the traditional drawbacks of simply taking the static data, such as students' information, as the object of research. In this paper, we manage the collected data by constructing a database containing activity objects, time, space and activity attributes such as student number, time, location, and specific content description. Let the set of all students in the data be U , the set of all locations be L , and the set of specific descriptions of all activities be D , in this paper, we regard a data record as the smallest data unit, defined as an activity a_i .

2.1.1 Definition of Activities

An activity is a specific action of a student under the influence of a specific behavioral condition, i.e., any human behavior consists of a set of related activities. Activities contain attributes such as student number, time, place and a specific description of the activity, an activity a_i is formally defined as follows:

$$a_i = (u, t, l, d) \tag{1}$$

$$A = \{a_1, a_2, ..., a_n\} \tag{2}$$

where $u \in U$ denotes the performer of the activity, t denotes the time of generating the activity, $l \in L$ denotes the location where the activity is generated, $d \in D$ denotes the specific description of the activity, and A denotes the set of all activities in the data, where l represents different zones in the spatial location of the campus.In this paper, the activity zones of the campus of the university students are divided into six categories according to the functional attributes of the buildings, namely Teaching and Research Area, Library Area, Dormitory Area, Food and Consumption Area, Sports Area and Other Area.

2.1.2 Behavioral Definitions

Behavior is a collection of externally manifested activities expressed under the conditions dictated by a particular idea, the sum of activities united by a common purpose and accomplishing a certain function, and a specific behavior consists of a series of activities with the same elements. The formal expression of a behavior consists of the set of activity elements that constitute the behavior, the relationship between different activities in the set of activities, and the specific semantic expression of the behavior. The formal definition of the behavior b_j is as follows:

$$b_j = \{A_{b_j}, R, S\} \tag{3}$$

$$A_{b_j} = Seq1(u, b_j) = (a_1, a_2, ..., a_k) \tag{4}$$

$$B = \{b_1, b_2, ..., b_m\} \tag{5}$$

where A_{b_j} denotes the sequence of activities associated with behavior b_j as a $k \times 3$ matrix. Activities with the same location position or the same description constitute a sequence of activities for a behavior, i.e., $Seq(u, b_j)$, which denotes that the j th behavior b_j of a student u is composed of a sequence of activities a_1 through a_k . R denotes the relationship between all the activity elements that make up the behavior, a $k \times k$ matrix generated by the algorithm when categorizing the activities into behaviors, which can either denote the temporal order relationship between the activities or the semantic descriptive correlation relationship between the activities. S denotes the semantic features of the behavior, different types of behaviors have different semantic expressions, the algorithm generates a semantic description of a behavior while categorizing the activity into a specific behavior, such as the behavior of attending classes, borrowing and reading, eating and drinking. B denotes the set of all behaviors in the student's campus life, usually a series of related but different activities constitute a common behavior.

2.1.3 Definition of Abnormal Behavior

Abnormal behavior is defined as behavior that is irregular within a certain time frame, i.e., individual behavior that is different from the group or does not conform to a certain code of conduct. Abnormal behavior as a label for data is the result of judging the data based on a set of norms. In this paper the label e is utilized to represent whether the behavior is abnormal or not, and the resulting formal definition of behavioral data with labels is as follows:

$$b_j = \{A_b, R, S, e\} \tag{6}$$

where $e \in \{0,1\}$ indicates the label of the behavioral data, when $e=1$, it represents that the behavioral record is an abnormal behavior, when $e=0$, it represents that the behavioral record is a normal behavior. The criteria for judging this abnormal behavior are labeled strictly according to the students' schedule of classes, the school's prescribed work and rest time, and other aspects of the activity schedule, using the fusion correlation of the data to further expand the feature dimensions of the data.

2.2. Spatio-Temporal Feature Extraction of Students' Behavioral Trajectories

Student behavior spatio-temporal feature extraction is an important step in this study, specifically, it is divided into two steps: spatial feature processing and temporal feature processing.

The first is spatial feature processing. That is, locations with the same or similar semantic information and functions are grouped into one category of locations, and their semantics usually represent the same interest information. Campus locations can be divided into five categories according to their functions: rest, study, food, sports and consumption. Among them, resting locations include male and female dormitories. Learning locations include teaching buildings, research buildings, etc. Dietary locations include student cafeterias and staff cafeterias. Sports venues include playgrounds, basketball courts and gymnasiums. Consumption category includes the school supermarket and so on.

Then for temporal feature processing. The trajectory data sequence of each user is arranged by date and feature extraction is performed in days and weeks, respectively. We denote the total number of weeks contained in a year as N.

From this we define the distribution matrix of the residence time of a student u at different types of locations in the n th week of a year as shown in equation (7).

$$X^{(n)} = \begin{bmatrix} Y_{11}^{(n)} & \cdots & Y_{1j}^{(n)} \\ \vdots & \ddots & \vdots \\ Y_{i1}^{(n)} & \cdots & Y_{ij}^{(n)} \end{bmatrix} \quad \begin{cases} 0 < i \le I, I = 7 \\ 0 < j \le J, J = 4 \\ 0 < n \le N \end{cases} \tag{7}$$

where, $\gamma_{ij}{}^{(n)}$ denotes the percentage of the user's individual u 's stay at location j on day i in week n as a percentage of this day, as shown in equation (8). Where, $t_{ij}{}^{(n)}$ denotes the user's stay at location j on day i in week n.

$$Y_{ij}^{(n)} = \frac{t_{ij}^{(n)}}{\sum_{0<j\le J} t_{ij}^{(n)}} \tag{8}$$

Define P_u as a vector of spatio-temporal characteristics of user u, $P_u = [P_{u1}, P_{u2}, \cdots]$. where P_{uj} denotes the average weekly percentage of time spent at location j in a total of N weeks in a year by an individual user u on that day, as shown in equation (9).

$$P_{uj} = \frac{\sum_{0<n\le N} \frac{\sum_{0<i\le I} Y_{ij}^{(n)}}{I}}{N} \tag{9}$$

Thus, the feature vector P_u representing the behavioral pattern of the student u is obtained, which calculates the percentage of time that the student spends in different types of locations on campus on an average day, and contains information about the interests embodied in the student's behavior.

2.3. Dynamic Prediction Model of College Student Behavior Based on Improved Dynamic GCN

Through the above spatio-temporal feature extraction process, we have successfully transformed the raw student activity data into feature vectors that can reflect their behavioral patterns and interest preferences. In order to utilize these features for dynamic prediction of future behaviors and effectively fuse dynamic semantic information and multi-scale temporal dependencies in behaviors, there is an urgent need to design a powerful prediction model. To this end, this study proposes an MF-STGCN model based on an improved dynamic graph convolutional network.

2.3.1. MF-STGCN Network Structure

Behavioral prediction refers to the prediction of future behavior by extracting the spatio-temporal and cyclical features of the raw data, which are analyzed in a specific way and the conclusions are used. Assuming that behavioral prediction is performed for student u, the equation expression for this process can be shown in equation (10).

$$\begin{cases} b_{u,\tau} = f(B_u) \\ B_u = Seq2(u, b_{u,\tau}) = (b_1, b_2, b_3, \cdots b_n) \end{cases} \tag{10}$$

In Eq. (10), $b_{u,\tau}$ denotes the behavior of student u at moment t; B_u denotes the sequence of activities associated with the behavior; and $f(\ )$ denotes the filtering function.

Combining this formula, the study proposes a dynamic semantic spatio-temporal graph convolutional neural network (STGCN) that combines multilayer segments, i.e., MF-STGCN.Relative to STGCN, the network is able to segment the data into multiple segments, which can then be fused for global and local feature capture.The structure of MF-STGCN is shown in Fig. 1.

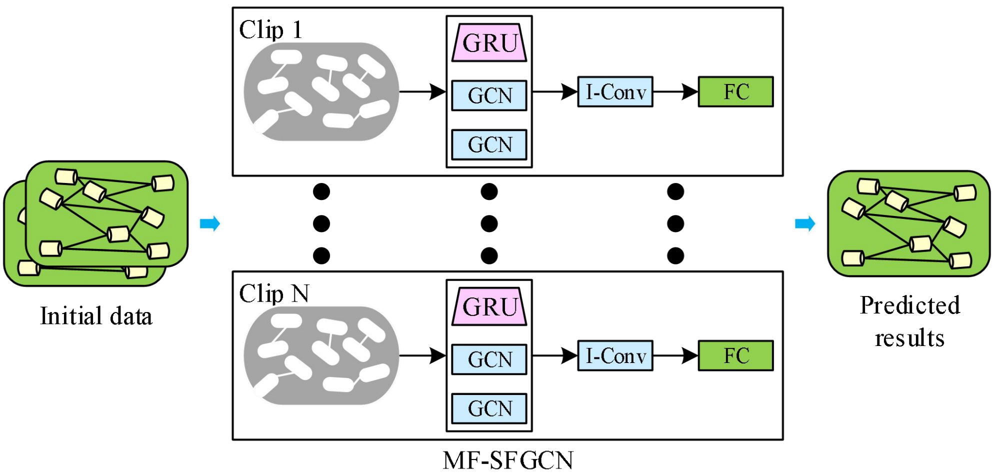

Figure 1. The structure of the MF-STGCN network.

As can be seen from Fig. 1, the whole network structure consists of multiple sets of segments, and each set of segments consists of a dynamic semantic graph convolutional layer, a temporal convolutional layer, and an all-connected layer. The dynamic semantic features of the data are firstly extracted using the dynamic semantic graph convolutional layer, and the meaningful semantic features of the behavioral data are extracted through the natural language processing technology, so as to reduce the interference of irrelevant factors on the prediction model. By analyzing students' schedule and attendance records, students' study habits and attendance patterns can be deduced. Secondly, capturing the adjacent time features of the fragment data and realizing the temporal characterization of the behavioral data through the temporal convolution layer, fully considering the temporal regularity of the behavior.

Then, by analyzing students' WIFI usage trajectories in different time periods, we can infer their work and rest time and campus activity patterns. Finally, by fusing multi-fragment periodic features, the capture of global and local features is achieved through the all-connected layer, which further improves the robustness and generalization ability of the model. For example, by fusing students' behavioral features in different time cycles of day, week, and month, their future behavioral trends can be predicted more accurately. Therefore, the study conducts behavioral prediction analysis with three cycle segments: daily, weekly, and monthly. Among them, the formula for the daily cycle segment is shown in equation (11).

$$X_d = \left(X_{t_p - t_d + 1}, X_{t_p - t_d + 2}, \cdots, X_{t_p - t_d + n} \right) \tag{11}$$

In Equation (11), t_p denotes the current time period; t_d denotes the daily cycle time period; and X_d denotes the characteristics of the first d days of the predicted time period. The formula for the weekly cycle segment is shown in Eq. (12).

$$X_w = \left(X_{t_p - t_w + 1}, X_{t_p - t_w + 2}, \cdots, X_{t_p - t_w + n} \right) \tag{12}$$

In Eq. (12), t_d denotes the specific number of days of collection; X_w denotes the characteristics of the first w weeks of the predicted time period. The formula for the monthly cycle segment is shown in Equation (13).

$$X_m = \left(X_{t_p - t_m + 1}, X_{t_p - t_m + 2}, \cdots, X_{t_p - t_m + n} \right) \tag{13}$$

In Eq. (13), t_m denotes the number of weeks of the specific collection; X_m denotes the features in the m months before the predicted time period. Combining the above 3 types of periodic features, the study is carried out to learn the dynamic graph structure with GCN units, which is completed to capture the spatial features in a spectral way. The expression of STGCN in this process is shown in equation (14).

$$X_t^l = F\left(X_t^{l-1}, A_t, W_t^l \right) = \sigma\left(D_t^{-\frac{1}{2}} (D_t - A_t) D_t^{-\frac{1}{2}} X_t^{l-1} W_t^l \right) \tag{14}$$

In Eq. (14), D_t denotes the diagonal matrix consisting of nodes; A_t denotes the campus spatial road network; X_t^{l-1} denotes the node features of the $l-1$ th layer at t time; W_t^l denotes the weight matrix of the convolution operation performed at the l th layer at t time; σ denotes the activation function; and F denotes the node's F dimensionality Eigenvalues.

2.3.2. Prediction of College Students' Daily Campus Behavior by Joint MF-STGCN

Combining the above formulas, the study proposes a predictive management model for college students' daily campus behavior using the MF-STGCN algorithm. The flow of the model is shown in Figure 2.

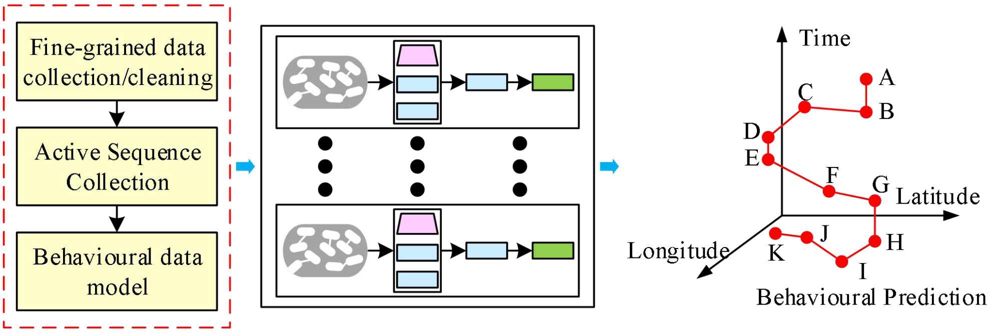

Figure 2. The student behavior prediction process combined with MF-STGCN.

As can be seen in Figure 2, first, student data collection and cleaning are performed through the spatio-temporal fine-grained standard, which integrates multiple data sources such as campus card consumption records, library borrowing records, cafeteria dining records, campus WIFI usage trajectories, course schedules and attendance records, grade information, online platform data, and monitoring system data, to reduce the bias and error that may be caused by a single data source. Then, the cleaned data are clustered in the form of activity sequences to constitute a complete behavioral data model. The study segments the historical student behavior data into daily, weekly and monthly segments through multi-layer segment feature extraction to capture the temporal patterns of student behavior and reduce the impact of behavioral fluctuations on the prediction results in the short term.The MF-STGCN model contains dynamic semantic graph convolutional layer, temporal convolutional layer and fully connected layer, which are responsible for extracting the dynamic semantic features, capturing the temporal features, and fusing the global and local features, respectively, to improve the robustness and generalization ability of the model. Through the methods of multi-source data fusion, fine-grained data processing, and multi-segment feature extraction, the study achieves significant results in avoiding the influence of many factors on students' daily behaviors, and meets the prediction of future behaviors.

3. Empirical Evidence and Behavioral Law Mining of MF-STGCN Model Based on Civic Education

Based on the MF-STGCN prediction model framework constructed in Chapter 2, this chapter will launch an empirical study relying on real campus multi-source data. The correlation between students' Internet use behavior and academic performance under the information management system of ideological and political education is deeply explored. Eventually, combined with the results of behavioral spatio-temporal feature extraction, the effectiveness of MF-STGCN in behavioral law mining and dynamic prediction will be verified.

3.1. Data Sources

The original data includes a total of 794 undergraduates of the 2024 class of the university from September 2024 to January 2025, a total of 327345 records, as well as student achievement data, physical exercise record data, and campus wireless network access record data. After integrating the data, the data was converted and imported into the distributed cluster HDFS by using the Sqoop tool, and the data was preprocessed based on Spark such as data cleaning and specification, and then statistical analysis was carried out to establish the characteristics of students' network use behaviors and five types of behavior trajectories based on the information management of ideological and political education.

3.2. Correlation Analysis of Students' Internet Use Behavior and Academic Achievement

Using the above data to summarize and process the students' WIFI connection records and input them into the classifier, this chapter mainly consists of 2 parts, the first part is the validation of the reasonableness of the samples, the raw data are processed and refined to obtain a number of dimensional features as a preparation for the subsequent training of the model, the features with a high contribution need to be subjected to a certain amount of statistical analysis before modeling, and the visualization is a better visualization is a better way to ensure that the collected samples are balanced and unbiased. The second section presents the results of the experiments and the conclusions drawn from them and the analysis of student behavior.

3.2.1. Sample Reasonableness Verification

After serializing the raw data, the entropy value mean value of each day in the experiment is counted in the unit of days for students in different score bands in the unit of one week. The distribution of students' scores and entropy value from Monday to Sunday in the information management system of ideological and political education is shown in Figure 3, which shows that students in the high score band have more sample points than students in the low entropy value part, and the sample points are mainly converged in the upper left and lower right, which indicates that there exists a law that the students in the high score band use WIFI less frequently, and the students in the low score band use WIFI more frequently. A small number of outliers exist. Behavioral differences during the week are not too different, block out the information that it is the day of the week, only from the dimension of whether it is the weekend to observe, mid-week and weekend student achievement and ordered distribution as shown in Figure 4, it can be seen that most of the high scoring students weekend entropy value is higher than the entropy value of the week, the low-scoring students do not have a significant difference, we believe that the entropy value of this feature for the achievement of a certain predictive ability. At the same time, it can be seen that the students' scores under the informationized management system of ideological and political education are mostly more than 70 points, and the entropy value of network use is smaller, which indicates that the ideological and political education has a good influence on the students' scores and the control of network use.

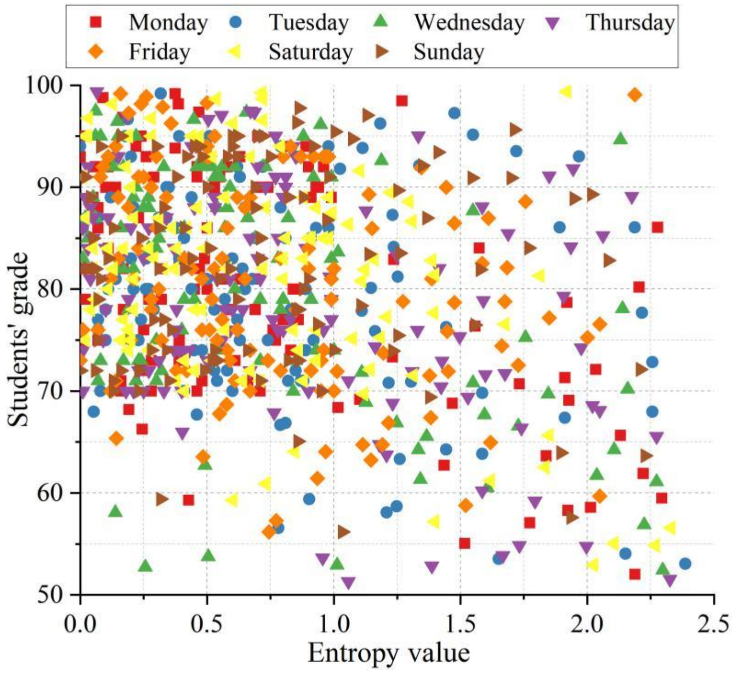

Figure 3. Distribution of students' grades and entropy values from Monday to Sunday.

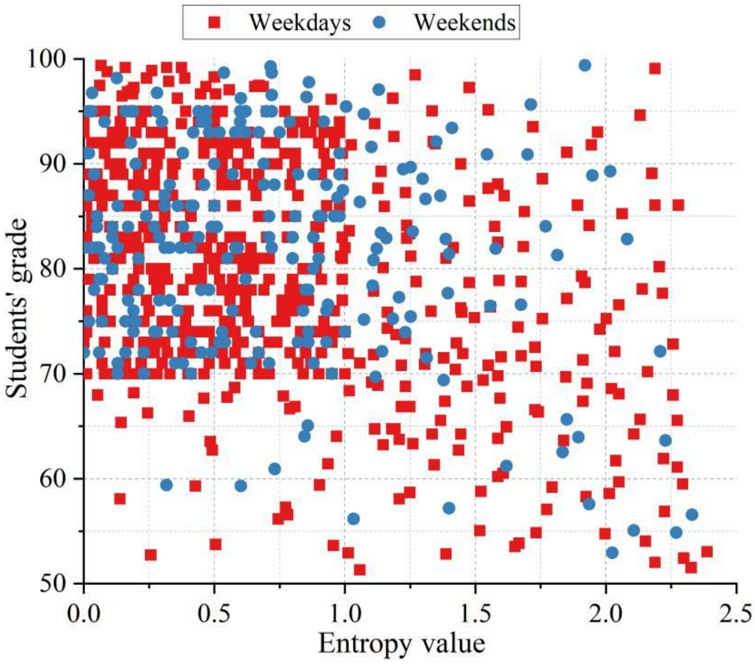

Figure 4. Distribution of grades and entropy values of weekdays and weekend.

3.2.2. Spatio-Temporal Distribution of Student Entropy Value in Different Scenarios

The average entropy value of students' WIFI usage in each week under the three scenarios of studying, resting, and eating (corresponding to the records of the three locations of the academic building, dormitory, and cafeteria, respectively) was counted for a total of 18 weeks from September 7, 2024 to January 11, 2025, and the distribution of weekly entropy averages under the three scenarios is shown in Figure 5.

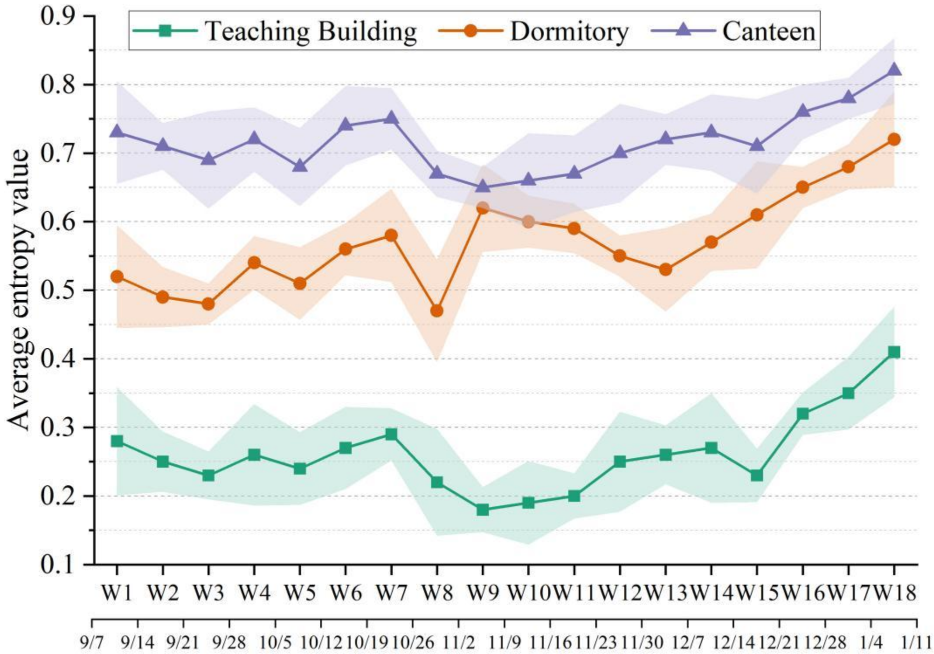

Figure 5. The weekly average entropy distribution in the three scenarios.

The closer the entropy value is to 0, the more regular the behavior is, and the closer it is to 1, the more random it is. From the entropy value distribution, the following conclusions can be intuitively obtained. the entropy value of the teaching building is the lowest among the 3 scenarios (0.18-0.41): constrained by the course schedule, the students' class time is fixed, and the behavior of WiFi connection is highly regular (especially on weekdays). Dormitory has medium entropy value (0.47-0.72): high frequency of use at night and on weekends but the time is scattered, and the regularity of behavior is weaker than that of the teaching area. Cafeteria has the highest entropy value (0.65-0.82): dining time is concentrated but the stay time is short, WiFi connection is random, which is consistent with the characteristics of fragmented behavior.

From the time dimension, the entropy value of the teaching area gradually decreases (0.28→0.23) at the beginning of the semester (Week1-3), reflecting students' gradual adaptation to the schedule pattern. The entropy value of the teaching area plummets to 0.18 during the midterm exam week (Week9-11) (regular revision before the exam), and the entropy value of the dormitory area rises to 0.62 (staying up late to prepare for the exam breaks the routine). Final week (Week16-18): the entropy value of the teaching area jumps to 0.35+ (end of the course, increase in behavioral randomness), and the entropy value of the dormitory area breaks through to 0.68 (disruption of work and rest, all-night revision/entertainment). In particular, the entropy values of all areas peaked in week 18 before the vacation (teaching 0.41/dormitory 0.72/cafeteria 0.82), reflecting the discrete behavior before leaving school.

3.3. Behavioral Characterization of College Students

After clarifying the association between online behaviors and grades, further quantification of the spatio-temporal patterns of campus behaviors is needed. Based on the extracted probability distributions of the five types of behaviors, this section reveals the structured characteristics of student behaviors on the day-cycle scale.

3.3.1. Characteristics of University Students' Campus Behavior Patterns

Continuing with the five types of behaviors (rest behavior, study behavior, eating behavior, exercise behavior and consumption behavior) in the above 794 college students' campus daily behavior data samples, the spatio-temporal feature extraction through the MF-STGCN model was used to obtain the probability distributions of the five types of behaviors of the students in a day under the information-based management system of ideological and political education, as shown in Table 1, and plotted as a line graph in order to show it more intuitively as shown in Figure 6.

Table 1. The probability distribution of four behaviors of students within a day.

Time	Rest behavior	Learning behavior	Dietary behavior	Motor behavior	Consumption behavior
7:00	41.83%	12.07%	28.95%	6.42%	10.73%
8:00	18.64%	45.22%	22.16%	5.31%	8.67%
9:00	5.29%	73.85%	7.62%	3.18%	10.06%
10:00	4.37%	76.91%	6.83%	2.75%	9.14%
11:00	6.12%	67.43%	14.28%	4.05%	8.12%
12:00	3.58%	9.74%	69.35%	1.97%	15.36%
13:00	24.16%	47.85%	11.29%	6.04%	10.66%
14:00	7.25%	72.63%	7.81%	3.42%	8.89%
15:00	6.38%	65.17%	8.93%	9.26%	10.26%
16:00	9.14%	54.32%	7.05%	17.83%	11.66%
17:00	8.47%	11.25%	60.84%	9.73%	9.71%
18:00	7.92%	6.38%	43.16%	34.25%	8.29%
19:00	11.63%	58.42%	10.57%	8.35%	11.03%
20:00	16.85%	59.37%	7.64%	5.92%	10.22%
21:00	38.26%	32.14%	7.83%	8.45%	13.32%
22:00	61.74%	14.53%	4.62%	5.18%	13.93%
23:00	77.35%	4.86%	3.79%	2.97%	11.03%

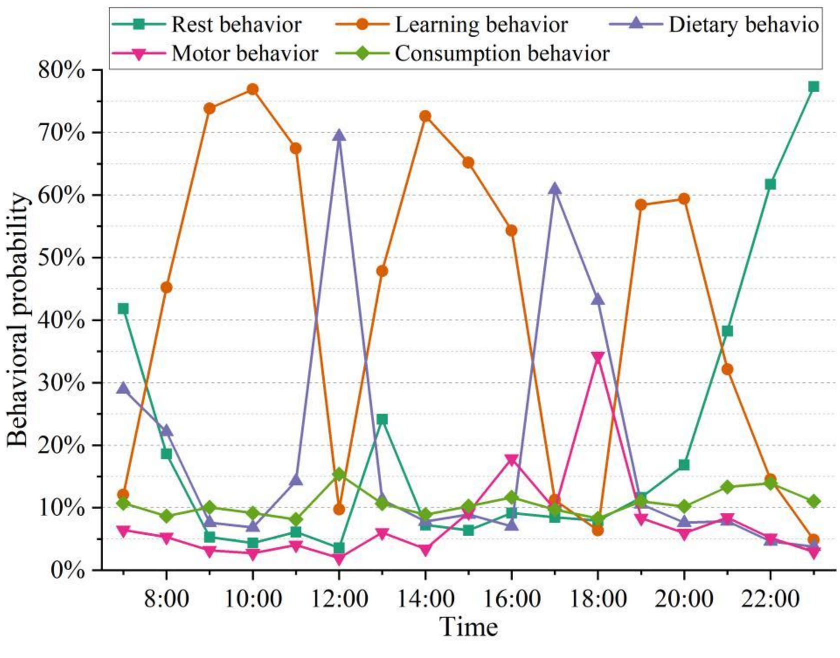

Figure 6. Five behavioral probability distributions.

Based on the data in Table 1, student behaviors showed significant time regularity and scene dependence. Study behavior concentrated time (9:00-16:00), with the peak percentage of study behavior from 9:00-11:00 a.m. reaching 73.85% (9:00) and 76.91% (10:00), reflecting the concentration during the course-intensive time. Afternoon 14:00-15:00 maintains a high level (72.63%→65.17%), reflecting the continuity of self-study or laboratory courses; while the bimodal characteristics of eating behavior, the breakfast period (7:00-8:00) accounted for 22.16%-28.95%, the lunch (12:00) plummeted to 69.35%, and the dinner (17:00-18:00) amounted to 60.84%→ 43.16%, in line with the law of three meal times; exercise behavior is active in the evening, the probability of exercise jumps to 17.83% at 16:00 (free activities after school), and reaches a full-day peak of 34.25% at 18:00 (centralized exercise period); rest behavior is dominant at night, with the proportion of rest behavior at 22:00-23:00 accounting for 61.74%→77.35%, which is in line with the lights-out work and rest period and late-night studying behavior drops to 4.86%; consumption behavior drops to 4.86%; and consumption behavior drops to 4.86%; and consumption behavior drops to 4.86%. Behavior drops to 4.86%; Consumption behavior is scattered and distributed, with small peaks in dining time (15.36% of consumption behavior at 12:00) and evening leisure time (13.32% at 21:00), reflecting the characteristics of fragmented consumption.

The data verifies the strong temporal regularity of campus behaviors and highly matches with the teaching schedule and life rhythm.The MF-STGCN model provides a quantifiable dynamic baseline for abnormal behavioral alerts by capturing multi-scale cyclical features.

3.3.2. Characteristics of University Students' Campus Behavior Patterns

Schools, food, sports and consumer places were selected to study the distribution of the time interval between card swipes at the places where these four types of behaviors occur, and Table 2 and Figure 7 show the distribution of the time interval between card swipes at the places where each type of behavior occurs under the information management system of ideological and political education, respectively.

Table 2. The distribution of card swiping time intervals of various behaviors occur.

Time interval	Learning behavior	Dietary behavior	Motor behavior	Consumption behavior
0-1min	2.17%	38.45%	5.32%	51.26%
1-5min	3.84%	41.73%	8.15%	32.18%
5-30min	12.56%	16.28%	24.37%	12.59%
0.5-1h	28.39%	3.62%	35.84%	3.71%
1-2h	32.15%	0.14%	21.93%	0.22%
2-3h	15.73%	0.05%	4.29%	0.04%
More than 3h	5.16%	0.03%	0.1%	0

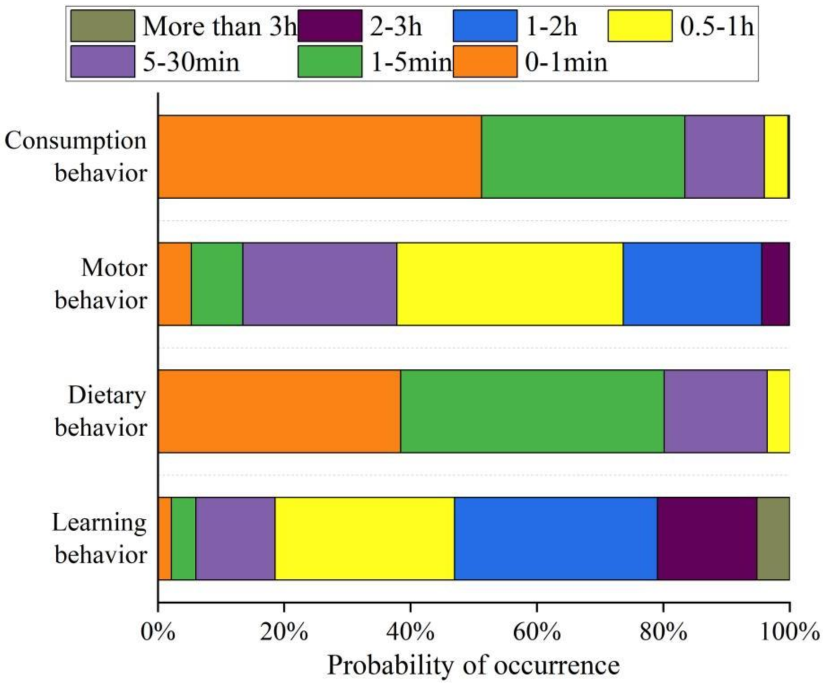

Figure 7. The distribution of card swiping time intervals for various behaviors.

Based on the data in Table 2, the four types of campus behaviors show significant differences in the distribution of time intervals, reflecting the behavioral characteristics of different scenarios. The long-time characteristics of learning behaviors, concentrated in periods of more than 30 minutes: 30 minutes-1 hour (28.39%), 1-2 hours (32.15%) and 2-3 hours (15.73%) accumulated 76.27%, in line with the continuity needs of the classroom/study scenarios, and reflecting the time-sinking characteristics of in-depth learning activities; and fragmented characteristics of eating and drinking behaviors, dominated by ultra-short time: 0-1 minutes (38.45%) and 1-5 minutes (41.73%) accumulated 80.18%, reflecting the fast dining mode (e.g., code-sweeping to pick up food), which is highly consistent with the queuing scene during the peak period in the cafeteria. Time period concentration of exercise behavior, bimodal distribution: 5-30 minutes (24.37%, warm-up/short exercise) and 30 minutes-1 hour (35.84%, regular exercise) while 1-2 hours accounted for 21.93% (team games/specialized training), reflecting the hierarchical differentiation of the intensity of the exercise; instantaneous characteristics of consumption behavior, 0-1 minutes accounted for 51.26% (code-sweeping payment dominant The instantaneous nature of consumption behavior, 51.26% for 0-1 minute (dominated by code payment) and 32.18% for 1-5 minutes (product selection + payment), reflects the “choose and go” characteristic of campus consumption.

3.4. Behavioral Prediction and Early Warning Analysis of College Students Based on Campus Big Data

Based on the above multi-source college campus big data, including students' network use behavior and five types of behavioral characteristics analysis, the article will use the Improved Dynamic GCN model based on the improved dynamic GCN model to carry out the dynamic prediction of college students' behaviors under the information management of ideological and political education. And the correctness of MF-STGCN model and traditional K nearest neighbor regression algorithm as well as the original GCN model to predict students' behavior is compared and analyzed by cross-validation method, and the model capability is reflected by the average relative error and standard error RMSE between the prediction results of students' behavioral characteristics and the real value. The larger the error, the lower the correctness, and the smaller the error, the higher the correctness, in order to verify the scalability of the model at the same time. This paper analyzes the correctness of the two based on the comparison of different sizes of students, and the test data sizes are 100, 500, 1,000, 2,000, 5,000, 8,000, and 10,000 students, respectively.

3.4.1. Relative Error Analysis

The average relative errors of the predictions of the three models are shown in Figure 8.

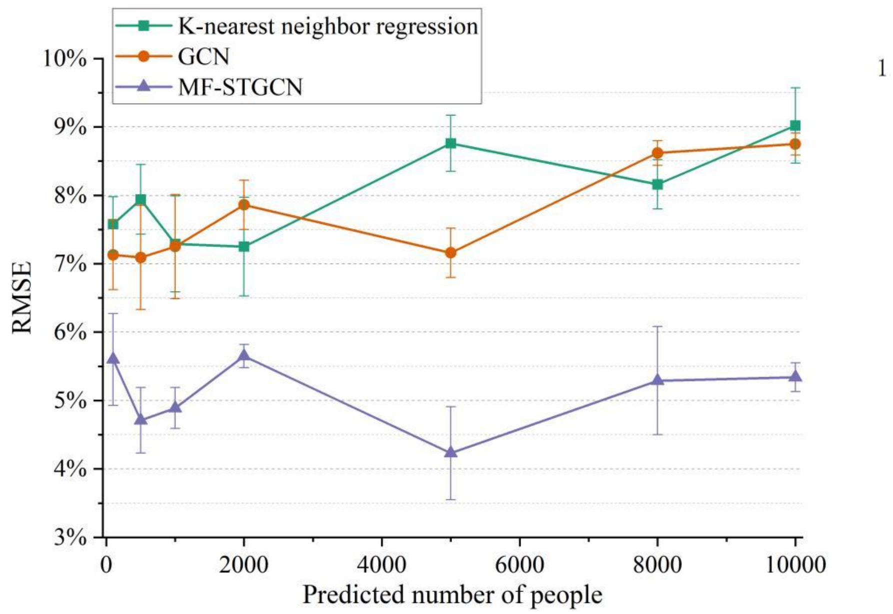

Figure 8. The predicted average relative errors of the three models(%).

MF-STGCN significantly outperforms the baseline model, consistently having the lowest average relative error (4.23%-5.65%) across all test sizes (100-10,000 people), especially at 5,000 people

(4.23%), which is 4.53% and 2.93% lower than K-nearest neighbor regression (8.76%) and GCN (7.16%), respectively.) and GCN (7.16%), respectively. k-nearest neighbor regression error rises with scale (100 people: 7.58% → 10,000 people: 9.02%), suggesting that it is difficult to adapt to large-data-volume scenarios. the GCN model is similarly fluctuating (7.09%-8.75%), with the highest error (7.86%) at the 2,000-person size the highest error (7.86%), which is not stable enough.The scalability advantage of MF-STGCN, even if the number of predicted people increases to 10,000, its error is only 5.34%, which is much lower than that of K-nearest neighbor (9.02%) and GCN (8.75%), which verifies the robustness of the model under large-scale data.

Regarding the average relative error of the MF-STGCN model in this paper for each characteristic indicator of student behavior (final grade, study hours, physical activity indicator, dietary pattern and consumption prediction) is shown in Table 3.

Table 3. The RMSE of characteristic f student behavior of MF-STGCN model(%).

Predicted number of people	Final grade	Study duration	Physical exercise indicators	Regular diet	Consumption forecast
100	5.36	4.38	3.77	3.37	3.83
500	5.64	5.14	3.74	4.06	4.87
1000	6.71	4.85	2.71	4.25	4.77
2000	6.37	5.60	3.59	4.17	4.23
5000	6.91	5.57	4.14	3.91	5.19
8000	6.03	4.18	3.94	4.77	4.02
10000	6.81	5.68	4.03	5.47	4.82

Among them, the "physical activity indicator" has the lowest error (2.71%-4.14%), especially at the size of 1,000 people, which is only 2.71%, reflecting the regularity of sports behavior and easy to capture. The error of "final grade" is the highest (5.36%-6.91%), which is difficult to predict dynamically because the grade is affected by many factors (e.g., temporary revision and examination status). Most of the indicators have the lowest error (e.g., 2.71% for physical exercise and 4.77% for consumption prediction) at the scale of 1,000 people, which may be the optimal scale for the model.

3.4.2. Comparison of Student Prediction Model Execution Times

In addition to analyzing the average relative error between student behavior prediction models reflecting model accuracy metrics, the runtime complexity of the models was analyzed next, and Figure 9 analyzes the runtime efficiencies of the different models in predicting different sizes of student volumes.

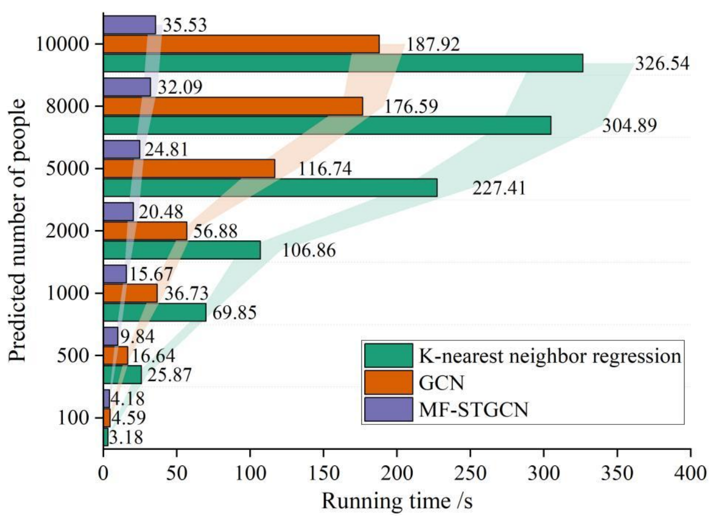

Figure 9. The running time efficiency of different models(s).

The efficiency advantage of MF-STGCN is also outstanding, in the scale of 500 people or more, the running time of MF-STGCN is significantly lower than that of other models, such as the model running time at 5000 people is 24.81s, while that of K-nearest neighbor is 227.41s, and that of GCN is 116.74s. Thanks to the design of MF-STGCN model's multi-fragment fusion and dynamic graph convolution, the model grows in the data volume when it still maintains a low time overhead, with an increase of only 31.35s for 100→10000 people. When the size increases to 10,000 people, MF-STGCN takes only 35.53 s, while K-nearest neighbor takes 326.54 s (9.2 times difference). K-nearest neighbor regression time grows linearly with the size (100 people: 3.18 s → 10,000 people: 326.54 s), which makes it difficult to be scaled up. -5 times (e.g., 8000 people: 176.59s vs 32.09s).

4. Conclusion

In this study, we constructed a three-level data system of “activity-behavior-abnormal behavior” based on multi-source campus data, and proposed a multi-layer fragment dynamic semantic spatio-temporal graph convolution network (MF-STGCN) to realize the dynamic and accurate prediction of student behavior in colleges and universities. The main conclusions are as follows

The regularity of student behavior is remarkable. The entropy value of teaching area is the lowest (0.18-0.41), which is strictly constrained by the course schedule; the entropy value of dormitory area is medium (0.47-0.72), which rises to 0.68 at the end of the semester; and the entropy value of cafeteria area is the highest (0.65-0.82), which is in line with the characteristics of fragmentation. Instructional entropy dropped from 0.28 to 0.23 at the beginning of the semester; weekly dormitory entropy broke 0.68 at the end of the semester.

Behavioral pattern quantitative characteristics. The percentage of five types of behaviors were peak study behavior: 9:00-10:00 up to 73.85%-76.91%; bimodal eating behavior: lunch (12:00) 69.35%, dinner (17:00) 60.84%; and dominant nighttime rest: 23:00 dormitory behavior accounted for 77.35%. Behavioral duration distribution was 76.27% for study behavior >30 minutes (long duration); 80.18% for eating behavior <5 minutes (fragmentation).

The MF-STGCN model has excellent performance, with an average relative error of only 4.23%-5.65% at a scale of 100-10,000 people, which is 2.93%-4.53% lower than that of the K-nearest neighbor and GCN models. The prediction time consumed grows linearly with the scale, and it takes only 35.53s for 10000 people, which is 9.2 times faster than that of K nearest neighbor, and the running time at 10000 people is 35.53s and 326.54s, respectively.

The model provides a quantitative baseline for abnormal behavioral warning, promotes the transformation of Civic Education management from empirical decision-making to data-driven, and supports precise intervention.

References

1. Lin, M., Wu, H., & Li, H. (2022, December). Research on Precise Ideological and Political Education in Colleges and Universities Based on the Analysis of Student Group Portraits. In International Conference on 5G for Future Wireless Networks (pp. 221-229). Cham: Springer Nature Switzerland.
2. Huang, Q. (2023). Research on Precise Ideological and Political Education Based on Improved K-means Algorithm for College Students’ Portrait Construction. Journal of Artificial Intelligence Practice, 6(4), 58-64.
3. WANG, S., & ZHAO, L. T. (2017). Data portrait”: The accurate characterization of college students’ ideological dynamic in the big data age. DEStech Transactions on Social Science Education and Human Science (iced).
4. Wang, J., Zhang, S., Hu, Y., & Li, L. (2024). Research on Precision Ideological and Political Education in Universities Based on Student Group Profiling Analysis. JOURNAL OF SIMULATION, 12(2), 39.
5. Liu, T. (2024). Approaches to Integrating Ideological and Political Education into Student Management in Colleges and Universities. Educational Innovation Research, 2(2), 103-108.
6. Lin, L., & Hai, S. (2016). Research on the multi-dimension ideological and political education of college students based on computer platform. Revista Ibérica de Sistemas e Tecnologias de Informação, (E7), 244.
7. Xu, C., Wang, C., & Yang, N. (2019, October). The supply-side precision reform and innovation of ideological and political education in colleges based on big data technology. In 4th International Conference on Modern Management, Education Technology and Social Science (MMETSS 2019) (pp. 637-643). Atlantis Press.
8. Luo, Y. (2021). Artificial intelligence model for real-time monitoring of ideological and political teaching system. Journal of Intelligent & Fuzzy Systems, 40(2), 3585-3594.
9. Xiaoyang, H., Junzhi, Z., Jingyuan, F., & Xiuxia, Z. (2021). Effectiveness of ideological and political education reform in universities based on data mining artificial intelligence technology. Journal of Intelligent & Fuzzy Systems, 40(2), 3743-3754.
10. Li, K., Jing, M., Tao, X., & Duan, Y. (2023). Research on online management system of network ideological and political education of college students. International Journal of Electrical Engineering & Education, 60(2_suppl), 377-388.

11. Zhou, W., & Yang, T. (2021, May). Application analysis of data mining technology in ideological and political education management. In Journal of Physics: Conference Series (Vol. 1915, No. 4, p. 042040). IOP Publishing.
12. Qiu, J. (2017, October). A Kind of Ideological and Political Education Management Platform Design Based on Web Technology. In 2017 International Conference on Robots & Intelligent System (ICRIS) (pp. 1-4). IEEE.
13. Han, W. (2020, April). Using information technology to improve ideological and political education and student management efficiency in colleges. In Journal of Physics: Conference Series (Vol. 1533, No. 2, p. 022126). IOP Publishing.
14. Huang, L. (2016, August). Teaching management data clustering analysis and implementation on ideological and political education of college students. In 2016 International conference on smart grid and electrical automation (ICSGEA) (pp. 308-311). IEEE.