future internet

MDPI

Article

The Effects of Non-Directional Online Behavior on Students' Learning Performance: A User Profile Based Analysis Method

Kun Liang, Jingjing Liu * and Yiying Zhang

College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China; liangkun@tust.edu.cn (K.L.); yiyingzhang@tust.edu.cn (Y.Z.)

* Correspondence: antares@mail.tust.edu.cn; Tel.: +86-152-3010-0726

Abstract: Network behavior analysis is an effective method to outline user requirements, and can extract user characteristics by constructing machine learning models. To protect the privacy of data, the shared information in the model is limited to non-directional network behavior information, such as online duration, traffic, etc., which also hides users' unconscious needs and habits. However, the value density of this type of information is low, and it is still unclear how much student performance is affected by online behavior; in addition there is a lack of methods for analyzing the correlation between non-directed online behavior and academic performance. In this article, we propose a model for analyzing the correlation between non-directed surfing behavior and academic performance based on user portraits. Different from the existing research, we mainly focus on the public student behavior information in the campus network system and conduct in-depth research on it. The experimental results show that online time and online traffic are negatively correlated with academic performance, respectively, and student's academic performance can be predicted through the study of non-directional online behavior.

Keywords: undirected online behavior; multinomial regression; feature extraction; correlation analysis

check for updates

Citation: Liang, K.; Liu, J.; Zhang, Y. The Effects of Non-Directional Online Behavior on Students' Learning Performance: A User Profile Based Analysis Method. *Future Internet* **2021**, *13*, 199. https://doi.org/10.3390/fi13080199

Academic Editor: Paolo Bellavista

Received: 28 June 2021
Accepted: 28 July 2021
Published: 31 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In the network environment, users' casual, fragmented online behavior information is recorded, which can directly or indirectly reflect users' personality, characteristics, preferences, attitudes, and habits, etc. The research of network behavior is closely related to sociology, psychology, and anthropology, etc. It studies the regularity of network behavior in order to control and predict network behavior. User portraits [1,2] have not been a technology focused on by network behavior research in recent years, the goal of which is to extract the multidimensional attribute information (such as gender, age, and educational background) of users from massive data for mining and analysis, and to predict the characteristics of users and the laws behind their behaviors.

The behavior characteristics of campus network users are more unique [3]. The campus network provides online services for students, and the authentication gateway system records the log of students' network behavior, which has a huge amount of data and hides the objective law of students' network behavior [4]. These data have no obvious regularity, so it is difficult to directly divide users with similar characteristics into categories according to the original data. Students' online behavior can be divided into directional online behavior and non-directional online behavior. Directional online behavior refers to the user's specific network behavior dynamics, such as browsing websites and comments, etc. It is obvious that more user characteristics can be obtained by analyzing the data of directional behavior. However, such directional online behavior often involves revealing too much of users' privacy that cannot be disclosed to the public. When students surf the Internet at school through the school gateway, students' online data can truly reflect their online behavior. It is feasible to analyze and study students' online behavior by using their online data [5]. The log records the user's operation of using the network, such as login

time, logout time, usage time, and usage flow. Although this kind of data is easy to obtain, the data structure is complex, the value density is low, and the increment is rapid, which is often ignored by people, and there is little research on this kind of data [6]. In fact, these data often contain a lot of hidden information related to learning and life. If we can analyze these data scientifically and effectively, and make reasonable use of the analysis results, it will play a great role in promoting the school's teaching management [7].

This paper studies the influence of students' non-directional online behavior on learning and proposes an online behavior and score combined (OBSC) model based on the user portrait. Section 4 is used to extract the characteristics of the user's non-directional online behavior attributes, describes the user's online behavior preferences through statistical and cluster analysis, and determines whether the students tend to indulge in the Internet and bad online habits, to support the relevant teaching decisions or proceed with the corresponding educational intervention [8]. The fifth section of the paper uses a polynomial regression model based on the least square method to realize the correlation analysis and prediction of the user's academic performance.

The main contributions of this paper are as follows:

(1) Introduce the related research work.
(2) The concept of non-directional Internet behavior is put forward, and the user profile technology is used to analyze the user's Internet data.
(3) The feature extraction of users' non-directional Internet behavior is carried out by cluster analysis.
(4) The method of polynomial regression is proposed to predict students' academic performance, and the influence of non-directional Internet behavior on students' learning is analyzed.

2. Related Work

Whether online behavior is scientific and reasonable is one of the important factors affecting the development of the physical and mental health of contemporary college students. Any measure of mobile phone use, whether considered normative or problematic, quantifies the extent to which a person uses a phone, feels an emotional or other dependence on a phone, or categorizes the types of use and situations in which use occurs [5]. Most studies support the hypothesis that there is a negative correlation between Internet dependence behavior and students' academic performance [9,10]. They found valuable information in huge data and studied the characteristics of user's network behavior [11], so as to make judgments on network optimization and search engine optimization. Fan [12] used the user behavior log as the basic data set to study the user's personal preferences and analyze the potential purchase demand, so as to realize the digital research on user demand, and to realize the digital research on the needs of users. Qiao [13] proposed a new hybrid model called OBLD (User Online Behavior Linkage over Domains), which links the online behavior of cross-domain users with network traffic. This model derives several important attributes from the user's online behavior, such as the user's digital identity, and their various fingerprints on the terminal and browser.

In the face of massive amounts of information, how to enable users to obtain the required information quickly and accurately is a difficult problem currently faced by information retrieval. Building user portraits helps to quickly mine different characteristics of user groups in massive data to meet personalized needs [14]. Srisura B et al., proposed a network usage log mining framework that can mine, track, and verify the dynamic multifaceted user profile information [15]. Wang Lee proposed a user profile method based on the online behavior log. Firstly, the user feature set is constructed by feature selection and feature extraction, and then the user profile model is constructed by using the technology of model stacking to combine multiple single classifiers. This method can greatly improve the accuracy of identifying the gender, grade, and age attributes of users [16]. A cross-modal learning idea was proposed, and a user profile model based on multimodal fusion was designed [17]. The stacking integration method was used to integrate multiple multimodal

learning joint representation networks to learn the corresponding model combination; the attention mechanism, introduced to enable the model to learn the contribution of different modal representations to the prediction results, was different.

With the development of artificial intelligence technology, the machine learning algorithm has been gradually applied in the field of behavior analysis. Scholars can analyze social media and extract emotional information from it, which can predict user demands [18]. K. Ikeda et al., mined the text data and interactive data of Twitter users, and performed clustering analysis on the collected data to generate Twitter user portraits. The user portraits intuitively showed the characteristics of users using Twitter and other microblog social networks [19]. Grieve studied Snapchat, an instant messaging software based on image social tools, made a portrait of its user group, and found that Snapchat's audience is mainly young people who prefer the use of image communication [20]. At the early stage of user profile development, this was mostly used in the field of e-commerce [21]. Some colleges and universities in China also apply user profiles to library services, pre-warning of failed subjects and pre-warning of student status, providing a dynamic analysis of thoughts [22]. Chen et al. analyzed the basic information, the online learning behavior, and classroom performance of learners under the open teaching, combined with brain cognitive experiments; they explored the characteristics of learners' interests, hobbies, and learning ability from the perspective of data mining and cognitive psychology, and summarized and depicted their personalities in the form of labels [23]. Liang analyzed the relation indicators of E-Learning to build the student profile and proposed the intelligent guide model to guide learners to improve online learning according to the E-Learning resources and learner behaviors [22]. All of these studies have promoted the development of user profiles in the field of education, but there are few types of research on mining and depicting user profiles from the data in the network log, and on the correlation analysis of learning achievement.

3. Model Structure

In this section, the processes of Online Behavior and Score Combined (OBSC) model construction are described. Model construction consists of three parts, which are data processing, feature acquisition, and Behavior-Score analysis. The OBSC model based on user profiles is shown in Figure 1.

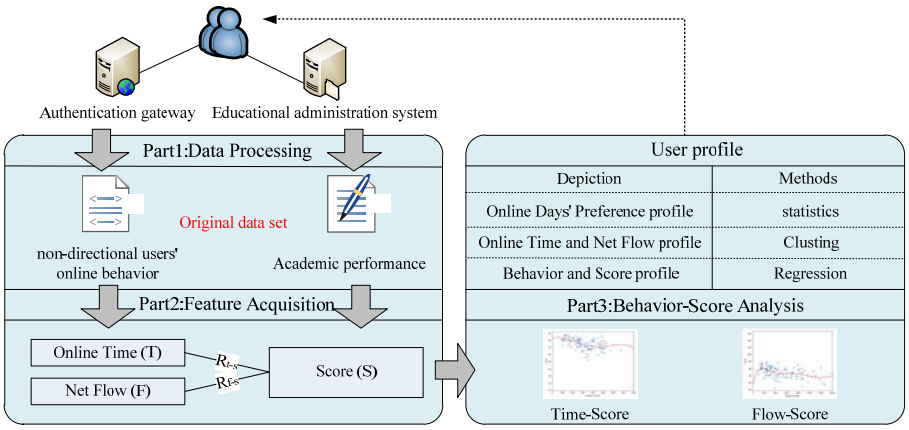


Figure 1. User profile based structure of OBSC model.

(1) Data processing is used to collect the required raw data. The data source includes two parts: one is the non-directional online behavior data from the campus network authentication gateway log, the other is the student academic performance data from the educational administration management system. By removing the null value, data standardization, and other operations to clean and organize the original data, we can obtain effective online behavior data and academic performance data.

(2) In the feature acquisition part, we select and extract the feature of the original data to build the tag database, and extract the feature of online time, flow, and terminal examination score. The K-MEDIODS clustering algorithm is used to obtain the user's preference features of online behavior, and these preference features are classified and marked to depict the user's profile.
(3) The behavior-score analysis algorithm uses the polynomial regression method based on the least square method, through the training of the sample set, to predict the students' learning performance.

3.1. Data Source and Description

The network behavior that this research focuses on specifically refers to a series of data information generated by users through interaction and online behaviors in the network system, such as login, logout, mouse clicks, page views, online reviews, online duration, and traffic usage. For example, in an education website, the possible network behavior attributes are shown in the following Table 1. However, since the private data is protected, the data obtained in this study can only be the displayed flow and length of online time, and specific website interaction information cannot be obtained.

Table 1. The features of network behavior.

Network Behavior	With/Without Interaction	Directed/Non-Directed
Click	With interaction	Directed
Browse		
Slide		
Movie watching		
Online evaluating		
Flow	Without interaction	Non-directed
Length online time		
Login time		
Logout time		

3.1.1. Data Description of Non-Directional Internet Behavior

There are more than 20,000 students in a college in Tianjin. The school allowed us to access the school's certification gateway log and collect the original data set of non-directional users' online behavior between 2015 and 2019, which contained 11 attributes of 9950 science and engineering students (without distinguishing majors and grades), and approximately 12.5 million records. The original data had the typical time series characteristic [24], which records the user behavior data of each login, that is, the same user corresponds to multiple login records, including user ID, login time, logout time, length of login time, total traffic, IP address, MAC address, international upward traffic, international down traffic, domestic upward traffic, domestic down traffic, where total traffic = upward traffic + down traffic, and the length of login time = logout time − login time. The missing values and irrelevant features in the original data were cleaned to make the processed data more complete and obvious, which is convenient for further calculation and conversion to the key features of the user profile [25]. Since the IP addresses in the records were the same, and the MAC addresses were all 0, these two attributes had little impact on the research, so they were cleaned out through data processing.

The original data set described the user's non-directional online behavior. The user attributes included user ID, Login_time, Logout_time, Length_time, Flow, etc. The meaning of each attribute is shown in Table 2.

Table 2. Several important attributes and explanations of non-directional online behavior.

Attribute Name	Description
User ID	The user's account number is unique for each campus network user.
Login_time	The time each user logs in to the campus network.
Logout_time	The time when the user logs off from the campus network.
Length_time	The duration of each login to the campus network, in minutes.
Flow	The network flow is used for each login, in MB.
Flow_up_I	The international uplink flow is used for each login, in MB.
Flow_down_I	The international downlink flow is used for each login, in MB.
Flow_up_N	The domestic uplink flow is used for each login, in MB.
Flow_down_N	The domestic downlink traffic is used for each login, in MB.
IP	Internet Protocol Address.
MAC	Media Access Control Address.

3.1.2. About Academic Performance

The academic achievement data comes from the educational administration system of a university in Tianjin of China, which collected 82 attributes of 9950 student users, including students' personal private information, such as name, gender, and age, etc.; and score information for more than 40 courses between 2015 and 2019, and score statistics information. Each user has a record corresponding to the data user's non-directional online behavior. Since this study only analyzed the impact of student users' online behavior on their academic performance, the data collection in this paper did not distinguish professional grades and treated all subjects equally. There were four attributes of the original data, taken from all courses of the University during the four years that were studied, such as Course ID, Course Name, User ID, and average grade.

There were two types of academic achievement, one was digital, the other was grade. In order to standardize the performance, we converted the excellent, good, medium, pass, and fail grades into the digital grades 90, 80, 70, 60, and 50 according to the scope of the school performance evaluation standard. Some subjects had blank scores, indicating that students had not taken the course.

3.2. Attribute Analysis and Standardization

In the calculation of campus network flow [26], there is no distinction between international flow and domestic flow, so the international upstream flow and domestic upstream flow can be combined into upstream flow "Flow_up", and the international downstream flow and domestic downstream flow can be combined into the downstream flow "Flow_down". The upstream flow generally includes the flow consumed by users sending data requests to the server through the computer and the flow consumed by uploading data. Generally, the consumption of users is less, and individual users uploading data to cloud tools such as the network disk may generate a large amount of upstream flow. Downstream flow generally refers to the flow consumed by data transmission from the network end to the user, including downloading data, watching videos, and data transmission from the network server or other computers to the user's computer. In addition, the particularity of authentication gateway data in a campus network is that the upstream flow is not included in the usage flow, so the user flow is mainly composed of all the downstream flow, that is, Flow = Flow_down_I + Flow_down_N, which can be regarded as a key feature of non-directional online behavior.

Considering the validity of user attributes, we filtered all the collected records according to the following rules.

Rule 1: the user's login time is more than 10 min, i.e., Length_time > 10;

Rule 2: the user's usage flow is greater than 10 MB each time, i.e., Flow > 10.

In order to integrate data with different attributes, the first step is to nondimensionalize [27], which is to transform the original data with a measurement into data without a unit. The nondimensional data processing solves the comparability between the network

behavior data of different attributes of campus network users, which makes the fusion analysis of different attributes possible. For example, in the original data, the feature scales and units such as Length_time and Flow were not consistent, so the original data needed to be centralized and standardized so that different features had the same scale. The data with the mean value of 0 and a standard deviation of 1 are calculated by the Formula (1).

$$x' = \frac{x - \mu}{\sigma} \quad (1)$$

where, x is the attribute value of different dimensions of each record, i.e., sample point; μ is the mean value; σ is the standard deviation.

In many users, there were some similar groups, which were clustered. There were also some abnormal users outside of the groups, which were treated as extreme values. However, taking into account the possible surge in online time or the use of traffic was precisely the main factor affecting academic performance. Therefore, the maximum value was retained. In the clustering results, these users were also analyzed as a group in order to find their abnormal behavior patterns and compare them with other user groups. The minimum value of 0 was filtered through the above rules. After the above steps, a new effective data set was finally generated. The above processes are the most commonly used data processing methods. This paper will not discuss the process of removing missing values and standardizing.

4. User Feature Model

4.1. Label Library Construction

According to the user attributes and dynamic behavior characteristics, the user profile summarizes the user's individual or group preference characteristics and labels them. The user attributes here mainly include non-directional online behavior attributes and user academic performance attributes. Users' attributes are abstracted from labels to extract features with a high ability to distinguish labels. The label is the symbol identification of user characteristics, which has two important characteristics [28]. First, it has a certain population, which can sample and summarize the characteristics of things to a certain extent [23]; second, it can use symbols to represent a certain kind of characteristic of users, such as Chinese, English, or numbers. Label library is the centralized management of labels, which is used to mark user behavior and attributes.

Obtaining feature labels is the key to building a user profile. Let user feature set D include two parts: labeled user feature set D^A (training set) and unlabeled user feature set D^U (test set), $D = D^A \cup D^U$. In order to realize the automatic marking of the $D^U \epsilon D$, we trained the users of the given training set $D^A \epsilon D$, and clustered the user attributes of the training set, so as to label the user $D^U \epsilon D$ with the same characteristics.

4.2. Feature Extraction of Non-Directional Online Behavior

The user portrait in the campus network environment mainly describes the user's non-directional online behavior attribute, learning attribute, and the relationship between them. There are two key characteristics of non-directional online behavior attributes: online time and net flow. The clustering method can be used to extract user behavior characteristics and to observe the impact of user's online behavior habits on their academic performance. The user profile is mainly constructed from three dimensions: online time, net flow, and performance. The user feature set is expressed in a quadruple form, i.e., D = ((T,F),S,R), where T is the online time, F is the net flow, S is the academic performance, and R represents the relationship between the two features.

According to the natural habit, a day is divided into 24 time intervals, with time label Hi = {0:00–1:00,1:00–2:00, . . . , 23:00–0:00}, where i = 1,2, . . . ,24. The number of days online is expressed as d. According to the attribute "Login_time", the cumulative time of each time interval is t_i by summing the user's online time, and the cumulative flow of each time interval is $flow_i$ by summing the net flow. The user's net flow in the different time

interval in a period time (such as days d) is $F = \sum_{0}^{d} flow_i$. Total time is $T = \sum_{0}^{d} t_i$. Finally, the processed feature set (T, F) is obtained.

4.3. Online Behavior Preference Algorithm Based on Clustering

Clustering can group objects with similar characteristics in one class, and objects in the same class have high similarities. In this paper, we use the K-MEDIODS algorithm to cluster [29]. Its basic idea is to reduce the overall loss value of the data set, improve the quality of the cluster, and calculate the square sum of the error of the data set as the loss value of each cluster. Observing the online time of users in different time intervals can reflect the user's online time preference and the intensity of using the network [10,28].

Therefore, the K-MEDIODS algorithm is used to cluster the user's online time in different time intervals, to divide the users with similar behaviors into the same cluster, and to divide the users with different behaviors into different clusters, which can more intuitively observe the user's behavior differences in different time intervals. Because the amount of data filtered in this paper conforms to the characteristics of small datasets, the algorithm will have little impact on the system running speed, but the selection of K value needs to be tested repeatedly. The user's online time t in each time interval is taken as the clustering feature, and the process of the online time preference algorithm is shown in Figure 2.

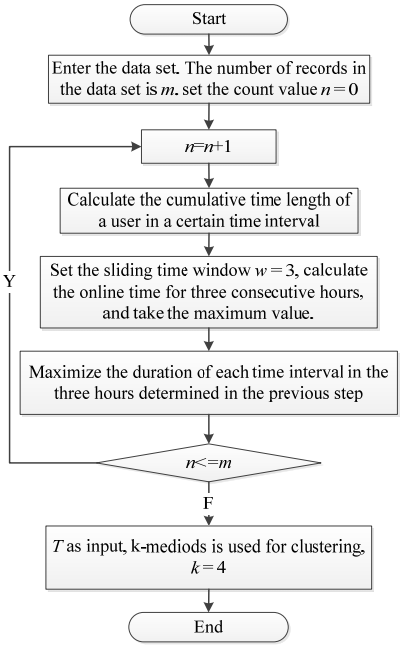


Figure 2. Online Behavior Preference algorithm flow chart.

Step 1: calculate the cumulative time $t_{n,i}$ of the user in the time interval i to form the matrix T_1.

$$T_1 = \begin{bmatrix} t_{1,1} & t_{1,2} & \cdots & t_{1,i} \\ t_{2,1} & t_{2,2} & \cdots & t_{2,i} \\ \vdots & \vdots & \vdots & \vdots \\ t_{n,1} & t_{n,2} & \cdots & t_{n,i} \end{bmatrix}$$

T_1 is the matrix of n rows and 24 columns and $t_{n,i}$ refers to the cumulative online time t_i of the n-th student in the i hour of d day. For example, $t_{2,3}$ is the online time of the second student in the hour 2:00–3:00.

Step 2: according to expert opinions, a person's continuous online time is best within 1–3 h, set a sliding time window W = {$w1$,$w2$, . . . ,$w24$}, with a length of 3, is used to select

the three consecutive hours of online time in a day. The time window starts from the three periods of $w1$ = 23, 24, 1, each time sliding backward for one-time interval, the next time window will be the three-time intervals of $w2$ = 24, 1, 2, and so on, until $w24$ = 22, 23, 24. Every time the window slides, calculate the online time length by $\sum_{i}^{i+2} t_{n,i}$, that is, by summing the three consecutive elements of each row in T_1, that the matrix T_2 is obtained

$$T_2 = \begin{bmatrix} t_{1,23}+t_{1,24}+t_{1,1} & t_{1,24}+t_{1,1}+t_{1,2} & \cdots & t_{1,22}+t_{1,23}+t_{1,24} \\ t_{2,23}+t_{2,24}+t_{2,1} & t_{2,24}+t_{2,1}+t_{2,2} & \cdots & t_{2,22}+t_{2,23}+t_{2,24} \\ \vdots & \vdots & \vdots & \vdots \\ t_{n,23}+t_{n,24}+t_{n,1} & t_{n,24}+t_{n,1}+t_{n,2} & \cdots & t_{n,22}+t_{n,23}+t_{n,24} \end{bmatrix} = \begin{bmatrix} t_{1,w1} & t_{1,w2} & \cdots & t_{1,w24} \\ t_{2,w1} & t_{2,w2} & \cdots & t_{2,w24} \\ \vdots & \vdots & \vdots & \vdots \\ t_{n,w1} & t_{n,w2} & \cdots & t_{n,w24} \end{bmatrix}$$

The maximum value $\max(t_{1,w1} : t_{1,w24})$ of each line of T_2 is calculated as the maximum online time of the user for three consecutive hours, assuming that the result obtained after the maximum value is decomposed as shown in matrix T_3

$$T_3 = \begin{bmatrix} t_{1,w24} \\ t_{2,w4} \\ \vdots \\ t_{n,w7} \end{bmatrix} = \begin{bmatrix} t_{1,22} & t_{1,23} & t_{1,24} \\ t_{2,2} & t_{2,3} & t_{2,4} \\ \vdots & \vdots & \vdots \\ t_{n,5} & t_{n,6} & t_{n,7} \end{bmatrix}$$

Step 3: in matrix T_3, compare the online time length $t_{n,i}$ of each time interval to obtain the period time where the maximum $t_{n,i}$ is located, assuming that its maximum value is as shown in the matrix T

$$T = \begin{bmatrix} t_{1,23} \\ t_{2,2} \\ \vdots \\ t_{n,7} \end{bmatrix}$$

Step 4: take t as input, use K-MEDIODS to cluster, and set K value as 4.

5. Behavior-Score Analysis Model

The purpose of analyzing non-directional online behavior is to understand students' online behavior preferences and rules, and to make an accurate prediction of their academic performance and verify the impact of online behavior on them. The Behavior-Score analysis algorithm can realize the prediction of a user's academic performance and minimize the error.

5.1. Analysis Method of Correlation of Learning Achievement

We used the least square method [30] in the regression model of learning achievement prediction. The idea is to find the best function matching of data by minimizing the square sum of errors. By using the least square method, the unknown data can be simply obtained, and the sum of squares of the errors between the obtained data and the actual data can be minimized. Usually, in the study of simple one-dimensional data, the purpose of prediction is achieved by fitting accuracy.

The general form of the least square method is as follows:

$$\text{objective } function = \sum (\text{objective } value - theoretical)^2$$

The observation value is a group of samples, and the theoretical value is a hypothetical fitting function. The objective function is the loss function in machine learning [31]. For example, when we study the relationship between two variables X, Y, we can usually obtain a series of pairs of data $(X_1, Y_1; X_2, Y_2; \ldots; X_m, Y_m)$. When these data are depicted

in the rectangular coordinate system, a straight line can be fitted near these points. As shown in Formula (2)

$$\hat{f}(x) = \sum_{i=1}^{b} \alpha_i \varphi_i(x) \tag{2}$$

The least square formula (3) is as follows

$$J_{LS}(\alpha) = \sum_{i=1}^{n} \left(\hat{f}(x_i) - y_i\right)^2 = ||X\alpha - \mathrm{y}||^2 \tag{3}$$

where $X = \varphi_i(x)$. The slopeα of the fitting line is $\alpha = argmin J_{LS}(\alpha) = \left(\mathrm{X^T X}\right)^{-1} \mathrm{X^T}$, x represents two behavior attributes(T,F), y_i represents the user's score, $\hat{f}(x)$ represents the curve fitted by the model, and α_{LS} is the vector to be learned by the model. The data set is divided into two parts, 70% of which is the training data set, and the vector α_i to be learned is obtained by training. The remaining 30% of the data is used as the test set to predict the academic performance of users, and the impact of user behavior on their learning will be obtained.

By using this method, we can obtain the unknown parameter, which makes the loss function minimum, and then obtains the best fitting curve. This method can also be extended to the nonlinear fitting of multiple sample features.

5.2. Behavior-Score Analysis Model Based on Polynomial Regression

To analyze the impact of non-directional online behavior on students' academic performance, it is necessary to analyze the relationship R between two key features of non-directional online behavior and their score, that is "time-score" $R_{d\text{-}s}$ and "flow-score" $R_{f\text{-}s}$.

There are 82 attributes in the score set. Except for the user name attribute, the other 81 attributes are all courses. The corresponding score of each course is S_x, (x = 1,2, . . . ,81). Due to the different subjects and numbers of final examinations for each user, to simplify the scores, we first calculated the average final score of each user, which is recorded as AVE_{score}, and the calculation method is shown in Formula (4).

Then, we added up the daily net flow in all the effective online records of the same user, divided by the online days, and obtained the daily average flow of each user, which is recorded as AVE_{flow}, and the calculation method is shown in Formula (5). These two types of data are used as the basis for the correlation analysis of user average score and daily average net flow.

In the same way, we added the daily online time in all the effective online records of the same user and divided by days to obtain the daily average online time of different users, which is recorded as AVE_{length}. The calculation method is shown in Formula (6). This kind of data is taken as the basis of the correlation analysis between the average score of users and the average online time.

$$AVE_{score} = \frac{\sum S_x}{m} \tag{4}$$

where m is the number of scores.

$$AVE_{flow} = \frac{F}{d} \tag{5}$$

$$AVE_{length} = \frac{T}{d} \tag{6}$$

where d is the number of days users are online, m is the number of courses.

According to empirical assumptions, the polynomial regression equation of flow-score is

$$y = ax^3 + bx^2 + cx + d \tag{7}$$

The polynomial regression equation of length-score is

$$y = mx^4 + nx^3 + px^2 + qx + s \tag{8}$$

6. Experimental Results and Analysis

6.1. Online Days Preference Profile of Individuals and Groups

User preference is a type of behavior preference that the user shows inadvertently [9]. According to the online time of student users, we can mine their online habits and preferences, and define the user behavior feature labels. Online days preference is based on the online time interval i to calculate the personal online days in three months and the overall average online days to compare the preference difference. We chose 132 students of the same grade and major and compared one student's online time with the group in order to finally obtain the user's profile regarding their online days' preference compared with individuals and the group.

In this experiment, we used Python 3.7 to calculate the online time preferences of individuals and groups based on the statistical method. The line chart is shown in Figure 3. The X-axis is the online time interval, and the Y-axis is the online days.

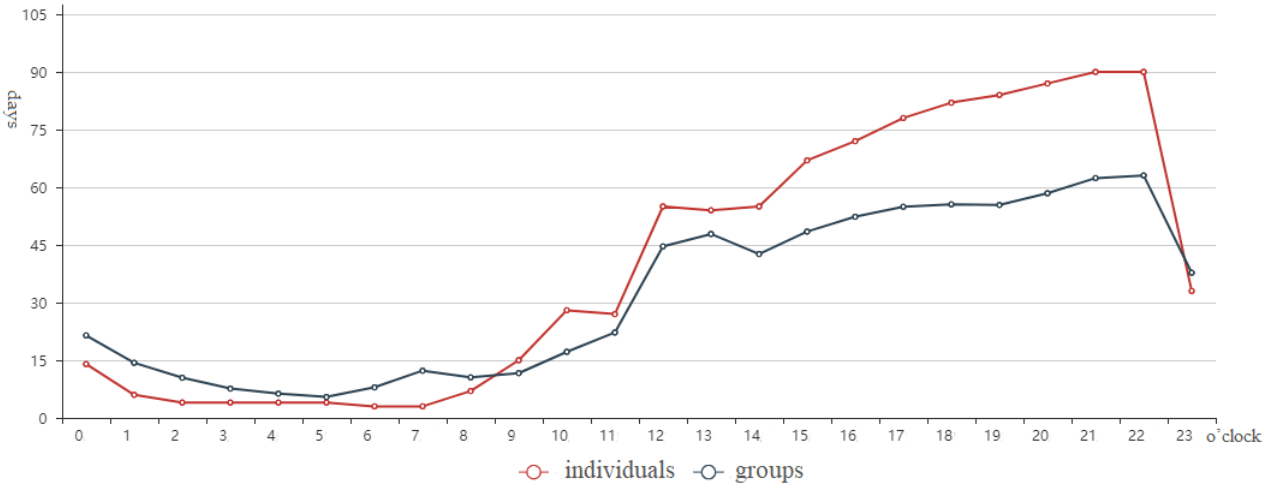


Figure 3. Comparison of time preference between individuals and groups.

Figure 3 shows that from 8:30 in the morning, the number of online days of a student was significantly more than that of the group, especially after midday. In addition, due to the influence of the school power supply time, the number of online days increased gradually between 12:00 and 22:00. It can be seen that the online time preference of student users was between 12:00 and 22:00, and reached the maximum value at 22:00. At 23:00 there appears an obvious downturn, which coincided with the school's required rest time. The rest time of the individual student was similar to that of the group, but the student was online between 21:00 and 22:00 every day for almost three months, and the average online days of the group in this period was approximately 47 days. This proves that the students are very dependent on the network, which requires the counselors and teachers to take effective measures to intervene and give more attention and correct guidance.

However, this experiment was mainly based on the statistical method, from the perspective of online days to observe the difference in online behavior between individuals and groups; thus, the extracted preference information is limited. However, in the attribute of non-directional online behavior, there was an important feature of net flow. Therefore, we used the clustering method to describe the user profile from the two dimensions of net flow and online time.

6.2. Online Time and Net Flow Profile Based on Clustering

To some extent, the online time and duration of users reflect their dependence on the network. Clustering the user's online time and duration can divide the user group's online behavior preferences into different levels, thus tagging the user profile.

Based on the experiment in Section 6.1, we further calculated the online time length of each time interval in a month and used the K-MEDIODS algorithm to cluster the online time and the duration. Due to a large amount of data, the clustering results of high-density points affected the observation. In order to make the clustering results clearer, we also performed further processing of the data. We used the steps in Section 4.2 to calculate the online time of users in a month. After repeated tests, the clustering effect was most obvious when the final K value was 4, as shown in Figure 4.

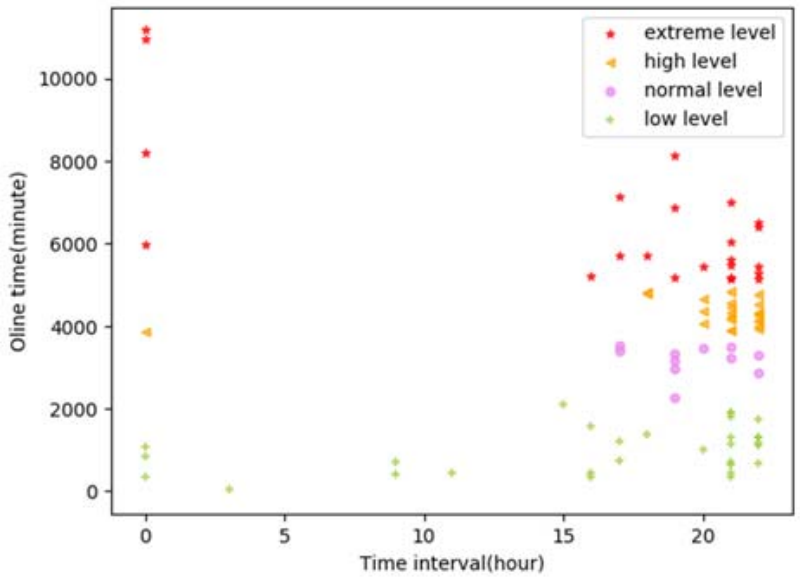


Figure 4. Online time preference by clustering.

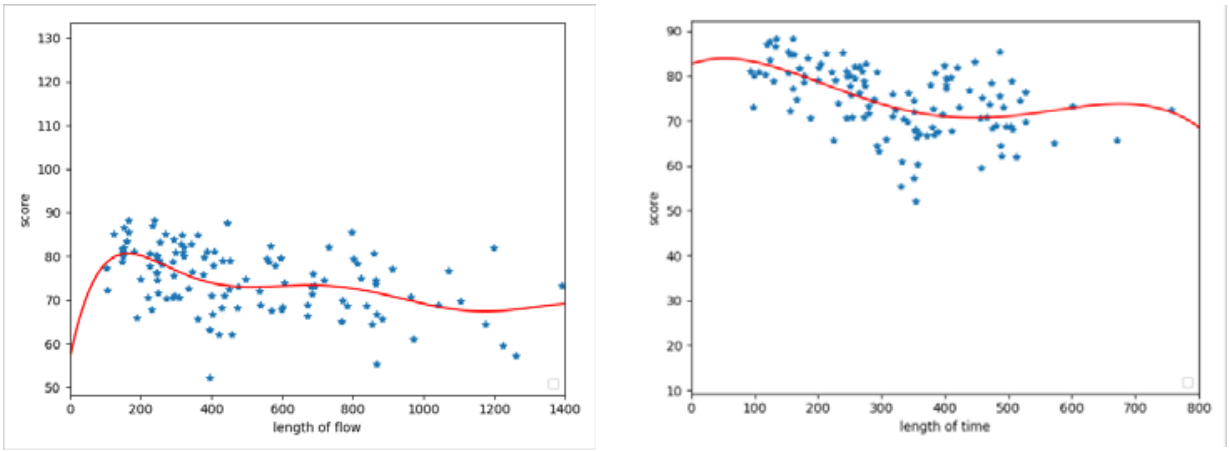
In Figure 4, the X-axis represents the online time interval, the Y-axis represents the total time used in the time interval, each point represents a user, four different colors are different clusters, and their online preference levels are low, normal, high, and extreme. For example, most users liked to be online after 19:00, which is a break without classes. Such students were labeled as "self-disciplined". Some users were online at 0:00, they preferred to stay up late, and they used a lot of net flow during this time interval, thus, such students were labeled as "night owls".

6.3. Regression Results of Non-Directional Online Behavior and Score

The experimental purpose of the polynomial regression model [32] is to fit the curve trend between performance and online behavior by training the features of the user's non-directional online behavior in the user profile, and take the regression curve as the analysis model so as to predict the student's academic performance according to the user's non-directional online behavior.

We calculated the students' average score AV_{score} and daily average flow AVE_{flow} as the training data. After many times of training, when the degree was 7, the fitting effect was the best. Figure 5a shows the quartic polynomial regression model for score and net flow. The X-axis is the average daily net flow (MB), the Y-axis is the score, and the flow-score regression model is

$$R_{f-s} = -19x^7 - 2.197E - 15x^6 + 5.87E - 12x^5 - 8.004E - 9x^4 + 5.899E - 6x^3$$



(**a**) (**b**)

Figure 5. Regression model of non-directional online behavior and score. (**a**) Flow-score quartic polynomial regression model. (**b**) Time-score cubic polynomial regression model.

By calculating the average score AV_{score} and daily average online time AVE_{length} as the training data, after many experiments, when the degree was 4, the fitting effect was the best. Figure 5b shows the cubic polynomial regression model of time and score. The X-axis is the daily average online time (minutes), the Y-axis is the performance, and the time-score regression model is

$$R_{t-s} = -7.633E-10x^4 + 1.196E-0.6x^3 - 0.0005505x^2 + 0.04846x + 82.68$$

The experimental results show that the longer users are online or the more traffic they use, the lower their academic performance is. This proves that users' irregular online behavior will harm their learning. Through the analysis of the model, we can find out whether users are addicted to the network in time, and draw the attention of teachers or schools to this in order to formulate relatively strict online behavior management strategies for them.

7. Conclusions

In order to solve the problem of whether non-directional Internet behavior has any influence on a learner's academic performance, this paper constructs an OBSC model based on user profile, which comprehensively covers the variables needed for the study of this issue, and proposes to use network behavior data as a new index to evaluate students' learning performance. In this paper, the learning behavior records of each student were collected from the school certification network log, and the data was sorted using Python technology. In the process of solving the problem, we first used the K-MEDIODS algorithm to extract user features and then used the polynomial regression model to test the relationship between online behavior and academic performance. The results show that there is a significant difference in learning performance among students based on non-directional online behavior. Compared with other studies, this paper achieves a significant effect of predicting academic performance through qualitative analysis of limited behavioral information under the premise of protecting student users' online privacy. However, this study has several limitations. First of all, because of the diversity of the research structure on the generation of non-directional Internet behavior, a single generalization effect cannot be obtained. The second limitation is the quality of the data included in the study. Although the samples for studying educational achievements include students' academic achievements in various disciplines, there are still many indicators that can be used to test students' learning level, which have not been included in this

study. However, the final effect mainly depends on these factors. High-quality sample data research is necessary and will not be greatly affected by smaller studies.

In this study, through the analysis of non-directional online behavior, it was found that online time and traffic usage harm the final online learning performance to a certain extent. Students' non-directional online behavior can be used to predict their final learning performance. In the process of online learning, this analysis is conducive to the early warning of learning risk, helping teachers to guide students to produce benign online learning behavior, the early detection of Internet addiction, and timely intervention. In the future, we will further explore the potential impact of learners' network behavior on their studies, pay attention to and cultivate teachers' correct guidance for students' network use, and put forward new opinions on early warnings of academic risk.

Author Contributions: Conceptualization, K.L. and Y.Z.; methodology, J.L.; validation, J.L.; resources, K.L.; writing—original draft preparation, J.L. and K.L.; review, Y.Z.; visualization, J.L.; funding acquisition, K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (61807024).

Data Availability Statement: The data source studied in this paper are the original data set of non-directed users' online behavior collected through the log of a college certification gateway.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jiang, M.; Miao, D.; Luo, S.; Zhao, C.R. Multi granularity user portrait based on granular computing. *Pattern Recognit. Artif. Intell.* **2019**, *32*, 691–698. (In Chinese)
2. Ouaftouh, S.; Zellou, A.; Idri, A. Social recommendation: A user profile clustering-based approach. *Concurr. Comput. Pract. Exp.* **2019**, *31*, e5330. [CrossRef]
3. Liu, X. Analysis of Network User Behavior on Campus Network. *Comput. Inf. Sci.* **2018**, 98–103.
4. Liang, K.; Liu, J.; Pang, H. Visualization and Analysis of Users' Online Behavior in Campus Network. In Proceedings of the ICNCC 2019: 2019 The 8th International Conference on Networks, Communication and Computing, Luoyang, China, 13–15 December 2019.
5. Hu, Z.; Jie, S. Research on online behavior Analysis and data Mining of College students. *Distance Educ. China* **2017**, *2*, 26–32. (In Chinese)
6. Kates, A.W.; Wu, H.; Coryn, C.L.S. The effects of mobile phone use on academic performance: A meta-analysis. *Comput. Educ.* **2018**, *127*, 107–112. [CrossRef]
7. Wu, L.; Lao, C.; Liu, Q.; Cheng, Y.; Mao, G. Research on online Learning behavior Analysis Model and its Application in Web-based Learning Space. *Mod. Educ. Technol.* **2018**, *6*, 46–52.
8. Zhang, M.; Du, X.; Li, H. Research on Early Warning of Achievement Based on the Analysis of Students Behavior Patterns. Available online: https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CAPJ&dbname=CAPJLAST&filename=JSGG20210406003&v=3Qinehu7XHJPX3fibC28GjbxFazuMUYY%25mmd2BJ9hp0LD%25mmd2FTpWUwyLCBKIdTm7ZnMoCkAl (accessed on 30 July 2021). (In Chinese).
9. Labriji, A.; Charkaoui, S.; Abdelbaki, I.; Namir, A. User interest center based on a semantic user profile. In Proceedings of the International Conference on Multimedia Computing & Systems, Marrakech, Morocco, 29 September–1 October 2016.
10. Na, D.; Wei-Na, L.; Bo-Tao, H. Modeling Method of Network Abnormal Behavior Based on Big Data. *Electr. Power Inf. Commun. Technol.* **2018**, *1*, 6–10. (In Chinese)
11. Hejun, Z.; Liehuang, Z.; Meng, S.; Salabat, K. Online and automatic identification and mining of encryption network behavior in big data environment. *J. Intell. Fuzzy Syst.* **2018**, *34*, 1111–1119. [CrossRef]
12. Fan, J. *Research on Log Analysis System Based on User Behavior*; Jilin University: Changchun, China, 2018.
13. Qiao, Y.; Wu, Y.; He, Y.; Hao, L.; Lin, W.; Yang, J. Linking user online behavior across domains with internet traffic. *J. Univers. Comput. Sci.* **2018**, *24*, 277–301.
14. Wu, W. User portrait technology and its application analysis. *Chin. Sci. Technol.* **2019**, *21*. (In Chinese). Available online: https://m.fx361.com/news/2019/1230/6243788.html (accessed on 30 July 2021). (In Chinese).
15. Srisura, B.; Wan, C.; Sae-Lim, D.; Meechoosup, P.; Win, K.M. User Preference Recommendation on Mobile Car Parking Application. In Proceedings of the IEEE International Conference on Mobile Cloud Computing, Bamberg, Germany, 26–29 March 2018.
16. Le, W. User portrait method of online behavior log based on model stacking. *J. Shandong Univ. Sci. Technol.* **2018**, *37*, 70–79.
17. Zhang, Z. User portrait method based on multimodal fusion technology. *J. Peking Univ.* **2020**. [CrossRef]
18. Teo, J.; Hou, L.; Tian, J.; Mountstephens, J. Classification of Affective States via EEG and Deep Learning. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 5. [CrossRef]

19. Li, J.; Wang, X.; Chang, Y.; Zhang, C. Research on the use behavior portrait of WeChat elderly users based on mobile terminal logs. *Libr. Inf. Work* **2019**, *63*, 31–39.
20. Farnadi, G.; Tang, J.; De Cock, M.; Moens, M.-F. User profiling through deep multimodal fusion. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Marina Del Rey, CA, USA, 5–9 February 2018.
21. Grieve, R. Unpacking the characteristics Snapchat users: A preliminary investigation and agenda for future research. *Comput. Hum. Behav.* **2017**, *74*, 130–138. [CrossRef]
22. Liang, K.; Zhang, Y.; He, Y.; Zhou, Y.; Tan, W.; Li, X. Online Behavior Analysis-Based Student Profile for Intelligent E-Learning. *J. Electr. Comput. Eng.* **2017**. [CrossRef]
23. Chen, H.; Dai, Y.; Han, D.; Feng, Y.; Huang, H. A study of learner's portrait and individualized teaching under Open Teaching. *Open Educ. Res.* **2017**, *3*, 105–112. (In Chinese)
24. Mias, G. Time Series Analysis. *Math. Bioinform.* **2018**, 329–373.
25. Mecca, G.; Papotti, P.; Santoro, D. Schema Mappings: From Data Translation to Data Cleaning. In *A Comprehensive Guide through the Italian Database Research over the Last 25 Years*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018.
26. Chowdhury, C.; Hahn, D.A.; French, M.R.; Vassermann, E.Y.; Manadhata, P.K.; Bardas, A.G. eyeDNS: Monitoring a University Campus Network. In Proceedings of the IEEE International Conference on Communications, Kansas City, MO, USA, 20–24 May 2018.
27. Yu, C.; Tian, X.; Guo, Y.; An, L. Research on User Portrait Based on Behavior-Content Fusion Model. *Libr. Inf. Work.* **2018**, *13*, 54–63.
28. Wei, H.; Zhang, F.; Yuan, N.J.; Cao, C.; Fu, H.; Xie, X.; Rui, Y.; Ma, W.-Y. Beyond the words: Predicting user personality from heterogeneous information. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, Cambridge, UK, 6–10 February 2017.
29. Jyoti, V.; Vineet, R. A Review: Salient Feature Extraction Using K-Mediods Clustering Technique. *Asian J. Comput. Sci. Inf. Technol.* **2013**, *2*.
30. He, H.; Wang, L. Static error modeling of five axis machine tool based on least square method. *Chin. J. Constr. Mach.* **2019**, *3*, 268–271. (In Chinese)
31. Chen, S.; Liu, X.; Li, B. A Cost-Sensitive Loss Function for Machine Learning. In *Database Systems for Advanced Applications*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 10829, pp. 255–268.
32. Zhang, L.; Luo, T.; Zhang, F.; Wu, Y. A Recommendation Model Based on Deep Neural Network. *IEEE Access* **2018**, *6*, 9454–9463. [CrossRef]