Research on Information Literacy of College Students Based on Data Mining

Hong Juan

Guangdong University of Science and Technology, Dongguan, Guangdong ,China

Abstract: With the rapid development of the Internet and information technology, it has brought convenience to the learning and life of college students, but it has also triggered a series of problems, such as online fraud, online gambling, and addiction to online games, online violence, and online rumors. This article uses data mining techniques and constructs data mining models to deeply explore the internet behavior of college students, timely discover abnormal internet behavior of students, analyze the reasons, and propose measures for cultivating information literacy of college students, providing reference for organizing network management activities and regulating internet behavior of college students.

Keywords: Data Analysis; Data Mining; Internet Behavior; College Students; Information Literacy

1. Introduction

With the continuous improvement of information technology, the Internet has covered all aspects of people's production and life. For college students, the internet brings convenience to their lives and studies, but some students may develop unhealthy online behavior due to their own and environmental factors. In recent years, cases of college students being deeply involved in online loans, internet addiction, and illegal activities have repeatedly sounded the alarm bell, leading to the premature withering of their lives[1,2]. Therefore, we need to provide students with information literacy, enable them to learn how to use the internet correctly, and fully leverage the advantages of online education resources. Conversely, it will lead students to lose themselves in the online world, affect normal learning, waste a lot of time, energy, financial resources, and even lose their lives. Whether from the perspectives of society, law, school, family, or individual students, analyzing and researching the online behavior of college students has practical significance.

Related concepts and significance of research

2. Related Concepts and Significance of Research

Data mining is the process of extracting knowledge, information, and patterns from large-scale datasets[3]. It is a technology that utilizes interdisciplinary knowledge such as computer science, artificial intelligence, machine learning, and statistics. Data mining technology can be applied in various fields, such as business, science, healthcare, social networks, etc., to extract valuable information for analysis such as prediction, classification, clustering, and association rule mining. In the analysis of online behavior among college students, data mining techniques can help us understand their online behavior and habits, which can be used for student behavior prediction, educational improvement, and security management[4]. Specifically, data mining can help us analyze the online behavior of students. Through data mining techniques, we can analyze the behavior patterns of students on campus networks, including frequency of use, online time, website browsing types, and so on. Predicting students' online behavior can be achieved through data mining to establish predictive models and predict their future online behavior, in order to better guide them. By detecting abnormal behavior and monitoring students' online behavior, inappropriate and abnormal behavior can be detected in a timely manner, leading to early warning and intervention to ensure campus safety[5]. And by analyzing students' online behavior, we can understand their reactions and evaluations towards different courses, thereby helping teachers improve their teaching strategies. To better understand the behavior habits and needs of students, in order to better serve them, improve their learning

outcomes and quality of life.

Information literacy, as an important literacy, has extremely high value in the field of education both domestically and internationally. This concept originated in the United States and is based on book retrieval skills. Paul Zekowski defined information literacy as the skill of using a large number of information tools and primary sources to answer questions[6]. In the field of domestic education, Professor Wang Jiqing defined it in his "Information Literacy Theory": "Information literacy is a cultivable cultivation and ability related to obtaining, analyzing, and utilizing information in the information society.". This definition is considered an interpretation of domestic research on information literacy and lays the foundation for the development of information literacy[7]. For college students, information literacy refers to the basic abilities and cultivation that college students should possess in the process of cognition, internalization, application, and creation of information. This is a narrow definition of information literacy. In the era of rapid development of the Internet, the cultivation of information literacy among college students is particularly important to adapt to the development trend of education and ensure the effectiveness of education[8].

Information literacy plays a significant role in promoting the utilization of internet resources by college students. Information literacy can help college students understand how to effectively search and filter information on the Internet to obtain accurate, reliable, and useful information they need; Improve the ability of college students to evaluate the quality and reliability of information on the Internet, and avoid them being misled by inaccurate or deceptive information; Help college students understand how to protect their personal privacy and data security, enabling them to communicate and learn information more securely on the Internet; Cultivate the awareness of college students to follow network norms and ethics, help them better master the skills of online communication, and enhance their online literacy; Promote the innovative thinking and creative ability of college students, enabling them to use the Internet and information technology more creatively to solve practical problems; Help

college students effectively utilize online resources, improve their academic level and research ability, and enable them to have a deeper understanding of the cutting-edge trends in their field of study[9].

3. Constructing a Data Mining Model for Abnormal Student Network Behavior

Feature selection and preprocessing of abnormal online behavior among students. When conducting feature selection, the indicators of abnormal network behavior should be determined first, and features should be selected based on these indicators[10]. At the same time, it is necessary to consider the correlation between features, avoid selecting highly correlated features, avoid introducing redundant information, and perform classification preprocessing on the obtained data.

Establish a classifier. When establishing a classifier, supervised learning methods can be used for training. Common classifiers include decision trees, naive Bayes, support vector machines, neural networks, etc. It should be noted that the number and quality of training samples have a significant impact on the performance of the classifier, so it is necessary to sample and annotate the data reasonably.

The decision tree algorithm can classify students' online behavior based on different characteristics. The following is a virtual decision tree algorithm formula:

```
if the internet duration is>4 hours then
if the internet frequency is>20 times/day then
if online location==on campus then
return "High risk"
else
if social network usage==frequent then
return "High risk"
else
return "Medium risk"
else
if online content==adult content then
return "High risk"
else
if game usage==frequent then
return "Medium risk"
else
return "Low risk"
else
if online location==on campus then
if social network usage==frequent then
```

Table1. Internet Abnormal Behavior Indicator Form

Indicator Name	Indicator Definition	Data Source
Online Duration	Daily time spent online	Self-reported data, Network monitoring system
Online Frequency	Number of times accessed the Internet per day	Self-reported data, Network monitoring system
Online Location	Location where the Internet is accessed	Network monitoring system, GPS positioning
Online Content	Content accessed online	Network monitoring system, Network logs, Self-reported data
Social Network Usage	Frequency, time and content of social network usage	Self-reported data, Network monitoring system, Social network data
Game Usage	Frequency, time and content of game usage	Self-reported data, Network monitoring system, Game platform data
Malware Usage	Usage of malware, including download and installation	Network monitoring system, Antivirus software logs
Online Transaction	Online shopping and transaction behavior	Self-reported data, Network monitoring system, E-commerce platform data
Online Scam Behavior	Visits to websites related to online scams, communication of information	Network monitoring system, Network logs, Self-reported data
Online Violence Behavior	Visits to websites related to online violence, communication of information	Network monitoring system, Network logs, Self-reported data

Establish a classifier. When establishing a classifier, supervised learning methods can be used for training. Common classifiers include decision trees, naive Bayes, support vector machines, neural networks, etc. It should be noted that the number and quality of training samples have a significant impact on the performance of the classifier, so it is necessary to sample and annotate the data reasonably.

The decision tree algorithm can classify students' online behavior based on different characteristics. The following is a virtual decision tree algorithm formula:

if the internet duration is>4 hours then
if the internet frequency is>20 times/day then
if online location==on campus then
return "High risk"
else
if social network usage==frequent then
return "High risk"
else
return "Medium risk"
else
if online content==adult content then
return "High risk"
else
if game usage==frequent then
return "Medium risk"
else
return "Low risk"
else

if online location==on campus then
if social network usage==frequent then
return "Medium risk"
else
return "Low risk"
else
if network transaction behavior==frequent then
return "Medium risk"
else
if online fraud==frequent or violent online behavior==frequent then
return "High risk"
else
return "Low risk"

Based on this decision tree algorithm, we can predict students' online behavior and classify it as low-risk, medium risk, or high-risk. For example, if a student spends more than 4 hours online per day, uses the internet more than 20 times per day, and is online on campus with frequent use of social networks, it is classified as high-risk.

Model evaluation. When evaluating a model, techniques such as cross validation should be used to evaluate the model and test its predictive accuracy and stability. Common evaluation indicators include accuracy, recall, F1 value, ROC curve, etc. If the model performs poorly, optimization methods such as adjusting model parameters, changing feature

selection methods, and increasing training samples can be considered.

In summary, when constructing a data mining model for student abnormal network behavior, it is necessary to comprehensively consider multiple aspects such as data features, algorithm selection, and model evaluation to improve the predictive ability and reliability of the model. Meanwhile, in practical applications, continuous optimization and improvement are needed based on specific situations.

4. Strategies for Cultivating Information Literacy Among College Students

4.1 Create a Good Atmosphere for Cultivation

Parents should neither be complacent and conservative due to computers and ubiquitous networks, nor should they relax their children's education and become addicted to the internet due to the development of information. Parents should constantly recharge themselves, have the ability for lifelong learning, keep up with the pace of digital society development, keep their thoughts in the same direction and pace with their younger generation, and provide information literacy education to their children and younger generations in a harmonious family atmosphere. The cultivation of information literacy through social synergy requires the participation of the entire society to create a positive and healthy social atmosphere. Strictly implement the norms of online civilized behavior, and utilize the power of social public opinion to create a good and healthy moral atmosphere. Maintain the dominant position of values and ensure that the online order operates within a reasonable range[11]. Promote positive social energy and implement full coverage of supervision over campus networks and local area networks. The existing network legislation should be supplemented, modified, and improved in a targeted manner to ensure the applicability of network laws and keep up with the times. Special laws targeting college students will be introduced in zones and stages, which are in line with the characteristics of students, in line with the laws of network information development, strengthen law enforcement efforts, and comprehensively serve the improvement and cultivation of college

students' information literacy.

4.2 Improve Campus Informatization Construction

In the context of digital education, it is very important to improve campus information construction. With the continuous development of information technology, schools need to keep up with the pace of the times and carry out various information technology projects, such as smart campuses, micro courses, MOOCs, academic management systems, etc., to build campuses into information centers. Through information technology, schools can manage teaching more efficiently, achieve the sharing of teaching resources, and improve the efficiency of teacher-student interaction. In addition, schools also need to establish a sound information management system, create a healthy learning environment, and cultivate students with a correct view of information. Schools can promote information knowledge through various channels, such as Weibo, WeChat, expert lectures, and competitions organized on National Security Education Day, to gradually help students master correct information acquisition and usage skills, as well as awareness of information security and protection. In student management, schools can organize themed class meetings, speech competitions, and other activities to help students understand the importance of information literacy, master the basic operating methods and skills of information technology, and cultivate their awareness and initiative in information literacy. In addition, university libraries also need to strengthen information technology construction, increase multimedia reading rooms, improve students' information knowledge, skills, and ethical cultivation level, so that students can continuously improve their information literacy in the process of reading and learning. In short, strengthening campus informatization construction and cultivating students' information literacy is an important measure to improve the comprehensive strength of schools and the quality of talent cultivation.

4.3 Guide Students to Have Awareness of Obtaining Information

To effectively improve the information literacy of college students, the first step is to guide them to have awareness of obtaining

information. From a person's sensitivity to information, one can see the strength of their information awareness. If a student has a good awareness of acquiring information, in the era of big data, they can have a keen ability to capture information and effectively assist in learning. In most universities, there are courses on information literacy for college students, which can be used as a medium to cultivate their information awareness and gradually strengthen their information literacy abilities. In addition, universities can carry out a series of distinctive educational activities on the cultivation of information literacy. For example, regularly holding special academic lectures, relying on the latest literature published on the library website, teaching college students some information retrieval tips, and exercising their ability to use information technology.

4.4 Utilize Classroom Teaching to Enrich the Information Literacy Knowledge of College Students

In the era of big data, how university teachers enrich the information literacy knowledge of college students in classroom teaching is a difficult point in the current process of university education and teaching. Information knowledge, literally interpreted, refers to modern information technology and some knowledge related to analysis and processing of modern information technology. In short, the prerequisite for cultivating information literacy is a certain mastery of information knowledge theory. The author believes that universities should provide students with corresponding information and knowledge based on their majors to promote their professional learning. So, in order to enrich the knowledge of these college students in the field of information, universities can regularly organize students to participate in professional information knowledge training and hold diverse forms of specialized academic lectures. When explaining professional databases to students, it is necessary to conduct in-depth mining based on their different majors, in order to maximize the utilization of university information resources.

4.5 Intensify Training on Information Skills for Students

The cultivation of information literacy is mainly aimed at enabling college students to master information skills and improve their level of application of information technology. Among them, it involves multiple modules such as information processing, information acquisition, and information utilization. In today's data age, college students must ensure the timeliness, accuracy, and relevance of information acquisition, and be able to independently complete tasks such as inference, comparison, analysis, summarization, evaluation, and induction to obtain information. Therefore, universities should increase the training of information skills for students, organize them to participate in various information literacy competitions, and provide targeted training before the competitions.

5. Conclusion

With the rapid development of the Internet and information technology, digitalization of education has arrived, and the cultivation of information literacy among college students has become increasingly important. In the outline of the medium and long-term education reform and development plan, the country emphasizes that "comprehensively implementing the reform and development of quality education" is the strategic theme, and information literacy education is an important way to solve the current difficulties in education development. Information literacy education is an important opportunity and force to assist in the current reform of higher education curriculum and teaching, and further improve the quality of talent cultivation. Therefore, cultivating the information literacy of college students has become a requirement of the times for quality education in universities. Cultivating the information literacy of college students, especially improving their learning and innovation abilities in an information-based and networked environment, is an important direction of quality education in universities.

and Analysis"(Code:23JDSZ3152) rescarch results.

References

[1] Huo Lijuan, "On the Philosophy,Framework and Strategies of Vocational Education Learning Space Design in the Era of Digital Transformation",Vocational and Technical Education,2021,42(10):25-31.

[2] Xiang Jie, Zhang Bin, "Research on Strategies for Improving Digital Literacy of Higher Vocational Normal College Students Based on SPOC", Journal of Kaifeng Vocational College of Culture and Arts, 2020,40(01):167-168.

[3] Chen Chen, "Research on the Curriculum Evaluation System of Civil Aviation Safety Technology Management in Higher Vocational Education", Invention and Innovation (Vocational Education), 2021,8(05):114-115.

[4] Yi GaiZhen, "Visual Analysis of Internet Behavior Data of College Students Based on SAS", Modernization of Education, 2017,4(15):154-156.

[5] Hou Qifeng, Yang Hao, Chen Lanlan, "A Study on the Relationship between Internet Addiction and Self Efficacy of Internet Refusal among College Students", Journal of Zhejiang University (Science Edition), 2013,40(4):483-488.

[6] Long Jianyi, Wang Kunzhong, "A Study on the Addiction Behavior of College Students to Mobile Internet", Heihe Journal, 2011(8):197-198.

[7] Wang Ningbang, "Analysis and Response to Network Public Opinion in the Era of Self media in Universities under the Background of Big Data", Journal of Yunnan Agricultural University (Social Sciences), 2021,15(1):134-143.

[8] Yang Tingwei, "Reflections on the Characteristics of Internet Behavior among College Students and Countermeasures for Ideological and Political Education", Scientific consulting (technology management), 2016(5):81:82.

[9] Yang Shuang, Zhou Zhiqiang, "Research on the Construction of Digital Literacy Evaluation Indicators for College Teachers", Modern Information, 2019.

[10] Xu Huan, Wen Shangyi, "A Review of the Development of Digital Literacy Training Models in the United States, Europe, Japan, and China", Library and Information Work, 2017,61(16):98-106.

[11] Ma Teng, Sun Ling, "Research on the Evaluation of Data Literacy of College Students from the Perspective of Information Ecology", Information Science, 2019,37(08):120-126.