ID3 algorithm-based research on college students' mobile game preferences and analysis of circumvention paths

Ren Yaxin*, Wang Shaoyu, Luo Yiting, Chen Shuyu

School of Management, Jiangsu University, 212013, Zhenjiang
* Email: 1679976711@qq.com

ABSTRACT. *Research on ID3 algorithm-based mobile game preference research and circumvention path analysis for the college students, consult relevant literature for in-depth understanding of classic algorithms and learn effective evasion path methods; by extracting data from college students 'questionnaire surveys and establishing attribute analysis models, use The data mining method mines the preference attributes of the game, establishes the ID3 decision tree model with the psychological state, and finally forms the decision rules, which aims to analyze the strongly related game attributes that affect the psychological state of the college students, and uses the avoidance path method to circumvent the relevant attributes of the game Research results to provide game application management decisions to minimize the adverse effects of games on college students.*

KEYWORDS: *ID3 algorithm, mobile game, circumvention path*

1. Overview

With the continuous development of information technology and the Internet, smart phones, as one of the most influential communication devices, have been widely used in all aspects of life. Mobile learning, mobile Internet, mobile shopping, mobile payment, mobile games are convenient and fast. The mobile game mode is increasingly favored by many players [1]. As a college student, it is more favored for mobile games, and playing games using mobile phones has become the main purpose of college students using mobile phones. Reasonable use of games can make people relax and be happy. However, excessive use of mobile games will cause some negative effects on the mental health of college students. In this form, how to effectively enable college students to use their spare time to reasonably treat mobile games to make them more conducive to physical and mental development is a problem we urgently need to solve. The application of data mining technology can analyze each data to find its law from a large amount of data, so that colleges and universities can find an effective evasion path. Effectively mine the mobile game

preference data of college students, analyze the data through the relevant decision tree algorithm, and grasp the rules and relationships existing in the data to provide a basis for colleges to deal with this problem reasonably, accurately and efficiently.

In this context, colleges and universities need to choose appropriate and scientific data mining techniques to grasp the connection between mobile games and mental states, and thus provide more reliable help and reference for college students to treat mobile games correctly.

2. Data mining

(1) Data mining technology

The essence of data mining is to conduct algorithm operations from a large number of noisy, inaccurate, vague, and actual business data, and finally find out that the data knowledge that has not yet been recognized or cannot be clearly recognized, and has certain practical meaning the process of. The complete data mining process should consist of six steps: data collection, data selection, preprocessing of data, establishment of data mining models, use of established data mining models, data mining operations on target data, and data Mining the results to explain and express the work [2]. Decision Tree (Decision Tree) is one of the most important and commonly used methods in data mining, mainly used for classification and prediction in data mining. Decision tree is a way of presenting knowledge. The path from the item point to each node in the decision tree is a classification rule [3].

(2) Introduction to ID3 algorithm

ID3 decision tree algorithm is an important inductive learning method, which is often used for learning and prediction based on a large number of examples. The ID3 algorithm uses the concept of entropy in information theory as a base station and builds a decision tree by selecting attributes with high information content [4].

The basic idea is: the ID3 algorithm calculates the information gain of each attribute, thereby selecting the test attribute, and using this attribute to create a node, the test attribute selects the attribute with the highest information gain, and establishes each value of the attribute A branch, and then divide the sample according to the attribute value. It starts from the root node and recursively builds the decision tree from top to bottom [5]. The process is as follows:

The first step: define the target attribute of the data set calculation.

Step 2: Calculate the information entropy and GINI coefficient for each attribute in the reference attribute list in the data set, and calculate the arithmetic average of GINI coefficients.

In the third step, the attributes whose GINI index is lower than the arithmetic mean are selected as candidate attribute sets.

In the fourth step, the candidate with the largest information gain value is

selected as the division attribute. If the value is larger, it means that the target attribute loses more information entropy on the reference attribute, then the attribute should be at the upper level of the decision tree.

The fifth step is to judge whether it meets the conditions for constructing a decision tree.

3. Performance analysis of ID3 algorithm in mobile game preferences

3.1 Data processing

(1) Data collection

In the early stage, through reading relevant literature and interviewing to understand the relevant factors that affect college students' mobile game preferences, and then designing a questionnaire survey and issuing it, the questionnaire has a total of 16 questions, including gender, grade, mobile game design tendency, mobile game time period, and self-control about How much money do you spend on rival games.A total of 130 pieces of data were collected and collated, and the data came from the students of Jiangsu University, which had obvious authenticity.

(2) Data cleaning

The main purpose of this process is to eliminate redundancy, errors, and noise in the data. Data cleaning is mainly to filter and remove duplicate data, supplement and improve incomplete data, and correct or delete erroneous data [6]. Questionnaire for college students According to the survey's return visit, some students did not play mobile games for 3 hours a day, and they generated noise data; some students did not know how often they played mobile games longer than expected. If they did not fill in the corresponding attributes, Partially incomplete data will appear; duplicate data is mainly information with the same attribute value; incomplete and invalid data is caused by students' inattention or other reasons, which will cause a lot of inaccurate noise data.

In order to make the collected data real and effective, four levels of "frequent", "occasional", "rarely", and "no" were set for "Do you often play longer than expected" during the questionnaire survey and return visit? Directly delete the collected noise data, incomplete data and abnormal data, including gender attributes, for example. Since the two influencing factors of the size of frustration and stress in life are similar, it can be regarded as a repeated attribute setting, so only the data collected in the questionnaire "whether to choose mobile games to vent due to frustration in life", A total of 101 valid data sets after cleaning.

(3) Data specification

After data cleaning, the data attributes include: grade, attraction of mobile games, frustration, time period of mobile games, self-confidence, self-control, investment of money, and time. In the study, the data column of "investment time" was used as the category identification attribute, and the remaining 12 data columns were used as the

decision attributes. In this paper, the data normalization process includes data discretization, data classification, and format normalization.

The final data processing result is shown in Figure 1:

	NJ	HM	CJ	PU	PK	JS	ZQ	QT	CZ	TM	ZK	TR	TE
0	4	1	1	1	0	0	0	0	2	2	2	2	1
1	1	0	1	0	1	1	0	0	3	3	3	2	1
2	4	0	0	0	0	0	1	1	1	1	1	2	0
3	1	0	1	0	1	0	1	0	3	3	3	1	0
4	1	0	0	1	0	0	0	1	2	1	2	3	1
...	...	...	...	...	...	...	...	...	...	...	...	...	...
96	4	0	0	0	0	0	0	1	1	1	3	1	0
97	4	1	1	0	0	0	0	0	1	3	3	2	1
98	2	1	0	1	0	1	0	0	1	3	3	1	0
99	3	0	0	0	0	0	0	1	3	3	1	1	1
100	2	1	0	0	0	0	0	0	1	3	2	1	1

Fig. 1 Data processing table

3.2 Build a decision tree

Calculate the information gain of each attribute, calculate the information gain of the attribute "grade (NJ)" is 0.0334, the information gain of the attribute "exquisite picture (HM)" is 0.0297, and the information of the attribute "a sense of accomplishment (CJ)" The gain is 0.0051, the information gain of the attribute "whether the classmates are playing (PU)" is 0.0062, the information gain of the attribute "sports (PK)" is 0.0228, the information gain of the attribute "role playing (JS)" is 0.0404, the attribute " The information gain of "Make Money (ZQ)" is 0.012, the information gain of the attribute "Other (QT)" is 0.0283, the information gain of the attribute "Frustration (CZ)" is 0.1465, and the information gain of the attribute "Mobile Game Time Period (TM)" The gain is 0.001, the information gain of the attribute "self-controlling force (ZK)" is 0.091, and the information gain of the attribute "investment of money (TR)" is 0.0624. The order of the obtained information gain values is CZ, ZK, TR, JS, NJ, HM, QT, PK, ZQ, PU, CJ, TM in descending order. In order to improve the purity of the decision tree, only the

attributes with large information gain of CZ, ZK, TR, JS, NJ, and HM are retained, and other attributes that have little effect on the decision result are deleted. The decision tree model constructed with Python and Graphviz software is shown in Figure 2.
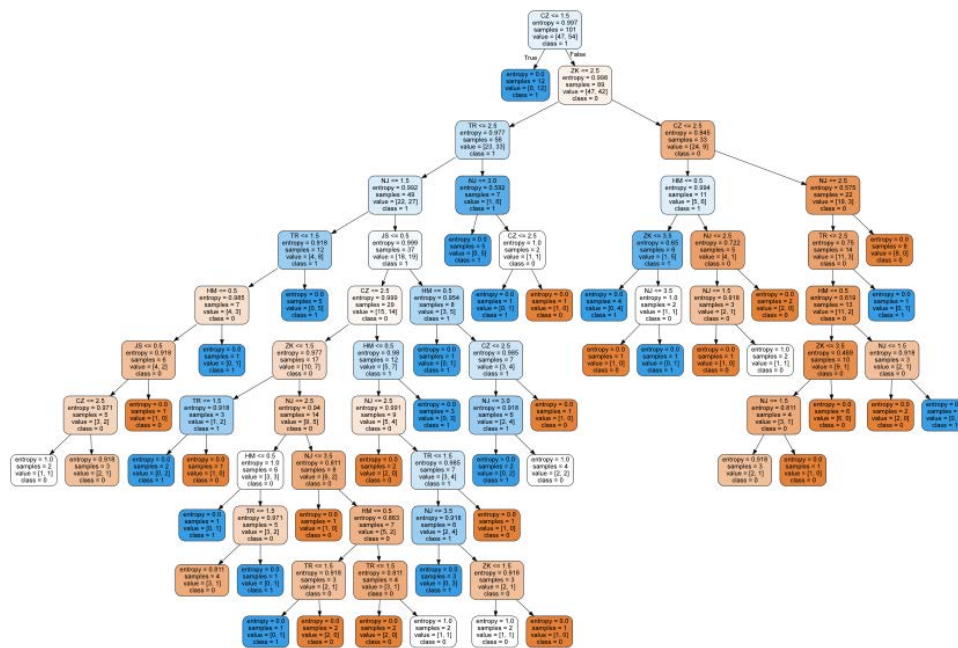

Fig.2 Decision tree

4. Data mining analysis

Extracting some rules from the decision tree for analysis is mainly to dig out the influence degree and the relationship between the selected observation points and the time spent on college students' mobile game input. Extract the classification rules from the generated decision tree. The main rules are:

If CZ <= 1.5 then mobile game time = normal;

If CZ> 1.5 and ZK <= 2.5 and TR <= 2.5 and NJ <= 1.5 and TR <= 1.5 then hand

Swimming time = abnormal;

If CZ> 1.5 and ZK <= 2.5 and TR <= 2.5 and NJ> 1.5 and JS <= 0.5 and CZ <= 2.5 then mobile game time = normal;

If CZ> 1.5 and ZK <= 2.5 and TR <= 2.5 and NJ> 1.5 and JS> 0.5 and HM <= 0.5 then mobile game time = abnormal;

If CZ> 1.5 and ZK <= 2.5 and TR <= 2.5 and NJ <= 3.0 then mobile game time = abnormal;

If CZ> 1.5 and ZK <= 2.5 and TR <= 2.5 and NJ> 3.0 and CZ <= 2.5 then mobile game time = normal;

If CZ> 1.5 and ZK> 2.5 and CZ <= 2.5 and HM <= 0.5 and ZK <= 3.5 then mobile game time = abnormal;

If CZ> 1.5 and ZK> 2.5 and CZ <= 2.5 and HM <= 0.5 and ZK> 3.5 and NJ <= 3.5 then mobile game time = normal;

If CZ> 1.5 and ZK> 2.5 and CZ <= 2.5 and HM <= 0.5 and NJ <= 2.5 then mobile game time = normal;

If CZ> 1.5 and ZK> 2.5 and CZ> 2.5 and NJ <= 2.5 and TR <= 2.5 and HM <= 0.5 and ZK <= 3.5 then mobile game time = normal;

If CZ> 1.5 and ZK> 2.5 and CZ> 2.5 and NJ <= 2.5 and TR <= 2.5 and HM <= 0.5 and NJ <= 1.5 then mobile game time = normal;

If CZ> 1.5 and ZK> 2.5 and CZ> 2.5 and NJ <= 2.5 and TR <= 2.5 and HM <= 0.5 and NJ> 1.5 then mobile game time = abnormal;

If CZ> 1.5 and ZK> 2.5 and CZ> 2.5 and NJ <= 2.5 and TR> 2.5 then mobile game time = abnormal;

If CZ> 1.5 and ZK> 2.5 and CZ> 2.5 and NJ> 2.5 then mobile game time = normal;

The above results are analyzed: under normal circumstances, the frustration degree has the greatest influence on the time invested in mobile games. The smaller the frustration degree, the higher the probability of not being addicted to games; In the case of low self-control in games, age is the main influencing factor; in addition, the higher the student's investment in mobile games, the greater the chance of indulging in games.

5. Conclusion

In response to the occurrence of the above phenomena, schools can carry out targeted development based on this. Carry out timely psychological health counseling and prevention for students, pay attention to students' attitudes towards frustration, self-control ability of junior students, and the amount of money invested in mobile games, and take effective measures, such as pushing more mobile games on the public account. Related articles and timely psychological counseling. These measures have great practical significance for the future development of students' physical and mental health.

Acknowledgement

References

[1] Sun Lisha, Song You. Analysis of the Impact of Online Games on Interpersonal Communication of College Students [J]. Journal of Economic Research, 2016 (9): 145-146.

[2] Wang Hai, "Research and Implementation of Multidimensional Data Clustering Algorithm": [Master Thesis], Shenyang, Shenyang University of Technology, 2004.

[3] Zhang Xiao. Research on decision tree ID3 classification algorithm [D]. Zhejiang: Zhejiang University of Technology, 2014. [1] Zhang Xiao. Research on decision tree ID3 classification algorithm [D]. Zhejiang: Zhejiang University of Technology, 2014.

[4] Sheng Wenshun, Sun Yanwen. An improved ID3 decision algorithm and its application [J]. Computer and Digital Engineering, 2019, 47 (12): 2943-2945. DOI: 10. 3969 / j. -9722. 2019. 12. 001.

[5] Wang Yingyun. Student score mining and analysis based on decision tree algorithm [D]. Anhui: Anhui University, 2012. DOI: 10.7666 / d.y2114137.

[6] Meng Longmei. Research on data cleaning in the process of data mining [J]. Journal of Tonghua Teachers College, 2015.36 (04): 7-10.