Construction of Big Data Analysis Model for High School Students' Career Planning Interest Events Based on Weight Algorithm

Xin Zhao[1*], Lunjun Cheng[2]

[1]Dean's Office Undergraduate admissions office, University of Electronic Science and Technology of China, Chengdu
[2]College Entrance Examination Planning Research Institute

zx@uestc.edu.cn, 934621273@qq.com

Abstract. Interest is one of the most important factors in senior high school students' career planning, which is related to the accuracy of senior high school students' career planning conclusions. At present, Holland Interest Assessment is the most commonly used tool for senior high school students' interests. However, senior high school students are often difficult to devote themselves to the assessment, which often leads to some deviation in the results. Based on this, this paper uses the principle of Xiangyang's career interest event analysis method for reference, introduces the structure of weight algorithm and big data analysis model, and constructs a big data analysis model of high school students' career planning interest events, which provides an effective reference for analyzing interest factors in the process of high school students' career planning.

Keywords: Career Planning, Interest, Weight Algorithm, Big Data Analysis Model

1 Introduction

Interest is one of the most important factors in senior high school students' career planning, which is related to the accuracy of senior high school students' career planning conclusions. At present, Holland Interest Assessment is the most commonly used tool for senior high school students' interests. However, senior high school students are often difficult to devote themselves to the assessment, which often leads to some deviation in the results. Based on this, this paper uses the principle of Xiangyang's career interest event analysis method for reference, introduces the structure of weight algorithm and big data analysis model, and constructs a big data analysis model of high school students' career planning interest events, which provides an effective reference for analyzing interest factors in the process of high school students' career planning.

H. Wu et al. (Eds.): CSIEDE 2022, ACSR 103, pp. 526–532, 2022.
https://doi.org/10.2991/978-94-6463-108-1_60

2 Interest Event Analysis Method

In the work of career planning for senior high school students, the core problem is to help senior high school students scientifically and reasonably determine their target careers according to their personal conditions, and then to launch target majors and colleges. In terms of determining the target occupation, interest is relatively easy to be found by high school students, parents and middle schools. Therefore, it is necessary to use the "interest event analysis method" for the mining of high school students' interests. The interest event analysis method is a simple and quick method of career positioning created by Xiang Yang. It identifies a person's interest events through three dimensions of activity duration, frequency and involvement, and then analyzes the type of interest and career positioning.

3 High School Students' Interest Weight Algorithm

The weight reflects the probability that the items accessed by senior high school students may be accessed by users in all events. The higher the weight is, the higher the senior high school students' interest in the event is, and vice versa. According to the size of the weight, it can conveniently and effectively reflect the degree of interest of high school students [1]. This paper mainly divides the interest of high school students into two aspects:(1) Explicit weight and explicit information are mainly reflected in the interest tags selected by high school students;(2) Implicit weight, implicit information is mainly shown in the high school students' access time and frequency to a tag. Through the calculation of the two aspects, the weight of each project visited by senior high school students is obtained, and the interest degree of senior high school students is calculated comprehensively.

3.1 Explicit Weight Calculation for Senior High School Students

Explicit weight is mainly reflected by events marked by senior high school students themselves [2]. In the project system, there are four relationships among senior high school students, events, weights and scenarios, which can be expressed as<Hi, Ei, Wi, Si>, which means that the weight of the project Ei of senior high school students Hi under the scenario Si is Wi.

(1) Build a project situation matrix to reflect the situation that the same event has been marked many times. The matrix is

$$H_{is} = \begin{cases} 0 \\ 1 \end{cases} \tag{1}$$

Where: H_{is} refers to the annotation of event i by senior high school students in situation s. The corresponding event element value after annotation is 1, otherwise it is 0.

(2) The proportion of the frequency of each event labeled in all situations in all labeled events of the high school student reflects the importance of the event, that is, the weight of the event, that is

$$F_{ih} = i_h / \sum_i^n i_h \tag{2}$$

Where: F_{ih} represents the explicit weight of high school student h to event i; The numerator i_h indicates that senior high school student h marks event i in all situations; The denominator $\sum_i^n i_h$ represents all the dimensions made by the user.

3.2 Implicit weight algorithm for senior high school students

The implicit weight of senior high school students is mainly calculated through the historical information such as the records and behavior information of senior high school students' participation in events [3]. Therefore, the implicit weight calculation is mainly divided into high school students' stay situation item, visit event frequency and timeliness factor. After considering the situation that high school students' interests will change with time, time weight item is added [4].

(1) High school students stay in media items, which can make the recommended information closer to the needs of high school students under the premise of different situations. Therefore, the frequency of visiting a situation reflects the degree of interest of senior high school students in a certain medium.

$$R(h_s) = \sum h_s / \sum_{s=1}^N h_s \tag{3}$$

Where: $R(h_s)$ represents the proportion of high school students' access frequency to media in all media; The numerator represents all the visits of senior high school student h to the media s; The denominator indicates the number of times high school student h has visited all media.

(2) The frequency of interview events can accurately show the items that senior high school students visit most and less frequently. The proportion of the frequency of visiting a certain event in the frequency of all interview events can fully show the interest of senior high school students in the event. The larger the proportion, the greater the interest. On the contrary, the lower the interest.

$$R(h_{si}) = \sum h_{si} / \sum_{s=1}^N h_{si} \tag{4}$$

Where:

$R(h_{si})$ represents the proportion of the frequency of high school student h's access to event i in all interviews under the s situation; The numerator represents the number of times senior high school student h visited event i in the context of s; The denominator indicates the number of times senior high school student h has visited all events under the s situation.

(3) The timeliness factor is a special item introduced in the context of protecting the privacy of high school students [5]. In order to protect the privacy of high school students, this paper does not use high school students' participation records but their access records to events in the process of data mining.

$$\lambda = \begin{cases} 0, \Delta t < 0.5 \\ \frac{\Delta t}{5}, 0.5 \leq \Delta t < 5 \\ 1, \Delta t \geq 5 \end{cases} \tag{5}$$

Where λ represents the aging factor. As for the relationship between the timeliness factor and the degree of interest of senior high school students, when senior high school students stay less than 0.5h before an event, it is considered that senior high school students are not interested in the event, which is recorded as 0; when the stay time is 0.5h-5h, it is considered that senior high school students are more interested in this kind of time, whose degree of interest is $\Delta t / 5$, and Δt is the specific time that senior high school students stay (unit: h); When the senior high school students stay for more than 5 hours, it is considered that the senior high school students are very interested in this kind of time, and it is recorded as 1.

(4) The time weight is considered from the decay of information value of high school students. The process of information forgetting is regarded as the decay of information value, that is, the degree of information forgetting represents the reference value of information for current recommendations. Therefore, relevant scholars propose to use the time weighted function method to keep its value between (0,1].

$$f(t) = e^{\lambda - t} \tag{6}$$

$$\lambda = \frac{\ln 0.5}{T}$$

Where: f (t) represents time weight; e is the natural logarithm; , T is the half life of information; T refers to the length of the last interview of senior high school students, that is, $t = t_{now} - t_{h.i}$, $t_{u.i}$ refers to the last scoring time of senior high school students.

The timeliness factor fully takes into account the influence of the dynamic attenuation of high school students' interest degree over time, and effectively improves the accuracy of implicit weight.

To sum up, the recursive formula of user implicit weight P_{ih} is:

$$P_{ih} = R(h_s)R(h_{si})\lambda t f(t) \tag{7}$$

Explicit weight reflects the needs of high school students themselves and is a clear choice made by high school students themselves. However, with the participation of high school students, through data mining and algorithm calculation, the implicit interest of high school students is gradually discovered. Both of these two aspects are indispensable. They are the key indexes indicating the high school students' interest in an event. Therefore, the parameter ∂ is introduced.

$$W_{ih} = \partial F_{ih} + (1 - \partial) P_{ih} \tag{8}$$

When the explicit information of senior high school students can fully show their interests and when the participation records of senior high school students are insufficient to calculate the implicit weight, the value of parameter ∂ is larger.

4 Construction of Big Data Analysis Model for Interest Events

On the basis of establishing the interest weight algorithm for senior high school students, this paper uses the framework of big data analysis model to build an interest event analysis model. The big data analysis model of interest event analysis model mainly includes three parts: high school students' behavior data source, HDFS distributed storage and management, and MapReduce parallel computing. The framework and specific content of high school students' behavior big data analysis model are shown in Figure 1:

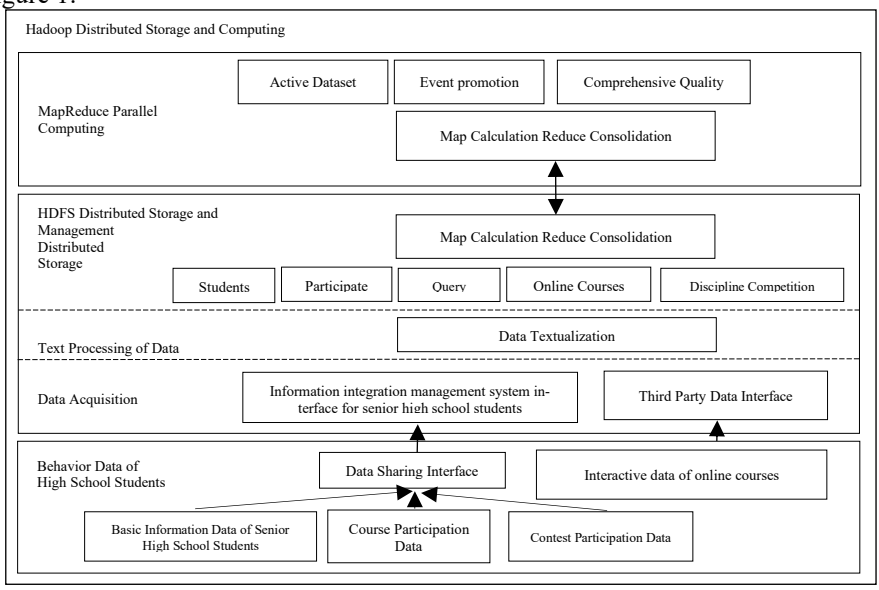

Fig. 1. Big data analysis model of high school students' behavior [Owner-draw]

In order to build a scientific and reasonable model, this paper collected a large number of high school students' data as the basis for analysis. The types and contents of relevant big data are shown in Table 1:

Table 1. Types and Contents of High School Students' Behavior Big Data [Owner-draw]

Data Type	Data content
Basic Information Data of Senior High School Students	Name, Gender, Student Number, Grade

Course Participation Data	Student Number, Course Name, Time of Participation, Completion Time and Times of Participation

After the classification and sorting of big data types are completed, HDFS distributed storage and management can be established to complete the collection, integration, data storage and management of high school students' behavior data (the process is shown in Figure 2), which also completes the construction of high school students' career planning interest event big data analysis model.

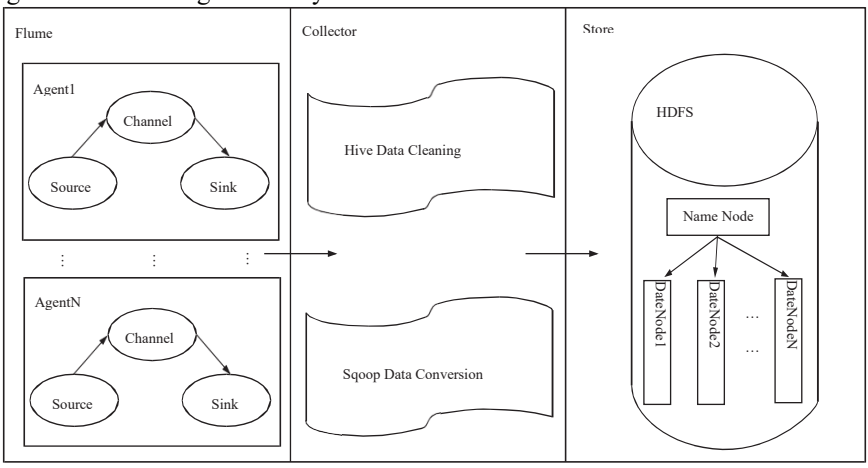

Fig. 2. HDFS distributed storage and management [Owner-draw]

5 Conclusion

This paper uses the principle of interest event analysis method for reference, calculates the interest weight of high school students, and analyzes it from three aspects: duration, frequency and involvement. At the same time, based on the big data analysis model, this paper constructs the big data analysis model of interest events, which provides a certain reference for the analysis of interest factors in the career planning process of high school students.

References

1. Huang Chen, Shao Xiaoming. Construction of Unbalanced Engineering Cost Management and Multivariate Statistics System Based on Random Matrix Weight Algorithm[J]. Mathematical Problems in Engineering,2022,2022.
2. Jia Guangyu, Lam Hak-Keung, Althoefer Kaspar. Variable weight algorithm for convolutional neural networks and its applications to classification of seizure phases and types[J]. Pattern Recognition,2022,121.

3. Akanksha Garg, Deepak Murugan, Dharmendra Singh. Development of Optimal Weight Algorithm for Efficient Application of Dual Tree Complex Wavelet Transform for Resolution Enhancement of Satellite Images[J]. Current Science,2019,117(12).
4. Weisi Dai. Improvement and Implementation of Feature Weighting Algorithm TF-IDF in Text Classification[C]//.Proceedings of 2018 International Conference on Network, Communication, Computer Engineering (NCCE 2018).,2018:598-602.
5. Tsung-Ching Lin, Hung-Peng Lee,Hsin-Chiu Chang, Trieu-Kien Truong. A cyclic weight algorithm of decoding the (47, 24, 11) quadratic residue code.[J]. Inf. Sci.,2012,197.