Research on Distributed Software of College Online Education Based on Information Gain Optimization Algorithm

Lihua Chen
Quanzhou Arts and Crafts Vocational College
Quanzhou, Fujian 362500, China

Abstract:The paper analyzes the shortcomings of the information gain method, applies frequency, concentration and dispersion to the information gain method, and proposes a feature optimization selection method based on information gain. By introducing the service resource framework WSRF technical specification, a new grid distance education system model GEM based on the open grid service structure OGSA. According to the existing problems of modern distance education, a distributed distance education system based on object Web computing is studied by using CORBA/Java. On this basis, a new dynamic distributed system structure technology and idea based on object-oriented technology is applied to develop and realize the distance education question bank system.

Keywords: Distributed Software, College Online Education, Information Gain Optimization Algorithm

1. INTRODUCTION

Object-oriented technology represents a brand-new programming idea and method of observing, expressing and dealing with problems [1]. It is different from the traditional process-oriented development method. Object-oriented programming and problem solving strive to conform to people's daily natural thinking habits. In response to the above problems [2], combined with the characteristics of distance education, the thesis starts from the video streaming channel scheduling and distributed system perspective, and studies a distributed video-on-demand system suitable for large-scale user access in distance education [3]. To this end, the research starts from the actual application of VOD, the fifth stage, the focus of which is to establish an online automatic response system [4].

In a word, distance education in the 21st century is aimed at open, flexible and lifelong educational development. It is the continuation of traditional education, and at the same time, it is [5] also a huge change to traditional education. Some continuous attributes in the monitoring data of new energy smart vehicles the accuracy of the value of is too high, which will cause [6] the subsequent mining algorithm to take up too much space and time, and the trained model is prone to overfitting. The purpose of data discretization is to convert continuous attribute values into discrete interval values [7]. The basic algorithm of decision tree is greedy algorithm, which uses top-down recursion to construct decision tree. Ideally, when all the leaves are is a pure node, that is, when the instances of each leaf node belong to the same class, the decision tree stops growing [8].

Feature selection is mainly to select a set of features that are most effective for classification from the original feature space [9], which can effectively reduce the dimension of the feature vector space of the text, delete redundant features [10], and reduce the interference of irrelevant information on the text information processing process [11]. Its theory is clear, the method is simple, and the learning ability is strong. It is suitable for dealing with large-scale learning problems [12]. It has solved many problems in practical applications, especially

for non-incremental learning tasks [13]. Good choice. Grid computing is a support platform for distributed and parallel computing, a seamless, integrated computing and collaborative environment [14].

It can be used as a virtual whole in geographically dispersed computing resources [15]. The grid computing system based on it not only enables people to gather dispersed computing resources. The global network is that all devices and software connected online become a vast resource shared by the world [16]. The environment has also developed from a centralized to a distributed open system, so that users can transparently apply different models made by different manufacturers [17]. At the same time, the key technologies such as resource scheduling, video storage and program distribution in the VOD system are in-depth Research, and put forward the following improvements and innovations [18]: the use of distributed on-demand system to replace the traditional centralized on-demand system. Modern distance education is just a simple extension of traditional classroom education on the Internet [19].

Judging from most of the so-called "distance education" at present, the teaching form and content are no different from traditional classroom education [20]. At present, the commonly used supervised discretization algorithms can be divided into: discretization algorithms based on statistics, discretization algorithms [21] based on class and attribute correlation, and discretization algorithms based on information entropy [3]. Among them, the discretization algorithm based on statistics. At present [22], the more commonly used processing method for this situation is to use the majority voting method to determine the classification of this sample, that is, to convert the node in the decision tree into a leaf node and use the majority of the samples. The class it is in marks it. Not enough attention is paid to the word frequency of the feature item [23].

The second: wrongly increases the weight of feature items that appear infrequently in one category but frequently appear in other categories [24]. Aiming at these shortcomings, a feature optimization selection method based on information

gain is proposed. That is, the optimization algorithm of twice information gain. Whenever a new attribute is selected, the algorithm not only [25] considers the information gain brought by this attribute, but also Consider the information gain brought by the attributes that continue to be selected after selecting this attribute, that is, consider the two-level nodes of the tree at the same time. Computer network is a typical heterogeneous system. Different models, different operating systems, different computer programming languages, application software running on various models and operating systems, etc., make it very difficult to develop distributed system software.

2. THE PROPOSED METHODOLOGY

2.1 The Information Gain Optimization Algorithm

In information theory, the amount of information refers to the measure of information required to select an event from four equally possible events, event a. The amount of information can be measured by a log: P(ai), where P(aj) represents the probability of the occurrence of event ai. Information gain is an important concept in information theory and is widely used in the field of machine learning. For the classification system, calculating the information gain is for each feature item, it calculates the information of the feature item t on the category c by counting the number of documents in which a feature item t appears or not in the category C. Gain, the core of the algorithm is to select attributes at all levels of nodes in the decision tree, and use the information gain rate as the criterion for attribute selection, so that when each non-leaf node is tested, the largest category information about the tested example can be obtained. This difference indicates the amount of information provided by the attribute pair classification.

Therefore, information gain can be used to quantify the relevance of an attribute to a given class or concept, with greater information gain indicating greater relevance to the classification task. That is, the number of texts that do not contain feature t is divided by the total number of texts; p(c. Idiao means that the text does not contain feature items and belongs to the conditional probability of class q, that is, the number of texts that do not contain feature t and belong to class G is divided by The number of texts that do not contain the feature f; the book is the number of categories. The two-time information gain optimization algorithm is based on the foundation. Whenever a new attribute is selected, the algorithm not only considers the information gain brought by the attribute, but also considers the selection of the The information gain brought by the selected attribute after the attribute.

On the basis of ID3 algorithm, the performance of ID3 algorithm is improved by combining attribute reduction based on information gain and minimum distance classification method. The attribute reduction Ho based on information gain is based on the size of attribute gain and the correlation coefficient between attributes.

2.2 The Online Education In Colleges and Universities Based on Information Gain Optimization Algorithm

Current distributed computing models make the following assumptions in order to reduce their complexity. There is no concept of a network and no remote address space. All components utilize a common language - an interface language and are made independent of their specific programming language through the interface language. The same Each host in the subnet must respond to broadcasts, causing unnecessary host interruptions and wasting processor resources; in the current network with a switch-centered flat structure, the broadcast method is prone to the well-known "broadcast storm" problem. Teaching resources Low level of sharing and low quality. Since there is no accurate standard for learning resources and courseware, the quality of learning courseware published online is often low, or even a simplified version of the syllabus. The ideal selection strategy for candidate breakpoints should ensure the discrimination of the original information system. On the premise of the relationship, select the fewest candidate breakpoints.

The bottom layer of the grid is the resource layer, which is a collection of distributed resources in the grid, mainly including the computing resources, storage resources and other educational resources of various departments and institutions of the school. Since these resources belong to different colleges or institutions, various resources meet user needs in different virtual organizations according to certain sharing strategies.

2.3 The Distributed Software for Online Education in Colleges and Universities

The deeds of Tarjing are often less than the disturbance it brings. Because the value of attributes often has the same impact on the classification and attributes of instances, the existing decision tree inductive learning algorithms only pay attention to the selection of attributes, and put the value of attributes in a secondary position. It is a toolkit and version One of the core components is an information service framework based on describing the overall grid environment, and its goal is to effectively represent a large number of geographically distributed, heterogeneous and also dynamic resources and services in the grid computing environment.

3. CONCLUSION

Aiming at the defect that the algorithm proposed in the literature is not effective for multi-valued attributes, this paper proposes an information gain optimization algorithm for attribute 0-value pairs by taking advantage of its advantages for two-valued attributes. Some related key technologies involved in distributed VOD system are studied, which lays a theoretical foundation for proposing a VOD system suitable for distance education. Since the system is based on J2EE and XML platforms, the concepts of movable code, movable data and strong type robustness requirements in J2EE platform make it possible to build a truly dynamic distributed system. Instead, the minimum distance classification method is used to determine its category.

4. REFERENCES

[1]Chen Jilong, Zhang Hongyu, Zhang Congyi, et al. Design of distributed traffic real-time analysis system based on online learning [J]. Technology and Market, 2022, 29(3):2.

[2] Liu Kai, Hu Jing. Theoretical Framework of Artificial Intelligence Education Application: The Symmetry Hypothesis of Learners and Educational Resources——Interview with Professor Hu Xiangen, Expert of Intelligent Learning Guidance System [J]. Open Education Research, 2018.

[3] Li Hailong, Li Xinlei. Discussion on SPOC Experiential Learning Based on Distribution Flip in the "Post-MOOC" Era [J]. 2021(2015-11):44-50.

[4] Li Qianshu. The path analysis of blockchain technology to help the development of online education---and the practical ideas of Changzhou Open University [J]. Journal of Xiamen City Vocational College, 2019, 21(2):6.

[5] Huang Zhenyu. Research on the Construction of Distributed Laboratory in Distance Education [J]. Journal of Changsha Vocational and Technical College of Communication, 2018.

[6] Xu Yuhong. An online learning system based on distributed stream processing: CN108924581A[P]. 2018.

[7] Tu Mingjiang, Liu Yibing, Wu Nanzhong. Distributed teaching based on VR: Theoretical model and implementation strategy [J]. Electronic Education Research, 2021, 42(1):8.

[8] Zhang Peng. Research and Design of Distributed Management System Architecture [J]. Information Technology and Informatization, 2019(7):3.

[9] Li Shuai. Research on Microservice Architecture and Implementation of Online Education System. Beijing University of Posts and Telecommunications, 2018.

[10] Shen Weiji. Exploration on the Construction and Application of Smart Campus in the Big Data Environment [C]// Proceedings of the 2020 "Basic Education Reform and Innovation in the Internet Environment" Seminar Proceedings. 2020.

[11] Huang Zhenyu. Research on the teaching mode of the practice link of open education online courses [J]. Journal of Guangxi Radio and Television University, 2019, 30(3):3.

[12] Liu Kai, Hu Jing. Theoretical Framework of Artificial Intelligence in Education Application: Assumption of Symmetry between Learners and Educational Resources——Interview with Professor Hu Xiangen, Expert of Intelligent Guide System [J]. Open Education Research, 2018, 24(6):8.

[13] Zhang Zheyuan, Gu Xingsheng. Research on low-voltage fault ride-through of DFIG based on distributed deep neural network [J]. Journal of East China University of Science and Technology (Natural Science Edition), 2022, 48:1.

[14] Wang Wendong, Wu Haini, Hou Mian. Simulation of personalized scheduling of distributed online learning resources under the Internet of Things [J]. Computer Simulation, 2019, 36(1):5.

[15] Wang Ne, Zhou Ruiting, Li Zongpeng, et al. An online scheduling method for distributed machine learning tasks based on edge-cloud collaboration: CN112579270A[P]. 2021.

[16] Wang Yifei, Wu Siyao, Li Chunli. Research on the application of online education in colleges and universities under the concept of "Internet + education" [J]. 2021.

[17] Yong Ren Zhu. Research on the new form of classroom teaching with a mix of online and offline—taking the course "Capp Software Application" as an example [J]. Modern Education Forum, 2020, 3(1).

[18] Li Jingwei, Han Shuhe. Research on the Design Method of Online Education Platform Based on Microservices and TCC Distributed Transactions [J]. Journal of Nantong Shipping Vocational and Technical College, 2020, 19(2):4.

[19] Yang Doudou. Research on Online Consulting System Based on Web Socket and Collaborative Filtering Algorithm [D]. Hunan University, 2019.

[20] Wang Hengtao, Kong Fei, Chen Yaobin, et al. Online intelligent learning decision-making optimization method and system for gas distributed energy system: CN107844869A[P]. 2018.

[21] Zhu Xiping, Li Weiqin, Deng Kui, et al. Remote Distributed Multilingual Learning Guidance System:, CN109859543A[P]. 2019.

[22] Dong Ningning. Design and Implementation of Kubernetes-based Software Engineering Education Cloud Basic Container Platform Subsystem [D]. Nanjing University, 2019.

[23] Wei Liang. Research on data-driven online adaptive learning system in higher vocational colleges [J]. Education Observation, 2021, 10(18):5.

[24] Li Guangxia, Shen Yulong, Liu Jia. IoT data analysis method using distributed asynchronous update online machine learning:, CN110175680A[P]. 2019.

[25] Qiao Zhi. The development and enlightenment of regional open online courses [J]. Educational Information Technology, 2018(4):4.

[26] Li Qing, Zhang Xin. Blockchain: Promoting the Openness and Credibility of Education with Technology [J]. 2021(2017-1):36-44.