Check for updates

Mining Recessive Teaching Resources of University Information Based on Machine Learning

Zheng Jingya(✉) and Jichao Yan

Zhengzhou Technical College, Zhengzhou 450121, China
zhengjingya222@yeah.net

Abstract. The accuracy of mining implicit teaching resources in traditional universities is low, so a method of mining implicit teaching resources in universities based on machine learning is designed. Firstly, it designs the process of data mining, define the problem, collect and preprocess the data, execute the mining algorithm and then explain and evaluate it. The classification method of data mining is optimized. In this paper, the classification technology is neural network, and the artificial neural network unit is built by biological neuron structure, and the classification is completed by biological transfer and activation function. Finally, the machine learning algorithm is improved, and the ight is updated by introducing momentum scalar factor. In the contrast experiment, it chooses the data set and train the parameters, design the process of data mining, and count the relevant parameters of the data set. The experiment results show that the accuracy of the designed method is 4.03% higher than that of the traditional method.

Keywords: Machine learning · Informatization teaching · Recessive resources

1 Introduction

With the advent of the information age, the application of information technology has been slowly penetrated into all areas of people's lives and quietly changing the way of people's lives. In this era of knowledge explosion, it is easy for people to get lost in the huge amount of information, it is difficult to find effective information, which requires people to have a higher information literacy [1, 2]. For a long time, the cultivation of college students' information literacy mainly relies on information technology courses. Hover, the lack of information technology curriculum hours has become a major problem in the cultivation of college students' information literacy, at the same time, it is not realistic to expand information technology curriculum hours in a short time. Therefore, looking for another way - information technology hidden curriculum, to cultivate college students' information literacy is particularly important.

At present, there are two main definitions of IT implicit curriculum: (1) IT implicit curriculum may be those that appear in the formal school teaching plan under the name of other courses, but in fact carry out IT education, including those that are not included in

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2021
Published by Springer Nature Switzerland AG 2021. All Rights Reserved
W. Fu et al. (Eds.): ICMTEL 2021, LNICST 388, pp. 178–188, 2021.
https://doi.org/10.1007/978-3-030-82565-2_15

the formal teaching plan but are provided by the information environment construction of the school. (2) The hidden curriculum of information technology mainly refers to the non-public educational experience that is consciously or unconsciously conveyed or implied to students in the soft and hard environment of information technology education in the school context. It is irreplaceable and has the function of "teaching without teaching". It is an intangible asset of the school [3, 4]. In fact, the definition of these two courses is not contradictory. The former mainly discusses the process of information education inside and outside classroom teaching, while the latter focuses on the software and hardware environment of information technology. In the traditional process of mining the recessive resources of IT teaching in colleges and universities, the accuracy of the results is low, so a mining method based on machine learning is designed, it is innovation lies in the use of biological neuron structure to establish artificial neural network unit, and the use of biological transfer function and activation function to complete the classification. Its focus is to improve the machine learning algorithm and introduce momentum scalar factor. The final result is to modify the machine learning algorithm to ensure the effect of teaching resource mining.

2 Mining Recessive Teaching Resources of University Information Technology Based on Machine Learning

2.1 Design Data Mining Process

Data mining is a complete process, through continuous interaction with the user, the results of mining at different stages of continuous feedback is completed, the whole process is shown as follows (Fig. 1):

Problem definition: First, clear the actual work of data mining requirements; Second, to determine the available learning algorithm. Data preprocessing: generally includes eliminating noise, deriving calculation gap data, eliminating duplicate records and completing data type conversion. Mining algorithm execution: First, to determine the mining tasks; Second, to decide which algorithm to use [5, 6]. Interpretation and evaluation: Eliminate redundant or irrelevant patterns; for patterns that do not meet user requirements, you need to go back to the previous stage. The effectiveness of the data mining technology adopted and the quality and quantity of the data used for data mining are two main factors that affect the quality of data mining.

The core idea of this kind of method is to imitate some behavior activities of biology, and simulate these activities by computer program, mainly including neural network method class and genetic algorithm. Some researchers use genetic algorithm to optimize the search engine and resource scheduling, improve the search and sharing of network teaching resources, so that the majority of teachers and students in the teaching process to obtain quality teaching resources. In the application of RTVU pilot summing-up evaluation system, the fuzzy BP neural network method is used to develop the sub-modules of FBPNN learning, knowledge base and FBPNN reasoning machine. The fuzzy neural network module is responsible for knowledge acquisition, storage and solution.

Normalization of knowledge representation and expression transformation is the responsibility of the input/output schema transformation. In the evaluation application,

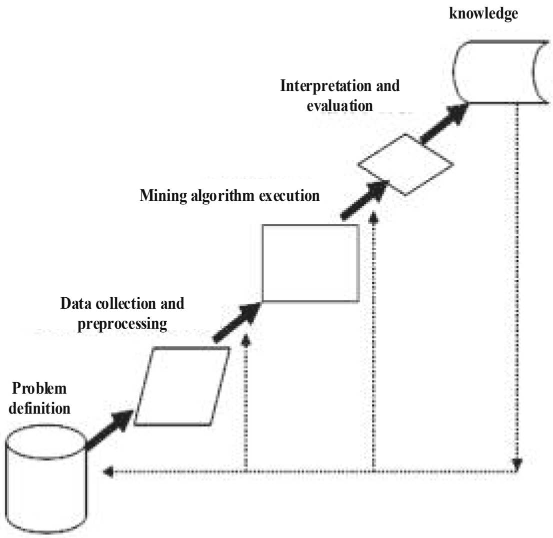

Fig. 1. Process diagram for data mining

it can better realize the summative evaluation of TV University, reduce the interference of artificial uncertainty factors, and enhance the scientific and standard [7–10]. Aiming at the advantages of neural network, the BP neural network is used in the teaching quality evaluation system of modern distance education to construct mathematical models, input different evaluation indexes, and output teaching effect, so as to evaluate teaching quality and teaching effect of modern distance education scientifically and accurately. In order to solve the problems in reality and highlight their own advantages, genetic algorithm and neural network are combined to form a novel evolutionary neural network research field, and many valuable conclusions and results are obtained.

2.2 Optimize Data Mining Classification Methods

Data mining includes a variety of analysis methods to mining data sets analysis, get patterns and apply them, of which classification is occupying a place, and classification methods have been well known. How to classify the data correctly will directly affect the accuracy of the mining results and the efficiency of mining patterns.

Applications of classification include a variety of problem areas, such as text, multimedia, social networking, and biological data. In addition, there may be different problems in many different scenarios. Classification is a fairly diverse topic, and its underlying algorithm relies heavily on data domains and problem scenarios. Classification algorithm is also one of the most important research fields. The overall goal of the data mining approach is to extract information from the information set and associate it with a comprehensive structure for future use.

Classification is a very important mining method in DM. It is a process of looking for classifiers. Objects in the dataset are assigned to different classes by some constraints.

It uses a given class tag to analyze the objects in the dataset, usually using a training set where all the objects are already associated with known class tags. The classification algorithm learns and builds a model from the training set, and then uses the model to classify new objects. In other words, it can say that classification is the process of summarizing data according to different classes. Classification techniques can handle a wider range of data and are becoming increasingly popular [11–15]. The basic principle of the artificial neural network is to imitate the structure and function of human brain, and the computer system is made up of several simple processing units connected in some way. Nerve cells constitute the basic unit of the nervous system, called biological neurons, neurons for short. Neurons consist mainly of three parts: (1) cell body, (2) axon, and (3) dendrite [16–19]. The following figure shows (Fig. 2):

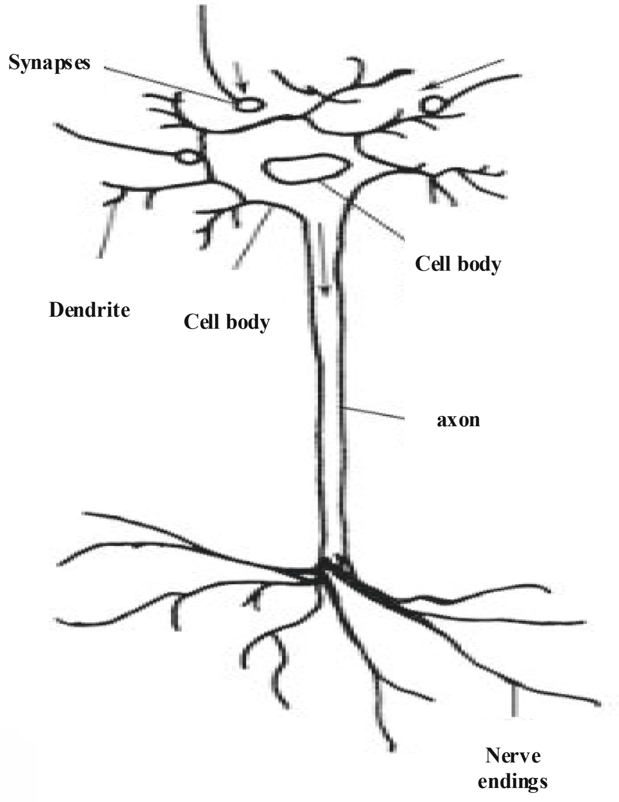

Fig. 2. Structure of biological neurons.

In the neuronal structure shown above, a neuron consists of a cell and its many processes. There is a nucleus in the cell, synaptic function is to transmit information. Several of the processes that introduce an input signal are called "dendrites," while only one of the outgoing processes is called an "axon." The process of neuron transmitting information can be regarded as a dynamic process of a nonlinear system with multiple inputs and single outputs [20–24]. The artificial neural network processing unit thus produced is shown in the following figure (Fig. 3):

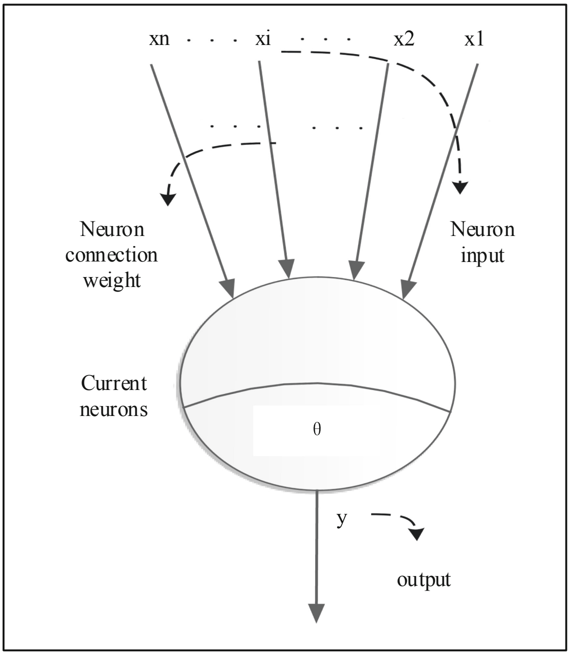

Fig. 3. Artificial neural network unit

Pictured above is a classic M-P neuron model that mimics the basic structure of biological neurons. Where,… is the input to a neuron, and for each neuron it receives input signals from multiple other neurons, and the synaptic intensities are expressed in real coefficients that represent the ighted values of the actions of each neuron on one another. The primary function of ights is to influence the input from a higher level, and the ights can be modified according to rules. All inputs to the neuron can be calculated using a summation formula, that is, the ighted sum of the input signals. The resulting value is compared to the threshold value to determine whether to output, and when the sum value is greater than the threshold value, the output signal is generated, otherwise no output [25–29]. It is the activation function of the processing unit, usually a nonlinear function, representing the threshold of the hidden layer neural node. To sum up [30], the input process of M-P neuron can be described as signal and its corresponding ighting coefficients afferent neuron processing unit, summing them up and substituting activation function to produce output and complete classification of data mining.

2.3 Improved Machine Learning Algorithm

In this paper, the design and improvement of machine learning algorithm is divided into two stages: feedforward stage and back-propagation stage. In the feedforward stage, input samples are imported from the input layer, processed layer by layer, and then transmitted to the output layer. When the actual output value of the output layer does not agree with the desired result, it is corrected by the back-propagation error. The back-propagation of error is to back-propagate the output error from the hidden layer to the input layer, and then distribute the error to all neurons passing through each layer to obtain the error signal of each layer. Then the error signal is used to correct the ight of each unit. When the output error is within the controllable range, that is, the error is less than a set threshold until the end of the cycle. In this paper, the momentum scalar factor is introduced, the ight is defined by the use of momentum term factor, and the batch learning technique is used to make the ight more accurate. The mathematical representation is as follows:

The training data X with P sample is given by the following formula:

$$X = \left\{x_p, d_p\right\} \; p = 1, 2, ..., P \tag{1}$$

Where x_p is the input vector for the sample p with n feature (or dimension):

$$x = \{+1, x_1, x_2, ..., x_n\}^T \tag{2}$$

d_p is the vector to which it is associated with the expected output:

$$d = \{d_1, d_2, ..., d_k\}^T \tag{3}$$

In the feedforward stage, for neurons J in the hidden layer, compute (P, J) the dimension input matrix u:

$$u = XV \tag{4}$$

In the above expression,X is the matrix representing the $(P, n + 1)$ dimension training data, V is the dimension ight vector matrix $(n + 1, J)$ in the hidden layer. Two differentiable activation functions are used in the algorithm: hyperbolic tangent function and logistic model function. Output matrix y of hyperbolic tangent activation function in hidden layer:

$$y = \frac{2}{\left[1 + \exp(-u)\right] - 1} \tag{5}$$

Similarly, the signal matrix of the hidden layer is calculated as the product of the derivative output matrix and the ight. By using machine learning algorithm, the network ights are updated iteratively, and the LMS algorithm is used to find the corrected ights of each ight matrix. The correction ight vector is the correlation error signal of the input layer of the product matrix multiplied by a scalar selection learning rate. In order to accelerate the learning process, this paper introduces a momentum scalar factor. In order to use them in the adaptive process of network ights, the previous batch of network ights are saved separately, so the LMS algorithm and momentum factor are used to adjust the network ights.

3 Experiment

3.1 Experimental Environment Parameter Settings

In this chapter, it use TREC conference to provide four representative AD hoc information retrieval test datasets: diskl & 2, disk4 & 5, WT10G and GOV2. The size and type of these four resource datasets are different. For example, diskl & 2 and disk4 & 5 resource datasets mainly include many teaching platforms, the quality is relatively high, while WT10G and GOV2 datasets are obtained from the Internet through the crawler, the quality of its teaching resources cannot be guaranteed.

In this experiment, only the title field is used to construct the query for retrieval. In the process of indexing and query, only two simple pretreatments are carried out: (1) using the Porter word drying tool; and (2) using the standard discontinued word table in the InQuery system to remove discontinued words. The MAP values of 1000 teaching resources are used as evaluation criteria, which is the most commonly used evaluation index in TREC evaluation.

3.2 Parameter Training

From the above, this paper can see that there are some parameters to be adjusted in different feedback models. In order to find the best parameter setting, the comparison model and the model proposed in this paper are all obtained by parameter training. This method is a common method for establishing strong baselines in information retrieval field.

The entire excavation process is shown below (Fig. 4):

In order to verify the effectiveness of the improved method, some data samples from the UCI machine learning database are used in this paper. The following table shows the descriptive information for the experimental datasets used (Table 1):

The values of the fixed parameters used in the experiment are as follows: The momentum factor is used to accelerate the convergence of the cost function to the minimum, and η_0 is a value of 0.75 is good for convergence purposes. *kw* is A value of 0.1 is used to initialize the ight vector in the hidden layer, leaving it in a small range $[-0.1, +0.1]$. The K-multiple parameter refers to the cross-validation of 10 times during training.

In the experiment, 11 samples are set up, and 10 samples are used as training set and the other one is used as testing set. The average error percentage is obtained and compared with the results of traditional mining method.

3.3 Experimental Results and Analysis

Under the above experimental conditions, for all datasets trained in the MLP structure, the parameters are constant. Using the above resource dataset to train and study the mining method of this paper and the traditional mining method, the accuracy of the results are as follows (Table 2):

In this paper, the hidden resource mining method based on machine learning information teaching has higher classification accuracy than the traditional algorithm. In the experimental data set, this method and the traditional algorithm are used for training

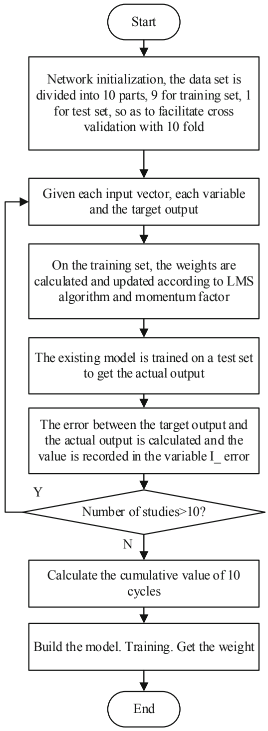

Fig. 4. Mining process

Table 1. Summary of experimental dataset information

Resource data set	Resource data scale	Number of attributes	Number of categories
Aural	160	5	3
Wine	280	8	5
Heart	297	13	5
Cancer	198	32	2
B can	232	19	2
Head	178	15	3
Arm	146	2	3
Craw	989	23	3
Liver	218	60	2
Vote	2533	21	6
Letter	378	35	4

Table 2. Comparison of accuracy of execution results

Resource data set	Accuracy of traditional mining method 1	Accuracy of traditional mining method 2	Accuracy rate of mining method in this paper
Aural	99.67	98.67	99.35
Wine	88.35	87.35	96.54
Heart	95.46	95.56	98.51
Cancer	100	100	100
B can	83.24	83.14	92.33
Head	100	100	100
Arm	87.65	87.35	94.34
Craw	98.35	98.15	98.35
Liver	62.45	62.45	77.35
Vote	97.33	97.23	97.68
Letter	91.23	91.13	95.63
Average value	91.25	91.15	95.28

and learning, respectively, so that the classifier can predict the results. Accuracy refers to the percentage of the ratio between the predicted value and the actual value (or the percentage of the ratio between the actual value and the predicted value, which is subject to the observation results and less than 1). In view of the current academic level, the teaching and research of hidden resources mining is difficult to achieve substantial upgrading, which is still the direction of follow-up action.

4 Closing Remarks

Under the impetus of global informationization, information technology has developed rapidly. Fully excavating and reasonably utilizing the recessive curriculum resources can enrich the curriculum resources and teaching contents, enhance the teaching vitality, and stimulate the students' interest and thirst for knowledge. In this context, under the guidance of the new curriculum reform theory, machine learning and data mining in the processing of mass data, make it easy to find useful information in a large number of data become a reality.

But there are still some deficiencies in this paper, the original data set in this paper is the standard operation order of order of magnitude to be unified between [0, 10], and then data mining work. Hover, there are some limitations, and the [0, 10] data range is not necessarily the most reasonable choice for different data sets. The scope of the specification should be different for different data sets, and this can be explored further.

References

1. Zhang, J.L.-L., Duan, Y.-F.: Analysis on the informatization teaching of specialized English in higher vocational colleges—taking the shipping management major as an example. J. Tianjin Vocat. Inst. **20**(10), 83–87 (2018)
2. Zhang, Z., Wu, F.-F.: Research on precision teaching model based on big data and data mining technology. Inf. Technol. Inf. **12**(10), 130–131,134 (2018)
3. Hong, Z.H.O.U.: Flipped-classroom-based implicit knowledge construction of business English. J. Shenyang Agric. Univ. (Social Science Edition) **20**(04), 463–467 (2018)
4. Wang, L.-P.: User implicit information requirement mining in reading promotion —based on censydiam model. Libr. Theory Pract. **234**(04), 75–79+103 (2019)
5. Tang, Q., Cheng, L., Zhang, L.: Rescheduling mode selection under recessive disturbance accumulation via machine learning. China Mech. Eng. **30**(04), 472–479 (2019)
6. Yi, I.: The impact of informal institutions on health resource allocation decision making—reflections on the implicit priority setting rules. Chinese Public Adm. **1**(03), 99–103 (2019)
7. Fu, W., Liu, S., Srivastava, G.: Optimization of big data scheduling in social networks. Entropy **21**(9), 902 (2019)
8. Liu, S., Li, Z., Zhang, Y., et al.: Introduction of key problems in long-distance learning and training. Mobile Netw. Appl. **24**(1), 1–4 (2019)
9. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. Complex Intell. Syst. **1**, 23 (2020). https://doi.org/10.1007/s40747-020-00161-4
10. Li-you, W.A.N.G.: The design and implementation of big data platform of university smart campus. J. Hebei Norm. Univ. Natl. **162**(02), 94–99 (2020)
11. Ang, L.M., Ge, F.L., Seng, K.P.: Big educational data & analytics: survey, architecture and challenges. IEEE Access **8**(99):1–1 (2020)
12. Zhou, Y., Song, Z.: Effectiveness analysis of machine learning in education big data. J. Phys.: Conf. Ser. **1651**(1), 012105 (7pp) (2020)
13. Bulltail, G., Walter, M.T.: Impacts of coal resource development on surface water quality in a multi-jurisdictional watershed in the Western United States. J. Contemp. Water Res. Educ. **169**(1), 79–91 (2020)
14. Soini, K., Korhonen-Kurki, K., Asikainen, H.: Transactional learning and sustainability co-creation in a university - business collaboration. Int. J. Sustain. High. Educ. **20**(6), 965–984 (2019)
15. Hu, J., Huang, L., Sun, T., et al.: Proactive planning of bandwidth resource using simulation-based what-if predictions for Web services in the cloud. Front. Comp. Sci. **15**(1), 1–28 (2021)
16. Jones, L., Credo, J., Parnell, R., et al.: Dissolved uranium and arsenic in unregulated groundwater sources – western navajo nation. J. Contemp. Water Res. Educ. **169**(1), 27–43 (2020)
17. Hao, D., Tu, S., Zhang, C.: Experimental study on the effect of moisture content on bituminous coal porosity based on 3D reconstruction of computerized tomography. Nat. Resour. Res. **29**(3), 1657–1673 (2020)
18. Caron, J., Asselin, H., Beaudoin, J.M.: Attitudes and behaviors of mining sector employers towards the Indigenous workforce. Resour. Policy **61**(10), 108–117 (2019)
19. Wang, T., Zhang, H., Gamage, R.P., et al.: The evaluation criteria for rock brittleness based on double-body analysis under uniaxial compression. Geomech. Geophys. Geo-Energy Geo-Resourc. **6**(3), 1–19 (2020)
20. Hellqvist, M.: Teaching sustainability in geoscience field education at falun mine world heritage site in Sweden. Geoheritage **11**(4), 1785–1798 (2019)

21. Helal, S., Li, J., Liu, L., et al.: Identifying key factors of student academic performance by subgroup discovery. Int. J. Data Sci. Anal. **7**(3), 227–245 (2019)
22. Yu, J.: Analysis and design of course website for software testing based on SPOC. J. Phys. Conf. Ser. **1187**(5), 052015 (2019)
23. Turner, P.J., Thaler, A.D., Freitag, A., et al.: Deep-sea hydrothermal vent ecosystem principles: Identification of ecosystem processes, services and communication of value. Marine Pol. **101**, 118–124 (2019)
24. Patra, R., Saha, S.K.: A hybrid approach for automatic generation of named entity distractors for multiple choice questions. Educ. Inf. Technol. **24**(2), 973–993 (2019)
25. Lynch, S.: Education resources for the boiler supply chain. Eng. Syst. **36**(5APP.), 3–3 (2019)
26. Ma, T., Antoniou, C., Toledo, T.: Hybrid machine learning algorithm and statistical time series model for network-wide traffic forecast. Transp. Res. Part C Emerg. Technol. **2020**(111), 352–372 (2020)
27. Mu, A.Y.: A hybrid machine learning model with cost-function based outlier removal and its application on credit rating. J. Phys. Conf. Ser. **1584**(1), 012001 (2020)
28. Chefira, R., Rakrak, S.: Accuracy assessment of applied supervised machine learning models on usual data probability distributions. J. Phys.: Conf. Ser. **1743**(1), 012011 (13pp) (2021)
29. Gosal, A.S., Ziv, G.: Landscape aesthetics: spatial modelling and mapping using social media images and machine learning. Ecol. Ind. **117**(0), 106638 (2020)
30. Sengupta, P.P., Shrestha, S., Béatrice, B., et al.: Proposed requirements for cardiovascular imaging-related machine learning evaluation (PRIME): a checklist: reviewed by the American college of cardiology healthcare innovation council. JACC. Cardiovasc. Imaging **13**(9), 2017–2035 (2020)