Research on College Students' English Online Autonomous Learning Based on ID3 Algorithm

Zhangyan Li(✉)

Guilin University, Guilin Guangxi 541006, China
115249936@qq.com

Abstract. The purpose of this project is to study the use of ID3 algorithm in online learning environment. The purpose of this study is to study whether autonomous learning based on ID3 algorithm can be used as an effective tool for College English teaching. This research will provide us with information about how students learn and use their online resources, what impact it has on them, and whether there are any negative effects of using these tools. This study aims to determine whether students can learn English more effectively when they are able to work by themselves, rather than when they are assigned tasks that require them to complete in a group environment. The study will involve two groups: one group will work independently, and the other group will work together using an online program called English Online Autonomous Learning (eoal). We intend to provide an effective and efficient English teaching method by using ID3 algorithm. The main purpose of this study is to find out how ID3 can be used as an autonomous learning system for English teaching through web-based multimedia tools such as video clips and audio files.

Keywords: ID3 algorithm · Online learning · College English · Reform in education

1 Introduction

Autonomous learning has become a hot topic in the field of social research since the 1950s. With the development and progress of Chinese society, communication with foreign countries has become increasingly close, and English autonomous learning has been paid more and more attention by English learners. Henry holee once pointed out in 1981 that learner autonomy is the ability of students to manage language learning. In this situation, the majority of English educators began to pay attention to and actively engage in the research of autonomous learning. This paper will study college students' English Online Autonomous Learning Based on ID3 algorithm [1].

According to the spirit of the outline of basic education curriculum reform (for Trial Implementation), it can be understood that we just want to change the past tendency of "overemphasizing acceptance learning", rather than completely denying the acceptance learning method, and advocate students to learn how to study independently.

M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2022, LNICST 465, pp. 586–595, 2023.
https://doi.org/10.1007/978-3-031-23950-2_64

Meanwhile, in the process of foreign language teaching, when we advocate cultivatingcollege students'autonomous learning ability,we cannot ignore the factors of Englishclassroom environment.The factors that affect the students'autonomous learning mainlyinclude three categories: physical environment,social environment and psychologicalenvironment. In terms of physical environment, the campus environment and hardwarefacilities can affect students'autonomous learning. The social environment mainly includesteachers' support, harmonious relationship between teachers and students. The psychologicalenvironment involves class involvement, teachers'innovation and so on.

So far, researchers have carried out a great deal of studies on college students'perception of classroom environment and the autonomous learning ability, but most of thestudies only tend to one aspect of them.Among these studies, some are about the relationshipbetween classroom environment and learning achievement, or classroom environment itself inuniversities or senior high schools, others are about the correlation between the autonomouslearning ability and achievement or its influence in the process of English learning.However,the analysis of the correlation between the classroom environment and the students'autonomous learning ability is relatively few. Thus, this paper aims to investigate the potentialcorrelation between these two aspects. Figure 1 below shows the flow chart of autonomous learning.

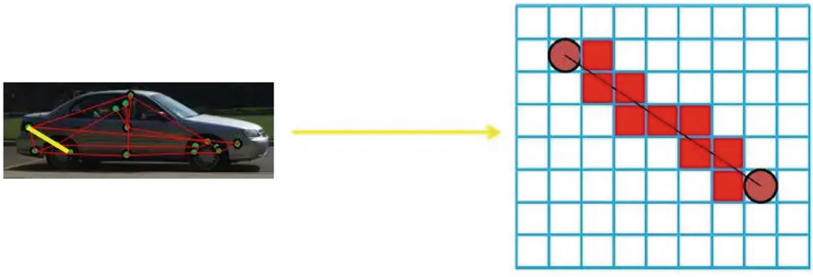

Fig. 1. Autonomous Learning flow chart

2 Data Mining and Decision Tree

2.1 ID3 Algorithm

ID3 (iterative dichotomizer 3) algorithm was proposed by Quinlan in 1986. Based on the theory of information entropy, Fig. 2 below shows the ID3 algorithm judgment process.

ID3 is a typical decision tree learning system. Its core is to select attributes on all levels of nodes of the decision tree, and use information gain as the attribute selection standard, so that when testing on each non leaf node, the largest category information about the tested example can be obtained [2]. After using this attribute to divide the instance set into subsets, the entropy of the system is the smallest. It is expected that the average path of the non leaf node to each descendant leaf node is the shortest, which makes

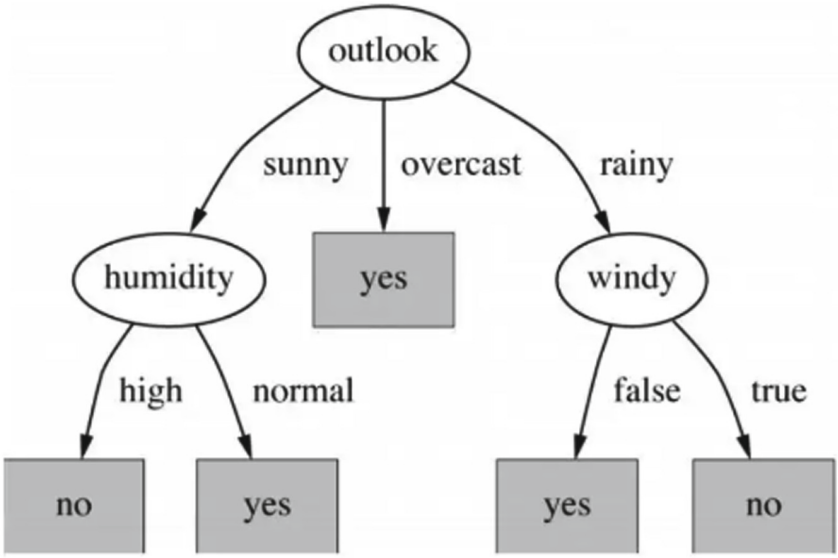

Fig. 2. ID3 algorithm judgment process

the average depth of the generated decision tree smaller and improves the classification speed and accuracy. The basic algorithm of ID3 algorithm is greedy algorithm, which uses top-down recursion to construct decision tree. Its principle is to summarize and summarize a large number of data, extract a universal and general description (i.e. the attribute laws of things), and express these laws in the form of decision tree.

$$E(A) = \sum_{i=1}^{v} \frac{S_{1i} + \ldots + S_{mi}}{S} - I(S_{1i} + \ldots + S_{mi}) \tag{1}$$

ID3 adopts this top-down strategy to search part of the whole space, which ensures that fewer tests are done, so the classification speed is also fast. ID3 learns by constructing a decision tree from top to bottom [2]. All current training samples are used in each step of the search. Because the statistical attributes of all samples are used, the sensitivity to errors of individual training samples is greatly reduced, so the algorithm is suitable for learning with errors in training samples. The effectiveness of ID3 on autonomous learning is shown in Fig. 3 below.

2.2 Improved ID3 Algorithm

Set a as the selection attribute, a has v attribute values, and the corresponding weight is ω_1, ω_2,…,ω_i. The attribute a is extended according to ID3 algorithm, and the corresponding information entropy is $E(B1)$, $E(B2)$,…, $E(Bv)$. Then the weighted entropy is defined as:

$$E(A)^* = \sum_{i=1}^{v} \omega_i \times E(B_i) \tag{2}$$

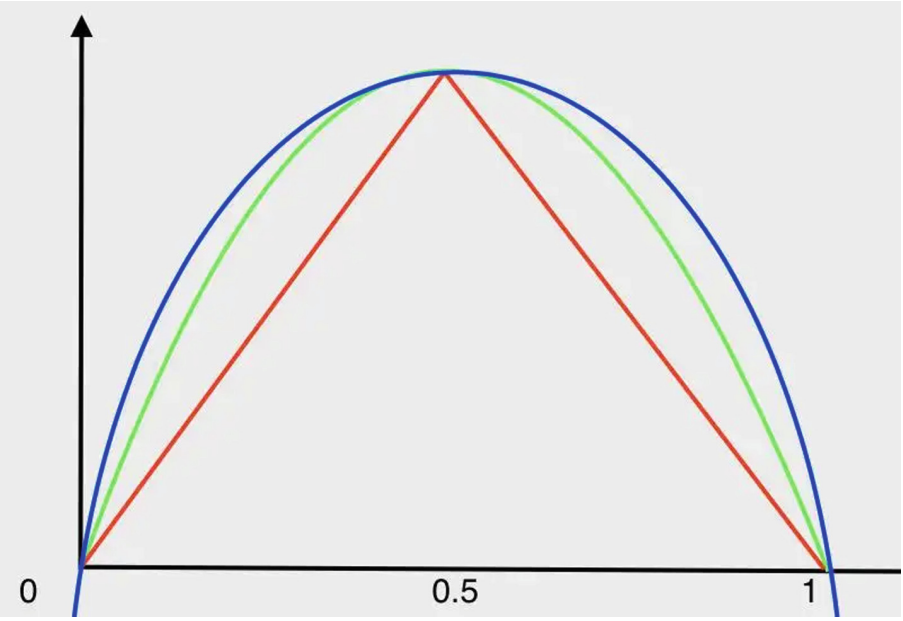

Fig. 3. The effectiveness of ID3 for autonomous learning

Directly copying a data set takes a lot of memory. If the depth of a tree is k, the data star of the original data set is n, and the dimension of each data is m, when the data itself is big data, M may be very large, then the total spatial complexity of the whole decision tree is O (n*k*m) (the worst case is a full tree with depth K, that is, all leaf nodes are at the bottom level, and each level is a division of the whole data set at this time) [3]. The robustness of ID3 algorithm framework is shown in Fig. 4 below.

(1) .n refers to the data scale. If the amount of data is large enough, the method of importing memory in batches can be adopted, but the reading operation of data sets on the Weka platform is transparent to learners. This point is difficult to import memory in batches without looking at the underlying code of the data import implementation mechanism, so it is not optimized here

(2) K is the depth of the tree obtained according to the established decision tree algorithm ID3, which can be optimized in combination with the decision tree algorithm taught by the teacher, such as 4.5 and 4.8 algorithm, which many students have done here, and there is ready-made code on the Weka platform [4]. I think it is absolutely effective to invent a new algorithm that can significantly shorten the depth of the decision tree theoretically, and the effect is immediate, but for beginners, my theoretical knowledge reserve is insufficient, It is difficult to find a universal algorithm that abandons experience and avoids generalizations, but rises to the theoretical level. This can be achieved by teachers, so I won't teach others. I will talk about how to optimize the spatial complexity of the algorithm from the perspective of a programmer. Of course, I will start from the perspective of algorithm implementation

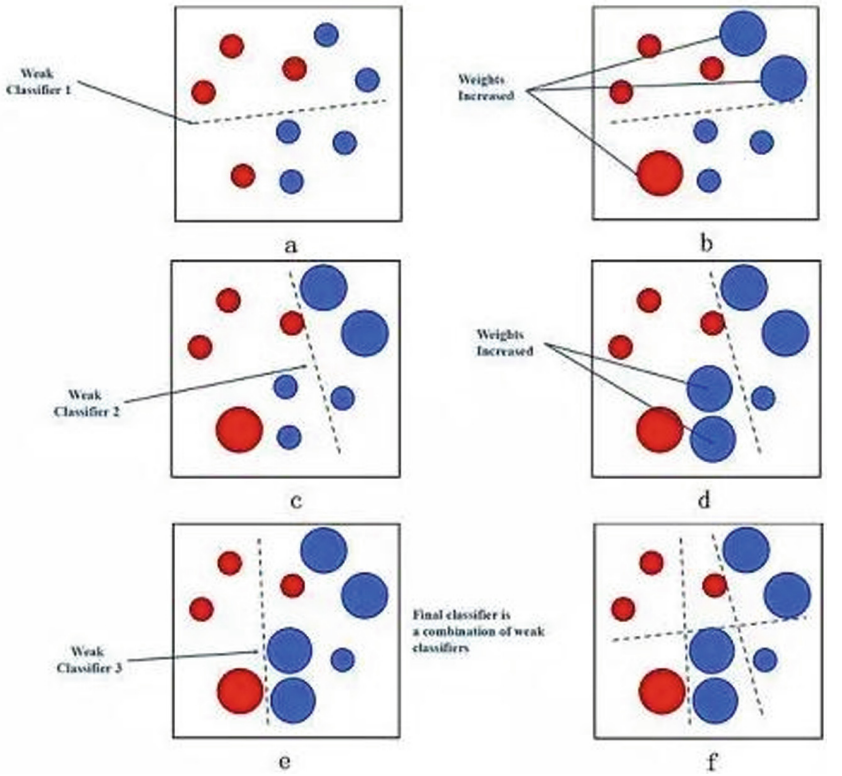

Fig. 4. Robustness of ID3 algorithm framework

based on details.

$$\pi(\theta|x) = \frac{p(x|\theta)\pi(\theta)}{m(x)} = \frac{p(x|\theta)\pi(\theta)}{\int\limits_{\emptyset} p(x|\theta)\pi(\theta)d_\theta} \tag{3}$$

(3) .m is the dimension, which is an important factor affecting the spatial efficiency of the algorithm when the data is complex. Imagine that M is 101001000… Then the space consumption of the corresponding algorithm should be increased by 10 times, 100 times, 1000 times. Opening memory on all nodes to store the corresponding data sets is very memory consuming. How to optimize it? It is absolutely feasible to store only the index of the corresponding data set in the original data set on all nodes, and determine which data is divided into the corresponding child nodes through the index. In this way, the spatial complexity of the algorithm is effectively reduced to o (n*k), which can completely avoid the risk of insufficient memory caused by excessive data dimension in the data set, and improve the robustness and robustness of the program. Moreover, it is faster to find corresponding elements and divide elements into corresponding subtrees through subscript indexing than to directly operate the whole element, because what we directly operate is an index (probably an int type data), rather than a user-defined object with complex structure with large

dimensions [5]. There is no need to participate in the preparation of constructors, copy constructors, assignment functions, etc. related to object initialization, which simplifies programming; Moreover, the instantiation of objects requires the new operator, which consumes computer resources in any programming language. Why not optimize the execution time and space of the algorithm?

3 College Students' English Autonomous Learning Under the Network Environment

3.1 Current Situation of College Students' English Autonomous Learning

(1) Positive results. By using network information, students' learning efficiency is greatly improved. The application of network technology in English teaching provides real communication needs and language learning atmosphere, and transforms abstract and boring learning contents into vivid and vivid learning contents [6]. Therefore, it enhances students' interest in learning English. At the same time, through the resource sharing on the network, if students encounter problems in the process of autonomous learning, they can query relevant information on the Internet at the first time, which is conducive to obtaining a large amount of knowledge and cultivating autonomous learning ability [7]. Multimedia and network technology provide students with a relaxed learning environment. Without face-to-face communication, students are more likely to release emotional anxiety. Therefore, learning efficiency has been improved and the effect is more positive. Figure 5 below shows the English autonomous learning process.

(2) Existing problems. Influenced by the traditional teaching mode, many students have poor autonomous learning ability and can not adapt to the teaching mode of autonomous learning. This means that the teaching mode of teachers' autonomous learning in the classroom is difficult to get the cooperation of students. Some students have utilitarian motivation to learn English. Their motivation to learn English is to get CET-4 and CET-6 certificates. This utilitarian motivation is not conducive to cultivating students' autonomous learning ability in the network environment. At the same time, under the influence of this traditional education model, not only students do not adapt to the teaching model of autonomous learning, but also teachers can not adapt to the transformation of roles in the autonomous teaching model [8]. This also hinders the implementation of the autonomous learning teaching model to a great extent. Some teachers go too far, directly abandon the traditional teaching platform and completely rely on network resources. In this way, all learning tasks are pressed on students, ignoring the guiding role of teachers in the whole autonomous teaching model, which will make students learn blindly and find no learning direction.

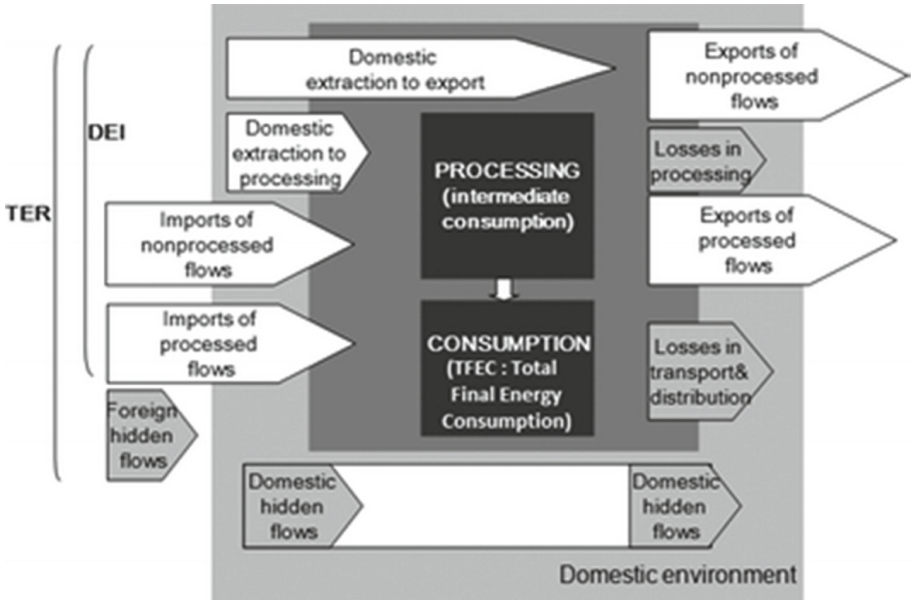

Fig. 5. English autonomous learning process

3.2 Countermeasures for Cultivating College Students' English Autonomous Learning Ability Under the Network Environment

(1) Create a good e-learning platform. An advanced e-learning platform is an important guarantee to improve the efficiency of College Students' English autonomous learning [9]. The platform can adopt a more advanced, efficient and humanized software system. Its function should include the retrieval of curriculum learning resources, have a good interactive function between teachers, students and students, and can reflect all links of the teaching process. In the learning process, it can retrieve problems at any time, has flexible and convenient upload and download methods, and establishes an effective test and Q & a system. Quickly enrich teaching resources in multiple ways, so that all kinds of course materials can be uploaded and updated in time.

(2) Cultivate students' ability to correctly use e-learning. Most students are influenced by the traditional teaching mode and are used to "walking" passively with teachers in the classroom [10]. Therefore, we should pay attention to correcting students' bad habits, encourage them to give full play to their main role in the classroom, change their utilitarian learning motivation, and formulate learning objectives and learning plans in line with students' reality under the guidance of teachers. At the same time, guide students to use network resources for independent learning, so that students can achieve the purpose of full preparation before class and full review after class through the network.

(3) Change the role of teachers in College English classroom. As the manager of e-learning platform, teachers should not only make full use of and master the functions

of the platform, but also constantly optimize and improve the platform [11]. In College English classroom, teachers should let students play the main role of learning, and can not exaggerate the role of network resources, completely rely on network resources, and ignore their guiding and supervising role for students.

3.3 Using ID3 Algorithm to Study College Students' English Autonomous Learning

This paper analyzes and Studies College Students' English autonomous learning by selecting the five attributes of the longest login time, the cumulative online time, the number of visits to the learning resource network, questions and posts, and the progress of autonomous learning as candidate attributes and autonomous learning results as class label attributes. The data records are from the College English learning database, which holds the learning records of College English in general liberal arts and science and Engineering in the past two years [12].

The decision tree generated by ID3 algorithm is shown in Fig. 6:

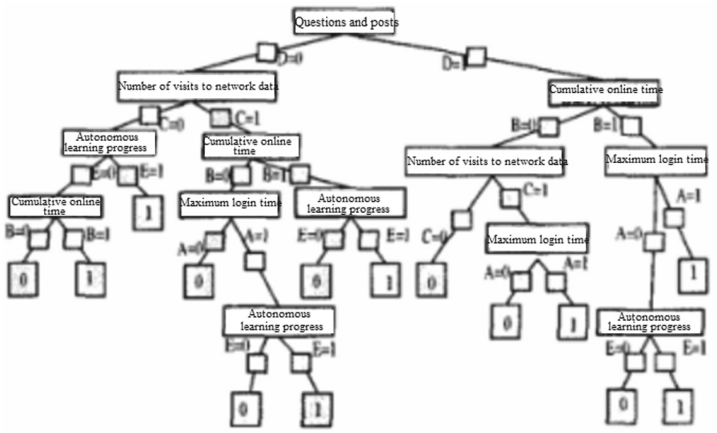

Fig. 6. Decision tree generated by ID3 algorithm

Using analysis of variance (ANOVA) to learn college students' English Autonomous Learning in a classroom environment the purpose of this chapter is to provide you with an overview of ANOVA. It also provides you with a brief description of the software that can be used for statistical analysis and how to apply it to all aspects of learning classroom teaching. English classroom teaching is a unique and effective way to teach English to non-native speakers [13]. It combines the best aspects of traditional classroom teaching with the latest research on language learning, which shows that students will learn more effectively when they actively participate in the learning process.

College English Online Autonomous Learning (oalce) is based on ID3 algorithm, which is used to identify and mark audio files. The ID3 standard was developed by Michael sweet in 1994 as a way for digital music players to recognize tracks and make

them available for playback. Since then, it has been adopted by many other applications, including software that can automatically recognize speech from microphones or video files. The oalce system works on a similar principle: it uses its own proprietary technology to identify spoken words in recordings, and then marks them with metadata that describes their meaning so that they can be searched later [14].The decision tree generated by the improved ID3 algorithm is shown in Fig. 7:

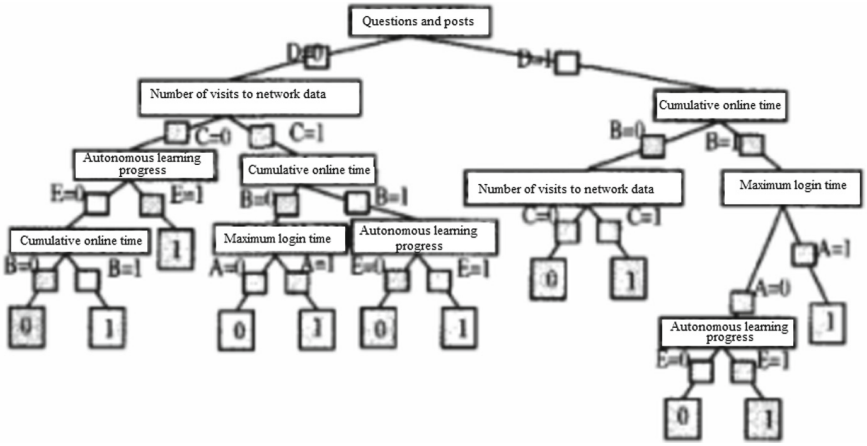

Fig. 7. Decision tree generated by improved ID3 algorithm

4 Conclusion

Online Autonomous Learning of College Students' English based on ID3 algorithm is a unique way of English learning. It focuses on students' ability to learn and use English in real life. The program is based on the latest technology and allows students to learn by themselves without the supervision of any teacher or tutor. The plan consists of two parts: the first part is an online course, which contains all the necessary materials for self-study; The second part is an interactive learning environment, which helps students master new skills through practice and interaction with other learners. Students use this program to learn language and improve their speaking, listening, reading and writing skills. Students can use this program to learn more effectively because it is designed by experts who know how to teach people how to speak English. Based on the latest technology, Online Autonomous Learning of College Students' English based on ID3 algorithm enables learners from all over the world to connect with each other through video conferences and chat rooms.

References

1. Lu, Y.U.: Investigation and analysis of college students' English autonomous learning based on online open courses. J. Changzhou College Inf. Technol. (2019)

2. Wu, C.: Effect of online and offline blended teaching of college english based on data mining algorithm. J. Inf. Knowl. Manag. **21**(Supp02) (2022)
3. Ma, X.: Study on college English online teaching model in mixed context based on genetic algorithm and neural network algorithm. Discr. Dyn. Nature Soc. 2021 (2021)
4. Wang, Q.: Construction of two way interactive and cooperative teaching mode between college English teachers and students under the network autonomous learning platform. In: ICIMTECH 21: The Sixth International Conference on Information Management and Technology (2021)
5. Sang, H., Jiang, M., Lu, Z., et al.: Data mining algorithm for influencing factors of university students' position promotion based on decision tree. J. Beihua Univ. (Nat. Sci.) (2019)
6. Zhang, B., Fang, J., Xiong, Z.: Online English vocabularies autonomous learning model based on EM algorithm. J. Phys. Conf. Ser. 1168 (2019)
7. Qian, W.: Challenges and strategies of online English learning for college students in the post epidemic era **12**(2), 5 (2022)
8. Li, J.: Mobile-assisted autonomous English learning in college students' extra-curriculum time: a case study in China. Technium Soc. Sci. J. 19 (2021)
9. Chen, L., Bai, J.L.: Research on the training mode of "the Combination of College Students'Extracurricular English Autonomous Learning and Chinese Culture" based on WeChat platform. Overseas English (2019)
10. Ding, J.: A study on vocational college students' English learner autonomy based on three-dimensional conceptual model. J. Ningbo Inst. Educ. (2019)
11. Zhang, F.: A study of college students' autonomous learning based on project-based learning. J. Qiqihar Junior Teachers' College (2018)
12. English hybrid teaching model analysis teaching based on association algorithm. J. Phys. Conf. Ser. **1881**(2), 022041 (2021). (7pp)
13. Li, M., University, J.: A study on strategies of training college students' English learning autonomy based on network environment. J. Heihe Univ. (2019)
14. Xiao-Fan, L.I.: Research on college english autonomous learning model under the background of big data. J. Hubei Open Vocational College (2019)