Research on College Students' English Online Autonomous Learning Based on Big Data Analysis

Jie Wu(✉)

English Department, School of Foreign Languages,
Dalian Polytechnic University, Dalian 116033, Liaoning, China
wujie@dlpu.edu.cn

Abstract. This paper analyzes some problems existing in the process of College Students' English autonomous learning, and improves the ID3 algorithm in the field of data mining, and applies the traditional ID3 algorithm and the improved ID3 algorithm to the research of College Students' English autonomous learning respesctively. This paper holds that only when college students have a certain degree of autonomous learning ability, can they effectively carry out Network-based College English learning. The key to improve autonomous learning ability is students' self-monitoring. Therefore, after improving the theory and awareness of self-learning, College English teachers should actively strengthen students' self-learning awareness, cultivate students' self-monitoring ability, help students complete the transformation from other control to self-control, from conscious to automatic, from local to overall, so as to improve their autonomous learning ability and realize the goal of College English online teaching, That is to achieve the best learning effect of students through personalized learning.

Keywords: College Students' English autonomous learning · Data mining · ID3 algorithm

1 Introduction

The new model of College Students' English Autonomous Learning Based on ID3 algorithm is favored by colleges and universities because of its strong interactivity, more choices, less restrictions, wide information and high efficiency, and most of them have mastered more mature theoretical methods and practical skills. However, from the current situation of the learning mode in Colleges and universities, there are still some problems, such as the lag of school hardware and software facilities, the lack of teachers' auxiliary effect, the limited ability of students' autonomous learning, etc., which directly limit the advantages of the learning mode. Therefore, it is necessary to study the Network-based English autonomous learning of college students in order to promote the further improvement of the learning mode, provide help for the improvement

M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2021, LNICST 391, pp. 494–505, 2021.
https://doi.org/10.1007/978-3-030-87900-6_57

of College Students' Comprehensive English application ability, and lay a foundation for the lifelong learning of college students [1–5].

In nearly 10 years of English teaching career, I have been exploring, creating and researching to make English class more interesting, more wonderful and more delicious, so that more students can enjoy English class, so that they can gradually fall in love with English. In teaching, teachers should gradually improve the guidance plan of each semester and make exquisite courseware for each class; Often tell students English stories, sing English songs, guess English riddles, tell English jokes, talk about English life, put funny English short films to show their unique problem-solving thinking [6–8].

Modern teachers use computer-aided instruction more and more widely, although teachers spend a lot of effort, and students like and recognize it, but in fact, students are still less active in learning, and students are still in a passive state of acceptance. So how to use the advantages of information technology to give full play to students' active participation and learning ability? It's a good way to let computer network into English class [9–12]. At present, the online "English Paradise" which is suitable for students' autonomous learning is deeply liked by students and has received good results. It makes English more interesting and operational, and cultivates students' ability to use their brains and hands to learn English.

English online "Le Xue Yuan" mainly consists of three parts [13–19], one is "knowledge in mind", the other is "testing I can do", the third is "learning I break through". Learning in the English online "Le Xue Yuan" can be said to be playing in learning and learning by playing. This kind of emotional learning makes English learning no longer boring and boring. Now the specific contents, main functions and operation of the three plates are introduced as follows.

(1) Knowledge is in the heart

It mainly includes pre class, in class and after class learning content, which is compiled into preview case, communication case and test case for students to use. The key and difficult problems are recorded in the "micro class" so that students can learn repeatedly and understand the problems thoroughly. It is necessary to upload students' current typical problems, such as students' homework, test papers, etc., and of course, it also shows the excellent students' test papers, homework, test papers, etc. Methods of solving problems. In this way, comparative learning, mutual learning, can promote learning.

For some classroom experiments, whether they are demonstration experiments or students' experiments, they are more or less confused after learning, and students want to do experiments again to solve their doubts. There is a very good software "junior high school English NB laboratory", which can complete all the experiments of the whole junior high school. Students can operate freely in the "Le Xue Yuan" and complete the experimental exploration, so that they can have a deep understanding of the problems and promote the smooth completion of learning. In the study of electricity, drawing circuit diagram and connecting physical diagram is the basis of learning electricity well. In order to let students have a better understanding, judge the circuit, analyze the circuit, circuit fault, ammeter, voltmeter, the use of sliding rheostat and other electrical problems, we can do it in the "electrical virtual laboratory" on the online "Le Xue Yuan". Teachers often use the software in teaching, which is convenient for teaching and makes the problems

intuitive and clear. If students encounter some problems in electricity, they can solve them by using their hands in "Le Xue Yuan" "Knowledge in mind" column, can help students to complete the problems difficult to solve in the classroom, let students free, participate more, experience more, students in real learning, to achieve flexible use of knowledge, they learn more confidence, more fun.

(2) I can do it

As long as it is training and hands-on operation, there will be a special evaluation, which can reflect the students' achievements in time. When assigning homework at the weekend, you can set the time to complete the homework. If it exceeds, the system will stop and you need to do it again. Teachers can know when each student does it, how much time it takes to do it, which questions are wrong, whether to modify it, whether to retest, etc. all these are recorded. Parents can be informed of the unfinished or completed situation in time to understand the children's learning status and learning effect. Of course, parents can log in to the english paradise on the parents' side, and see the homework done by their children on the Internet, so as to better communicate with their children. The whole class has homework feedback form, the highest score, the lowest score, the average score, the number of people in each section, etc. there are comparisons between classes. Very good analysis of homework or examination, easy to teach. Online homework, students can complete more seriously, dare not lazy, dare not careless, also can let students correct their mistakes in time, the system will remind, and the system will automatically put each wrong question together, generate the wrong question book, convenient for students to review. This plate has powerful detection function, thorough analysis, very direct and fast understanding of students' homework and detection situation, better discovery of problems, targeted communication with students, and also reduce the burden of teachers in performance analysis.

2 Introduction of Data Mining and Decision Tree

The platform subverts the previous classroom teaching based course teaching form. Students can easily log on to the learning platform anytime and anywhere to learn a course through modern network means. The process is to evaluate and analyze students from logging on the network learning platform → learning course knowledge points → completing homework → course test and comprehensive test → ending course learning.

Auxiliary

The use of autonomous learning platform should be used as the auxiliary and extension of conventional classroom teaching, and can not completely replace classroom learning. Teachers upload the relevant information to the teaching platform, which can help students preview before class and consolidate their learning after class, expand the course content and complement the traditional teaching.

Autonomy

Autonomous learning platform learning lacks effective learning process control mechanism for students. It is difficult to monitor students' learning time and learning effect in a short period of time. Cultivating students' autonomous learning ability is the key to bring the effect of e-learning platform into full play.

Repeatability

Traditional classroom teaching requires students to be highly focused in class, which is often difficult to achieve, and the difficulties of classroom content are difficult to master at one time through classroom teaching. The online autonomous learning platform provides technical support for students' repeated learning. Teachers can record course videos, organize course related materials and upload them to the online learning platform, which can make the classroom reappear and provide students with the opportunity of repeated learning for two or more times.

Asynchronous

Traditional teaching is the synchronization of classroom teaching and learning, students and teachers must be synchronized in time, ideas, in order to achieve better learning effect, and autonomous learning platform can achieve asynchronous teaching and learning, as long as teachers arrange learning tasks and carry out appropriate training, students can reasonably arrange their own time to complete learning within a given period.

Open

The traditional teaching mainly focuses on teaching materials. At the same time, due to the constraints of classroom time, the teaching content is greatly limited. The online learning platform can provide students with an open learning environment. Through the guidance of teachers, the curriculum content can be effectively expanded and extended, especially in combination with the real world. The content is no longer abstract and dogmatic, Let students realize the usefulness of the course, stimulate students' interest in learning.

Interaction

The communication and communication in traditional classroom teaching is very limited. Autonomous learning platform can use various communication tools to improve the space and time of communication, so that teachers can fully grasp students' learning state and effect, better grasp the difficulties of teaching, and promote students' autonomous learning ability, And anonymous interaction can greatly reduce the psychological pressure of students in communication.

Online lesson preparation and online teaching

Autonomous learning platform through the course content management, resource management, learning plan management, work management, question and answer management and learning monitoring management and other functional modules to complete online lesson preparation, online teaching, learning control and a series of teaching activities.

Online learning

Students can carry out autonomous learning, cooperative learning and research-based learning through online courses, topic discussion, friend exchange, question and answer interaction, work mutual evaluation and other functional modules.

Intelligent comprehensive evaluation

In the process of practice evaluation such as classroom practice and synchronous homework, and inspection evaluation such as mid-term test, final test and simulation test, the self-learning platform adopts the latest technologies such as intelligent test paper generation and intelligent evaluation to provide a complete set of solutions for each evaluation stage.

Perfect evaluation management

All the data of teaching and learning process are automatically recorded in the whole process, through which the evaluation scheme is made, and the system generates process evaluation data according to the evaluation scheme, so as to ensure the objectivity and fairness of the evaluation.

Improve teaching efficiency and reduce the burden of lesson preparation

In the process of teaching, teachers can focus on the teaching according to the learning situation monitored by the system and the relevant data recorded, so as to improve the teaching efficiency; In addition, the subject teaching plan and course content are uniformly allocated by the lesson preparation team leader, and teachers can set them according to their own needs, so as to reduce the burden of lesson preparation.

Resource CoConstruction and sharing

The self-learning platform system has the functions of uploading teaching resources, adding test questions, making works and voting evaluation. Test questions, resources and works can be recommended and reviewed level by level, realizing the co construction and sharing of excellent teaching resources.

SNS concept

The system aggregates resources, students and teachers through tags, wikis and other technologies, so as to form the knowledge network and crowd network for autonomous learning and cooperative learning, and provide a good tool and open platform for students' autonomous inquiry learning.

The comprehensiveness, openness and timeliness of evaluation

The system has the process evaluation of students' online learning and the summative evaluation of online unified examination, and supports the open evaluation of works. The relevant examination data is managed uniformly to form students' comprehensive evaluation files.

Student portfolio

The platform automatically records the whole process of students' learning and growth, and finally forms a personal growth portfolio.

Data mining is a process of selecting, exploring and modeling a large number of data in order to discover unknown rules and relationships in advance. The purpose of data mining is to obtain clear and useful results for the owner of data. Data mining is the core technology of discovering knowledge from database. It is developed from machine learning of artificial intelligence. Combined with traditional statistical analysis method, fuzzy mathematics method and visualization technology of scientific calculation, data mining method and technology are formed by taking database as research object.

Data mining methods include decision tree method, set theory method, neural network method, genetic algorithm and so on. The decision tree method uses the principle of information theory to establish a decision tree. This method has good effect and great influence. The representation form of knowledge obtained by this method is decision tree.

The decision tree is a tree structure with the attributes of samples as nodes and the values of attributes as branches. The root node is the most informative attribute among all the samples, the middle node of the tree is the attribute with the largest amount of information in the sample subset contained by the subtree with the node as the root, and the leaf node is the category value of the sample(shown in Fig. 1).

2.1 ID3 Algorithm

ID algorithm is the earliest and most influential decision tree method in the world. Its basic algorithm is greedy algorithm. It constructs decision tree by top-down recursion. The information gain measure is used to select test attributes on each node of the tree. Select the attribute with the highest information gain as the test attribute of the current node.

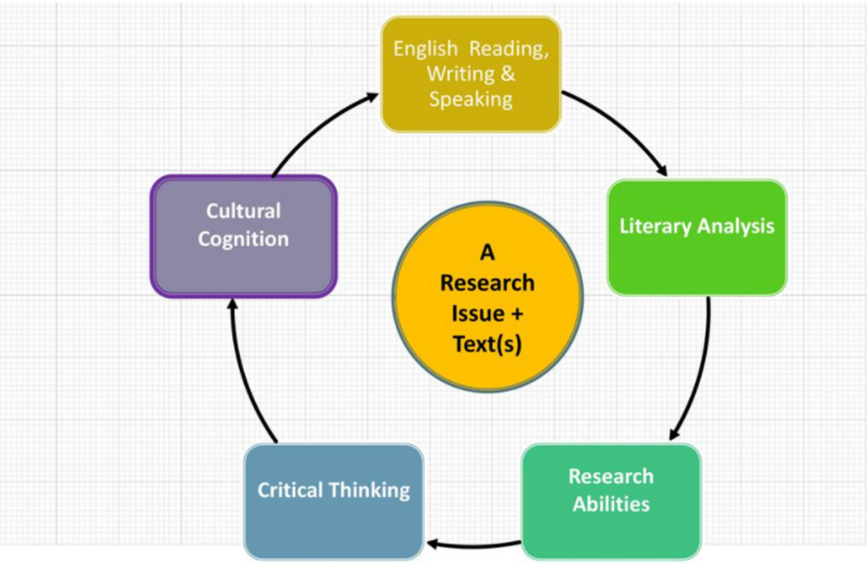

Fig. 1. English online autonomous learning

Let S be the set of s data samples. Suppose that the class label attribute has m different values, define m different Ci (I = 1,2, …, m) If Si is the number of samples in class Ci and Pi is the probability that any sample belongs to Ci, then the expected information for a given sample classification is as follows:

$$I(s_1, s_2, \cdots s_m) = -\sum_{i=1}^{m} P_i \log_2(P_i) \tag{1}$$

Let attribute A have v different values, We can divide s into v subsets {S1, S2,… Sv}, where Sj contains some samples in S that have a on A. If a is selected as the test attribute, these subsets correspond to the branches growing from the nodes containing the set S. Let Sij be the sample number of class Ci in the subset Sj, then the entropy or expected information divided into subsets according to attribute a is as follows:

$$E(A) = \sum_{i=1}^{v} \frac{S_{ij} + \cdots + S_{mj}}{S} I(S_{ij}, \cdots, S_{mj}) \tag{2}$$

The smaller the entropy, the higher the degree of subset division. Then, the information gain of branch A is as follows:

$$G(A) = I(s_1, s_2, \cdots, s_m) - E(A) \tag{3}$$

This kind of ID3 algorithm tends to choose the attribute with more values, because the weighted sum method makes the classification of instance set tend to abandon the data tuples with small amount of data. However, the attribute with more values is not

always the optimal attribute. In other words, according to the principle of minimizing entropy and maximizing information gain, ID3 algorithm lists the attributes that should be selected by ID3 algorithm, and testing it will not provide too much information.

2.2 Improved ID3 Algorithm

Let A be the selection attribute, A has five attribute values, and the corresponding weight is $\omega_1, \omega_2, \cdots, \omega_v$. According to the algorithm of ID3, attribute a is extended, and the corresponding information entropy. Then the weighted entropy is defined as:

$$E(A)^* = \sum_{i=1}^{v} \omega_i \times E(B_i) \tag{4}$$

Among them (B1, B2, Bv) is the attribute selected for v nodes, and ω_i refers to the weight of the subset. The weight ω_i is calculated by the proportion of the subset Bi in the whole set, and then the weighted entropy is calculated, and the value of the attribute is selected by comparing the weight entropy.

In addition, the improvement of ID3 algorithm is used to simplify the complexity of information calculation.

Firstly, the basic properties of logP function are studied. Through the research, it can be proved that the information calculation formula is a kind of convex function, so we can use the unique properties of the convex function to improve the calculation formula of information quantity.

2.3 Functional Modules of Autonomous Learning System

(1) Management interface: teachers log in to the system through the management interface module, use the management interface to make and arrange educational materials, convert educational materials into appropriate learning objectives, and store them in the learning objectives database. These learning objectives form the basis of students' personalized learning content; The processing interface also outputs the learning objectives in the database to the textbook generation engine module to generate the textbook database.

(2) Textbook generation engine: the primary function is to check whether the input learning objectives meet the SCORM standard, and arrange the learning objectives that meet the scrom standard to generate textbooks and store them in the textbook database.

(3) Curriculum architecture engine: the primary function is to build personalized learning content adaptively according to the personal information and learning records of learners in the student file database and student record database, and combined with the learning content in the textbook database, and output the results to the curriculum visualization engine to show to learners.

(4) Course visualization engine: it is mainly used for the emergence of personalized learning content suitable for different learners. Different learners will see different learning content and arrangement forms, which reflects the adaptability and personalization of the system.

(5) Evaluation Center: it is mainly used to evaluate the learning effect of learners. According to the preset test questions, the learners are evaluated. The evaluation results, as the evaluation of the learners' learning ability and common sense, are also the important basis for the dynamic arrangement of learning objectives, and the corresponding learning content, personalized learning environment and learning strategy guidance appear.

3 Research on College Students' Autonomous English Learning by Using Decision Tree Method

How to transfer inquiry learning and extended learning to the Internet to develop students' scientific literacy, humanistic literacy and information literacy?

Online inquiry learning and extended learning is an exploratory research, which needs to be formed and developed in the process of curriculum development and implementation. We have tried novel coronavirus pneumonia online learning course, online learning course "harmony between humans and animal", online English appreciation course, and the online learning mode of "curriculum resource package + learning guidance".

(1) The theme resource package based on network resources is constructed by acquisition, screening, reorganization and a small amount of creation.
(2) Design program: background - knowledge learning - deep learning - interactive learning based on network learning space.
(3) Use learning guide (learning task book) to guide learning.
(4) Extensive reading (text and non text reading) is the basis of extended learning.
(5) Resources should be integrated into the process of students' autonomous learning to guide problems, stimulate thinking and active discussion.

By selecting the five attributes in the process of College Students' English autonomous learning, which are the longest login time, the cumulative online time, the number of visits to the learning resources network, the situation of questions and posts, and the progress of autonomous learning as candidate attributes, and the results of autonomous learning as class label attributes, this paper analyzes the autonomous learning of graduate students. The data record is from the College English learning database, which keeps the College English learning records of the College of foreign languages in recent two years.

Through the decision tree generated by ID3 algorithm, it can be found that if the number of questions and posts of a student reaches or exceeds 5, it indicates that the communication between the student and the teacher is relatively active, and the number of visits to the learning resource network reaches or exceeds 20 times, the autonomous learning progress is normal or ahead, and the accumulated online time reaches or exceeds 28 h, then the student's English autonomous learning result is qualified, The student can be upgraded and qualified for the next College English course.

Using ID3 algorithm and improved ID3 algorithm respectively, we can find that the decision tree generated by the improved I3 algorithm is relatively simple with fewer

branch nodes. The simpler the decision tree is, the less the cost of storing the decision tree, Moreover, the cost of transferring information between two entities is smaller. Moreover, the correct classification rate of the improved algorithm is 71.20%, which is higher than that of the improved algorithm (65.23%). Therefore, compared with the traditional ID3 algorithm, the improved ID3 algorithm has more advantages in the field of College Students' English autonomous learning.

The quality of learning materials is the premise of improving learning efficiency in college students' Online Autonomous Learning. Therefore, it is necessary to improve students' information literacy ability in order to obtain high-quality learning resources. Information literacy ability mainly includes four aspects: information acquisition ability, information immunity ability, information integration ability and information transformation ability. Information acquisition is the premise of using information. Students should know how to use various search technologies to find English learning materials that meet their own needs, including quickly and effectively finding useful learning resources in the network learning platform. In addition, there is no lack of negative and junk information in the massive network information, so students need to cultivate immunity to bad learning materials. The ability of information integration refers to the classification of English learning materials obtained by students according to the principle of "removing the rough and reserving the essence, eliminating the false and retaining the true", so as to ensure the practicability, correctness and value of the resources, so as to absorb and digest them, and integrate them into their own English knowledge structure system by using transformation ability. The above four abilities are progressive and indispensable, and these abilities are also the necessary skills for students' lifelong learning after graduation.

The concept of learning community was first proposed by the famous Japanese educator Sato, which is an efficient classroom teaching mode focusing on the interaction between teachers and students. "Changing a school requires teachers to open the door to the classroom," he said In the new era, with the development and driving of technology, the availability of education related resources is getting higher and higher. The unprecedented richness of resources and technology provides great convenience for teachers' classroom teaching, students' learning and parents' education and support for students. At the same time, the complexity of the impact on classroom teaching is also greatly improved. This kind of classroom breaks the original classroom boundary only constructed by teachers and students, expands into a star network topology composed of student-centered, teachers, parents, technology and resources, activates the connection between nodes, and straightens out the relationship between nodes, which has become an important goal to improve classroom efficiency and learning effectiveness. With the increase of community building elements, the number of connections between the elements will increase geometrically. How to help students actively activate the connection with each node, help students learn, and promote students' life growth will be the focus of our research.

According to this model, the elements interact, interconnect and promote each other. The development and drive of technology provide a convenient channel for teachers' teaching, students' learning and parents' support. The influx of resources also brings

great trouble for teachers and parents to choose resources suitable for students or students to choose their own resources, In this kind of classroom, which takes students' learning and growth as the core, independent choice becomes the key to improve classroom efficiency. According to the learning situation and their own teaching methods, teachers can choose the appropriate resources independently, and parents can choose the way that conforms to the children's learning habits according to their own ability and children's situation. More importantly, students should actively choose the learning resources suitable for themselves from the resources provided by teachers, given by parents and contacted by their own immediately related technologies.

4 Conclusion

Through the application of ID3 algorithm and improved ID3 algorithm to the research of College Students' English autonomous learning, the root nodes of the two decision trees generated are "questions and posts". It can be seen that in the process of College Students' English autonomous learning, the network communication between college students and College English teachers is very important, Therefore, it is necessary to ensure that college students have a relatively superior network infrastructure environment in the process of English autonomous learning, and the teaching role of College English teachers should not be ignored because it is autonomous learning, but should be paid more attention and applied, so as to better guide college students' English autonomous learning, monitor and master the situation of College Students' English autonomous learning, To improve the efficiency of College Students' English autonomous learning, achieve the expected effect of College Students' English autonomous learning, and ensure the normal progress of College English autonomous learning, so as to improve college English teaching activities.

Autonomous learning system is a kind of learning support channel. The channel takes learners as the main body, dynamically adjusts learning content and learning progress according to learners' cognitive ability and common sense level, and provides learning content, environment and strategy support for learners' individual needs. After personalized evaluation, the system actively adjusts personalized learning plans, Self control the whole learning process.

The main characteristics of autonomous learning system are as follows.

(1) According to the test results of learners' cognitive level, cognitive style and learning style, combined with the characteristics of the learning content, the system actively presents the most suitable learning content for learners.
(2) Autonomy, autonomous learning system can make learning as the main body to participate in learning, and can choose the corresponding learning ways and learning strategies according to their own learning network.
(3) The purpose of developing autonomous learning system is to let learners fully grasp the subject common sense system, which requires the system to comprehensively integrate teaching resources to get used to learners' autonomous selection and arrangement.

References

1. Xu, C., Hou, T., Huang, X., et al. Analysis of network teaching strategy structure. Distance Education in China (Comprehensive Edition), 10 (2006)
2. Sun, Y., Huang, H.: Research on network teaching strategy based on data mining technology. Teaching Method Research of Education and Occupation, 12 (2009)
3. Yuan, F.: Practical Data Mining. Electronic Industry Press, Beijing (2004)
4. Chen, X., Liu, J.: Construction of evaluation index system with decision tree method. Comput. Appl. 2 (2006)
5. Chen, L.: On the activation and improvement of students' autonomous learning system in TVU open and distance education. Mod. Dist. Educ. Res. **04**, 32–35 (2002)
6. Fang, L., Wang, K.: Personalized knowledge push service in network autonomous learning system. Res. High. Eng. Educ. **5**, 145–148 (2008)
7. Yi, H., Yang, J.: Design and development of Internet dynamic autonomous learning system. Comput. Eng. Des. **023**(007), 60–62 (2002)
8. Cao, W.: Design of personalized autonomous learning system based on e-learning. Vocat. Tech. Educ. **26**, 85–87 (2009)
9. Chen, L.: On the activation and improvement of students' autonomous learning system in open and distance education. China's Distance Education (Comprehensive Edition) (2002)
10. Chen, L.: On the activation and improvement of students' autonomous learning system in open and distance education. China Distance Education (2002)
11. Zhang, C.: Research and Design of Network Autonomous Learning System Based on Concept Map. Central China Normal University (2007)
12. Wang, J.: English autonomous learning system for postgraduates. Audio Visual Foreign Lang. Teach. 5, 56–58
13. Cai, J., Ruan, X.: OCPA bionic autonomous learning system and its application in robot posture balance control. Pattern Recogn. Artif. Intell. **024**(001), 138–146 (2011)
14. Gu, J.: Construction and application of English audio visual network autonomous learning system. Research on Higher Education of Science and Technology (2006)
15. Wang, H., Wang, Y., Jiao, Y., et al.: Design and implementation of web based personalized autonomous learning system. In: 13th Annual Conference of Computer Aided Education Professional Committee of Chinese Society for artificial intelligence (National CBE Society) (2008)
16. Li, J., Hu, Z.: On the improvement of online autonomous learning system – based on the research of online autonomous learning system of international accounting. Second Half of Entrepreneur World (Theoretical Edition) (2007)
17. Chi, S.: Development and implementation of web based online autonomous learning system. Value Eng. **34**, 183 (2010)
18. Wang, C.: An AI autonomous learning system and method based on quantum theory. cn110969255a (2020)
19. Zhang, S., Que, H., Yu, W.: A flexible riveting autonomous learning system based on visual guidance. cn108480546a (2018)