International Journal of Education and Humanities
ISSN: 2770-6702 | Vol. 13, No. 3, 2024

English Question Library Generation Algorithm Based on Machine Learning

Feiyan Wang*, Yaling Jiang

School of General Education, Nantong Institute of Technology. Nantong 226000, China
* Corresponding author

Abstract: In English learning, the teacher wrote the text of the test source according to their own experience. To reduce the complex and repetitive tasks in teaching, teachers can focus on the preparation, teaching and response of teaching, and further improve their teaching ability. It has received the source of test questions through multiple channels and created the test library quickly and efficiently has become the fast-paced requirement of English writing. Machine learning is a mathematical training method that explores how computers mimic human writing behavior and constantly accumulate experience and improve their performance. This is one of the first and most important cases of AI. The purpose of this book is to read the problem of creating algorithms for English question libraries based on machine learning. This project focuses on the creation of machine-based English question library algorithms, hoping to use the rich information resources on the Internet, to quickly create a large test question library using machine learning technology.

Keywords: Machine Learning, English Question Library, Teaching Resources, Artificial Intelligence.

1. Introduction

In English education, teachers usually build a question bank based on their experience. They have long been engaged in the education of this research, and have accumulated many sources of test questions. If you make full use of the available resources, effective test libraries can be quickly created, thus reducing the complexity and repetitive work in teaching, making teachers focus on academic preparation, teaching and response, and effectively improve their teaching ability. Therefore, quickly storing the existing test sources in the test library can improve the effectiveness of English teaching [1].

In the research of English intelligent response system of intelligent fuzzy decision tree algorithm, many scholars study it and achieve good results, for example, Chuang put forward the concept of template automatic generator, believes that TTA G can automatically learn tree template from training web pages, and TTAG has a good extraction effect [2] in query-based and fast update web pages.Hsu et al. first proposed data mining based on the template method, designing different templates to describe the logical structure of web pages, [3].

This paper studies English question library generation algorithm based on machine learning. The first aspect is to introduce machine learning theory, including basic principles of VC dimension, empirical risk and structural risk; then design English question library system in the form of web client to show the specific functions and manifestations of the question bank model.The English question bank is the implementation of the assessment of students through the computer and the network, and the construction significance has the following points: save human costs, replace the teachers by machines, ensure the fairness and justice of the examination, eliminate human factors, realize the separation of teaching and examination, and realize the standardized of the examination .

2. Research on the Machine Learning-Based English Question Library Generation Algorithm

2.1. Production Algorithm Objectives of English Question Library

The use of the traditional question bank is concentrated in the examination period, and the purpose is to conduct the teaching test and evaluation through the examination. People often ignore a major advantage of the question bank itself: the question bank itself is a very good learning platform[4].The question bank stores the test questions and their answers, which can open up the non-test questions and previous test questions to develop online test functions. The informatization of English question database in primary and secondary schools has the following objectives:

(1) Scientific classification of test questions, the development of online examination function, the first need to classify the test questions, which is conducive to distinguishing the use of the test questions, and the establishment of a special examination function branch on each classification. According to the purpose, the test questions are divided into three major categories[5]: test questions, after-class questions, and ability expansion questions. The examination questions are used for the examination of the group papers, applied to different sizes of the examination, the scheduled midterm final examination, monthly examination, simulation examination, etc. After-class questions are related to the course, taken from the questions of the textbook and the questions written by teachers, which can be used by students after class, and can also be used for teachers to leave homework. According to the requirements of education and teaching, write and expand students a certain aspect of the topic, can be interesting English questions, scene to topic; can be completed by one person, or multiple people, the purpose is to exercise students' practical use of English, oral ability, collaboration ability and

other [6-7].

(2) Rationalization of question bank evaluation. Domestic traditional examinations tend to test students' academic performance, can not comprehensively evaluate students, and put too much emphasis on selection and selection. Through evaluation, we should ensure that students can constantly understand themselves, find themselves, improve themselves, achieve the teaching predetermined goal, and promote the continuous improvement of students' comprehensive quality.

(3) Standardization of question bank construction refers to the standardization of question type and organizational form, and the unified management of test questions. Standardized examination has the characteristics of convenient organization, economic convenience, science and stability. Test questions usually include subjective questions and objective questions, respectively organizing subjective questions and objective questions into volumes for separate printing. The standardization of marking papers can not only facilitate teachers to review and judging papers, but also be conducive to the standardization of tools. The use of standard scores is an important marker of standardized examinations, which can improve the comparability of measurements, make the addition of scores of different subjects more reasonable, and provide more accurate information about [8].

(4) Open maintenance of the question bank, the construction of the question bank is semi-closed, usually by the education department to formulate relevant standards, implemented by teachers or relevant examination departments, open question bank construction refers to the opening of the test questions entry entrance, so that the majority of users to participate in the maintenance of the test questions to [9].

(5) The application of question bank is popularized, and education informatization uses modern scientific and technological means and network means to facilitate knowledge dissemination and sharing. After the standardization of question bank, it can be popularized on the Internet.The popularization of question library is that the standardization of question library is more conducive to the implementation of unified standards; reduce the difference of regional education, students and teachers can use a unified question library; convenient for students to conduct self-test, under the unified standard, students are easier to test their true level [10].

2.2. Application of the English Question Library Generation Algorithm Based on Machine Learning

Stochastic gradient descent SGD is an algorithm used for almost all deep learning algorithms. Loss functions in machine learning algorithms can often be broken down into the sum of the loss function for each sample. For example, the negative conditional log-likelihood of the training data can be written as:

$$J(\theta) = E_{x,y\sim P_{data}} L(x,y,\theta) = \frac{1}{m}\sum_{i=1}^{m} L\left(x^{(i)}, y^{(i)}, \theta\right) \quad (1)$$

Where L is the loss function for each sample

$$L(x,y,\theta) = -\log p(y|x;\theta) \quad (2)$$

For these additive loss functions, the gradient descent needs to be calculated

$$\nabla_\theta J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \nabla_\theta L\left(x^{(i)}, y^{(i)}, \theta\right) \quad (3)$$

The calculation cost of this operation is O (m).As the training program develops, it takes a considerable time to complete a gradual process.Using small samples you can approximate expectations.At each step of the algorithm, a small number of samples were uniformly drawn from the training procedure.The sample size in this paper is always small, ranging from one to several hundred.Basically, this article usually keeps up with the data samples.Slope calculation can be expressed as:

$$g = \frac{1}{m}\nabla_\theta \sum_{i=1}^{m} L\left(x^{(i)}, y^{(i)}, \theta\right) \quad (4)$$

Data from the bulk samples were used.The stochastic gradient descent algorithm then uses the gradient descent estimates as follows

$$\theta \leftarrow \theta - \varepsilon \cdot g \quad (5)$$

Here is the learning rate.

Stochastic algorithms have many different applications.It is an advanced method for training a large number of line models in large databases.For fixed-size models, the statistical values for each step of the random background updates do not depend on the size m of the training set.The time required for the model to reach a consensus usually increases as the training database grows.However, while I wait for infinity, the model finally achieves the best test error for the approximate bottom line described in the training set before each test.It still takes a long time to reach the best test error of the model.The asymptotic cost of training the model with the SGD can be considered as the order of O (1), as a function [11-12] of m.

3. Design Experiment of English Question Library Generation Algorithm Based on Machine Learning

3.1. Web-page Extraction of Data

In different web pages, the English content in the web pages is extracted according to the machine learning technology. First, count the number of test papers in the sample web page. Then, through the web content extraction, the number of correctly extracted papers TP, the number of wrongly extracted papers FP, and the number of unextracted papers FN is analyzed. Finally, the accuracy p and the recall r were calculated using the following expressions.

3.2. Practical Application of the Test Bank

After the initial introduction of the question bank, the number of questions basically met the requirements of the routine and final exams. As the customization continues, be sure to read the specific numbers of questions. This project selects the comprehensive English level 3 exam before the final exam of the semester to predict the students of English

professors in two classes. For prediction issues, this book provides in-depth research and time management. According to the English language course and students' academic conditions, the scores of questions used in English synthesis 3 are as follows: simply choose to evaluate words and words first, about 2 / third of the textbook, the remaining third of the extra-curricular materials. Some lattice questions were matched from after-school exercises, and some from full exam questions. There are three items: multiple-choice questions, answer questions, justice and fallacy questions. Explain both chapters from suggestions on the text and work issues. Both English and English translations examine the keys and difficulties of the text so that students can understand the content of the text. Writing primarily involves topics related to the topics taught, and evaluates students' participatory thinking, subject understanding, and basic writing skills. The test questions are based on textbooks, focusing on the students' basic language ability, analytical ability and critical thinking ability.

4. Experimental Analysis of the English Question Library Generation Algorithm Based on Machine Learning

4.1. Data Extraction

The purpose of this section is to test the effectiveness of the web paper content extraction method based on the template method. According to the body weight rules, the paper weight in Eq. Extract one by one according to the website number, and extract and recall differently according to different question types. Table 1 can be made according to the experimental data. According to Table 1, Figure 1 can be obtained.

Table 1. Sample Web Page Sets (Unit: Sets)

Website Number	Level 4	Level 6	Title English	GRE	PETS	Specialized 4th and 8th
1	27	62	10	80	5	87
2	264	128	428	94	126	153
3	11	13	61	76	49	77
4	28	21	0	1	41	57
5	25	15	266	0	164	367

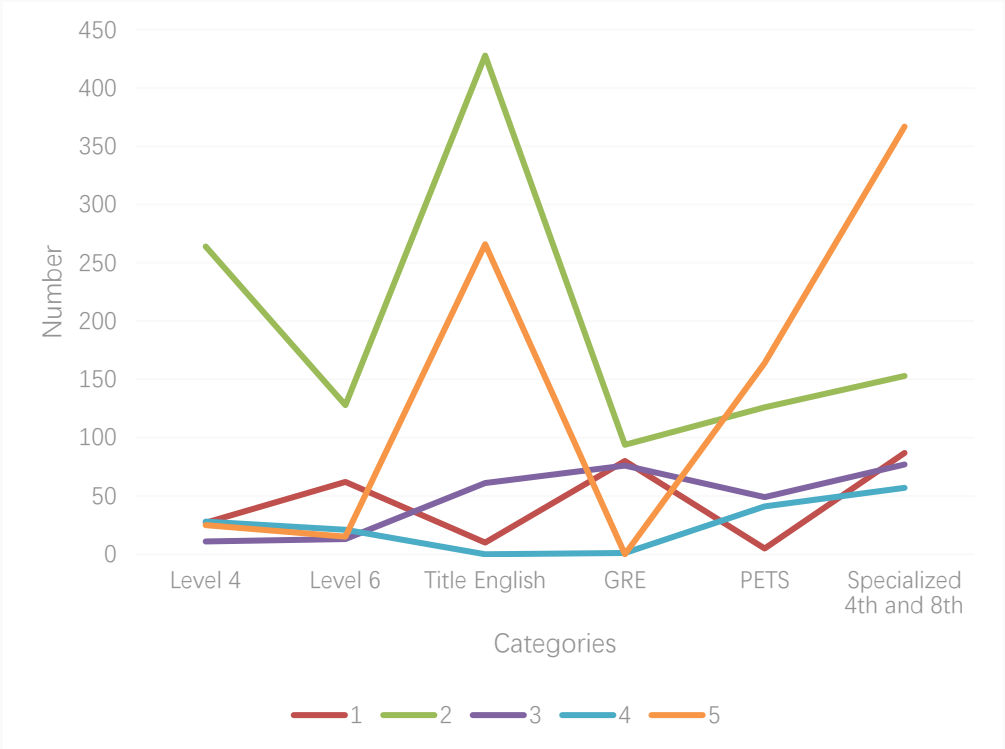

Figure 1. Sample Web Page Sets (Unit: Sets)

In this paper method, the recall rate reached 98.37%, and the accuracy rate reached 98.42%. According to Figure 1, after the machine learning is trimmed with a template, the extraction rules are extracted, which has a high accuracy. However, when the system handles some web pages, it will also appear with incorrect extraction results. This is due to the complexity and heterogeneity of the web pages.

4.2. Application of the Test Question Library

From the selected questions, they improved the English and English translation and interpretation rates, tested the students' understanding of the textbook, and included a fourth major reading question. In the case of the increase of individual questions, the students' answers are good, the exclusion rate is not high, but the content of the examiner expectations, emphasizing the importance of the content and

the contradiction between improving the test results. Although some tests have been added (such as fourth-grade reading questions and answers), student acceptance is realistic and fast. For the questions raised by the students, such as some single choice ambiguity, the question bank group members corrected the questions, and some answers to the English interpretation questions were inappropriate. For example, the students wrote the meaning but used the expression that was different from the original answer, they were deducted to different degrees when marking, and the question bank group also supplemented the answer for this. At the same time, according to the situation of the volume, the question bank group to supplement the topic of writing. The resulting test data are made in Table 2. FIG. 2 can be made according to the data in Table 2.

Table 2. After improvement, the final exam is extracted from the question bank, and the test results of the main exam

Score segment	outstanding	good	secondary	pass a test	fail
number of people	0	6	15	10	1
percentage	0	18.75	46.88	31.55	3.13

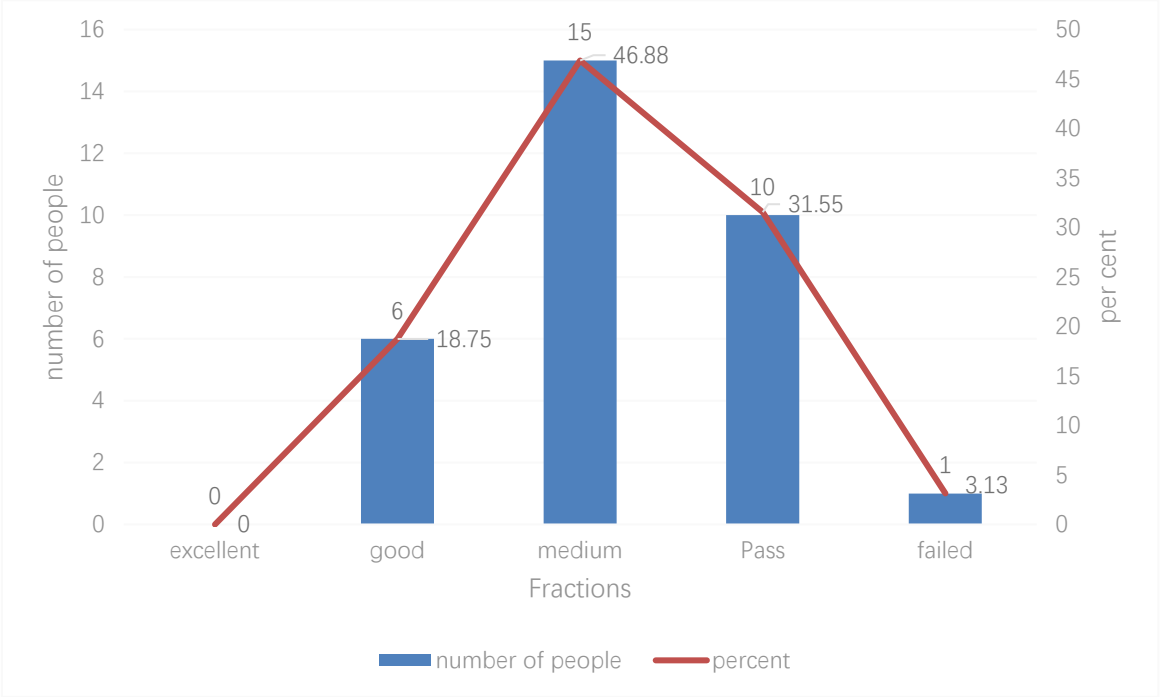

Figure 2. After improvement, the final exam is extracted from the question bank, and the test results of the main exam

As far as Figure 2 shows, compared with the preliminary test, the difficulty and differentiation difference are not big, the stability of the test questions is good, and the reliability is also high. Because before the exam has been told the students to randomly draw the questions from the question bank, so the exam is no longer delimit the specific scope, students pay more attention to the comprehensive review, no longer rely on the teacher to guess the questions, carefully review the knowledge learned, which also has a positive impact on the teaching.

5. Conclusions

This paper studies two core algorithms in machine learning, and proposes the information extraction rules for the information of English test paper. Using this rule, you can accurately extract the test information in the page. As long as the web page to be extracted is similar to the template structure, the extraction rules of this article can be used to quickly extract the target information in the new web page. This process is much simpler and faster than the template generation process, namely bulk web information. The test results show that the English question library generation system has high accuracy and recall. And it can correctly analyze and extract the results, and deposit them in the database according to the standard, eliminating the work of manual labeling. At the same time, it is also found that the system will have incorrect extraction results for some web pages. This is due to the complexity and heterogeneity of the web pages. A lot of Chinese confusion under the Java platform causes errors in web content extraction or a complex hierarchy to capture too deep nodes when trathrough the DOM tree.

References

[1] Xuan W , Zeng F . Design of electro-hydraulic servo loading controlling system based on fuzzy intelligent water drop fusion algorithm[J]. Computers & Electrical Engineering, 2018, 71:485-491.

[2] Sengor I, Chuang, Guner S , Erdinc O . Real-Time Algorithm based Intelligent EV Parking Lot Charging Management

Strategy Providing PLL Type Demand Response Program[J]. IEEE Transactions on Sustainable Energy, 2020, PP(99):1-1.

[3] Yu H , Lu J ,Hsu, Zhang G . [IEEE 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE) - Nanjing (2017.11.24-2017.11.26)] 2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE) - Learning a fuzzy decision tree [J]. 2017:1-7.

[4] Saleh E , Valls A , Moreno A , et al. A Hierarchically _L-Decomposable Fuzzy Measure-Based Approach for Fuzzy Rules Aggregation[J]. International Journal of Uncertainty, Fuzziness, and Knowledge-based Systems, 2019, 27 (Dec.Supplpll1):59-76.

[5] Mahela O P , Shaik A G . Recognition of power quality disturbances using S-transform based ruled decision tree and fuzzy C-means clustering classifiers[J]. Applied Soft Computing, 2017:243-257.

[6] He Z , He Y , Liu F , et al. Big Data-Oriented Product Infant Failure Intelligent Root Cause Identification Using Associated Tree and Fuzzy DEA[J]. IEEE Access, 2019, PP(99):1-1.

[7] Cheraghalipour A , Hajiaghaei-Keshteli M , Paydar M M . Tree Growth Algorithm (TGA): A novel approach for solving optimization problems[J]. Engineering Applications of Artificial Intelligence, 2018, 72(JUN.):393-414.

[8] Ai B , Li B , Gao S , et al. An Intelligent Decision Algorithm for the Generation of Maritime Search and Rescue Emergency Response Plans[J]. IEEE Access, 2019, PP(99):1-1.

[9] R Abcan J , Levashenko V , Zaitseva E , et al. Application of Fuzzy Decision Tree for Signal Classification[J]. IEEE Transactions on Industrial Informatics, 2019:1-1.

[10] Kanimozhi U , Manjula D . An Intelligent Incremental Filtering Feature Selection and Clustering Algorithm for Effective Classification[J]. Intelligent Automation & Soft Computing, 2018, 24(4):701-709.

[11] Mcglinn K , Yuce B , Wicaksono H , et al. Usability evaluation of a web-based tool for supporting holistic building energy management[J]. Automation in Construction, 2017, 84:154-165.

[12] Qiao J , Niu Y , Kifer T . Intelligent optimization algorithm for global convergence of non-convex functions based on improved fuzzy algorithm[J]. Journal of Intelligent & Fuzzy Systems, 2018:1-9.