Research and Design of Online System Based on Subjective Questions Automatic Scoring

Yang-Xia SHU
Jiangxi Institute of Fashion Technology, Nanchang Jiangxi, China
shuyx2002@qq.com

Keywords: Online Test System, Degree Similarity, Automatic, Scoring, Arithmetic.

Abstract. The examinations of college public courses are great challenges for teachers who are responsible for large numbers of college students. With the traditional examination model, teachers need to set questions, set types, invigilate, correct, record scores and add up the score. The complex and various work costs them a lot of energy and spending. Online examination system is a system based on web technology that expands the flexibility of the examination and ensures the safety of the answers. However, at present, the questions in the online examination system are too single and they also need to be graded by teachers manually. Hence, this paper aims to develop an automatic scoring system for online subjective questions in the basic courses tests.

Based on the research of the automatic grading technology at home and abroad, the online examination system realize automatically grading function of subjective questions, combined with Chinese word segmentation technology, keyword extraction algorithm, the statement in the matching processing algorithm, degree theory in fuzzy mathematics thought and algorithm analysis. This research simulates the artificial evaluation steps, and detailing the evaluation process step by step. According to the particularity of natural language, it establishes a synonym thesaurus for keyword matching in order to improve the accuracy of matching. Finally combing the close of the sentence, it calculates score by putting keywords as a scoring point, and the system is feasible through testing.

Introduction

With the development of information technology, modern teaching methods have changed a lot with the existence of the distance teaching platform and online studying system one after another. Due to the increasing number of college students, the teaching burden of the teachers is more and more heavy and the traditional teaching model doesn’t work for the social development. As a result, many online examination systems are constantly developed and they are applied to different types of examinations. But most online examination systems only involve the testing and automatic grading of the objective questions. There are few systems that involve the subjective questions. If any, the subjective questions are scored in the way of manual grading because it is difficult for the computer to realize the automatic scoring of subjective questions. The difficulty of this kind of automatic scoring lies on the complex and abundant features of nature languages, especially the Chinese meanings, making it ineffective to do some researches on it at home. Though the research is difficult, it doesn’t mean that it is meaningless to do this research. So this paper aims to find a best solution for the online examination system by researching relevant materials and studying the system at home and abroad.

Research Background

As a professional institute of fashion technology, Jiangxi Institute of Fashion Technology has a steadily increasing enrollment of students with the development of the school in recent years. The present online examination system is in mature period in grading the objective questions, such as true or false questions, choice questions and gap filling questions. But when it refers to the subjective
questions, such as question tests and essay questions, which do not have the only answer, the system cannot do automatic scoring. So it will combine the automatic scoring of objective questions and manual scoring of subjective questions to add up the scores.

The research of the automatic scoring at abroad is more extensive. In the 1990s, the Electronic Essay Rater (E-Rater) system that was developed by Jill Burstein in the US Education Service Center was used for The Us MBA Entrance Examination and TOEFL. E-Rater mainly combines IR and NLP to develop the system and analyze the article with the natural language processing kit. According to the statistical tests, the accuracy of grading in this system is rather high and is up to 97 percent. At the beginning of 21 century, Conceptual Rater (C-Rater) system that was developed by E-Rater group synthesized the E-Rater technology and NLP to realize the function of automatic scoring by extracting the key words of reference versions and matching the answers of the students with the reference versions. Because this kind of system is relevant to the understanding of natural language and the complexity of natural language conditions, the accuracy of it is not very high based on the statistic. Someone once used this system to do a test of 170000 essay questions and found the accuracy was just 85 percent.

Based on the research of automatic scoring of subjective questions, people at abroad classify it into two major types: one is the automatic scoring of shot passages, such as explanations of nouns and question tests. This kind of grading does similarity comparison of the answers of students with reference versions to calculate the corresponding grades by the comparison. The other is the automatic scoring of long passages that means to grade the article. This kind of questions needs to extract the important information of the reference versions and then to do the scoring according to the semantic situations. Also, it can give some prepared answers to match to calculate the scores the students get based on the matching degree.

**Methods of the Automatic Scoring of Subjective Questions**

Due to the fuzzy and multiple features of the subjective questions and the difficulty of understanding the natural language of artificial intelligence technology, there are a lot of problems of the online automatic scoring of subjective questions to be solved. By now, the researchers and the scholars have done many relevant researches on it. The followings are the theoretical results:

**Calculating the Similarity of the Sentence by VSM**

The calculation adds up the scores of students with the statistic of weight value by considering the frequency of the key words in the sentences, analyzes the feature vector of the key word, and calculates the similarity of the words by making use of the similarity vector. This kind of method has its limit. It doesn’t work for the longer sentences and its accuracy is not too high for the semantics. Meanwhile, the similar meanings cannot be identified and it cannot judge the place of the key word, so the weight value is not standard.

**Calculating the Similarity by the Resources of Hownet**

The method makes use of the resources of Hownet based on the world knowledge to calculate the similarity with the similarity of the meanings of two words. This method shows us the abundance of the Chinese semantic, but due to the complexity of the words’ definition, it makes it inaccurate to calculate the similarity.

**Basing on the Weighted Semantic Similarity Model**

On the basis of traditional similarity model of word frequency, it can calculate the semantic similarity between the questions of computer users and those in the database by calculating the semantic similarity with the length and the height of semantic tree.
Methods of Dynamic Planning

Build up the key word matrix with the methods of dynamic planning to calculate the sentence similarity in the research of the best way in the matrix. Then it will finish calculating the similarity of the key word matching and statement matching.

String Matching Algorithm of One-Way Inclination Extent

The algorithm discloses the sentences in the form of string and matches them according to the sequence of characters in the sentences. So it can get the inclination extent. If the inclination extent is greater than or equal to the full marks, it will be equal to the full marks; otherwise it will be what it is.

The Progresses of Automatic Scoring of Subjective Questions

The progresses of automatic scoring of subjective questions can be divided into the following steps:

Preprocessing of Clauses

Processing of clauses divides a text into many sentences with a critical limitation of punctuation. Generally, it regards full stop, exclamatory mark and question mark as end marks of clauses. But this processing has difficulty in matching later. In order to reduce the difficulty of processing answers and shortening the length of clauses, we also recognize comma, semicolon and colon as the end mark. Finally, the processing of dividing the sentences into clauses is finished on the basis of dividing the students’ answers with the mentioned punctuations.

Preprocessing of Participle

In Chinese, participles refer to countless singer meaningful words that are divided in the coherent characters sequence. At present, algorithm of Chinese clauses can be divided into three categories: the algorithm of clauses based on the matching of strings, the algorithm of clauses based on the statistics and the algorithm of clauses based on understanding.

Processing of Key Words

When correcting the answers of students, teachers will find out the key words at the first glance of the answers because key words are the important points of scoring. The computer makes the judgment by the key words when matching the students’ answers with the reference versions. According to the meanings of extracting key words, we classify key words into two types: one is the general key word; the other is the necessary key word. General Key words are the key words of the text rather than the point scores while the necessary key words are the essential parts and are the point scores of the students’ answers. Hence, when establishing the answer bank, it is necessary to make annotation that needed to be commented in the rating scales.

Processing of Matching

At present, the arithmetic of the similarity of matching Chinese texts can be divided into three types: the first is the arithmetic based on semantic information, for example the calculating methods by using Hownet and Cilin to match; the second is the arithmetic based on the semantic structures of sentences, for example the calculating methods which are based on the dependent semantic similarity; the third is a very traditional calculating methods, that is the matching arithmetic based on the key words, for example the calculating methods based on the vector space model.

Calculating Inclination / Calculation of Inclination

According to the results of matching students’ answers with the standard answers, it is available to calculate the similarity of each sentence by applying to the inclination arithmetic of fuzzy mathematics.
Dealing with Scoring

Multiply the similarity of each sentence with the corresponding weighted value to add up the scores of each sentence, and it is the final score of the student.

The Design of the Arithmetic of Automatic Scoring of Subjective Questions

The Design of the Arithmetic of Clauses

This paper mainly recognizes semicolon, colon, question mark and full stop as the separator of the clause.

The progress of reducing the CRLF with the use of preprocessing the text:

Set s1 to answer string, char to variable characters
(1) Save the first character of s1 into char;
(2) Judge if char is CRLF, if so, removes the character off s1 and keep operating (1); otherwise, it illustrates that there is no empty line before the string, and then keep going on (3);
(3) Save the last character of s1 into char and judge if there is a CRLF in char. If so, remove the character off s1 and keep operating (3); otherwise, it illustrates that there is no empty line before the string, and then keep going on (4);
(4) Detect if there is empty line before or after s1 and return and save s1.

The processing of the clauses of the text:
Set s1 to answer string and set char to variable characters
(5) Read s1 and judge if it is empty, if so, end operating, otherwise turn to (6);
(6) Read the first character of s1 and save it into char, judge if char is the last character, if so, turn to (7); otherwise turn to (5);
(7) If char is the ending punctuation, mark the clause array to +1 and delete the punctuation from s1;
(8) If char is CRLF, delete the first character of s1 and turn to (5);
(9) Merge char into clause array, delete the first character of s1 and turn to (5).

Participle and Part-of-Speech Tagging Technology

Nowadays, there is not a participle system that has the accuracy up to 99.9 percent but the rate of ICTCLAS3.0 participle is up to 999.63KB/S and the accuracy of participle is up to 96.56%. It can realize the function of Chinese-English compound participle. ICTCLAS uses C/C++ to do programming and to support various developing languages that make it is in a higher level. This research is to do the automatic scoring to the test papers of students, and the accurate scores are the most important part of this system. So it is necessary to choose the good participle software and choose more accurate ICTCLAS system when selecting the participles for the text.

Design of Arithmetic of Extracting the Key Words

Extracting the key words is based on analyzing sentences, and it only takes the words which can express the meanings of the sentence. These words are mainly content words including nouns and verbs. The arithmetic of extracting key words:
(1) Flip through the answer text of the student, if it is the end of the word, end operating, otherwise turn to (2);
(2) Judge the part-of-speech of the word, if it is verb or noun then turn to (3); otherwise turn to (!);
(3) Extract the word and turn to (1).

Design of Arithmetic of Preprocessing the Matching of Sentences

Based on the sentence matching methods, the followings are the matching arithmetic:
(1) Match the key words of students’ answers with those of reference version. If it can be find the corresponding in students’ answers, turn to (2), otherwise end operating;
(2) Judge if the corresponding is up to 60 percent, and turn to (3), otherwise turn to (4);
(3) Calculate the degree of coverage in the matrix and choose to match the most key words, turn to (4);
(4) Mark the matched clauses and do the +1 processing to the frequency of the matched sentence. Finish a matching progress, and see if there is any clause that isn’t matched, turn to (1) and do the next matching, or end operating.

**Design of Calculating the Inclination Degree**

The followings are the design of inclination arithmetic

1. Fill in the synonyms of the key words in the reference version;
2. Extract the key words of students’ answers and see if they are the score point, if they are and then turn to (5), otherwise there is no score and turn to (3);
3. Compare the key words of reference version with those of students’ answers, and judge if the key words of reference version are Chinese, and if they are, turn to (4), otherwise turn to (5);
4. Deal with Chinese with double byte, turn to (6);
5. Deal with non-Chinese with single byte, turn to (6);
6. Strictly calculate the inclination of students’ answers and reference version and get the scores of questions of students. If the scores of questions are greater or equal to the full mark, the scores are full; otherwise it will be marked down.

Combined with the above algorithm design and development of this system using the. Net three-tier architecture, and then all function modules are divided into three levels: The first is user experience layer, its main function is to provide friendly operation interface in the top architecture; The second is the business logic layer, it is as a form of data conversion and to achieve the mutual change between user’s data and database data in the middle layer of architecture; The third is data access layer, it is mainly to perform data management operations of database in the bottom layer of architecture, including data increase, data delete, data modify, data query and so on.

**Conclusion**

Although this system can realize automatic scoring of subjective questions, the accuracy of scoring needs to be further tested. The understanding degree of computer to natural language decides its realization of the system. It is because of the complexity of natural language that the system cannot be used for some difficult technology which needs to be further researched:

The system considers the weight of key words and sentences when scoring, while teachers speculate weight with the certain standard. So the scores may be influenced by teachers’ experience. Above all, the scoring standard needs to be rationalized.

Preprocessing the matching sentences with key words cannot take the rationality of the sentence and the sequence into consideration, and the matching results are not ideal. So it is necessary to find a more intelligent solution.

There are a lot of methods of calculating including the technology of dealing with the natural language by the arithmetic in Math. It is important to combine a variety of arithmetic to improve the modern technology when designing the system.

It needs to do a thorough test to increase the accuracy, so the system can be applied to teaching practice.

**References**


