A Classification Technique for Answering Multi-choice World History Questions FRDC_QA at NTCIR-11 QA-Lab Task

Shuangyong Song, Yao Meng, Zhongguang Zheng, Jun Sun
{shuangyong.song, mengyao, zhengzhg, sunjun}@cn.fujitsu.com
Fujitsu R&D Center Co., Ltd.

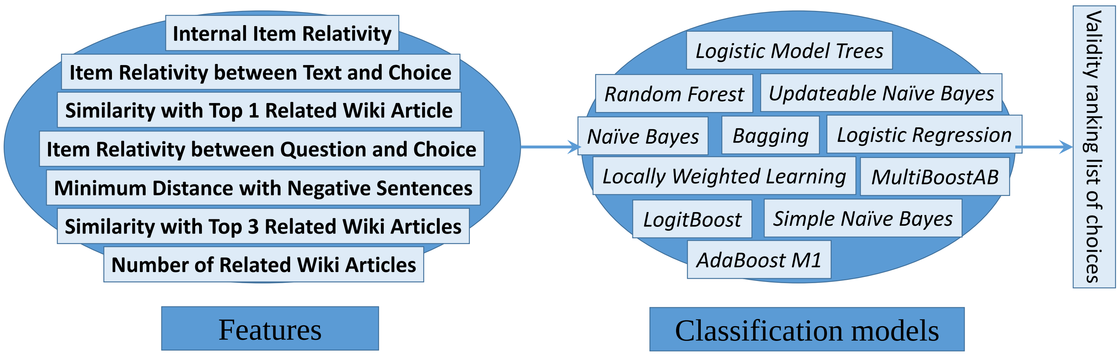
Introduction

■ Our FRDC_QA team participated in the QA-Lab English subtask of the NTCIR-11. In this paper, we describe our system for solving real-world university entrance exam questions, which are related to world history. Wikipedia is used as the main external resource for our system.

■ Since problems with choosing right/wrong sentence from multiple sentence choices account for about two-thirds of the total, we individually design a classification based model for solving this type of questions. For other types of questions, we also design some simple methods. Descriptions of methods, experimental results and conclusions are given in next sections.

Methods

➢ Framework for questions with multiple sentence choices



Features

Classification models

◆ Each classifier can get an accuracy probability for each choice, and the average value of the accuracy probability from all classifiers will be taken as the final accuracy probability of a choice.
◆ If the question is asking us to choose the right choice with the keywords 'correct', 'correctly' or 'appropriate', we choose the choice with highest accuracy probability as the final answer.
◆ If the question is asking us to choose the wrong choice with the keywords 'incorrect', 'incorrectly' or 'mistake', we choose the choice with lowest accuracy probability as the final answer.

➢ Frameworks for other types of questions

Questions with chronological sequence

We utilize the 'Lucene Index of Item Time' to search timestamp of each event in the choices, and rank them with the chronological order, then we can choose the right answer according to this order easily.

Questions with Term Choices

We detect items contained in background text and the question with Maximum Matching Method, then calculate the relativity between those items and the choice item. Finally, the choice with highest relativity will be chosen.

Questions with judging true/false sentences

We use the same features in above section to train SVM classification model to handle this type of questions by directly output the 'true of false' result of each choice instead of the accuracy probability.

Other types of questions

We choose final answer with the random selection method for other types of questions, which need image analysis technology. In particular, we set a specified random seed to keep the stability of the results given by our system.

Results

Types of questions	Number of correct answer / Total number	Score of correct answer / Total score
Type 1 questions	10/23	28/62
Type 2 questions	0/0	0/0
Type 3 questions	3/7	9/20
Type 4 questions	0/0	0/0
Other questions	0/7	0/18
Total	**13/36**	**37/100**

Table 1. Evaluation results of our system in phase 1 English task

■ For types 'Questions with multiple sentence choices' and 'Questions with term choices (without images)', we achieve a precision of about 45% on both 'Number of correct answer' and 'Score of correct answer', which shows the much better effectiveness than random method, since we think the precision of random method should be 25% on four-choice questions. However, the real result of random method on 'other types of questions' is not as good as our thought. We got wrong answers on all the seven 'other types of questions' with the random method, which makes our total result getting a precision of 37%, far below the 45%.

Future work

Several attempts can be tried to improve the system performance in our future work:
◆ more useful external resources can be utilized, such as query results from Google like search engines, etc.
◆ more reasonable and intelligent combination way for different classification models should be tried;
◆ different writing styles for timestamps, locations and personal names should be considered.

Furthermore, a unified domain insensitive system for choosing wrong/right answer from multiple sentence choice will be a trial in our future work.