Journal of Artificial Intelligence Practice (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/jaip.2022.050407
ISSN 2371-8412 Vol. 5 Num. 4

An Automated English Translation Judging System Based on Feature Extraction Algorithm

Ruichao Li*

School of Translation Studies, Xi'an Fanyi University, Xi'an, Shaanxi, China

280663560@qq.com

**Corresponding author*

Keywords: Feature Extraction Algorithm, English Translation, Automatic Trading, and Evaluation Criteria

Abstract: As an important and necessary element in the assessment of students' English language ability, translation is a comprehensive indicator of students' mastery and ability to use English vocabulary, sentence structure, grammar, and other indicators. Compared to other types of questions, the task of marking translation is more demanding and time-consuming, and the objectivity and fairness of marking are more difficult to ensure because the assessment criteria for translation are relatively flexible. Feature extraction algorithms, which have developed rapidly and been widely used in recent years, can learn and extract development patterns from information such as images and texts, and make judgments in the corresponding context, while English translation(ET) texts have the characteristics of diverse and quantifiable feature points related to grading, and the feasibility of automatic grading exists. Therefore, this paper carries out the design and implementation of a relevant automatic RATING system based on feature extraction algorithms, with a view to improving the fairness, accuracy and efficiency of translation rating.

1. Introduction

The English examination is also an important examination in China's basic education system with a large number of participants, in which the marking of ETs is time-critical and tasking. In order to reduce the workload of the marking teachers and improve the efficiency and fairness of the marking, this paper designs and implements an automatic marking system for ETs in Chinese.

With the development of natural language processing technology, the features extracted from translated texts can basically cover grammar, lexis, and other features that directly reflect the translation content. Some scholars have pointed out in their research that ET scoring mostly considers words and grammar, and the consideration of semantics is extremely limited, so two attributes, semantic coherence and consistency, are proposed to enrich the feature extraction of translation [1]. The process of classification is to extract the features that reflect the quality of the translation, and then classify the translation into several different human grades according to these features, after which the manually annotated data samples are used as the training set to train the classification model of the translation. Since then, many scholars have conducted research on the basis of scoring systems and used various classification techniques for automatic scoring of translations, with certain achievements [2-3].

In this paper, we firstly describe the features of TEXTS and the features of translation scoring, then construct the automatic scoring model and the functional modules of the rating system, then analyse the implementation process of the login module, and finally analyse the practical English text scoring application of the system and compare the effects of different feature groups on the scoring results.

2. Characteristic Analysis

2.1 Textual Characteristics of ETs

Lexical features: the form of words in English is relatively strict, for example, the singular and plural forms of nouns have clear usage scenarios.

Phraseological features: there are many forms of English phrases, such as verb phrases and noun phrases, which have a fixed form of collocation, in that through phrases form part of the translated long sentences.

Syntactic features: The basic sentence structure is subject-predicate, but there are also more complex subordinate clauses in English to express rich content, and the existence of subordinate clauses is also more flexible in form.

Structural features: the articulation between sentences and sentences and between paragraphs in ET is the basis for the coherence of the translated text, and coherence is an important prerequisite for a well-structured text[4].

2.2 Translation Scoring Features

In the actual marking environment, the traditional translation assessment standard “*Xin Da Ya*” is often concretized, refined, and modified to a certain extent to form the manual marking standard for translation questions in the examination, in order to meet the marking requirements and achieve the purpose of assessing the candidates’ language mastery ability [5]. In terms of accuracy, in the manual marking, if the translation to be evaluated has a high degree of literal overlap with the reference translation, the accuracy of the translation can be judged to be high, while the accuracy of translation of key phrases and collocations and the semantics of the translation will also be combined to judge the accuracy of translation. From a comprehensive point of view, the manual evaluation points of translation accuracy include three points: literal overlap, semantic accuracy, and keyword translation accuracy, and the analysis and design of automatic marking features are carried out for these three points [6-7].

(1) In terms of literal overlap, since the N tuples formed after the text is divided and intercepted contain word order and combination forms, which are more representative of the structure and meaning of the translation than individual words or phrases, the N tuples matching rate between the translation to be evaluated and the reference translation is used instead of a single word or phrase matching rate to quantify the literal overlap between the two [8].

(2) In terms of semantic accuracy, latent semantic similarity improves the traditional vector space by mathematically extracting the shallow semantics of sentences. The word vector space constructed according to word2vec mines the deeper meanings of words and interword connections, and the resulting sentence vectors have more depth in terms of semantic representation, based on which the semantic relations of sentences are represented using vector distance. In view of the importance of semantic accuracy in the evaluation of translation quality, it is considered that both shallow and deep semantic levels are examined, i.e., the potential semantic similarity and textual

distance between the translation to be evaluated and the reference translation are used to assess the semantic accuracy of the translation at the same time [9-10].

(3) As the manual evaluation of the accuracy of the translation of key phrases also involves the understanding of the meaning of words, it is clearly not reasonable to judge the correctness of the translation of key words only from the word side when automatic scoring is carried out, so the semantic similarity of key words is included in the evaluation criteria. Combined with the English-Chinese dictionary, the average translation accuracy of keywords is calculated according to the word2vec word vector space equation (1), where N is the number of keywords.

$$keyTrans=\frac{\sum_{n=1}^{N} accuracy_n}{N} \tag{1}$$

3. System Design Based on Feature Extraction Algorithm

3.1 Automatic Scoring Model Construction for ET

For the linguistic features of the translation, manual extraction is adopted to obtain information about the translated text in terms of vocabulary, sentences, grammar, etc. There are two ways of English language text preprocessing: one is the rule-based method, using the rules of human language to process the text, but due to the complicated rules of human language, among which there are also such exceptions as polysyllabic words, polysemous words and colloquialisms, this method is time-consuming and laborious, and has been rarely used; the other is the statistical-based method, firstly, the existing corpus is processed manually, and then the processed corpus is Statistical analysis is carried out and new texts are processed automatically according to the statistical results [11-12].

Since whether a translation is too long or too short is relative to the length of the reference translation, the ratio of the average length of the translation to be evaluated to that of the reference translation is used to calculate the effect of length on the rating, as shown in equation (2), where N is the number of reference answers. It is important to note that in the case of multiple reference translations, the difference in length is minimal as the only difference between the reference translations is that there are multiple approximate translations of individual word combinations.

$$lenRatio=\frac{len_{candidate}}{\frac{\sum_{n=1}^{N} len_{reference_n}}{N}} \tag{2}$$

3.2 Functional Module Design

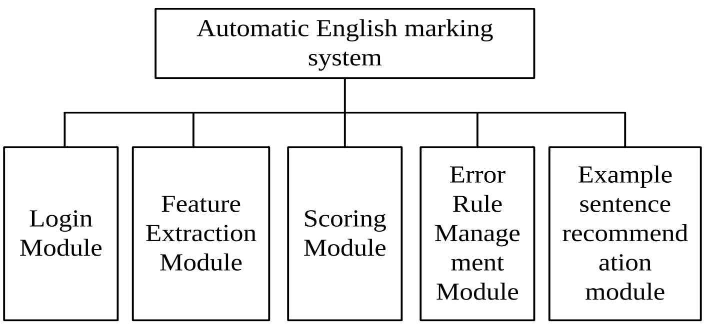

Figure 1: Functional modules of the system

The functional modules of the system in this paper are shown in Figure 1 and are described as follows.

(1) Login module

This module manages the login method and password, through which the user obtains the system usage rights and can change the password.

(2) Feature extraction module for translated text

When candidates have a large vocabulary, the words in their translations tend to be more diverse and spelling is rarely wrong, so features such as the number of words and the average word length are defined as the main features to examine the vocabulary of translation; considering that the better the candidates' translation level is, the richer their sentence types and sentence contents will be, when extracting sentence features, the main features to be examined are the length of sentences, the number of sentences and the use of complex sentences In the feature extraction, the number of sentences in the translation is reflected by the number of punctuation marks such as full stops, question marks and exclamation marks, and the complexity of the translated sentences is reflected by the number of commas and the number of inverted commas. In order to better assess the quality of the translated text, the translated text is transformed into a vector to facilitate a comprehensive examination of the translation content, while the semantic features of the translated text are extracted through the n-gram; the translation structure usually takes into account the context of the translated text and the writing style, and usually translations with a clear central intention and rich content have a more rigorous structure, so the translation structure features are mainly derived from the translation content features laterally The structural features of the translation are therefore mainly reflected in the content features of the translation.

(3) Marking Module

When assessing an ET, English teachers focus on the clarity of the structure of the translation, the correct use of vocabulary, the fluency of the language, and the neatness of the handwriting. If the translated text is obtained by manual entry, there is no need to disregard the neatness of the handwriting. As it is difficult for the program to determine whether the content of the translation is relevant to the translation requirements, it is assumed that the final score of the ET is only related to the quality of the vocabulary, the quality of the phrasing and the quality of the structure, but the user can judge the quality of the content of the text by the subject words extracted by the system.

(4) Error rule management module

This module is used to determine whether there is a grammatical error in the input English statement, and if it matches a certain error rule, then the statement is judged to have a grammatical error. In order to ensure that the error rules in the rule base can accurately identify incorrect English statements, all error rules in the system are written under the guidance of professional English teachers.

(5) Example sentence recommendation module

The users of the system can be either teachers or students. Teachers can use the system as an aid to evaluate students' translation scores; students can use the system to check for problems in their translations and improve their translation based on the revision suggestions given back to them by the system. If the system checks that a sentence has a grammatical error, it will recommend 3 correct statements as a reference.

3.3 Database Design

The core entities in the system are divided into users, questions, candidates, responses, and scoring datasets for model training, while the entity attributes and interentity relationships are extracted. This is mainly due to the fact that in the actual translation scoring, all scoring scores are uniformly scored by the same user using the system to meet the scoring requirements, and there is no need for multiple users to handle different exam scoring at the same time. The model training dataset entities are not really linked to other entities such as exams and students.

4. System Implementation and Application

4.1 Implementation of the Login Function

The user enters the system, enters the user name and password, and then clicks "Login", the background will verify the user name and password data submitted by the user, if the verification fails, an error message will be returned and the page will show the message "User name or password error". The user will need to reenter the correct username and password and resubmit the information. If the verification is successful, the system will be entered and the home page will be displayed, which gives the general operation process for automatic marking of a translation test, as shown in Figure 2.

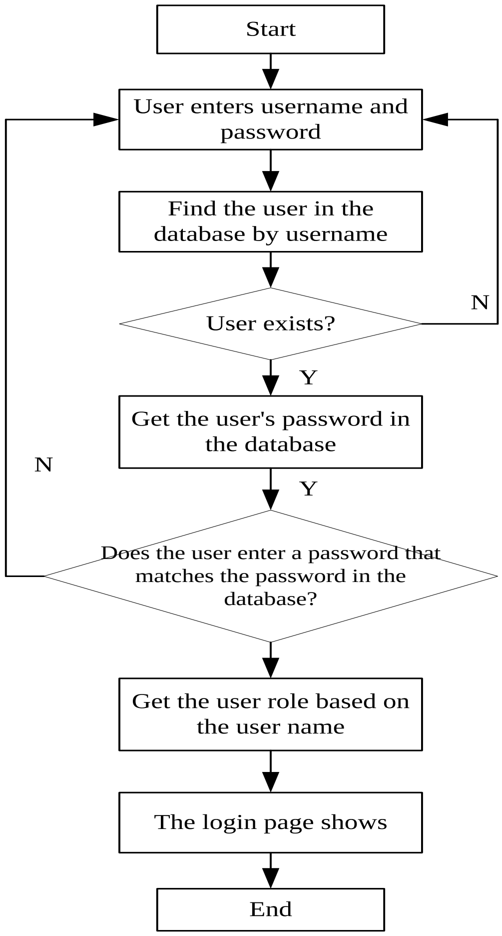

Figure 2: Flow chart of the implementation of the login module

4.2 Results and Analysis of Automatic Scoring of ETs

The criteria commonly used in automatic translation scoring to assess the merits of automatic scoring models for translation are Pearson's correlation coefficient and quadratic weighted kappa (QWK). The QWK is therefore also used in the experiments to assess the agreement between manual scoring and the automatic scoring system scoring, and takes a value between 0 and 1, with 0 indicating complete inconsistency and 1 indicating complete agreement.

In order to better compare the impact of different feature items on the scoring results, the experiment was divided into four data sets and QWK values were obtained for each data set on three sets of translated texts. The first group contains all features of the translated text, the second group contains semantic, global, and local features, the third group contains linguistic and global features, and the fourth group contains linguistic and local features. The experimental results of the four groups are shown in Table 1 and Figure 3.

Table 1: Experimental results

	Text 1QWK	Text 2QWK	Text 3QWK	Average
Group 1	0.823	0.805	0.816	0.8147
Group 2	0.766	0.814	0.794	0.7913
Group 3	0.817	0.776	0.798	0.7970
Group 4	0.804	0.783	0.791	0.7927

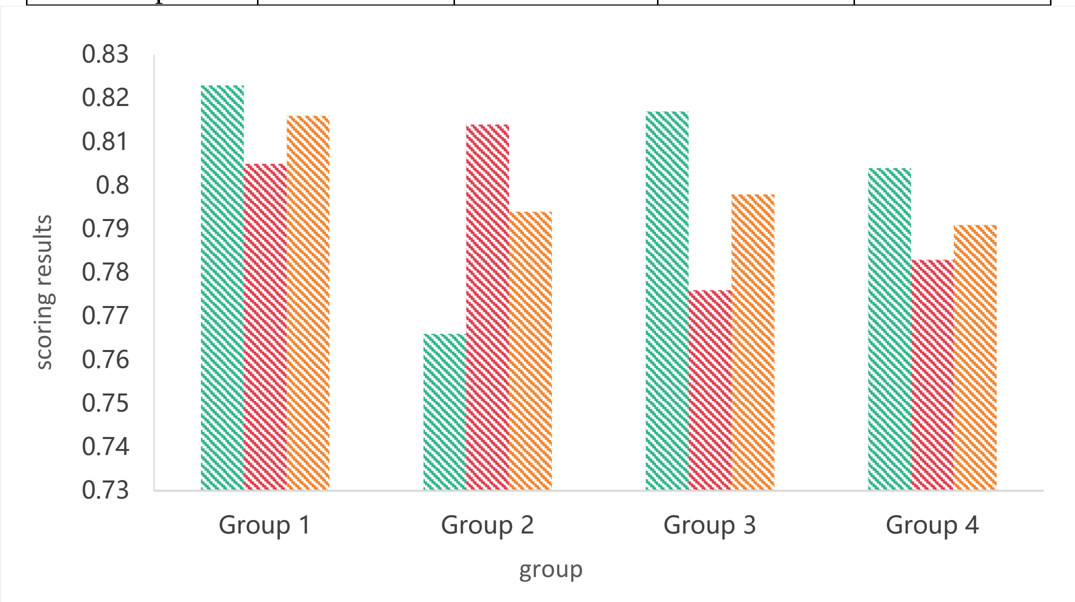

Figure 3: Automatic scoring results

Analyzing the experimental results, it can be seen that the QWK values of all four groups achieved a high degree of consistency in scoring, confirming the validity of the scoring model. The average QWK values of the first group in the three data sets were higher than those of the other three groups, indicating that all three feature terms played a role in the scoring process. However, the difference between the mean QWK values of all four groups was not significant. Therefore, the majority of students in the experimental dataset were uniform in their mastery of linguistic skills and content expression, and translations with good language quality tend to have less poor content quality. Therefore, the difference between the scores of the experimental group with only some of the features extracted and the first group is not significant. However, a more comprehensive feature extraction can help to identify translations that imitate good translations but are not at the same level of overall translation ability, and if the proportion of such compositions in the data set increases, the difference in the scoring results between the four experimental groups will increase accordingly.

5. Conclusion

In English language teaching, ET can better reflect a student's overall mastery of the language, but the open-ended nature of ET texts and the subjective nature of markers in the marking process, such as mood and perception, make it difficult to ensure consistency in marking criteria. The automated marking system for ET designed in this paper, however, is fast and efficient, and can greatly reduce the time teachers spend on marking ETs, allowing teachers to spend more time on teaching their students, as well as allowing English learners to train their ET skills in a targeted manner, while applying it to large scale examinations can greatly reduce human and material resources and increase efficiency, while ensuring the fairness of marking.

Acknowledgements

This work was supported by the 2020 Annual Project of the 13th Five-Year-Plan for Education and Science in Shaanxi Province (Number: SGH20Y1509) and the research project of Xi'an Fanyi University (2021HZ-841).

References

[1] Rajitha Jasmine Rajappan, Thyagharajan Kondampatti Kandaswamy. (2022) A Composite Framework of Deep Multiple View Human Joints Feature Extraction and Selection Strategy with Hybrid Adaptive Sunflower Optimization-whale Optimization Algorithm for Human action Recognition in Video Sequences. Comput. Intell, 2, 366-396.

[2] Asma Naseer, Sarmad Hussain, Kashif Zafar, Ayesha Khan. (2022) A Novel Normal to Tangent Line (NTL) Algorithm for Scale Invariant Feature Extraction for Urdu OCR. Int. J. Document Anal. Recognit,1, 51-66.

[3] Lata Jaywant Sankpal, Suhas Haribhau Patil; Squiride Rank. (2022) Squirrel Ride Rank Algorithm-based Feature Extraction for Re-ranking of Web Pages. Int. J. Web Portals, 1, 1-23.

[4] Hassan Habib, Rashid Amin, Bilal Ahmed, Abdul Hannan. (2022) Hybrid Algorithms for Brain Tumor Segmentation, Classification and Feature Extraction. J. Ambient Intell. Humaniz. Comput, 5, 2763-2784.

[5] Najme Mansour, Gholam Reza Khayati, Behnam Mohammad Hasani Zade, SeyedMohammad Javad Khorasani, Roya Kafi Hernashki.(2022) A New Feature Extraction Technique Based on Improved Owl Search Algorithm: a Case Study in Copper Electrorefining Plant. Neural Comput. Appl, 10, 7749-7814.

[6] Kumod Kumar Gupta, Ritu Vijay, Pallavi Pahadiya, Shivani Saxena. (2022) Use of novel Thermography Features of Extraction and Different Artificial Neural Network Algorithms in Breast Cancer Screening. Wirel. Pers. Commun, 1,495-524.

[7] Nadezhda S. Lagutina, Ksenia V. Lagutina, Elena 1. Boychuk, Inna A. Vorontsova, llya V. Paramonov. (2020) Automated Rhythmic Device Search in Literary Texts Applied to Comparing Original and Translated Texts as Exemplified by English to Russian Translations, Autom. Control. Comput. Sci, 7,697-711.

[8] Alexander Zabolotskikh, Anna Zabolotskikh, Tatiana Dugina, Daria Tavberidze. (2021) Creating Individual Learning Paths in the Moodle Plugin for Undergraduate Students to Study English Grammar. Educ. Inf. Technol, 1,617-637.

[9] Ivana Simonova. (2019) Blended Approach to Learning and Practising English Grammar with Technical and Foreign Language University Students: Comparative Study. J. Comput. High, Educ, 2, 249-272.

[10] Ummadi Reddy, B. Venkata Ramana Reddy, B. Eswara Reddy. (2021) LESH- Feature Extraction and Cognitive Machine Learning Techniques for Recognitionof Lung Cancer Cells. Int. J. Comput, Aided Eng, Technol, 1, 32-45.

[11] P. Ajitha, A. Sivasangari, R. Immanuel Rajkumar, S. Poonguzhali. (2021) Design of Text Sentiment Analysis Tool Using Feature Extraction Based on Fusing Machine Learning Algorithms. J. Intell. Fuzzy Syst, 4, 6375-6383.

[12] Bhawna Ahuja, Virendra P. Vishwakarma. (2021) Deterministic Multikernel Extreme Learning Machine with Fuzzy Feature Extraction for Pattern Classification, Multim, Tools Appl, 21, 32423-32447.