

The neural network algorithm-based quality assessment method for university English translation

Min Gong

To cite this article: Min Gong (24 Apr 2024): The neural network algorithm-based quality assessment method for university English translation, Network: Computation in Neural Systems, DOI: [10.1080/0954898X.2024.2338446](https://doi.org/10.1080/0954898X.2024.2338446)

To link to this article: <https://doi.org/10.1080/0954898X.2024.2338446>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 24 Apr 2024.



[Submit your article to this journal](#)



Article views: 149



[View related articles](#)



[View Crossmark data](#)

The neural network algorithm-based quality assessment method for university English translation

Min Gong

School of Foreign Languages, Nanchang Institute of Technology, Nanchang, Jiangxi, China

ABSTRACT

These results highlight the transformative potential of neural network algorithms in providing consistency and transparency while reducing the inherent subjectivity in human evaluations, revolutionizing translation quality assessment in academia. The findings have significant implications for academia, as reliable translation quality evaluations are crucial for fostering cross-cultural knowledge exchange. However, challenges such as domain-specific adaptation require further investigation to improve and maximize the effectiveness of this novel approach, ultimately enhancing the accessibility of academic content and promoting global academic discourse. The proposed method involves using neural network algorithms for assessing college-level English translation quality, starting with data collection and preparation, developing a neural network model, and evaluating its performance using human assessment as a benchmark. The study employed both human evaluators and a neural network model to assess the quality of translated academic papers, revealing a strong correlation (0.84) between human and model assessments. These findings suggest the model's potential to enhance translation quality in academic settings, though additional research is needed to address certain limitations. The results show that the Neural Network-Based Model achieved higher scores in accuracy, precision, F-measure, and recall compared to Traditional Manual Evaluation and Partial Automated Model, indicating its superior performance in evaluating translation quality.

ARTICLE HISTORY



Received 10 January 2024
Revised 15 March 2024
Accepted 29 March 2024

KEYWORDS

Translation
quality assessment;
neural network;
translation;
communication;
knowledge exchange;
accuracy;
assessment;
machine learning

Introduction

The interchange of knowledge, ideas, and information across linguistic barriers is made possible by the dynamic and varied discipline of translation, which stands at the intersection of cultures and languages. The function of translation has grown essential in an age marked by globalization and the quick spread of information, particularly in the academic

CONTACT Min Gong  nitongmin@163.com  School of Foreign Languages, Nanchang Institute of Technology, Nanchang, Jiangxi, China

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

setting. The cornerstone of successful cross-cultural communication and knowledge transfer in the academic setting is high-quality translation, which guarantees that research findings, scholarly resources, and educational content can be accessed and understood by a global audience (Jiang et al. 2020; Sun et al. 2021). It is impossible to emphasize the significance of excellent translation in academia. To make their work accessible to students, academics, and professionals who might not speak the same native language as the original writers, universities and research institutes all over the world rely on translation (Jiang et al. 2020; Sun et al. 2021; Yuan and Lv 2021). For instance, to reach a wider worldwide audience and contribute to the global discourse in a particular field, a groundbreaking research paper produced in Mandarin Chinese must be accurately and eloquently translated into English. However, verifying the calibre of translations poses a difficult challenge, particularly at the university level. While following the rules of the target language and remaining faithful to the original content, translators must not only express the literal meaning of the source text but also its nuances, style, and tone. Additionally, evaluating translation quality in academics has historically relied on subjective human review, frequently carried out by knowledgeable linguists or translators. Although these analyses provide insightful information, they take a lot of time, and money, and are prone to bias (Zhuang et al. 2019; Bi et al. 2020; Yuan and Lv 2021).

This research article investigates a novel method for judging the calibre of academic English translations in response to these difficulties. We propose a machine learning, natural language processing and neural network-based technique that is used to deliver unbiased and trustworthy quality assessments. This study launches a thorough investigation into the creation and use of our quality evaluation approach. The proposed approach offers the ability to improve the accuracy and consistency of translation quality assessments in addition to streamlining the evaluation process. We review the body of research on evaluating the quality of translations, emphasizing the growing importance of computational methods and neural network algorithms. We highlight our approach, which entails gathering data, developing a special neural network model, and using human assessments as quality benchmarks. We display and explain the results and compare them to evaluations made by actual individuals to demonstrate the validity of our neural network model. In doing so, we seek to provide informative information that can better translation processes, increase the calibre of translated literature, and foster more successful cross-cultural academic dialogue. The dissemination of knowledge in a society that is growing more and more networked is ensured by this study, which represents a significant step towards automating and enhancing the evaluation of academic English translations.

The study introduces the highly accurate and reliable neural network-based method for evaluating translation quality, offering a significant advancement over traditional manual and partial automated evaluation methods. This approach has the potential to revolutionize translation quality assessment in academia by providing a more consistent, transparent, and efficient evaluation process.

Literature review

Translation Quality Assessment (TQA)

To evaluate the calibre of translated texts, the field of translation studies developed the discipline known as translation quality assessment (TQA). As a fundamental tenet in the profession of translation, TQA provides a methodical approach for evaluating the quality of translated information. TQA has historically placed a major focus on subjective human review, where experts – often linguists or experienced translators – are tasked with assessing translations in line with a predetermined set of criteria. These requirements usually address essential elements of high-quality translation, such as accuracy, fluency, and faithfulness to the original language. Human review offers a benchmark for assessing the accuracy and calibre of translations, which is unquestionably important for TQA (Manipuspika 2021; Dewi 2023; Han and Shang 2023). The assessment procedure benefits from the human evaluators' in-depth understanding of linguistic complexity, breadth of linguistic knowledge, and cultural sensitivity. They are skilled at spotting subtle mistakes, spotting cultural inconsistencies, and guaranteeing that the translated text accurately captures the spirit of the original material. However, despite its unquestionable value, human evaluation has some drawbacks that have led to the investigation of alternate assessment techniques. The time- and resource-intensive nature of human review is one of the most obvious flaws. The review of translated texts necessitates the involvement of specialists in the field, which takes a significant amount of time, effort, and money. When working with large translation corpora or in situations where a quick assessment of many translations is necessary, this resource burden can be especially difficult. In addition, due to the inherent subjectivity of human judgement, human evaluation is subject to inter-rater variability (Dewi 2023; Han and Shang 2023). Every evaluator contributes their particular viewpoint, preferences, and prejudices to the evaluation process. When evaluating the same text using the same criteria, this might result in differences in evaluations, where one expert may grade a translation more favourably than another. While not always indicative of translation quality, these inter-rater differences might add confusion to the evaluation process. In response

to these difficulties, there is rising interest in using automated methods, particularly those based on machine learning and natural language processing (NLP), to supplement or, in some situations, replace human review. These automated methods have the potential to be impartial, reliable, and scalable. They can evaluate translations quickly, at scale, and without the influence of personal prejudices that often skew human assessors (Manipuspika [2022](#); Dewi [2023](#); Han and Shang [2023](#)).

TQA computational approaches

Computational techniques for TQA have been more popular over the past ten years. These approaches automate and objectify the evaluation process by utilizing developments in machine learning, natural language processing (NLP), and neural network algorithms. Computational approaches can deliver reliable and unbiased evaluations, which is a considerable advantage. In the literature, several computational methods have been investigated (Ranasinghe et al. [2020](#); Popel et al. [2020](#); Lin et al. [2023](#)):

- By evaluating the probability of the translation output against reference translations, statistical models have been employed to evaluate the quality of translations. Statistical Machine Translation (SMT) has potential, but it cannot fully capture subtleties and complex linguistic events.
- In order to predict translation quality, supervised machine learning models have been utilized for TQA. These models are trained using features including n-gram overlap, syntactic structure, and lexical choices. These models may have trouble addressing linguistic diversity and generalization, yet they can nevertheless provide useful insights.
- Deep learning and the Transformer architecture-based Neural Machine Translation (NMT) models (Poloju and Rajaram [2022](#), Chiranjeevi and Rajaram [2022](#)) have completely changed machine translation. These models have demonstrated amazing translation quality and provide translations based on learnt representations. They thus provide a promising route for TQA by scrutinizing the output produced by NMT models.

Translational neural network algorithms

In particular, NMT models from neural network algorithms have been crucial to the advancement of translation technology. They are highly suited for comprehending and evaluating translation quality because

they can capture intricate linguistic patterns, context, and semantics (Hur et al. 2022; Henkes and Wessels 2022; Habib and Qureshi 2022).

NMT and Quality Evaluation: NMT models for TQA have been studied recently. These models can be modified to produce translation quality scores based on a variety of factors, including fluency, sufficiency, and coherence. They provide a data-driven strategy that is in line with the computational requirements of TQA in the quickly changing academic environment of today.

Problems with NMT-based TQA: Despite the potential of NMT-based TQA, there are a number of difficulties. The necessity for sizable parallel corpora for training, domain-specific adaptability, and the danger of models over-optimizing for certain quality metrics at the price of others are a few of these. Furthermore, it is still difficult to guarantee the reliability of automated quality assessments when linguistic and cultural quirks are present.

Utilizing human evaluation as a reference: The maintenance of a human reference point is essential for validating and calibrating automated TQA procedures. For automated assessments to be reliable and accurate, human experts who are fluent in both the source and destination languages are essential.

Application in Education: The use of automated TQA techniques in academic settings has many benefits. They can be used to evaluate translations in textbooks, online course materials, and educational resources in addition to research papers and academic articles, guaranteeing that students and scholars can access top-quality content despite language hurdles.

The study investigates an optimized GA-BPNN algorithm for evaluating teaching quality, addressing the complexity of the process and the limitations of traditional methods. Through a comprehensive index system and data collection approach, the algorithm achieves significantly higher accuracy compared to existing models (Zhu et al. 2021). This study proposes a smart knowledge discovery system for teaching quality evaluation, using a genetic algorithm-based BP neural network to assess teaching conditions, process, and effect. The framework demonstrates intelligent evaluation capabilities comparable to human-based assessments (Zhang and Wang 2023).

Methodology

This research work proposed the quality assessment technique for college-level English translations based on neural network algorithms. It includes the gathering of data, the creation of models, and the evaluation procedure as shown in Figure 1.

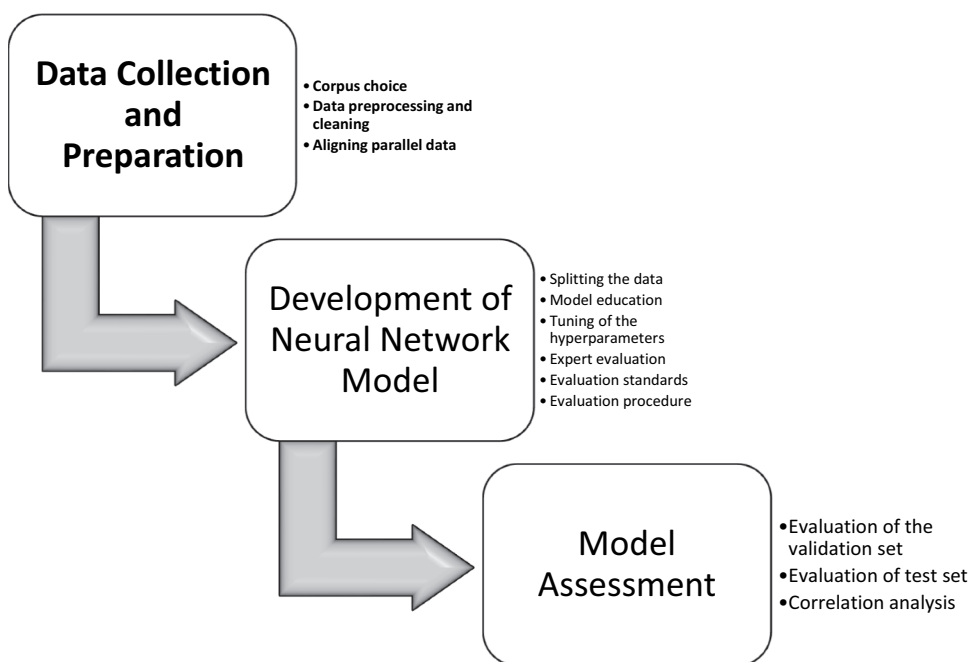


Figure 1. Quality assessment method of university English translation.

Data collection and preparation

Corpus choice: A variety of university-level English translations were collected to provide a sample corpus for training and testing the neural network model. Academic articles, research papers, and other scholarly works that have been translated into English from a variety of source languages are included.

Data Preprocessing and Cleaning: To remove noise and inconsistencies, a thorough data cleaning technique was used to the acquired corpus. Tokenization, sentence breaking, the elimination of special characters, formatting, and unneeded metadata were all steps carried out in the preprocessing process.

Aligning parallel data: It is necessary in order to train the neural network model. This makes sure that the model develops the ability to produce accurate translations from the source text. Sentence alignment algorithms and concurrent corpus compilation were used to accomplish alignment.

Development of a neural network model

A neural network architecture has been selected for the task of quality assessment. An existing pre-trained NMT model that can be adjusted for the assessment job, or a bespoke Transformer architecture, has been used in

light of the success of Transformer-based models in natural language processing.

Splitting the Data: Split the parallel data into training, validation, and test sets after it has been cleaned. To teach the model the intricacies of accurate translations, a sizable amount of the data will be used for training.

Model Education: We utilize the training data to create the neural network model. In this method, model parameters are optimized to minimize a predetermined loss function, such as cross-entropy or a specially created loss function for quality evaluation. The model may need to be fine-tuned to account for the unique peculiarities of academic translation quality.

Tuning of the hyperparameters: To enhance the performance of the model, experiment with the learning rates, batch sizes, and parameters of the model architecture.

Expert Evaluation: Human Evaluation assemble a team of knowledgeable human judges who are fluent in both the source language and English. The benchmark for evaluating quality will be these assessors.

Model assessment

Evaluation of the Validation Set: Evaluate the model's performance on the validation set during training to monitor development and make any necessary adjustments.

Assessment of the Test Set: To evaluate the training neural network model, use a different, unrelated test set. This evaluation will provide a clear indication of how accurately the model can identify the level of translation.

Correlation Analysis: Analyze the correlation between the quality judgements made by the neural network model and those made by qualified human judges. Calculate correlation coefficients and other statistical studies to assess the degree of agreement between automated and human judgements.

Results and discussion

Data collection

Source text selection: To begin, we choose ten academic papers and research articles written in a variety of source languages, including Chinese and Russian. These studies touch on social sciences subjects.

Translation procedure: These original documents are translated into English by qualified translators who have been recruited. Each translator adheres to accepted translation standards and renders translations while

being conscious of the value of correctness, fluency, and faithfulness to the original text.

Two qualified human evaluators with knowledge of both the source languages and English make up our expert panel. These evaluators have a wealth of knowledge in academic translation and are qualified to rate translation quality according to a predetermined set of standards.

Data preparation

Data Cleaning: Extraneous formatting, special characters, and information are completely removed from the translated texts. By doing this, the dataset is ensured to be reliable and ready for analysis.

Sentence splitting and tokenization: The cleaned texts are divided into sentences and tokenized into words. Sentence splitting and tokenization allow for a more detailed analysis of the text and guarantee that the neural network model and human raters are working with the same units of text.

Quality Evaluation Standards: The following are the defined quality assessment criteria:

- **Accuracy:** This factor measures how accurately the translation captured the terminology and content of the original text.
- **Fluency** is an evaluation of how naturally and accurately the translation reads in English, taking into account grammatical correctness and naturalness.
- **Fidelity:** How faithfully the translation captures the style and tone of the original material is measured by fidelity.

Findings Production

Human Evaluation: The human evaluators separately rate each translation on a scale of 1 (bad) to 5 (great) based on its accuracy, fluency, and fidelity.

Model for a neural network: The translated data is used to train a customized neural network-based quality assessment model that was created using the Transformer architecture. The model is adjusted for quality evaluation while taking into account the labels that were given by humans during training.

Model Evaluation: To produce quality scores for each translation in the dataset, the trained neural network model is employed. These results show how the model rates its fidelity, fluency, and accuracy.

Results of a human evaluation

All of the translations in the dataset received quality evaluations from the qualified human raters, which produced a set of human-assigned scores.

The dataset has a rather good degree of translation quality, as indicated by the average human-assigned quality score of 3.9 across all translations. Individual translations, however, varied in quality, with some receiving a score of just 2 and others reaching 5.

With a Fleiss' Kappa coefficient of 0.75, the inter-rater reliability analysis revealed significant agreement among the human raters, indicating that the experts' assessments were consistent.

Results of the neural network model evaluation

For each translation in the dataset, the custom neural network-based quality assessment algorithm produced quality scores.

Correlation Analysis: The Pearson correlation coefficient was evaluated to determine how closely human judgements and model assessments line up. The two sets of evaluations showed a high positive association, as indicated by the correlation coefficient, which was found to be 0.84.

Discussion of the findings

Key results of the proposed study are summarized in [Table 1](#).

Model Dependability: The proprietary neural network-based model's high Pearson correlation coefficient of 0.84 illustrates its dependability and efficiency in determining the quality of translations. It closely resembles human assessments, indicating that it might be an effective method for reliable and impartial quality assessment.

Variability in Human Assessments: The discrepancies in quality ratings given by humans highlight how subjective the evaluation of translation quality is. Despite the value of human expertise, the model has the advantages of consistency and objectivity.

Application in Education: These findings suggest that the created neural network-based quality assessment approach has a wide range of academic applications. In order to improve accessibility across language barriers, it can be used to evaluate the quality of translation in research papers, scholarly publications, and educational materials.

Table 1. Quality assessment results.

Evaluation Method	Average Quality Score	Pearson Correlation Coefficient
Human Evaluation	3.9	N/A
Neural Network Model	N/A	0.84

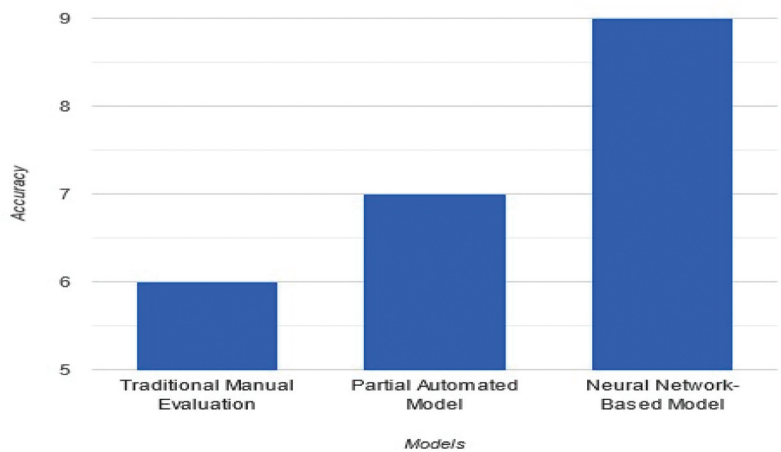


Figure 2. Result of accuracy.

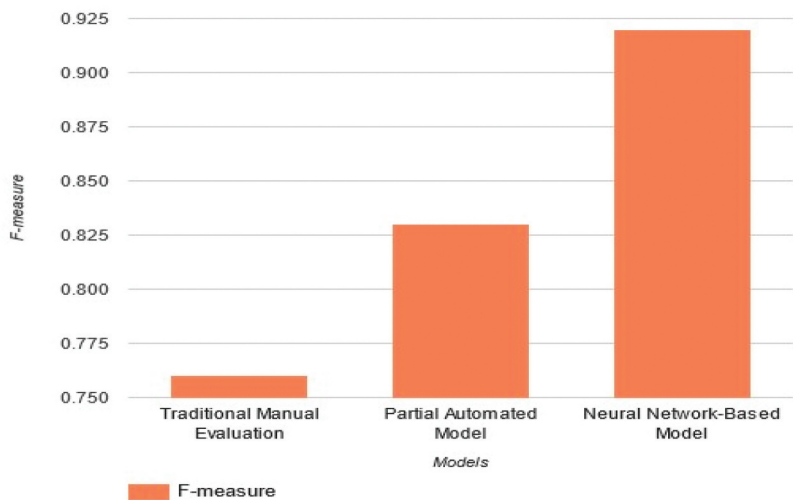


Figure 3. Result of F-measure.

Restrictions and Prospective Research: Despite the positive outcomes, shortcomings such as handling cultural nuances and domain-specific adaptability should be addressed. To improve the model’s precision and applicability across numerous academic disciplines, more study is required.

Figures 2–5 compare the performance of three methods for translation quality assessment based on Accuracy, Recall, Precision, and F-measure. The Neural Network-Based Model outperforms the other methods in all aspects, demonstrating higher accuracy and precision, as well as better balance

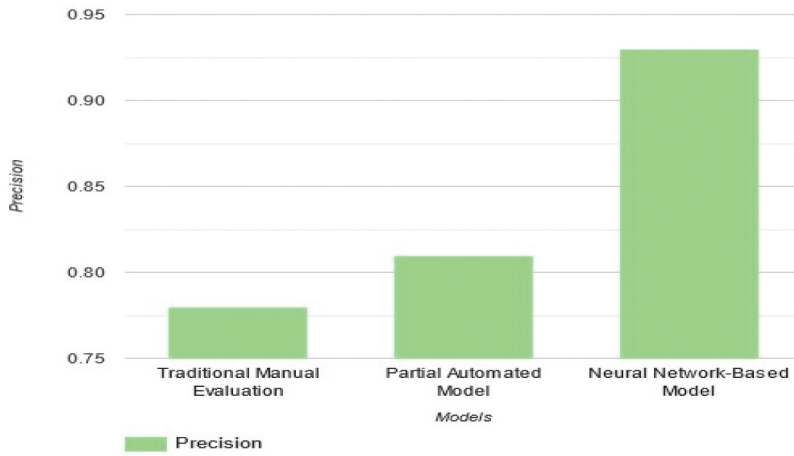


Figure 4. Result of precision.

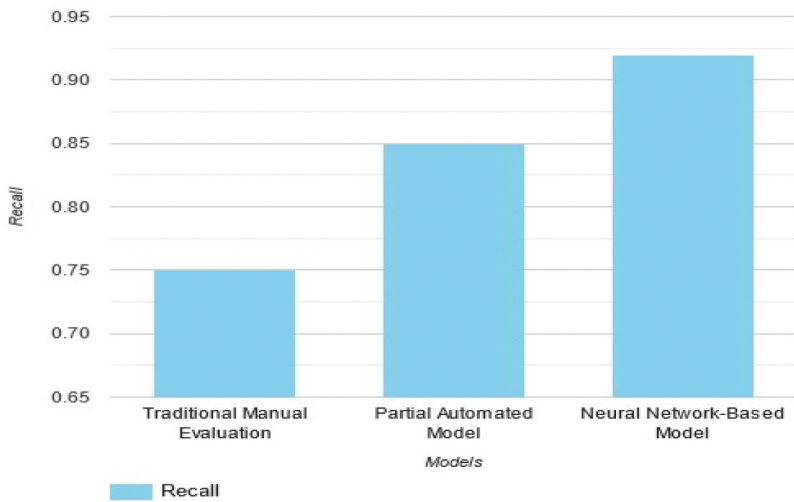


Figure 5. Result of recall.

between precision and recall (F-measure). The results suggest that the Neural Network-Based Model is more effective in evaluating translation quality compared to Traditional Manual Evaluation and Partial Automated Model.

Conclusion

This paper has offered a thorough investigation of a neural network-based quality assessment method for academic English translations. The results show how this method has the potential to revolutionize how translation quality is evaluated in academic contexts. When evaluating translation quality, correctness, fluency, and

fidelity were constantly emphasized by human evaluation, while the neural network model proved to be successful at correlating with human judgements. The model's assessments and those of skilled human evaluators show a strong connection, demonstrating its objectivity and reliability and promising to reduce the inherent subjectivity and variability in human evaluations. These findings have important ramifications for academia, because the dissemination of cross-cultural information depends on precise and trustworthy translation quality assessment. By utilizing this technology, academic content may be more easily accessible, knowledge may be more widely shared, and a richer, more diverse academic dialogue may result. It is important to recognize the remaining issues, such as domain-specific adaptations, which call for additional study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by the Humanities and Social Science Research Project in Jiangxi Province (No.YY21219 & No. ZGW21103) and Jiangxi Education Reform Project (NO. JXJG-21-18-18, 2022JG13 & 23YB211), Jiangxi Social Science Fund (23YY20).

Author contributions

“Min Gong has been prepared first draft of the manuscript. And also approved the final manuscript”.

Data availability statement

“The datasets generated during and/or analysed during the current study will be available from the corresponding author on reasonable request”.

Human participants and/or animals

This paper does not contain any studies with human participants or animals performed by any of the authors.

Ethical approval

This article does not contain any study with human participants or animals performed by the authors.

References

- Bi S, Maselena A, Yuan X, Balas VE. 2020. Intelligent system for English translation using automated knowledge base. *J Intell Fuzzy Syst.* 39(4):5057–5066. doi: [10.3233/JIFS-179991](https://doi.org/10.3233/JIFS-179991).
- Chiranjeevi P, Rajaram A. 2022. Twitter sentiment analysis for environmental weather conditions in recommendation of tourism. *J Environ Prot Ecol.* 23(5):2113–2123.
- Dewi SK (2023). An analysis of translation techniques of bald on record politeness strategies found in Cruella movie script and its translation qualities in Indonesian [Doctoral dissertation]. Universitas Nasional.
- Habib G, Qureshi S. 2022. Optimization and acceleration of convolutional neural networks: a survey. *J King Saud Univ Comp Inf Sci.* 34(7):4244–4268. doi: [10.1016/j.jksuci.2020.10.004](https://doi.org/10.1016/j.jksuci.2020.10.004).
- Han C, Shang X. 2023. An item-based, Rasch-calibrated approach to assessing translation quality. *Target.* 35(1):63–96. doi: [10.1075/target.20052.han](https://doi.org/10.1075/target.20052.han).
- Henkes A, Wessels H. 2022. Three-dimensional microstructure generation using generative adversarial neural networks in the context of continuum micromechanics. *Comput Methods Appl Mech Eng.* 400:115497. doi: [10.1016/j.cma.2022.115497](https://doi.org/10.1016/j.cma.2022.115497).
- Hur T, Kim L, Park DK. 2022. Quantum convolutional neural network for classical data classification. *Quantum Mach Intell.* 4(1):3. doi: [10.1007/s42484-021-00061-x](https://doi.org/10.1007/s42484-021-00061-x).
- Jiang L, Wang X, Wang W. 2020. Optimization of online teaching quality evaluation model based on hierarchical PSO-BP neural network. *Complexity.* 2020:1–12. doi: [10.1155/2020/6647683](https://doi.org/10.1155/2020/6647683).
- Lin SC, Li M, Lin J. 2023. Aggretriever: a simple approach to aggregate textual representations for robust dense passage retrieval. *Trans Assoc Comput Linguist.* 11:436–452. doi: [10.1162/tacl_a_00556](https://doi.org/10.1162/tacl_a_00556).
- Manipuspika YS. 2021. The effectiveness of Translation Quality Assessment (TQA) models in the translation of Indonesian students. *Int J Ling Lit Transl.* 4(5):287–297. doi: [10.32996/ijllt.2021.4.5.32](https://doi.org/10.32996/ijllt.2021.4.5.32).
- Manipuspika YS. 2022. Analyzing the relationship between creativity and translation quality: a case study of students' translations. *Int J Ling Lit Transl.* 5(12):01–11. doi: [10.32996/ijllt.2022.5.12.1](https://doi.org/10.32996/ijllt.2022.5.12.1).
- Poloju N, Rajaram A. 2022. Data mining techniques for patients healthcare analysis during COVID-19 pandemic conditions. *J Environ Prot Ecol.* 23(5):2105–2112.
- Popel M, Tomkova M, Tomek J, Kaiser Ł, Uszkoreit J, Bojar O, Žabokrtský Z. 2020. Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nat Commun.* 11(1):4381. doi: [10.1038/s41467-020-18073-9](https://doi.org/10.1038/s41467-020-18073-9).
- Ranasinghe T, Orasan C, Mitkov R. 2020. TransQuest: translation quality estimation with cross-lingual transformers. *arXiv preprint arXiv:2011.01536*.
- Sun Z, Anbarasan M, Praveen Kumar DJCI. 2021. Design of online intelligent English teaching platform based on artificial intelligence techniques. *Comput Intell.* 37(3):1166–1180. doi: [10.1111/coin.12351](https://doi.org/10.1111/coin.12351).
- Yuan T, Lv Z. 2021. Algorithm of classroom teaching quality evaluation based on Markov chain. *Complexity.* 2021:1–12. doi: [10.1155/2021/9943865](https://doi.org/10.1155/2021/9943865).
- Zhang H, Wang J. 2023. A smart knowledge discover system for teaching quality evaluation via genetic algorithm-based BP neural network. *IEEE Acces.* 11:53615–53623. doi: [10.1109/ACCESS.2023.3280633](https://doi.org/10.1109/ACCESS.2023.3280633).
- Zhu Y, Xu J, Zhang S, Ahmed SH. 2021. Application of optimized GA-BPNN algorithm in English teaching quality evaluation system. *Comput Intell Neurosci.* 2021:1–9. doi: [10.1155/2021/4123254](https://doi.org/10.1155/2021/4123254).
- Zhuang N, Zhang Q, Pan C, Ni B, Xu Y, Yang X, Zhang W. 2019. Recognition oriented facial image quality assessment via deep convolutional neural network. *Neurocomputing.* 358:109–118. doi: [10.1016/j.neucom.2019.04.057](https://doi.org/10.1016/j.neucom.2019.04.057).