Hindawi
Journal of Sensors
Volume 2021, Article ID 7062511, 7 pages
https://doi.org/10.1155/2021/7062511

Research Article

Model and Verification of Medical English Machine Translation Based on Optimized Generalized Likelihood Ratio Algorithm

Peng Yu[1] **and Youyu Zhu**[2]

[1]*School of Foreign Languages, Chaohu University, Chaohu, Anhui 238024, China*
[2]*School of Basic Medical Sciences, Anhui Medical University, Hefei 230032, China*

Correspondence should be addressed to Youyu Zhu; zhuyy0103@126.com

Received 24 July 2021; Revised 26 October 2021; Accepted 6 December 2021; Published 28 December 2021

Academic Editor: Kelvin Wong

Phrase identification plays an important role in medical English machine translation. However, the phrases in medical English are complicated in internal structure and semantic relationship, which hinders the identification of machine translation and thus affects the accuracy of translation results. With the aim of breaking through the bottleneck of machine translation in medical field, this paper designed a machine translation model based on the optimized generalized likelihood ratio (GLR) algorithm. Specifically, the model in question established a medical phrase corpus of 250,000 English and 280,000 Chinese words, applied the symbol mapping function to the identification of the phrase's part of speech, and employed the syntactic function of the multioutput analysis table structure to correct the structural ambiguity in the identification of the part of speech, eventually obtaining the final identification result. According to the comprehensive verification, the translation model employing the optimized GLR algorithm was seen to improve the speed, accuracy, and update performance of machine translation and was seen to be more suitable for machine translation in medical field, therefore providing a new perspective for the employment of medical machine translation.

1. Introduction

Affected by the raging novel coronavirus the world over, medical English translation has become an active and important communication medium in the fight against the epidemic among the countries. In recent years, the number of machine translation applications has witnessed a boom due to the fact that education and technology develop at the fastest speed that we have ever seen [1]. These applications, however, are mainly concentrated in the translation of such fields as cultural exchanges, economy, politics, and academic literature. Less attention was seen to be paid to the special field, for example, medical English translation in question. Moreover, existing machine translation technology did exhibit some drawbacks when applied to the field of medical translation. For instance, there are terms representing categories and concepts in medicine, whose semantic relationship is complicated within these phrases, posing problems for the phrase identification in current machine translation. Accordingly, the accuracy of the identification in machine translation is seen to fail to meet the standards of medical translation. As we all know, phrase identification plays a crucial role in machine translation. Notably, one of the difficulties of the current English-Chinese machine translation is the resolution of phrase ambiguity [2]. Furthermore, medical terms usually display a high degree of ambiguity in both English and Chinese languages, which makes the syntactic analysis in machine translation extremely complicated. But this ambiguity, to a large extent, can only to be solved by phrase identification, and thus, machine translation is inseparable from phrase identification. Along this line of consideration, the core issue affecting the quality of machine translation is the machine's performance to deal with ambiguities based on employing appropriate phrase identification. To be specific, structural ambiguity is one of the most complex ones, and previous researchers studied this phenomenon from multiple angles and proposed a variety of methods for phrase identification and disambiguation [3].

Joty et al. [4] applied a rule-based algorithm to phrase identification via calling the rules in question to obtain correct labeling, trying to establish a complete and accurate set of labeling rules. This algorithm could accurately describe the certain phenomenon between part of speech collocations. However, it was not seen to be a satisfying solution to structural ambiguity of the phrases since the language coverage of the rules was limited, the compilation and the maintenance of the huge rule base were overwhelming, and the priority and the conflict between the rules were not easy to be settled.

Banik et al. [5] used statistical algorithm for phrase identification. The algorithm was described to collect the language information in the training corpus via statistical methods. To be specific, the information of the language in statistical algorithm was used as an automatically "summed up" language phenomenon and was applied to the test corpus to obtain the correct part-of-speech tagging [6–9]. Evidently, this algorithm considered the dependence among parts of speech from a macro perspective, which was seen to cover most language phenomenon, thus possessing an overall higher accuracy and stability. Comparatively, the accuracy of statistical algorithm in describing the phenomenon of determining part of speech collocation was not as good as that of the rule-based algorithm.

Hybrid algorithm was also employed by the researchers for phrase identification. The hybrid method, as the name implied, referred to the combination of rule-based algorithm and statistical algorithm. Namely, its part-of-speech tagging model combined those of the two algorithms, which was regarded as the most effective tagging method based on the statistical algorithm tagging mode via the rule-based algorithm. Nevertheless, hybrid algorithms still could not resolve structural ambiguities to a large extent. In summary, in these automatic phrase identification algorithms, some structures that were extremely simple in artificial translation could not yet be accurately identified.

After reviewing the above literature, it is not difficult to find that the intelligent identification of phrases is to recognize and summarize the phrases in the sentence, to mark their part of speech and syntax, and automatically to combine and translate them against the corpus, eventually obtaining the corresponding translation results. Evidently, nowadays, there are quite many English translation model designs, most of which were designed based on word-sense disambiguation and semantic role labeling. To a certain extent, they could partly meet the needs of users. However, medical translation is different from other general translation activities. It has higher requirements for accuracy and professionalism. Therefore, common part-of-speech identification technologies could hardly meet the requirements of medical English translation.

As mentioned in the foregoing sections, intelligent phrase identification was regarded as the core of medical English translation since it could facilitate the selection of translation samples and the precise alignment of parallel corpus. Furthermore, the use of phrase intelligent identification technology could effectively remove structural ambiguity. Along this line of consideration, the present paper used a machine translation model on the basis of an optimized likelihood ratio (GLR) algorithm [10]. To be specific, the algorithm in question constructed a medical phrase corpus of approximately 250,000 English and 280,000 Chinese words labeled, making the phrases searchable. These phrases, like those of vocabulary, were made to possess such features as subcategories, morphology, semantics, and other characteristics. In effect, these features were mainly reflected by the central word of a phrase. Accordingly, the part-of-speech identification result was obtained while recognizing the short syntactic structure, and the ambiguity of the English-Chinese structure in the part-of-speech identification was corrected in accordance with the syntactic function of the parsing linear table. Finally, the recognized content was obtained, and the actual range of the position of the phrase in translation was therefore determined. Therefore, the model based an optimized GLR was assumed to alleviate the structural ambiguity in the current medical translation to a certain extent and to improve the accuracy of phrase identification.

2. Intelligent Modes Based on Optimized GLR

2.1. Construction of Medical English Intelligent Translation Model. The machine translation model based on the bilingual corpus is seen to make its translation more accurate via the identification of phrases, thus contributing more help to translators. Therefore, corpus, especially bilingual corpus, is increasingly gaining attention and application in current intelligent translation models. To be specific, accurately labeling the English-Chinese bilingual phrase corpus and storing it in the corpus would, to a large extent, improve the accuracy and efficiency of the phrase identification algorithm in the machine translation process, which would serve as an effective auxiliary tool for translators to improve translation quality and efficiency [11–13]. Corpus, however, is a multiangle, multilevel, and multidomain research tool, whose classification is intricate and still seems to be an open question. In spite of that, the English-Chinese bilingual medical phrase corpus is homogenous, that is, it only collected the same type of content. Accordingly, this type of corpus would be more accurate and professional when applied to machine translation in specific fields, and meanwhile, the probability of ambiguity in semantic identification would also be reduced.

Notably, the following three aspects were considered in the construction of the English-Chinese bilingual medical corpus in question. First is the field of the corpus. Medical field is regarded as an important one of machine translation applications. Communication in medicine, as we know, is often carried out among hospitals, firms, and individuals using different languages, especially English and other languages. Therefore, a certain demand for machine translation cannot be avoided in such an information-explosion era. From a linguistic point of view, medical English is unique in stylistic features, i.e., obvious syntactic and morphological features, such as rich terminology, rigorous long sentence structure, and standardized wording. Moreover, its written medical tests are stylized. Thus, these features make it more suitable for the research and application of machine translation. Second are the size, the genre, and the style of the corpus. Due to the limited time and manpower, the scale of the English-Chinese bilingual medical phrase corpus in this paper was positioned at 15,000 sentence pairs, with the genre of the corpus being medical language and with the style being written and spoken language. Third is the collection and the sorting of corpus. The collection and arrangement of corpus was composed of five processes: corpus collection, clauses, English-Chinese alignment, deduplication, and proofreading, separately. To be specific, the source of the corpus was from publicly issued books and electronic journals, and the corpus itself was in terms of sentence-level parallel. Moreover, the original corpus collected initially was paragraphs, and then, the phrases of which were divided into sentences. The division of sentences, however, was mainly in terms of English ones. Furthermore, in the English-Chinese alignment stage, Chinese sentences were matched to their English counterparts, and after the alignment of English and Chinese, the duplicates were removed. Therefore, there were no repeated English sentences in the corpus. Furthermore, the final process was proofreading, while other aspects remained the original appearance of the corpus. Thus, the authenticity of the corpus was assured.

Accordingly, the phrase corpus of medical translation model constructed in this paper contained 250,000 English words and 280,000 Chinese counterparts, which could meet the needs of constructing 10,000 sentences and 5000 phrases. As is shown in Table 1, the medical phrase corpus was homogeneous, mainly focusing on medical-related professional terms, and could be translated between English and Chinese in various medical fields such as clinical, pharmacy, and imaging. To be specific, the English phrase corpus and that of Chinese were marked separately, meanwhile distinguishing the tenses of different phrase corpus. Evidently, the corpus processing method was composed of three parts: data, level, and processing mode, separately. Specifically, the type of data was text format, and the level of part of speech and alignment were selected. Additionally, the processing method adopted direct interaction between human and machine, carrying out a series of operations of translation and promoting the authenticity and accuracy of phrase corpus translation. The specific corpus information is shown in Table 1.

2.2. The Optimized Algorithm Employed in the Model. As mentioned in the foregoing sections, phrase-level syntax analysis was the core of the intelligent identification algorithm of machine translation, while the GLR algorithm was a commonly used algorithm in part-of-speech identification [14]. To be specific, this algorithm was the one that identified context-independent languages via the analysis tables of "action" and "goto." Furthermore, each table entry contained multiple shifts or reduction actions in which each entry and each exit of the stack existed in terms of the state symbol pair. However, when there was ambiguity between advancement and statute, the GLR algorithm would apply the graph structure stack technology to copying the analysis stack, allowing each analysis stack to complete an action in the analysis table, while retaining multiple possibilities to generate multiple identification results. Then, an independent analysis would be carried out on these identification results. Particularly, when an error occurred in one of the analysis stacks, this analysis stack was discarded and other analysis results were output [15, 16].

Therefore, when a machine translation model using the GLR algorithm was applied to the translation in medicine, the following problems would arise. First, the number of identification results given by the GLR algorithm was uncertain, and there would be overlapping data in the identification results, which affected the accuracy of the identification results and thus hindered the quality of translation. Second, in the results of the GLR algorithm, each chunk was not seen to be compatible with one another, that is to say, phrases, unlike those of vocabulary, did not have semantic, morphological, and subcategory characteristics. Finally, the central word of the syntax structure was not specified in the results of the GLR algorithm.

To avoid the problems mentioned above, this paper, however, used a GLR algorithm that had been expanded and optimized. Specifically, this algorithm in question employed a context-independent grammatical form in the system and expanded its start symbol S and production formula P. Moreover, it analyzed the structure of the phrase via phrase, which effectively reduced the probability of overlapping data points. Its algorithm form was a quaternion, as shown in

$$G=(V_{\mathrm{N}}, V_{\mathrm{T}}, S, \alpha). \tag{1}$$

In Equation (1), V_{N} represented a nonterminal symbol set, which was a nonempty finite set; V_{T} represented a terminal symbol set, which was likewise a nonempty finite one, and the elements in V_{T} and V_{N} did not overlap. S stood for the start symbol set, an element in V_{N}, and a syntactically recognizable phrase symbol set. α represented the set of productions. Assuming that P was any action in α and $P \in V_{\mathrm{N}}$, the production (2) could be obtained:

$$P \longrightarrow \{\theta, C, \beta, \gamma\}. \tag{2}$$

In Equation (2), θ, C, β, and γ represented the right symbol string, center symbol, restriction condition, and target conversion mode of the action, respectively. Among them, θ and C belonged to both V_{T} and V_{N}, and γ could belong to both V_{T} and V_{N}. The improved GLR algorithm stipulated that the top symbol of the linear table of the identification result was consistent with θ, the restriction condition β should be true, and the center symbol C should be a numeric value, not a null value. Only the identification result that met the above three criteria was the result of phrase part-of-speech identification.

2.3. The Process of Algorithm Designed in the Model. In current English-Chinese machine translation algorithms, the part-of-speech identification result of the phrase corpus was usually output as the final result of translation, which

Table 1: Corpus information of English-Chinese bilingual medical phrases.

Element	Nature	Content
Corpus composition	Scale	250,000 English words 280,000 Chinese words,
	Scale of use	Clinical, pharmacy, imaging, inspection, etc.
	Style	Spoken and written
	Tense	Past, present, future
Corpus processing	Data	Text
	Level	Part of speech, alignment
	Processing	Man-machine communication
Corpus application	Scale	Medical English translation

mainly relied on the part-of-speech analysis of the corpus. However, the identification in question did not improve the structural ambiguity between English and Chinese languages, and thus, it hindered the accuracy of the translation results. Therefore, it was difficult to meet the high-accuracy and high-precision requirements of medical English translation.

Along this line of consideration, it was essential to correct the results of identification in the process of machine translation [17–19]. Therefore, this paper further considered the correction of the results of the identification and identified phrase actions via the analytic linear table in the process of performing part-of-speech analysis against the optimized GLR algorithm. In addition, errors of the identification would be analyzed via such pointers as advancement, specification, acceptance, termination, error, and correction due to the fact that the analytic linear table also owned the function of syntactic identification. These errors were finally to be corrected by searching the marked content in the phrase corpus. The detailed phrase correction algorithm flow is shown in Figure 1.

In Figure 1, there were 6 actions involved in the entire algorithm, namely, advancement, statute, acceptance, terminator, error, and correction. Moreover, the relationship between advancement and statute could be evidently observed in which lay the similarities and essential differences. On the one hand, the similarity was that the two functions were similar both of which were to replace the position of the terminator in the analytical linear table. On the other hand, the difference between them was that the advancement referred to putting the current state and symbols on the stack and moving down the analysis pointer. However, statute referred to reinvoking the constraint condition function to check the rule condition. If the conditions were met, each subnode would be popped from the symbol stack to form a nonterminal syntactic structure tree. At the same time, the identification pointer of the central word was pointed to the corresponding central one, eventually generating the translation of the current nonterminal character in accordance with the mode of translation. Conversely, if the conditions were not met, the terminator pointer was directly placed. Specifically, terminator replacement meant that if the terminator pointer was not placed, the current system terminator was mapped to the analysis table terminator via the symbol mapping function; if the termination pointer was entered, the current system terminator was directly mapped to the analysis table terminator.

It should be pointed out that before the terminator replacement the type of pointer was required to be identified in the optimized and expanded GLR algorithm. To be specific, suppose that it was a statute pointer and then whether the constraint function of the pointer belonged to the phrase corpus should be checked; if not, the termination pointer was directly placed. The terminator generally appeared at the backup point with structural ambiguity. Therefore, when it was queried, a phrase structure tree would be formed, and the central symbol would be check whether it was placed on the correct sentence structure. If it was not correct, then the algorithm in question would call up error pointer to correct the identification result of the part of speech. As shown in the figure, there were multiple phrase identification outputs in the entire correction process of the algorithm, and one acceptance pointer only output one identification result. However, when multiple identification results appeared at the same time, the correction process would write them into the same node of the phrase structure tree, and the receiving pointer then would automatically treat it as one identification result.

3. Model Design Verification

In order to detect the actual effect of the medical English-Chinese translation model on the basis of the optimized GLR algorithm, relevant evaluations were carried out in the research. Furthermore, the main performance indicators of the evaluation included the accuracy of the translation results, the translation speed, and the update ability, separately. Specifically, the evaluation team of the experiment was composed of 5 English-Chinese translation machines, 5 professional medical translators, and professional scorers. Among them, 5 English-Chinese translation machines chose rule-based algorithm, statistical algorithm, hybrid algorithm, GLR algorithm, and optimized GLR algorithm, respectively. Moreover, 5 professional medical translators all owned more than 10 years' experience in medical translation and worked together as a team, negotiating, and forming the only version of the tested material.

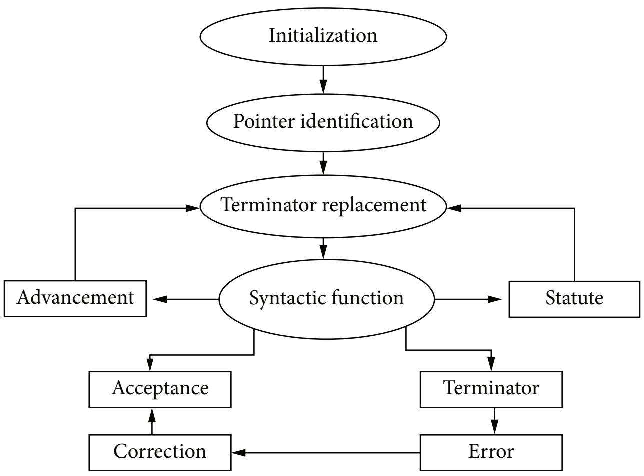

Figure 1: Intelligent identification algorithm correction flow.

In our paper, 5 English-Chinese machine translators translated the designated 70 medical terms and 70 randomly chosen medical English sentences in the evaluation process. Likewise, professional medical translators translate the same 70 phrases and 70 randomly chosen sentences. Then, the scorers would score the results of the machine translators, respectively, in accordance with certain rules. Specifically, the score would be given according to such rules as translation accuracy, translation speed, and update performance. To put it concretely, the translation accuracy was scored based on the clarity and accuracy of the translation, and the total score was 100 points. Furthermore, the translation speed was based on the total identification time multiplied by the weight, and then, the sum was divided by the number of phrase identification. The update capability, however, depended on the total update time multiplied by the weight, and then, the sum was divided by the number of phrase identification. Additionally, the weight of each score was the translation accuracy of 0.6, the translation speed of 0.2, and the update performance of 0.2.

4. Results and Discussion

The detailed experimental results are shown in Figures 2 and 3.

From the results in Figure 2, the machine translation based on the optimized GLR algorithm was seen to be the best of its kind in terms of translation accuracy, translation speed, and update performance. Furthermore, as the comprehensive evaluation results showed in Figure 3, the optimized GLR algorithm ranked the highest with a score of 94.4, while the statistical algorithm ranked the lowest with a score of 79.4. However, the hybrid algorithm was not much different from the optimized GLR algorithm in the final test score. The main gap between the two was centered on the score in update performance. Combined with Figures 2 and 3, it is obvious that the optimized GLR algorithm had obvious performance advantages over other algorithms, which was seen to be more suitable for medical translation.

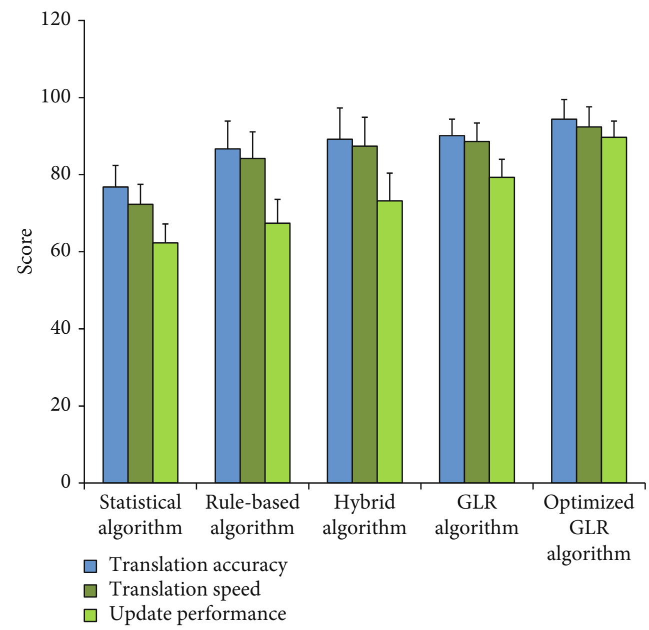

Figure 2: Results of different English-Chinese translation algorithms.

In order to test the performance of removing the structural ambiguity among different algorithms in real cases, this paper also employed a Chinese sentence that is related in medicine. "Tóutòng zhīqián de zhèngzhuàng y□u kěnéng shì yóu dànǎo bùfèn qūyù gōngxiě shùnjiān jiǎnshǎo su□ dǎozhì de" was selected for translation, and the results were compared among the translation model based on rule-based algorithm, statistical algorithm, hybrid algorithm, GLR algorithm, optimized GLR algorithm, and artificial translation. The results are shown in Table 2.

It can be found from Table 2 that translations based on statistical algorithms, rule-based algorithms, hybrid algorithms, and GLR algorithms were basically correct from a grammatical perspective, but from a semantic point of view, they were not very complete. In particular, the translation results of statistical algorithms were ambiguous, and the translation results were not very accurate. In addition, of the 5 algorithms, four did not translate the Chinese word "zheng zhuang," symptom, into English. Evidently, only the optimized GLR algorithm translated it into English. From the comparison of the semantics of the translation results, only the machine translation based on the optimized GLR algorithm was the closest to the artificial translation. Therefore, compared with the machine translation of other algorithms, it can be clearly seen that the machine translation result of the optimized GLR algorithm was more accurate in part-of-speech identification, the translation result was the closest to the artificial result, and the identification accuracy reached more than 96. This showed that the optimized GLR algorithm was more suitable for machine translation.

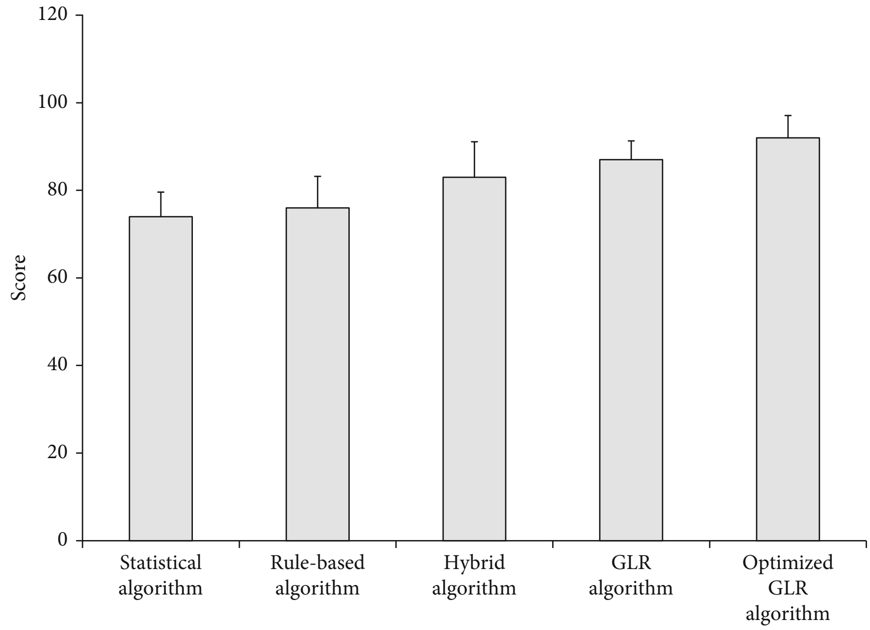

Figure 3: Comparison of comprehensive test scores of different English-Chinese translation algorithms.

Table 2: Results of translation examples.

Method	Content
Statistical algorithm	The headache aforementioned has possibility that it is caused by the brain's parts of blood supply reducing suddenly.
Rule-based algorithm	The headache before may be caused by an instantaneous reduction in blood supply to parts of the brain.
Hybrid algorithm	It is likely that the headache may be caused by a sudden reduction in blood supply to parts of the brain.
GLR algorithm	The headache before is probably caused by an instantaneous reduction in blood supply to parts of the brain.
Optimized GLR algorithm	It is possible that the symptoms before the headache may be caused by an instantaneous reduction in blood supply to parts of the brain.
Artificial translation	It is possible that the symptoms preceding the headache result from a transient decrease in blood supply to areas of the brain.

5. Conclusion

In order to improve the performance of machine translation in medicine field, this article designed an intelligent medical English translation model via expanding and optimizing the traditional GLR algorithm, which was seen to be capable of removing the structural ambiguity of English and Chinese medicine terms. The algorithm in question constructed the phrase structure through the phrase center point and endowed a phrase with such characteristics of a word as semantics, morphology, and subcategory, thus improving the accuracy of the phrase identification. Particularly, when this algorithm was applied to machine translation in medicine, correction pointer was added in the identification process. Therefore, when structural ambiguities were encountered, the syntactic function of parsing linear tables was to be used to correct the English and Chinese structural ambiguities in the results of part-of-speech identification. Notably, this algorithm largely changed the low accuracy of phrase part-of-speech identification among traditional algorithms and improved the accuracy of machine translation's performance in medicine. The results of the evaluation showed that, compared with other algorithms, the translation model on the basis of the optimized GLR algorithm was more accurate in identification, faster in translation speed, and stronger in update performance. Accordingly, it was seen to be more suitable for medical English machine translation. An intelligent medical English translation model based on deep learning algorithms [20–22] may be developed in future implementations.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] J. Downie and A. Dickson, “Unsound evaluations of medical machine translation risk patient health and confidentiality,” *JAMA Internal Medicine*, vol. 179, no. 7, pp. 1001-1002, 2019.

[2] A. Panayiotou, A. Gardner, S. Williams et al., “Language translation apps in health care settings: expert opinion,” *JMIR mHealth and uHealth*, vol. 7, no. 4, article e11316, 2019.

[3] S. van de Velde, L. Macken, K. Vanneste et al., “Technology for large-scale translation of clinical practice guidelines: a pilot study of the performance of a hybrid human and computer-assisted approach,” *JMIR Medical Informatics*, vol. 3, no. 4, article e33, 2015.

[4] S. Joty, F. Guzmán, L. Màrquez, and P. Nakov, “Discourse structure in machine translation evaluation,” *Computational Linguistics*, vol. 43, no. 4, pp. 683–722, 2017.

[5] D. Banik, A. Ekbal, and P. Bhattacharyya, “Statistical machine translation based on weighted syntax–semantics,” *Sadhana*, vol. 45, no. 1, p. 191, 2020.

[6] C. C. Chua, T. Y. Lim, L. K. Soon, E. K. Tang, and B. Ranaivo-Malançon, “Analogical-based translation hypothesis derivation with structural semantics for english to malay example-based machine translation,” *Advanced Science Letters*, vol. 24, no. 2, pp. 1263–1267, 2018.

[7] C. C. Chua, T. Y. Lim, L.-K. Soon, E. K. Tang, and B. Ranaivo-Malançon, “Meaning preservation in example-based machine translation with structural semantics,” *Expert Systems with Applications*, vol. 78, pp. 242–258, 2017.

[8] K. Welnitzová and D. Munková, “Sentence-structure errors of machine translation into Slovak,” *Topics in Linguistics*, vol. 22, no. 1, pp. 78–92, 2021.

[9] K. Rajan, A. Zielesny, and C. Steinbeck, “Stout: smiles to iupac names using neural machine translation,” *Journal of Cheminformatics*, vol. 13, no. 1, p. 34, 2021.

[10] M. A. Stoto, R. D. Fricker, A. Jain et al., “Evaluating statistical methods for syndromic surveillance,” in *Statistical Methods in Counterterrorism: Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication*, pp. 141–172, Springer, New York, NY, 2006.

[11] N. Önder, “Compiling a written corpus of english as an academic lingua franca: medical research articles of turkish academics,” *Boğaziçi Üniversitesi Eğitim Dergisi*, vol. 29, 2012.

[12] I. Taavitsainen, T. Hiltunen, A. Lehto et al., “Late modern English medical texts 1700–1800&58; a corpus for analysing eighteenth-century medical English,” *ICAME Journal*, vol. 38, no. 1, pp. 137–153, 2014.

[13] E. E. Sweany, “Unsettling comparisons,” *English Language Notes*, vol. 58, no. 2, pp. 83–100, 2020.

[14] D. Vivek and P. Balasubramanie, “An expressive phrases identification supported with feature prediction consuming unstructured data collection,” *Multimedia Tools and Applications*, vol. 79, pp. 1–16, 2020.

[15] M. Barash and A. Okhotin, “Generalized lr parsing algorithm for grammars with one-sided contexts,” *Theory of Computing Systems*, vol. 61, no. 2, pp. 581–605, 2017.

[16] A. Okhotin, “Lr parsing for conjunctive grammars,” *Grammars*, vol. 5, no. 2, pp. 81–124, 2002.

[17] M. Murata, M. Utiyama, K. Uchimoto, H. Isahara, and Q. Ma, “Correction of errors in a verb modality corpus for machine translation with a machine-learning method,” *Acm Transactions on Asian Language Information Processing*, vol. 4, no. 1, pp. 18–37, 2005.

[18] N. Ehsan and H. Faili, “Grammatical and context-sensitive error correction using a statistical machine translation framework,” *Software: Practice and Experience*, vol. 43, no. 2, pp. 187–206, 2013.

[19] J. E. Denny and B. A. Malloy, “The IELR(1) algorithm for generating minimal LR(1) parser tables for non-LR(1) grammars with conflict resolution,” *Science of Computer Programming*, vol. 75, no. 11, pp. 943–979, 2010.

[20] Z. Tang, G. Zhao, and T. Ouyang, “Two-phase deep learning model for short-term wind direction forecasting,” *Renewable Energy*, vol. 173, no. 72, pp. 1005–1016, 2021.

[21] K. K. L. Wong, “Bridging game theory and the knapsack problem: a theoretical formulation,” *Journal of Engineering Mathematics*, vol. 91, no. 1, pp. 177–192, 2015.

[22] K. K. Wong, “A geometrical perspective for the bargaining problem,” *PLoS One*, vol. 5, no. 4, article e10331, 2010.