2022 5th Asia Conference on Machine Learning and Computing (ACMLC) | 979-8-3503-3392-3/22/$31.00 ©2022 IEEE | DOI: 10.1109/ACMLC58173.2022.00014

The Design of English Translation Software Based on Machine Learning Technology

Xiaoshan Zeng
School of Foreign Languages
Shaanxi University of Technology
Hanzhong, Shaanxi, China
xiaoshanzeng@126.com

***Abstract*—With the increasing frequency of our country's international communication and the popularization and penetration of mobile Internet into people's work and life styles in modern society, the level of social informatization has also increased. In order not to be eliminated by this era, people must follow the pace of development of this era, always keep an eye on and receive the latest information from all over the world. Most of these materials are published on the Internet in foreign languages. Therefore, language has become the biggest obstacle for people to obtain information. As machine translation technology is restricted in terms of convenience and cost control, people's need for machine translation or machine translation technology has become more and more urgent. This paper studies the design of English translation software (ETS) based on machine learning technology (MLT). By introducing MLT into ETS, a new neural machine translation method is proposed, and related experiments are used to test the effectiveness of the translation method. The designed translation software has been evaluated for translation quality. The experimental results show that as the arc length distribution increases, the translation quality (TTQ) decreases. The English translation software designed in this paper is of great importance to the research and development of machine translation.**

Keywords—machine learning, English translation, software design, machine translation

I. INTRODUCTIONS

As an important application field of artificial intelligence technology [1-2], has begun to attract the attention of academia and the software industry. On the one hand, machine translation, as an application field in the computer field, is a level of artificial intelligence technology has great scientific research significance [3-4]. On the other hand, because there are still many insurmountable cultural differences between different countries and regions in our real life [5-6], such as language barriers, in the new state of social and cultural diversification, automatic translation software becomes more and more, and it becomes more and more important [7-8].

For English translation research, some researchers have used MLT and embedded technology to investigate the language characteristics of different translations, and found that there are few comparative studies for different language translation types, and there are few related corpus studies [9]. Some researchers have also designed an English handheld translation device that can carry out oral communication, using n-gram large vocabulary for speech recognition, but the system still has some problems, because the irregular English accent will cause English translation errors [10]. Some researchers have also proposed a method of neural machine translation based on weight-sharing resources on how to use the language of parallel corpus resources to improve the performance of ETS based on MLT, and use the method of weight sharing to improve English neural machine translation. The performance of this method is verified through experiments [11]. Researchers have also conducted research on the application of MLT, and proposed a method of applying less supervision to improve the blindness, inefficiency, redundancy and superficiality in the training of machine translation models. Some researchers have also constructed a deep learning network neural translation model, and analyzed the practicality and feasibility of the translation model. The results show that the translation ability of the neural network translation model is nearly one higher than that of the traditional N-tuple translation model. The accuracy rate based on the neural network translation model reached 95.21%, and the adaptability was 0.4 times that of the traditional N-tuple model [12].

This paper first, it explains the problems of ETS based on the literature, and then analyzes the application of MLT in ETS, and then analyzes the problems according to the problems ETS based on MLT is designed, and a new neural machine translation method is proposed. Then, experiments are designed to verify the proposed translation method, and the translation software designed in this paper is tested, and draw relevant experimental conclusions.

II. RESEARCH ON ETS

A. Problems with ETS

(1) Lack of proper semantic representation: What automatic translation actually performs is the process of regenerating understanding at the semantic level. The statistical method of automatic translation is mainly to convert two language texts at the word, phrase, and syntax level. The disadvantage of automatic translation is poor expressive ability and lack of appropriate expression to support equivalent conversion to a meaningful level [13].

(2) It is difficult to make full use of different localities and contexts: the traditional methods mainly use grammars that are not related to the environment to design and search algorithms

979-8-3503-3392-3/22/$31.00 ©2022 IEEE
DOI 10.1109/ACMLC58173.2022.00014

to find the optimal path from a high-resolution space. For example, the n-gram method generally only takes the first 3-5 fields of the current word as a historical data. Therefore, it is difficult for machine translation to effectively collect accurate data of the proposal in the modeling process.

(3) Data isolation: A statistical method based on machine translation usually regards a word (statistical machine translation based on a word) and a phrase (statistical machine translation based on a phrase) as an independent standard. Even if it is a model that uses large-scale data training, it is difficult to effectively alleviate some sparse data problems.

B. Application of MLT in ETS

In the processing of natural language, syntactic analysis mainly includes two manifestations: simple phrase structure analysis and dependency analysis. No matter what kind of analysis method is adopted, the proposition is considered to be a traceable tree structure, so the result of the analysis will be to create a syntax tree corresponding to the proposition containing all the corresponding syntax information. Therefore, the use of neural network algorithms for grammatical analysis can reduce the waiting time for syntactic analysis.

C. Translation Algorithm Based on MLT

The repetitive neural network consists of an input layer x and an output layer y, one or more hidden layers h. Taking the sequence labeling task as an example, the operating mechanism of the repetitive neural network is described as follows: When the time sequence reaches time t, the input plane receives the attribute vector of the t word Wt to represent Xt. The attribute expression here can be a single encoding of attribute words, or dense vector attributes or sparse encoding. The size of the input plane and the dimension of the feature vector must be the same. The output level gives the distribution probability of the word Wt assigned to each label, and its dimension is consistent with the number of labels. Unlike neural networks, cyclic neural networks establish a connection between the previous h-1 state of the hidden layer and the ht state of the current time. Because of this, the design allows the iterative neural network to store historical information, that is, it has memory capabilities. The hidden layer h and the output plane y can be calculated by the following formula:

$$h(t) = \text{sigmiod}(Ux_t + \text{W}h_{t-1}) \quad (1)$$

$$y_t = softmax(Vh_t) \quad (2)$$

Among them, U, W, and V are the weight parameters that need to be optimized during the network training process. The calculation formulas of the nonlinear activation functions sigmoid and softmax are as follows:

$$\text{signoid}(z) = \frac{1}{1+\text{e}^{-\text{z}}} \quad (3)$$

$$\text{sofe max}(z_m) = \frac{\text{e}^{\text{z}m}}{\sum_n^{\text{e}^{\text{z}_\text{k}}}} \quad (4)$$

III. ETS DESIGN BASED ON MLT

A. English Corpus

The process of English translation can be simply regarded as a process of reasoning and analysis of applied knowledge. Knowledge representation is the basic requirement of this process. We divide the knowledge expression used in the English translation process into internal knowledge and external knowledge. External knowledge is stored in a knowledge base and managed by language workers. It is mainly a kind of basic knowledge used to describe the grammar and cultural and semantic characteristics of translation and analysis sentences, such as tree diagrams, cultural characteristics structure and its semantic network.

The external knowledge display of this translation software is mainly processed through the knowledge base in the translation software system. The knowledge base in the translation system mainly includes various language models, dictionaries, different rule bases and professional words. These dictionaries can be roughly divided into two categories: basic bilingual dictionaries and specialized monolingual dictionaries. The content of the rules includes simple phrase rules, sentence pattern rules, topic matching rules, English-Chinese conversion rules, etc. All the rules adopt a single data structure. Therefore, the local related rules in the use of the dictionary are exactly the same as the ordinary rules, and the local rules are required to take precedence over the ordinary related rules we use. This is particularly useful for dealing with specific vocabulary closely related to specific vocabulary. The part of dictionary and its rule maintenance refers to the function of dynamic query, modification, addition, and deletion of various dictionary rules. The case warehouse is used to store example sentences and related materials in both English and Chinese bilinguals.

B. Translation Algorithm

This paper introduces MLT into translation software. Keep special signs or keywords and form a sentence box with the rest. Next, use the bilingual extension MNP to train TNMT model. Finally, separate the longest noun phrase and sentence frame for translation and reorganize to get a complete translation. This work introduces a strategic division and conquest, which ultimately effectively alleviates the sensitivity of neural machine translation along the sentence length.

C. Interface Function

The user interface is an important factor in the realization of human-computer interaction. According to the functions of the translation software, the software interface in this article includes user login, English-Chinese translation, and Chinese-English translation functions.

IV. TESTING OF ETS

A. Test Design

Based on the test of ETS, TNMT method based on the longest noun phrase divide-and-conquer strategy proposed in this article should be tested first, and then the quality of the software translation should be tested.

B. Translation Method Test Results

This paper uses TNMT technology method based on the longest time noun phrase divide-and-conquer strategy to extract 5 experimental data from a corpus in the software. It can be calculated by extracting the result of the software mnp

and the result of manual labeling. The performance of system mnp recognition. Related data results are shown in Table I:

TABLE I. TRANSLATION METHOD

	Accuracy	Recall rate	F value
1	73.4%	72.5%	73.09%
2	72.4%	72.3%	73.04%
3	74%	72.4%	73.01%
4	73.5%	73%	73.03%
5	73.6%	72.6%	73.04%

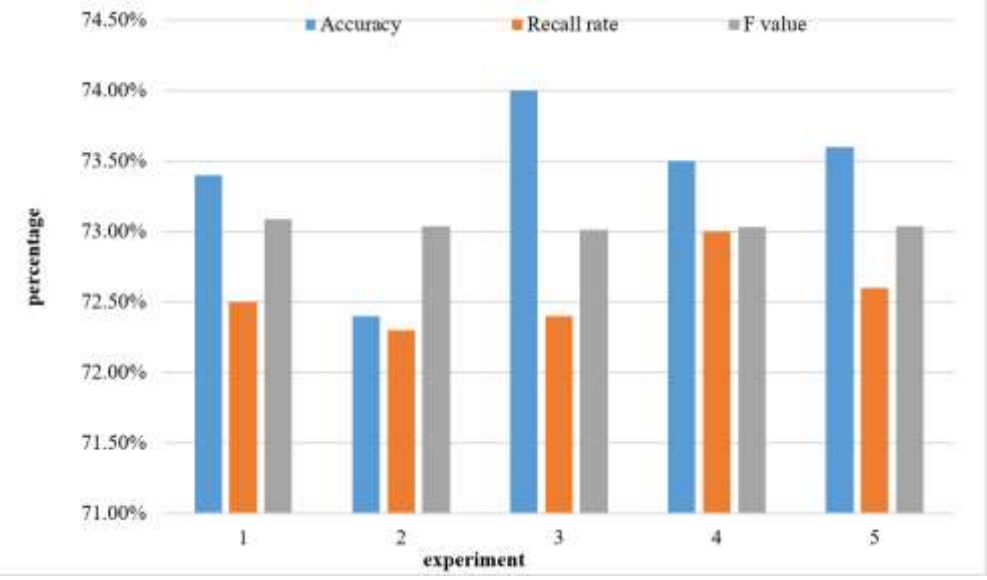

Fig. 1. Translation method test results.

It can be seen from Fig. 1 that the accuracy rate fluctuates around 73%, and the recall rate fluctuates around 73%. This shows that the recognition performance of the translation method is equivalent.

C. Translation Quality Test

This paper uses the case-insensitive 4-gram BLEU method integrated in the Moses open-source system to automatically evaluate the quality of the translation. The relevant data results are shown in Table II:

TABLE II. TEST RESULTS

	Baseline	"MNPi" logo	MNP core words
0-20	38	36	35
20-40	28	27	28
40-60	26	27	28
60-80	23	25	26
80-100	18	20	23

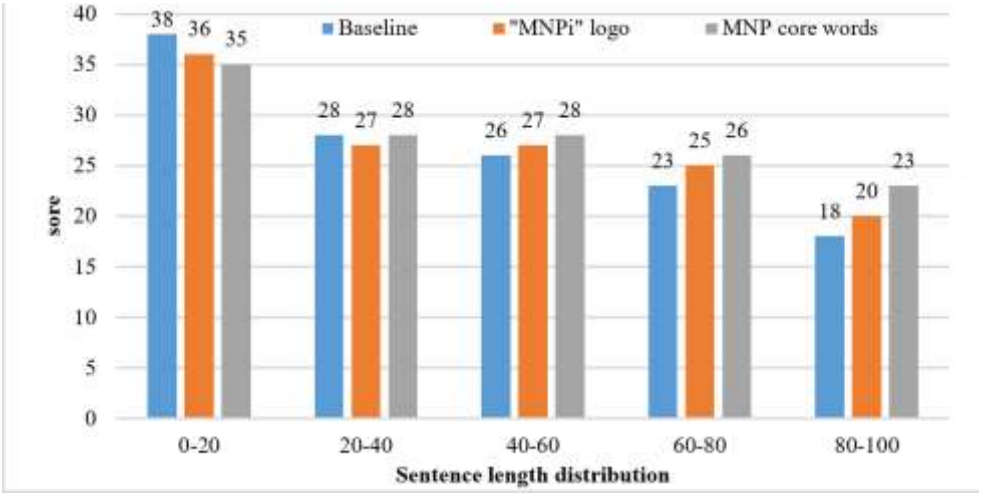

Fig. 2. Translation quality test

It can be seen from Fig. 2 that with the increase of sentence length distribution, TTQ of translation software is decreasing. When the sentence length is 0-20, TTQ is better, but as the sentence length is 20 or later, TTQ decreases significantly dropped by about 10 points. The experimental results show that the translation method proposed in this article has a recognition error rate of about 27%, but it affects subsequent translation. The quality of the translation increases with the distribution of sentence length, the quality of translation is also reduced.Thus the limitation of this research paper is it has a lot time complexity.

V. Conclusions

This article is based on the design and research of ETS based on MLT. By summarizing the problems of translation software, introducing machine translation technology into translation methods to solve some related problems, and then testing the designed translation software, mainly to propose the translation method is tested and the English quality of software translation is evaluated. The final experimental results show that our MLT produces a significant improvement in the translation performance of the machine translation system. From the point of view of the completeness of the theoretical framework, a complete translation system requires two corresponding theoretical systems for the incoming and outgoing languages, and our future work will focus on the establishment of a rule system for syntactic construction.

References

[1] Hasler B S, Salomon O, Tuchman P, et al. “Real-time gesture translation in intercultural communication.” Ai & Society, vol. 32, no. 1, pp. 25-35, 2017.

[2] Ali M, Rahman M L, Chaki J, et al. “Machine translation using deep learning for universal networking language based on their structure.” International Journal of Machine Learning and Cybernetics, no. 3, pp. 1-12, 2021.

[3] Ajmi C, Zapata J, Elferchichi S, et al. “Deep Learning Technology for Weld Defects Classification Based on Transfer Learning and Activation Features.” Advances in Materials Science and Engineering, vol. 2020, no. 1, pp. 1-16, 2020.

[4] Dwyer D B, Falkai P, Koutsouleris N. “Machine Learning Approaches for Clinical Psychology and Psychiatry.” Annual Review of Clinical Psychology, vol. 14, no. 1, pp. 91-118, 2018.

[5] Helma C, Cramer T, Kramer S, et al. “Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds.” J Chem Inf Comput, vol. 35, no. 4, pp. 1402-1411, 2018.

[6] Baydin A G, Pearlmutter B A, Radul A A, et al. “Automatic differentiation in machine learning: A survey.” Journal of Machine Learning Research, vol. 18, no. 153, pp. 1-43, 2018.

[7] Poret N, Twilley R R, Coronado-Molina R M. “Object-based correction of LiDAR DEMs using RTK-GPS data and machine learning modeling in the coastal Everglades.” Environmental Modelling and Software, vol. 112, no. 3, pp. 491-496, 2018.

[8] Lamperti F, Roventini A, Sani A. “Agent-Based Model Calibration using Machine Learning Surrogates.” Journal of Economic Dynamics & Control, vol. 90, no. MAY, pp. 366-389, 2018.

[9] JF Hernández, Z Díaz, Segovia M J, et al. “Machine Learning and Statistical Techniques. An Application to the Prediction of Insolvency in Spanish Non-life Insurance Companies.” The International Journal of Digital Accounting Research, vol. 5, no. 9, pp. 1-45, 2020.

[10] Adebayo G O, Yampolskiy R V. “Estimating Intelligence Quotient Using Stylometry and Machine Learning Techniques: A Review.” Big Data Mining and Analytics, vol. 5, no. 3, pp. 163-191, 2022.

[11] Haedar A, Jorge M, Javad R. “Patient privacy in smart cities by blockchain technology and feature selection with Harris Hawks Optimization (HHO) algorithm and machine learning.” Multim. Tools Appl, vol. 81, no. 6, pp. 8719-8743, 2022.

[12] Ahmad A, Andy S, Michael U. “Automatically extracted parallel corpora enriched with highly useful metadata? A Wikipedia case study combining machine learningand social technology.” Digit. Scholarsh. Humanit, vol. 36, no. 1, pp. 1-15, 2021.

[13] Masoomeh J, Shaahin H, Nader B. “Enhancing Reliability of Emerging Memory Technology for Machine Learning Accelerators.” IEEE Trans. Emerg. Top. Comput, vol. 9, no. 4, pp. 2234-2240, 2021.