

Applications Research of Machine Learning Algorithm in Translation System

Lu Yang¹, Da Chen¹, Wenxue Wu²

Corresponding Author: Lu Yang

1. School of Foreign Languages and Cultures, Xihua University, Chengdu, China
2. School of Science, Xihua University, Chengdu, China

Keywords: Machine learning algorithm, Deep learning, Translation system

Abstract: In recent years, machine translation has made outstanding achievements, but there are also the problems such as fuzzy rules, insufficient data and improper order in the machine translation practice, which makes it still unsatisfactory. Machine learning can abstract learning features, and establish complex mapping relationship between input and output signals, and effectively improve the problems existing in the translation system. The concepts and common models of machine learning and deep learning are given in this paper. Based on the neural model, the paper analyses the rules, structure and models of machine translation system. According to the requirement of the translation system, this paper develops a translation system based on machine learning, and gives the design process of the pre-processing module, the coding module, the attention module and the decoding module. Practice has proved that the system has a good performance in translation performance.

1. Introduction

Machine translation refers to the process of automatically translating a natural language into another natural language with the same meaning. Machine translation is an important issue in the field of artificial intelligence. It is one of the main tasks in natural language processing. It is closely related to the centrally theoretical topics of syntactic analysis, semantic understanding and natural language generation. Machine translation is one of the most beautiful ideas at the beginning of the birth of the computer and even the whole of the whole human being. It is becoming more and more alive with the development of the world economy and the urgent need of cultural exchange. Machine Translation has gone through a tortuous road of development. As for the current situation, there are still many problems to be solved in machine translation. In recent years, researchers have found that deep learning can better alleviate the problems such as linear uncertainty, lack of appropriate semantic representation, difficult features of design, difficult to make full use of non-local context, data sparsity and error propagation, so the Machine Translation based on depth learning has become the present research in Machine Translation. Look at the hot spot. Machine translation can be roughly divided into two types of Machine Translation based on rules and corpus. The former advocates that experts make translation rules and conversion process of computer, and the advantage lays in the ability to explicitly describe the law of language transformation and the strong interpretability of the deep language. The disadvantages are the high dependence on human experts and the long system development cycle. The latter advocates the use of computer to automatically discover and induce the law of translation from the corpus, and to solve the bottleneck problem of the traditional methods of translation knowledge acquisition, which has become the mainstream method of machine translation.

Machine translation still faces serious challenges in the following aspects. Machine translation performs a process of semantic re generation. The statistical machine translation method mainly implements the conversion of two language texts in words, phrase fragments and syntactic levels. The disadvantages are poorly expressive ability, low computability, and lack of appropriate translation to support semantic level. Traditional methods rely on linguists to characterize various rules of translation in the form of features, such as in a reverse transcriptional grammar-based order

model, which usually designs features based on the artificial experience of a word string. However, the transformation of language structure is very complicated, and the characteristics of artificial design are difficult to cover the complicated linguistic phenomena. The statistical machine translation method usually deals with words and phrases as independent symbols. Even if the large-scale corpus training model is used, it is difficult to alleviate the serious problem of data sparsity. Traditionally, the translation process is usually divided into several modules according to human experience. For example, syntactic based machine translation methods usually perform lexical analysis and syntactic analysis first, and then transform the tree to string or tree to tree in the result of syntactic analysis. Any errors in the translation pipeline will propagate backwards, which will seriously affect the translation performance of the whole system. In this paper, a neural machine translation method based on the divide and conquer strategy is proposed to alleviate the sentence-length sensitivity of the neural machine translation based on the shortcomings of the current end to end machine translation.

2. Theory of machine learning algorithm

2.1. Machine learning

With the birth and rapid popularization of computer, the machine learning algorithm has been gradually discovered and improved by scientific researchers. It has been applied in all fields of human life, which has rapidly promoted the progress of human society. Machine learning is a new multi-field interdisciplinary subject, which involves the important theories of probability theory and statistics. It has an important contribution to the problem of data mining. Through the training of many known data classifiers or algorithms, through the accumulation of continuous data obtained by the system itself, and then self-improvement and continuous improvement of performance, it can solve the decision function which is beneficial to solve the classification or regression problems, and quickly and accurately predict similar unknown samples. Step by step, the machine can replace manual work to complete repetitive and complex tasks. However, because the specific data mining problem is different, and the time and storage cost of the solution have different requirements, so the use of machine learning algorithms is completely different. According to the learning form of machine learning algorithm, it can be roughly divided into supervised learning and unsupervised learning. Popularly speaking, supervised learning is more than unsupervised learning to label the data. We have a data set. If we want to judge the label of every single data according to its feature vector, then we have supervised learning. In unsupervised learning, all data only have no tags, but it can be found that these data are clustered in structure, and the essence is that a similar type will gather together. Clustering these unlabelled data into one combination is clustering. Google news, for example, collects a lot of news every day, then clusters them all, and automatically divides them into dozens of different groups, each of which has a similar content structure.

2.2. Deep learning

Deep learning is a kind of machine learning method. It is used to deal with the analysis and classification of various patterns by multi-layer nonlinear transformation and the automatic learning and hierarchical abstract representation of the features of original input. Corresponding to deep learning is all kinds of traditional machine learning methods. The shallow machine learning method generally uses a linear function to establish the connection between the input characteristic space and the output. Because there is a general nonlinear relationship between the original input signal and the output to obtain better learning effect, the traditional learning method needs to convert the original input signal to the output approximate line explicitly or implicitly. The characteristic space can be divided by sex. Artificial feature transformation or nonlinear kernel function design requires a lot of domain knowledge, and even some tasks are difficult to find effective methods. Unlike these shallow methods, the depth learning directly starts from the original input signal and combines the low-level features into the more abstract features, and the complex mapping relationship between the learning input and the output. Compared with shallow structure, deep structure can use fewer

parameters to approximate more complex mapping relationship, which is a more compact representation. Deep learning can automatically learn distributed representation of features and can better promote global generalization. Deep learning can automatically learn hierarchical feature abstraction, reduce workload of artificial feature changes, and generate new useful features. The intermediate representation learned from a task can be applied to another field, which is a natural multi-task learning method. Deep structure has some biological support. It is a direction worth exploring in the field of artificial intelligence. The current upsurge in neural network research is mainly caused by its good experimental results on various artificial intelligence tasks, and most of the research focuses on the design of a suitable deep neural network structure for specific tasks.

2.3. Common neural network model

Convolution neural network is a special kind of neural network. It is different from the general fully connected neural network. The convolution neural network has a certain topology, a neuron is connected only with the input unit in a certain area, and the parameters of the different neuron connections are shared. Convolution neural network greatly reduces the number of parameters in the neural network, and this structure has the explanation of local feature extraction in the image recognition task. Multilayer convolution neural network is the first successful training of deep neural network and continues to play an important role in the field of image processing. The convolution neural network is a non-fully connected multilayer neural network, which is usually composed of the coiling layer, the lower sampling layer and the fully connected layer. First, the original image is convoluted with the coiling layer and the filter, and several feature maps are obtained. Then the feature is blurred through the lower sampling layer, and then a group of special links is used to output a group of special images. Eigenvector is used for classification and recognition of classifier. The hidden layer of convolution neural network is composed of the volume layer and the down sampling layer alternately. The input target image is convoluted by convolution kernel, and the layer with characteristic graph is obtained. Then the image in the feature graph is pooled, and the size of the pooling scale can be set according to different needs, and then the feature layer with the pooled feature is obtained. Finally, the feature of image recognition is obtained through the full link layer, and the features obtained by multi class classification are classified. The connection between neurons is not fully connected, and the weights of connections between neurons in the same feature graph are shared. The network structure with incomplete connection and weight sharing reduces the complexity of the network model, the number of weights, and facilitates parallel learning.

3. Translation system analysis based on deep learning

3.1. Rule analysis of translation system

Machine translation is a study spanning many fields. Machine translation rules system can be divided into two categories, binary translation rules and ternary translation rules. The binary translation rule system refers to the system consisting of two elements, corpus and machine grammar. The so-called ternary element translation rule system refers to the system consisting of three elements, corpus, grammar rule table and interpreter. When using this structure, the difference between the different translation requirements system is only different from the corpus and the grammatical rule table. The interpreter can be used as a general program, which is more suitable for the modular development idea of the program and improves the efficiency and maintainability of the program development. The basic model of machine translation system is natural language processing system. The basic principle of the machine translation system is the principle of element synthesis. First, it decomposes the sentence into the basic constituent elements, and uses syntactic analysis, semantic analysis and context analysis to exclude improper ambiguities to form the internal representation of the original text. Then, the structure of the sentence structure is obtained, and the appropriate translation words are selected. After that, word order adjustment, function word addition and deletion, and morphological changes will finally get the surface sentence of the

translation. In addition to the executable unit support, such a system needs a language knowledge base, including static dictionaries, syntax rules, etc., and dynamic context related information. The early statistical machine translation model is a lexical based model. The lexical based translation model takes the translation pairs as the basic translation unit. As the lexical level of translation is ambiguous, the order phenomenon is obvious, and the lexical translation model does not record the context information. These factors make the translation model based on the vocabulary difficult to achieve a very good effect. The lexicon-based translation model is not the mainstream statistical machine translation model.

3.2. Structure analysis of translation system

According to the guidance of the above linguistic algorithms, we can understand that the development of the Machine Translation software system is related to the content of the translation rules, and on the other hand, what development tools are used in the design of the program. The development tool used here is Delphi, which makes it easier to realize the independence of modules and syntactic rules in the translation system. In the model, the language rules and algorithms are separated and the algorithm becomes the controller and rule interpreter of the system. The rule interpreter is the core of the whole system. In addition to good algorithm support, a good translation system needs efficient language information library support. In this way, the coding and natural language mathematical modelling are used. The expression is designed to be a marked tree structure. The markers on the node are mainly three levels: grammatical category, syntactic function and logical semantics, which not only make addition to the grammatical category, the syntactic function and the logic semantics. It works flexibly and ensures the continuity of algorithm when there is a lack of information. X is a grammatical category, Y is syntactic function, and Z is a logical semantic relation. There are two kinds of X values: one is the category name, the other is the K, and the other is the category of the word category, which is the ultimate symbol and is expressed by K. Grammar and grammar are the static set of rules in grammar rule base. Grammar itself can not affect the process of the execution of the rules. However, the entire analysis algorithm is connected to each part of the algorithm, which is left and right, combined to detect the various structures of the semantics, which is easy to cause the combination explosion. In this way, it is difficult to use the dichotomy and multilevel indexing to speed up the translation process in the whole grammar. Figure 1 is the basic structure of the machine translation system.

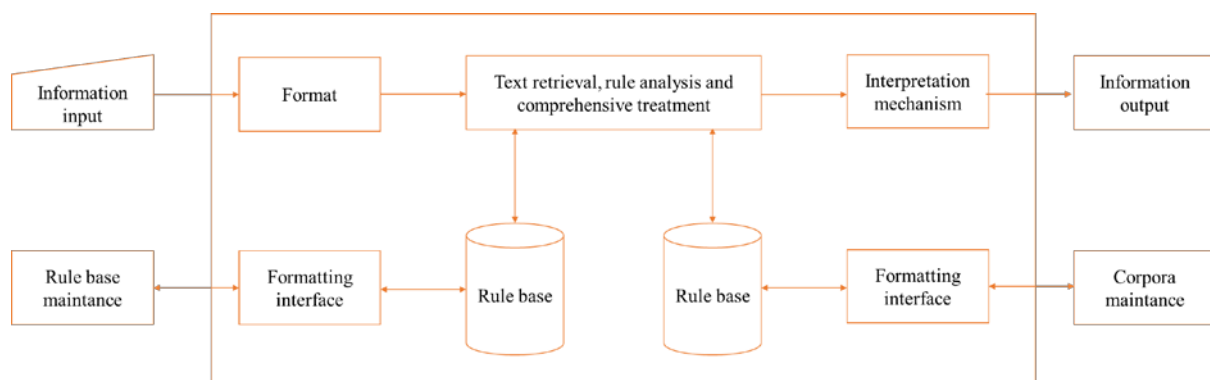


Figure 1. Structure of machine translation system

3.3. Module analysis of translation system

The word order model is used to deal with word order differences in different languages. The problem of ordering is a difficult point in statistical machine translation research. In the translation system of a tree to a string or a tree, the translation model itself contains the order information, which may not be particularly used in these translation systems. The translation system based on phrase and formal grammar often has a special ordering model. The order model can be used as a characteristic of the translation log linear model, and the translation candidate is evaluated in the

process of translation decoding. The commonly used order models include lexicalization order models, based on multiple sparse features, and so on. We propose a deep neural network-based translation model and achieve good results. The search space of the machine translation system is huge, especially after the introduction of the global features such as the meta grammar model, and the lack of effective dynamic programming means. Therefore, it is necessary to adopt the approximate decoding method. Commonly used machine translation decoding algorithms include finite state transition. The two algorithms can be regarded as implicit construction and traversing hypergraph. The decoding algorithm based on this algorithm can consider various decoding constraints flexibly, which is a commonly used decoding algorithm in practice. To speed up the decoding process, the two decoding algorithms need a lot of pruning. Machine translation automatic evaluation standard is a necessary condition for machine translation model discriminant training, and an important index to quickly measure the quality of machine translation system. The error rate minimized and the error function used in statistical machine translation is a function specific to the parameter, which can only find the local optimal, and each optimization result varies greatly with the initial value because of the error rate minimized and the error function used in statistical machine translation. When the number of parameters is small, good results can be achieved, but for many parameters, it is difficult to achieve very good results.

4. Design and implementation of translation system based on machine learning algorithm

4.1. Design

The design and implementation of the Chinese English neural machine translation system mainly includes pre-processing module, encoding and decoding module and attention module. The sequence of each module in accordance with the system includes pre-processing module, encoding module, attention module and decoding module. The four modules work together to achieve end-to-end neurologic Machine translation process. The functions of each part are described as follows. Pre-processing module: this stage is based on the MNP segmentation method proposed in this paper, the phrase structure parser is used to analyse the phrase information, and then the MNP can be extracted and the sentence framework is obtained. In view of the multi sequence coding method which introduces speech and dependency information, the language text of the source language text is tagged and dependent on the dependency syntactic analysis by using the language technology platform of Harbin Institute of Technology. The above two steps need to be built based on Chinese word segmentation. Coding module: This module is part of the end-to-end machine translation's encoder decoder" framework. A bidirectional GRU network is used to map the ID sequence corresponding to a sentence into a continuous high-dimensional vector representation. Attention module: this part mainly realizes the dynamic and selective attention to each part of the source language sentence in the generation of target language words to obtain more accurate context vector as the basis for translation. Decoding module: corresponding to the encoding module, it is another part of the framework. The function of this module is to map the semantic representation of vector form into natural language sentence. We use nonlinear models instead of linear models to solve the problem of linear inequalities of large scale linguistic data in statistical machine translation. We use a single complex neural network instead of complex hidden structure pipelining, which greatly simplifies the translation process and solves the problem of error propagation in the multi module collaboration process of statistical machine translation. We use a recursive neural network substitution method to capture the infinite length from the context of a certain window size. Historical information alleviates the problem that statistical machine translation cannot make full use of non-local context.

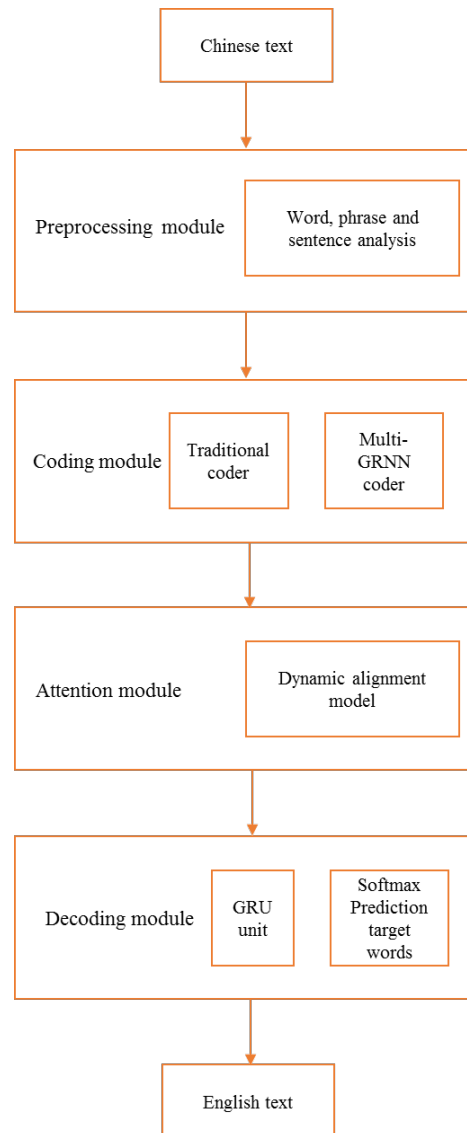


Figure 2. System modules design

4.2. Implementation

Overall, divide and conquer the longest noun phrase and sentence frame by the divide and conquer strategy, and in part use the method of multi sequence coding to translate the longest noun phrase and sentence frame. In the previous part of the experiment, we have introduced two methods to pre-process data. The process of integration can be described as follows. The common user interface is the main way for ordinary users to use the system, mainly including the source language box and the target language box, as well as translating and emptying buttons. After the input source language sentence, click the text translation button, a translation request can be generated, the background translations results are arranged to return to the final translation to the target language box. The left side frame of the main translation interface is the source language frame and the right is the target language frame. The phrase structure parser can only analyse the boundary of the noun phrase, prepositional phrase and so on. When extracting the longest noun phrase, we need to extract noun phrases which are not nested by other noun phrases according to their definitions. In view of the MNP based method, we use Berkeley Parser to analyse the phrase structure syntactic analysis of Chinese sentences with length. Then, in view of the multi sequence coding method, the Chinese sentences are divided, part of speech tagging and dependency analysis using the language technology platform tool of Harbin Institute of Technology. Finally, the Chinese sentences are divided into word segmentation, part of speech tagging and dependency analysis. According to the result of the phrase structure analysis, the MNP is extracted in Chinese sentences and the corresponding parts of the sequence of word and epistasis are extracted and the sequence of word

and epistasis of MNP and sentence frame are formed. In this paper, GRNN is used to replace the traditional RNN, to alleviate the problem of "gradient disappear" in the training process. In GRNN, considering the comparison of LSTM, the design of GRU network is more succinct, and it can reduce the time and space cost of network training. Therefore, this paper uses the bidirectional GRU network to encode the source language end of the system and uses the GRU network to decode the target language end. To achieve better translation performance, the system introduces the attention mechanism, that is, dynamically aligning the source language sentences when generating target words. To improve the performance of the system, the lexical and syntactic information are incorporated into the neural network translation model by giving the sentence frame, tagging the word and the epistatic information. From the experimental results, the method based on MNP divide and conquer strategy can alleviate the problem of Machine Translation's sensitivity to sentence length to a certain extent.

5. Conclusions

Based on the concept of machine learning, the specific design methods and the implementation processes of machine translation system are presented in this paper. The main conclusions are as follows:

(1) This paper analyses the similarities and differences between machine learning and deep learning and gives the common neural network modules including convolutional neural network.

(2) This paper explores the translation rules of the system (binary translation rules and ternary translation rules), structure (Figure 1) and modules of machine translation system (order module, language module and translation module).

(3) The design and implementation of pre-processing module, the coding module, the attention module and the decoding module are introduced in detail in this paper.

References

- [1] Haque A U, Mandal P, Meng J, et al. Wind speed forecast model for wind farm based on a hybrid machine learning algorithm[J]. *International Journal of Sustainable Energy*, 2015, 34(1): 38-51.
- [2] Bahar P, Alkhouli T, Peter J T, et al. Empirical investigation of optimization algorithms in neural machine translation[J]. *The Prague Bulletin of Mathematical Linguistics*, 2017, 108(1): 13-25.
- [3] Wu Y, Schuster M, Chen Z, et al. Google's neural machine translation system: Bridging the gap between human and machine translation[J]. *arXiv preprint arXiv:1609.08144*, 2016.
- [4] Balahur A, Turchi M. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis[J]. *Computer Speech & Language*, 2014, 28(1): 56-75.
- [5] Van Merriënboer B, Bahdanau D, Dumoulin V, et al. Blocks and fuel: Frameworks for deep learning[J]. *arXiv preprint arXiv:1506.00619*, 2015.
- [6] Schmidhuber J. Deep learning in neural networks: An overview[J]. *Neural networks*, 2015, 61: 85-117.
- Jean S, Cho K, Memisevic R, et al. On using very large target vocabulary for neural machine translation[J]. *arXiv preprint arXiv:1412.2007*, 2014.
- [7] Najafabadi M M, Villanustre F, Khoshgoftaar T M, et al. Deep learning applications and challenges in big data analytics[J]. *Journal of Big Data*, 2015, 2(1): 1.

- [8] Podolsky M D, Barchuk A A, Kuznetsov V I, et al. Evaluation of machine learning algorithm utilization for lung cancer classification based on gene expression levels[J]. *Asian Pacific Journal of Cancer Prevention*, 2016, 17(2): 835-838.
- [9] Lochner M, McEwen J D, Peiris H V, et al. Photometric supernova classification with machine learning[J]. *The Astrophysical Journal Supplement Series*, 2016, 225(2): 31.
- [10] Shen S, Cheng Y, He Z, et al. Minimum risk training for neural machine translation[J]. *arXiv preprint arXiv:1512.02433*, 2015.
- [11] Wong T Y, Bressler N M. Artificial intelligence with deep learning technology looks into diabetic retinopathy screening[J]. *Jama*, 2016, 316(22): 2366-2367.
- [12] Wohlberg B. Efficient algorithms for convolutional sparse representations[J]. *IEEE Transactions on Image Processing*, 2016, 25(1): 301-315.
- [13] Yan Z, Zhan Y, Peng Z, et al. Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition[J]. *IEEE transactions on medical imaging*, 2016, 35(5): 1332-1343.
- [14] Chang X, Ma Z, Lin M, et al. Feature interaction augmented sparse learning for fast kinect motion detection[J]. *IEEE Transactions on Image Processing*, 2017, 26(8): 3911-3920.
- [15] Deng L. A tutorial survey of architectures, algorithms, and applications for deep learning[J]. *APSIPA Transactions on Signal and Information Processing*, 2014, 3.
- [16] Truong H M. Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities[J]. *Computers in human behavior*, 2016, 55: 1185-1193.
- [17] Zhuang L, Chan T H, Yang A Y, et al. Sparse illumination learning and transfer for single-sample face recognition with image corruption and misalignment[J]. *International Journal of Computer Vision*, 2015, 114(2-3): 272-287.