On Application of Natural Language Processing in Machine Translation

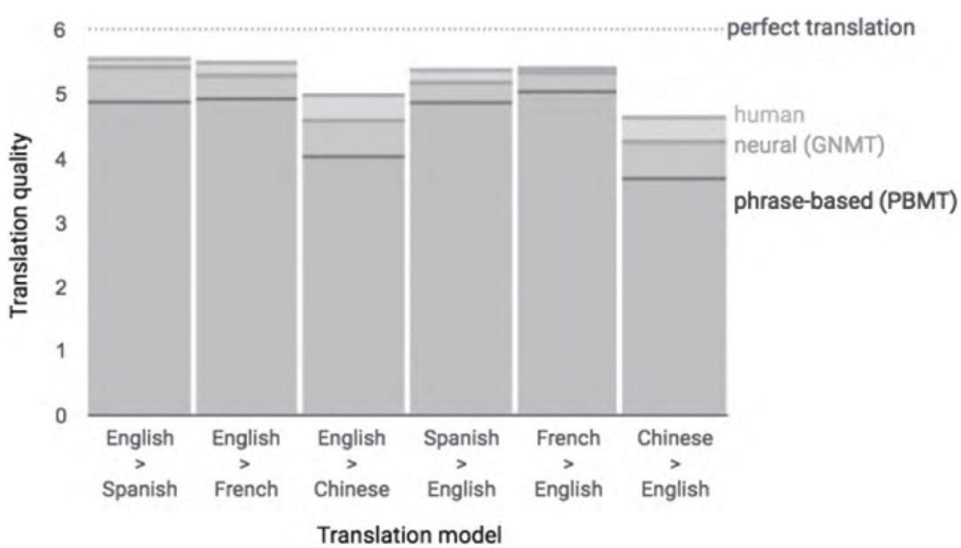
Zhaorong Zong[1] Changchun Hong[2]
School of Foreign Languages, Huangshan University, 39 Xihai Road, Tunxi District, Anhui Province, China
Corresponding author e-mail: Zongzhaorong@hsu.edu.cn hongcc@hsu.edu.cn

Abstract: **Natural language processing is the core of machine translation. In the history, its development process is almost the same as machine translation, and the two complement each other. This article compares the natural language processing of statistical corpora with neural machine translation and concludes the natural language processing: Neural machine translation has the advantage of deep learning, which is very suitable for dealing with the high dimension,label-free and big data of natural language, therefore, its application is more general and reflects the power of big data and big data thinking.**

Key Words: natural language processing; machine translation; decoding

I. Introduction

On September 27, 2016, scientists Quoc V. Le and Mike Schuster from Google Brain Group published a blog post on Google Research Blog: "A product-scale neural network for machine translation", declaring that Google once again made a major breakthrough in the field of machine translation and introduced a new machine translation system called GNMT after Google introduced the phrase-based machine translation system "Google Translate" a decade ago. (Le & Schuster 2016). Google also published a paper on arXiv that detailed the technical mechanism of GNMT's working mechanism (Wu et al. 2016). Experimental results using Wikipedia and news corpus as test data show that the classic phrase-based statistical machine translation model GNMT significantly reduced the translation error rate between several key language pairs by 55% to 85%. Figure 1 shows that the quality of machine translation from French to English, English to Spanish is very close to the quality of human translation[1].



II. Machine Translation and Natural Language Processing in Statistics Corpus

Prior to 1990, machine translation systems could be grouped into three basic types: direct translation, intermediate language translation, and transfer translation (Hutchins, 2009: 505-509). On the Fourth high-level meeting of Machine Translation held in Kobe, Japan in July 1993, the famous British scholar W.J. Hutchins declared in his special report that since 1989 the development of machine translation has entered into a new era. An important symbol of the new era is the introduction of a corpus in machine translation technology, namely the corpus converted database into machine translation by means of natural language processing[2]. In recent years, corpus-based machine translation systems have developed rapidly and achieved outstanding results.

978-1-5386-8481-8/18/$31.00 ©2018 IEEE
DOI 10.1109/ICMCCE.2018.00112

As early as 1947, Warren Weaver put forward his method of machine translation using the method of decoding passwords in his memorandum entitled "Translation". This so-called method of "interpreting the password" is essentially a kind of natural language processing. He wanted to use a statistics-based method to solve machine translation problems[3]. Based on Weaver's ideas, IBM's Peter Brown and other scientists put forward the mathematical models for statistical machine translation. Statistical-based machine translation treats machine translation problems as a noisy channel problem, as shown in Figure 1.

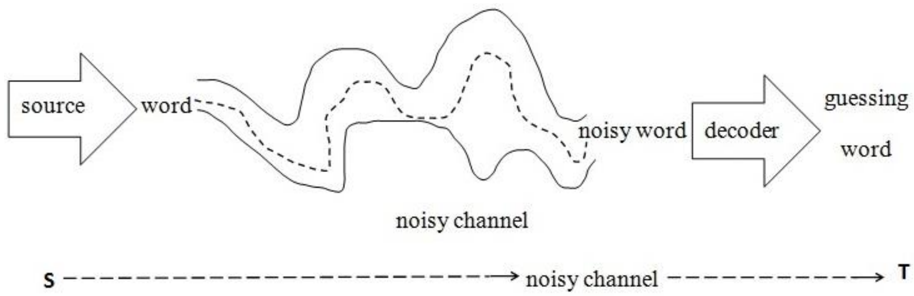


Fig 1. noisy channel model

Machine translation can be seen in this way: Source language is distorted due to passing through a noisy channel and appears as Target language at the other end of the channel. The translation problem is actually how to recover the most likely Source language from the observed Target language. The Source language is the input in the channel sense and the Target language in the sense of translation, while the Target language is the output in the channel sense and the Source language in the sense of translation. From this point of view, any sentence in a language may be a translation of some sentences in another language, but the possibilities of these sentences are different. Machine translation is to find out the sentence with the highest possibility, that is, calculates the most probable one as the translation of the source language T for all possible target languages S. Due to the large number of S, stack search can be used. The main data structure of stack search is the table structure, which contains the most promising S corresponding to T. The algorithm continues to loop, and each cycle extends some of the most promising results until the table contains an "S" with significantly higher score than other results. Stack search does not guarantee optimal results for it can lead to erroneous translations and is therefore only a suboptimal algorithm[4].

It can be seen that the task of the statistical machine translation system is to find the sentence with the highest probability as a translation result in all possible target languages (A target language in the sense of translation is a source language in the sense of a noise channel model). The probability value can be obtained using the Beyes formula (T in the formula is the target language in the sense of translation, S is the source language in the sense of translation):

$$P(T|S) = \frac{P(T)P(S|T)}{P(S)}$$

Since the denominator P(S) on the right side of the equation is independent of T, finding the maximum value of P(T|S) is equivalent to finding a T so that the multiplication P(T)P(S|T) of the numerator on the right side of the equation is the maximum, that is:

T=argmax P(T)P(S|T)

Where P(T) is the language model of the target language and P(S|T) is the translation model of S under given T. According to the language model and the translation model, the process of solving the target language sentence T closest to the real target in the given source language sentence S is equivalent to the decoding process in the noise channel model.

Based on the idea of statistical machine translation and using British and French bilingual parliamentary debate records as a bilingual corpus, Braun and other researchers from IBM developed an English-French machine translation system, Candide.

Table 1 Comparison of Systran & Candide

	Fluency		Faithfulness		Time Ratio	
	1992	1993	1992	1993	1992	1993
Systran	.466	.540	.686	.743		
Candide	.511	.580	.575	.670		
Transman	.819	.838	.837	.850	.688	.625
Manual		.833		.840		

The table in Table 1 is the test results of ARPA (United States Department of Defense Advanced Research Projects Agency) on several machine translation systems. The first line is the translation result of the famous Systran system, the second line the translation result of Candide, the third line the result of a manual proofreading of Candide and the fourth line the result of a purely human translation. There are two evaluation indicators: Fluency and Adequacy. Transman is a post-translation editing tool developed by IBM. TimeRatio shows the ratio of the time used by Candide plus Transman to manually proofreading and the time spent on pure manual translation. From an index point of view, Candide's natural language processing system has surpassed the traditional system Systran.

III. Neural Machine Translation & Natural Language Processing

Since 2014, end-to-end neutral machine translation has achieved rapid development. Compared with statistical machine translation, the translation quality has been significantly improved. Figure 2 shows the comparative experimental results of statistical machine translation and neural machine translation in 30 languages [5]. Neural machine translation is much better than statistical machine translation in 27 of the natural language processing. Therefore, neural machine translation has replaced statistical machine translation as the core technology of commercial online machine translation systems such as Google, Microsoft, Baidu, and Sogou.

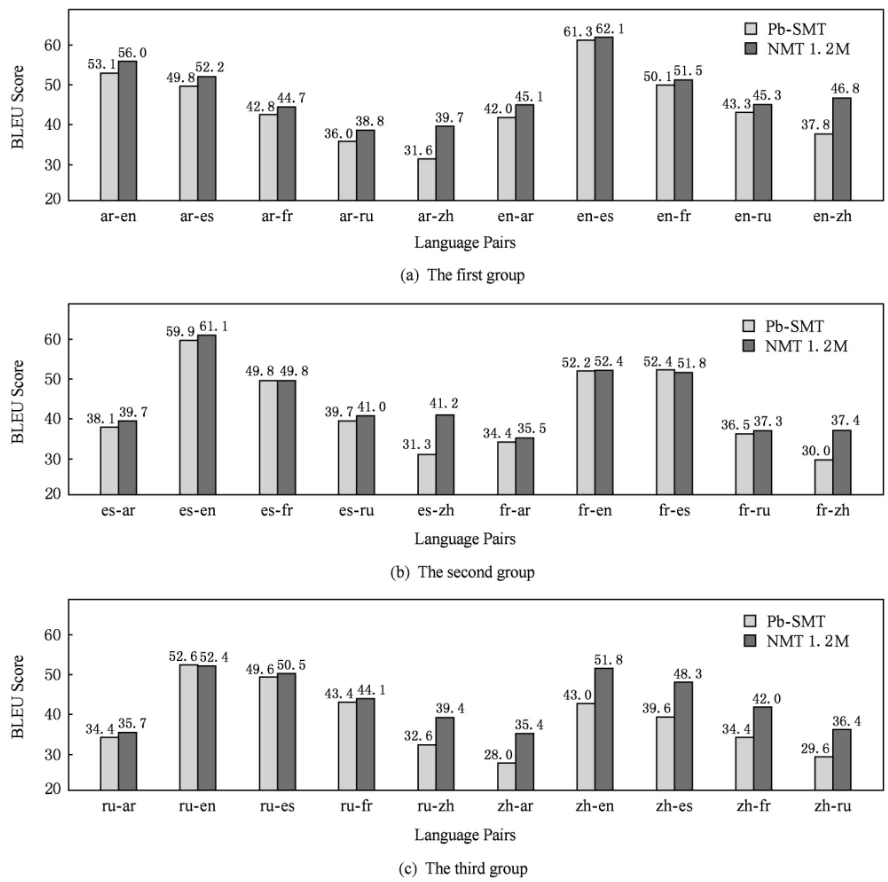


Fig 2. Comparison between statistical machine translation and neural machine translation on 30 language pairs

The basic idea of end-to-end neural machine translation is to directly implement automatic translation between natural languages through neural networks. For this reason, neural machine translation usually uses an encoder-decoder framework to achieve sequence-to-sequence conversion [6]. Taking Figure 3 as an example, given a Chinese sentence "布什 与 沙龙 举行 了 会谈," the encoder-decoder framework first generates a vector representation for each Chinese word and then goes from left to right through a recurrent neural network generating a vector representation of the entire Chinese sentence. Among them, "</s>" indicates the ending terminator. We call the recursive neural network used by the source language as an encoder, that is, encode the source language sentence into a dense, continuous real number vector.

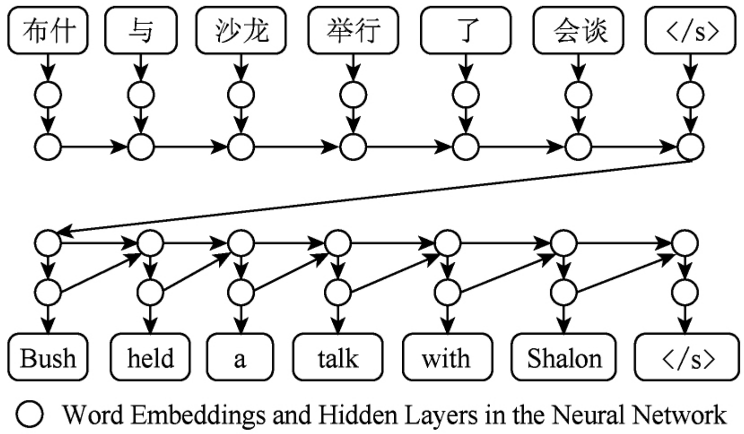


Figure 3 The Encoder-decoder Framework

After that, the target language uses another recursive neural network to reversly decode the source language sentence vector to generate the English sentence "Bush held a talk with Shalon </s>". The entire decoding process is generated word by word, and when the end of the sentence "</s>" is generated, the decoding process terminates. The recurrent neural network used by the target language is called a decoder. It should be noted that each newly generated English word is used as the historical information to generate the next English word. Therefore, the decoder can be regarded as a language model containing the target language of the source language information. Neural machine translation based on the attention mechanism uses a completely different encoder whose goal is no longer to generate a vector representation for the entire source language sentence but a vector representation containing global information for each source language word.

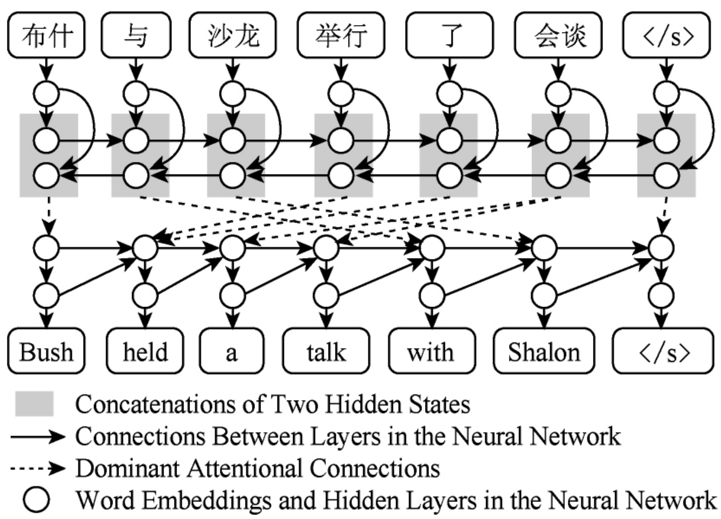


Fig.3 Attention-based Neural Machine Translation

The advantage of this approach is that the vector representation of each source language word contains its left and right context information. On the target language side, the decoder dynamically finds the source language context associated with each target language word. For example, when the English word "Bush" is gererated, the Chinese word "布什" is the most relevant, and the words "Hold" and "Last" may not be relevant. It is only necessary to pass the "Bush" vector representation as the source context to the target. When the English word "held" is generated, the most relevant Chinese words are "举行" and "了." Therefore, the attention mechanism changes the way of information transmission and can dynamically calculate the most relevant context, so as to better solve long-distance information transmission problems and significantly improve the performance of neural machine translation. Therefore, the encoder-decoder model based on attention mechanism has become the mainstream method of neural machine translation and has been widely used.

IV. Conclusion

Natural language processing is a prerequisite and condition for machine translation. In terms of natural language processing, neural machine translation not only has more generality, but also reflects the power of big data and big data thinking. In addition, neural machine translation requires more powerful computing power than statistical machine translation.With the continuous development of artificial intelligence technology, new technologies for natural language processing will change the industrial construction of translation, and the translation industry chain may experience a major industrial change. At present, neural machine translation has been widely used because of the advantage of deep learning and the characteristics of handling high dimension, label-free and big data of natural language.

Acknowledgments

This paper is financially supported by the following : Anhui Social Sciences Key Project Funding SK2016A0878 Shanghai Foreign Language Education Press Project Funding SK2015A812

References

[1] Le, Quoc V. and Mike Schuster. 2016. Neural Network for Machine Translation, at Production Scale.n.d.Sep.27,2016.https://research.googleblog.com/2016/09/a-neural-network-for-machine.htm.

[2] Hutchins, W. John. 1995. Machine Translation: A Brief History. In E. F. K. Koerner and R. E Asher (eds.), Concise History of the Language Sciences: From the Sumerians to the Cognitivists. Oxford: Pergamon Press.

[3] Weaver, Warren. 1955. Translation. In William N. Locke and Andrew Donald Booth (eds.), Machine Translation ofLanguages: Fourteen Essays. Cambridge: MIT Press.

[4] Foster I, Kesselman C, Nick J, et al. Computer Grid Services for Distributed System Integration[J]. IEEE Computer, 2002, 35(6)

[5] Bahdanau D, Chok, Bengio Y. Neural Machine Translation by jointly learning to align and translate [J]. arXiv:1409.0473,2014

[6] Junczys-Dowmunt M, Dwojak T, Hoang H. Is Neural Machine Translation Ready for Deployment? A Case Study on 30 Translation Directions [J]. arXiv:1610.01108v2,2016