Computer Translation

Yuanyuan Zhou

§1 Introduction

This paper discusses ways to improve machine translation. Machine translation is a method of translation using a computer.

Machine translation has recently become an important subject. There are two main reasons for this. Firstly, machine can translate more sentences in a shorter time than humans. Practically, Google Translate (https://translate.google.co.jp) translated a Japanese essay of 1000 words in a moment. The same essay took the present writer 4 hours to translate. Secondly, machine translation has improved rapidly recently, thanks to deep learning. Deep learning can find feature of value from a large amount of data.

This research subject was chosen because of the difficulty in communicating with unfamiliar languages which may not even be understood. Many people might have the experience of having difficulties in communicating, despite their best efforts. Translation can solve this problem.

§2 Literature Review

In order to improve machine translation, the following questions were asked and answers sought in literature books, and interviewing a professor of Cambridge University.
1. What is important for improving computer translation?
2. What is important for improving dictionaries?
3. What is important for improving syntax analysis?
4. What difficulties are there in improving computer translation?
5. How does deep learning work?
6. Is the singularity coming?
Singularity is a point that a computer can produce better one than itself.

§3 Methodology

The following predictions were made to the questions.
1. What is important for improving machine translation?

It is important to improve dictionaries and syntax analyzing for improving machine translation because they are the main process of machine translation.

2. What is important for improving dictionaries?

It is important to gather more data of on the meaning of words.

3. What is important for improving syntax analysis?

It is important to gather more data for deciding on sentence structure.

4. What difficulties are there in improving computer translation?

The ambiguity of language is the main difficulty in translation. People rarely speak in perfect sentences like the example of sentences in textbooks, and even a perfect sentence, can be misunderstood.

5. How does the deep learning works?

Deep learning can be applied to image recognition, however, its exact process is unknown to the answer at the time of writing.

6. Is the singularity coming?

The singularity is theoretically possible, but physically impossible.

Programs or algorithm, which produce more refined versions of themselves, already exist in the real world, for example, alpha Go, but no matter how well they make algorithms, they cannot survive without electricity. So, at least to the day when computers can create electricity by themselves, the singularity

will not come.

However, people are already controlled by many factors; nature, media, or family. Consequently, the singularity is a case where the controller changes. It may seem frightening, but humanity has adapted to many changes in the past.

§4 Results

Professor Copestake (see fig.1) answered the questions from §2.

1. What is important for improving computer translation?
2. What is important for improving dictionaries?
3. What is important for improving syntax analyzing?

Prof. Copestake talked about DELPH-IN, a project for improving machine translation, especially syntax analysis. It searches for the most likely sentence structure. There is a large corpus of data taken from various texts, such as newspapers, chemistry papers, and online wiki, and given to DELPH-IN. These data have information about the structure of sentences,

Fig 1. Professor Copestake(left), Cambridge University.

so the program will learn what words in what row, are most likely to play a certain role in the sentence. When a new sentence is inputted, it searches and shows every possible structure of the sentence, and lets a human operator to choose the right one, then it learns from the feedback, with deep learning.

It is important to register idioms in the dictionary, or data used for analyzing sentences. This is a big key to improving computer translation.

4. What difficulties are there in improving computer translation?

Prof. copestake said machine translation works almost perfectly when it is translating between European languages, but when it comes to translating less related languages, such as Japanese and English, the accuracy decreases.

Machine translation works better in translating legal documents than novels. This is because sentences used in legal documents are often formulaic.

6. Is the singularity coming?

A time frame of 50 years was suggested by Prof. Copestake because improvement in technology is not faster enough.

This answer was from books or papers.
5. How does the deep learning works?

According to the two websites, (http://qiita.com/icoxfog417/items/5fd55fad152231d706c2, https://deeplearning4j.org/ja/neuralnet-overview), deep learning extracts features from a large amount of data, by imitating the neurons of the brain.

It is configured in layers, and the layers are a mass of nodes. Nodes are connected to the between layers, and it transfers the information of inputted data. Data inputted in weight layer would be changed, following the function of weight layer. The node received the data from weight layer decides what data to transfer to the next. In this way, it extracts its feature. The feature changes according to the function.
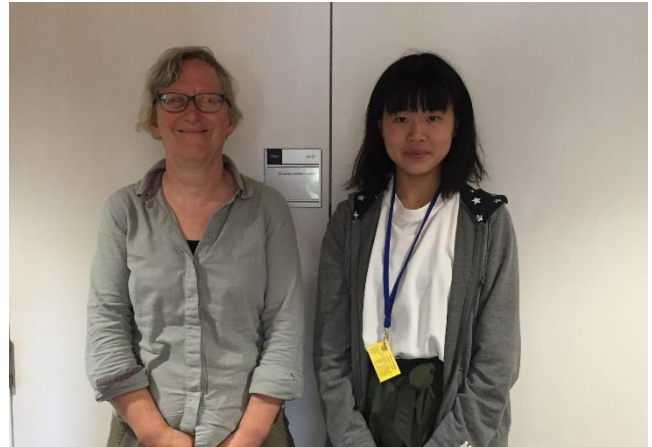
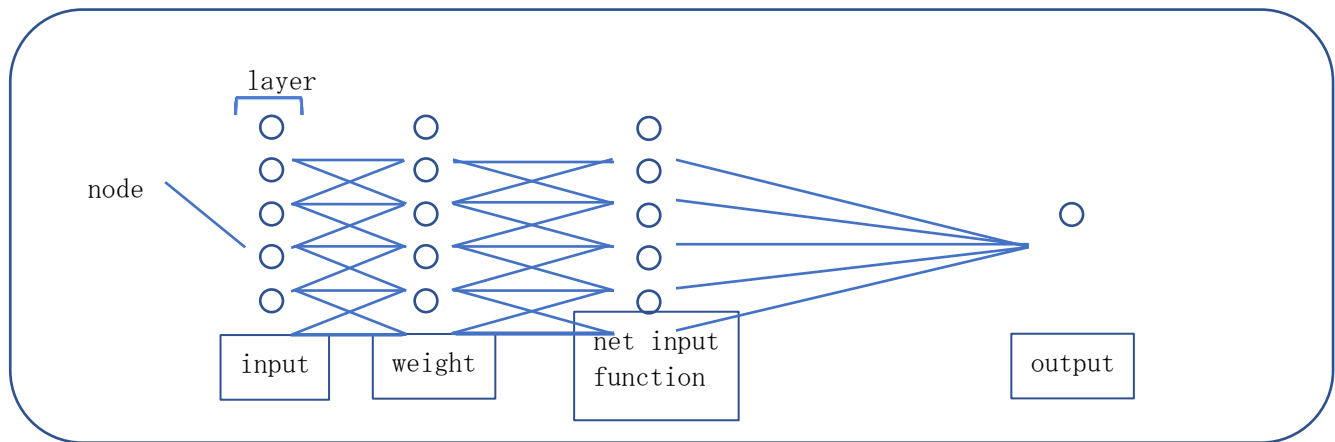Fig 2. Diagram showing the structure of deep learning.

§5 Discussion

§5-1 The System of Machine Translation

According to the book of Nitta, (2012), machine translation now works in the following way. First, it analyzes the structure of a sentence. It recognizes words, then finds out what role each word plays in this sentence and decide the structure of this sentence. This is called "syntactic analysis." Next, it transfers sentences into target language. There are two ways to do this. One is called direct transfer. It transfers the sentence directly from source language to target language, using a dictionary for these two languages.
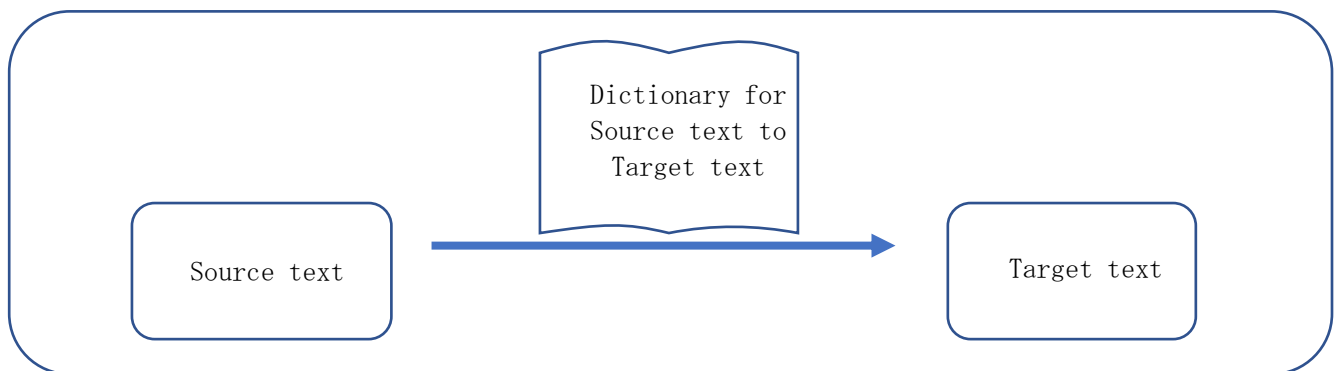

Fig 3. Diagram showing the structure of direct transfer.

The other method is interlingual transfer. It first transfers source text into a pivot language, then transfers to the target language. Pivot language is a language neither the source nor target language, not only natural languages but also artificial languages. Two dictionaries, source-pivot dictionary and pivot-target, are needed, but when the number of languages which have to be translated increases, interlingual transfer needs fewer dictionaries than direct transfer.
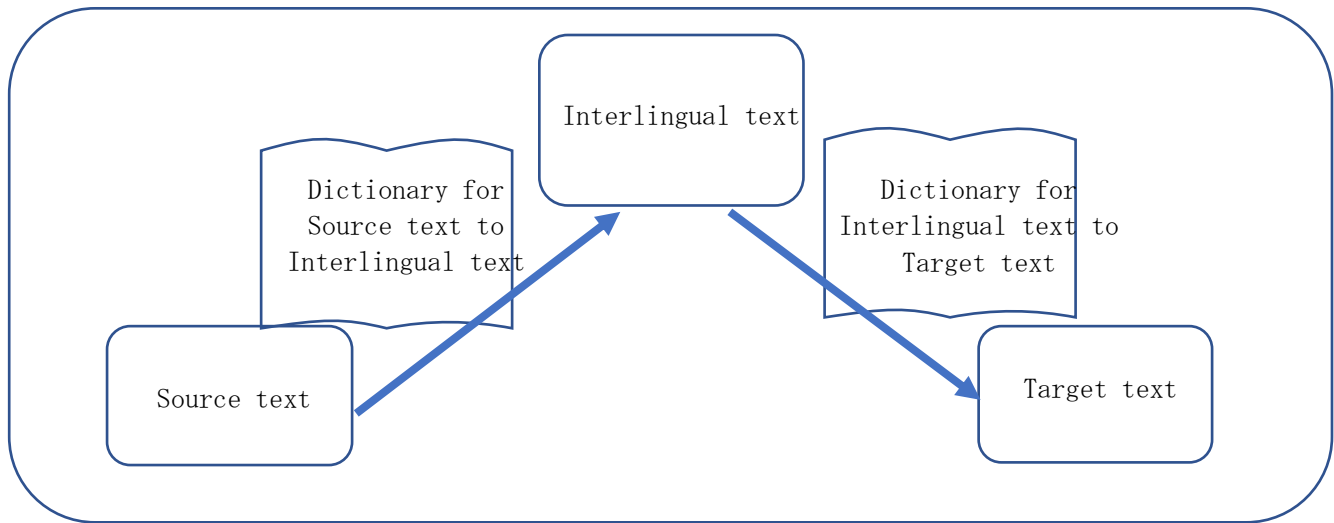
Fig 4. Diagram showing the structure of interingual transfer.

§5-2 Ambiguity of Syntax Analyzing
The problem of ambiguity of words is faced in translation. For example,
        I saw a dog running on the grass.
This sentence can be read in several ways. This is because "saw" is a word with many different meanings, and there are two candidates of the adverbial phrase, "on the grass."

  <The difference caused by "saw">
        I saw a dog running on the grass. (using my eyes)
        I saw a dog running on the grass. (with a saw)
        I saw a dog running on the grass. (saw ≒ meet)
  <The difference caused by "on the grass">
        I saw a dog which was running on the grass.
        I, on the grass, saw a dog running.

Usually almost all English-speaking people can understand this sentence in one way; I saw a dog which was running on the grass, with my eyes. However, computers cannot so easily understand because computers do not have common knowledge which most people usually have. We know what could happen and what rarely happens, from our daily experience, but it is troublesome to give computers the same knowledge. Therefore, a way to teach computers by giving them a large corpus of data was created. This is deep learning and helps computers learn for more effectively than ever before.

  §5-3 Is it possible to use pictures as an interlingual language and solve ambiguity of languages?
  While searching for how to improve computer translation, this idea seemed viable.

    §5-3-1 The Image Processing Technology
  In order to know how to automatically describe the content of an image, in paper of Vinyals and his colleagues (2015), they presented a generation model based on a deep recurrent architecture that combines recent advances in computer vision and machine translation. That can be used to generate natural sentences describing an image.

    §5-3-2 Problems with this method
  The phase of syntax analysis is the cause of the problem of ambiguity.
  It may be useful to generate an image from the fundamental meaning of a word. For example, the fundamental meaning of "see" is "face to an object and get its image." This makes it possible to get

rid of the need to decide only one meaning, but identical words with different meanings, for example, "saw (see)" and "saw (with a saw)" are still cause of concern.

Using image captioning and image recognition will be useful in the following two ways. First, pictures can be used as a kind of interlingual language. Second, it can help communicate with children who still cannot read letters, or with people who use a language which cannot be easily translated.

§6 Conclusion

Machine translation consists of syntax analysis and transferring words, and the improvement of these two processes improves the accuracy of translation. Syntax analysis and transferring words can be improved by adding more data. Also, deep learning helped a lot for improving machine translation.

It is difficult to make use of images as an interlingual language for improving machine translation because of the ambiguity caused by the fundamental system of translation. However, the idea of using image for communication is good for small children or users of a language which cannot yet be easily translated.

If voice recognition and screen review utilities are added, simultaneous translation can be provided, but still human interpreters are needed for two reasons. One is that there is a limit to the accuracy of machine translation. The other is that it is required to make clear where the responsibility lies in serious occasions.

§7 Addenda

新田義彦, 2012, 機械翻訳の原理と活用法 古典的機械翻訳再評価の試み, 明石書店
松尾豊, 2015, 人工知能は人間を超えるか ディープラーニングの先にあるもの, KADOKAWA/中経出版
清水亮, 2016, よくわかる人工知能 最先端の人だけが知っているディープラーニングのひみつ, KADOKAWA
天野真家, 石崎俊, 宇津呂武仁, 成田真澄, 福本淳一, 2007, IT Text 自然言語処理, 株式会社オーム
http://qiita.com/icoxfog417/items/5fd55fad152231d706c2, 2017/08/27
MachineTranslation.ppt, 2017/08/27
https://deepmind.com/research/alphago/
https://deeplearning4j.org/ja/neuralnet-overview
Vinyals, O., Toshev, A., Bengio, S., Erhan, D., 2015. Show and Tell: A Neural Image Caption Generator.
arXiv:1411.4555v2 [cs.CV] 20 Apr 2015