A Preliminary Study on the Present Situation and Future Development of Natural Language Processing

Liu Junlan
Xi’an International University, Xi’an, Shaanxi, 710077, China

Keywords: Natural Language; Processing; Present Situation Study; Future Development

Abstract: Natural language processing is an important research direction in the field of artificial intelligence and computer science. Natural language processing is a science that integrates mathematical science, language science and computer science. It contains the research and processing of words, sentences, paragraphs and chapters in the subject of language, so its research involves the language of our daily life, and there is also a very important relationship with linguistics. But there is also a certain difference, because it is not just a study of linguistics. This paper briefly expounds the technology and development history of natural language processing, analyzes the present situation of the development of natural language processing, and probes into the development trend of natural language processing in various fields, so as to promote the development of natural language processing.

1. Introduction

In recent years, the development of computer science is getting faster and faster, and at present, it has been developed by leaps and bounds. The application of computer technology in all aspects of people's daily life has a certain impact on our life and work. In the information age, the Internet is becoming more and more popular; it likes a huge information base that provides people with all kinds of convenience, and the information is generally expressed in natural language. When using these information bases, people usually use search engines to obtain information and translate the information by machine translation. In addition, smart phones with certain advanced technology have also been widely used in people's lives; smart phone has a great advantage that is smart voice, smart voice assistants can bring a lot of conveniences to our works, studies and lives. Therefore, natural language processing still has a great development prospect in the future.

2. Natural language processing technology

The language in our daily life is called natural language, such as Chinese, Japanese, Korean, English, German, and so on. Using computer technology to use, process and interpret natural language is called natural language processing technology, which is abbreviated as NLP, and which is a kind of discipline that combines linguistics and artificial intelligence\(^\text{[1]}\). Natural language processing is the main research direction in the field of artificial intelligence, its main purpose is to let the human beings can communicate with the computers in natural language. Human beings is the only specie with the ability of language, by which they can express their own ideas. therefore, in order to communicate with computers in natural language, natural language must be integrated into computer technology. Natural language contains many aspects, in general, it involves these aspects ---- generating, understanding and cognition. Understanding and cognition is to use computer technology to convert the input language into readable symbols, and then to carry out purposeful technical processing. Generating is to convert the vast amount of information contained in the computer’s database into natural language. In this way, we do not need to expense a lot of energy to study the difficult computer language when using the computer, and we can realize the direct connection between the computer language and the natural language, so it is of great significance to study the natural language processing technology. At present, the natural language processing technology is mainly about grammatical analysis, lexical analysis and chapter interpretation. From
the specific application point of view, natural language processing technology is the use of information retrieval, intelligent translation and filtering, language recognition, character recognition and so on, which has the significance of popularization and application prospect.

3. The development history of natural language processing

The concept of natural language processing was first put forward in the early 1950s. The earliest study of natural language processing technology was the study of artificial intelligence translation. In 1950, Alan Turing put forward the “Turing Test” in his paper Computing Machinery and Intelligence. It is used to test whether the computer has a function that is similar as artificial intelligence translation[2]. As soon as this research concept is put forward, it has been vigorously popularized, and machine translation and chess are taken as the main research objects. In 1997, the dark blue supercomputer developed by IBM beat a world champion of chess named Kasparov, but the level of machine translation was not as good as artificial translation, this showed natural language processing is of great significance. in 1954, George City University, in cooperation with IBM, successfully translated 60 sentences of Russian into English by using computer technology. Although there were still some shortcomings in the technology, the researchers at that time were full of passion, they also claimed that machine translation technology will be able to reach a certain stage of improvement in the next three to five years. However, in the actual developmental process, the actual situation did not exceed the expectation, the expected results have not been achieved after ten years, which results a significant decline in research funds. Before the 1980s, many natural language processing methods were limited to artificially formulated complex rules, so that the natural language processing technology were in the stagnation of development until the late 1980s. In late 1980s, machine translation technology had only made some progress, because in this stage of natural language processing, computer algorithm had been used[3]. At present, computer technology is becoming more and more advanced, and the cost of manufacturing computer is becoming lower and lower. Researchers pay more and more attention to natural language processing technology and computer learning algorithm. In addition, the researchers realized that the machine translated language must keep a high degree of homogeneity with the original text, so the analysis of implied meaning has become the main research object of the researchers. According to the relevant research data, by means of the statistics and analysis of massive language data, the computer language processing can play its powerful role better.

4. The present situation of natural language processing research

Although The research time of natural language processing is not very long, but it has made remarkable achievements and establish its significant position in some fields, which proves that natural language processing has certain application value and development prospect. According to the current research results, by virtue of the natural language processing, many problems have been solved, it can provide more convenience and help for people, such as electronic dictionary, text input, document translation and search engine, etc. However, with the more and more urgent social needs such as information technology, communication services, national security and so on, many aspects of the research need to be more perfect, such as fast and efficient translation and semantic calculation and expression, etc. At present, there are still many problems in the study of natural language processing, and the study lacks some practical theoretical bases. For example, the problem of meaning interpretation of word has not been solved, it has begun to study the deeper subject; the application of many technologies can not be completely consistent with the actual situation, and so on; these problems need to be deeply studied and solved by researchers.

There are still many problems in natural language processing, and the primary cause of these problems is the ambiguity in the dialogue of natural language. Ambiguity refers to the indistinct existing in the process of word analysis; in order to solve this problem better, a large number of knowledge words are needed. In artificial intelligence translation, a certain knowledge vocabulary reservation and strong knowledge background are necessary[4]. For example, in English, “Would
you like some tea?” , the general answer “Oh thanks” means declining with thanks. So, for the
intelligent machines, in the process of translation, the machines should have a database with
tremendous knowledge. Human expressions and activities are complex, so natural language is full
of ambiguity; in addition, because the grammar and lexical rules in language are limited, it is
possible that the same sentences express a variety of meanings. Taking Chinese as an example,
paragraphs are composed of sentences, sentences are composed of words, the meaning of a sentence
in different context is also different; what is more, it is possible for different forms of sentences to
express the same meaning; so these are the charm of the language. But at the same time, it also
brings some difficulties to language processing. In Chinese, the key point to eliminate ambiguity is
word partition, the smallest unit of a sentence is a word, word partition is an urgent problem to be
solved in natural language processing; in the process of expressing and writing a sentence, words
are coherent. English is different from Chinese, English has some natural words that can be the
partition, but Chinese does not has, so there are some obstacles in dealing with Chinese. In the
process of word partition, the computer must add a separator after each word, and sometimes the
meanings of some sentences are ambiguous, so it is difficult to add a separator. For example, the
Chinese sentence “Chongqing Yangtze River Bridge” can be understood not only as “the mayor of
Chongqing named Jiang Daqiao”, but also as “a bridge across the Changjiang River in Chongqing”.
So, in order to divide a sentence properly and then know the exact meaning of the sentence, it is
necessary to understand and analyze the sentence in combination with the actual context, which is
still a challenge for the computer.

5. The future development prospect of natural language processing

At present, as science and technology become more advanced, people gradually know that the
traditional words, sentence rules are too complex, and the empiricism of statistics can only obtain
limited information. With the database becomes more powerful, more and more language data have
laid the foundation for natural language processing. In addition, the statistical method of the data
has also been widely concerned and valued by researchers, and the method of automatic learning
and acquiring language knowledge by machine has also been widely used in natural language
processing. Moreover, in natural language processing, researchers also pay more and more attention
to the application of words. In the development of natural language processing, the construction of a
powerful knowledge database has become a hot issue[5]. At present, the research of natural language
processing is not only the study of characters, but also the research of syntax analysis, voice
recognition, machine translation, information retrieval and so on, and its application in practice is
becoming more and more wide. It has promoted the development of other disciplines and improved
the ability of computer language processing, so that it has become the focus point of people’s
research.

In the long run, natural language processing has amazing application value and development
prospect. As a research field with multiple disciplines at the same time, the development of natural
language processing can bring some help and influence to other related disciplines. With the
improvement of the database that contains semantic and grammatical knowledge and the analysis
technology, and with the emergence of new methods and the continuous updating of research
models, these can not only lay the foundation for theoretical research, but also consolidate the rapid
development of natural language processing, and let it occupy an important core position in the
whole scientific and technological research[6]. Finally, according to the analysis of the present
situation of natural language processing in this paper, we can see that in the future research, the
research methods will pay attention to the combination of theory and practice, to the combination of
technology and many fields, to the combination of module construction and tool environment, and
to the combination of knowledge meaning and application, so as to are realize the long-term
development of natural language processing.
6. Conclusion

Natural language processing is a new subject in the present age, and it is also a subject that develops relatively rapidly. In the process of development, natural language processing technology faces a series of challenges and setbacks, and also has some opportunities; by the research on various methods and models, it provides some convenience and support for the development of natural language processing. Although voice recognition and intelligent translation have not been developed as advanced as we expect, but natural language processing is more and more used in people's life widely, which proves that natural language processing technology plays a very important role in the whole science and technology field. In the near future, natural language processing technology will become more and more advanced, its application will be more widely, and it has a promising development prospect.

References


