

The Development and Significance of NLP From the Perspective of Linguistic Development and Social Influence

CHEN Meng

University of Shanghai for Science and Technology, Shanghai, China

As a subfield of artificial intelligence (AI), natural language processing (NLP), especially natural language generation (NLG), has developed rapidly in recent years, and natural language processing has played an important role in the development of linguistics and society. This paper aims to study the development process of NLP in the above two levels. I hope to provide a new perspective to look at linguistics, and even combine linguistics with the most advanced technology to radiate new vitality.

Keywords: natural language processing, linguistics, developmental influence

The Definition of NLP

NLP, namely natural language processing, is an important research and application object in the field of artificial intelligence and computer science. The main research goal and content of natural language processing is how to use natural language for real and effective communication between human and computer. Precisely, it includes:

(1) Syntactic and semantic analysis: for a given sentence, word segmentation, part of speech tagging, named entity recognition and linking, syntactic analysis, semantic role recognition and polysemy disambiguation.

(2) Information extraction: extract important information from a given text, such as time, place, person, event, reason, result, number, date, currency, proper noun, etc. Generally speaking, we need to know who, when, why, to whom, what we have done, and what the result is. It involves entity recognition, time extraction, causal relationship extraction, and other key technologies.

(3) Text mining (or text data mining): including text clustering, classification, information extraction, summary, emotion analysis and visualization, interactive expression interface of mining information and knowledge. At present, the mainstream technology is based on statistical machine learning.

(4) Machine translation: automatically translate the input source language text into another language text. According to different input media, it can be divided into text translation, voice translation, sign language translation, graphic translation, and so on. From the earliest rule-based method to the statistical method 20 years ago, and then to the neural network (encoding decoding) method today, machine translation has gradually formed a more rigorous methodology.

(5) Information retrieval: index large-scale documents. We can simply index the words in the document with different weights. We can also use the technology of 1, 2, 3 to build a deeper index. When querying, we

CHEN Meng, postgraduate, Institute of Foreign Languages, University of Shanghai for Science and Technology, Shanghai, China.

analyze the input query expression, such as a key word or a sentence, then find the matching candidate documents in the index, and then sort the candidate documents according to a sorting mechanism, and finally output the document with the highest ranking score.

(6) Question and answer system: For a question expressed in natural language, the question and answer system gives an accurate answer. It is necessary to do some semantic analysis on natural language query statements, including entity linking and relationship recognition, to form logical expressions, and then search for possible candidate answers in the knowledge base and find the best answer through a sorting mechanism.

In the fields of finance, law, medical health, natural language processing technology has been widely used. In the financial field, natural language processing can provide various analysis data for securities investment, such as hot spot mining, public opinion analysis, financial risk analysis, fraud identification, etc. In the field of law, natural language processing can help to search cases, predict judgments, automatically generate legal documents, translate legal texts, and answer intelligent questions and answers. In the field of medical health, natural language processing technology has a wide application prospect, such as the auxiliary entry of medical records, medical data retrieval and analysis, auxiliary diagnosis, and so on. Modern medical data are vast; new medical means and methods have developed rapidly. No doctor or expert can master the development of all the medical developments. Natural language processing can help doctors quickly and accurately find the latest research progress of various difficult diseases, and make the people who get sick enjoy the achievements of medical technology progress fastest.

The Development History of NLP

As we all know, language behavior is the essential characteristic of human beings to distinguish from other animals. Language carries the logical way of thinking of human beings and is an important channel of communication. From the perspective of the occurrence and purpose of language, language act can also be regarded as an act. At this level, strictly speaking, language is equivalent to the instruction used in the process of human communication, that is, natural language. With the development of science and technology, the purpose of real and effective communication between human and computer cannot be fully achieved and satisfied only by the application of computer language. Therefore, from the beginning of computer science and the use of natural language, it is urgent to meet the needs of human beings. NLP (natural language processing) is a process that integrates natural language used in human daily life into computer science to achieve free communication between human and machine. As an interdisciplinary subject, natural language processing mainly focuses on linguistics, involving computer science, mathematics, psychology, philosophy, logic, statistics, electronic engineering, biology, and other fields. The diversity of subjects involved in it makes the study of natural language processing complex and special, and it is doomed to be difficult to realize the process of communication between human and machine. From the perspective of theory and technology, it is still a long-term goal for the scientific community to develop an efficient and high-quality natural language processing system. But for some low demand applications, practical systems with basic natural language processing capabilities have been successfully developed, and some have even been commercialized, generalized, or even industrialized. Typical examples of using natural language processing system in human daily life are: various machine translation systems, mobile phone voice assistant, automatic summarization system, etc. The universal application of these products is the witness of the achievements of natural language processing in computer science.

Up to now, the development of natural language processing has gone through four important periods, namely, the germination period, the rapid development period, the low development period, and the recovery and integration period. Before 1956, it can be regarded as the initial stage of natural language processing. On the one hand, human civilization, after thousands of years of development, has accumulated a lot of raw materials in linguistics, mathematics, physics, and other areas, and consolidated the basic theoretical knowledge of natural language processing. On the other hand, the birth of the first generation of electronic computers not only provides a material basis for the application of natural language processing, but also leads to the social needs of machinetranslation, and promotes the basic research of natural language processing. In addition, the research on speech recognition system by Bell laboratory in 1952 and the birth of artificial intelligence in 1956 also provide the technical foreshadowing for natural language processing. The second stage is the rapid development period from 1957 to 1970. During this period, natural language processing has been rapidly integrated into the field of artificial intelligence with the rapid development trend. The main researchers are divided into two groups because of the different research methods: One is the sign school based on rule method; the other is the random school adopting probability method. Semiotics focused on the theory of formal language and the study of generative syntax, and in the late 1960s, they also studied formal logic system. However, the stochastic scholars have made great progress in this period by adopting the statistical research method based on Bayesian method. In the following years from 1971 to 1993, the research of natural language processing has reached a low speed development period because it is gradually found that the application of natural language processing is difficult to realize, and it is difficult to achieve significant research results. After the mid-1990s, the research of natural language processing has been revived and developed, stimulated by the rapid development of computer and the increasingly complete network technology. So far, many landmark research achievements have been achieved, such as neural network, attention mechanism, and pre training language model of NLP.

Influences

As a sub field of artificial intelligence (AI), natural language processing is a new high-tech technology. The significance of natural language processing lies in deep learning, rapid application, employment and feedback on linguistic development.

(1) A number of useful applications with great growth potential have been developed from the research field of natural language processing. Here are several: spelling check, keyword search, synonym search, information extraction from the website, analysis of spoken dialogue system of machinetranslation, and combination of a complex system. In fact, these applications have been widely used in the industry, from search (written and oral) to online ad matching.

(2) New models and algorithms with advanced functions and improved performance: It can flexibly represent the algorithm itself, more effectively end-to-end joint system learning, more effective transfer learning method between tasks, and better regularization and optimization methods. Most machine learning methods work well. Deep learning provides a very flexible, universal, and learning framework for presenting visual and linguistic information in the world.

(3) The development of artificial intelligence is the general trend. People engaged in the Internet have already produced a large amount of data, and will continue to continue in the future, including structured data, semi-structured and unstructured data. The integration of big data and cloud computing technology makes the

structured data technology mature and stable. However, semi-structured and unstructured data have great difficulties and challenges in the present and future applications due to its own complexity. At present, the market is very urgent for NLP technical talents, and this state will last for 5-10 years. Most enterprises need people who understand NLP technology to deal with large amounts of unstructured data. In this market gap, this technology will generate a lot of talents. This is not only for their own affirmation, but also for their future planning. This kind of talent is undoubtedly preemptive, well paid and has great development potential. In boss directly hired, the salary of NLP development engineer is only 30-60 k in the treatment demand for NLP technicians.

(4) It is generally believed that natural language processing is an interdisciplinary subject of artificial intelligence and linguistics. However, from the perspective of its development, it is still a research field dominated by computer scientists. Linguists often play an auxiliary role. Even in many stages of its development history, there is no direct involvement of linguists. However, it seems that linguistic theory can be seen in terms of its development, current situation, and development strategy. At present, the research results of computer learning and using natural language will provide new inspiration for the study of language acquisition, language learning and language use.

Conclusions

It can be seen that with the development of artificial intelligence in full swing today, natural language processing seems to have been unveiled by people. This paper only lists a few of them, but it is believed that there will be some theories to be found out. At that time, natural language processing will add a lot of color to our life!

In natural language understanding of sentences, we first interpret the meaning of the words themselves, while syntactic analysis is only used when it is necessary to confirm and remove ambiguity. But the feature of formal language expression is to use syntactic differences to understand and express various relationships, structures, and changes, and then carry out logical processing according to the structural relationships.

At present, there are still many problems to be solved in natural language processing and natural language generation. For example, they are not intelligent enough, unable to understand the implicatures, difficult to generate sentences in line with the specific context, and lack of word selection grammar. Lack of correlation, strong sense of violation. Many sentences cannot be completely close to natural language, human language. But the generated language has begun to integrate into every part of our daily life, so it is necessary to interpret it from the perspective of behavioral linguistics. I believe that in the near future, natural language processing will bring us real artificial intelligence.

References

- Chen, J. H., Su, M. C., Azzizi, V. T., Wang, T.-K., & Lin, W.-J. (2021). Smart project management: Interactive platform using natural language processing technology. *Applied Sciences*, 11(4), 1597.
- Kim, S., & Yun, J. H. (2020). Motion-aware interplay between WiGig and WiFi for wireless virtual reality. *Sensors*, 20(23), 6782.
- Levinson, S. C. (1987). Pragmatics and the grammar of anaphora: A partial pragmatic reduction of binding and control phenomena. *Journal of Linguistics*, 23(2), 379-434.
- Levinson, S. C. (2001). *Pragmatics*. Beijing: Foreign Language Teaching and Research Press.
- Pereira, D. A. (2020). A survey of sentiment analysis in the Portuguese language. *Artificial Intelligence Review*, 54(2), 1087-1115.
- Tusi, B. M. (1996). *English conversation*. Oxford: Oxford University Press.
- Wooffitt, R. (2006). *Conversation analysis and discourse analysis*. London: Sage Publications.