Big Data with natural language processing

YuanMing Huang
Indiana University
resident student
Bloomington, Indiana 47408
huang226@iu.edu

ABSTRACT

With the development of science and technology of today, natural language processing has been very widely used. And the application of big data is also more and more wide, while,in various areas,the role and impact of big data is also growing rapidly. In this study, it answered the question of "How big data affects the natural language processing."

KEYWORDS

Big Data, NLP, Artificial Intelligence, I423, hid230

1 INTRODUCTION

Natural language processing is an important direction in the field of computer science and artificial intelligence. It studies various theories and methods that enable effective communication between people and computers in natural language. Natural language processing is a fusion of linguistics, computer science, mathematics in one of the science.Just like "Natural language processing (NLP) is a form of artificial intelligence that helps machines "read" text by simulating the human ability to understand language."[1] .Big data refers to a collection of data that can not be captured, managed, and processed with regular software tools over a certain period of time. It is a need for a new processing model to have greater decision-making power, insight into discovery and process optimization.High growth rates and diversified information assets.for example "the phrase "big data" describes the growing volume of structured and unstructured,multi-source information that is too large for traditional applications to handle."[1] .In many areas, the application of big data play an important role for natural language processing,such as Search engine, voice assistant and machine translation."Natural language processing for big data can be leveraged to automatically find relevant information and/or summarize the content of documents in large volumes of information for collective insight."[1]. "Today around 80 percents of total data is available in the raw form. Big Data comes from information stored in big organizations as well as enterprises. Examples include information of employees, company purchase, sale records, business transactions, the previous record of organizations, social media, etc."[2] and "NLP can solve significant problems of the business world by using Big Data. Be it any business of retail, healthcare, business, financial institutions."[2] which indicates the importance of big data in natural language processing.

2 BIG DATA APPLIED IN MACHINE TRANSLATION

Google Translate is one of the most commonly used translation tools that can quickly translate web pages and short text snippets. Google is a very good example in the use of big data to improve the machine translation capabilities. Although machine translation has been around for a long time, the accuracy of machine translation has lagged far behind manual translation. Because there are many development problems in many machine translation software, for example, how to define the grammar and vocabulary of different languages, this is a very difficult problem in machine translation.But the application of big data helps and better improve accuracy of machine translation.There is a very famous engineer ,Franz Och, using a new method, using a purely statistical method, making Google in this area has a huge breakthrough.by translating a large number of available translations, the translation between English and French is much better than the old algorithm-driven translation method. The larger the available text database for parallel processing, the better the translation effect. This is one example of the application of large data in machine translation, as well as the role of large data in natural language processing.

3 BIG DATA APPLIED IN SPEECH RECOGNITION

In the field of speech recognition, there are similar development issues. Baidu is a very good example, Baidu technology company use large data to help improve the voice recognition technology, and improve the accuracy of voice recognition.

3.1 The relation between AI and NLP

Artificial intelligence is a branch of computer science that attempts to understand the essence of intelligence and produce a new intelligent machine that can react in a similar way to human intelligence. The field of research includes robotics, language recognition, image recognition, Natural language processing and expert systems.It can be divided into three types,like Artificial Narrow Intelligence,Artificial General Intelligence,Artificial Super Intelligence.[2]. artificial intelligence builds systems to do something intelligent, and also natural language processing builds systems to understand languages. "It is a subset of Artificial Intelligence".[2].

3.2 Stages of Artificial Intelligence

In the Artificial intelligence area,according to the "Overview of Artificial Intelligence"[2].There are three main stages will experience during the process.Stage 1 is Machine Learning, Machine learning is a set of algorithms used by intelligent systems to learn from experience.Stage 2 is Machine Intelligence,These are the advanced round of algorithms used by machines to learn from experience.such as Deep Neural Networks.Artificial Intelligence technology is currently at this stage.Stage 3 is Machine Consciousness,this process is self-learning from experience without the need of external data.

3.3 Using big data and machine learning to improve speech recognition

Data are very important in speech recognition area as well,"Data is the lifeblood of colleges and universities. Much like the circulatory system that carries our blood to all parts of our bodies, data flows through our institutions, in and out of the hands of the many individuals and departments that collect it and use it."[4]. A major breakthrough in speech recognition technology is also on the computing power of big data platforms. This calculation is determined by the platform capabilities and algorithms, not only requires a large cloud data processing capabilities, but also a new algorithm. In order to improve the ability of speech recognition, Baidu technology company researchers used a new algorithm to improve innovation.In the traditional language recognition technology, it must divide the modeling unit into multiple independent state, however, Baidu's researchers have used The overall model, the model of the unit can be used vowels, phonemes, syllables, and even can be mixed with each other. And then Baidu used their own the big data resource, and ultimately completed a major breakthrough. using big data to achieve the personalized voice recognition in a short time.Every time Users using the voice recognition system will further improve their account "model" to improve the accuracy of speech recognition. This is another example that how big data improved the natural language processing again.It is very similar to the sense of "Natural language processing employs computational techniques for the purpose of learning, understanding, and producing human language content."[3]

4 CONCLUSION

With the development of science and technology, big data has begun to use more and more widely in many different directions of artificial intelligence.like this study mentioned area, machine translation and voice recognition.All of them used big data to improves the accuracy of natural language processing. The relationship between Big data and natural language processing has become increasingly close.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregor von Laszewski for his support and suggestions to write this paper.

REFERENCES

[1] 2016. NLP for Big Data: What everyone should know? (2016). http://www.expertsystem.com/nlp-big-data-everyone-know/

[2] 2017. Overview of Artificial Intelligence and Role of Natural Language Processing in Big Data. (2017). https://www.xenonstack.com/blog/amp/overview-of-artificial-intelligence-and-role-of-natural-language-processing-in-big-data

[3] Christopher D. Manning Julia Hirschberg1. 2015. Advances in natural language processing. *Science, Volume 13: Issue 3* (2015).

[4] James Wiley. 2016. Do Your Know Where Your Data is Going? (2016). http://www.eduventures.com/2016/09/where-is-your-data-going/

5 BIBTEX ISSUES

Warning–no key, author in Scagliarini2017

Warning–to sort, need author or key in Scagliarini2017

Warning–no key, author in Jagreet2017

Warning–to sort, need author or key in Jagreet2017

Warning–no key, author in Jagreet2017

Warning–no key, author in Jagreet2017

Warning–no key, author in Scagliarini2017

Warning–no key, author in Scagliarini2017

Warning–empty author in Scagliarini2017

Warning–no key, author in Jagreet2017

Warning–empty author in Jagreet2017

Warning–no number and no volume in Hirschberg12015

Warning–page numbers missing in both pages and numpages fields in Hirschberg12015

(There were 14 warnings)

6 ISSUES