NATURAL LANGUAGE PROCESSING: AN OVERVIEW

Lyudmyla Guryeyeva

Ph.D., Associate Professor,

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

One of the main tasks of artificial intelligence is to make human-machine communication easier, in particular, to provide the ability to communicate in natural language. The development of natural language interfaces – applications that process a natural language text and either translate it into some formal language or perform some action according to the received message, help achieve this goal. A natural language interface is a rather complex program consisting of several components that perform sequential processing and test recognition based on linguistic methods and use one or more dictionaries (Park, 2010).

One of the most critical components of a natural language interface is a rule-based component designed to process production rules that define linguistic rules and a dictionary component that works with existing dictionaries.However, considering that natural languages have complicated structure, it creates a challenge for many natural language tasks withinmathematically precise algorithmic solutions. Therefore, statistical machine learning approaches for natural language processing (NLP) tasks problems are applied to solve this problem (Gudivada, Rao, Raghavan, 2015).

The fundamentalgoal of a linguistic processor (LP) is to transform a natural language (NL) text into a semantic networkrepresenting the text meaning.A linguistic processor is a complex of programs that provides analysis and synthesis of texts in natural language. The primary task of the linguistic processor is to parse and "understand" an input phrase in NL (in the case of analysis) or to construct an NL phrase corresponding to a formal description of its meaning (in the case of synthesis).

The language processor includes three key components (Turchin, 2021):

- linguistic (formal model of NL, dictionaries, grammars, linguistic tables, rules etc.);

- mathematical and algorithmic (translators of formal languages, algorithms for processing texts):

- software.

The perception of natural language information by a machine in a broad sense consists of recognising the meaning of the text, which is carried outby automatic dictionaries and formal grammars. In this context, a text can be defined as a redundant multi-level system for storing and transmitting information (Zarcone, 2020).

It should be noted that the linguistic processor itself has no means of interacting with the user; the software environment of the LP should provide such means. Therefore, storage of the initial and output data is simplified: the source texts are taken from ordinary text files. Finally, the analysis result (a fragment of the semantic network) is written to a binary file in the internal format of the semantic analyser.

A classical problem in NLP is text classification or text categorisation. It involves assigning labels or tags to textual units (i.e. sentences, queries, and documents and has awide range of applications (Minaee, 2021). As mentioned above, a text is an extremely deep source of information. However, extracting structured information applying mathematical solutions and algorithms can be challenging and time-consuming due to its multi-layered nature.

Classical machine learning required more sophisticated and high-speed tools to solve natural language processing and text classification tasks. However, considering the recent upgrade in computation power, the deep learning methods have made a breakthrough. As a result, deep learning-based models have excelled in various text classification tasks, including sentiment analysis, news categorisation, question answering, and natural language inference (Minaee, 2021).

One of the most criticaldevelopments in deep learning research in the last decade is the attention mechanism. It emerged as an upgrade of the encoder/decoder-based neural machine translation system in natural language processing (Hore P., 2019). As a result, the attention mechanism is now used in many applications, including text, image and speech processing etc. However, regarding the role the attention mechanism plays in solving complicated tasks in diverse application areas, including

psycholinguistics, machine translation and language model, this question requires further development.

References

Hore P., C. S. (2019, November 20). *A Comprehensive Guide to Attention Mechanism in Deep Learning for Everyone*. [Blog post]. Retrieved from https://www.analyticsvidhya.com/blog/2019/11/comprehensive-guide-attention-mechanism-deep-learning/

Minaee, S. K. (2021, January). Deep Learning Based Text Classification: A Comprehensive Review. *ACM Computing Surveys (CSUR), 54*(3), 1-40. https://doi.org/10.1145/3439726

Park, S. B. (2010, November). Natural language-based user interface for mobile devices with limited resources. *IEEE Transactions on Consumer Electronics, 56*(4), 2086-2092. Doi:10.1109/TCE.2010.5681076

Turchin, A. (2021). Using Natural Language Processing to Measure and Improve Quality of Diabetes Care: A Systematic Review. *Journal of Diabetes Science and Technology, 15(3),* 553-560. Doi: https://doi.org/10.1177/19322968211000831

Gudivada, V. N., Rao, D., Raghavan, V. (2015). Big Data Driven Natural Language Processing Research and Applications. *Handbook of Statistics, 33,* 203-238. https://doi.org/10.1016/B978-0-444-63492-4.00009-5

Zarcone, R. E. (2020). Analog Coding in Emerging Memory Systems. *Scientific Reports, 10*. https://doi.org/10.1038/s41598-020-63723-z