International Journal of Research in Engineering, Science and Management
Volume-1, Issue-9, September-2018
www.ijresm.com | ISSN (Online): 2581-5782

Natural Language Processing

Sneha Utekar[1], Tejali Shinde[2]
[1,2]*Student, Department of Computer Engineering, MGMCET, Navi Mumbai, India*

***Abstract*—With the increasing usage of the Web and other text application areas, the demands in both text mining and natural language processing (NLP) have been increasing. Researchers in text mining have hoped that NLP—the attempt to extract a full meaning representation from free text—can provide useful improvements to text mining applications of all kinds. The primary goal of Natural language processing (NLP) is to implement within computers the skill to understand a normal human language or natural language. It is related to the field of computer-human interaction. One of the motivations of NLP is for the society whose access to web information is obstructed simply by inability to use the key-board and operating system. Natural language comes under the domain of artificial intelligence with the goal of understanding and creating meaningful expressions in human language. Artificial intelligence is the capability of a machine to imitate intelligent human behavior. Therefore NLP uses Artificial Intelligence and issued to recover information in data mining.**

***Index Terms*—Natural Language Processing, Artificial Intelligence, Data mining, Text mining, Human Language**

I. Introduction

The concept of natural language processing is to develop and computer systems that can analyse, understand and synthesis natural human languages. Natural language falls within the domain of artificial intelligence with the goal of understanding and creating meaningful expressions in the human language. Natural language processing came into existence to ease the user's work and to satisfy the wish to communicate with the computerlanguage. Since all the users may not be well-versed in machine specific language, NLP caters those users who do not have enough time to learn new languages or get perfection in it. A language can be defined as a set of rules or set of symbol. Symbol are combined and used for conveying information or broadcasting the information. The input and output of an NLP system can be –

- Speech
- Written Text

Natural Language Processing basically can be classified into two parts i.e. Natural Language Understanding and Natural Language Generation which evolves the task to understand and generate the text.

- NLP Terminology:

Phonology: It is study of organizing sound systematically.
Morphology: It is a study of construction of words from primitive meaningful units.
Morpheme: It is primitive unit of meaning in a language.
Syntax: It refers to arranging words to make a sentence. It also involves determining the structural role of words in the sentence and in phrases.
Semantics: It is concerned with the meaning of words and how to combine words into meaningful phrases and sentences.
Pragmatics: It deals with using and understanding sentences in different situations and how the interpretation of the sentence is affected.
Discourse: It deals with how the immediately preceding sentence can affect the interpretation of the next sentence.
World Knowledge: It includes the general knowledge about the world.

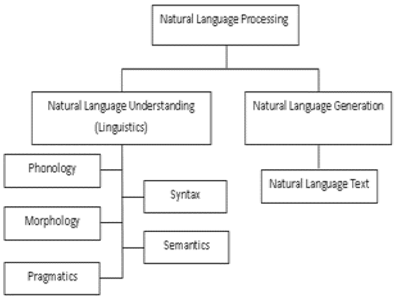

Fig. 1. Natural language processing

II. Steps in NPL

There are general five steps:

- *Lexical Analysis:* It involves identifying and analyzing the structure of words. Lexicon of a language means the collection of words and phrases in a language. Lexical analysis is dividing the whole chunk of text into paragraphs, sentences, and words.
- *Syntactic Analysis (Parsing):* It involves analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words.

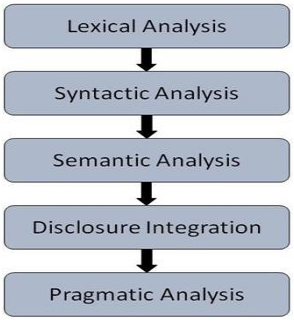

Fig. 2. Steps in NPL

- *Semantic Analysis:* It draws the exact meaning or the dictionary meaning from the text. The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain. The semantic analyzer disregards sentence such as "hot ice-cream".
- *Discourse Integration:* The meaning of any sentence depends upon the meaning of the sentence just before it. In addition, it also brings about the meaning of immediately succeeding sentence.
- *Pragmatic Analysis:* During this, what was said is re-interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.

III. Procedure

A typical human-computer interaction based on NLP might go as follows:

1. The human says something to the computer
2. The computer captures the audio
3. The captured audio is converted to text
4. The text's data is processed
5. The processed data is converted to audio
6. The computer plays an audio file in response to the human.

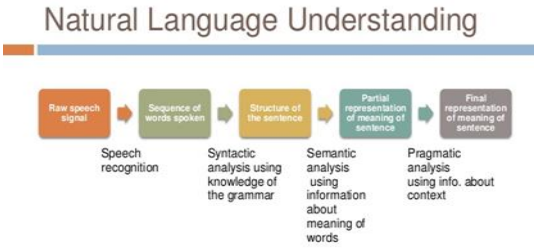

Fig. 3. Natural language understanding

IV. Approaches used in NLP

Four categories of approaches to semantic analysis in NLP:

A. Distribution

It employs large-scale statistical tactics of Machine Learning and Deep Learning. Distributional approaches include the large-scale statistical tactics of machine learning and deep learning. These methods typically turn content into word vectors for mathematical analysis and perform quite well attacks such as part-of-speech tagging (is this a noun or a verb?), dependency parsing (does this part of a sentence modify another part?), and semantic relatedness (are these different words used in similar ways?). These NLP tasks don't rely on understanding the meaning of words, but rather on the relationship between words themselves such systems are broad, flexible, and scalable .They can be applied widely to different types of text without the need for hand-engineered Features or expert-encoded domain knowledge. The downside is that they lack a true understanding of real-world semantics and pragmatics. Comparing words to other words or words to sentences, or sentences to sentences can all result in different outcomes.

B. Frame Based

"Aframeisa data-structure for representing a stereotyped situation," explains Marvin Minskyin hisseminal1974paper called "A Framework for Representing Knowledge." Think off ramesasa canonical representation for which specific scan be interchanged. Liang provides the example of commercial transaction as a frame. In such situations, you typically have a seller, a buyers, goods being exchanged, and an exchange price. Sentences that are syntactically different but semantically identical–such as "Cynthia sold Bob the bike for $200" and "Bob bought the bike for $200 from Cynthia" –can be fit into the same frame. Parsing then entails first identifying the frame being used, then populating the specific frame pa rameters–i.e. Cynthia, $200.The obvious down side of frames is that they requires upper vision. In some domains, an expert must create them, which limits the scope of frame-based approaches. Frames are also necessarily incomplete. Sentences such as "Cynthia visited the bike shop yesterday" and "Cynthia bought the cheapest bike" cannot be adequately analysed with the frame we defined above.

C. Model Theoretical Approach

The third category of semantic analysis falls under the model-theoretical approach. To understand this approach, we'll introduce two important linguistic concepts: "model theory" and "compositionality". Model theory refers to the idea that sentences refer to the world, as in the case with grounded language (i.e. the block is blue). In compositionality, meanings of the parts of a sentence can be combined to deduce the whole meaning. Liang compares this approach to turning a guage into computer programs. To determine the answer to the query "what is the largest city in Europe by population", you first have to identify the concepts of "city" and "Europe" and funnel down your search space to cities contained in Europe. Then you would need to sort the population numbers for each city you've shortlisted so far and return the maximum of this value.

D. Interactive Learning

Paul Grice, a British philosopher of language, described language as a cooperative gamebet we enspeaker and listener. Liang is inclined to agree. He believes that a viable approach to tackling both breadth and depth in language learning is to employ interactive, interactive environments where humans teach computer gradually .In such approaches, the pragmatic needs of language inform the development.

V. Application

- *Communication:* Many communication applications such Facebook Messenger are already using artificial intelligence. On the whole, Facebook looks very interested

in AI. Some months ago, Facebook announced its M service that promises to become your personal assistant "M can do anything a human can." When you request something that M can't do on its own, it sends a message to a Facebook worker and, as they work with the software, the AI begins to learn. Another interesting application of natural language processing is Skype Translator, which offers on-the-fly translation to interpret live speech in real time across a number of languages. Skype Translator uses AI to help facilitate conversation among people who speak different languages. This is great news! Without language barriers, people can communicate using the language they are comfortable with, which will in turn speed up a range of businesses processes.

- *Faster diagnosis:* Examples of natural language processing systems in artificial intelligence are also in hospitals that use natural language processing to indicate a specific diagnosis from a physician's unstructured notes. For example, NLP software for mammographic imaging and mammogram reports support the extraction and analysis of data for clinical decisions, as a study published in Cancer affirms. The software is able to determine breast cancer risk more efficiently, decrease the need for unnecessary biopsies and facilitate faster treatment through earlier diagnosis. According to the study, artificial intelligence reviewed 500 charts in but a few hours, saving over 500 physician hours. "Accurate review of this many charts would be practically impossible without AI," the author Stephen T. Wong said.
- *Customer Review:* Natural language processing in artificial intelligence applications makes it easy to gather product reviews from a website and understand what consumers are actually saying as well as their sentiment in reference to a specific product. Companies with a large volume of reviews can actually understand them and use the data collected to recommend new products or services based on customer preferences. This application helps companies discover relevant information for their business, improve customer satisfaction, suggest more relevant products or services and better and understand the customer's needs.
- *Virtual digital assistants:* Thanks to smart phone, virtual digital assistant (VDA) technologies (automated software applications or platforms that assist the human user by understanding natural language) are currently the most well-known type of artificial intelligence. Many companies are understanding the importance of the VDAs for their businesses and are investing significant resources to stay up to date. According to a study published by Research and Markets, VDA users will grow from 390 million in 2015 to 1.8 billion worldwide by 2021, while enterprise VDA users will rise from 155 million in 2015 to 843 million over the same period. VDAs are able to assist the consumers with transaction activities or optimize the call center operations to offer a better customer experience and reduce the operational costs. We will increasingly see these applications in other devices such as PCs programs, smart home systems, automobiles and in the enterprise market.

If these simple examples of natural language processing applications in artificial intelligence are any indication, the next Artificial Intelligence software and applications will improve our ability to transform unstructured data into valuable business insight and make smart automated decision-making part of our everyday lives.

VI. Future Scope

Supporting invisible UI: Human communication both conversation and text are part of most every interaction we h we have with machines. Amazon's Echo is just one example of the move toward a future that puts humans more directly in contact with technology. The concept of an invisible or zero user interface will rely on direct interaction between user and machine, whether through voice, text or a combination of the two. NLP that leverages a greater contextual understanding of human language, in other words, as it gets better at understating us what we say no matter how we say it, and what we are doing will be essential for any invisible or zero UI application.

Smarter search:

The future of NLP is also for smarter search—something we've been talking about here at Expert System for a long time. Applied to search, the same capabilities that allow a chatbot to understand a customer's request can enable "search like you talk" functionality (much like you could query Siri) rather than focusing on keywords or topics. Recently, Google announced that it has added NLP capabilities to Google Drive to allow users to search for documents and content using conversational language.

Intelligence from unstructured information:

An understanding of human language is especially powerful when applied to extract information and reveal meaning and sentiment in large amounts of text content (aka unstructured information), especially the types of content that must be manually examined by people. Analysis that accurately understands the subtleties of language—the choice of words, the tone used—can provide useful knowledge and insight of information, especially in the carefully worded language of annual reports, call transcripts and other investor-sensitive communications, as well as legal and compliance documents.

More effective and accurate understanding between humans and machines will only strengthen the efficiencies of both. No matter where it is applied, NLP will be essential in understanding the true voice of the user and the customer and facilitating more seamless interaction on any platform where language and human communication are used.

VII. Conclusion

Natural language processing is a branch of artificial intelligence

& computer science and it uses text mining to make the interaction between human and computer, though its purpose is to have interaction among natural language of human beings and computers. Current research in NLP is shown more interest in learning different algorithms which are based on unsupervised learning.

REFERENCES

[1] Nihar Ranjan, Kaushal Mundada, Kunal Phaltane, Saim Ahmad, "A Survey on Techniques in NLP". International Journal of Computer Applications, Volume134–No.8, January 2016.

[2] Ann Cope stake "Natural Language Processing". http://www.cl.cam.ac.uk/users/aac/. Copyright AnnCopestake, 2003–2004(7-6-17).

[3] Chomsky, Noam, 1965, Aspects of the Theory of Syntax,Cambridge, Massachusetts: MIT Press.

[4] Rospocher, M. ,van Erp, M. ,Vossen ,P., Fokkens ,A. ,Aldabe ,I. ,Rigau ,G. , Soroa ,A., Ploeger ,T. ,and Bogaard, T. (2016) .Building event-centric knowledge graphs from news. Web Semantics: Science, Services and Agents on the World Wide Web, In Press.

[5] Shemtov,H.(1997).Ambiguitymanagementinnaturallanguagegeneration. Stanford University.

[6] Emele, M. C., & Dorna, M.(1998,August) .Ambiguity preserving machine translation using packed representations. In Proceedings of the 36th Annual Meeting of the Association.