

Overview of Natural Language Processing

Deeksha Dwivedi¹, Mahendra Singh Sagar²

¹Student, Teerthanker Mahaveer University, Bagarpur Moradabad U.P India

²Associate Professor, Teerthanker Mahaveer University, Bagarpur Moradabad U.P India

diksha.dwivedi4@gmail.com

mahendra.singh12jan@gmail.com

Abstract— Natural Language Processing is a field of Computer Science, Artificial Intelligence and Computer Linguistics concerned with the interactions between human who have natural language provided by the nature and computational devices which are manmade and in, particular concerned with the programming computers to process the large natural languages corpora. Formerly many languages processing task typically involve the direct hand coding of the roles which is not in general robust to natural language variation. Challenges in natural language processing involve natural language understanding, natural language generation, connecting language and machine perception, managing human computer dialog system etc.

Keywords— Corpora, human computer dialog system, machine perception, computer linguistic.

I. INTRODUCTION

Natural Language Processing simply combination of three words which are Natural which means not artificial or machine made, Language means the method of human communication, either written or spoken and Processing means to generate.

And hence Natural Languages are those which are either spoken or written by humans.

The topic “NATURAL LANGUAGE PROCESSING” girdles everything a computer needs to be understand not only understand but to generate also.

Natural Language Processing is a subfield of the Artificial Intelligence and Computer Linguistics which is develop to make the computer to understand the human or natural languages.

Natural Language Processing is came in existence when we want to communicate with computer (machines) because we can't force humans to learn all the machine specific code or the language. A language is a system of symbols and rules.

Symbols are combined and used for conveying information and Rules are made to handle them.

II. STRUCTURE OF NATURAL LANGUAGE PROCESSING

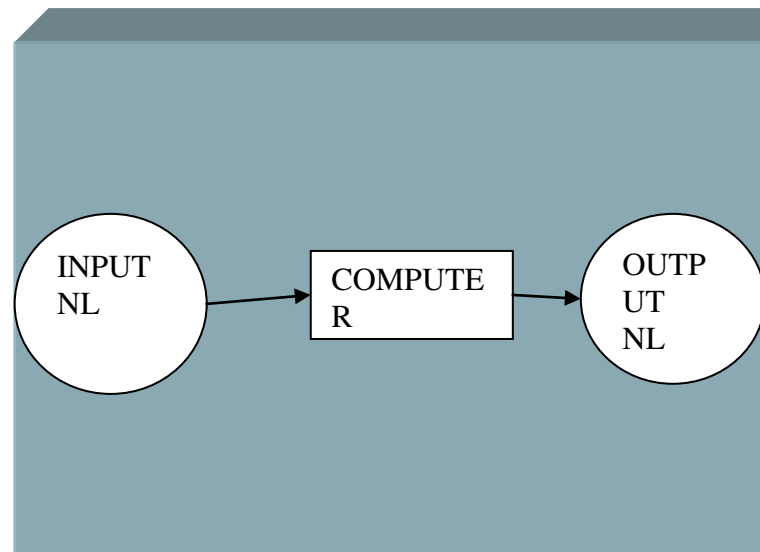


Fig 1. Structure of NLP

There are two stages:-

A. Natural Language Understanding

Whatever input is provided by user as Natural Language, the machine or the computer need to understand that and this stage is known as Natural Language Understanding.

It will analyse the different aspects of the language and mapping of the input language into useful representations.

B. Natural Language Generating

Whatever input is provided by the user on that basis machine or computer will generate some output that should also be in natural language so that user can understand the output and this is

known as Natural Language Generating. it involves Text Planning, Sentence Planning, Text realisation. NLU is harder than NLG.

III. DIFFICULTIES IN NATURAL LANGUAGE UNDERSTANDING

Natural Language has very rich forms and structure, which is very ambiguous.

There can be any type of ambiguity:-

- Lexical Ambiguity
- Syntax level Ambiguity
- Referential Ambiguity

IV. HISTORY OF NATURAL LANGUAGE PROCESSING

In 1948s, first NLP application was proposed which is dictionary looked up system and it was developed at Birkbeck College, London.

In 1950s, Alan Turing published an article titled "Machine and Intelligence" which is now called Turing Test as a subfield in intelligence.

In 1954s, Georgetown Experiment which involved fully automatic translation of more than sixty Russian sentences into English.

In 1960s, some notably successful NLP systems were developed in which there was SHRDLU, a natural language system working in restricted block words with restricted vocabularies and ELIZA.

In 1970s, many programmers begin to write "conceptual ontologies", which structured real-world scenario into computer understandable form.

Up to the 1980s, many NLP systems were made which was typically based on many complex handmade rules this is because of the steady increase of the computational powers.

Many of the notable early successes occurred in the field of machine translation, due to the work of IBM research where many statistical models were developed.

Recent research has increasingly focused on supervised or unsupervised learning algorithms and these algorithms involve the data that has not been hand-annotated with the desired answers.

In recent years, there has been many of result showing deep learning techniques in many natural

language processing task, for example in language modelling, parsing etc.

V. TERMINOLOGY IN NATURAL LANGUAGE PROCESSING

A. Phonology

Study of Organisation of sound systematically.

B. Morphology

Study of construction of words from primitive units or morphemes.

C. Morpheme

It is primitive unit of meaning in a language.

D. Syntax

It refers to arranging words to make a sentence, also involves the determining the structural role of words.

E. Semantics

Study for how to combine the word to form a meaningful sentence.

F. World knowledge

Includes the general knowledge about the world.

VI. LINGUISTICS AND LANGUAGE PROCESSING

Linguistics is the Science of Languages. It includes the study of following:-

- Sounds which refers to phonology
- Wordformation refers to morphology
- Sentence structure refers to the syntax
- Meaning refers to semantics
- Understanding refers to pragmatics

VII. LEVELS OF LINGUISTIC ANALYSIS

- Higher level corresponds to SPEECH RECOGNISATION
- Lower level corresponds to NATURAL LANGUAGE PROCESSING

VIII. STEPS FOR NATURAL LANGUAGE PROCESSING

A. Morphological and Lexical Analysis

It involves the analysing and identifying the structure of words which are being used to make a sentence or phrases. Lexicon of language means the

collection of smallest unit that is word which are semantically to make a sentence.

In this step whole chunk of text into paragraphs, words, and sentence.

B. Syntactical Analysis

It is also known as parsing.

It involves the analysis of the words in the sentence for grammar and arranging words to make the relationship among them.

C. Semantic Analysis

It will extract the meaning from the text. The text is checked for meaningfulness and it is done by mapping syntactical structures and objects in the task domain.

D. Discourse Integration

The meaning of any sentence depends upon the meaning of the sentence just before it which means it brings the actual meaning of succeeding sentence.

E. Pragmatic Analysis

In this phase, what was said is re-interpreted on what is actually meant. It involves the deriving those aspects of language which require real world knowledge.

IX. NATURAL LANGUAGE PROCESSING USING MACHINE LEARNING

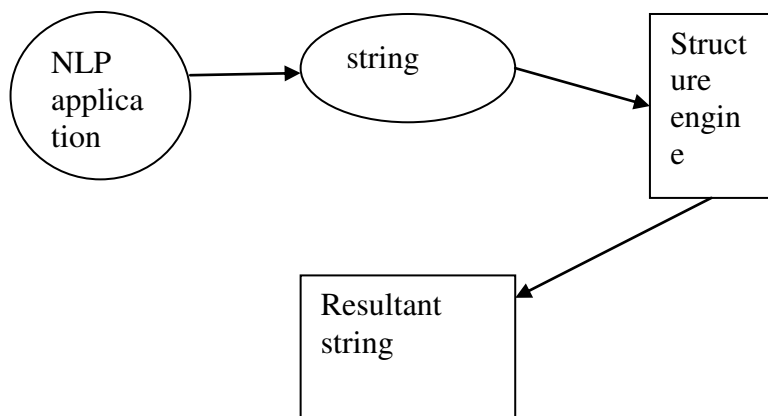


Fig2. NLP procedure

Many NLP algorithms are based on machine learning and mainly statistical machine learning and the working of the machine learning is different

from natural language processing. Prior implementations for natural language processing involve the direct hand coding of large set of rules.

Many algorithms of the machine learning are applied to NLP tasks in which algorithm take input a large set of characteristics that are generated from the input source.

Nowadays research has focused on Statistical models, which make probabilistic decision taking based on attaching real valued weight on each input.

X. MAJOR TASK IN NATURAL LANGUAGE PROCESSING

There is a list of the major task which are performed under natural language processing:-

A. Automatic Summarization

It produces automatically the summary of the set of text which is understandable.

B. Conference Resolution

It refers the sentence or a long set of sentence or text that determines which word refer to the same type of the objects.

C. Discourse Analysis

This is to identify the discourse structure of connected text and the nature of relationship between the sentences.

D. Machine Language

Automatically translates the text from one human language to another.

E. Named Entity Recognition

It describes the stream of text, determine the which item in the text relates to proper names and what the type of each name we are referring to is.

XI. STATISTICAL NATURAL LANGUAGE PROCESSING

This includes of taking some data generated in accordance with some unknown probability distribution and making inferences.

Statistical natural-language processing using some random, probabilistic and statistical methods to settle some of the difficulties especially the ones which arise because longer sentences are highly ambiguous when processed with realistic grammars

XII. APPLICATIONS OF NATURAL LANGUAGE PROCESSING

- Building a natural interface with database
- It has programs for classifying and retrieving the documents by content.
- It can be used for machine translation.
- It has advanced word processing tools.

XIII. WHAT MAKES NATURAL LANGUAGE PROCESSING A COMPUTATIONAL CHALLENGE

Ambiguous nature of Natural Language.

There are varied applications for language technology knowledge representation is a difficult task and there are different levels of encoding in our language.

XIV. FUTURE OF NATURAL LANGUAGE PROCESSING

Human level natural language processing is an AI-complete problem. It is equivalent to solving the central artificial intelligence problem and making computers as intelligent as people so that they can solve problem like human as well as perform activities that human can't perform and making it more efficient than humans. NLP future likely to be linked in Artificial Intelligence.

XV. ADVANTAGES OF NATURAL LANGUAGE PROCESSING

- It relieves burden of learning syntax.
- It requires no training
- Highly expressive
- Permits a variety of access point
- Highly flexible
- Represents many point of view
- High degree of exhaustion

XVI. DISADVANTAGES OF NATURAL LANGUAGE PROCESSING

- very difficult to make generic searches
- problem with synonyms
- problems with homograph
- problem with false drop
- ambiguous, fuzzy and soft
- not very compact
- not-standardized

XVII. CONCLUSION

The strength to use natural language query specification and retrieval bags over the keyword, keyphrase approaches. The believe that the restricted use of the natural language in captions for multimedia data abstraction and we have a system that can be judged and so built upon not only for abstracting images but also the form so multimedia data or input source as well.

XVIII. REFERENCES

- [1] Artificialneuralnetwork, http://en.wikipedia.org/wiki/Artificial_neural_network
- [2] Neuralnetwork, www.en.wikipedia.org/wiki/Neural_network.
- [3] Guoqiang Zhang, b. Eddy Patuwo, Michael Y.Hu*(july 1997), forecasting with the Artificial Neural Networks, kent State University, Kent, USA.
- [4] <http://www.slidehare.net/jhonerehmat/naturallanguageprocessing>.
- [5] Artificial intelligence by Elaine Rich and Kevin Knight, 1991, McGraw-Hill.
- [6] Artificial Intelligence-A Modern Approach by Russel and Norvig, 1995, prentice Hall.
- [7] Natural Language Processing an introductin www.ncbi.nlm.nih.gov v.18(5); sep-oct 2011.
- [8] NaturalLanguageProcessing, www.myreaders.info/html/artificial_intelligence.html.
- [9] NaturalLanguageProcessing-computerscience andengineering www.cse.unt.edu/~rada/CSC5290/Lectures/Intro.ppt.
- [10] NLP, www9.geogretown.edu/faculty/mda87/05/nlp/slides/Intro.ppt.