CMIS 4114 Index : 122139

Natural Language Processing

Content

- Introduction
- Components and Process
- Conclusion

INTRODUCTION

- What is Natural Language Processing
- NLP for machines
- Why NLP
- History of NLP

What is natural language processing?

- Process information contained in natural language text
- Also known as Computational Linguistics (CL), Human Language Technology (HLT), Natural Language Engineering (NLE)

NLP for machines…

- Analyze, understand and generate human languages just like humans do
- Applying computational techniques to language domain
- To explain linguistic theories, to use the theories to build systems that can be of social use
- Started off as a branch of Artificial Intelligence
- Borrows from Linguistics, Psycholinguistics, Cognitive Science & Statistics
- Make computers learn our language rather than we learn theirs

Why NLP?

- A hallmark of human intelligence
- Text is the largest repository of human knowledge and is growing quickly
- computer programmes that understood text or speech

History of NLP

- In 1950, Alan Turing published an article titled "Machine and Intelligence" which advertised what is now called the Turing test as a subfield of intelligence
- Some beneficial and successful Natural language systems were developed in the 1960s were SHRDLU, a natural language system working in restricted "blocks of words" with restricted vocabularies was written between 1964 to 1966

COMPONENTS AND PROCESS

- Components of NLP
- Linguistics and Language
- Steps of NLP
- Techniques and Methods

Components of NLP

- Natural Language Understanding
  - Taking some spoken/typed sentence and working out what it means
- Natural Language Generation
  - Taking some formal representation of what you want to say and working out a way to express it in a natural (human) language (e.g., English)

Components of NLP (cont.)

- Natural Language Understanding
  - Mapping the given input in the natural language into a useful representation
  - Different level of analysis required:
    - morphological analysis
    - syntactic analysis
    - semantic analysis
    - discourse analysis

Components of NLP (cont.)

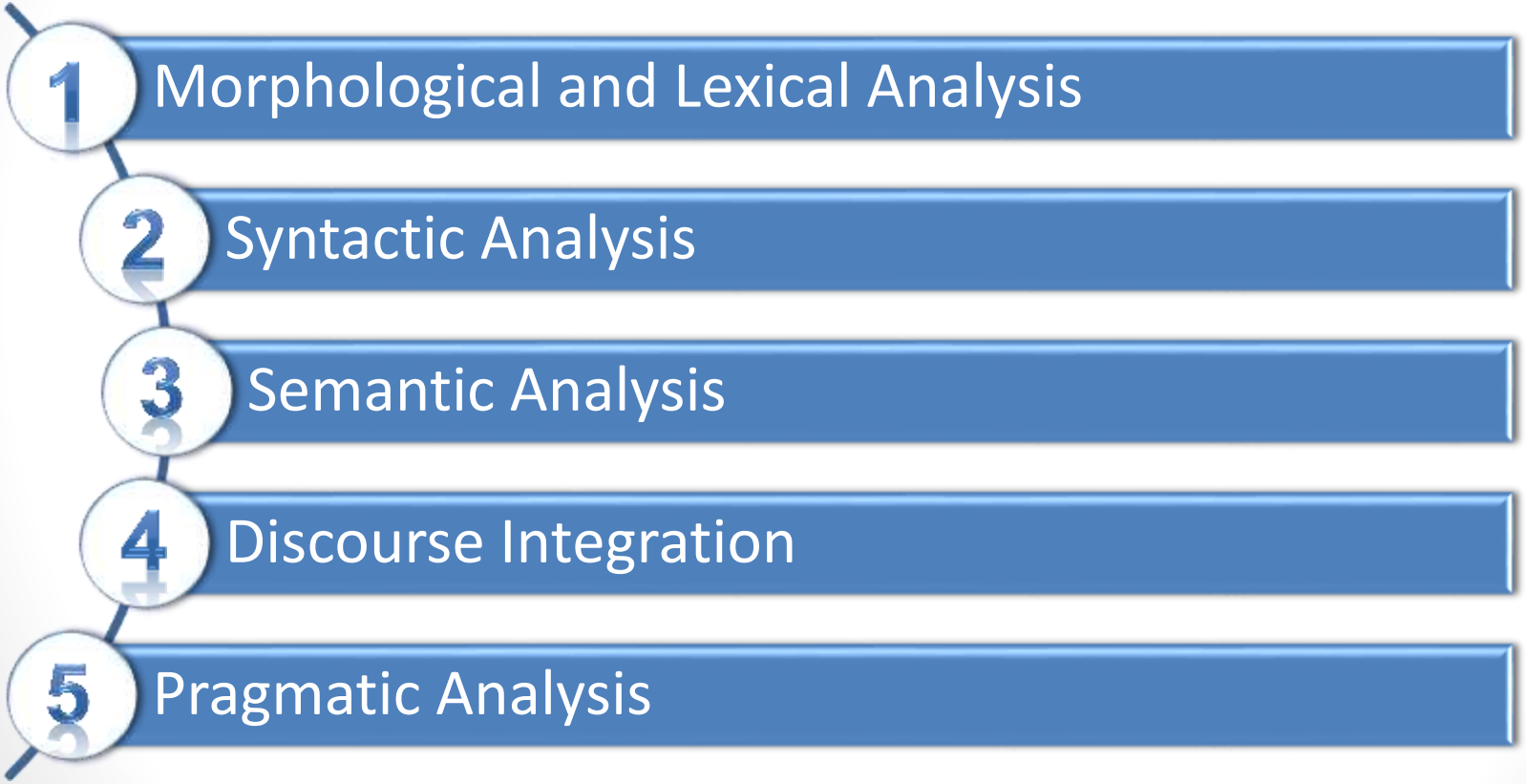
- Natural Language Generation
  - Producing output in the natural language from some internal representation
  - Different level of synthesis required:
    - deep planning (what to say)
    - syntactic generation

- NL Understanding is much harder than NL Generation. But, still both of them are hard

Linguistics and language

- Linguistics is the science of language
- Its study includes:
  - Sounds which refers to phonology
  - Word formation refers to morphology
  - Sentence structure refers to syntax
  - Meaning refers to semantics
  - Understanding refers to pragmatics

Steps of NLP


1
Morphological and Lexical Analysis
2
Syntactic Analysis
3
Semantic Analysis
4
Discourse Integration
5
Pragmatic Analysis

Morphological and Lexical Analysis

- The lexicon of a language is its vocabulary that includes its words and expressions
- Morphology depicts analyzing, identifying and description of structure of words
- Lexical analysis involves dividing a text into paragraphs, words and the sentences

Syntactic Analysis

- Syntax concerns the proper ordering of words and its affect on meaning
- This involves analysis of the words in a sentence to depict the grammatical structure of the sentence
- The words are transformed into structure that shows how the words are related to each other
- Eg. “the girl the go to the school”. This would definitely be rejected by the English syntactic analyzer

Semantic Analysis

- Semantics concerns the (literal) meaning of words, phrases, and sentences
- This abstracts the dictionary meaning or the exact meaning from context
- The structures which are created by the syntactic analyzer are assigned meaning
- E.g.. “colorless blue idea” .This would be rejected by the analyzer as colorless blue do not make any sense together

Discourse Integration

- Sense of the context
- The meaning of any single sentence depends upon the sentences that precedes it and also invokes the meaning of the sentences that follow it
- E.g. the word “it” in the sentence “she wanted it” depends upon the prior discourse context

Pragmatic Analysis

- Pragmatics concerns the overall communicative and social context and its effect on interpretation
- It means abstracting or deriving the purposeful use of the language in situations
- Importantly those aspects of language which require world knowledge
- The main focus is on what was said is reinterpreted on what it actually means
- E.g. “close the window?” should have been interpreted as a request rather than an order

Natural Language Generation

- NLG is the process of constructing natural language outputs from non-linguistic inputs
- NLG can be viewed as the reverse process of NL understanding
- A NLG system may have two main parts:
  - Discourse Planner

    what will be generated. which sentences
  - Surface Realizer

    realizes a sentence from its internal representation
  - Lexical Selection

    selecting the correct words describing the concepts

Techniques and methods

- Machine learning
  - The learning procedures used during machine learning
  - Automatically focuses on the most common cases
  - Whereas when we write rules by hand it is often not correct at all
  - Concerned on human errors

Techniques and methods

- Statistical inference
  - Automatic learning procedures can make use of statistical inference algorithms
  - Used to produce models that are robust (means strength) to unfamiliar input e.g. containing words or structures that have not been seen before
  - Making intelligent guesses

Techniques and methods

- Input database and Training data
  - Systems based on automatically learning the rules can be made more accurate simply by supplying more input data or source to it
  - However, systems based on hand- written rules can only be made more accurate by increasing the complexity of the rules, which is a much more difficult task

CONCLUSION

- NLP vs. Computer Language
- Future of NLP
- Summery

Natural language vs. Computer language

- Ambiguity is the primary difference between natural and computer languages
- Formal programming languages are designed to be unambiguous
  - They can be defined by a grammar that produces a unique parse for each sentence in the language
- Programming languages are also designed for efficient (deterministic) parsing
  - They are deterministic context-free languages (DCLFs)

Future of NLP

- Human level or human readable 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
- Make computers as they can solve problems like humans and think like humans as well as perform activities that humans cant perform and making it more efficient than humans

Cont….

- NLP's future is closely linked to the growth of Artificial intelligence
- As natural language understanding or readability improves, computers or machines or devices will be able to learn from the information online and apply what they learned in the real world
- Combined with natural language generation, computers will become more and more capable of receiving and giving useful and resourceful information or data

Summery

- The need for disambiguation makes language understanding difficult
- Levels of linguistic processing:
  - Syntax , Semantics, Pragmatics
- Statistical learning methods can be used to:
  - Automatically learn grammar
  - Compute the most likely interpretation based on a learned statistical model
  - Make intelligent guesses

End of Presentation

THANK YOU

Q & A