Natural Language Processing

Course Code : DLO8012

Dr. Rahul Khokale
Head of the Department
Department of Computer Engineering
St. John College of Engineering and Management, Palghar

What is NLP ?

- Natural language processing (NLP) is the ability of a computer program to understand human language as it is spoken.
- The process of computer analysis of input provided in a human language (natural language), and conversion of this input into a useful form of representation.

- NLP is a component of artificial intelligence (AI).
- Natural Language Processing (NLP) refers to AI method of communicating with an intelligent systems using a natural language such as English.
- Processing of Natural Language is required when you want an intelligent system like robot to perform as per your instructions, when you want to hear decision from a dialogue based clinical expert system, etc.

- The field of NLP is primarily concerned with getting computers to perform useful and interesting tasks with human languages.
- The field of NLP is secondarily concerned with helping us come to a better understanding of human language.

• Forms of Natural Language

- The input/output of a NLP system can be:
  - **written text**
  - **speech**
- We will mostly concerned with written text (not speech).
- To process written text, we need:
  - **lexical, syntactic, semantic knowledge about the language**
  - **discourse information, real world knowledge**
- To process spoken language, we need everything required to process written text, plus the challenges of speech recognition and speech synthesis.

• Components of NLP

- **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***, …
- **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.

• Why NL Understanding is hard?

- Natural language is extremely rich in form and structure, and **very ambiguous**.
  - How to represent meaning,
  - Which structures map to which meaning structures.
- One input can mean many different things. Ambiguity can be at different levels.
  - Lexical (word level) ambiguity -- different meanings of words
  - Syntactic ambiguity -- different ways to parse the sentence
  - Interpreting partial information -- how to interpret pronouns
  - Contextual information -- context of the sentence may affect the meaning of that sentence.
- Many input can mean the same thing.
- Interaction among components of the input is not clear.

• Knowledge of Language

- **Phonology –** concerns how words are related to the sounds that realize them.

- **Morphology –** concerns how words are constructed from more basic meaning units called morphemes. A morpheme is the primitive unit of meaning in a language.

- **Syntax –** concerns how can be put together to form correct sentences and determines what structural role each word plays in the sentence and what phrases are subparts of other phrases.

- **Semantics –** concerns what words mean and how these meaning combine in sentences to form sentence meaning. The study of context-independent meaning.

• Knowledge of Language (cont.)

- **Pragmatics –** concerns how sentences are used in different situations and how use affects the interpretation of the sentence.

- **Discourse –** concerns how the immediately preceding sentences affect the interpretation of the next sentence. For example, interpreting pronouns and interpreting the temporal aspects of the information.

- **World Knowledge –** includes general knowledge about the world. What each language user must know about the other's beliefs and goals.

• Ambiguity

- **I made her duck.**

- How many different interpretations does this sentence have?
- What are the reasons for the ambiguity?
- The categories of knowledge of language can be thought of as ambiguity resolving components.
- How can each ambiguous piece be resolved?
- Does speech input make the sentence even more ambiguous?
  - Yes – deciding word boundaries

• Ambiguity (cont.)

- Some interpretations of : **`I made her duck.`**
  1. I cooked *duck* for her.
  2. I cooked *duck* belonging to her.
  3. I created a toy duck which she owns.
  4. I caused her to quickly lower her head or body.
  5. I used magic and turned her into a *duck*.
- `duck` – morphologically and syntactically ambiguous:
- noun or verb.
- `her` – syntactically ambiguous: dative or possessive.
- `make` – semantically ambiguous: cook or create.
- `make` – syntactically ambiguous:
  - –. Transitive – takes a direct object. => 2
  - –. Di-transitive – takes two objects. => 5
  - –. Takes a direct object and a verb. => 4

• Ambiguity in a Turkish Sentence

- Some interpretations of: **Adamı gördüm.**
  1. I saw the man.
  2. I saw my island.
  3. I visited my island.
  4. I bribed the man.
- Morphological Ambiguity:
  - ada-m-ı ada+P1SG+ACC
  - adam-ı adam+ACC
- Semantic Ambiguity:
  - gör to see
  - gör to visit
  - gör to bribe

• Resolve Ambiguities

- We will introduce *models* and *algorithms* to resolve ambiguities at different levels.
- **part-of-speech tagging** -- Deciding whether `duck` is verb or noun.
- **word-sense disambiguation** -- Deciding whether `make` is `create` or `cook`.
- **lexical disambiguation** -- Resolution of part-of-speech and word-sense ambiguities are two important kinds of lexical disambiguation.
- **syntactic ambiguity** -- `her duck` is an example of syntactic ambiguity, and can be addressed by probabilistic parsing.

• Resolve Ambiguities (cont.)

- **I made her duck**
- S S
- NP VP NP VP
- I V NP NP I V NP
- made her duck made DET N
- her duck