Introduction to Natural Language Processing

Dereje T. Abzaw

Outline

Introduction to NLP

- Definitions
- Scope and Coverage of NLP,
- Application of NLP,
- Approach to NLP,
- Methods and resource,
- Levels of Language Processing,
- NLP challenges

What is NLP ?

- Is a subfield of **linguistics**, **computer science**, **information engineering**, and **artificial intelligence** concerned with the interactions between computers and human and/or natural languages.
- The ultimate goal of NLP is to read, decipher, understand, and make sense of the human languages.
- Extract meaningful information from natural language input and/or producing natural language output through computer program.
- Most NLP techniques rely on machine learning to derive meaning from human languages.

Contd …

How humans interact with machines using NLP - A simple application would be:

- A human talks to the machine
- The machine captures the text/audio
- Audio to text conversion takes place
- Processing of the text’s data
- Data to audio conversion takes place
- The machine responds to the human by playing the audio file

What can be the challenges of NLP?

- **Understanding Ambiguity and Context**
  - The Nature of Human Language:
    - Rules are complex and High-leveled. Consider Sarcasm for example
    - The letter “s” on end of a word , may signify plurality or simply part of the word itself.
  - Human language, in general , requires understanding both the words and how the word should be used in a sentence.
- **Bias in Training Data:**
  - NLP models are trained on massive datasets of text and speech. As in most ML models , If this data contains biases, the models can learn and perpetuate those bias.
- **Accounting for errors and non-standard language**
  - Slang
  - Colloquialisms
- **Limited Understanding of rare languages:**

Scope and Coverage

The scope and coverage of NLP are vast and continually expanding, with applications in various domains including:

1. Text Analysis and Classification
    - Sentiment Analysis
    - Topic detection
    - Categorization of texts
2. Language Translation
    - Google Translate
3. Speech Recognition
    - Siri & Alexa
4. Chatbots and Virtual Assistants
5. Information Extraction
    - Names
    - Dates
6. Text Generation
7. Spell Checking and Grammar Correction
8. Search and Recommendation Systems

Contd ...

Levels of Language Processing

- Phonetics and Phonology: The study of linguistic sounds.
- Morphology: The study of the meaningful components of words.
- Syntax: The study of the structural relationships between words.
- Semantics: The study of meaning.
- Pragmatics: The study of how language is used to accomplish goals.
- Discourse: The study of linguistic units larger than a single utterance.

Reading Assignment: *Explore in detail what the levels are by reading in depth about each study*

Contd …

- Acoustic, Lexicon and Language Models
  - Multi-modality
  - Speech and Character Recognition (what is said/written)
- Grammar, Lexical Meaning
  - Speech and text Analysis (what is meant)
- Discourse Context and Knowledge about Domain
  - Speech Understanding

Methods and Resource

- Natural Language technology come from several disciplines:
  - Computer Science
  - Computational and theoretical linguistics,
  - Mathematics (How?),
  - Electrical Engineering
  - Psychology.
- These discipline mainly concerned with the interactions between computers and human languages.

Contd …

- **Programming languages** and **algorithms** for generic data types from the point of Computer Science methods.
- **Algorithms**: Parsing, translation, for morphological and syntactic processing
- **Mathematical methods** : Statistical techniques have become especially successful in ASR (Automatic Speech Recognition) and IR (Information Retrieval)
- **Linguistic knowledge resources** : dictionaries, morphological and syntactic grammars, rules for semantic interpretation, pronunciation and intonation.
- **Corpora** and **corpus tools** for the acquisition and testing of statistical or rule-based language models.