

Introduction to NLP (CS7.401)

Spring 2022, IIIT Hyderabad
19 Apr, Tuesday (Lecture 20)

Taught by Prof. Manish Shrivastava

Tasks

Classification

The simplest ML problem we can think of is the classification task. Specifically in the context of NLP, it can occur at many levels:

- document classification (which tends to be multiclass and relevance-based)
- sentence classification
- group-of-word classification (idiom extraction, NP-extraction, chunking, etc.)
- word classification (sometimes formulated as a sequence labelling problem)

There are many metrics we can use to evaluate performance on classification tasks, most commonly precision, recall and F_1 . These are defined as

$$P = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

$$R = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

$$F_1 = \frac{2PR}{P + R}$$

Syntactic Annotation

Syntactic annotation can happen at many levels, the most obvious of which is parsing.

There are two main types of approaches to syntactic parsing – constituency-based and dependency-based.

Semantic Parsing

Semantic parsing is the semantic analogue of finding syntactic structure. A part of this problem is semantic or thematic role labelling – assigning roles (like *agent*, *theme*, *instrument*, etc.) to entities participating in the action of a sentence.

Information Retrieval

Many tasks involve extracting some relevant subset of information from text, even if they are not always thought of in this framework. This includes named entity recognition, sentiment analysis, question analysis, searching, entity linking (an extension of coreference resolution), and knowledge base population.

Information representation (in terms of the relations among entities) is also a major component of this. Relation identification is done in terms of entity-relation-entity *triples*, forming a semantic web.

Natural Language Inference/Recognising Textual Entailment

This task relies on drawing conclusions based on natural language understanding. This usually involves identifying a relation between a premise sentence and a hypothesis sentence – whether it supports, contradicts or is neutral towards it.