Sequence Labeling: POS Tagging

Natural Language Understanding Lab

Evgeny A. Stepanov,
Mahed Mousavi, Gabriel Roccabruna

SISL, DISI, UniTN & VUI, Inc.
evgeny.stepanov@unitn.it

Objectives

- Understanding:
  - Relation between Classification and Sequence Labeling
  - Relation between Ngram Modeling and Sequence Labeling
  - General setting for Sequence Labeling
  - Universal Part-of-Speech Tags
- Learning how to:
  - perform POS-tagging using NLTK
  - perform POS-tagging using spacy
  - train and test (evaluate) POS-tagger with NLTK

Outline

1. Sequence Labeling
2. Part-of-Speech Tagging
   - Universal Part of Speech Tags
   - Part-of-Speech Tagging with spaCy
   - Part-of-Speech Tagging with NLTK
   - Rule-based POS-Tagging with NLTK
     - *Exercise*: 20 min
   - Training POS-Tagger with NLTK
3. ***Lab Exercise***: Comparative Evaluation of Taggers: 40 min

Recommended Reading

- Dan Jurafsky and James H. Martin. Speech and Language Processing (3rd ed. draft)
  - Chapter 8: Part-of-Speech Tagging (HMMs)
- Steven Bird, Ewan Klein, and Edward Loper. Natural Language Processing with Python
  - Chapter 5: Part of Speech Tagging