Unsupervised Learning in Computational Linguistics

Course syllabus WS 2017–2018

Course Description

Unsupervised machine learning is a collection of methods for inferring (hidden) structure from 'unlabeled' data. Considering the labor-intensive and time-consuming nature of creating labeled data and the abundance of unlabeled data, it is clear that unsupervised methods are attractive in many fields, including in computational linguistics (CL) and natural language processing (NLP). Besides these practical motivations, unsupervised learning is also instrumental in investigating many problems of linguistics and cognitive sciences.

In this course we will study unsupervised methods for solving some of the typical NLP tasks such as tokenization, part-of-speech tagging, morphological analysis and parsing. We will also review some of the research-oriented applications of unsupervised methods in linguistics. For example, their use in modeling human language processing and acquisition, and investigating linguistic variation.

The course will take a practical approach. As well as reading and discussing some important and/or recent research, we will build practical models/applications during the course.

The course language is English.

Prerequisites

This is an advanced seminar for Master's and advanced Bachelor's students. Prior exposure to machine learning and fluency in programming are required.

Course work and evaluation

- Active participation during the class
- Leading discussion in one of the topics
- Project work

The normal coursework is worth 6ECTS, the students may extend their project and write a term paper for additional 3ECTS.

Practical information

Refer to the course web page for the course schedule, announcements and up-to-date information

Instructor	Çağrı Çöltekin ⟨ccoltekin@sfs.uni-tuebingen.de⟩
Course hours/location	Mon 10:00–12:00 & Wed 12:00–14:00 (room 1.13)
Course web page	http://sfs.uni-tuebingen.de/~ccoltekin/courses/ucl/
Office hours	Mon 12:00–14:00 (room 1.09)