PA164 Natural Language Learning

Lecture 01: The Course Overview

Vít Nováček

Faculty of Informatics, Masaryk University

Autumn, 2023

M U N I

Outline

The Goal of the Course

- Learn about various techniques. . .
- that can be used. . .
- to:
    - understand,
    - generate,
    - or "intelligently" process in any other way
- natural language text. . .
- by means of machine learning.
- Or, more succintly:
    - Learn how to apply machine learning (ML) methods. . .
    - to solve natural language processing (NLP) tasks.

Examples of "Classic" NLP Tasks

- Disambiguation
  - ▶ Morphological, synactical, word sense, . . .
- Parsing
  - ▶ Shallow, dependency, . . .
- Text classification
  - ▶ Spam filtering, document classification, sentiment analysis, named entity recognition, . . .
- High-level NLP tasks
  - ▶ Machine translation, text entailement, text understanding, text generation, knowledge extraction, . . .

Examples of ML Techniques

- Supervised ML
  - Learning to predict output class labels (classification) or numerical values (regression) that are associated with input objects (typically represented by so called feature vectors of predictor variables)
  - Typically trained and tested on two independent sets, where the correct output values are hidden in the test set
  - Some popular methods: naïve Bayes, support vector machines, decision/regression trees, (deep) neural networks
- Semi-supervised and unsupervised ML
  - Only some (semi-supervised) or no (unsupervised) output values known
  - Learning patterns (e.g., clusters or distribution) in the data
  - Some popular methods: unsupervised neural networks (Boltzman and Helmholtz machines, autoencoders), probabilistic methods (PCA, cluster analysis)
- Reinforcement learning
  - Agents learning how to take actions in an environment. . .
  - to maximise the cumulative reward function
  - Some popular methods: Monte Carlo, Q-learning, deep refinforcement learning

Outline

Lectures

- 13 lectures in total
  - 2 introductory ones (overview of the course and of relevant ML techniques)
  - 4 lectures illustrating the evolution of the current approaches to using ML in NLP
  - 2 lectures dedicated to poster sessions
  - 2 lectures on sample applications of ML to specific NLP tasks
  - 2 guest lectures (one from academia, one from industry; speakers and topics TBC)
  - 1 lecture dedicated to project presentations

Labs

- 6 labs in total, covering selected topics from the lectures
- Focused on hands-on assignments to be solved using off-the-shelf models and libraries (all in Python)
- If needed, some time during the labs may be dedicated also to your posters and projects

Posters

- Split into groups of up to 4 people
- In each group, pick a paper related to the field of natural language learning
  - The criteria for picking the paper are summarised in the interactive syllabus
  - Please consult your picks with the teacher before starting to work on the poster!
- Study the paper and present it to the class via a poster

Projects

- Try to reproduce the approach and the experiments from the paper you chose for the poster
- Identify possible areas where the approach could be improved
- Try to implement the improvements
- Summarise your results (even the negative ones!) and lessons learned in the form of a presentation to the class

Evaluation and Grading

- The maximum number of points you can gain is 100, split as follows
  - 20 for your poster
  - 50 for your project
  - 30 for your final oral exam
    - Obligatory (unless you go only for the credit)
    - You need at least 15 points here to pass
- Grading
  - A: 90 or more points
  - B: 80-89 points
  - C: 70-79 points
  - D: 60-69 points
  - E: 50-59 points
  - credit only: at least 45 points

Outline

The Lectures and the Labs

- Lectures
  - ▶ Weekly on Mondays at 2pm, starting on 18/09/2023, in A318
- Labs
  - ▶ Bi-weekly on Wednesdays at 6pm, starting on 27/09/2023, in B117
- Exam
  - ▶ Oral, based on "Questions and tasks" in the interactive syllabus
  - ▶ Organised during the standard exam term

The Posters and the Projects

- Both are homework assignments, to be worked on at your leisure
- You can work on them in the labs as well, though, where we can discuss whatever issues as needed
- Posters
  - You can start working on them as soon as you'll pick the topic (ideally by the third week of the semester)
  - Must be finished by your poster session (organised at one of the mid-term lectures)
- Projects
  - You should start working on them in parallel with the poster preparation
  - Must be finished by the project presentations (the last lecture)
- The exact requirements, expectations and deadlines will be fine-tuned and discussed as we go, based on the size, number, experience, skills and preferences of the particular student groups