

PA164 Natural Language Learning

Strojové učení a přirozený jazyk

Luboš Popelínský, Vít Nováček

Faculty of Informatics, Masaryk University

Autumn, 2021

MUNI

Outline

- 1 Welcome. Text mining and Natural language learning
- 2 Goal of this lecture
- 3 Organisation
- 4 Study materials

Goal of this lecture

- to understand a text
- by means of machine learning

better:

- use machine learning
- to solve NLP tasks

NLP tasks:

- disambiguation (morphological, synactical, word dis.)
- filtering and classification (spam fil., sentiment analysis, name entity recognition)
- highlevel NLP tasks (text understanding)
- maybe more

Organisation

- Lecture weekly
- Labs biweekly
- Poster presentation
- Project
- Final exam

Lecture

- Preprocessing. Distributional semantics, LSA, word embeddings (2 lessons)
- ML techniques for NLP (2 lessons) including RNN and others
- Ensemble learning, outencoders and outlier detection (2 lessons)
- Poster presentation (2 lessons)
- Learning language in logic. ILP. Keyness in text
- Text summarization
- Sentiment analysis
- Knowledge extraction from text. Knowledge integration

Poster presentation, labs and the project

Poster presentation

- Find a NLP task
- Present a poster in 30 minutes talk

Labs and the project

- The labs will have two parts
 - 1 A number of more or less open-ended exercise(s) in topics roughly following the lectures
 - 2 Some time dedicated to discussion of your projects
- The project can be picked from a selection of three possible assignments
- Projects may be elaborated in groups (more ambitious solution and fancier presentation will be expected then)
- Details on the assignments in the first labs next week

Study materials

- lecture materials in pdf, video
- readings, interesting projects (not for an exam)
- all at Interactive syllabus

Literature

the basic book

Charu C. Aggarwal, Machine Learning for Text. Springer 2018

more on deep learning for NLP:

Li Deng, Yang Liu (Eds.) Deep Learning in Natural Language Processing. Springer 2018

Additional materials

A Primer on Neural Network Models for Natural Language Processing
Yoav Goldberg. Arxiv 2015

CS224n: Natural Language Processing with Deep Learning Stanford
<http://web.stanford.edu/class/cs224n/>

Chollet, François *Deep learning v jazyku Python : knihovny Keras, Tensorflow* Grada Publishing, 2019

Charu C. Aggarwal, et al. (Eds.), *Mining Text Data*. Springer 2012