COURSE DESCRIPTION CARD - SYLLABUS

Course name

MACHINE LEARNING METHODS IN NATURAL LANGUAGE PROCESSING

Course

Proposed by Discipline

Information and communication technology

Year/Semester

II/4, III/6

Type of studies

Doctoral School

Course offered in

English

Form of study

full-time

Requirements

elective

Number of hours

Lecture

4

Tutorials

Projects/seminars

Number of credit points

1

Lecturers

Responsible for the course/lecturer:

dr hab. inż. Mikołaj Morzy, prof. PUT

email: mikolaj.morzy@put.poznan.pl

phone: +48 61 665 2961

Faculty of Computing and Telecommunications

Poznan University of Technology

ul. Piotrowo 2, 60-965 Poznan, Poland

Responsible for the course/lecturer:

Prerequisites

Knowledge: very basic understanding of algorithms, ability to read and understand very simple blocks of computer pseudo-code, understanding of basic statistics.

Skills: ability to transfer knowledge between domains, and to apply learned patterns to different domains.

Social competencies: thinking outside the box to solve various problems using available textual resources.

Course objective

The main aim of the course is to present the newest developments in the area of natural language processing (NLP) using algorithms and techniques of machine learning (ML). The majority of human knowledge is currently stored in the form of unstructured text. Abstracts, reviews, descriptions, posts, emails, tweets, all create a huge corpus of data which cannot be analyzed manually. Such textual corpora exist in almost all domains of science and technology. Computer methods for text analysis are collectively known as NLP. In the recent years we are witnessing a true revolution in NLP due to the development of machine learning methods designed specifically to tackle NLP challenges. During the lecture the students will learn basic NLP methods (tokenization, lemmatization, stemming), basic representation methods (one-hot encoding, TF-IDF), as well as methods based on neural networks (word and sentence vectors, transformer language models). We will discuss methods for sentiment analysis in text, named entity recognition, neural translation, sequence to sequence learning, and more.

Course-related learning outcomes

Knowledge

A PhD student who graduated from doctoral school knows and understands:
1) current achievements in the combined fields of machine learning and natural language processing, they understand basic principles of algorithms used to extract useful knowledge from unstructured text, [P8S_WG/SzD_W01]
2) the current developmental trends in machine learning and natural language processing, and can identify research questions in their scientific domains that can be addressed using machine learning and natural language processing. [P8S_WG/SzD_W02]

Skills

A PhD student who graduated from doctoral school can:
1) has the knowledge of machine learning and natural language processing to collect new data and new insights in their respective scientific disciplines, [P8S_UW/SzD_U01]
2) design new distributed representations of data in their scientific disciplines using the paradigm of encoder-decoder neural network architecture. [P8S_UW/SzD_U03]

Social competences

A PhD student who graduated from doctoral school is ready to:
1) acknowledge the importance of natural language processing methods by designing a research question involving their own discipline that can be addressed using machine learning and natural language processing. [P8S_KK/SzD_K03]

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

PQF code	Methods for verification of learning outcomes	Assessment criteria
W01, W02	Writing a short project description of how to use selected NLP methods creatively to answer scientific questions relevant to student's own discipline	Quality, size, and availability of text corpora proposed in the project description
U01, U03	Writing a short project description of how to use selected NLP methods creatively to answer scientific questions relevant to student's own discipline	Appropriateness of the proposed methods for defining semantic similarity in the area of student's discpline
K03	Writing a short project description of how to use selected NLP methods creatively to answer scientific questions relevant to student's own discipline	Relevance of the proposed NLP method to the selected scientific question in the area of student's discipline

Programme content

1. Natural Language Processing (Representation of text, Stemming, lemmatization, tokenization, Vector-space models of text).
2. Machine learning (Introduction to machine learning, Neural networks, Classification and clustering, Distance functions).
3. Distributed representations (Word vectors (word2vec, GloVe), Sentence vectors, Language models).
4. Advanced topics (Sentiment analysis, Named Entity Recognition, Neural Translation).

Teaching methods

Lecture: multimedia presentation including illustrations and examples.

Bibliography

Basic

1. Collobert, Ronan, et al. "Natural language processing (almost) from scratch." Journal of machine learning research 12. Aug (2011): 2493-2537.

Additional

1. Mikolov, Tomas, et al. "Distributed representations of words and phrases and their compositionality." Advances in neural information processing systems. 2013.

2. Le, Quoc, and Tomas Mikolov. "Distributed representations of sentences and documents." International conference on machine learning. 2014.

3. Bengio, Yoshua, et al. "A neural probabilistic language model." Journal of machine learning research 3.Feb (2003): 1137-1155.

Breakdown of average student's workload

	Hours	ECTS
Total workload	26	1.0
Classes requiring direct contact with the teacher	6	0.5
Student's own work (literature studies, preparation for tutorials, project preparation) [1]	20	0.5

[1] delete or add other activities as appropriate