CS918-15 Natural Language Processing

22/23

Department
Computer Science
Level
Taught Postgraduate Level
Module leader
Gabriele Pergola
Credit value
15
Module duration
10 weeks
Assessment
Multiple
Study location
University of Warwick main campus, Coventry

Description

Introductory description

Knowledge of the fundamental principles of natural language processing.

Module aims

The aim of the module is to equip students with a fundamental understanding of automated methods for processing linguistic data in textual form (natural language processing) from different sources (newswire, web, social media, academic publications) and associated challenges. The module will also provide students with the skills to analyse textual data and familiarise them with state of the art tools and applications.

Outline syllabus

This is an indicative module outline only to give an indication of the sort of topics that may be covered. Actual sessions held may differ.

The module will address core methodologies in natural language processing and related tools and will proceed to examine current applications. The syllabus may cover:

- Regular expressions, word tokenisation, stemming, sentence segmentation
- N-grams and language models
- Part-of-Speech Tagging
- Hidden Markov Models and Maximum Entropy Models
- Semantics: Lexical Semantics, Distributional Semantics, Word Sense Disambiguation and Vector Space Models
- Text classification
- Sentiment analysis
- Information Extraction: Named Entity Recognition, Relation Extraction
- Syntactic Parsing
- Semantic Parsing
- Question Answering and Summarisation
- Recommender systems

Learning outcomes

By the end of the module, students should be able to:

- Demonstrate knowledge of the fundamental principles of natural language processing.
- Understanding of methods and algorithms used to process different types of textual data as well as the challenges involved.
- Understanding of the state of the art in the core areas of Natural Language Processing such as Language Models, Part-of-Speech tagging, Named Entity Recognition, Syntactic Parsing, Information Extraction, Text Classification, Distributional Semantics and Vector Space Models.
- Working knowledge of state of the art tools available for analysing linguistic data in the context of the above mentioned areas.
- Computational skills to create NLP processing pipelines using existing NLP libraries, retrain models and extend existing NLP tools.

Indicative reading list

Please see Talis Aspire link for most up to date list.

View reading list on Talis Aspire

Research element

Students need to do some research about features used for sentiment classifier training in Assignment 2

Subject specific skills

- Have knowledge of the fundamental principles of Natural Language Processing (NLP).
- Understanding of methods and algorithms used to process different types of textual data as well as the challenges involved.
- Understanding of the state of the art in the core areas of Natural Language Processing such

as Language models, Part-Of-Speech tagging, Named Entity Recognition, Syntactic Parsing, Information Extraction, Text Classification, Distributional Semantics and Vector Space Models.
- Understanding of the state of the art in current application areas such as Semantic Parsing, Sentiment Analysis, Social Media analysis, Summarisation, Question Answering, Information Extraction.
- Working knowledge of state of the art tools available for analysing linguistic data in the context of the above mentioned areas.
- Computational skills to create NLP processing pipelines using existing NLP libraries, retrain models and extend existing NLP tools.

Transferable skills

- Analytical skills – Examine NLP problems thoroughly with attention to details
- Research skills – Identify relevant resources and background information to be used in coursework projects
- Problem solving skills – Think creatively and apply sensible approaches to solve the NLP problems given
- Communication skills – Present approaches and findings in a coherent manner in coursework reports

---

Study

Study time

Type	Required
Lectures	20 sessions of 1 hour (13%)
Seminars	8 sessions of 1 hour (5%)
Supervised practical classes	9 sessions of 1 hour (6%)
Private study	113 hours (75%)
Total	150 hours

Private study description

Background reading.
Coursework completion (including programming and report writing).
Revision.

Costs

No further costs have been identified for this module.

Assessment

You do not need to pass all assessment components to pass the module.

Students can register for this module without taking any assessment.

Assessment group D2

	Weighting	Study time
Assessed practical coursework	30%	
Assessed practical coursework. This assignment is worth more than 3 CATS and is not, therefore, eligible for self-certification.		
In-person Examination	70%	
CS918 exam		

- Answerbook Pink (12 page)
- Students may use a calculator

Assessment group R1

	Weighting	Study time
In-person Examination - Resit	100%	
CS918 resit exam		

- Answerbook Pink (12 page)
- Students may use a calculator

Feedback on assessment

Students will receive written feedback on coursework.

Past exam papers for CS918

Availability

Pre-requisites

Self-contained module but it would be helpful to take in conjunction with CS910 and/or CS909.

Courses

This module is Optional for:

- TCSA-G5PD Postgraduate Taught Computer Science
    - Year 1 of G5PD Computer Science
    - Year 1 of G5PD Computer Science
- Year 1 of TCSA-G5PA Postgraduate Taught Data Analytics

This module is Core option list A for:

- Year 1 of TPSS-C803 Postgraduate Taught Behavioural and Data Science

This module is Core option list C for:

- Year 1 of TPSS-C803 Postgraduate Taught Behavioural and Data Science

This module is Option list B for:

- Year 1 of TIMA-L981 Postgraduate Social Science Research