Statistical Natural Language Processing (COMP0087)

Description

Aims:
The module introduces the basics of statistical natural language processing (NLP) and machine learning techniques relevant for NLP.

Learning outcomes:
On successful completion of the module, a student will be able to understand relevant ML techniques, in particular in deep learning, what makes NLP challenging (and exciting), how to write programs that process language, and how to address the computational challenges involved.

Content:
NLP is domain-centred fields, as opposed to technique centred fields such as ML, and as such there is no "theory of NLP" which can be taught in a cumulative technique-centred way. Instead this course will focus on one or two NLP end-to-end pipelines (such as Machine Translation and Machine Reading). Through these applications the participants will learn about language itself, relevant linguistic concepts, and Machine Learning techniques. For the latter an emphasis will be on deep learning prediction. Topics will include (but are not restricted to) machine translation, sequence tagging, constituent and dependency parsing, information extraction, semantics. The course has a strong applied character, with coursework to be programmed and lectures that mix practical aspects with theory and background.

NLP Tasks:
- Language Models;
- Machine Translation;
- Text Classification;
- Sequence Tagging;
- Constituency Parsing;
- Dependency Parsing;
- Information Extraction;
- Machine Comprehension;

NLP and ML methods:

Key information

Year: 2019/20
Credit value: 15 (150 study hours)
Delivery: PGT L7, Campus-based
Reading List: View on UCL website
Tutor: Prof Sebastian Riedel
Term: Term 2
Timetable: View on UCL website

Assessment

Coursework: 30%
Coursework: 40%
Coursework: 30%

Find out more

For more information about the department, programmes, relevant open days and to browse other modules, visit ucl.ac.uk
-Encoder/Decoder Architectures;
-Feature Engineering;
-Deep Neural Networks;
-RNNs, CNNs;
-Attention;
-Word Vectors;
-Pretraining;

Requisites:
In order to be eligible to select this module, a student must be registered on a programme for which it is a formally-approved option or elective choice AND must have (i) an understanding of Basic Probability Theory (e.g. Bayes Rule), Linear Algebra and Multivariate Calculus; (ii) proficiency in coding in Python; and (iii) the ability to install libraries on a computer.
Statistical Natural Language Processing (COMP0087)

**Description**

**Aims:**
The module introduces the basics of statistical natural language processing (NLP) and machine learning techniques relevant for NLP.

**Learning outcomes:**
On successful completion of the module, a student will be able to understand relevant ML techniques, in particular in deep learning, what makes NLP challenging (and exciting), how to write programs that process language, and how to address the computational challenges involved.

**Content:**
NLP is domain-centred fields, as opposed to technique centred fields such as ML, and as such there is no "theory of NLP" which can be taught in a cumulative technique-centred way. Instead this course will focus on one or two NLP end-to-end pipelines (such as Machine Translation and Machine Reading). Through these applications the participants will learn about language itself, relevant linguistic concepts, and Machine Learning techniques. For the latter an emphasis will be on deep learning prediction. Topics will include (but are not restricted to) machine translation, sequence tagging, constituent and dependency parsing, information extraction, semantics. The course has a strong applied character, with coursework to be programmed and lectures that mix practical aspects with theory and background.

NLP Tasks:
- Language Models;
- Machine Translation;
- Text Classification;
- Sequence Tagging;
- Constituency Parsing;
- Dependency Parsing;
- Information Extraction;
- Machine Comprehension;

NLP and ML methods:

**Assessment**

<table>
<thead>
<tr>
<th>Coursework</th>
<th>30%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coursework</td>
<td>40%</td>
</tr>
<tr>
<td>Coursework</td>
<td>30%</td>
</tr>
</tbody>
</table>

**Key information**

- **Year:** 2019/20
- **Credit value:** 15 (150 study hours)
- **Delivery:** UGM L7, Campus-based
- **Reading List:** View on UCL website
- **Tutor:** Prof Sebastian Riedel
- **Term:** Term 2
- **Timetable:** View on UCL website

**Find out more**

For more information about the department, programmes, relevant open days and to browse other modules, visit ucl.ac.uk
- Encoder/Decoder Architectures;
- Feature Engineering;
- Deep Neural Networks;
- RNNs, CNNs;
- Attention;
- Word Vectors;
- Pretraining;

**Requisites:**
In order to be eligible to select this module, a student must be registered on a programme for which it is a formally-approved option or elective choice AND must have (i) an understanding of Basic Probability Theory (e.g. Bayes Rule), Linear Algebra and Multivariate Calculus; (ii) proficiency in coding in Python; and (iii) the ability to install libraries on a computer.