EPFL

EE-724   **Human language technology: applications to information access**

Popescu-Belis Andrei

Cursus	Sem.	Type
Electrical Engineering		Opt.

Language of teaching	English
Credits	4
Session	
Exam	Multiple
Workload	120h
Hours	**56**
Courses	28
TP	28
Number of positions	

Frequency

Every 2 years

Remark

Next time: Spring 2024

Summary

The Human Language Technology (HLT) course introduces methods and applications for language processing and generation, using statistical learning and neural networks.

Content

The methods, presented in the HLT course, enable accessing to textual information across three types of barriers: the quantity barrier (large repositories), the cross-lingual barrier (different languages), and the subjective barrier (opinions and interactions).

After a brief introduction to the basic stages of natural language processing and their challenges, the course will present through lectures and practical work (50% of the time each) the following approaches that overcome the three barriers to information access:

• The quantity barrier: vector space models for information retrieval; word vectors and embeddings; document classification and similarity using non-contextual embeddings; learning to rank in information retrieval; question answering using Transformer-based models.

• The cross-lingual barrier: brief history of machine translation (MT) with n-gram models; decoding; recurrent neural models with attention; the Transformer for MT; cross-lingual transfer and multilingual MT; translation biases and evaluation issues.

• The subjective barrier: neural models for sentiment analysis; language in social media analysis; text generation using recurrent NNs or attention-only models; response generation for chatbots using RNNs; neural dialogue modeling.

• Issues in data-driven HLT, especially for very large models: capabilities, power consumption, and ethical problems.

Keywords

Human language technology, language engineering, neural networks, machine translation, information search and retrieval.

Learning Prerequisites

Recommended courses

At least one prior course in statistics, machine learning, or computational linguistics. Ability to use Python for simple projects based on existing libraries.

Learning Outcomes

By the end of the course, the student must be able to:

• Explain the main neural network architectures used for human language technology

• Categorize HLT tasks and list state-of-the-art solutions to solve them

• Match in creative ways existing HLT building blocks to achieve new functionalities

• Assess / Evaluate critically the impact of training data on the resulting systems, the related ethical issues, and bias correction strategies.

Assessment methods

Project report and oral presentation.

Resources

Websites

• http://iict-space.heig-vd.ch/apu/

Moodle Link

• https://go.epfl.ch/EE-724