SD TSIA 214 – Machine learning for natural language processing

Chloé Clavel,
chloe.clavel@telecom-paristech.fr,

Telecom ParisTech, France

Natural Language processing ?

In French : Traitement Automatique du Langage Naturel (TALN)

Crossroads of :

- Artificial Intelligence
- Linguistics
- Machine learning

Natural Language processing ?

Objectives :

- Extract meaning from textual data
- Speech synthesis, natural language generation

`https://www.youtube.com/watch?v=Ea_ytY0UDs0` Luc Steels - BREAKING THE WALL TO LIVING ROBOTS. How Artificial Intelligence Research Tries to Build Intelligent Autonomous Systems - 1 min 52

NLP applications

?

NLP applications

- Automatic translation (Google translate)
- Textual data mining/ document classification / information extraction
- Spell-checkers
- Automatic summary
- Human-Computer interactions
- Speech recognition
- Speech synthesis
- Opinion analysis (from social media)

The textual Data and its challenges

Challenges : Moving away from the academic writing to spontaneous expressions (abbreviations, hashtags, acronyms, typos/mistakes, oral transcript)

Lectures and pedagogical team

1. Introduction to Natural Language Processing **CLAVEL Chloe**
2. Natural Language Preprocessing and resources **CLAVEL Chloe**
3. Syntax and Parsing **Jean-Louis Dessalles**
4. Text clustering and text categorization **CLAVEL Chloe**
5. Deep learning for NLP **CLAVEL Chloe**
6. Hidden Markov Models **Laurence Likforman**
7. Non-negative Matrix Factorization **Slim Essid**
8. ML for opinion analysis **Chloé Clavel**

Lab sessions

1. Syntax and Parsing **Jean-Louis Dessalles**
2. Naive Bayes for opinion categorization **Chloé Clavel**
3. Text segmentation using Markov Models **Laurence Likforman**
4. NMF For Topic Modelling **Laurence Likforman**

Evaluation

- Multiple Choice Questions Test
- Labs

At the end of the course...

- You will be able to describe and implement the different methods for text representation into vectors
- You will master the main linguistic issues for NLP
- You will be able to build a text classification framework
- you will master more involved machine learning methods

Prerequisites

TSIA-SD 210 : supervised machine learning methods including neural networks

Materials

See pedagogical website