INF2499M: Natural Language Processing
Overview

Hrishikesh Terdalkar
hrishikesh.terdalkar@univ-lyon1.fr

Laboratoire d'InfoRmatique en Image et Systèmes d'information
Université Claude Bernard Lyon 1
https://moodle.univ-lyon1.fr/course/view.php?id=9902

2024-2025

Tue 0945-1300 at Nautibus TD001 (RDC)

Rules

- "Prerequisites"
  - Proficiency in Python (or the willingness to acquire it)
  - Understanding of Linear Algebra and Probability is *helpful*
  - Familiarity with basic ML algorithms (e.g. Regression, SVM)
- Participate
  - Attend classes on time
  - Ask and answer questions
- Do assignments individually unless asked otherwise
- Discussion with peers is allowed
  - However, copying is not allowed
- Email: `hrishikesh.terdalkar@univ-lyon1.fr`
- Subject must start with `[INF2499M]` to avoid landing in spam

Grading Policy

- Assignment: 20%
  - Mid-term individual coding assignment(s)
- Project: 80%
  - Groups of *at most* two
  - Form your own idea and get it approved
  - Just implementation or survey will not be enough
  - Proposal (10), Implementation (40-50), Presentation (10), Report (10-20)
- Evaluation Criteria
  - Clarity of proposals
  - Implementation effort
  - Robustness of methodology
  - Presentation of results

Project Deliverables

- Proposal (10%)
  - Presentation in the class
  - 5 min per group
  - Outline the project, dataset, and proposed approach
- Implementation (40-50%)
  - Functional implementation pipeline
  - Well documented code
- Final Presentation (10%)
  - Presentation in the class
  - 10 min presentation + 2 min questions
  - Explain the problem, approach, results and key takeaways
- Final Report (10-20%)
  - Introduce and motivate the problem
  - Describe methodology, challenges, solutions
  - Results and ethical considerations

Course Material

- Slides
- Books
  - *Speech and Language Processing*. Daniel Jurafsky and James H. Martin. https://web.stanford.edu/~jurafsky/slp3.
- Documentation of standard python libraries
- Research articles from journals/conferences
  - ACL, NAACL, EACL, EMNLP, CL, COLING, WWW, etc.
- References mentioned in classes

Course Contents I

1. Introduction to NLP
   - What is NLP?
   - Linguistic concepts: syntax, semantics, rule-based NLP
   - Text preprocessing: tokenization, stemming, lemmatization, stopwords
   - Regular expressions, Edit distance
   - Classical models for NLP: Naive Bayes for Classification
2. Word Embeddings and Feature Representations
   - Introduction to word embeddings: Word2Vec, GloVe, FastText
   - Vector representation of words and phrases
   - Importance of context in embeddings: ELMo

Course Contents II

3. Language Models and Sequence Modeling
   - Introduction to Language Models
   - Statistical n-gram models, Markov Models
   - Recurrent Neural Networks (RNN)
   - Gated Recurrent Units (GRU)
   - Long Short-Term Memory (LSTM) Networks
   - Language generation
   - Sequence-to-sequence modeling
4. Transformer Architecture and Attention Mechanism
   - Limitations of RNNs/LSTMs
   - Motivation behind Transformers
   - Breakdown of Transformer architecture
   - Attention mechanism

Course Contents III

5. BERT and Transfer Learning in NLP
   - Introduction to BERT (Bidirectional Encoder Representation from Transformers)
   - Pre-training and fine-tuning: Transfer learning in NLP
   - Applications of BERT in tasks such as sentiment analysis, question answering
6. Large Language Models (LLMs) and Generative AI
   - Overview of Large Language Models: GPT-2, GPT-3, GPT-4, Multimodal LLMs, Gemini, Claude, Llama, etc.
   - Text generation, few-shot learning, and applications in real-world scenarios
   - Ethical considerations: Bias, fairness, and safety in LLMs
7. Special Topics in NLP
   - Multilingual NLP and Machine Translation
   - Low-resource NLP tasks and Zero-shot learning
   - Explainability in NLP: Making models more transparent

Course Schedule

Tentative

Date	Type	
7 Jan	CM + TP	Introduction
14 Jan	CM + TP	Embeddings
20 Jan	CM + TP	Language Models
21 Jan	CM + TP	Sequence Modeling
		Project Proposal
28 Jan	CM + CM	Transformers
		Assignment Due
4 Feb	CM + TP	Transfer Learning
11 Feb	CM + TP	LLMs
18 Feb	CM + TP	LLMs
25 Feb	CM + TP	Special Topics
4 Mar	TP + TP	Project Presentation