

ITCS 4111/5111: Intro to Natural Language Processing

Final Project

Razvan C. Bunescu

Department of Computer Science @ CCI

razvan.bunescu@uncc.edu

Models to be covered

- **Recurrent Neural Networks (RNNs, LSTMs):**
 - Sections 9.1 to 9.6 in J&M.
 - Sequence labeling and Classification:
 - Text Classification, e.g. Sentiment Analysis.
 - Relation extraction.
- **Word Embeddings and Pre-training for NLP:**
 - **Word2vec.**
 - **Transformer, GPT, and BERT.**
 - Sections 9.7 to 9.9 in J&M.
 - Many NLP tasks.

Implementations, publicly available

- **RNNs, LSTMs:**
 - <https://pytorch.org/>
 - https://pytorch.org/tutorials/beginner/nlp/sequence_models_tutorial.html
- **Transformer-based models:**
 - <https://nlp.seas.harvard.edu/2018/04/03/attention.html>
 - https://pytorch.org/tutorials/beginner/transformer_tutorial.html
 - <https://huggingface.co/models>
 - **BERT:** <https://huggingface.co/bert-base-uncased>
 - **GPT2:** <https://huggingface.co/gpt2>
 - **T5:** <https://huggingface.co/t5-small>
 - **OpenAI GPT:** <https://platform.openai.com/>

NLP Applications

- **Text classification:**
 - Sentiment analysis.
 - Document classification.
- **Information Extraction:**
 - Named Entity Recognition, Relation Extraction, Named Entity Linking, Coreference Resolution.
 - Chapter 17 in J&M, chapter 17 in Eisenstein.
- **Question Answering:**
 - Chapter 23 in J&M.
- **Text generation:**
 - Summarization (extractive vs. abstractive).
 - Dialogue (chatbots vs. task-oriented).
 - Chapter 19 in Eisenstein, chapter 24 in J&M.

NLP Applications & Models with my students

- **2023: Socratic debugging**
 - A dialogue system that helps beginner programmers fix their buggy code on their own using guidance from a LM.
- **2023: Bayesian surprise and serendipity for recommender systems**
 - Figuring out how to recommend books that a user finds surprising and that are likely to be rated positively by the user.
- **2022: Measures of surprise in creative language**
 - Creative use of language, such as metaphors and humor, relies on word choices that are unexpected (surprising).
- **2022: Semantic parsing of mathematical proofs into code**
- **2022: Creating variations of short stories**
 - Changing the narrative perspective.

My Group's NLP Applications & Models

- 2022: Conversational AI, automated TA for coding classes, semantic parsing of mathematics into code, humor generation, ...
- 2021: Style transfer, e.g. changing the narrative perspective.
- 2019: Question answering on medical data, using semantic parsing.
- 2019: Diagnosis of speech disorders, using sequence models.
- 2019: Caption generation for figures.
- 2017, 2010: Question ranking in community question answering.
- 2016: Tone classification in Mandarin Chinese.

My Group's NLP Applications & Models

- 2016: Information retrieval in software engineering.
- 2015: Mapping bug reports to source code files.
- 2013: Sense disambiguation and sense clustering in Wikipedia.
- 2012: Coreference resolution using adaptive clustering.
- 2012: Identifying interlingual links in Wikipedia.
- 2011: Grading short answer questions on student exams.
- 2010: Question focus identification.
- 2007: Biomedical relation extraction using Multiple Instance Learning.

My Group's NLP Applications & Models

- 2007: Relation extraction from the web using minimal supervision.
- 2006, 2005: Extraction of protein-protein interactions from Medline abstracts.
- 2006: Named entity linking using Wikipedia.
- 2005: Relation extraction using subsequence and dependency kernels.
- 2003: Sentiment analysis.
- 2003: Associative anaphora resolution.
- 2001: Question answering.
- 2001: Coreference resolution.

Proposal: Your NLP Application or Model

- Submit a 1-2 page PDF document detailing your proposal for the final project.
 - What **problem** you want to address.
 - What **dataset(s)** you plan to use.
 - A rough **plan** for how you will pursue the project (e.g., “we propose to download X system, run it, then implement our system on top of their framework and compare the results”).
 - You can also elaborate on **ML model / architecture, features, ...**
 - What is the **novel contribution** of your project. For example:
 - A new application (=> new dataset).
 - An experimental comparison of state-of-the-art models on an NLP task.
 - A new NLP model, an application of NLP techniques to extract linguistic insights from data, ...
- While you don't need a full **related work** section, you should mention a few pieces of prior work and state how your project relates to them.

Proposal: Your NLP Application or Model

- Only submit a proposal if you are confident that:
 - You can make a **novel contribution**.
 - It represents a **substantial amount of work**:
 - Substantially more than a regular homework assignment.
 - You can **finish** it by the final exam date:
 - Code, data, experimental evaluations, and **project report**.
- If the proposal is not convincing, the recommendation will be to take the final exam.
- On the course website, under the Final Project heading:
 - Tips for choosing a project topic.
 - Project report guidelines.