CS 584 Natural Language Processing

Course Project

Yue Ning
Department of Computer Science
Stevens Institute of Technology

Project

- **An application research project**: A project demonstrates the application of some techniques discussed in class in natural language processing. Properties, drawbacks, advantages of the used techniques are analyzed within the context of the explored application.
- **A theoretical or methodological research project**: A study of different classes of models and approaches discussed in class; proving either theoretically or experimentally properties of known algorithms; designing a new approach.

Project Milestones

- **Proposal report (2 pages)**:
  - Introduction, background, data, your method, planned evaluation.
  - Due: Nov. 4, 2020
- **Project final report (6-10 pages)**:
  - Introduction, background, your method, experimental design, analysis of the results, conclusion and future work
  - Due: Dec.16, 2020
- **Project presentation**: Dec. 9, 2020 and Dec. 16, 2020:
  - 10 minutes total
  - If it is a team work (up to 2), both have to present

Problems & Data

- Kaggle examples (some are expired/completed):
  - Real or Not? NLP with Disaster Tweets: Predict which Tweets are about real disasters and which ones are not.
  - TREC-COVID Information Retrieval: Build a pandemic document retrieval system.
  - Google QUEST Q&A Labeling: Improving automated understanding of complex question answer content.
- Pros of Kaggle:
  - Some have money prize
  - Training and testing data provided
  - Strict evaluation metrics

Problems & Data (Cont.)

- Codalab Competition:
  - Reading Comprehension of Abstract Meaning: In the task, computers are given passages to read and understand. If a model can digest the passages as human do, we expect it can predict abstract words that human being use to write summaries after understand the passage.
  - Evaluating grammatical error corrections
  - MeasEval - Counts and Measurements MeasEval is a new entity and semantic relation extraction task focused on finding counts and measurements, attributes of these quantities, and additional information including measured entities, properties, and measurement contexts.
  - Toxic Spans Detection Highlighting toxic spans in toxic comments.
- Others (collect your own data)
  - Rock music lyrics generation
  - Text summarization

Course Project Proposal

Provide details and a literature review of the problem:

- Introduction (see later pages)
- Background/Related work (see later pages)
- Data sets: provide the details of the datasets including statistics and features.
- Method: present the set of methods you plan to design/apply and explain why.
- Evaluation: what evaluation metrics you plan to use.

Final Course Project Report

Should have the structure of a conference paper:

- Introduction
- Background/Related work
- Your approach
- Experimental design
- Experimental results
- Analysis of the results
- Conclusion and future work

Sections

- Introduction: Describe the problem; why is it important; context; motivating examples; state and summarize the scope and objectives of the project.
- Background/Related work: brief summary of previous work done in the specific area; emphasis is on the limitations; use this section to demonstrate the relevance of the problem you want to work on.
- Your approach: Your point of view of the problem; scope and objectives of the project; your effort; proposing a new approach? comparing existing approaches? evaluating in terms of accuracy, efficiency ...? proposing an analysis to achieve a better understanding?

Sections (cont.)

- Experimental design: software; algorithms; data sets used in your experiments; specify sources; software publicly available used; software/algorithms that you implemented; experimental setting; training/testing. cross-validation; parameter setting; validation measures: accuracy, precision, recall, RMSE, MSE, running time etc; Do NOT write the steps to install the software you used and similar system issues.
- Experimental results: Describe and comment the results obtained. You should be able to elaborate and answer the questions/issues raised in the introduction/approach sections.
- Conclusion and future work: additional directions worth exploring; results obtained suggest new directions?

Grading Principles on Reports

- Your approach/objective + experimental results is the core of the paper
- Well organized
- Well written
- Ideas are clearly stated
- Concepts are formally stated
- Correctness
- Be precise and concise
- Max 10 pages (including references)
- Use the provided Overleaf template on Canvas

Code

- Each project should involve some programming.
- You are required to turn in your code as well.
- You can use scikit-learn, keras, pytorch, or other tools. But they should NOT be the focus of your project.
- Your code should have no errors and be able to run on a different machine.
- You need to provide detailed comments in your code.