CS 559 Machine Learning

Course Project

Yue Ning
Department of Computer Science
Stevens Institute of Technology

Project

- An application research project: The project demonstrates the application of some techniques discussed in class in an application domain (e.g., text mining, bioinformatics, computer vision, image processing, artificial intelligence etc.). Properties, drawbacks, advantages of the used techniques are analyzed within the context of the explored application domain.
- A theoretical or methodological research project: A study of different classes of models and approaches; proving either theoretically or experimentally properties of known algorithms; designing a new approach.
- Might be on your resume.

Project Milestones

- Proposal report (1-2 pages):
  - Introduction, background, data, your method, how to evaluate
  - Due: Oct. 31, 2024
- Project final report (6-10 pages):
  - Introduction, background, your method, experimental design (data, metrics), analysis of the results, conclusion and future work
  - Ideally ready to publish
  - Due: Dec. 19, 2024
- Video:
  - 10 minutes total
  - If it is a team work (up to 2), both have to present
  - Due: Dec. 19, 2024

Problems & Data

- Kaggle examples:
  - House Prices: Advanced Regression Techniques: predict sales prices and practice feature engineering
  - ImageNet Object Localization Challenge: identify the objects in images
  - Movie Review Sentiment Analysis: classify the sentiment of sentences from the Rotten Tomatoes dataset.
- Pros of Kaggle:
  - Some have money prize
  - Training and testing data provided
  - Strict evaluation metrics

Problems & Data (Cont.)

- KDD cup examples:
  - 2011: Predict music ratings and identify favorite songs
  - 2012: Predict the click-through rate of ads given the query and user information
  - 2022: Spatial Dynamic Wind Power Forecasting Challenge
  - 2024: Amazon LLM Multi-Task Shopping
- Pros of KDD cup:
  - High impact problems
  - High reputation
  - Strict evaluation metrics
  - Some have money prize
  - Data provided

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 avenues worth exploring; results obtained suggest new directions?

The whole paper

- 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)
- Latex is highly recommended: You can use this template on Overleaf.

Code

- Each project should involve some programming.
- You are required to turn in your code as well.
- You can use scikit-learn, or other tools. But they should NOT be the focus of your project.

Video presentation

- Each team should make a video for the final presentation to talk about the project you worked on and the results.
- It should be about 10 mins.
- Upload the video to Canvas.