Midterm Project

(CMPSC 497; modified 2/11/2025)

Project description: Pls find any NLP task you are interested (other than POS tagging, Sentiment Classification and Paraphrase Identification we worked on in classroom); Applying Word Embeddings and Convolutional Neural Networks or Recurrent Neural Networks to solve it.

Participants: individually or as a group (max 3 members in a group)

Data: you decide

Requirements: i) you can use publicly available word embeddings, or train your word embeddings with tools such as Word2Vec, GloVe. ii) Your algorithm can be pure convolutional or recurrent NN, or their combination. iii) no other algorithms are allowed, e.g., SVM, Transformer, GPTs

What you need to submit (deadline 11:59pm on 2/28/2025):

URL of your github repository, including

- **A PDF (with your name or all members in your group) file describes everything**
- **Code files** for the whole project with a clear readme regarding how to reproduce your results.

Submit through Canvas assignments.

Evaluation:

- **PDF quality (80%)**: we score by the following five aspects:
  - Problem definition and Dataset curation (20 points)
  - Word Embeddings, Algorithm and how you train it (i.e., your optimization details) (20 points)
  - The results and how they are presented (20 points)
  - Any in-depth analyses/experiments (20 points)
  - Lessons&Experience you learned in this project (20 points)

  For individual participants, points will be deducted in units of **5**; for group participants, points will be deducted in units of **4+2*G** where "G" refers to group size (2 or 3).
- **(optional) Presentation (20%)**: Each participant (individual or group) can choose whether you want to present or not. You can get a default score **18** (out of max 20) for the presentation part if you choose NOT (such as your scores of other parts can secure an "A" for you). If you choose to present, the evaluation criteria include: slide quality, work

quality, presentation skills, question answering, etc.) For individual presentations, points will be deducted in units of **2**; for group presentations, points will be deducted in units of **4**.

- **If it is a group project, each team member gets the same score**.