

CS 584-A: Natural Language Processing

Homework 2

Due Date: Monday 11:59 AM (noon) EST, October 30, 2023

Total: 100 points

Goals

The goal of HW2 is for you to get hands-on experience of implementing RNN and CNN models for the classification task. You will get a deeper understanding of how these models are applied to text data. The skills you learn from using RNN and CNN models can be applied to a wide range of NLP tasks beyond text classification. Additionally, the skills you acquire will be transferable to more advanced topics, such as attention mechanisms, transformers, and other complex architectures. Please feel free to use any packages or libraries in your implementation.

Similar to HW1, all questions are open questions and there is no fixed solution. The difference in data processing, parameter initialization, data split, etc., will lead to the differences in predictions and evaluation results. Therefore, during the grading, the specific values in the results are not required. It is important that you focus on implementing and setting up the pipelines of applying these models to solve the tasks.

Dataset

Please use the same product review dataset as in HW1.

Task: Sentiment Analysis with Text Classification

The task is still to perform sentiment analysis, which predicts the sentiments based on the free-text reviews, which is the same as in HW1.

Differences from HW1: If you worked on the multi-class classification problem in HW1, please change to work on the binary classification problem in HW2. If you worked on the binary classification problem in HW1, please change to work on the multi-class classification problem in HW2. Please find the details about how to process the labels below. ***In the submission, please make sure to clearly mention which task you are solving and how did you prepare the class labels.***

For the multi-class classification problem, we will have 5 classes in total. To simplify the task as a binary classification problem, we will consider ratings 4 and 5 as positive, and rating 1 and 2 as negative. For data samples with neutral rating 3, you can discard them or simply consider them as either positive or negative samples.

1. Data preparation (10 points)

- 1) **Data preprocessing (5 points):** Please use the same data processing steps as you did in HW1. Please make sure to process the class labels based on the binary/multi-class classification task that you will be working on in HW2.
- 2) **Data split:** Please use the same data split as you did in HW1.

- 3) **Data statistics (5 points):** Since you are using the same data split as in HW1, most of the data statistics will not be changed except for the number of reviews of each class in training/development/testing since the class labels are changed in HW2. Please provide the number of reviews of each class in training/development/testing for the classification task in HW2.

2. Sentiment Analysis with RNN (40 points)

Please write the code to perform the sentiment analysis task you formulated in question 1. During the implementation, you will need to follow the requirements listed below.

- 1) You can select to implement 2-layer LSTM or GRU (you can directly call packages in Pytorch).
- 2) Please use **SGD** during optimization.
- 3) Please initialize the word embeddings **randomly** and learn them during the model training.
- 4) You can decide other parameters.

3. Sentiment Analysis with CNN (40 points)

Please write the code to perform the sentiment analysis task you formulated in question 1. During the implementation, you will need to follow the requirements listed below. Feel free to use any packages and libraries.

- 1) Please use **mini-batch gradient descent** method during optimization with batch size 20.
- 2) Please initialize the word embeddings with the **pre-trained glove embeddings** you used in HW1 and update them during the model training.
- 3) You can decide other parameters.

4. Evaluation (10 points):

Train the model on the training set, select the best model based on the validation set, and evaluate your model on the testing set.

- 1) Evaluate the model performance using metrics for classification, such as accuracy, precision, recall, F1-score, and AUC. Report your results for both methods. You can call classification report in sklearn. (4 points)
- 2) Have a brief discussion to compare the performance of the two models. It should be noted that there is no fixed answer for the results. You will need to report the exact results returned in your experiments. The discussions should only be based on your own experimental settings and returned results. (3 points)
- 3) Are there any differences between the results you obtained in HW1 and HW2? Which model performs best? Can you please provide some discussions about your findings? It would be great to think about and discuss the underlying reasons for the outperformance of the best model. (3 points)

Submission guidelines:

1. In your codes, please try to add **clear comments and detailed steps** to get all the partial points. Also, include the required discussions and results analysis in the text cell.
2. Please use Jupyter Notebook for your programming. Before submission, please make sure to clear all existing outputs and rerun all the codes. Once it is done to run all the codes, save it as a pdf file. You will need to **submit both the .ipynb and .pdf files**. Both files should include all the codes, comments, discussions, and answers for the questions.
3. You can also organize all the results and analysis in a separate pdf file. This is optional.
4. Submissions should be made on **Canvas**.
5. *Each student needs to submit their own answers and codes. Similar or the same answers/codes will not be graded. You will directly fail the course if cheating/plagiarism is found.*
6. **Properly cite any resources** you used for your submission, including but not limited to websites, GitHub repositories, publications, generative AI techniques, etc.