CS5740: Assignment 2

GITHUB_REPOSITORY

Full Name
Net ID

Full Name
Net ID

TODO: Update your name and details above. Names must be presented in alphabetical ascending order by the last name. If not filled correctly, we will subtract 2pt

TODO: Make sure to use the standard notation when defined (`https://www.overleaf.com/read/kbrdcdbtvhgt`). Check for the updated version of the notation.

1 Introduction (5pt)

TODO: The first paragraph is to briefly describe task and data

TODO: The second paragraph should describe your main experiments and results, including mentioning the data you use

2 Model (10pt)

TODO: Formally describe your model. This must include well defined notation and equations of your objective. If you use multiple objective functions, describe all of them. It should be clear what are the different variants your model enables, so when you refer to them later in your experiments it will be clear to the reader (not only to you) what each variant means.

3 Experiments (12pt)

TODO: Describe formally and carefully the expriments you will conduct (the results come in the results section). For each experiment, describe clearly what you vary and what are the different options.

4 Data

Data (5pt) **TODO:** Describe the data you use, including how many examples are in the training, development, and test sets. Please also provide shallow statistics of your data, with at least the vocabulary size, document length (when relevant). It is best to report all the statistics, including counts, in a table. Please describe how you compute statistics that can be computed in different ways (i.e., vocabulary size).

Pre-processing and Handling Unknowns (10pt) **TODO:** Describe how you pre-processed the data, including how you treated casing, tokenization, and anything else that you did to the raw data before computing features from it. Explain your choices, and show the data statistics after pre-processing. If you are using a subset of the data, explain why and discuss tradeoffs (you will also need to back them with experiments). It is also critical that you discuss how you handle unknown words, and what are the tradeoffs of your approach. We will penalize for choices that are obviously wrong.

5 Implementation Details (3pt)

TODO: Briefly describe the implementation details. No need to copy details already specified in the assignment. Include any hyper-parameters the model has. If there are any optimizations that you introduced, this is the place to describe them.

6 Results and Analysis

Test Results (3pt) **TODO:** Provide the test results from the leaderboard. This is best done in a table, and the table must referred from the text.

Development Results (15pt) **TODO:** Describe the results from your experiments. This is best done in a table/plot (or tables/plots), and the table must referred from the text. Make sure it is clear what experiment each result comes from. The text should provide a clear description of each experiment and what conclusions it

provides. Make sure to describe the sensitivity of your model to various conditions, such as using different amounts of data, using different number of training epochs, etc. Naturally, discuss in detail how different choices influence your model performance.

Qualitative Analysis (10pt) **TODO:** Qualitative analysis of selected failure examples. Show and discuss error examples from your development set. You must identify certain classes of errors and use the examples to illustrate them. This is often best to show in a table.

7 Conclusion (3pt)

TODO: Brief conclusion summarizing findings (from both numerical results and qualitative analysis).