UDACITY

PROJECT

Capstone Proposal

A part of the Machine Learning Engineer Nanodegree Program

PROJECT REVIEW

NOTES

SHARE YOUR ACCOMPLISHMENT!

Requires Changes

1 SPECIFICATION REQUIRES CHANGES

Hello Udacian,
You made a very great trial in this section and topic you selected is very interesting. There is only one minor mistake you need to amend, I believe you already understand most of the concept in this project. Looking forward to your next submission. 😉

I added some comments and links, I hope they will be helpful.

- https://www.kaggle.com/c/word2vec-nlp-tutorial/details/part-2-word-vectors
- https://code.google.com/archive/p/word2vec/
- https://deeplearning4j.org/word2vec.html#word2vec
- https://deeplearning4j.org/bagofwords-tf-idf

Good luck!

Project Proposal

Student briefly details background information of the domain from which the project is proposed. Historical information relevant to the project should be included. It should be clear how or why a problem in the domain can or should be solved. Related academic research should be appropriately cited. A discussion of the student's personal motivation for investigating a particular problem in the domain is encouraged but not required.

- Excellent job with the Domain Background.
- A great job was made in providing a thorough overview of the project.
- You did well in mentioning some academics research related to the same topic of NLP.

Student clearly describes the problem that is to be solved. The problem is well defined and has at least one relevant potential solution. Additionally, the problem is quantifiable, measurable, and replicable.

- The problem is clearly defined here.
- Awesome!

The dataset(s) and/or input(s) to be used in the project are thoroughly described. Information such as how the dataset or input is (was) obtained, and the characteristics of the dataset or input, should be included. It should be clear how the dataset(s) or input(s) will be used in the project and whether their use is appropriate given the context of the problem.

> The dataset we are going to use for our analysis is the Quora Question Pairs dataset (https://www.kaggle.com/quora/question-pairs-dataset). The problem we are trying to solve The dataset consists of 404350 question pairs which as labeled as duplicates (1) or not (0) which 149306 are in the 'are duplicates' class. There are 789801 different questions in the dataset.

You made a thorough description of the dataset to use in the project.
In this section, it would be good to provide some example extracted from the dataset.

Student clearly describes a solution to the problem. The solution is applicable to the project domain and appropriate for the dataset(s) or input(s) given. Additionally, the solution is quantifiable, measurable, and replicable.

> We can improve on our benchmark model by using advanced word to vector conversion models such as word2vec and then use an advanced deep learning model to better predict if the questions are duplicate or not within each pair.

Required:

Please clearly mention which advanced deep learning is mentioned in the report. Is it LSTM, or hybrid LSTM with CNN encoding or CNN? I suggest below readings that are also some related solution.

- https://web.stanford.edu/class/cs224n/reports/2759336.pdf
- https://web.stanford.edu/class/cs224n/reports/2748045.pdf

A benchmark model is provided that relates to the domain, problem statement, and intended solution. Ideally, the student's benchmark model provides context for existing methods or known information in the domain and problem given, which can then be objectively compared to the student's solution. The benchmark model is clearly defined and measurable.

The benchmark you selected seems good to me.

Student proposes at least one evaluation metric that can be used to quantify the performance of both the benchmark model and the solution model presented. The evaluation metric(s) proposed are appropriate given the context of the data, the problem statement, and the intended solution.

You are on the right track, the metrics are clearly defined and a justification was provided.

Student summarizes a theoretical workflow for approaching a solution given the problem. Discussion is made as to what strategies may be employed, what analysis of the data might be required, or which algorithms will be considered. The workflow and discussion provided align with the qualities of the project. Small visualizations, pseudocode, or diagrams are encouraged but not required.

A good strategy was conducted to approach the solution in the report.

Proposal follows a well-organized structure and would be readily understood by its intended audience. Each section is written in a clear, concise and specific manner. Few grammatical and spelling mistakes are present. All resources used and referenced are properly cited.

- The report is clear, well organized, readable, and easy to understand.
- Some references are provided at the end of the report.

Awesome work is done here.

RESUBMIT

DOWNLOAD PROJECT

Best practices for your project resubmission

Ben shares 5 helpful tips to get you through revising and resubmitting your project.

Watch Video (3:01)

RETURN TO PATH

Student FAQ