Capstone Proposal

REVIEW

CODE REVIEW

HISTORY

Meets Specifications

Congratulations Minh Gia Hoang,

The dog breed classification is an interesting project, and you demonstrated that have familiarity with the strategy that are you going to use in this problem. Just keep one eye on the class label distribution to see if your benchmark works or if it is necessary to adjust as commented in the specific rubric.

I hope the suggestions are useful in the development of your final project.

Happy studies,

Project Proposal

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.

*Awesome**
The problem is clearly defined as a multi-class classification task, which is based on dog and human images to predict the resembling dog breed.

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.

Awesome
You did a great job presenting the domain background:

- You clearly mentioned the relevance of the problem domain and its applicability to machine learning.
- There is an introduction about CNN and some different approaches that have been applied to dog breed classification. It is really interesting because it justifies the use of CNN.
- You made use of academic research and properly cited it.

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.

Awesome
The solution to the problem is very well explained and it is a good strategy to work on. Well done!

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.

Suggestion
You defined as a benchmark the CNN model from scratch (10% of accuracy) and a CNN pre-trained model (60% of accuracy). However, the benchmark model must be objectively checked against the models you produce in this work in the same metric. The values of accuracy for the models as a benchmark just work if your dataset is balanced or slightly imbalanced. This is why it is relevant to plot the class label distribution.

If you identify that your data is strongly imbalanced, you could set as a benchmark the simplest model. For instance, CNN with 1 Convolutional Layer and use other metrics instead of accuracy that distinguishes correctly between the numbers of correctly classified examples of different classes, as you mentioned (precision and recall).

Metrics for an imbalanced dataset: https://towardsdatascience.com/metrics-for-imbalanced-classification-41c71549bbb5
Accuracy alone not good for an imbalanced dataset: https://tryolabs.com/blog/2013/03/25/why-accuracy-alone-bad-measure-classification-tasks-and-what-we-can-do-about-it/

The important thing is that you have a well-defined benchmark model that yields results that can be compared against your final solution. This comparison should make some sense, that is, you will be able to discuss its outcome (example: "The final model exceeded the benchmark proposed by so-and-so, so we can conclude that ...").

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.

Awesome
Precision and recall work well as the metric to measure the performance of the models (benchmark and final solution).

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.

Suggestion
You described the dataset with sufficient details, mentioning the source and some relevant characteristics. Additionally, it is also recommended to provide some comments about:

- some images as a sample (to demonstrated different backgrounds).
- details about the image sizes (width and length).
- to plot the class label distribution. You mentioned that the data is imbalanced; however, it is also important to demonstrate it.

Class label distribution:
https://www.cognitivecoder.com/2018/03/29/3-quick-ways-to-create-graphs-of-your-class-distributions-in-python/
http://doc.perclass.com/perClass_Toolbox/guide/visualization.html

Student summarizes a theoretical workflow for approaching a solution given the problem. A 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.

Suggestion
The theoretical workflow with a discussion for each step is very well presented. Additionally, also consider to include comments about:

- Data exploration - Are there any characteristics of the data that you would like to observe? Such as the class label distribution.
- Preprocessing phase - Are you doing to transform the input data somehow? For example, resizing all images so that they all have the same width and height.

The 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.

Awesome
You followed the template and the proposal is very well written.

DOWNLOAD PROJECT

RETURN TO PATH

Rate this review

START