April 5, 2021

Milestone 1 - Team Ornithomimus

State of the Project

For our image classifier, we created a web scraper to pull photos from AllTrails. Using this scraper, we downloaded 3 images from each hike in British Columbia, for a total of over 8000 images; these photos will be used for our image classifier dataset, as well as for the images that will be shown to the user when they rate each hike. For our recommender system, we created another scraper to download the characteristics of each hike. Each hike is separated by name and stored with its attributes in a CSV file.

While all the data that will be used in our datasets have been gathered, our datasets themselves are not complete. We will work out the exact implementation of each dataset as we develop our machine learning models. For more information on the dataset challenges, see the section on current challenges.

Feature Changes

In our original proposal, we specified that the recommender system would use a collaborative filtering method for recommending hikes. This would have required downloading over a hundred thousand unique reviews from several thousand users. To reduce the size of our dataset, we will instead use a content-based filtering system. Our image classification dataset consists of over 8000 photos, and since we will be combining the results of both models, we are not concerned that changing the recommendation method will have a significant impact on the effectiveness of the system as a whole.

Additionally, we are considering using Django for our backend instead of Flask. We consulted with a mentor who pointed out that Django offers more features that will allow us to have a functioning app faster, so we need to consider which framework best fits our needs as a team. We will give an update on our decision at Milestone 2.

Current Challenges

As our app runs, the user makes decisions about which hikes they like and dislike. Based on these preferences, our machine learning models need to update their datasets in real time, labelling new samples with the user's preferences. Our biggest challenge right now is how to handle these dynamically updated datasets, and how to retrain the models.

In our experience, we have always retrained the model from scratch until it reaches convergence. Now that we have a dataset that grows each time the user makes a decision, we need to decide whether to partially train the model or to train the model from scratch each time. Further research will be required to find a solution to this problem, and we may attend lab sessions for advice.

Team Contributions and Upcoming Tasks

Andrew Forde

Contributions: Made image scraper for pulling hike photo links from AllTrails; gathered URLs for hike images, saved in CSV format; downloaded photos for easy hikes.
In-progress: Planning wireframe for Flutter application, collaborative development of content-based recommender system.

Taylor Mcouat

Contributions: Developed web scraper for pulling hike information from AllTrails. Downloaded information for ~3,000 hikes. Downloaded ~3,000 photos for moderate hikes.
In-progress: Researching web-application framework (Django vs. Flask), researching/developing recommender-system ML model.

Leo Sun

Contributions: Explored app API connections between front-end and back-end.
In-progress: Drafting object model, planning and developing ML models.

Jayden Wong

Contributions: Downloaded photos for hard hikes, merged branches, created git projects and group chat.
In-progress: Collaborative development of content-based recommender model and Flutter front-end, ideating/developing object model and app wireframe.