OnSight: Outdoor Rock Climbing Recommendations

Brian Cheng brc042@ucsd.edu, Eric Liu eyl052@ucsd.edu, Brent Min bjmin@ucsd.edu

ABSTRACT

Recommendations for outdoor rock climbing has historically been limited to word of mouth, guide books, and most popular climbs. With our project OnSight, we believe we can offer personalized recommendations for outdoor rock climbers.

INTRODUCTION

Recently, rock climbing has grown in popularity. Most of the growth has been in indoor gym climbing, while outdoor climbing hasn't experienced the same growth. We think the reason for the lack of growth in outdoor climbing is related to the lack of good recommendations. Currently, there is no established way of getting recommended personalized climbs. The traditional ways would be through word of mouth, guide books, or websites like MountainProject.com. However, each of these options have drawbacks.

For word of mouth, it requires you to have contact with knowledgeable outdoor climbers, which many indoor climbers do not have. Additionally, knowledgeable outdoor climbers can only have complete knowledge of one or two areas, so cross area recommendations are impossible.

In the case of guide books, the only form of recommendation they offer are the author's ratings for the climb. Obviously there is no way to tailor a guide book for an individual. Furthermore, guide books are very rarely free, and only cover one climbing area, so that cross climbing area recommendations are impossible.

The final place to get outdoor climbing recommendations are online websites, such as Mountain Project. Similar to guide books, these websites only offer the most popular climbs based on the community's ratings. On no website is there any way to get tailored recommendations, even though the internet is the perfect medium.

These downsides are a large part of our motivation to create a better outdoor rock climbing recommender.

By creating a recommender system for outdoor rock climbing routes, we would help grow the outdoor climbing scene by letting current indoor climbers transition to outdoor climbing. It would also improve the experience for outdoor climbers by allowing them to find good and relevant climbing routes.

DATA ACQUISITION

Data for this recommender system came from MountainProject.com. Mountain Project is a site where users can submit data about climbing routes they found and subsequent users can rate those routes. We chose Mountain Project because it has a large catalog of climbs located in the United

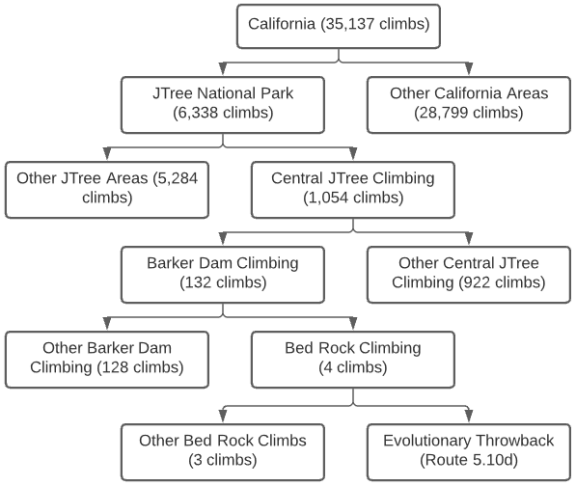
States and comprehensive data about climbing routes.

Mountain Project contains data on all types of rock climbs: boulder, route, aid, ice, and mixed. Of these five types of climbs, we only scraped data for boulder and route climbs due to limitations in our data storage and memory limits. Boulders are climbs done without a rope, and can range from less than 10 feet to 50 feet (where serious injury or death may occur). Boulder difficulty starts at V0 at the easiest, and goes to V16 at the hardset. Routes are climbs done with a rope, and can range from 20 feet to multi-pitch climbs thousands of feet long. Route difficulty goes from 5.0 to 5.9, then goes 5.10a, 5.10b, 5.10c, 5.10d, then 5.11a-5.11d, all the way to 5.15d.

Even though much of Mountain Project's database is user submitted, the verification system on the website gives us high confidence that every climb on the website actually exists.

Although Mountain Project did have a data API, it was deprecated right before we started working on OnSight. As a result, we obtained data from Mountain Project by scraping their website. From this scraping, we obtained data for every route and boulder in the United States. For every route and boulder, we obtained a unique climb identifier, the difficulty of the climb, the average rating and number of ratings, latitude/longitude of the climb, number of pitches, height of the climb, and the climb description.

Specifically, scraping was done using the Beautiful Soup python library. The script started at the state level, and did a depth-first search through the various nested climbing areas, until climbs were found. For example, a tree could look like:



This generated a list of climb urls which were then used to scrape data for each climb. Then, raw data was cleaned and uploaded to MongoDB.

Data cleaning was a relatively simple process. First we turned the difficulty strings for boulders and routes into integers. Then we did data imputation on climbs that were missing difficulty or height by filling in with -1. Finally various string data such as climb description was cleaned to remove extraneous new lines, spaces, and odd characters.

Generally, we don't recommend running the scraping, cleaning, and uploading scripts since they take a very long time. For example, in the scraping tree above, California has 35,137 climbs. Scraping all those climbs took about one full day. Because the cleaned data is hosted on

MongoDB, we recommend just querying the data.

There is also a script to scrape user profiles, specifically a user's past climbs. Whenever a user wants personalized recommendations, they will input their Mountain Project profile page. Their past climbing history, or 'ticks', will be used to recommend a set of similar climbs based on how the user rated each climb.

CREATING THE PRODUCT

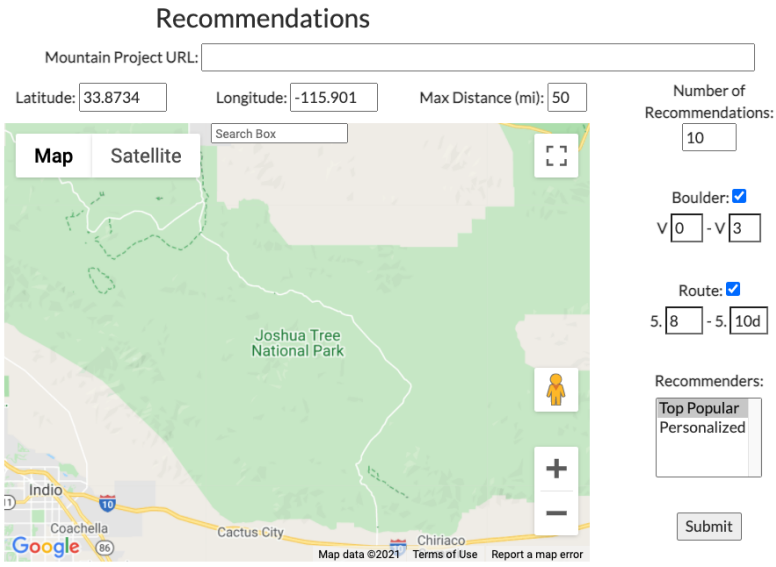
Our web application was created full-stack using Django. The frontend portion are Django templates with the layout determined by Bootstrap 4's grid system.

Our data is hosted separately with MongoDB. MongoDB allows for free data storage and has a simple interface for reading and writing data. Their free data storage limit of 512MB was more than enough for this project, provided that we only scraped boulders and routes.

We are currently using Heroku's free tier to host our web application. Heroku allows for easy web hosting and simple debugging. However, it's free tier only allows for 512MB of memory across all users using the website, so there can be issues when large numbers of people are trying to use the website at once. Additionally, Heroku has a default website request timeout of 30 seconds, which limits the complexity of the recommendation algorithms we use.

THE RECOMMENDER

Our recommender system is delivered as a web application. You can check out our site at https://dsc180b-rc-rec.herokuapp.com/. On the site, you will be asked to input the number of output recommendations, the target location, the maximum range to search for in miles, the climbing difficulty, the climbing type, and the algorithm of the recommender.



After inputting your preferences, the site will output the number of recommendations that satisfy your specifications. Each output will consist of the climbing route's name, the corresponding Mountain Project URL, the difficulty grade, and the description of the climb (limited to about 600 characters).

The application uses the Google Maps API to specify location input. Users have the option to input a location by searching for it in the search bar, by typing in latitude and longitude coordinates, or by dragging the map interface to the target location (the

center of the map will be the target location).

The maximum distance field allows users to restrict location of the recommended climbs. Only climbs within the specified radius of the specified location will be recommended. Users can search within a small radius around the target area or expand the search further away.

The difficulty range input lets users restrict their search to only routes with difficulties users are interested in. There are two inputs for difficulty: rock climbing routes and bouldering routes. Each of the two inputs follow different rating systems. If users are unfamiliar with the rating systems of each style of climb, an article written by REI is provided for reference. If users are only interested in rock climbing or bouldering routes, they can specify to only receive certain styles of climbs by checking and unchecking the styles (ie Boulder and Route). The climbs with unchecked styles will be ignored when producing the recommendations.

The website currently has two implemented recommender systems. A non-personalized recommender which recommends the most popular highly rated climbs at the target area, and a personalized recommender which recommends climbs similar to those the user has enjoyed in the past.

Additionally, for personalized recommendations, users will be asked to input their Mountain Project user profile url in order to find their most liked climbs, as determined by their highest rated. If users do not have a Mountain Project account, or have not given any ratings, they are unable to get personalized recommendations.

Users that are new to outdoor climbing will have to use the non-personalized recommender, as they will not have a Mountain Project profile or will not have ratings.

TOP POPULAR RECOMMENDER

Both the top popular and personalized recommenders first filter the entire database based on certain user input parameters like difficulty level, location, and style of climbing.

The maximum distance input is special because part of it happens before querying any data. Before we apply most of the user input filterings, what we query from the database is dependent on the maximum range input. This is absolutely necessary so that we do not have to download the entire database to the website, nor hold the entire database in memory. To do this, we create a latitude/longitude approximation square, and query the database for climbs inside the square. Typically, this cuts down on climbs from the hundreds of thousands to just a few thousand or even a few hundred.

The distance filter is then finished by using the highly accurate Haversine formula. Climbs located beyond the maximum distance are removed from consideration.

Next, climbs are filtered by the type of climbs requested (boulder and/or route), and then by the difficulty of climbs requested.

Finally, the top popular recommender has an additional filter where only routes with an average rating of 3.5 out of 4 or higher are recommended. We have this filter just in case we recommend a universally hated route even though it has more votes than others.

In the end, we rank the recommendations based on the number of votes and only return the top N ones based on the user's input of number of recommendations.

PERSONALIZED RECOMMENDER

The personalized recommender starts the same as the top popular recommender by applying filters such as maximum range, type of climb, and difficulty level to the entire database.

The recommender then takes the input user profile URL and scrapes it for the user's favorite climbs. We define favorite climbs as all those climbs which have the max rating the user has ever given. Additionally, all climbs the user has done are all removed from the filtered database (we would not want to recommend a climb the user has already done!).

For each of these identified favorite routes and all the possible filtered recommendation candidates, we create a cosine similarity matrix with the aid of Scikit-Learn. We then return the routes with the highest cosine similarity scores. That way, we are recommending routes that are the most similar to the user's favorite routes, which potentially can become their favorites too.

CONCLUSION

With OnSight, we aimed to contribute to outdoor rock climbing by providing both personalized and non-personalized recommendations.

Even though our non-personalized recommender is a top popular recommender, it already goes beyond the functionality of the Mountain Project top popular recommender. Mountain Project limits recommendations to climbing areas, so for example, you would have to choose between California and Oregon climbing areas, and could not search for climbs on the border. OnSight allows for this functionality.

Our personalized recommender is a huge step towards customized outdoor rock climbing recommendations. It particularly thrives in helping climbers find climbs similar to those they have enjoyed in the past. None of the previous recommenders systems such as word of mouth, guide books, or online recommendations have this ability.

DISCUSSION

In this project, a major challenge we had was obtaining the data. When we initially built the prototype of our recommender, we were only basing off of the data of the Yosemite National Park region, and that data took about 45 minutes to scrape. We were a little skeptical as to whether it was possible

to scrape data for the entire United States with just the three of us. Thankfully, with teamwork, we were able to coordinate the scraping effort between us three and obtained data of the entire United States over about a week of non-stop running.

With rock climbing, especially outdoors, there is an inherent risk that is taken when you decide to climb. A boulder climb as short as a few feet can cause injury with a bad landing, and rope climbing gear can always fail. By making rock climbing recommendations, some users may assume that the climbs we recommend are safe, which can never be the case due to inherent risk. We tried to solve this issue by adding a disclaimer to the website that appears upon opening for the first time.

Moving forwards, our future steps include incorporating more climbing types such as ice, aid, or mixed climbs. We will also look into recommending top popular based on a highest rated recommender, taking into account number of ratings with a binomial confidence interval, such as the Wilson score interval.

Once the pandemic lessens out, we will be posting our recommender on platforms such as Reddit inviting interested people to try out the recommendations and let us know if it is helpful or need further modifications.

REFERENCES

Our web application can be found at: https://dsc180b-rc-rec.herokuapp.com/

The original project proposal can be found at: https://docs.google.com/document/d/1PdplYHX-FMLyc7BOgW8-nO3_VWmvvx4BwifSTmlt9RU

Scikit-Learn's formula for cosine similarity: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.cosine_similarity.html