

Machine Learning Engineer Nanodegree

Capstone Proposal

Dominik Lindenberger November, 2018

Proposal

The goal of this project is to build a recommender system for climbing routes within a given climbing area. The system will take a climbers preference and his or her stats into consideration.

Domain Background

My Capstone Project is located in the world of **rock climbing**.

Rock climbing is an activity in which participants climb up, down or across natural rock formations or artificial rock walls. The goal is to reach the summit of a formation or the endpoint of a pre-defined route without falling. ¹

Here is a video of a rock climber on a famous route called [Action Directe](#).

In recent years rock climbing has gained a lot of traction and is becoming more popular as an outdoor sports for a larger number of people. As such, there exists a potentially large audience who could be interested in this project.

Unlike the climber from the video above most people do rock climbing as a leisure activity. Therefore, time they can spend to go climbing is limited. As a hobby climber planning a climbing trip it is quite common to have a time window of round about 1-3 weeks at the climbing destination of choice. The majority of climbing areas boast a large number of routes to climb, far greater than anyone can achieve within a single trip. For example, the climbing area 'Frankenjura', one of Germany's largest climbing area, contains more than 10,000 climbs. It is therefore in the interest of the climber to know and attempt those climbs that are most enjoyable to her or him and just skip the rest.

Often times there is a common agreement on what are the 'best' routes in a certain area and they are common knowledge. Most guide books contain some sort of list of the "Top climbs" in that region. However, climbers often differ in their personal preferences. Some prefer to climb on small crimp holds, while others enjoy long dynamic movements on large holds.

Additionally, for some type of rock, the climbs deteriorate over time as heavy traffic 'polishes' holds with formerly good and important friction.

Therefore, climbers would benefit from a personal recommendation of the best climbs within a climbing

area, that match their climbing level as well as their personal preference.

Meanwhile there are many websites where climbers can track their climbs and ascents in a virtual logbook and also rate the quality of the route. Most famous websites of this kind are

- 8a.nu
- theCrag.com
- UKClimbing.com

With this project, I will explore the database of one of these websites to establish a personal recommendation system for climbers venturing into new climbing areas.

As I am an avid climber myself, this project is something I take a strong personal interest in.

Problem Statement

The problem climbers face when first entering a new climbing area is which routes to go for. Typically there are a lot more climbs available than anybody can manage to climb in a typical climbing vacation. And not all climbs are worthwhile of doing (e.g. boring route, dangerous to climb, lot of dirt and vegetation, etc.). For any particular climber it would be very interesting to have a list of the best routes in each area. Ideally sorted by attractiveness and matched with the climbers skill level. So we are looking at a **ranking** or **recommendation problem**.

Since not all climbers enjoy the same type of climbing, such a list of attractive climbs would ideally be tailored to every individual climber. Such a system of personal recommendation does not exist so far.

Datasets and Inputs

On Kaggle a big data set of logged ascents from the popular website 8a.nu is available². The data set contains *all of the publicly available information from <http://www.8a.nu>, the world's largest rock climbing logbook*. Climbers from all over the world have logged ascents in various climbing areas on this website.

In more detail the data set contains four tables with information as following.

Table	Description
ascent	<p>Holds information about individual ascents (= completed climbs) climbers have made. Has fields such as</p> <ul style="list-style-type: none"> <code>crag</code> - the climbing area <code>sector</code> - a specific sector within a climbing area <code>name</code> - the name of a route in a particular sector <code>rating</code> - user rating for the route
grade	<p>Contains a list of all the different climbing grades, i.e. the difficulty of a climb. (The difficulty of the route "Action Directe" from the video above is 9a which is out of reach for 99% of all climbers.)</p>
method	<p>Defines the different styles of ascents. See this external article³ if you are interested in more detail</p>
user	<p>Information about climbers such as <code>birth</code> , <code>country</code> , <code>height</code> , etc.</p>

The original data set from Kaggle is 196 MB. For this project I will use only a subset, the area "Frankenjura" (crag = "Frankenjura") which is one of the largest and most famous climbing areas in Germany.

For this proposal I will provide a subset of the data and tables already joined.

Solution Statement

My goal is to create a recommender system for climbers that can suggest a list of n climbing routes within a particular climbing area. For an individual climber the recommender should base recommendations on the types of routes the climber liked before and based on what other climbers with a comparable climbing history have liked. The recommended climbs should be ordered by their attractiveness to the climber.

It will be interesting to see if any of the climbers stats (e.g. age, height) have an impact on the recommendations. In the climbing community short and tall climbers are always picking on each other ("You are not climbing, you are just reaching up for the high holds because you are so tall.")

For climbers without any personal climbing record, a simple recommendation should give a list of the top n climbs within the area based on average ratings from all climbers.

Benchmark Model

There are no benchmark models for personal recommendations as this has never been done before.

We will therefore compare our model to the results of

- Random model - model makes random recommendations
- Top average model - model recommends top rated climbs based on mean average of all ratings

- Magazine rated - model pulls recommendations from the Top 100 climbs of Frankenjura as chosen by klettern.de⁴, a well known German climbing magazine.

Evaluation Metrics

Ranking and recommendation problems in general lend themselves to using the following metrics

- Mean average precision (MAP)
- Normalized discounted cumulative gain (NDCG)
- Root mean squared error (RMSE)

During the course of the project I will explore which of these metrics is best applied in our case.

Project Design

When executing this project, I will follow those main steps

1 - Data exploration

I will start with some data exploration to gain insights about data distribution, outliers (if any) and missing values.

2 - Data preparation

During preparation I will try to extract a list of unique routes. In the dataset a record contains currently the route that was climbed and if the climber provided it, also a rating. As such there will be many duplicate routes. String similarity measures like Sørensen–Dice coefficient or Levensthein distance should help here.

Additionally, I will transform data, e.g. one-hot encoding of categorical attributes like `country` or `city` and decide how to do handle missing data.

3 - Model exploration

During the model exploration I will first look at exploring data with KMeans and try to find natural clusters of climbers with a preference for certain types of climbs.

In case results are not satisfactory, I may explore further using Collaborative Filtering (e.g. using Surprise⁵, Association Rules... and decide on the best model.

4 - Fine-tuning of the model Once the most promising model is found, I will continue to fine-tune the model to improve results. At this stage I will also employ evaluation metrics as mentioned in an earlier paragraph.

5 - Presentation of the solution

Finally, I will present results of the entire project and especially the solution to the problem.

- ## 1. What is Rock Climbing? ↩

2. [8a.nu Climbing Logbook ↩](#)
3. [Styles of ascent in sport climbing ↩](#)
4. [klettern.de's Top 100 Routes ↩](#)
5. [surpriselib.com/ ↩](#)