Machine Learning Model for Difficulty Assessment of Climbing Routes

Mert Türedioğlu
Middle East Technical University
mertturedioglu@gmail.com

Abstract

The climbing route difficulty assessment is a challenging problem. In this research, the difficulty of the climbing route will be evaluated by computational methods. Climbing routes will be chosen from the routes of the MoonBoard climbing board. MoonBoard is a popular climbing board with global prevalence. Among climbing boards, MoonBoard is the pioneer. The use of boards in climbing training is becoming more common day by day. As climbing has become popular and has become an Olympic sport, new training boards have emerged. Hundreds of new routes are uploaded to the applications of these training boards by people from the climbing community every day. With the increase in the number of these boards and the number of routes, the need to automate the route difficulty assessment has arisen. In recent years, research in this area has begun to emerge. In this research, collection of MoonBoard routes and evaluation of route difficulties will be performed in Python. Random Forest algorithm, one of the machine learning techniques, is used.

Keywords

Climbing, MoonBoard, Training Board, Difficulty Assessment, Machine Learning, Random Forest

1 Introduction

At the very beginning, indoor climbing emerged as a training for outdoor climbing. However, at the point reached today, indoor climbing has ceased to be a training for outdoor climbing and has become a sports discipline on its own. It even gained Olympic sport status with the Tokyo Olympics in 2020.

The routes selected in the research will be chosen from the MoonBoard routes, which is a climbing training board. Board training is the most efficient type of training in climbing, with its simple structure and oriented towards the training needs of a climber. The grid structure of the training board provides simplicity. The ledges of the training board held by the climbers are called the hold. Holds at specified distances and angles are placed on training boards around the world in the same way. This simple structure of the training board makes the difficult problem of routing difficulty assessment workable.

The fact that these training boards have a worldwide standard and have the same structure everywhere makes the output of this research generalizable. This research is intended to provide a formal aspect to the climbing route difficulty assessment. It is expected that the method in this research, which is made specifically for MoonBoard routes, can be applied to other climbing boards with some modifications.

1.1 Research Question

The following questions are planned to be answered in this research.

- What are the factors that determine the climbing route difficulty?
- What are the features that determine the difficulty of a climbing holds?
- What is the relationship between climbing holds and climbing route difficulty?

1.2 Methods and Results

As detailed below, IJCAI has prepared and made available a Random Forest algorithm was used in this research. There are multiple reasons for this choice. One of the reasons is that the Random Forest algorithm can be explained and interpreted. The other reason is that the random forest algorithm is suitable for multi-classification problems. The last reason is that it was expected to benchmark to other machine learning algorithms. Unfortunately, the expected result could not be obtained in the research. The overall accuracy of the model is 27%. The reason for this low accuracy rate will be explained in the following sections.

2 Related Work

The climbing route difficulty assessment is just beginning to appear in the literature. This task was studied with machine learning algorithms. Some of the research is only for the evaluation of climbing route difficulty, the other part is for creating new climbing routes. Due to its simple nature, the majority of research is directed towards training board routes. As in this research, it mainly works on MoonBoard routes. The reason for this is that MoonBoard is the first of the worldwide training boards.

Duh et al approached the route difficulty assessment and generation using neural networks in their research. They have established 3 separate neural network models: BetaMove, GradeNet, DeepRouteSet. BetaMove is developing a solution to the given route problem. GradeNet makes estimations of climbing route difficulty taking into account the beta provided by BetaMove. DeepRoute set produces climbing routes. They preferred the Recurrent Neural Network algorithm because of the sequential nature of the climbing routes. GradeNet achieved an accuracy rate of 64.3% in the training set and 46.7% in the test set. Climbing routes are labeled as discrete. However, the routes in the same difficulty label differ in some difficulty. Therefore, Duh et al also reported an accuracy rate of +-1 in their research. GradeNet's +-1 accuracy rate was reported as 91.3% on the training set and 84.8% on the test set. The +-1 accuracy rate is a good test indicator in terms of capturing the small differences in the difficulty of the climbing routes according to the differences in the heights of the climbers. The quality evaluation of the routes generated by DeepRouteSet was made through a survey. Routes were evaluated under 4 headings in the survey: unnecessary holds, strange moves/fluency of moves, reasonable route, route quality. Routes produced by DeepRouteSet are presented for evaluation together with MoonBoard routes. DeepRouteSet was found to be more successful than MoonBoard routes in 4 headings. Researchers believe that the success of GradeNet and DeepRouteSet lies in the success of BetaMove. It has been successful in producing and evaluating routes suitable for climbing with betas offered by BetaMove to routes.

Stapel preferred a heuristic approach to his research. In the research, it is aimed to generate new climbing routes. A heuristic method was followed. To achieve this purpose, climbing holds and climbing moves are classified. The effects of these classifications on the difficulty of climbing routes were evaluated. These classifications are also expected to assist future machine learning research. As in this research, the scope of the research has been reduced to MoonBoard routes. It is assumed that there is a relationship between the climbing hold difficulty and the route grade difficulty. Climbing holds were graded with the help of a questionnaire. The climbers evaluated all the holds on the MoonBoard and suggested a number. With the help of these difficulty suggestions and the climbing hold type, the next possible holds to the climb route hold sequence are selected by a Greedy algorithm. In addition, the distance and rotational angle of the next hold are presented to the model for this Greedy algorithm to consider. The evaluation of the climbing routes produced by this algorithm was made by the climbers with a questionnaire. Climbing routes were evaluated on the basis of grade, route flow and enjoyment of the climber. The flow of the generated routes was found to be worse than the flow of the MoonBoard routes. The reason for this is thought to be due to the following deficiency of the algorithm. The algorithm does not consider the climber's last position when adding a new hold to the hold sequence. This disrupts the flow of the route. Another shortcoming is that the footholds that the route may need are not taken into account. This is considered to be a more difficult task than the other deficiency. In order to produce a quality route, the algorithm must somehow take into account the climber's move sequence information.

3 Dataset

The routes on which difficulty estimations will be made in the research were chosen among the MoonBoard routes. MoonBoard has multiple training boards. The routes used in this research were selected from the MoonBoard 2016 board. This MoonBoard is the first training board. Currently, there are more than fifty thousand routes made for this board. Climbing routes are designed by climbers and can be shared with other climbers via the MoonBoard app. From the standardized structure of the training board, any climber can access this route information and climb these routes.

Access to MoonBoard routes is possible in two ways. Information on climbing routes can be accessed via MoonBoard's phone application or website. The routes to be used in this research were scrapped from the MoonBoard website. By means of a bot implemented in Python, the hold information and difficulty level of the climbing routes were obtained. The coding of this bot was supported by Python's Selenium package. The resulting routes have the Benchmark tag. The benchmark tag was given to the routes that were evaluated and appreciated by the MoonBoard team. Due to the challenging and time-consuming aspect of route quality and difficulty assessment, only 450 of the fifty thousand routes mentioned have benchmark tags at the moment. Other routes used in this research were filtered from the routes shared by the Enzo SR account, which can be accessed via GitHub.

In addition to benchmarked routes, the need for new route information resulted from the low number of benchmark routes. The small number of routes and the large number of categories that need to be classified greatly reduced the accuracy of the algorithm. It is possible to create routes in approximately 13 different difficulty levels on the MoonBoard training board. This range starts from 6B+ difficulty and goes up to 8B+. However, there are very few routes with benchmark tags at 8A+ and above difficulty levels. To be specific, there are currently 2 routes in 8A+ difficulty, 1 in 8B difficulty, and no routes in 8B+. These difficulty levels were not used in the development and testing of the model. It has not been possible to separate such a small number of routes into training and test sets.
The distribution of the routes that make up the data set into difficulty categories is unbalanced. The distribution of routes with benchmark tags is as follows.

- 6B+ 66 routes
- 6C | 63 routes
- 6C+ | 52 routes
- 7A 59 routes
- 7A+ - 69 routes
- 7B 49 routes
- 7B+ 23 routes
- 7C 33 routes
- 7C+ 26 routes
- 8A 11 routes

In order to increase the number of routes for categories with high difficulty levels, support was obtained from the data set mentioned above. However, there are 2 possible problems with these newly added routes. Due to the fact that they are not benchmarked, the difficulty rating labels of the routes may be incorrect, and these routes may not meet the quality expectations stated in the other research above. To overcome these 2 problems, a filter was applied to the routes. Although this filter was able to solve one of the 2 problems a little, it was not enough to solve the other problem. This filter eliminated newly added routes based on the number of times they were climbed by other climbers. It was expected that the routes climbed by a certain number of climbers would be quality routes. The quality expectation here is in line with the expectations in the titles mentioned above: flow, redundant holds. This filter provided routes that have been climbed by more than 100 climbers. As a result of this filter, the total number of routes available has been doubled. However, approximately one thousand routes were not sufficient to classify the algorithm in 10 different categories. It is essential for future research to evaluate the quality of the routes taken from the mentioned GitHub account with filters other than the mentioned filter and to increase the data set.

3.1 Feature Set

Various processes have been made to provide features to the routes that make up the dataset. The main point of these route features is the holds that make up the routes and the features derived from them. forms.

3.2 Hold Features

Holds have been physically studied and classified. These classification headings are as follows: hold's rotational angle, parallel to the ground and outwards force application capacity (z score) and color. Multiple hold features are derived from this rotational angle information and transferred to the route features. The handle color may seem irrelevant, but the 3 different colors in the MoonBoard 2016 layout reflect some features. Yellow handles are very small in size, black handles are relatively large and slopier, and whites are of various character. Hold's z score is used as a factor in creating route features to give an idea of possible climbing move difficulty. Multiple hold features are derived from the rotational angle feature obtained by physical examination. These categories are categorical features of which direction the climber can apply force to the handle: only sideways, only upwards, downwards. On very overhanging training boards such as the MoonBoard, the direction in which force can be applied to the handle provides information in terms of route difficulty estimation.

3.3 Route Features

Some of the route features are derived from the hold features, some of them are derived from the physical location of the holds in the route, and some of them are derived from the combination of the two. MoonBoard has a grid structure. The distance between the handles is fixed. The route information scraped from the MoonBoard site contains information about which holds are included in the route. The 'number of holds' and 'mean distance of holds' features are derived from this information.

There are multiple route features related to the colors of the holds available in the route. 'Black count', 'white count', 'yellow count', ''black percent', 'white percent', 'yellow percent', ''b_w percent'.

There are also multiple route features associated with the rotational angle of the holds in the route: 'downwards_count', 'only_upwards_count', 'only_sideways_count', 'downwards_percent', 'only_upwards_percent'.

Finally, two route features are derived from the difficulty of possible moves on the route. Some of the hold features were used to determine the difficulty of these moves. These hold features are the hold color, z_score, and rotational angle. In determining these move difficulties, the information about the number of holds in the route was also used. These two possible move difficulty features are first move and last move difficulty. The names of these features are 'first move difficulty' and 'last move difficulty'. The factors that determine these difficulties are associated with the average of the distances of the 3 holds included in these moves, how many of the 3 holds are yellow, their rotational angles and z-scores. The equation of this ‘difficulty of first move’ feature and the explanations of the variables are as follows:

FD = first move difficulty score
md = mean distance between three holds
fy = is first hold yellow (0 or 1)
yc = yellow count of three holds (0,1 or 2)
tu = is third hold’s rotational angle upwards (0 or 1)
zc = total z_score of three holds
ly = is the last hold yellow (0 or 1)
n = the number of holds in the route

$$FD = md^{(1+fy+yc+tu)/\ln(2.72+zc)}$$

The equation of this ‘difficulty of last move’ feature and the explanations of the variables are as follows:

$$LD = md^{(1+ly+yc)/\ln(n+zc)}$$

4 Model

Random forest algorithm was preferred in the research. Its explainability and suitability for multi-class categorization played a role in this preference.

The class that the model is expected to predict is the difficulty grade. Two different grade systems are used on MoonBoard routes. One is the Fontainebleau bouldering difficulty system from France, and the other is the V-scale Hueco Tanks system from America. It is possible to convert between the two. Fontainebleau difficulty rating system is further segmented than V-scale. The range that the model in the research tries to predict starts from 6B+ and goes up to 8A+.

SkLearn's library is used in the Python implementation of the random forest algorithm. Due to the small size of the data set, a 6 to 4 training test split was made. Due to the nature of the Random Forest algorithm, there was no need to standardize the features. Target categories in string format are encoded numerically. Various experiments have been carried out in the tuning of hyperparameters. N_estimators and max_depth are kept small so that the model does not overfit the training data. In order to enlarge the sample size after the training test split, the training data was oversampled with the help of the SMOTE package written by Brownlee. It is aimed to increase the number of samples on routes with high difficulty levels.

5 Results and Discussion

5.1 Results

Confusion table showing the result of the model in the training set can be seen below.

	precision	recall	f1-score	support
0	0.91	0.98	0.95	212
1	0.96	0.91	0.93	101
2	0.94	0.90	0.92	101
3	0.95	0.92	0.93	115
4	0.93	0.90	0.92	92
5	0.86	0.87	0.86	75
6	0.93	0.93	0.93	75
7	0.83	0.83	0.83	59
8	0.94	0.94	0.94	50
9	1.00	0.89	0.94	18
accuracy			0.92	898
macro avg	0.92	0.91	0.92	898
weighted avg	0.92	0.92	0.92	898

Table 1: Training Set Confusion Table

Confusion table showing the result of the model in the test set can be seen below

	precision	recall	f1-score	support
0	0.59	0.61	0.60	124
1	0.17	0.12	0.14	57
2	0.04	0.06	0.05	35
3	0.17	0.11	0.13	53
4	0.20	0.17	0.19	46
5	0.06	0.03	0.04	31
6	0.05	0.08	0.06	12
7	0.13	0.25	0.17	16
8	0.00	0.00	0.00	10
9	0.09	0.17	0.12	6
accuracy			0.27	390
macro avg	0.15	0.16	0.15	390
weighted avg	0.27	0.27	0.27	390

Table 2: Test Set Confusion Table

5.2 Discussion

The model performed poorly on the test data. It is thought that one of the main reasons behind this is the insufficient size of the training and test sets. Oversampling was done in order to increase the size of the training set and to balance the difficulty level categories. However, it was not possible to train the model well with duplicate samples. Since oversampling is not possible in the test data, the training test split was made at a rate of 60% to 40%. However, even these rates did not make it possible to achieve a balanced distribution in terms of difficulty levels. The routes in easy grades were more numerous than the difficult ones. It was not possible to train the model well in difficult route grades.

During the experiments, the prediction success of the route features was checked. The most successful ones have seen features such as 'mean distance', which shows the average of the distances between their holds in the route, and the features that show the difficulty of the first move and the difficulty of the last move. These features have been shown to be promising in determining route difficulty. In future studies, these features should be considered even better and tried to be effective in increasing the accuracy rate.

References

[Stapel, 2020] Stapel F. T. A. A heuristic approach to indoor rock climbing route generation. BS thesis. University of Twente, 2020.

[Duh and Chang, 2021] Duh, Yi-Shiou, and Ray Chang "Recurrent neural network for moonboard climbing route classification and generation." arXiv preprint arXiv:2102.01788, 2021.

[Brownlee, J., 2021] Jason Brownlee Smote for imbalanced classification with python. Machine Learning Mastery. Retrieved June 21, 2022, from https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/, 2021, March 16.