Strategy analysis and rank prediction in 2000m rowing tournaments
C.I. Kettler

Abstract

Data of 20 tournaments from the period 2013 until 2016 is used in the analysis of strategies as performed in 2000m rowing races and their relation to the outcome of the race, where strategy is defined as being the different approaches by which effort can be divided throughout the race. Strategies as considered in this thesis consist of four continuous normalized gradients of each 500m part of the race, where the average stroke rate of the regarded boat in this race is used as normalization factor and the gradient is a measure of difference between the stroke rate at first measured point and the last measured point of the 500m. This is the first time such a large amount of data is used in an exploratory research of strategies, generating new insights on the relation between strategies and other race dependent features. The data shows that more than one strategy is applied in 2000m rowing races, either having a U-shape or an L-shape in the stroke rate over the distance. One version of the U-shaped strategies proves to be the strategy that occurs most often and not the often proposed even strategy, where the velocity is kept as constant as possible. The four gradients do not seem to have a strong relation and therefore they are regarded separately to see the influence of each feature on the strategy shape in each of the four race parts. The features that are known prior to the race, and therefore can be used in the planning of the strategy, and have an influence on the strategy are boat size, round and the number of races after the current race on the same day. Using these features a prediction is made of the strategy, testing the performance of K-Nearest Neighbors, Random Forest and Multi-Layer Perceptron Regression. This shows that the selected features are most influential on the first and last 500m of the race, and shows that crews probably take these features into account when deciding for a strategy. Since the main goal of a crew is to win the race, i.e. reach one of the highest three ranks, the relation between strategy and rank is researched by their correlation, comparison of strategies used within the different ranks and the value of strategies in the prediction of ranks. The first two are performed on the separate gradients while the last represents the influence of the combination of the four gradients. When comparing the strategies within the different ranks, the significant differences are mostly found in the middle part of the race. Low correlation values are found between the strategy and rank, which increase when looking per feature configuration, which is a combination of values of the boat size, rounds and number of races after the current race. A thorough research of Ordinal Logistic Regression, K-Nearest Neighbor Regression, Random Forest Regression and Multi-Layer Perceptron Regression is performed to generate a good prediction model. There are only three cases where the strategy has some predictive value, but in all other cases the average rank is a closer match to the achieved rank then the rank predicted by the models. Therefore the strategy as defined in this thesis proves not to be influential on the rank a team achieves.