Predicting the results of Hungarian football matches with probabilistic and machine learning models

Pintér József and Ragács Attila

Predicting the results of football matches has been a popular task since the middle of the 20th century. The first real breakthrough in this area was achieved by an English mathematician Maher, who used Poisson distribution to model the defending and attacking ability of teams and predict the amount of goals scored. Many attempts were made to fix the original model's errors, but the one that stands out is a model made by Dixon and Coles. In this paper we investigate how the above models perform on the data of the OTP Bank Liga. Furthermore we attempt to apply our own ideas to increase the performance of these models.

Nowadays it is common to tackle this task with machine learning methods due to the increased amount of sports data. We have the data of last three seasons of the Hungarian league, which we acquired using webscraping by Python. Following the necessary data cleansing and data exploration we performed the fitting and evaluation of the probabilistic and machine learning models

We measured the performance of the models by calculating the rate of correct final outcomes (home win, draw, away win). Comparing the models we found that the probabilistic models outperform the machine learning models.