

Analyzing Sports Training Data with Machine Learning Techniques

Rehana Mahfuz, Zeinab Mourad, Aly El Gamal
Department of Electrical and Computer Engineering, Purdue University

ABSTRACT

In the sports industry, there has not been enough effort in analyzing the personalized monitoring data of athletes collected during training sessions. This research is an attempt to find meaningful patterns in the Purdue Women's Soccer training data that could help the coach design more efficient training sessions. We are specifically interested in studying this problem as an unsupervised learning problem. Our initial attempt is to cluster the players as well as drills into groups using k-means, c-means and spectral clustering algorithms, combined with feature transformation and reduction steps. These basic algorithms serve as a benchmark to measure performance improvements when suggesting more advanced methods. In spectral clustering, the gaussian kernel similarity function was used, in which sigma and number of clusters were matched using the eigengap method. The Pearson correlation was used to eliminate highly correlated features, and Principal Components Analysis was used to find mutually orthogonal axes with maximum variance. Three features were eliminated with negligible loss in accuracy. Satisfactorily consistent clusters were identified, where by "consistent", we mean the clustering results that we get through multiple algorithms. The next step will be to give the clusters meaningful labels with expert help (in this case, the soccer team's coach). It is hoped that this is a good start in sports performance analysis.

KEYWORDS

Purdue Women's Soccer, Data Science, Unsupervised Learning