HONG KONG SPORTS INSTITUTE
香港體育學院

RESEARCH HIGHLIGHTS

An Application of Machine Learning to Notational Analysis

Angus T.K. Lam, Danny P.K. Chu, P.M. Cheung
Sport Biomechanics and Technology Centre, Hong Kong Sports Institute

MARCH 2019

Introduction

Data science has been a hot topic in recent years. This would be a useful tool to understand sports as well. However, the communication between sports coaches and data scientists has obstructed the development. The relevance of data science has not been communicated well to elite sports coaches, and likewise, sports coaches may not be able to transfer their valuable knowledge of a sport in words understandable to a data scientist. Machine learning (ML) was first proposed by Arthur Samuel in 1959, and is a field of study focused on devising methods to give computers the ability to learn without being explicitly programmed [1]. Data are of fundamental importance to a data scientist, and in this study, a new data-collecting application was proposed to enable formation of a badminton shuttlecock-path dataset, which was then analysed by a basic ML algorithm. The 2D position (i.e., distance and angle) of a returning badminton shuttlecock was predicted by linear regression via a supervised learning classification. In this way, an attempt was made to machine-learn the playing style of top badminton players via a bespoke data-collecting application.

Methodology

A smartphone application was developed to collect badminton shuttlecock-path data from video recordings. The application recorded the fingertip position of a player as it moved on the screen (Figure 1a). As well as providing summary data for each game, in terms of outcome, the playing technique, shuttlecock-hitting region, and active or passive playing-style of each player were also recorded (Figure 1b). In addition, a preliminary kinematic-data summary (Figure 1c) was instantaneously displayed to the user.

A men's singles badminton match involving two of the world's top-10 ranked players was randomly selected from the Internet. The match consisted of three games with the results of 21-16, 13-21, and 21-19, respectively. Altogether, data on 500 effective shuttlecock returns were collected, and characterised in terms of 11 features (Figure 2a) and two labels (Figure 2b). The features were game, score, score difference ('ScD'), number of turns ('NoT'), hit position X ('HPX'), hit position Y ('HPY'), body movement ('BM'), opponent body movement ('OBM'), shuttlecock approach distance ('SAD'), shuttlecock approach speed ('SAS'), and shuttlecock approach angle ('SAA'). The labels were shuttlecock return-distance ('SRD') and shuttlecock return-angle ('SRA'). Data were manipulated with a basic machine learning algorithm [2], a Matlab® script for linear regression with multiple variables under supervised learning classification. A recursive data-training session was conducted to ensure an optimal match of each feature with its respective label, which involved allocating a weight to each feature.

Results and Discussion

The playing style of a player was predicted in terms of SRD and SRA, by adjusting the feature weightings as a player's style varied during the game. The weightings of the corresponding features are shown in Table 1. Game, BM, OBM, and SAD were found to be the major contributors to SRD, while BM, OBM, SAD, SAS, and SAA most prominently affected the SRA. Thus, we may determine a badminton player's style by investigating the weighting of these parameters.

Table 1. The weights to the corresponding features

	Game	Score	ScD	NoT	HPX	HPY	BM	OBM	SAD	SAS	SAA
SRD	***-0.24***	-0.03	-0.05	0.02	0.00	0.00	***-0.33***	***0.14***	***0.90***	-0.05	0.01
SRA	-0.02	0.01	-0.05	0.04	0.04	0.02	***0.31***	***-0.15***	***-0.26***	***-0.32***	***0.71***

Conclusion and Recommendation

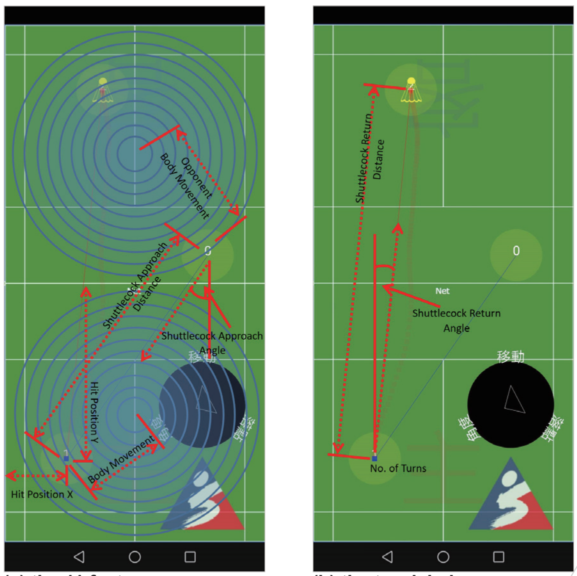
A new method for collecting ML-suitable data was demonstrated. Both breakdown and summary data were recorded in a rapid, simple, and convenient way, and a kinematic data summary was made immediately available to the user. The method involved use of a linear regression under supervised learning classification to predict badminton shuttlecock-paths in terms of returning distance and angle. The major contributing parameters could be easily determined by simply studying the weighting.

A database stores the competition data for later analysis, and this will enable regular data collection from players. As a player becomes more experienced and powerful over time, we expect that the weighting for recent games will be more important. The output result shown in Table 1 describes the playing style of a player as determined by this ML approach. This provides additional information for game analysis, and will therefore assist players to formulate their tactics for beating their next opponent, by ML analysis of this next opponent's playing style.

(a) fingertip positioning information (b) summary data for each score

(c) kinematic data summary (d) user interface

Figure 1. The user interface of the smartphone application



(a) the 11 features (b) the two labels

Figure 2. Graphical presentation of features and labels of smartphone application

Reference

1. *Samuel AL (1967). Some studies in machine learning using the game of checkers. II - Recent Progress. IBM Journal, November, 1967.*
2. *Andrew NG (2018). Machine learning. The Leland Stanford Junior University. https://www.coursera.org/learn/machine-learning*