Classification of top male tennis players

Ales Filipcic¹, Andrej Panjan² & Nejc Sarabon²³

¹University of Ljubljana, Faculty of Sport, Ljubljana, Slovenia
²S2P Ltd., Laboratory for Motor Control and Motor Behavior, Ljubljana, Slovenia
³University of Primorska, Science and Research Center, Institute for Kinesiology Research, Koper, Slovenia

Abstract

The main objective of this study was to define different quality groups of tennis players based on their position on the ATP ranking list. Ranking data on the top 300 players from 1990 to 2008 were used to conduct the study. The classification into quality groups was performed using six machine learning algorithms suitting such a task. Quality groups were formed better for each year separately than for all years together. Three clustering algorithms (k-means with a Euclidean metric, MDBC with a Euclidean metric, and Xmeans) were equally successful in the classification according to the criteria function. All three algorithms also created very similar quality groups. They are thus equally suitable for classifying tennis players into quality groups based on their ranking scores. Changes in the ranking system (in the year 2000) were also reflected in the differences in classification success between the two periods (before and after 2000). The boundaries between the quality groups were more stable for the period after 2000, and less stable for the period before 2000.

KEYWORDS: SPORT, CLASSIFICATION, RANKING, QUALITY GROUP, MACHINE LEARNING

Introduction

One of the most objective measures of an athlete’s competitive success is the annual ranking list used in many individual sports (golf, car racing, alpine skiing, tennis etc.). The Association of Tennis Professionals (ATP) was formed in 1972 and the first world ranking list was published in August 1973. The ATP publishes weekly rankings of professional players, the ATP Entry Ranking, a 52-week rolling ranking and, up until 2009, the ATP Race, a-year-to-date ranking. Every player starts collecting points from the start of the season. Tennis players collect ATP points at different tournament levels: Grand Slam (4), ATP Tour Finals (1), ATP World Tour Masters 1000 (9), the ATP World Tour 500 and 250 series, ATP Challengers, and ITF Futures tournaments. A player’s position on the ATP ranking list is defined by the number of ATP points collected and allows entry to and competing on different levels of tournaments.

A player’s competitive success is determined by three groups of indirect factors (Crespo & Miley, 1998):
1. player (playing standard, tactical understanding, technical competence, physiological development, mental characteristics, experience, game-style and training level, level of tournament);

2. opponent (same as for the player, but also ball trajectory, shot selection, position of the player, tactical intentions, and an opponent’s tactical strengths and weaknesses); and

3. environment (court surfaces, weather conditions, other environmental factors: spectators, umpires, time, psychological considerations etc.).

In recent years, we have witnessed a rapid increase in the volume of data in digital form. It is impossible to gain useful information from such a vast data set. We therefore obviously need tools that can effectively search for interesting information in large databases. Machine learning is an artificial intelligence field which deals with discovering knowledge in data. It is becoming an important tool for transforming such data into useful information. The growing expansion of machine learning is also reflected in a rising number of commercial systems within the industrial, medical, economic, banking, etc. sectors. The main principle of machine learning is the automatic modelling of data. Learned models attempt to interpret the data from which models were constructed. They can assist in making decisions when it comes to studying the modelled process in the future (predictions, diagnosis, control, verification, simulations etc.).

Many practical problems entail a need to classify given data into groups. Groups are formed based on certain criteria. In fact, this is one of the most primitive activities of human beings. In order to learn about a new phenomenon, people always try to identify descriptive features and further compare these features with those of known phenomena, based on their similarity or dissimilarity. Naturally, people have limited processing power and memory and are therefore unable to effectively classify large databases. Fortunately, we can use advanced machine learning tools that are suited to such tasks. Clustering is a process of identifying a finite and discrete set of ‘natural’ data structures from a finite, unlabeled set of data (Xu & Wunsch, 2009). Derived data structures are called clusters. In general, clustering techniques are classified as partitional clustering and hierarchical clustering, based on the properties of generated clusters. Hierarchical clustering groups data point into tree-like, nested structure partitions, while partitional clustering directly divides data points into a pre-specified number of clusters without a hierarchical structure.

In this study, the main objective was to define different quality groups of tennis players based on their position on the ATP ranking list. We were interested in discovering the borders between the quality groups, and the common characteristics of players in each group. This was performed for each year separately, and for all years together.

**Methods**

In the framework of our project the data for rankings, players, tournaments and matches were collected from the ATP webpage (ATP World Tour, 2009). All individual rankings for the years 1973 through 2008 were collected for the best 300 players. Rankings with ATP points information were available from 1990 on. We selected tournaments from 1968 to 2008, including Grand Slams, the ATP World Tour Masters 1000, the ATP World Tour 500, and the ATP World Tour 250. All matches collected were from 1991 to 2008 for all ATP tournaments previously mentioned. All of the collected data was stored in a specially designed database running on a MySQL 5.1 community database server. A custom-made application was used to
store the data in the database which was designed in a way that allows quick and easy queries across players, rankings, tournaments and/or matches.

For the purpose of this article, we used the following variables:

- ranking variables: points and position;
- match variable: round; and
- tournament variables: type of tournament and surface.

The quality groups were determined for every year separately and for all years together using different machine learning clustering algorithms based on rankings (points). We chose points over position because that reflects a real quality difference between players, whereas position only gives the order. There are many possible clustering algorithms (Xu & Wunsch, 2009). Since we had already anticipated having five desired clusters, we used partitional clustering algorithms: k-means with a Euclidean metric (KME) (Forgy, 1965), k-means with a Manhattan metric (KMM), MDBC with a Euclidean metric (Law et al., 2004), Xmeans (Pelleg & Moore, 2000), EM (McLachlan & Krishnan, 2008) and FarthestFirst (Hochbaum & Shmoys, 1985).

The metric, or distance function, is a function which defines the distance between elements of a set. Both Euclidean and Manhattan metrics are well known and frequently used. The K-means algorithm is a very simple method for grouping data into $n$ clusters, and tries to minimize the within-cluster sum of squares. It starts with $n$ randomly positioned clusters. All instances are then assigned to the closest cluster according to the chosen metric. Next, the centroid (mean) of the instances in each cluster is calculated. The whole process is repeated with new cluster centers. Iteration continues until the same instances are assigned to each cluster in consecutive rounds. But K-means does not guarantee that its solution is a global minimum. To increase the chance of finding a global minimum, the algorithm is run many times and the best solution is chosen (the one with the smallest total squared distance). The EM (expectation-maximization) algorithm is an iterative method which alternates between performing an expectation (E) step – which computes the expectation of the log-likelihood evaluated using the current estimate for the latent variables – and a maximization (M) step that computes parameters, maximizing the expected log-likelihood found in the E step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step. The algorithm iterates until the difference between the successive log-likelihood values is below a pre-defined, small value. The EM algorithm also does not guarantee that its solution is a global minimum. The same procedure employed with the k-means algorithm can be used to overcome this problem.

The quality of the clusters generated by each algorithm was estimated with the criteria function (CF) that most effectively separates the types of tournaments according to quality groups. First, the overall performance of the players in each quality group (QGP) was calculated as the average of the highest round played in tournaments of a specific type. Next, the CF was defined as the average of QGP differences between consecutive quality groups. Finally, the algorithm with the highest CF most effectively separates players into quality groups. In addition, differences in classification success between the two periods (before and after the year 2000) were examined due to changes in the ranking system.

**Results**

According to our criteria function, quality groups were formed better for each year separately than for all years together (Table 1). The best clustering algorithms for each year separately
were KME, MDBC, and Xmeans, whereas for all years together the best clustering algorithms were MDBC and Xmeans. Differences based on the criteria function between KME, MDBC, and Xmeans were minimal, thus the quality groups formed with these algorithms were very similar. The CF standard deviation was considerably smaller for all years together than for each year separately, indicating that the quality groups were formed better for each year separately. All algorithms (except KMM) had greater CF values for the period after 2000 compared to the CF values for the period before 2000 for each year separately. Thus, quality groups were formed better for the years after 2000. In contrast, there were no pronounced differences in the algorithms’ success within the two periods (before and after 2000) for all years together, except for the FarthestFirst algorithm.

Table 1. Evaluation of the derived quality groups with the criteria function (CF) for each year separately and all years together

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>CF</th>
<th>CF SD</th>
<th>CF after 2000</th>
<th>CF before 2000</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>YEARS ALTOGETHER</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KME</td>
<td>0.76</td>
<td>0.15</td>
<td>0.78</td>
<td>0.75</td>
</tr>
<tr>
<td>MDBC</td>
<td>0.84</td>
<td>0.16</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td>XMeans</td>
<td>0.83</td>
<td>0.16</td>
<td>0.83</td>
<td>0.83</td>
</tr>
<tr>
<td>FarthestFirst</td>
<td>0.76</td>
<td>0.40</td>
<td>0.84</td>
<td>0.69</td>
</tr>
<tr>
<td>KMM</td>
<td>0.65</td>
<td>0.09</td>
<td>0.64</td>
<td>0.67</td>
</tr>
<tr>
<td>EM</td>
<td>0.47</td>
<td>0.09</td>
<td>0.49</td>
<td>0.46</td>
</tr>
<tr>
<td><strong>YEARS SEPARATELY</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KME</td>
<td>0.84</td>
<td>0.25</td>
<td>0.91</td>
<td>0.77</td>
</tr>
<tr>
<td>MDBC</td>
<td>0.84</td>
<td>0.25</td>
<td>0.91</td>
<td>0.76</td>
</tr>
<tr>
<td>XMeans</td>
<td>0.84</td>
<td>0.25</td>
<td>0.91</td>
<td>0.77</td>
</tr>
<tr>
<td>FarthestFirst</td>
<td>0.78</td>
<td>0.39</td>
<td>0.82</td>
<td>0.75</td>
</tr>
<tr>
<td>KMM</td>
<td>0.54</td>
<td>0.09</td>
<td>0.54</td>
<td>0.55</td>
</tr>
<tr>
<td>EM</td>
<td>0.50</td>
<td>0.23</td>
<td>0.49</td>
<td>0.43</td>
</tr>
</tbody>
</table>

The average quality group profiles were similar for the KME, MDBC, and Xmeans algorithms (Figure 1), as were the differences between the quality groups for each year separately. On the other hand, the average quality group profiles across years of the other three algorithms (EM, KMM, and FarthestFirst) differed significantly.
Figure 1. Average quality groups profile for each algorithm for each year separately

The output of each algorithm was a set of quality groups divided by the boundaries between them. The boundaries for KME are plotted in Figure 2. The first quality group (the best players) contained the lowest number of players, while the fifth quality group contained the most players among all quality groups. The number of players in the other quality groups increased from the first to the fifth quality groups. Fluctuations of the boundaries between the quality groups were greater before the year 2000 and considerably smaller after that (Figure 2).

Figure 2. Profile of the quality groups produced by the KME algorithm

Discussion

When comparing the performance of clustering algorithms for each year separately and for all the years together, there was an evident difference between these two approaches. All algorithms, except the KMM, classified players into quality groups better for each year separately (Table 1). Such results were expected because quality groups can be more effectively adjusted for each year separately than they are independent from other years in this case. The adjustments are made as an adaptation of one quality group to adjacent ones. Hence, a boundary between two quality groups affects the adjacent boundaries so that all quality......
groups formed are the most compact regarding the selected metric function. In contrast, the classification into quality groups based on all years together yielded the same quality groups for all years. The boundaries between quality groups for all years together were very similar to those we obtained as an average of all of the boundaries between the quality groups for each year separately. The differences between boundaries were in ±3 ranking positions. Regarding the classification for each year separately, three clustering algorithms (KME, MDDB, and Xmeans) were equally successful according to the CF. All three algorithms also created very similar quality groups (Figure 1), which was expected because they are based on the same foundations and differ in the selection of the initial clusters. Thus, they are equally suitable for classifying tennis players into quality groups based on ranking scores.

The differences in classification success between the two periods (before and after 2000) were of interest due to changes in the ranking system designed to force top players to play in all the grand slam events and “Super 9” tournaments, or else suffer in the rankings (Mallett, 2009). The changes described here were also detected in our study. The boundaries between the quality groups were more stable for the period after 2000 and less stable for the period before 2000 (Figure 2). Moreover, the CF values were greater after 2000 than before, indicating the better formation of quality groups after 2000. We can therefore conclude that the ranking system changes in 2000 led to a more accurate ranking of tennis players. Moreover, no differences in the success of the clustering algorithms were detected between the two periods for each year separately, which means no algorithm is more suitable for one period than another.

The number of players in each quality group grows with decreases in quality. This behavior is completely natural and can be observed in all other sports. For example, only a few tennis players are able to perform on the highest level, while the number of players grows exponentially (the exponent is typically > 0) with decreasing quality. This behavior depends on the structure of tournaments (the exponential relation between the number of tournaments and prizes) that consist of a few tournaments with big prizes, thus all highly ranked players can participate in these tournaments. If highly ranked players had not been able to participate in all big prize tournaments, the relationship between the number of players and the quality would have moved towards a linear relation. Among all algorithms, it was the EM that had the most linear profile, while the FarthestFirst had an exponent with the highest exponential. All other algorithms were in between these two. The best three clustering algorithms had a similar profile, indicating they are equally suitable for classifying tennis players into quality groups.

**Conclusion**

The composition of individual quality groups is associated with a player’s performance at a specific level of tournaments (Grand Slams, ATP World Tour Masters 1000, ATP World Tour 500 and 250, Challengers, and Futures tournaments) and indirectly with their position on the ATP ranking list. Higher levels of tournaments or rounds of a tournament mean bigger prize money, more ATP points, and more competent opponents. The present findings suggest that each quality group of players is defined by specific tactical, technical, mental, and fitness competencies. Therefore, players must improve those competencies in order to be able to play on the level of the next quality group.

The performance of players can also be defined with an analysis of performance indicators. In the future, we want to determine the following: differences in performance indicators through different time periods, and differences between quality groups in performance indicators in
specific game situations. In this study, we classified players into five quality groups where the number of groups was chosen based on experts’ knowledge. However, we are aware that some other classification approaches could result in different conclusions.

Acknowledgement

The authors thank the International Tennis Federation for funding this project.

References


