Automatic Tennis Analysis with the use of Machine Learning and Multi-Objective Optimization

Miha Mlakar
Jozef Stefan Institute
1000-SI Ljubljana, Slovenia
Miha.mlakar@ijs.si

Mitja Luštrek
Jozef Stefan Institute
1000-SI Ljubljana, Slovenia
mitja.lustrek@ijs.si

ABSTRACT
Wearable devices for monitoring players' movements are heavily used in many sports. However, the existing commercial and research sports wearables are either not tennis-specific, or are worn on the wrist or in the racquet and thus offer too limited information. We therefore added tennis-specific information to a leading commercial device. Our solution is two-fold. Firstly, we developed a model for classifying shot types into forehand, backhand and serve. Secondly, we designed an algorithm based on multi-objective optimization to distinguish active play from the time in-between points. By combining both parts with the general movement information already provided by the device, we get a comprehensive set of metrics that are used by professional tennis players and coaches to objectively measure a player's performance and enable in-depth tactical analysis.

Categories and Subject Descriptors
I.2.6. Artificial Intelligence: Training

General Terms
Algorithms, Measurement, Experimentation

Keywords
Tennis; Wearable analytics; Shot detection; Optimization;

1. INTRODUCTION
The use of wearable sensors in sport is growing fast and can already be considered essential for success in some disciplines. In tennis the analytics started with computer vision and sensors for measuring shots. However, both of these approaches have limitations for professional use. The sensors worn on the playing wrist or built into the tennis racquets deliver information about the shots [6] or enable the analysis and modeling of different shot techniques [7]. However, the problem with this information is the lack of context (under what circumstances and where on the court did a specific shot occur), so it is not sufficiently actionable, i.e. cannot be used for tactical preparations or to significantly improve players' game. Video analysis offers better information, and there has been a lot of research on this topic [1, 5]. However, cheap solutions offer low accuracy, while better solutions are extremely expensive because they require advanced cameras with complex software for calibration. Additionally, they are bound to a specific court, so the information is not available whenever needed by the player or coach.

Due to these limitations, devices worn on the torso, and equipped with accelerometers, gyroscopes and GPS receivers are emerging as the new approach. These devices are perfect for determining the effort, distance covered, sprints analysis and much more. Here, the leading provider in the world is Catapult Sports, whose S5 product is currently used by the best tennis player in the world Andy Murray. Nevertheless, the problem with S5 is that it offers no tennis-specific metrics. That is why in our research we add tennis-specific information to the metrics already available in the Catapult S5 system, to produce a comprehensive solution that enables professional players to make better tactical preparations and to improve their game.

Our algorithm consists of two parts. In the first part, we detect when a tennis shot occurs and which type of shot it is. In the second part, we focus on detecting when the players actually play points (active play) and when they are in-between points. This allows us to determine the actual net playing time and real distance covered and also adds context to shots which enable complex analysis like "Is the player playing weaker shots, if the point is longer than 15 seconds?" With this solution the players and their coaches get a continuous comprehensive view of the player's game, both the physical and the tactical part of it.

2. DATA AQUISITION
To obtain sensor data we used the commercially available S5 device from Catapult. The position of the device was high on the player's back attached to a tight shirt. The device contains a 3D accelerometer (frequency 100 Hz), 3D gyroscope (frequency 100 Hz), 3D magnetometer (frequency 100 Hz) and GPS sensor, returning latitude and longitude (frequency 10 Hz).

We recorded 5 different professional tennis players for 6 hours in total. Due to the 100 Hz frequency, we obtained 2,172,363 data records. In this time, we recorded 1,373 shots. Each shot was labeled as a serve, forehand or backhand. As for detecting active play, we also manually labeled the beginning and end of each sequence of active play. Because we were interested in creating an algorithm for detecting shot types and active plays in actual matches, all the data were recorded during matches and none during predefined situations of practice sessions.

3. SHOT DETECTION
For every data point obtained from our device, we extracted a number of features used by the shot detection algorithm. We used supervised machine learning to train a model to detect shots. With this model we classified every data point and evaluated the shot detection.

3.1 Feature Extraction
To define informative features for shot detection, we visualized and examined the traces for the accelerometer and gyroscope. Since we saw that every shot is associated with body rotation, our main source for feature extraction was the gyroscope - more specifically angular speeds on axes 1 (Roll) and 3 (Yaw) - and not

the accelerometer. Figure 1 shows a typical trace of the gyroscope and accelerometer for a backhand shot.

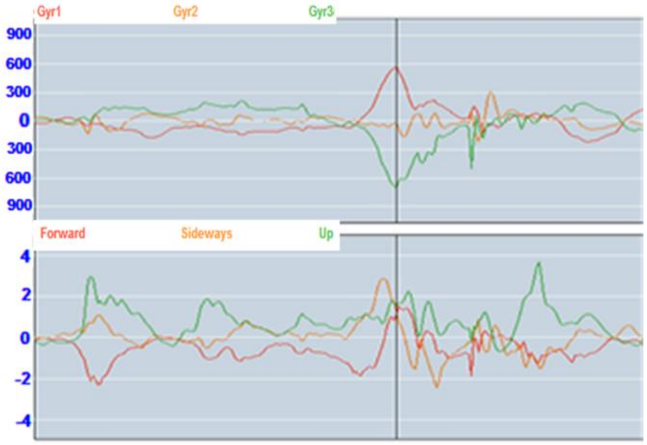

Figure 1: Gyroscope and accelerometer traces for backhand shot marked with the vertical line.

As our main feature, we calculated a feature called *Peak_strength* as follows:

- Calculate absolute sum of angular speeds on axes 1 and 3
- Raise it to the power 4, to emphasize higher values
- Apply Butterworth band-pass filter with high and low cutoff frequencies of 1.5 and 25 Hz.
- To get the final *Peak_strength* value, set the lower peak to zero when two peaks are too close (the distance was set by a domain expert to 1.3 s)

High *Peak_strength* values calculated in this way mark potential shots. Additionally, we calculated several other features. We set two different window sizes (0.8 s and 1.2 s) and calculated the average values, variances and standard deviations for each accelerometer and gyroscope axis. We added the sums of and differences between all pairs of gyroscope axis values and also between accelerometer axis values. We also calculated the speed of movement from the GPS coordinates. To illustrate its importance, Figure 2 shows how the combinations of *Peak_strength* and speed of movement separates shots and shot attempts (high *Peak_strength* values that are not shots).

3.2 Experimental Setup

We divided the evaluation in two parts. Firstly, we evaluated how well we can detect if a shot has occurred, and secondly, we tried to detect which type of shot was made.

For building the models, after empirical comparison of several algorithms, we chose the Random Forest (RF) [2] algorithm. Each RF consisted of 10 decision trees, the minimum number of samples required to split an internal node was 8, and the minimum number of samples required at a leaf node was 4.

We evaluated the models in two ways. Firstly, we performed 10-fold cross validation using Stratified shuffle split [3]. This procedure ensured equal class distributions between training and test sets. Secondly, we used the leave-one-player-out approach (LOPO), where we used one player's data for testing and the data from the other players for training. This approach enables us to estimate the accuracy of the models for previously unseen players with different shot techniques.

When evaluating the models, we classified each data entry (10 ms) as a shot or no-shot, and the type of shot. With this approach almost all the data points were classified as no-shots, so calculating the classification accuracy would be useless. We therefore focused on the precision and recall.

3.3 Results

The results for detecting shots and shot types for the cross-validation and for the LOPO approach are presented in Tables 1 and 2.

	Cross-validation	**LOPO**
Precision	97.3%	97.3%
Recall	96.6%	96.5%

Table 1: Precision and recall for detecting tennis shots

	Cross-validation			
	Foreh.	**Backh.**	**Serve**	**All**
Precision	95.3%	94.3%	99.1%	96.2%
Recall	91.4%	90.2%	99.3%	93.6%
	LOPO			
	Foreh.	**Backh.**	**Serve**	**All**
Precision	91.5%	93.6%	99.8%	95.0%
Recall	90.5%	90.6%	98.2%	93.1%

Table 2: Precision and recall for detecting types of tennis shots

As we can see, the precision and recall obtained with cross-validation and LOPO are very similar. This means that the built models are relatively independent from the type of player or his technique or style of play.

The main sources of errors are fast unnatural body rotation movements and special events that occur during the play. An example from our data set is a player warming up doing very similar body movements as during shots, or a player throwing his racquet at the fence with the same body movement as when serving.

4. DETECTING ACTIVE PLAY

The algorithm for detecting active play during a tennis game could only be developed after we have detected the shots. The reason is that we want our algorithm not only to have a high classification accuracy, but also to include as many shots as possible in the detected active play. In other words, misdetection of active play is less undesirable when no shots are made. So due to having two objectives, the detection was formulated as a multi-objective optimization problem.

4.1 Feature Extraction

The main idea when detecting the starting (end ending) point of a sequence of active play (rally) was that at this point the difference between the activity before and after would be the largest.

We used accelerometer values because they better represent the players' movement then gyroscope values, which primarily spike when making shots. From these values, we calculated a modified variance that gives more emphasis to the largest variations in data traces:

$$\text{var*} = \frac{\sum_{i=1}^{N}(x_i - \overline{x})^4}{N}$$

So for each data point, we calculated three additional features based on var*: the back overall variance (BV), the forward overall variance (FV) and the difference between these two (DV = BV – FV). FV and BV are calculated as the sum of the var* for each of the three acceleration axes, on the sequences immediately before (BV) and after (FV) a potential beginning or end of a rally (the size of the sequences was subject to optimization).

To be able to truly detect the best point describing the beginning or end of each rally, we also calculated peaks on the DV feature. Calculating the peaks was done the same way as for the shot detection. The minimum distance between peaks was subject to the optimization.

4.2 Problem Formulation

For each data point we calculated the previously described features, and set a rule for detecting the beginning of a rally and a rule for detecting the end of a rally. A data point is marked as the beginning of a rally if it satisfies the following rule:

$$(DV > p1)\ \&\ ((BV < p3)\ ||\ (FV < p4)).$$

A data point is marked as the end of the rally if it satisfies the following rule:

$$(-DV > p2)\ \&\ ((BV < p3)\ ||\ (FV < p4)),$$

where parameters $p1$, $p2$, $p3$, $p4$ were determined through optimization. Both rules consist of two parts. The first part determines the threshold for the change in activity before and after a potential beginning or end of a rally. For the beginning of a rally, this difference is usually larger because a rally often starts explosively and ends gradually, so the thresholds $p1$ and $p2$ can be different. The second part is the same for both rules and serves to remove false detections due to the variation in intensity during the rally by specifying that the activity, either before or after the beginning or end of a rally, should be low.

So altogether we optimized six input parameters: sequence size, minimum distance between peaks, $p1$, $p2$, $p3$ and $p4$.

4.3 Experimental Setup

To optimize the two objectives – classification error and the number of shots not inside the detected rallies – we used the well-known evolutionary multi-objective optimization algorithm called NSGA-II [4]. The population size was set to 25, the stopping criterion was set to 10,000 solution evaluations, and the tournament selection was used.

4.4 Results

The final front of the optimization can be seen in Figure 3. We can see a typical result for a multi-objective optimization problem, a non-dominated front showing a tradeoff between objectives. We can also see a knee on the front labeled with a circle. In this solution six shots are missed, since they occurred without the surrounding intense activity which accompanied other shots. An example is a player hitting the ball out of court after the rally finished. To include even such shots in the detected rallies, we would need to sacrifice a lot of classification accuracy.

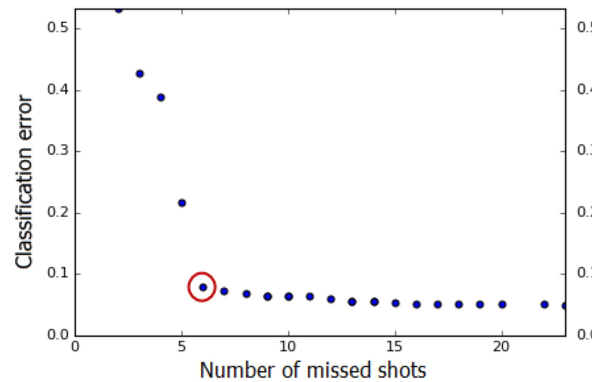

Figure 3: The final front showing the best solutions based on the classification error and the number of shots outside of detected active play.

Since our objective was to accurately detect the duration of the rallies, we chose one solution from the middle of the front and for this solution, we calculated the distribution of the durations of the rallies. The comparison with the manually labeled rallies can be seen in Figure 4.

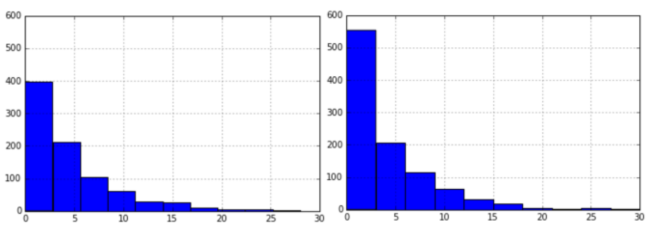

Figure 4: Comparing manually labeled (left) and automatically detected distributions for play durations.

We can clearly see the similarities between the distributions. The reason for the detected distribution having more very short rallies is that the algorithm detects even small starts of movement that we did not label as rallies because they were too short. For example, a server hitting the net with the first serve results in the returner making just a small movement.

By combining the classified shot types, detected active playing phases and locations from the GPS, we can calculate several useful metrics that help remove subjectivity from the game and allow for objective evaluation of different tactical approaches and training routines. An example of such a view can be seen in Figure 5, where we present the heat map of a player's position during active play and combine it with forehand and backhand shots as points of different color and size. We also included a dashed line that separates part of the court where more backhands are played from the one where more forehands were played. We can see that the player played more aggressively on the left side, thus his heat map is closer to the baseline. On the right side, a less aggressive approach allowed him to play more forehands and thus he dictated play by playing more often with his better shot.

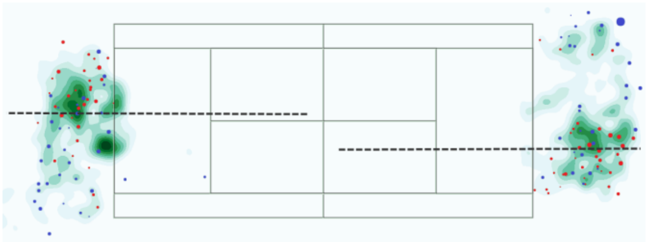

Figure 5: Heat map of a player's position during active play combined with shot locations (blue = forehand, red = backhand) and their *Peak_strength* (size of points).

5. CONCLUSION

In this article we presented a two-part algorithm for analyzing wearable sensor data for professional tennis players. Firstly, we detected and classified different shot types, and secondly, we distinguished the active playing phases from the time in-between points. By combining the procedures, players can get a unique perspective on their game which enables objective analysis in the tactical and physical sense.

For the future, we plan to equip both players with the same type of sensor, and by measuring the time difference between their shots and by calculating the distance between them, we will be able to calculate the average speed of ball and thus additionally quantify the quality of each shot.

6. ACKNOWLEDGMENTS

We would like to thank Catapult Sports for their sound advice and cooperation, and the national research agency (ARRS) for providing the research grant.

7. REFERENCES

[1] Renò, Vito, et al. "A technology platform for automatic high-level tennis game analysis." Computer Vision and Image Understanding (2017).

[2] Liaw, Andy, et al. "Classification and regression by RandomForest." R news 2.3 (2002): 18-22.

[3] Liberty, Edo, Kevin Lang, and Konstantin Shmakov. "Stratified sampling meets machine learning." International Conference on Machine Learning. 2016.

[4] Deb, Kalyanmoy, et al. "A fast and elitist multi-objective genetic algorithm: NSGA-II." IEEE transactions on evolutionary computation 6.2 (2002): 182-197.

[5] Yu, Xinguo, et al. "A trajectory-based ball detection and tracking algorithm in broadcast tennis video." Image Processing, 2004. ICIP'04. 2004 International Conference on. Vol. 2. IEEE, 2004.

[6] Pei, Weiping, et al. "An embedded 6-axis sensor based recognition for tennis stroke." IEEE International Conference on Consumer Electronics (ICCE), 2017.

[7] Yang, Disheng, et al. "TennisMaster: an IMU-based online serve performance evaluation system." Proceedings of the 8th Augmented Human International Conference. ACM, 2017