Deep Learning for Tennis Stroke Classification Using Tech Wearables

Ellen Roper; Tristan Gosakti
Stanford University

Abstract

Motion capture technology offers a myriad of applications in many different fields. TuringSense, a Santa Clara-based tech wearable company, created a product which uses motion capture to train athletes. Standard motion capture is often optical in nature, but TuringSense replaces this approach with digital technology, opting to use motion sensors instead. We use data collected with TuringSense's sensors to create a deep learning model which can predict and classify a tennis player's stroke. We trained a simple baseline model with one hidden layer, a deep network with four hidden layers, and a convolutional neural network on a multi-label stroke classification task, obtaining consistently high classification accuracy for five different tennis strokes on each, with the convolutional neural network achieving the highest accuracy.

Introduction

Currently, TuringSense uses their own mathematical models to process raw data from 6 sensors attached to different parts of the human body and the tennis racket for tennis stroke classification. The specifics of their objectives are undisclosed at the company's request, but it is clear that providing accurate data feedback to their users is one of TuringSense's most important goals. Motion capture as a whole has the potential to revolutionize training methods for athletes in many different physical activities.

Data

- 1 stroke is 0.5-1.5 sec of timestamped input from 6 sensors.
- Collected roughly 100 examples of each of five types of stroke
- Augmentation by adding normally distributed random noise to collected examples gave us 1421 training examples
- Validation/test sets of 50 samples; 10/stroke.

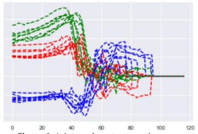

Figure 1. Joint angle output vs. time (x10 ms) of one sensor for forehand flat.

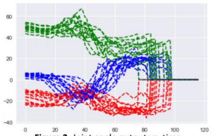

Figure 2. Joint angle output vs. time (x10 ms) of one sensor for forehand slice.

Procedure

- Raw data converted to angle data by TuringSense's algorithms.
- Extract individual strokes from metadata about stroke start/end.
- We pad each stroke with zeros to ensure all strokes are the same length, (determined by pre-analysis of our data).
- Sensor data at each timestamp is stacked vertically and appended into an 18 data point x 111 timestamps "image."
- We run this image through a convolutional neural network to output a one-hot vector with the predicted classification. Our five stroke type classifications are:

Class 1: Forehand Flat
Class 2: Forehand Topspin
Class 3: Backhand Slice
Class 4: Backhand topspin
Class 5: Forehand Slice

Model

Convolutional Neural Network With Two Convolutions and Max Pool

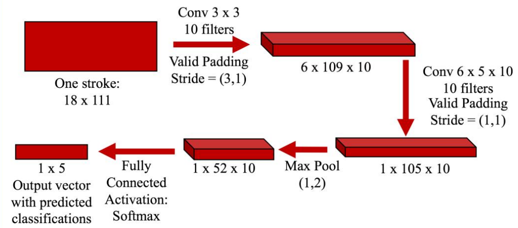

Table 1. Architecture Search. Different performances for three attempted architectures. Entries are the best results we were able to obtain with each model.

	Best Train Accuracy	Best Dev Accuracy	Best Test Accuracy
1 Hidden Layer	0.9754	0.94	0.96
4 Hidden Layers	0.9862	0.96	0.97
CNN	0.9938	0.98	0.98

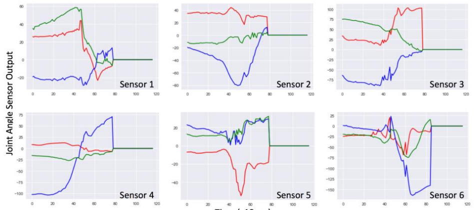

Chart 1. Output over time from each of six sensors on a forehand flat stroke.

Discussion and Results

The first models we tried were baseline neural networks, first with a single hidden layer and secondly with four hidden layers. Both were able to achieve high classification accuracy. However, due to constraints in data size, we decided to implement a CNN, as this likely will be better at accounting for variability in data if more is collected. This CNN performs consistently better than the simpler deep networks, with an average classification accuracy of 0.98 on the test set. As we run the model with augmented data of increasing variability (as would be seen in real life), the CNN continues to outperform other models, giving our best result.

Future Modifications and Special Thanks

Physical constraints requiring consistent repetitive motion of tennis strokes, fatigue, and considerable time means that we have a severe lack of data. Future work could involve:

- Collecting more data to further train and refine the results.
- Adding variability with data from more than one professional athlete.
- Expand the number of strokes classifications, as there are more than five strokes which are used in tennis.

We would like to thank TuringSense for their enthusiasm and support for our project. Please visit their website for more information and for updates on their amazing innovations.
www.turingsense.com

Contact

Ellen Roper
Stanford University
Email:
eroper@stanford.edu

Tristan Gosakti
Stanford University
Email:
tgosakti@stanford.edu

TuringSense
Website:
www.turingsense.com

References

[1] Connaghan, Damien, et al. "Multi-Sensor Classification of Tennis Strokes." *2011 IEEE SENSORS Proceedings*, 2011, doi:10.1109/icsens.2011.6127084.
[2] Jiang, Wenchao, and Zhaozheng Yin. "Human Activity Recognition Using Wearable Sensors by Deep Convolutional Neural Networks." Proceedings of the 23rd ACM International Conference on Multimedia - MM '15, 26 Oct. 2015, pp. 1307–1310., doi:10.1145/2733373.2806333.
[3] Ronao, Charissa Ann, and Sung-Bae Cho. "Human Activity Recognition with Smartphone Sensors Using Deep Learning Neural Networks." Expert Systems with Applications, vol. 59, 15 Oct. 2016, pp. 235–244., doi:10.1016/j.eswa.2016.04.032.
[4] Shah, Hitesh, et al. "Automated Stroke Classification in Tennis." *Lecture Notes in Computer Science Image Analysis and Recognition*, pp. 1128–1137., doi:10.1007/978-3-540-74260-9_100.
[5] TensorFlow. (2018). Get Started | TensorFlow. [online] Available at: https://www.tensorflow.org/get_started/