

A collection of sport activity files for data analysis and data mining

Authors:

- Samo Rauter (University of Ljubljana)
- Iztok Fister Jr. (University of Maribor)
- Iztok Fister (University of Maribor)

Contact details:

In case of any questions, please let us know at:

- samo.rauter@gmail.com or
- iztok.fister1@um.si

Donation details:

We invite everyone who is interested to donate data to this dataset to contact us.

Download details:

Dataset is available for download as a torrent file via:
<http://www.academicorrents.com/>

Citation details:

S. Rauter, I. Jr. Fister, I. Fister. A collection of sport activity files for data analysis and data mining. Technical report 0201, University of Ljubljana and University of Maribor, 2015.

Motivation

The increased popularity of smartphones and watches allows athletes to train smarter. More and more cyclists today use a different sports applications, that run on mobile devices, such as Strava, Endomondo, Garmin Connect and many others. Primary tasks of such applications is tracking and analyzing their workouts. These tracked workouts can also be exported as XML files and analyzed later. In other words, they offer an immense of different tasks of data analysis and data mining. Some ideas about using these methods were discussed in paper [2]. The mentioned data should be also used for planning the sport training sessions as discussed in [1].

In line with this, we had made some literature reviews about these research topic and realized that there is a lack of researches tackling the data analysis and data mining of sport activities created by sport trackers [3]. This was the primary fact for deciding us to collect training data of different outdoor athletes and make these publicly viewable in order to enable potential researchers to discover new ways of training data analysis. The majority of athletes captured in this dataset are cyclists with a large amount of experience with this sport.

Description of this dataset

Dataset consists of the data produced by nine cyclists. Data were directly exported from their Strava or Garmin Connect accounts. Data format of sport's activities could be written in GPX or TCX form, which are basically the XML formats adapted to specific purposes. From each dataset, many following information can be obtained: GPS location, elevation, duration, distance, average and maximal heart rate, while some workouts include also data obtained from power meters.

Ethics statement

Data in this dataset were voluntarily donated by cyclists. These cyclists would like to stay anonymous. Since these data are intended for research purposes, it is prohibited for competitors and rivals of these volunteers to analyse their data for own comparison.

References

- [1] Iztok Fister, Samo Rauter, Xin-She Yang, Karin Ljubič, and Iztok Jr. Fister. Planning the sports training sessions with the bat algorithm. *Neurocomputing*, 149:993–1002, 2015.
- [2] Iztok Jr. Fister, Dušan Fister, Iztok Fister, and Simon Fong. Data mining in sporting activities created by sports trackers. In *Computational and Business Intelligence (ISCBI), 2013 International Symposium on*, pages 88–91. IEEE, 2013.
- [3] Samo Rauter, Mojca Doupona Topič, and Iztok Jr. Fister. Mobile sport applications can make our cycling more sociable. In *Youth sport: abstract book, Ljubljana: University of Ljubljana, Faculty of Sport*, page 41, 2014.