

# *A collection of sport activity datasets for data analysis and data mining 2016b*

## **Authors:**

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

## **Contact details:**

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

- [iztok.fister1@um.si](mailto:iztok.fister1@um.si) or
- [samo.rauter@gmail.com](mailto:samo.rauter@gmail.com)

## **Donation details:**

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

## **Citation details:**

I. Jr. Fister, S. Rauter, D. Fister, I. Fister. A collection of sport activity datasets for data analysis and data mining 2016b. Technical report, University of Maribor, 2016.

## **Download details:**

Dataset is available for download as a torrent file via: <http://www.academictorrents.com> or direct [LINK](#)

## **Motivation**

The increased popularity of mobile smart-phones and smart watches has also extended the scope of usability to sports. Using this technology is especially evident in sports training, where it allows athletes to train smarter. In line with this, more and more athletes use a variety of sports applications. Sports applications, such as Strava, Endomondo, Garmin Connect and many others, are a perfect tool for tracking and analysing their workouts. Workouts should also be exported as XML files and analysed later on personal computers [4]. In other words, it offers an immense amount of different tasks for data analysis and data mining. Some ideas about how to use data mining methods were discussed in the paper [2]. Data should also be used for planning the sport training sessions as discussed in [1, 3].

On the other hand, we made some literature review about these research topics and realised that there is a lack of research about data analysis and data mining of sport activities created by sport trackers. This finding convinced us to collect some amount of training data of different outdoor athletes as the present collection.

## **Historical background**

The first collection of sport activity datasets was released in early 2015 [5], with the second in 2016 [6]. The first two collections consisted of both GPX, as well as TCX, types of training activity files and cover mostly cyclists (professionals and amateurs).

## **Description of this collection**

After the release of the first two collections, many researchers requested also training activity datasets covering other sports too. In this collection we try to eliminate this bottleneck and also offer researchers to use research data that was donated by some triathletes (swimming, cycling, running, and fitness sessions are included) and one runner. Thus, data were exported directly from their Strava and Garmin Connect profiles.

## **Main differences between previous collections**

In this collection, we do not include TCX files any more. Only GPX files are included. Some athlete's data that were included in previous collections are now updated with new data.

## Ethics statement

Data in this collection contains personal records of training and locations. All athletes donated data voluntarily **only for research purposes**. All athletes stay anonymous. Anyway, it is prohibited for anyone (especially for opponents and rivals of these volunteers) to analyse data deeply for their own comparison or similar tasks. Additionally, please do not post visualised routes of activities on social networks or sport websites.

## References

- [1] Iztok Fister, Samo Rauter, Xin-She Yang, Karin Ljubič, and Iztok Fister Jr. Planning the sports training sessions with the bat algorithm. *Neurocomputing*, 149:993–1002, 2015.
- [2] Iztok Fister Jr., 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] Iztok Fister Jr., Karin Ljubič, Ponnuthurai Nagaratnam Suganthan, Matjaž Perc, and Iztok Fister. Computational intelligence in sports: Challenges and opportunities within a new research domain. *Applied Mathematics and Computation*, 262:178–186, 2015.
- [4] Samo Rauter, Iztok Fister, and Iztok Fister Jr. How to deal with sports activity datasets for data mining and analysis: Some tips and future challenges. *International Journal of Advanced Pervasive and Ubiquitous Computing (IJAPUC)*, 7(2):27–37, 2015.
- [5] Samo Rauter, Iztok Fister Jr., and Iztok Fister. A collection of sport activity files for data analysis and data mining. *Technical report 0201*, University of Ljubljana and University of Maribor, 2015.
- [6] Samo Rauter, Iztok Fister Jr., and Iztok Fister. A collection of sport activity files for data analysis and data mining 2016a. *Technical report 0101*, University of Ljubljana and University of Maribor 2016a, 2016.