A collection of sport activity datasets for data analysis and data mining 2017a

Iztok Fister Jr., Samo Rauter, Dušan Fister, Iztok Fister

March 30, 2017

1   Description of this collection

This technical report presents a collection of sport activity datasets for data mining purposes. It is following the three previously published collections in 2015 [1] and 2016 [2, 3] and covers various sports and different types of athletes (professionals and amateurs). Data were exported from Strava, Garmin Connect and Movescount App profiles (last variant was also synchronized to Strava) and are available in a GPX format (some activities even contain cycling's power-meter values). Table 1 presents more details about athletes who donated data. Most of the data is donated by the same athletes than from previous collections, therefore, their data is updated with new activities.

Name	Sex	Age in 2017	Amateur/Pro	Primary sport
Athlete 1	Male	20	Amateur	Cycling
Athlete 2	Male	43	Amateur	Multisport
Athlete 3	Male	20	Amateur	Cycling
Athlete 4	Male	35	Amateur	Running
Athlete 5	Male	24	Amateur	Multisport[1]
Athlete 6	Male	41	Amateur	Cycling
Athlete 7	Male	22	Amateur	Cycling
Athlete 8	Male	19	Pro	Cycling
Athlete 9	Male	23	Amateur	Cycling
Athlete 10	Male	56	Amateur	MTB
Athlete 11	Male	30	Pro	Triathlon
Athlete 12	Male	46	Amateur	MTB
Athlete 13	Female	29	Pro	MTB
Athlete 14	Male	35	Amateur	Cycling and MTB
Athlete 15	Male	33	Pro	Triathlon

Table 1: Details of athletes

The Table 1 illustrates a deeper view of the collected data. The first column lists anonymous athletes, while the second identifies their sex. The third column describes the age of an athlete and the fourth defines the type of athlete - amateur or professional. The Table is completed with the fifth column, which specifies the primary sport of a particular athlete. The notation "primary sport" states the athlete's main sport's concentration, however some other sports can be found in his/her collection (especially during the dead season - winter). Multisport means that an athlete performs more sports (his concentration is not clearly obvious). It is worth mentioning that triathlon is not treated as a "multisport" in this dataset, but separately. MTB stands for mountain biking.

2 Ethics statement

We are very thankful to our valued athletes, who made this collection publishable by donating data (without them it would not be possible). Data in this collection contain personal records of training and locations. All athletes donated data voluntarily **for research purposes only**. All athletes are anonymous. It is prohibited for anyone (especially for opponents and rivals of these volunteers) to analyze data deeply for their own comparison or similar tasks. Additionally, please do not post visualized routes of activities on social networks or other websites. Other techniques that may reveal the identities or locations of these athletes are also prohibited!

3 Download details

The collection is available for downloading via http://www.academictorrents.com (torrent file) or direct LINK.

4 Citation details

I. Jr. Fister, S. Rauter, D. Fister, I. Fister. A collection of sport activity datasets for data analysis and data mining 2017a. Technical report, University of Maribor, 2017.

5 Future release

The next collection is expected to be compiled in October/November 2017. We are also trying to ensure some currently missing sports - skating, rowing.

References

[1] 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.

---

[1]There are also some sport flying activities in this dataset collection

[2] 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.

[3] Iztok Fister Jr., Samo Rauter, Dušan Fister, and Iztok Fister. A collection of sport activity datasets for data analysis and data mining 2016b. *Technical report, University of Maribor*, 2016.