SportyDataGen: An Online Generator of Endurance Sports Activity Collections

Iztok Fister Jr., Grega Vrbančič, Lucija Brezočnik, Vili Podgorelec and Iztok Fister

University of Maribor

Faculty of Electrical Engineering and Computer Science

Koroška cesta 46, 2000 Maribor, Slovenia

{iztok.fister1, grega.vrbancic, lucija.brezocnik, vili.podgorelec, iztok.fister}@um.si

Abstract. *Analyzing sport data becomes, every year, more interesting for a wide spectrum of researchers in the sports domain. Recently, more and more data relating to sports have become available to researchers due to the huge progress of information technologies. New wearable devices enable athletes to track performance data that are saved into sport activity datasets and later analyzed by sport trainers using data mining methods. These datasets are also very useful to researchers in the sports domain who usually employ a collection of them in their researches. Typically, these researchers are confronted with two problems: Firstly, how to gain the real data, and secondly, how to process them. In this paper, we propose a new tool for generating a collection of sport activity datasets online (named SportyDataGen), where data from various endurance sports disciplines, obtained from different mobile devices worn by amateur, as well as professional athletes are stored. The data are generated in CSV format according to user demands, and do not require any preprocessing. Here, the proposed tool is described systematically, while one example is also presented of a generated collection.*

Keywords. sport data analysis, data mining, triathlon, online generator

1 Introduction

Analyzing sports activity datasets has become an unavoidable tool for monitoring advances of athletes involved in the process of sport training. This statement is supported by the rising power of Information Technology (IT) on the one hand, and development of more sophisticated algorithms in the domain of Artificial Intelligence (AI) on the other. Simultaneously, a lot of people have been involving into sports, into which they are forced due to their, usually sedentary, modern life-style. Mainly, the endurance sports disciplines are preferred by the athletes, because those are individual sports and do not need any sports infrastructure (e.g., sports hall, special sports tools, etc.), where athletes train, when they have time, and therefore, there

is no need for special coordination between people as in the case of a team sports disciplines. Nowadays, these athletes use smart watches or smart phones during the training session in order to analyze tracked data about realized their sport activities after training. Typically, these devices are the origin of producing the huge amount of data. Indeed, data in this domain have been rising exponentially.

Swimming, biking, and running are typical endurance sports disciplines that are more eligible for amateur athletes. However, combining these three sports caused the formation of the more complex sports disciplines, such as, for example, triathlon. The word **triathlon** refers to the modern sport that is one of the more attractive sports disciplines recently, where the final time of the competition is the sum of all three sports disciplines, together with two transitions. Thus, the first transition is devoted to preparing an athlete from swimming to cycling, and the second from cycling to running. It is worth mentioning that there are various triathlon distances that spread from the short distance toward the medium and long distances, until the ultra-triathlons.

Although the triathlon is a very young sport, every year it attracts more and more competitors from all over the world. This trend is already evidenced by the following facts: (1) The majority of international races (e.g. IRONMAN triathlons) are filled with competitors from the whole world shortly after the Internet registrations are opened, (2) Many personal trainers have become more busy due to providing the process of sports training for an increased number of triathlon enthusiasts during the whole year, and (3) Triathlon has also become an interesting subject for intensive scientific researches [11, 16, 13, 10].

In the past years, most of the researches related to performance studies, tactics, and health in cycling, running, and triathlon. These sports also became interesting for researchers in Data Science and other AI domains by rising the volume of data produced during training sessions. For example, Matabuena et al. [14] showed that the performance of a group of athletes can be predicted without forcing them to fatigue by exer-

cises of low intensity. Not long ago, Mnadla et al. [15] outlined data concerning the related activities and interest for the Ironman competition on the Internet, while Knechtle et al. [12] analyzed participation and performance trends in ultra-triathlons. Moreover, analyzing cycling data [2, 17] also produced valuable research for a wide audience. On the other hand, Cintia et al. [3] shows how to exploit the massive data exported from Strava to construct personalized training programmes for cyclists.

Interestingly, public access to endurance sports data for research is still a very big bottleneck for researchers in this research domain, because a lot of data remain unreachable due to the policy of athletes, trainers, or even sport clubs. For that reason, studies cannot be replicated, while some studies cannot even be conducted because of the lack of data. In the past, we have tried to supply data produced by wearable devices (smart phones and smart watches) to researchers in some individual sports disciplines [18, 8]. In line with this, we were focused on collecting data obtained from real athletes. However, this task was not so easy, because some athletes did not want to share their activities publicly. Therefore, we have assembled sports activity data obtained from a group of athletes who were willing to share their data. Every year, we published a big collection of these data online. Fascinatingly, we received a very positive feedback over the years.

Unfortunately, these datasets were collected in their native format (i.e., TCX and GPX file formats), and, therefore, preprocessing (i.e., parsing) is needed before using them. Obviously, this task is far to be simple, especially for researchers without deep knowledge of computer programming. In order to help them, we decided to develop an online sports activity generator of sports activity collection, where users would be able to select what and how much of the generated data they want. This generator, named **SportyDataGen**, can be found at the Internet page `www.sport-slo.net`.

The aim of the paper is to present the features of the online generator of an endurance sports activity collection that are available to a potential user on the Internet. The structure of the remainder of the paper is as follows. Section 2 discusses problems arising in using existing collections of an endurance sports activity datasets. Generation of the endurance sports activity collection with SportyDataGen is described in Section 3. Examples of the generated endurance sports activity datasets are presented in Section 4. The paper is concluded with Section 5, which summarizes the performed work and outlines directions for the future.

2 Problems using existent collections of endurance sports activity datasets

SportyDataGen relies on real data that were obtained from mobile devices worn by amateur and professional athletes during their endurance sports training sessions. Most of the data were exported directly from Internet applications, like Garmin Connect [9], Strava Connecting the world's athletes [20], or Suunto MovesCount [21]. Athletes gave us their credentials in order to access their profiles on such applications. At the same time, they agreed that data can be used in this research. The exported datasets can either be in TCX or GPX file format. Actually, both formats basically have their roots in eXtensible Markup Language (XML).

Although these activity datasets obtained from real athletes were collected and made available publicly for the purpose of research, they were rarely used in practice, especially because they must be preprocessed before their use. However, the preprocessing is not a simple task, due to the fact that there is no sophisticated parser of these file formats available in particular programming languages. As a result, the parser needs to be developed from scratch, but this demands a sufficiently high level of knowledge in computer programming. Unfortunately, a lot of users suffer from a lack of this knowledge and, therefore, are unable to carry out such preprocessing.

Obviously, the authors of these collections are aware of these problems. Consequently, some previous versions of the collections allow downloading the activity datasets in a raw data format suitable for exploring without preprocessing [8]. Although these collections are very useful, they are too static, and do not consider any variability (e.g., using similar activity datasets from two different athletes). Indeed, this situation calls for using a generation tool for generation of the endurance sports activity collections that are suitable for generating data on user demand.

3 Generation of endurance sports activity collection

The SportyDataGen is a generator for endurance sports activity collection consisting of more components working together sequentially. Actually, the architecture of the generator that is presented in Fig. 1 is very complex. There are the following six components of the generator:

- Collection: This component enables automatic downloading of the endurance sports activities from online profiles to be at the disposal by producers of specific wearable devices, like Garmin, Strava, Suunto, etc.

- Processing: Downloaded raw endurance sports activities saved in TCX or GPX file formats are parsed by the mentioned components. This means that the most important features are extracted from raw files. Thus, features, like the average Heart Rate (HR), the total sports training Time Duration (TD), the total distance and consumed calories, characterize a typical endurance sports training session.

- Database: The purpose of the component is to store extracted features into a database for further analysis. Typically, each training session in the database is identified by the training load indicator - TRaining IMPulse (TRIMP), proposed by Banister [1]. This load indicator is expressed simply as:

$$\text{TRIMP} = TD \cdot HR, \qquad (1)$$

where TD denotes the time duration of the endurance sports training session in minutes, and HR is an average Heart Rate in beats per minutes.

- Generation: The corresponding endurance sports activities are selected from a database according to the user's preferences, like the sports discipline, the number of generated endurance sports activities in the collection, and the features that should be generated. In line with this, these activiis can also be clustered regarding the TRIMP measure.

- Output: The component is devoted for exporting the selected collection of measured sports activities in CSV format. Additionally, the table with the same features are also made available to the user.

- Web application: It allows control of the database and the generation process of the measured sports activity collection. On the other hand, the application enables users to interact with a User Application Interface (API) for tailoring the collection according to their needs.

The first three components are already part of the Artificial Sports Trainer (AST) proposed by Fister et al. [7], dedicated to acquire the sports training activities into a database. Indeed, the database serves as a basis for decision-making about realization of the sports training process by the AST. In line with this, the AST uses it for planning the sports training sessions [5], adapting the training sessions [6], proposing the dietary plan necessary for covering energy consumption as demanded by the specific training plan [4].

In this study, the database was exploited for generation of endurance sports activity collection, where the prescribed number of endurance sports activities are selected according to the user's preferences in a CSV file format that does not need any preprocessing and is, therefore, also available for use by users with less knowledge of computer programming. The following preferences are available to users by generating the collection of sports activities:

- the endurance sports discipline

- the number of generated activities

- features

- the number of clusters

- intensity of the generated endurance activity collection in TRIMP

- the generation mode

At this moment, two endurance sports disciplines are supported, i.e., cycling and running. However, more endurance sports disciplines could be supported in the future (e.g., triathlon). The number of generated endurance sports activities determines the endurance sports activities collection. Users can select between the following features: heart rate, duration, distance, calories, maximum and average altitude, and pace. The number of clusters prescribes those numbers of different clusters used by the k-means clustering algorithm, from which the activities could be selected. The intensity of the generated activities is optional, and actually refers to the generation mode. If the generation mode is purely random, then the value of zero is expected, which means that the intensity of all training activities in the collection is ignored. In contrast, when the evolutionary approach is selected, a minimum TRIMP intensity value I_{C_i} of the appropriate participating cluster is specified by the evolutionary optimization process.

As already mentioned, the SportyDataGen supports two modes of generating the in the remained of the paper sports activity collection, i.e., purely random, and evolutionary. In the remained of the paper, both modes are described in detail.

3.1 Generation modes

Two modes of generating the endurance sports activity collection exist. Actually, both modes start with clustering the endurance sports activities into a definite number of clusters. The clusters are determined according to the TRIMP measure, while the k-means clustering algorithm is applied for clustering. However, the number of clusters is crucial for the performance of the generation. The more clusters are selected, the more diverse activities can be generated. At the moment, the SportDataGen supports from 3 to 10 numbers of clusters. Typically, the proper number of clusters must be determined experimentally.

The set of endurance training sessions are clustered according to the TRIMP training load indicator in order to obtain groups of training sessions of similar intensities. As can be seen from Eq. (1), time duration and average heart rate have influence on the TRIMP training load indicator. However, the main disadvantage of this indicator is that it is insensitive to the different levels of training. Users of SportyDataGen can choose the
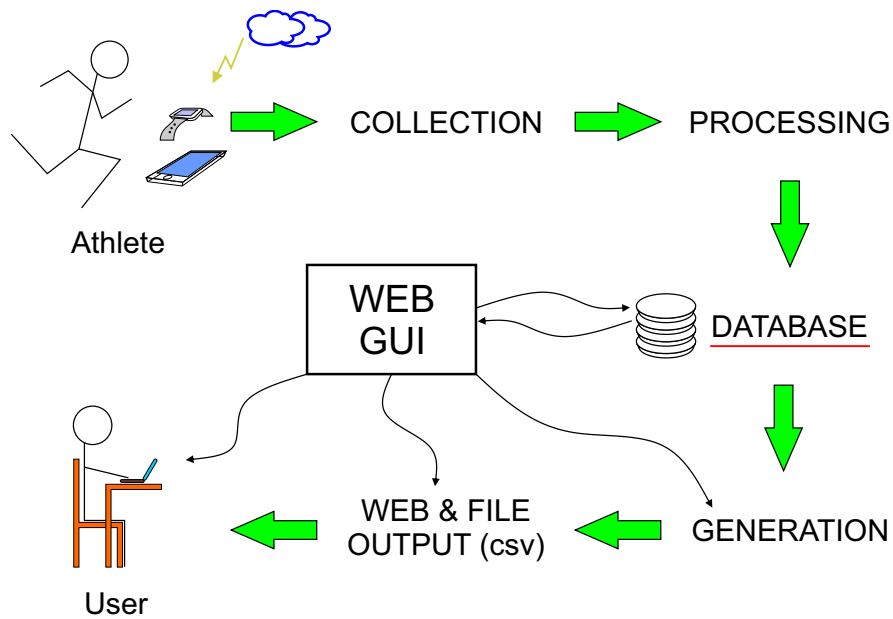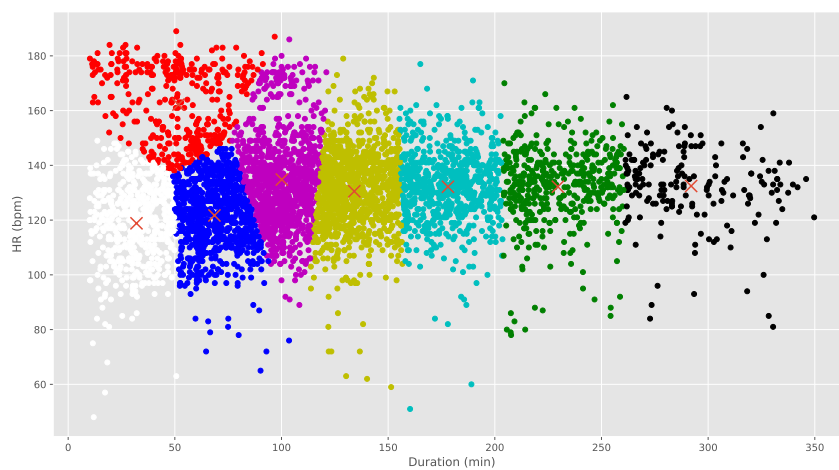
Figure 1: Architecture of SportyDataGen.

Figure 2: Example of clustering (C = 8).

desired cluster settings. Fig 2 presents an example of clustering, where the number of clusters is 8.

Generation of the endurance sports activity collection can be defined mathematically as follows. Let us assume a collection of endurance sports activities as a real-valued vector:

$$\mathbf{x}_i = [x_{i,1}, \ldots, x_{i,D}]^T, \quad \text{for } i = 1, \ldots, NP, \quad (2)$$

and a set of clusters are given:

$$C = \{C_1, C_2, \ldots, C_n\}, \quad (3)$$

where NP denotes the number of different collections, D is the number of endurance activities in the collection, n the number of clusters. Each cluster is defined as:

$$C_k = \{t_{k,1}, \ldots, t_{k,m_k}\}, \quad \text{for } k = 1, \ldots, n, \quad (4)$$

where $t_{k,l}$ denotes a specific endurance training session and m_k is the size of the k-th cluster. A vector denoting a sequence of clusters

$$\mathbf{c}_i = [c_{i,1}, \ldots, c_{i,D}]^T, \quad \text{for } i = 1, \ldots, NP, \quad (5)$$

is assigned to each of the vector $\mathbf{x}_i$ that determines a specific cluster from which the endurance sports activity can be selected. The sequence of clusters is then generated according to the following assumptions. Until the vector $\mathbf{c}_i$ is not full, a tournament is played between the specific cluster j and its corresponding threshold value is expressed as follows:

$$threshold(C_i, C_j) = \left| I_{C_i} - I_{C_j} \right| / 10. \quad (6)$$

If the value drawn from the uniform random distribution in interval $[0, 1]$ is less than the threshold value, the j-th cluster is placed on the observed position in the vector $\mathbf{c}_i$. The intensity cluster entered via user's API $I_{C_i} = 0$ favors those clusters C_j that gave the closest distances $|I_{C_i} - I_{C_j}|$ according to the following ordering:

$$I_{C_0} \succeq \ldots \succeq I_{C_{i-1}} \succeq I_{C_i} \succeq I_{C_{i+1}} \succeq \ldots \succeq I_{C_n}, \quad (7)$$

where $'\succeq'$ denotes the relation of 'is better than or equal to', and $I_{C_i} = 0$. Actually, a distance $|I_{C_i} - I_{C_{i-1}}| = 1$, or $|I_{C_i} - I_{C_{i+1}}| = 1$, a distance $|I_{C_i} - I_{C_{i-2}}| = 2$, or $|I_{C_i} - I_{C_{i+2}}| = 2$, while a distance $|I_{C_i} - I_{C_0}| = i - 1$, and $|I_{C_i} - I_{C_n}| = n - i - 1$.

Then, the corresponding sports activity $t_{c_{i,j},l}$ is determined, where l is calculated according to the following equation:

$$l = \lfloor x_{i,j} \cdot m_{c_{i,j}} \rfloor \cdot m_{c_{i,j}}, \quad (8)$$

If the purely random selection mode is drawn, each element $x_{i,j}$ is selected simply according to the following equation:

$$x_{i,j} = \text{rand}(0, 1), \quad (9)$$

where $\text{rand}(0, 1)$ generates the random number drawn from the interval $[0, 1]$.

The evolutionary approach is intended for producing more reliable and robust data. It is powered by Differential Evolution (DE) [19] that is considered as an Evolutionary Algorithm (EA). Due to the nature of EAs, DE is population-based, and consists of Np real-coded vectors representing the candidate solutions regarding Eq. 2. Each element of the solution is in the interval $x_{i,1}^{(t)} \in [x_i^{(L)}, x_i^{(U)}]$, where $x_i^{(L)}$ and $x_i^{(U)}$ denotes the lower and upper bounds of the i-th variable, respectively.

DE guided search is controlled by three operators:

- mutation,

- crossover and

- selection.

Mutation selects two solutions randomly, and adds the scaled difference between these to the third solution. This mutation is expressed as follows:

$$\mathbf{u}_i^{(t)} = \mathbf{x}_{r1}^{(t)} + F \cdot (\mathbf{x}_{r2}^{(t)} - \mathbf{x}_{r3}^{(t)}), \text{ for } i = 1, \ldots, Np, \quad (10)$$

where F denotes the scaling factor as a positive real number that scales the rate of modification, while $r1$, $r2$, $r3$ are randomly selected values in the interval $1 \ldots Np$. Note that, typically, the interval $F \in [0.1, 1.0]$ is used in the DE community.

Uniform crossover is employed as a crossover by the DE, where the trial vector is built from parameter values copied from two different solutions. It is expressed as follows:

$$w_{i,j}^{(t+1)} = \begin{cases} u_{i,j}^{(t)} & \text{rand}_j(0, 1) \leq CR \vee j = j_{rand}, \\ x_{i,j}^{(t)} & \text{otherwise}, \end{cases} \quad (11)$$

where $CR \in [0.0, 1.0]$ controls the fraction of parameters that are copied to the trial solution. Note, the relation $j = j_{rand}$ ensures that the trial vector is different from the original solution $\mathbf{x}_i^{(t)}$.

Selection is, in fact, a generalized one-to-one selection that is expressed mathematically as follows:

$$\mathbf{x}_i^{(t+1)} = \begin{cases} \mathbf{w}_i^{(t)} & \text{if } f(\mathbf{w}_i^{(t)}) \leq f(\mathbf{x}_i^{(t)}), \\ \mathbf{x}_i^{(t)} & \text{otherwise}. \end{cases} \quad (12)$$

In fact, each element $t_{c_i,l}$ denotes the l-th training session in the c_i-th cluster. As a result, the fitness function of the DE algorithm is expressed as follows:

$$f(\mathbf{x}_i) = \sum_{j=1}^{D} \text{TRIMP}_{t_{c_i,j}}, \quad (13)$$

where $\text{TRIMP}_{t_{c_i,j}}$ denotes the TRIMP training load indicator of the corresponding training session $t_{c_i,j}$ calculated by Eq. (13).

Let us mention that parameter settings for evolutionary approach are: $NP = 20$, N_{iter} [1] $= 200$, $F = 0.5$, $CR = 0.9$.

---

[1] number of iterations

3.2 Advances and weaknesses of Sporty-DataGen

SportyDataGen is easy to use. The user may only select the required parameters and sports data is generated. To use SportyDataGen, no annoying Register/Login features are needed, and no irritating advertisements are present. It generates real data obtained from different athletes. Thus, amateur, as well as professional, athletes are taken into account. SportyDataGen relies totally on references and, therefore, does not offer randomly generated data. Generated datasets are downloaded from a server. This means that they are easy to share. Each of them obtains a unique dataset link for the purposes of citing in a publication or sharing to your friends. On the other hand, they are also easy to reuse and for replication. In any case, reviewers may be equipped by a full dataset and convince themselves by reproducing your research. The generator supports a multiple format output. Striving to help researchers with time consuming tasks is taken into account. Sports data can, therefore, be downloaded in multiple formats, including CSV. More formats will be supported in the future. Generated datasets are stored on a server permanently, and daily backups are conducted. Data are added into our database regularly. To have access for sport volunteers for donating sports data is in accordance with the mission of SportyDataGen. The current version of SportyDataGen supports biking and running sports disciplines. However, SportyDataGen could support more sports disciplines in the future, including swimming and multi-sports (e.g., triathlon, duathlon). In other words, the current version of SportyDataGen is not finished yet. Indeed, new ideas will be implemented in the coming versions of SportyDataGen.

On the contrary, some weaknesses of the proposed method also exist. Due to the nature of our volunteers, some clusters are not supported by enough data. For example, according to the clustering in 10 clusters, the cluster with high intensities is not supported by enough activities. The main problems are that athletes do not perform extremely long training sessions with high average heart rate. One of the weaknesses that may also be considered is that raw data of a particular endurance activity is not presented online. Raw endurance activity consists of all GPS track points that some researchers can use for deeper analysis. However, this part is considered for future releases of SportyDataGen. Additionally, the current version is not concentrated on deeper validation of measures produced by wearable devices. In fact, some athletes do not use the newest wearable devices. For that reason, some measures may not be totally consistent due to the problems with devices. Anyway, we consider all original measures that are available in data. In the future, we intend to validate each activity deeply before saving it to the database.

4 Examples of generated endurance sports activity collection

Table 1 presents an example of a generated endurance sports activity collection in cycling. In this collection, the first row presents the ID of an activity and all the features that were selected by the user. It is worth mentioning that the value of duration is presented in minutes, while distance is presented in kilometers. Average HR is presented as beats per minute, altitude measures are presented in meters, while calories are based on kCal. Here, endurance activities are very similar, due to the selected parameter of intensity. These data can now be imported easily as test data for many tasks, e.g. planning sport training sessions with AI tools.

5 Conclusion

The objective of this paper was to present a very complex generator of endurance sport activity collection for research that is accessible online. In the past years, analyzing sport activities that were created by sport trackers has become a very interesting research area. However, data for analysis still represent a big bottleneck for many researchers. Although many datasets have been released in the past years, they were actually collections of raw activity data downloaded from various applications that producers of various wearable tracking devices offer their customers (i.e., athletes).

In line with this, raw data limited the use of such data due to the sophisticated processing. Some researchers are non-programmers and therefore can hardly deal with programming tasks. In this case, the proposed tool SportyDataGen is unavoidable. On the other hand, it really speeds up the preprocessing process, while it's web application is easy to use. Moreover, its architecture allows users to share data easily. Users can simply select parameters and thus they are capable of generating the various outputs.

In the future, we intend to improve this generator according to the user's feedback. More generation modes are going to be supported in a future release, along with more endurance sports disciplines. The next big challenge is to support generation of triathlon datasets.

Acknowledgments

References

[1] EW Banister. Modeling elite athletic performance. *Physiological testing of elite athletes*, pages 403–424, 1991.

ID	Duration	Distance	Average HR	Average alt.	Max alt.	Calories	Ascent	Descent
1	153.72	61.45	152.00	214.83	267.00	1401.00	785.80	785.80
2	147.07	28.23	138.00	808.41	938.60	782.00	1294.40	1293.80
3	160.42	66.54	135.00	207.46	285.20	1109.00	690.80	695.40
4	164.38	22.92	132.00	207.09	274.80	1230.00	626.40	617.60
5	169.02	61.38	141.00	131.24	203.80	1558.00	228.80	226.80
6	144.60	41.06	134.00	251.24	334.60	1115.00	919.20	933.00
7	146.43	40.03	141.00	29.21	105.60	780.00	803.80	801.20
8	155.92	46.42	130.00	481.50	687.00	1520.00	1051.60	1059.20
9	143.75	33.02	144.00	795.67	919.80	766.00	1262.40	1262.80
10	148.17	42.22	119.00	224.80	288.80	851.00	627.40	636.60
11	146.35	35.55	130.00	304.37	535.60	1133.00	709.00	720.80
12	146.65	70.07	140.00	378.05	447.60	980.00	549.20	518.80
13	165.53	78.82	128.00	281.12	378.00	1069.00	985.00	988.00
14	173.92	19.50	115.00	275.66	432.40	765.00	572.40	650.40
15	151.07	67.79	120.00	406.36	628.60	1845.00	866.80	870.00
Avg.	154.47	47.67	133.27	333.13	448.49	1126.93	798.20	804.01

Table 1: Example of generated dataset using pure random approach

[2] Hana Charvátová, Aleš Procházka, Saeed Vaseghi, Oldřich Vyšata, and Martin Vališ. Gps-based analysis of physical activities using positioning and heart rate cycling data. *Signal, Image and Video Processing*, 11(2):251–258, 2017.

[3] Paolo Cintia, Luca Pappalardo, and Dino Pedreschi. " engine matters": A first large scale data driven study on cyclists' performance. In *Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on*, pages 147–153. IEEE, 2013.

[4] Dušan Fister, Samo Rauter, Iztok Fister, and Iztok Fister Jr. Generating eating plans for athletes using the particle swarm optimization. In *17th International Symposium on Computational Intelligence and Informatics (CINTI)*, pages 193–198, 2016.

[5] 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.

[6] Iztok Fister Jr. and Iztok Fister. *Generating the Training Plans Based on Existing Sports Activities Using Swarm Intelligence*, pages 79–94. Springer International Publishing, Cham, 2017.

[7] 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.

[8] Iztok Fister Jr., Samo Rauter, Dušan Fister, and Iztok Fister. A collection of sport activity datasets for data analysis and data mining 2017a.

[9] Inc. Garmin. Connect. https://connect.garmin.com/en-US/, 1996–2017. [Accessed on: December 2017].

[10] Mark J Horne. The relationship of race discipline with overall performance in sprint and standard distance triathlon age-group world championships. *International Journal of Sports Science & Coaching*, page 1747954117738878, 2017.

[11] Nicholas E Kimber, Jenny J Ross, Sue L Mason, and Dale B Speedy. Energy balance during an ironman triathlon in male and female triathletes. *International journal of sport nutrition and exercise metabolism*, 12(1):47–62, 2002.

[12] Beat Knechtle, Patrizia Knechtle, and Romuald Lepers. Participation and performance trends in ultra-triathlons from 1985 to 2009. *Scandinavian journal of medicine & science in sports*, 21(6), 2011.

[13] Romuald Lepers, Christoph A Rüst, Paul J Stapley, and Beat Knechtle. Relative improvements in endurance performance with age: evidence from 25 years of hawaii ironman racing. *Age*, 35(3):953–962, 2013.

[14] Marcos Matabuena, Mario Francisco-Fernández, and Ricardo Cao. Predicting the physiological limits of sport stress tests with functional data. In *Functional Statistics and Related Fields*, pages 179–187. Springer, 2017.

[15] Sofiane Mnadla, Nicola Luigi Bragazzi, Mehdi Rouissi, Anis Chaalali, Anna Siri, Johnny Padulo, Luca Paolo Ardigò, Francesco Brigo, Karim Chamari, and Beat Knechtle. Infodemiological data of ironman triathlon in the study period 2004–2013. *Data in brief*, 9:123–127, 2016.

[16] Oliver Neubauer, Daniel König, and Karl-Heinz Wagner. Recovery after an ironman triathlon: sustained inflammatory responses and muscular stress. *European journal of applied physiology*, 104(3):417–426, 2008.

[17] Aleš Procházka, Saeed Vaseghi, Hana Charvátová, Ondřej Ťupa, and Oldřich Vyšata. Cycling segments multimodal analysis and classification using neural networks. *Applied Sciences*, 7(6):581, 2017.

[18] Samo Rauter, Iztok Fister Jr., and Iztok Fister. A collection of sport activity files for data analysis and data mining 2016a.

[19] Rainer Storn and Kenneth Price. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *J. of Global Optimization*, 11(4):341–359, December 1997.

[20] Inc. Strava. Connecting the world's athletes. https://www.strava.com/, 2017. [Accessed on: December 2017].

[21] Inc. Suunto. Suunto MovesCount App. http://www.suunto.com/movescountapp, 2017. [Accessed on: December 2017].