THE IMPORTANCE OF MONITORING AND MAINTAINING DATA IN SPORTS TRAINING PROCESS

Fister, I. Jr.[1], Fister, K.[2], Fister, D.[1], Fister, I.[1], Rauter, S.[3]

1University of Maribor, Faculty of Electrical Engineering and Computer Science, Maribor, Slovenia

2General Hospital Rakičan, Murska Sobota, Slovenia

3University of Ljubljana, Faculty of Sport, Ljubljana, Slovenia

ABSTRACT

The sports training process is an origin of huge amounts of data obtained using mobile devices equipped with various sensors. Heart rate monitors, GPS, step counters, velocity meters, cadence meters and power meters are just some of them. If interpreted carefully, collected data offer an outstanding insight into an athlete's current form. Thus, sports trainers can recognize an athlete's abilities, fatigue or performance of implementing each sports training and, consequently, the optimal training can be predicted based on the existing sports activities. In the past few years, many computational methods were developed for analysing sports training data automatically. These methods incorporate recent knowledge from disciplines like data mining and computational intelligence. As a result, they offer an automatic sports training planning, athlete's habits discovery and food prediction regarding the sports training plans. On the other hand, storing data during the athlete's sports training sessions could provide many statistical comparisons among various generations of athletes or even sports teams. Actually, it would be interesting to know how the training plan of a marathon athlete looked over a hundred years ago. The purpose of this paper is to show the importance of monitoring, maintaining and storing data obtained during the sport training sessions. In line with this, we discuss the sports activity datasets creation process. Then, the simple and one of the more efficient ways of maintaining data is presented. Finally, the social effects of these data are exposed, where activity datasets could be distributed via the Internet for research and analysis purposes.

Keywords: *Data mining, dataset, sports activity, sports training*

INTRODUCTION

"Information is power. But like all power, there are those who want to keep it for themselves." were the first two sentences in the Guerilla Open Access Manifesto that was written by Aaron Swartz back in 2008 (Swartz, 2008). He exposed the problem where much scientific literature is in the hands of big publishing corporations and, thus, locked by paywalls. Aaron Swartz has seen the potential of the open access to the scientific literature, books and data for the common good. He fought for the ideal that everyone in the world would have open access to the common data free of charge. During his life, he has organized many campaigns that promoted the access to open data and made people aware of the importance of this subject. His visions were very important for the whole world. In the past years, many organizations began releasing open data for research purposes. Eventually, open access publishers are also attracting more and more people as their venues for publishing their research.

Nowadays, sport represents the most important marginal thing in the world. For instance, if you turn the TV on, broadcasts of various sport competitions, interviews, news, and even talk shows about sport, are presented every day. Magazines, newspapers, health organizations, universities are

promoting a healthy lifestyle and the benefits of practicing in sports. On the other hand, successful professional sport athletes become heroes with whom mass peoples are identifying. For this role, however, these heroes are paid in amounts that were unimaginable earlier. Producers of sports equipment also want to exploit the benefits offered in modern sport. In line with this, bikes have never been as improved as they are nowadays. During the sports training sessions, a lot of data have been created by an athlete using mobile devices. Currently, every athlete trains with a smart watch or smartphone that is capable of producing an enormous amount of data. Data from heart rate monitors and GPS tracks are just a few of them. On the other hand, modern trainers monitor almost every state of the athlete, their performances, health status, characteristics, habits, and so on. As a result, both sources have collected a huge amount of data characterizing the behaviour of an athlete in sports training. Approximately 15 years ago, athletes and trainers were not aware about the opportunities of using their collected data, since computers were not as available as they are nowadays. The situation now is slightly different, although both trainers and athletes still do not give sufficient attention to the power of these data.

Although data obtained during sports training sessions are produced by a multitude of athletes around the world daily, these data are typically hidden or kept secret by trainers or the athletes themselves. They are hiding data because they think that these could serve to improve the performance of their competitors and rivals. However, they are not aware that hiding data also has many weaknesses. Instead of sharing and allowing the other athletes open access to data and, thus, improving competitiveness in specific sports disciplines, these stay hidden and often untreated on their web profiles or hard disks.

After studying this situation, we have realized some facts that prevent athletes from enabling open access to their data. Indeed, there are two important facts that must be taken into account. The first fact is connected with fear, where athletes do not want to release data because of fear of their opponents. Especially, professional athletes think that they have advances over rivals, when they hide data about their sports training performance. For example, data from sport trackers may serve for revealing their training characteristics, habits, and heart rate statistics and so on. Fortunately, some professionals do not care about these problems and share their data with the others without any problems. The second fact is connected with collection and preparation of activity datasets and their later sharing. According to our interviews, it can be concluded that more athletes are not aware why their data are so important for the future progress in some sports disciplines. On the other hand, most of the athletes also do not have any experience about how to collect data, what is important, how to anonymize some parameters and, finally, how to share data. In order to break this wall, the intention of this short paper is to write a guide (a short methodology) that may help athletes and trainers to prepare and share an archive of the sports activity datasets easily.

WHAT KIND OF DATA DO WE HAVE IN SPORT?

The process of sports training is the most important task of the coach. The process consists of four stages: (1) Planning, (2) Implementation, (3) Monitoring and (4) Evaluation. Although all four phases are equally important, the majority of coaches have been focused primarily on the first two phases in the past. With the rapid development of new technologies, which allows tracking of an athlete's indicator load during the training process, the third and fourth stages in this process (i.e., monitoring and evaluation) became manageable. However, both stages cannot be considered without data that can arise from various sources. In this study, we focus on the four data sources, as follows:

- Sports tracker data,
- Laboratory measured performance data,
- Laboratory measured health data,
- Sport results data.

In the remainder of this section, the mentioned data sources available for performance analysis in sports are presented in detail.

SPORTS TRACKER DATA

Data from sport trackers (Fister Jr., Fister, I., Fister, D., & Fong, 2013) and smart watches are essential in endurance sports. Loosely speaking, sports trackers appeared with the advent of smartphones. Most of the smartphones also contain a GPS receiver that is capable of recording the current position of a person. In line with this, a lot of applications (Ferrari & Mamei, 2013; Mutijarsa, Ichwan & Utami, 2016) devoted for tracking sport activities were developed on smartphones. In 2017, there is a bunch of such applications, where the more popular ones are: Strava, Endomondo, Runkeeper, Runtastic.

Additionally, there are also special sports watches (e.g., Garmin, Suunto, Polar) that also consists of a GPS receiver. The main advance of these sports watches over the smartphones is their convenience. Watches can be mounted easily on a hand and even a bike, and are easy to transport. They can also be used in the water. For that reason, professional athletes prefer to use watches instead of taking a smartphone on training. Roughly speaking, watches and smartphones data encompass GPS data, data from heart rate monitors, cadence, power-meters and even some data that is available via Internet connection (e.g. current weather). Data from GPS receivers show an exact map where athletes are performing a sports training session, while information like total ascent, descent are used for the later analysis. The average and maximal speed can also be obtained from GPS data. Other sensors (Novatchkov & Baca, 2013; Novatchkov & Baca, 2012), like power-meters, heart rate monitors, cadence sensors, also produce more parameters about the performed training that can be used as load indicators, and are important for performance analysis (Mutijarsa et.al., 2016).

LABORATORY MEASURED PERFORMANCE DATA

Monitoring and evaluation of sports training are very important parts for an athlete in the training process. Usually, measurements are pillars of an athlete's performance. Hence, they are conducted on various time intervals in dynamic conditions. For example, during the winter, cyclists conduct sports training on cycling treadmills. These treadmills are always situated in laboratory conditions (e.g., a gym), while athletes can set some parameters on the treadmill. Parameters are intended to change the difficulty of a workout, simulate mountain conditions, etc. Treadmills are very special, because trainers can see an easy comparison with previous sessions of an athlete on the same settings.

Table 1 presents a simple example of measurements that were performed in the Kolesarski Klub Tropovci in winter 2009/10. All athletes needed to ride 6.3 km on a Tacx treadmill with slope +1. The ride was scheduled every second week in the month. As we can see from Table 1, athletes and trainers gained a lot of knowledge on the performance of athletes.

Name of competitor	Time in December	Time in January	Time in February
Jernej	9m15s	9m5s	8m59s
Jan	9m25s	9m47s	9m5s
Bojan	11m2s	10m20s	9m53s
Dusan	9m2s	8m41s	8m39s
Matej	11m45s	10m50s	10m23s
Iztok	8m55s	9m20s	8m12s

Table 1: Example of field data of simulation of time trial test on 6.3 km.

Nowadays, a lot of cyclists uses power meters and, with them, can monitor every single training. The second example of monitoring training data is a power profile (maximal mean power output values of different duration) of a cyclist during the 2014 season (Table 2). Such kind of monitoring is very helpful for athletes and coaches to compare different periods during the same season. This kind of monitoring is very helpful for planning the sports training sessions and, in some case even, for predicting the sport results in competitions.

Date	MPO 15''	MPO 1'	MPO 5'	MPO 10'	MPO 15'
Jan.	9,1	6,8	4,6	4,6	4
Feb.	9,8	7,1	5,1	4,6	4,2
Mar.	10,4	7,4	5,3	4,9	4,4
Apr.	11,4	6,7	5,0	4,9	4,2
May.	11,0	7,5	5,2	4,6	4,4
Jun.	11,8	9,1	5,7	5,2	4,5
Jul.	11,9	8,0	5,5	5,1	4,9
Aug.	11,6	8,4	5,4	5,0	4,6
Sep.	12,7	9,0	5,9	5,2	5,1
Oct.	11,2	7,3	5,6	5,3	4,6
Dec.	8,6	6,6	4,6	3,7	3,4

Table 2: Example of training data of power output values during the whole season.

LABORATORY MEASURED HEALTH DATA

Health data are very important because, on the one hand, they contribute by evaluating the athlete's current form while, on the other hand, they can be an indicator of several health symptoms. An example of the kind of health data involved: An athlete's anthropometry, body composition, blood parameters etc. (Table 3). Typically, these data are obtained in medical laboratories.

	K_HB g/dl	K_HT %	MCH pg	MCHC g/dl	MCV fL	red cells 10^6/uL	K_Lkcu 10^9g/L	S_feritin ug/L
jan.	14,5	43	32,2	33,2	95,5	4,5	4,1	41
feb	13,5	41	31	33,2	94	4,3	4,8	60
mar.	14,1	41,9	32,3	34,2	96	4,37	5	36
apr.	14,7	42,9	33,4	34,2	98	4,39	6,9	41
may	14,4	42,8	32,7	34,2	97	4,41	8,8	40
june	14,9	43,1	31,7	34,9	92	4,69	4,6	67
july	14,1	42,5	31,5	33,3	94,9	4,48	6,7	61

Table 3: Laboratory tests.

SPORT RESULTS DATA

At the end, the most important evaluation process of the sports training are the sport results of the athletes achieved during the season. Nowadays, most of the serious races can be tracked live online (e.g. IRONMAN tracker), while all race results are usually published online within the competition day.

Analyzing the athlete's results is of the utmost importance, since someone can get a lot of insight about his/her performance when comparing these with the results of the opponents.

CREATING AN ARCHIVE OF SPORTS ACTIVITIES DATASETS

Some years ago, we were trying to perform some sport related research involving data from sports trackers. However, we have quickly identified that there is a lack of these files for serious research. Hence, there was not any special possibility then to begin collecting the sports activity datasets produced by athletes in various sports disciplines (e.g., cycling, running). We began with extensive calling of various athletes (amateurs, as well as professionals, participating in various endurance sports) in order to give us some data for research. Mainly, the aim was to create our own archive of the sports activity datasets that can be used openly for the general public. After some weeks, we received some valuable data of athletes that were well-formed and suitable for research. We downloaded data from Garmin Connect and Strava profiles of those athletes who allowed us access to their data. At the beginning, we released the data in TCX and GPX format while, later, we delivered data only in GPX format. Guidelines of how to deal with these data are available here (Rauter, Fister I.,
& Fister Jr. I, 2016), while a deeper description of data for the reader can be found in the following technical reports (Rauter, Fister I & Fister Jr. I, 2015; Rauter, Fister I & Fister Jr. I 2016).

PRACTICAL GUIDANCE FOR COLLECTING, STORING AND DISTRIBUTING DATA IN SPORT

Many athletes that we started to interview firstly raised the following questions: "Who will use my data?", "I am not a superstar. Therefore, why is my data so important for you?" Unfortunately, these athletes were not aware that these data are part of a very complex system and, therefore, very important for developing new tools enabling deep performance analysis. Today's computer can process enormous amounts of data. Hence, there is no problem if someone can provide some more MB's of sports activity datasets. All these data serve a common purpose, i.e., to make a huge archive of the sports activity datasets dedicated to current or future performance analysis.

A lot of professional athletes, especially in team sports, have realized that, after their career, they do not have any data with which ups and downs during their careers could be analysed. For instance, the professional volleyball player Claudio Carletti mentioned in his talk (February 2017) in Maribor that, after more than 20 years of playing volleyball he, in fact, did not have any data for describing his career in detail. He realized that he has no information about his training plans, no health data, no match statistics and even no recorded matches. Thus, he mentions that, today, he is like nobody since, when he left the club in which he played the last seasons, it erased all its data. After realizing these problems, he began to make people aware of collecting their sports data for later use. Additionally, he also started a big project, Vanda (https://github.com/CarlSpobble/vanda), where he wishes to collect sport data together to the archive and use it. Claudio's story is just one of the many stories that show us how important the sports data are. This means that collecting the sports activity datasets should start already at young ages.

The next step for someone who wishes to collect sports activity datasets, is to transform these data into a universal format. Usually, athletes and trainers are collecting all data in Excel tables or Word documents. Although Excel files are a very good tool, because of portability and easy processing, dealing with sport tracker data can be performed using many web applications available today. In this case, only download of this data needs to be conducted.

When sports activity datasets are collected into an archive, the easiest way to make these data accessible publicly is to put them on the Internet or save them into other storage warehouses, e.g.

Dropbox. However, the archive needs to be well documented in order to enable people to start work on this data.

CONCLUSION

This paper describes briefly creating and exploring the archiving of valuable sports data that are generated during the process of sport training. Some years ago, nobody had given any special attention to these data. With the rise of data mining and artificial intelligence methods, data have also been becoming very attractive for research purposes (e.g., performance analysis of realized sports training sessions).

In line with this, we have exposed systematically various types of sports data arising during the sports training and present a practical guidance for collecting, storing and distributing data arising in sports. The guidance is devoted to potential athletes that should contribute their sports data into the archive for the future research purposes.

REFERENCES

Fister Jr, I., Fister, I., Fister, D., & Fong, S. (2013). Data mining in sporting activities created by sports trackers. In *Computational and Business Intelligence (ISCBI), 2013 International Symposium on*, 88–91.

Swartz A (2008). Guerilla open access manifesto. *Online-Ressource, URL http://www. openeverything. eu/guerilla-open-access-manifest*.

Rauter, S., Fister, I., & Fister Jr, I. (2015). 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.

Rauter, S., Fister Jr, I., & Fister, I. (2015). A collection of sport activity files for data analysis and data mining. Ver 12.05. Ljubljana: University of Ljubljana.

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

Novatchkov, H., & Baca, A. (2013). Artificial intelligence in sports on the example of weight training. *Journal of sports science & medicine*, *12*(1), 27.

Novatchkov, H., & Baca, A. (2012). Machine learning methods for the automatic evaluation of exercises on sensor-equipped weight training machines. *Procedia Engineering*, *34*, 562-567.

Ferrari, L., & Mamei M. (2013). "Identifying and understanding urban sport areas using nokia sports tracker." *Pervasive and Mobile Computing* 9(5), 616-628.

Mutijarsa, K., Ichwan, M., & Utami, D. B. (2016). Heart rate prediction based on cycling cadence using feedforward neural network. In *Computer, Control, Informatics and its Applications (IC3INA), 2016 International Conference on* (pp. 72-76). IEEE.