

Data Mining in Sporting Activities Created by Sports Trackers

Iztok Fister Jr, Iztok Fister, Dušan Fister

University of Maribor

Faculty of Electrical Engineering and Computer Science

Smetanova 17 - Maribor, Slovenia

e-mail: iztok.fister@guest.arnes.si

Simon Fong

University of Macau

Faculty of Science and Technology

Av. Padre Tomas Pereira, Taipa, Macau SAR

e-mail: ccfong@umac.mo

Abstract—Pervasive technologies allow people to get information from everywhere and at anytime. The rapid development of smart mobile devices on the one hand and the progress of wireless networks on the other, has led to the widespread growth of pervasive technologies. Pervasive devices are small pieces of hardware that allow users to communicate over the Internet. Due to their capabilities, which enable users to develop the software they need, many applications for mobile devices have emerged. In this article, we will present data mining in sporting activities using mobile sport trackers. Each tracked sporting activity is saved into a file. Data mining could be applied to the collection of these files, which would help sportsmen analyze their workouts, predict their further training activities, advise about nutrition, etc.

Keywords: *sports activity; GPS; data mining;*

I. INTRODUCTION

Together with the ascendancy of smart mobile devices, mobile-applications have been developing rapidly [1]. In the beginning, only a few applications were executed on desktop computers. Nowadays, a lot of online applications, like news bloggers and many different social networks, have become part of our everyday life. These applications, which were primarily dedicated to running on desktop computers, have been steadily ported from this static platform to mobile platforms. In the year 2007, applications for tracking and navigating during different kinds of sport activities emerged (e.g., Endomondo, Runtastic, Sports Tracker, Nokia Sport etc.).

Sport trackers track and monitor sporting activities in real time. Thus, when an athlete moves during their sporting activity, information pertaining to the duration time, length and altitude of the course are recorded. For this, only a mobile device, including a GPS device, is needed. In order to monitor data about training activities online by trainer, a wireless communication over the Internet must be available. However, the training data is stored into a file in GPS exchange format (GPX). The GPX format is designed to describe waypoints, tracks, and routes, which is useful for different software applications. For example, the Google Maps application on the Internet allows users to track their route on a site map during a specific activity (cycling,

running, driving, etc.). Moreover, these files can be uploaded onto a web server and analyzed later in detail or shared with friends over social networks in order to compare their achievements with one another.

In modern sports, athletes hire personal trainers to worry about their training, draw attention to the training errors made in the past, and even prescribe a nutritional regimen, all in order to improve their results. Using personal trainers is not only typical for professionals but also for amateur sportsmen. However, the cost of this service on average amounts to more than a thousand Euros per month.

The motivation for our work is to build an automated personal trainer that would be able to aggregate data from GPX files and on the basis of an analysis advises athletes about what kind of training is recommended for them in the near and distant future. This training regimen should be based on the current readiness of the athlete, which sport (in the case of multiple sports) represents a potential weak point that should be improved before the start of a real competition, and what results can be expected by the athlete with regard to their history of results in training. Additionally, it should propose a diet that would improve the athlete's form according to the psycho-physical predispositions of athletes, etc.

At this moment, we are at the beginning of our research. Therefore this article focuses primarily on the designing of the methods that will be the most appropriate to accomplish all of the mentioned goals, whilst the implementation of the automated personal trainer algorithm is left for future work. In line with this, the data mining methods are selected in our planning phase to design this algorithm. Two methods of analyzing the large quantity of GPX files are proposed, i.e. clustering and classification. The former divides input files into clusters according to training disciplines, characteristics of the file owner, and the duration of training activities. The latter allows us to classify GPX files according to the type of training activity, whereas lower activity is followed by higher level ones in terms of time intervals and endurance. Lower level activities, by contrast, extend over a longer period of time and base, where the basic training activities are performed, etc.

On a basis of these classified GPX files, a prediction is made, where the type of training activity is proposed according to the characteristics of the sportsmen. These characteristics are obtained during the learning (adaptive) process, where the behavior of individuals in different, mostly stressed, situation can be predicted.

The structure of this article is as follows. In the next section, some information about the sports tracker is introduced. Section III reveals the design of an automated personal trainer using data mining methods on GPX files obtained by tracking sporting activities. At first, the data mining methods are represented in general then the first steps in clustering and classification of GPX files are described. The paper concludes with a discussion of the performed work, while directions for future work are outlined.

II. SPORTS TRACKERS

People have always liked entertainment. For example, some of them go to the cinema, some to the ballet and some practice sports. The roots of sports go back to ancient Greece, e.g. the amateur Olympics. Through different civilizations, sport has thrived to this day [2].

A lot of people only practice sports for their health and fun, but some of them are professionals and participate in order to earn money. According to the number of individuals, sports could be divided into: individual sports, like biking swimming, running, fencing, dancing, and team sports, like basketball, football, rugby, American football, etc. Sports became even more popular when mass sporting competitions began [3]. On the other hand, sports have also become more approachable for amateur athletes, who can compete for medals together with professionals in some sporting competitions, like triathlons [4].

Typically, tracking training activities cannot be imagined nowadays without the use of mobile applications running on mobile devices. These applications can help athletes stay healthy, regenerate after a long lasting activity, and remain fit. Many of these can even prescribe future training activities but in a limited scope. In the rest of this article, a short review of contemporary sports trackers will be provided.

The first family of devices for improving human performance was the heart rate monitor (the popular Finnish company Polar was the first one to sell it). The heart rate monitor consists of a belt and a hand watch. The belt is usually fixed around the chest. When measuring, it transmits signals to the watch, which then displays the corresponding data on the screen. Unfortunately, these devices have several shortcomings. First, only a few functions were implemented, e.g., maximum, minimum and average rate. Second, the watches for bikers allowed users to measure their speed according to the average length of the wheel and the measured time. However, the average length of the wheel depends on the pressure and type of tires. As a result, measuring speed can only be approximate. Finally, the distance of the running activities is estimated according to

the average step of the runner. This estimate is also unreliable because the step is different when running over flat terrain than when running uphill.

The next generation of mobile sport tracker devices was invented around the year 2005. These devices are based on GPS and play a great role in today's sporting world. They were worn as watches and were able to measure time and distance very precisely. For bikers, this means that they do not need to install any sensors on their bikes and swimmers can also measure their distance in lakes. Sports watches became portable, good-looking, and provide almost any information needed by athletes during competitions. Essentially, the performed activities can be uploaded onto the Internet for further analysis.

The newest family of sports trackers follows the great expansion of wireless technology that enables real-time support. For example, the tracker of DOS-RAS Extreme cycling race enables the tracking of competitors in real-time over the Internet. Nowadays, sports trackers are a piece of software that run on a several mobile devices, e.g., Nokia Sports Tracker [5], Blackberry, iPhone, Android, etc. Information about its location are obtained from GPS satellites, saved temporarily into memory, and transferred to the server's database over the Internet. The server can then analyze this data online, show graphs, compute the fastest pace, burned calories, altitude, maximum heart rate, etc. It can also conduct many popular sports tests, like Conconi.

Unfortunately, none of these services have so far been applied in order to build an application that could act as a personal trainer using data mining methods. Usually, athletes can have more than 2,000 sport's activities saved in GPX files. In fact, this amount of data files now represents a large amount of data. Such an amount of data cannot be analyzed and predicted without using data mining methods. Using these methods, predictions for trainings, injuries, adjusting, breaking and improving form could be set up automatically. Here, the use of an automated personal trainer is inevitable.

What is the future for sports? In sports, cheating is performed via doping. The most abused sports for achieving better performance are cycling, football, athletics, etc. In order to maintain the sport as clean as possible, an automated personal trainer could help analyze the results of athletes and detect any inconsistencies in the sudden improvement of their results in a competition, which could indicate that the athlete might be using banned substances.

III. THE DESIGN OF AN AUTOMATED PERSONAL TRAINER

Data mining is a very appropriate method for processing GPX sport activity files. Currently, there were some attempts in applying data mining to sporting activities [7, 8]. Some examples are evidence-based educational model by P. Hung [6] which uses video-camera to capture every moment of a human's movement activities for analysis.

A. Data Mining

There are many data mining techniques that have been developed in the past and which are now very useful in data mining projects:

Association: here, patterns are discovered based on the relationship between items in the same transaction. E.g.: We have a shop and customers. With association we can analyze and identify the products that customers frequently purchase together,

Classification: this is a data mining method based on machine learning. Usually, we use decision trees, neural networks, ensemble methods, and evolutionary algorithms. Here, we develop programs which are able to learn how to classify items into groups.

Clustering: this is a data mining method that creates a cluster of objects which have similar characteristics. Clustering defines classes and then puts objects in every class.

Prediction is a technique that discovers the relationship between dependent and independent variables.

B. First Steps in the Data Mining of Sports Activities

Clustering by:

- Sporting activities: to specify the type of activities between 1,000 unknown files;

- Performance: some athletes are top professionals, and some are amateurs; data mining tries to determine their characteristics;

- Duration: to determine the differences between long and short training activities;

Classification of training:

- Type: endurance, interval, base, slow, calculating their performance and effectiveness.

Predictions of:

- Type: sportsmen in a triathlon go through 15 training activities per week, therefore, data mining tries to determine the best type of training;

- Form: how to be in the best form on the day of the competition;

- Duration training: how to train the most effectively;

- Interval training: a lower level activity is followed by higher level ones for certain time intervals in order to train both types of muscle, i.e. quicker ones are more appropriate

for sprints while resistance is more appropriate for long-term competitions:

- Best heart rate through race: how to run a half-marathon with the most effective heart rate

Rule Learning:

- Aims: to determine the habits of athletes, i.e., one prefers endurance, one interval, and one enjoys both types of training activities.

Association Rule Learning:

Every athlete has a little different training habit. Some athletes train one activity per day, but others do more than one activity. Association rule learning in sport activities might be suitable in the next way. Let us have a data from one multisport athlete, who was training: swimming, biking, running, mountain-biking, climbing, Nordic skiing and other sports. Moreover, this athlete had more than one training per day.

With association rule learning we can determine, which training our athlete combined daily. For example, he usually liked to combine swim, run and climbing in one day.

C. First Step in Clustering

The proposed algorithm of Data Mining in Sport Activities consists of:

- Preprocessing: the processing of files and calculating the length, duration, pace, fastest 100 meters, altitude, etc.;

- Creating and applying rules for recognition: every sport has its own specifications. For example:

- Swimming: duration between 5 min and up to 2 hours, with a length from 100 m up to 8 km, fastest speed from 70 meters per minute up to 100 meters per minute, altitude from 0 up to 15 meters.

- Cycling: duration 10 min up to 10 hours (ultra), length from 10 km up to 300 km, fastest pace 60 km/h, altitude from 10 m up to 5km.

- Running: duration between 10 min up to 3 hours, length from 400 m up to 50 km, ultra marathons more than 50 km, with fastest speed 2:50 minutes per km by professionals, altitude from 0 up to 300 meters, except mountains run, which are rare.

- Classification: creating rules and preprocessing are saved to a flexible database, e.g., Apache Cassandra.

- Presentation of results: after some time, the algorithm gives us results with activity IDs, being sorted into different types

and being eliminated, if not satisfying conditions. The results are presented in some appropriate database tables.

IV. CONCLUSION

In this paper, we presented a problem that might be interesting in the near future, i.e., how to improve the results of athletes on the basis of their training activities, accumulated in GPX files. In line with this, an automated personal trainer is proposed that can analyze a large amount of data in GPX files using the clustering and classification data mining methods in order to make predictions for future training activities by athletes. These predictions are made based on their characteristics, which can be learned by the automated personal trainer using machine learning methods.

The automated personal trainer is especially useful for amateur sportsmen that cannot afford the cost of their own personal trainer.

REFERENCES

- [1] I. Fister in I. J. Fister, "A concept of drafting detection system in Ironmans," *Elektrotehnikni review*, vol. 78, no. 4, pp.217–222, 2011.
- [2] I. J. Fister, M. Mernik, I. Fister in D. Hrnčič, "Implementation of EasyTime Formal Semantics using a LISA Compiler Generator," *Computer Science and Information Systems*, vol. 9, no. 3, pp.1019–1044, 2012.
- [3] Samo Rauter and Mojca Doupona Topič, "Sports became even more popular when mass sporting", *Kinesiology*, Vol.43, 2011, Issue (1), pp.82-90.
- [4] Fister Iztok Jr., Fister Iztok, Mernik Marjan, Brest Janez, "Design and implementation of domain-specific language Easytime", *Computer Languages, Systems and Structures*, Volume 37, Issue 4, October 2011, pp.151–167.
- [5] L. Ferrari in M. Mamei, "Identifying and understanding urban sport areas using Nokia Sports Tracker," *Pervasive and Mobile Computing*, št., In press, 2012.
- [6] Jayshiro Tashiro, Patrick C. K. Hung, Miguel Vargas Martin, "Evidence-Based Educational Practices and a Theoretical Framework for Hybrid Learning", *ICHL 2011*, pp.51-72.
- [7] Leonardo De Marchi, "Data mining of sports performance data", *Erasmus computing 2010/2011*, University of Leeds, School of Computing Studies, 2011, 132 pages.
- [8] Jae Young Lee, Hoff, W., "Activity Identification Utilizing Data Mining Techniques", *IEEE Workshop on Motion and Video Computing (WMVC 2007)*, Feb. 2007, ISBN: 0-7695-2793-0, Austin, TX, USA, pp.20.