Am I overtraining? A novel data mining approach for avoiding overtraining

Iztok Fister Jr.
University of Maribor
Faculty of Electrical Engineering and Computer Science
Smetanova 17, 2000 Maribor
Slovenia
iztok.fister1@um.si

Goran Hrovat
University of Maribor
Faculty of Electrical Engineering and Computer Science
Smetanova 17, 2000 Maribor
Slovenia
goran.hrovat@um.si

Samo Rauter
University of Ljubljana
Faculty of Sport
Gortanova 22, 1000 Ljubljana
Slovenia
samo.rauter@fsp.uni-lj.si

Iztok Fister
University of Maribor
Faculty of Electrical Engineering and Computer Science
Smetanova 17, 2000 Maribor
Slovenia
iztok.fister@um.si

ABSTRACT

Overtraining is one of the biggest problems in process of sport training. Especially, freshmen's and amateur athletes who do not have enough knowledge about behavior of their body, training and who do not have personal trainers encounter overtraining. So far some theoretical and practical solutions for avoiding overtraining have been arisen. In this paper, we propose a novel automatic solution which helps athletes to avoid overtraining. It is based on data mining which is applied to athlete's workout history. First experiments and simulations showed the promising results.

Keywords

data mining, apriori, sport training, overtraining

1. INTRODUCTION

Recently, an expansion of massive sport events and competitions have been emerged. These events capture usually sport disciplines, like running, cycling and triathlon [1]. Consequently, more people joined to a sport and healthy lifestyle. In fact, a lot of health and medicine organizations in the world also encourage people to participate in such events and live a healthy lifestyle. When people start to deal with sport, they do not have any special goals. There, only one goal is presented, i.e., I have to finish a specific sport race/event in which I invested so much effort! However, when someone finish more races, he/she want to go forward and compare his/her achievement with other people. This stage naturally leads to a competition. To compete in sport and fight for a medals is not easy anymore. Firstly, athletes have a devotion to the sport and in line with this assume themselves tremendous amount of work. Secondly, they have to build/create a good and effective training sessions. To finish race or to win a race is a huge difference and this difference is usually reflected in training, eating and resting. When amateur athletes start to deal with a sport, they do not have enough knowledge about these three components of each preparation for sport competitions:

- Training: Usually, the following set questions is asked by athletes. What is a proper training? When I have to train? How much to train? What kind of training should I train? Are intervals better than long distance training? Where should I train? With whom should I train?
- Eating: The following question are to be answered in line with this. What I have to eat? Why I need to eat so much carbohydrates? Do I need to increase vegetables and fruits intake? Is it good to drink any powders or isotonic drinks? Why do I need to avoid eating sugar? Is it good to take additional proteins into my food? Why eating a lot of fish is useful?
- Resting: Here, the following questions could be emerged, e.g., Do I need to take a rest every week? When I need rest and free days? What is a good resting? Can I still have a simple and short run/cycle? Swimming helps for relaxation of muscles and body?

These three components define well trained athlete. Training and resting are equally important. If you train a lot you get overtraining. On the other hand, you cannot be trained enough and suffer in the races. Finally, eating has a great

impact to athlete's performance. It is important for recovery and regeneration. Nowadays, overtraining is still one of the biggest problems in sport. Almost every freshman encounter this problem. When freshman encounter this problem, they usually suffer for psychological problems in mind. Moreover, it is very hard to overtake this problem. So far many theoretical, practical and abstract solutions have been developed to deal with this problem. In this paper we propose a new solution to help athletes to avoid overtraining. Data mining methods are applied on athlete's workouts which were created by sport trackers. Our solution send an alarm to athletes before overtraining. Therefore, they can change a training or take a more rest and prevent themselves from overtraining. The paper is structured as follows: in the next section we describe some basics about sport training. The third section outlines overtraining, while section 4 reviews current overtraining prevention methods. Section 5 describes data mining methods. Section 6 is devoted to the sport trackers. Section 7 proposes a novel solution and the last section conclude the paper with some remarks for future work.

2. SPORT TRAINING

Definition of Sports Training is based on scientific and pedagogical principles of planning the systematic activities or processes in order to achieve the highest results in the chosen sports discipline. The final effect of the systematic training process can be manifested as

1. athlete's well sport form,
2. the increased capacity of the athlete's organism, or in the worst case
3. overtraining syndrome.

Sport form can be described as a phenomenon of short-term increased capacity of the athletes organism in relation to the expected competitive capacity that is perceived on the subjective level. In such conditions athlete feels that he/she overcomes a certain load of sport activity with a little effort or that he/she is able to overcome a higher load at the maximum effort [2]. In simple terms, it means "to be in the best shape at the right time." Efficiency of the process of sports training also means that the athlete rationalize the time devoted to the training [3]. All this is happening in a real process of sports training, where in order to achieve sports form, athletes include different methods in the process of sport training and intensity of workload. Also the rest period is very important. For instance, Table 1 presents an example of proper training two weeks before half marathon race.

2.1 Overtraining

Overtraining is a special state of an athlete occuring when someone undergoes a bigger amount of intensity training and consequently body is not able to recover. In such state, athletes are not able to compete on high level because their body is exhausted. Overtraining leads to overtraining syndrome. Overtraining syndrome manifests itself in physiological and psychological symptons and has an effect on athlete's performance [4]. These physiological and psychological indicators as proposed in [4] and presented in Table 2.

Physiological	Psychological
Higher resting heart rate	Sleep disturbances
Changes in normal blood pressure	Loss of self-confidence
Delayed return to normal heart rate	Drowsiness and apathy
Elevated basal metabolic rate	Irritability
Elevated body temperature	Emotional and motivational imbalance
Weight loss/excessive thirst	Excessive, prolonged weariness
Impedded respiration	Lack of appetite
Subcostal aching	Fatigue
Bowel disorders	Depression
	Anxiety
	Anger/hostility
	Confusion

Table 2: Physiological and Psychological indicators of overtraining syndrome

	Causes
1	Length of the competitive season
2	Monotony of training
3	Feelings of claustrophobia
4	Lack of positive reinforcement
5	Feelings of helplessness
6	Abusiveness from authorities
7	Stringent rules
8	High levels of competitive stress

Table 3: Causes of overtraining syndrome as proposed in [4]

As it can be seen from Table 2, there are many indicators telling the athletes are overtrained. However, better than curing overtraining is avoiding overtraining. In [4], authors presented some causes of overtraining syndrome.

A lot of research were done on overtraining in the past, e.g., in [5, 6, 7, 8, 9].

2.2 Current methods of avoiding overtraining

The main herald of overtraining syndrome is a hard training and an inadequacy of the rest. People transferred training or competition stress differently. Athlete tolerance on training stress varies throughout the season periods. In line with this, a training process should be adapted and varied through the season period. Especially, the right strategy in the competition period is very important. Too much high intensity workout with too short rest period may results in bad results at the competition or in the worst case may leads to overtraining. To avoid overtraining syndrome is not so easy. Athletes are most of the time very tired. It is hard to distinguish when the first signs of fatigue happen. Prevention requires good nutrition, complete hydration, and rest periods between exercises themselves [7]. The effective and good organize periodization process is necessary to ensure adequate adaption of the athlete organism to the requirements of the

DAY	METHODS	DURATION	INTENSITY
1	Race		
2	Easy jogging with some acceleration	30 min	Low
3	Rest day		
4	Running (interval training 2 x 10 min fast; between sets 5 min - 8 min of easy jogging)	45 min	Low & High
5	Rest day		
6	Running (interval training 4 x 4 min fast; between sets 3 min - 4 min of easy jogging)	45 min	Low & High
7	Running (interval training 2 sets of 3 - 5 x 90 sec accelaration)	45 min	Low & High
8	Rest day		
9	Rest day		
10	Endurance running	90 min	Low
11	Endurance running	75 min	Low
12	Rest day		
13	Endurance running	105 min	Low
14	Running (interval training 2 sets of 3 x 2 min very fast with 1 min recovery; between sets 5 min - 8 min of easy jogging)	75 min	Low & High

Table 1: Example of training for 21 km race (last 14 days before the competition)

competition in chosen sport discipline.

To avoid overtraining syndrome it can help an individual monitoring of:

- achievements,
- mood states and
- heart rate variability.

To overcome the overtraining syndrome it can help:

- low intensity workout, rest and relaxation strategy;
- exercise with very short high intensity sprints with long rest/low intensity period, the confidence to progress well.

2.3 Monitoring the sport activities

Recently, training sessions of athletes are monitored with sport trackers. Sport trackers are a combination of software and hardware which are used for tracking activities during sport workouts. Hardware is a smart phone, while software are applications which run on smart phones. For a more detailed description of sport trackers, readers are invited to read papers [10, 11].

3. DATA MINING OF MONITORED DATA

Data obtained during the sport session of athletes can be transferred to personal computers in a form of TCX data sets. Normally, these data sets are analyzed using different data mining methods. In computer science, the data mining (DM) is a method, where the main goal is extraction of information from a data sets and conversion of these data to a form understandable by humans. The more frequently used methods in data mining are: clustering, classification, regression, association rule mining. Data mining methods involve also some machine learning methods like decision trees, logistic regression, support vector machines, etc. But nowadays also some nature-inspired algorithms [12, 13] are employed in the data mining field. In data mining, a broad spectrum of many applications has been developed, e.g., these methods have been used for weather prediction, fraud detection, sensor data mining, surveillance and many more.

3.1 Association rule mining

One of the data mining tasks is association rule mining. It was discovered by Agrawal and Srikant [14, 15] who developed Apriori. Other algorithms for association rule mining also emerged e.g. FP-growth [16], which perform faster. Association rule mining is a task of discovering rules from set of transactions. The well known example is basket market analysis however association rule mining can be applied for other datasets such as sports data. Each transaction is consisted of items and is also called item-set. Discovered rules are the form $X \Rightarrow Y$, where X is subset of items and Y is usually only one item. As a result many rules are discovered, for this reason a lot of measures are proposed to evaluate and rank these rules. The most known interestingness measures are support and confidence [14], others are lift [17], all-confidence [18], conviction, leverage, χ^2 [19] etc. Support is proportion of transactions containing items of a rule and confidence is defined as $conf(X \Rightarrow Y) = supp(X \Rightarrow Y)/supp(X)$. In our case transactions is consisted of items e.g. training distance, average heart rate, motivation, eat-

ing, sleeping, etc., and an example of rule which can be discovered is $('MOTIVATION','EATING') \Rightarrow ('SLEEPING', 1.000)$, where confidence and support are 1 and means that everyone who well eats and is motivated also well sleeps.

4. PROPOSED METHOD FOR AVOIDING THE OVERTRAINING

Our proposed method consists of the following steps:

- definition and design of framework,
- discretization,
- creating transactions,
- applying Apriori algorithm and
- interpretation of results.

In the remainder of the paper, these steps are illustrated in detail.

4.1 Definition and design of framework

At first, a framework needs to be defined that enable an athlete to store, maintain and use data. Let us suppose that after every training session athlete uploads a file in TCX format. This file is automatically parsed and data (total duration, total distance, average heart rate, max heart rate) are stored into the database. During uploading the file, athlete needs to answer to the five predefined questions that characterize current training session. These questions are:

1. Did athlete train intervals? (Possible answer values: YES, NO)
2. Did athlete feel any special pains during the training session? (Values: YES, NO)
3. Do athlete have a great motivation for future trainings? (Values: YES, NO)
4. Did athlete sleep well today? (Values: YES, NO)
5. Can atlete eat normally? (Values: YES, NO)

When athlete answers to these questions, he/she is linked with today's workout.

4.2 Discretization

When all data are collected a discretization process is performed. In this process, numerical values have to be discretized, since Apriori algorithm works only with categorical values. An example of discretized attributes is presented in Table 4.

4.3 Creating transactions

After attributes are discretized, a transactions are created. In our case, we consider one transaction for one training. Therefore, after 2 months of training, there will be around 45 transactions. More transactions we have, more accurate rules can be inferred and in line with this also the overtraining can be predicted more precisely.

4.4 Applying Apriori algorithm

In this step, Apriori algorithm is launched using the prepared data (transactions). Apriori algorithm discovers rules as a result. For example, a rule $('MOTIVATION', 'EATING') \Rightarrow ('SLEEPING', 1.000)$ means that athlete who has a motivation and can eat well, can also sleep well.

4.5 Interpretation of results

The last step in the proposed approach is an interpretation of results. From the rules which we get with Apriori algorithm, we want to get a real situation.

5. EXPERIMENTS AND SPORT INTERPRETATION OF RULES

The first experiments of the proposed method for predicting the overtraining were conducted on a real dataset. This dataset was produced during the training sessions of a professional mountain biker. The total number of workouts/transactions were limited to 50. In Figure 1, some transactions are presented that were used in our experiments.

```
SHORT_RIDE,SHORT_DURATION,HIGH_RATE,INTERVAL,NO_PAINS,MOTIVATION,NO_SLEEPING,EATING
MEDIUM_RIDE,MEDIUM_DURATION,MEDIUM_RATE,NO_INTERVAL,NO_PAINS, MOTIVATION,NO_SLEEPING,NO_EATING
LONG_RIDE,LONG_DURATION,MEDIUM_RATE,NO_INTERVAL,NO_PAINS,MOTIVATION, SLEEPING,EATING
MEDIUM_RIDE,SHORT_DURATION,HIGH_RATE,INTERVAL, NO_PAINS,MOTIVATION,SLEEPING,NO_EATING
LONG_RIDE,LONG_DURATION,LOW_RATE,NO_INTERVAL,NO_PAINS,MOTIVATION,SLEEPING,NO_EATING
```

Figure 1: Example of transactions

From experiments, interesting results were obtained. For instance, the rules are inferred from the transactions, as follows.

1. Rule: $('EATING',) \Rightarrow ('MOTIVATION', 1.000)$ Interpretation: From this rule we can see that proper eating is connected with motivation. If we eat a lot we also have a motivation.
2. Rule: $('NO\_MOTIVATION',) \Rightarrow ('NO\_SLEEPING', 1.000)$ If athlete does not have a motivation, he can not sleep. It might means that he is overtrained.
3. Rule: $('LONG\_RIDE',) \Rightarrow ('NO\_INTERVAL', 1.000)$ From this rule we can see what is usually in practice. If we ride a long rides, we do not do intervals, but going the distance only.
4. Rule: $('SLEEPING',) \Rightarrow ('MOTIVATION', 1.000)$ Here we see can again that sleeping and motivation are connected. If athlete can sleep well, then he is very motivated and for sure not overtrained.
5. Rule: $('NO\_MOTIVATION',) \Rightarrow ('NO\_EATING','NO\_SLEEPING', 1.000)$ In this rule we see that athlete who does not have a motivation, can not eat and sleep. In this case, he is very overtrained.
6. Rule: $('LONG\_RIDE','NO\_EATING') \Rightarrow ('NO\_INTERVAL', 1.000)$ This rule defines long ride and that athlete can not eat. We can also see that in this case athlete were not doing an intervals.

From previous rules, it can be seen some rules which were discovered from a real dataset. In some cases we can see how to determine overtraining in rules. However, after running

Attribute	Current values	Discretized values	How to discretize?
Distance	Numerical (km)	LONG_RIDE MEDIUM_RIDE SHORT_RIDE	> 120 km ≤ 120 km and > 50 km ≤ 50 km
Duration	Numerical (min)	LONG_DURATION MEDIUM_DURATION SHORT_DURATION	> 300 min ≤ 300 min and > 150 min ≤ 150 min
Average heart rate	Numerical (BPM)	HIGH_RATE MEDIUM_RATE LOW_RATE	> 170 BPM ≤ 170 BPM and > 130 BPM ≤ 130 BPM
Intervals training?	Categorical (Yes, No)	INTERVAL NO_INTERVAL	
Pains?	Categorical (Yes, No)	PAINS NO_PAINS	
Motivation?	Categorical (Yes, No)	MOTIVATION NO_MOTIVATION	
Sleeping?	Categorical (Yes, No)	GOOD_SLEEPING BAD_SLEEPING	
Eating?	Categorical (Yes, No)	GOOD_EATING BAD_EATING	

Table 4: Discretization of our attributes

LONG_RIDE	MEDIUM_DURATION	HIGH_RATE	PAINS	MOTIVATION	GOOD_SLEEPING	NO_INTERVAL	BAD_EATING

Table 5: An example of one transaction

more experiments we encounter that more transactions we have, more rules we can obtain. In line with this, we do not obtain only rules connected with overtraining, but also some others tackling the basic habits of athletes, like rule 1. On the other hand, rule 5 shows a basic example of overtraining.

6. CONCLUSION

In this paper we developed a prototype solution for predicting overtraining of an athlete. We used an Apriori data mining method and applied it to the real dataset that was created by mountain biker. Experiments showed that this method is interesting for such manners. However there are also some limitations like size of the dataset. In our case we took 50 trainings into the account to get the first picture of an athlete. In the future, we will also perform more tests with different athletes.

7. REFERENCES

[1] Samo Rauter. Mass sports events as a way of life (differences between the participants in a cycling and a running event). *Kinesiologica Slovenica*, 2014.

[2] Kuno Hottenrott, Sebastian Ludyga, and Stephan Schulze. Effects of high intensity training and continuous endurance training on aerobic capacity and body composition in recreationally active runners. *Journal of sports science & medicine*, 11(3):483, 2012.

[3] Riggs J Klika, Mark S Alderdice, John J Kvale, and Jay T Kearney. Efficacy of cycling training based on a power field test. *The Journal of Strength & Conditioning Research*, 21(1):265–269, 2007.

[4] Mary Black Johnson and Steven M Thiese. A review of overtraining syndromeâĂŤrecognizing the signs and symptoms. *Journal of athletic training*, 27(4):352, 1992.

[5] Laurel T MacKinnon. Overtraining effects on immunity and performance in athletes. *Immunology and cell biology*, 78(5):502–509, 2000.

[6] Shona L Halson and Asker E Jeukendrup. Does overtraining exist? *Sports medicine*, 34(14):967–981, 2004.

[7] Lucille Lakier Smith. Overtraining, excessive exercise, and altered immunity. *Sports Medicine*, 33(5):347–364, 2003.

[8] Dianna Purvis, Stephen Gonsalves, and Patricia A Deuster. Physiological and psychological fatigue in extreme conditions: overtraining and elite athletes. *PM&R*, 2(5):442–450, 2010.

[9] Karen Birch and Keith George. Overtraining the female athlete. *Journal of Bodywork and Movement Therapies*, 3(1):24–29, 1999.

[10] Iztok Fister, Duan 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.

[11] Iztok Fister, Dušan Fister, Simon Fong, and Iztok Fister Jr. Widespread mobile devices in applications for real-time drafting detection in triathlons. *Journal of Emerging Technologies in Web Intelligence*, 5(3):310–321, 2013.

[12] Iztok Fister Jr, Xin-She Yang, Iztok Fister, Janez Brest, and Dušan Fister. A brief review of nature-inspired algorithms for optimization. *Elektrotehniški Vestnik*, 80(3):116–122, 2013.

[13] Simon Fong. Opportunities and Challenges of Integrating Bio-Inspired Optimization and Data Mining Algorithms. In Xin-She Yang, Zhihua Cui, Renbin Xiao, Amir Hossein Gandomi, and Mehmet Karamanoglu, editors, *Swarm Intelligence and*

Bio-Inspired Computation, pages 385–402. Elsevier, 2013.

[14] Rakesh Agrawal, Tomasz Imieliński, and Arun Swami. Mining association rules between sets of items in large databases. In *ACM SIGMOD Record*, volume 22, pages 207–216. ACM, 1993.

[15] Rakesh Agrawal, Ramakrishnan Srikant, et al. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, volume 1215, pages 487–499, 1994.

[16] Jiawei Han, Jian Pei, and Yiwen Yin. Mining frequent patterns without candidate generation. In *ACM SIGMOD Record*, volume 29, pages 1–12. ACM, 2000.

[17] Sergey Brin, Rajeev Motwani, Jeffrey D Ullman, and Shalom Tsur. Dynamic itemset counting and implication rules for market basket data. In *ACM SIGMOD Record*, volume 26, pages 255–264. ACM, 1997.

[18] Edward R Omiecinski. Alternative interest measures for mining associations in databases. *Knowledge and Data Engineering, IEEE Transactions on*, 15(1):57–69, 2003.

[19] Craig Silverstein, Sergey Brin, and Rajeev Motwani. Beyond market baskets: Generalizing association rules to dependence rules. *Data mining and knowledge discovery*, 2(1):39–68, 1998.