Towards automatic food prediction during endurance sport competitions

Iztok Fister Jr., Iztok Fister, Dušan Fister, Karin Ljubič
University of Maribor
Faculty of Electrical Engineering and Computer Science
Smetanova 17, 2000 Maribor, Slovenia
E-mail: iztok.fister1@um.si

Yan Zhuang, Simon Fong
University of Macau
Faculty of Science and Technology
Taipa, Macau SAR
E-mail: ccfong@umac.mo

***Abstract*—Endurance sport events have increasingly been gaining the popularity. Every year, more and more amateur athletes decide to participate in such events. During the race, proper eating is one of the most important components for achieving the good finish time and in this respect also the good place. In this paper, we examine possibility to predict what to eat at the moment. In line with this, machine learning method, i.e., decision tree was used that bases on the current athlete performance, his/her feeling, needs, and even weather. First simulations showed that this method may be suitable for future use in endurance events.**

Keywords-component; endurance sport event; machine learning; decision tree; food;

I. INTRODUCTION

Nowadays, during the very stress lifestyle, people are looking for some relaxation and stress free activities. Some people find inspiration in theater and cinema, some in reading books/magazines/stories, some likes walking, some fishing, and some practice and compete in sport. A lot of people have started to deal with different sport disciplines. Usually, they participate in running and biking competitions, but some compete also in another modern sport, like triathlon. Especially, triathlon is one of the hardest sports in the world and therefore people need a lot of training and experience in order to achieve well results. Triathlon can be divided in short, medium, long and ultra distances. Especially, ultra triathlons like Double ultra triathlon have a long duration, where participants are limited to 36 hours for finishing it.

For this type of trials, participants must be good prepared, healthy and powerful. One very important component in racing such event is also good eating during the race. Usually, participant has to eat every 30–45 minutes. On the other hand, there are also a lot of foods which have a high influence to the athlete's performance. Some preferred food by the ultra athletes are: bananas, power gels, power bars, chicken, pizza, orange, potatoes, pineapple and several kinds of tablets, e.g., magnesium, multi vitamins, minerals, salt, etc.

Recently, rise of mobile and pervasive computing radically have changed the life of modern human. These changes have also not gone past the sport. More and more people began to use smart phones in order to track activities and have real time picture of their own performance. Measures, like current heart rate of athlete, his/her position, elevation, total time duration and burned calories during the sport activity are only a few measures that can be obtained by the modern mobile devices.

In this paper, we try to develop a novel solution that should help athletes to automatically propose the proper food for eating during the competition based on athletes own performance. For example, athletes can feel good, falls into a crisis or is sensitive on any other conditions (like weather) during long bike rides. Moreover, they can also be dehydrated and suffer for a lack of salt, vitamins, and minerals in the body. Therefore, it is hard to decide what refreshment to take at the specific moment. Especially, amateur athletes who lack of experiences, hardly find proper food during the race.

In our proposed solution, we try to predict what to eat at the specific moment during the long duration race. In line with this, a decision tree to build that is a well-known tool taken from machine learning domain capable of prediction. We defined some attributes and create a training set, based on personal experience from these races. On the other hand, a test set can be generated automatically based on athletes current welfare on the race. This welfare is measured by the heart rate that can be get directly from activity tracker on smart phone. On basis of this and additional measures obtained from the mobile devices, the decision tree built during the test phase can quickly recommend the proper food for athlete during the long duration race. Furthermore, this method is also not time complex and can be easily implemented on the Android platform.

The structure of the paper is as follows. In Section II, we explain basics about double ultra triathlon, while Section III is concentrated on smart phones and activity tracking during the sport session. Section IV gives basics about a machine learning tool called decision tree, while section V presents a new solution. Initial simulations with results are subject of Section VI. Last section concludes with an overview of possible directions for the future work.

II. DOUBLE ULTRA TRIATHLON

Double ultra triathlon is a ultra triathlon race that is held under the cover of International Ultra Triathlon Association

978-1-4673-6751-6/14 $31.00 © 2014 IEEE
DOI 10.1109/ISCMI.2014.30

(IUTA). Every year, there are about 5 such competitions in the world. Of particular interest are also European and World Championships. For instance, the World Championship in 2014 will be hosted in Slovenia.

The Double ultra triathlon consists of a 7.6 km swimming, 360 km biking, and 84.4 km running. Average finish times are between 27–32 hours. To participate in this event, athletes must be well prepared, in good conditions, healthy, must be older than 21 years, have finished at least one Ironman race in the last year and have a well blood tests, e.g., Hematocrite must be sub 50, because above 50 might be suspicion on doping.

Figure 1: Double ultra triathlon

Usually, preparations for such races can take a lot of months and even years. Beside the physical condition, athletes must also be mentally well trained.

III. SMART PHONES AND SPORT ACTIVITY TRACKING

Smart phones are a new generation of mobile phones that consists of software and hardware. Hardware represents a mobile phone, while software consists of an operating system (OS) that allows users to upload and use different applications. Currently, Android OS is one of the most widely used OSs which is under general public license (GPL). It was developed by Google and there are millions and millions of different applications today that help users in their everyday lives.

One of the more interesting applications today are fitness and health applications, which encourage people to do more practice, have a healthy living style and live the sport life in general. Moreover, a lot of amateur athletes use sport trackers that track their sport activities/sessions in order to monitor their progress during training as well as during the real competitions. Modern trackers allow tracking the following data obtained during the training sessions:

- a current speed,
- a current time,
- a current heart rate,
- a total time duration,
- total kilometers,
- total calories burned,
- a total elevation,
- laps and some advanced trackers also any other measures.

This means that athletes have a real-time view of their performance. A good thing is that these devices enable athletes also to analyze data about their training sessions after sport activities. They can upload these data online and monitor performance graphs, share the obtained data with friends, predict training sessions, etc.

IV. DECISION TREE

Decision tree is one of the machine learning methods for a non-parametric supervised learning widely used for classification and regression problems. The goal of using the decision tree is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features [7]. Decision trees have been applied in many areas of human activities from industry to pharmacy and health applications. They were also used in sport [1][2][3][4]. Advantages of decision trees can be summarized in the following points [8][9]:

- they are very simple to understand by everyone,
- they can be visualized,
- they can handle multi-output problems.

An example of decision tree is presented in Figure 2, where we try to predict what to eat according to feeling and calories.

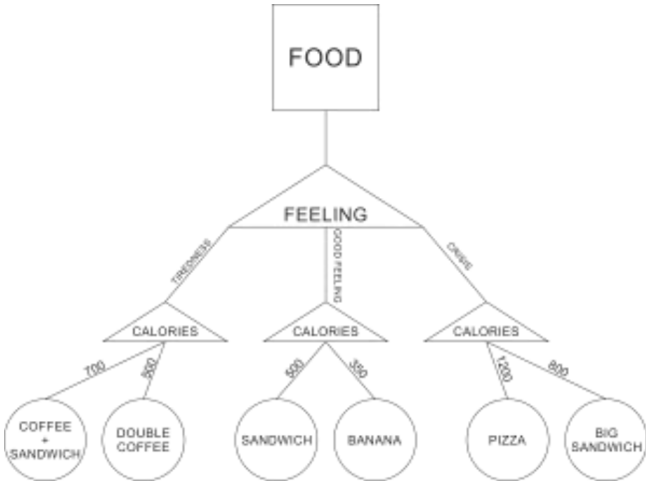

Figure 2: An example of simple decision tree

V. PROPOSED METHOD

The proposed method consists of three stages, as follows:

- definition of attributes,
- building the decision tree,
- classification with decision tree.

A first stage is definition of attributes. In this stage, parameters for decision tree are identified and their domain of values is defined that appear as nodes in the tree. In sport, there might be many different attributes, but we can take only the most important for our sport discipline. In the second stage, a building of decision tree starts, where creation of training set must be performed at first. The training set consists of six variables. On the basis of both

sets of variables, the decision tree is learned. In the last stage, the decision tree is successfully built and ready for classification. In line with this, a test set is needed to be defined. The test set is used as predictor for the proper use of food by athlete on long duration race at the specific moment.

A. *Definition of attributes*

At first, we need to define the input attributes for the decision tree. In this pilot study, we defined the following six attributes:

- **Current welfare**: This attribute defines the current feeling of an athlete. An athlete goes through many feelings during 360 km of biking, e.g., from welfare, to tiredness, and many similar situations in between these mentioned two. The welfare also depends on the weather. As a result, we defined the following values for these attributes:

Welfare={*very tired, good, moderate, craps, crisis*}.

- **Calories**: Calories symbolize the energy content of the food. Some foods like energy bars have a lot of calories, while another, like orange has very low calories intake. Values for calories are defined as numeric values drawn from interval $Calories \in [1,...,n]$.

- **Supplement**: This attribute reflects to an additional intake of food supplements, like magnesium, vitamins, multi-minerals, salt usually consumed in the form of tablets. The valid values of this attribute are $Supplement \in \{yes,no\}$.

- **Additional salt**: Especially, during the hot weather and high humidity, athletes lose a lot of salt and electrolytes. This means that they need to catch up these ingredients via salt or electrolyte drinks. The valid values of this attribute are $Salt \in \{yes,no\}$.

- **High sugar**: Several times, athletes need more sugar and we can catch up they with additional power gels, chocolate, etc. Values of this attributes are $Sugar \in \{yes,no\}$.

The result of prediction is predicted food. The food is a set of possible foods as recommended by a decision tree. This set of foods is defined as:

- **Food**: This attribute defines what food to take according to values of input attributes. A domain of valid values for this attributes denotes names of the most widely used foods for these purposes, e.g., $Food \in \{sandwich,energy\ bar,banana,...,pizza\}$.

B. *Building the decision tree*

First of all in this stage we have to build a training set. Training set must be very robust and data must be well defined if we want to get a good classification. Therefore, in this stage is good to have an expert from domain area who totally know domain and have experience. Since we have experience with ultra triathlon races, we defined it from our knowledge.

Training set (Fig. 3) consists of categorical and continuous attributes, while the last array is class label. When we have full training set, then we use it and learn a decision tree. After this, we get a model. On this model we later test it with test set and result of this step is predicted food.

```
attributes = ['Current welfare', 'Calories', 'Supplement', 'Additional salt', 'High sugar' 'Food

training = [
  ['good', '400', 'no', 'yes', 'no', 'small sandwich'],
  ['crisis', '700', 'yes', 'yes', 'yes', 'high energy bar'],
  ['moderate', '300', 'no', 'no', 'no', 'banana'],
  ['very tired', '350', 'no', 'no', 'no', 'double coffe'],
  ['craps', '50', 'yes', 'yes', 'no', 'magnesium tablet'],
  ['craps', '100', 'yes', 'yes', 'yes', 'banana'],
```

Figure 3: A snippet of training set

C. *Classification with the decision tree*

The test set is developed in a little bit different way. We try to include parameters directly from sport trackers and build a test set. If this solution is integrated into the sport tracker, it can high effective decide what to eat. By taking predefined attributes into the account, we can set the test set in the following way (Table 1):

- **Current welfare** is calculated according to the heart rate. If it is too low, it means that athlete is tired or goes very slow. On the other hand, if athlete has a moderate heart rate, means that he is feeling well. However, high heart rate symbolizes also very good feeling and possibility to race very fast.

- **Calories** can be also calculated from feeling, since we know what athlete must eat according to feeling. If athlete is tired, we have to wake up him, e.g., with coffee.

- **Supplements** are very connected with weather. Nowadays, getting weather information on smart phone is very easy. We can use free web services to fetch a current weather and calculate special needs. If it is very hot, we have to give athlete additional supplements of salt, magnesium, electrolytes and a lot of drink.

- **Additional salt** is also connected with weather and also feeling of an athlete. If he is very tired, we can him additional salt.

- **High sugar** is usually connected where athlete have a high speed and have a good heart race. If he is in good condition, he can take more "fast energy" to go faster. Moreover, it is also important in the final parts/segments of the race. We can get this information from tracker too.

The decision tree algorithm for automatic food prediction was implemented in Ruby using the Decision Tree Library (https://github.com/igrigorik/decisiontree) that on basis of the defined training and test sets predicts the proper food for athlete competing in the long duration sport race.

Table 1: Translation from tracker to attribute's values

Attribute	How to get it from tracker
Welfare	Get from heart rate
Calories	Approximate calculation according to welfare
Supplements	From weather
Salt	From weather
Sugar	From heart rate and distance

VI. First Simulations And Discussion

To test this initial idea, we defined some simulation scenarios. We defined the following athlete's scenarios:

- Scenario 1: heart rate *low*, weather *very hot,* and total kilometers *low*.
- Scenario 2: heart rate *high*, weather *cold,* and speed *very fast.*
- Scenario 3: heart rate *high,* weather *very hot,* and speed *slow.*
- Scenario 4: heart rate *high,* and weather is *cold.*

In the remainder of the paper, these scenarios were described in details.

Table 2: Scenarios

Scenarios	1	2	3	4
Welfare	*Tired*	*Moderate*	*Good*	*Good*
Calories	700	500	600	400
Supplements	*yes*	*no*	*yes*	*no*
Salt	*yes*	*no*	*yes*	*no*
Sugar	*yes*	*yes*	*no*	*yes*

A. *Scenario 1: heart rate low, weather very hot, and total kilometers low*

Athlete's heart rate is a very low, it is very hot and he/she at the beginning of the long duration race. In this case, the test set was built with the following values of attributes as illustrated in Table 2.

The predicted result in this scenario was high energy bar. If we examine, high energy bar, we can find, that it has a lot of calories, it has electrolytes which help with hydration and it some bars also have caffeine which help fighting with tiredness.

B. *Scenario 2: high heart rate, weather is not hot and athlete is very fast*

Athlete has a high heart rate and it is not hot. He is racing very fast. Values for attributes were defined in Table 2.

In this case decision tree did not find solution. It is probably, because our training set was not so huge. This simulation tells us that we have to expand training set to use it in real-world.

C. *Scenario 3: high heart rate and very hot weather, but slow speed*

Athlete has a nice heart rate, but it is very hot. He is racing slowly. Result of this prediction was: chocolate.

D. *Scenario 4: high heart rate and weather is not hot*

Athlete has a high heart rate, but it is not very hot.

Result of this prediction was: small sandwich. It is also a good prediction, because it is not hot and athlete does not need at the moment to take additional salt or magnesium. Therefore, good sandwich may help to the athlete to go normally along the next kilometers.

VII. Conclusion With Future Remarks

In this paper we examined a pilot solution for athlete's food prediction during endurance sport events. We took double ultra triathlon as an example of endurance sport event. Simulations showed that this method may be suitable for usage in such events, since decision tree decided proper food according to athlete's need. There are still some problems, because training set must be really expanded and more food added into the account.

In the future, there are still a lot of work to do in this domain, because lots of additional simulations and scenarios must be taken into account. It must also go to real scenario, where athletes race and listen to the predictions and eat predicted food to get better insight into the solution. Currently, it is also very synthetic based and every human behaves differently. Therefore, link with sport trainers and real athletes must be also included in this research.

ACKNOWLEDGEMENT

The authors are thankful for the financial support from the research grant of Grant no. MYRG152(Y3-L2)-FST11-ZY, offered by the University of Macau, RDAO.

References

[1] Willoughby, Keith A., and Kent J. Kostuk. "An analysis of a strategic decision in the sport of curling." Decision Analysis 2.1 (2005): 58-63.

[2] Parkka, Juha, et al. "Activity classification using realistic data from wearable sensors." Information Technology in Biomedicine, IEEE Transactions on 10.1 (2006): 119-128.

[3] Joseph, A., Norman E. Fenton, and Martin Neil. "Predicting football results using Bayesian nets and other machine learning techniques." Knowledge-Based Systems 19.7 (2006): 544-553.

[4] Fister, Iztok Jr., Dusan Fister, Iztok Fister and Simon Fong. "Data mining in sporting activities created by sports trackers."Computational and Business Intelligence (ISCBI), 2013 International Symposium on. IEEE, 2013.

[5] Fister, Iztok, and Iztok Fister Jr. "Concept of drafting detection system in Ironmans."Electrotechnical review", 78 (4), 218-222, 2012.

[6] Fister, Iztok, et al. "Widespread Mobile Devices in Applications for Real-time Drafting Detection in Triathlons." Journal of Emerging Technologies in Web Intelligence 5.3 (2013): 310-321.

[7] Fister Jr, Iztok, Iztok Fister, and Janez Brest. "Comparing various regression methods on ensemble strategies in differential evolution." Mendel conference on soft computing, Czech republic", 87 – 92, 2013

[8] Breiman, Leo, et al. Classification and regression trees. CRC press, 1984.

[9] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." The Journal of Machine Learning Research 12 (2011): 2825-2830.