Optimizing Interval Training Protocols
Using Data Mining Decision Trees

Myung-kyung Suh¹, Mahsan Rofouei¹, Ani Nahapetian¹,²,³, William J. Kaiser¹,², Majid Sarrafzadeh¹,³

Computer Science Department¹, Electrical Engineering Department², Wireless Health Institute³
University of California, Los Angeles
{dmksuh,mahsan,ani,majid}@cs.ucla.edu, kaiser@ee.ucla.edu

Abstract—Interval training consists of interleaving high intensity exercises with rest periods. This training method is a well known exercise protocol which helps strengthen and improve one’s cardiovascular fitness. However, there is no known method for formulating and tailoring an optimized interval training protocol for a specific individual which maximizes the amount of work done while limiting fatigue. But by using data mining schemes with various attributes, conditions, and data gathered from an individual’s exercise session, we are able to efficiently formulate an optimized interval training method for an individual.

Recent advances in wireless wearable sensors and smart phones have made available a new generation of fitness monitoring systems. With accelerometers embedded in an iPhone, a Bluetooth pulse oximeter, and the Weka data mining tool, we are able to formulate the optimized interval training protocols, which can increase the amount of calorie burned up to 29.54%, compared with the modified Tabata interval training protocol.

Index Terms—Data mining, heart rate limitation, interval training, wearable wireless sensors

I. INTRODUCTION

Recent advances in sensors, smart phones and wireless technology have made a new generation of health and fitness monitoring systems available. The SmartCane system (Vahdatpour [31]) helps patient rehabilitation monitoring on a cane, using accelerometer, pressure and gyroscope sensors. Dencker [32] uses accelerometers to evaluate daily physical activities in children aged 8 to 12. Previously, athletes had to visit athletic centers or hospitals to monitor their health and fitness using large fitness monitoring systems. However, with recent advances in technology, people supplement or even tailor this process using sensors and handheld systems in their very own home or in the field. Moreover, these systems have the power to give precise and real-time feedback, using the data collected from sensors such as accelerometers, pressure sensors, and gyroscopes. In this paper, we focus on an interval training system with wireless sensors and a smart phone.

Interval training consists of interleaving high intensity exercises with rest periods. It has been the basis for athletic training routines for many years. This training method is a well known exercise protocol which helps strengthen and improve one’s cardiovascular system ([1] – [7]). Moreover, it helps with weight loss, rehabilitation, general fitness, and the reduction of heart diseases. During interval training, the body’s energy production system is utilized, and both aerobic and anaerobic energy sources are activated. Energy from these two sources is then efficiently distributed throughout the body for the duration of the workout period.

Currently, there exists several interval training methods but there is no known one-size-fits-all optimal solution. Usually experts recommend “active” rest which means people should continue low intensity exercise during rest periods ([12] – [15]). Billat et al. [12], Newman et al. [13], Brooks et al.[14], and DC Poole et al. [15] all recommended such low intensity activities during rest intervals in order to remove blood lactate which causes muscles to ache and furthermore, causes a person to feel tired. As Borg et al. [16] mentions, there is a relation between heart rate and lactic acid. Therefore, by monitoring one’s heart rate during exercise, we can avoid exceeding the level at which the heart rate causes person to feel exhausted.

There are many studies related to data mining and the medical and fitness domain ([19]-[24]). However, we found no study using data mining for interval training monitoring and scheduling.

In this paper, we present the following key contributions. By using data mining decision trees, we find conditions which maximize an individual’s amount of work. In our limited testing of this approach on different individuals, we obtained up to 22.73 % increase in the amount of exercise. For an individual, we saw an improvement of up to 29.54 % in the amount of calories burned within 2 weeks.

II. EXERCISE AND HEART RATE MODEL

During exercise, the quantity of blood pumped by the heart increases to match the increased skeletal muscle demand. In addition, heart rate also acts as an indicator of exercise intensity. The more intense the activity, the faster your heart
will beat. Thus, after starting exercise session, the increase in heart rate variable can be observed. In contrast, immediately after stopping exercise activity, the heart rate should decrease (CR Cole [30]).

Fig 2.1 indicates heart rate variables during 22 minutes modified Tabata interval training (Table 3.2). We can observe that the heart rate increase after starting exercise and decrease after taking a rest. If we can increase the amount of time until the heart rate reaches a certain high level during exercise, it helps people exercise more without fatigue. Also, if we can decrease the amount of time until the heart rate reaches a certain level during rest, it will help increase the amount of exercise time within limited time.

![Heart rate variable during Interval training](image)

**Fig 2.1. Heart rate during 22 minutes Tabata interval training**

For example, if you can increase the time to reach heart rate 113 in the first exercise period of 22 minutes Tabata protocol from 135 seconds to 269 seconds, you can exercise 113 in the first 22 minutes Tabata protocol. Similarly, if you can reduce the amount of rest time, exercise time will be increased.

![Increase in the amount of exercise time](image)

**Fig 2.2 Increase in the amount of time to reach heart rate 113 in 22 minutes Tabata Interval Training Protocol**

III. TWO PHASES TO INCREASE THE AMOUNT OF WORK FOR AN INTERVAL TRAINING PROTOCOL.

Mizuo Mizuo et al. [10] denotes that rising heart rate curves have an exponential hyperbolic shape, and the falling heart rate curves are exponential when exercising.

\[
\text{HR of the Rising Curve} = Ae^{-\beta t} \sinh(\alpha t) + C
\]

\[
\text{HR of the Falling Curve} = A(1-e^{-\beta t}) + C
\]

However, when observing the same heart rate curve from the same individual (Jansen [25]) several times, the curves are dramatically different. Thus it is hard to predict and guess future heart rate values with only the above equations. Instead, it may be better to use a statistical and probabilistic approach since this approach can extract, compare and infer certain characteristics from the many different heart rate curves of an individual. By combining the information of the heart rate curve, such as the heart rate value and the time constant of the curve which characterizes the frequency response of the curve, along with other individual conditions, we can maximize the amount of work an individual performs during an interval training workout. As many studies have suggested including Christensen [8] there is a correlation between fatigue level and heart rate. Since each individual has his or her own unique conditions such as age, gender, level of activity, which will in turn affect the heart rate values, we will use this information to help in the calculation of a user’s optimal heart rate and fatigue level. Also, the use of pharmaceuticals a person takes regularly, and the elapse time after they are taken can drastically affect the heart rate curve (Y S Tuininga [27], Conny M. A [28], LUND-JOHANSEN P [29]). Additionally, we need to take into account the amount of time after having a meal since this could also affect the heart rate, fatigue level and other conditions related to exercise. In our experiments, we took into account all factors since these conditions can dramatically affect an individual's heart rate level. To monitor heart rate in this research, Alive Technologies Bluetooth pulse oximeter was used in our experiment, which sends heart rate variable data every 0.1 second.

Accelerometers are currently among the most widely studied wearable sensors for activity recognition. They are also a very useful sensor for interval training. By analyzing data obtained from 3 different axes, the accuracy of exercise and the calorie consumption can be calculated. Additionally, they are also widely embedded in smart phones such as the Apple iPhone, Google G1, and Nokia N95. Moreover, these smartphones can help individuals to follow up the scheduled speed during a specific time period, can also give feedback using the cell phone’s sound, vibration, and other graphical interfaces. Additionally, their calculation functionalities can also be leveraged.

In this study, we used the iPhone[33] for our development. The iPhone is very light, about 133g, has a 3.5 inch multi-touch display, and supports both Wifi and Bluetooth. Furthermore, the iPhone has an in-built accelerometer and proximity sensor. By using an embedded 3-axis accelerometer in the iPhone, we can detect one’s activity pattern.

A. Phase 1: Finding conditions which affect optimal time constants.

By modifying an already existing interval training method, Tabata Protocol (Table 3.2), we collected 10 sets of data which include several attributes which may affect an individual’s heart rate and exercise potential.
performing Tabata protocol, a participant should step every 0.5 second or 0.75 second during certain periods on a 25cm-tall stool. For rest periods, we assumed the participants took full rest during the rest period. Each 22 minute exercise session datasets includes 13219 attributes as shown in Table 3.1. One Tabata protocol is also divided into 4 exercise periods and 3 rest periods.

To maximize the amount of work, the time constant of an exercise period should be maximized until the heart rate reaches a certain high level. In contrast, the time constant of a rest period which has the exponential curve characteristic should be minimized in order to reduce the time to reach a certain low level heart rate. As you can see in Fig. 2.2, we should change the original exponential sine hyperbolic curve to the modified one which has a bigger time constant.

Before starting the exercise session, other attributes such as age, height, and weight can be recorded in the data base. During the 22 minute session, the user’s heart rate should be monitored and recorded, since it is coming into our system real time. Additionally, we should record the inaccuracy of exercise after finishing our workout session. In our data, the inaccuracy means the difference between the scheduled amount of exercise and the actual exercise performed. As mentioned, the speed or movement in the actual exercise can be obtained by using the 3-axis accelerometer on the iPhone.

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Value</th>
<th>Attributes</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>[male, female]</td>
<td>Days after period</td>
<td>numeric</td>
</tr>
<tr>
<td>Age</td>
<td>numeric</td>
<td>Number of exercises done today</td>
<td>numeric</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>numeric</td>
<td>Time</td>
<td>numeric</td>
</tr>
<tr>
<td>Weight(kg)</td>
<td>numeric</td>
<td>Inaccuracy of exercise (%)</td>
<td>numeric</td>
</tr>
<tr>
<td>Activity level</td>
<td>[low, medium, high]</td>
<td>Time constant</td>
<td>numeric</td>
</tr>
<tr>
<td>Disease name</td>
<td>string</td>
<td>Number of exercise period</td>
<td>numeric</td>
</tr>
<tr>
<td>Hours of sleep</td>
<td>numeric</td>
<td>Number of rest period</td>
<td>numeric</td>
</tr>
<tr>
<td>Hours after a meal</td>
<td>numeric</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Medicine</td>
<td>string</td>
<td>13200 Heart rate variables (every 0.1 s)</td>
<td>numeric</td>
</tr>
</tbody>
</table>

Table 3.1. 13219 attributes in one data set

<table>
<thead>
<tr>
<th>Attributes</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Warm up</td>
<td>stepping on the stair per 0.75 second for 5 minutes</td>
</tr>
<tr>
<td>6 set exercises</td>
<td>6 x (stepping on the stair per 0.5 second for 20 seconds + 10 seconds rest)</td>
</tr>
<tr>
<td>Rest</td>
<td>1 minute rest</td>
</tr>
<tr>
<td>6 set exercises</td>
<td>6 x (stepping on the stair per 0.5 second for 20 seconds + 10 seconds rest)</td>
</tr>
<tr>
<td>Rest</td>
<td>1 minute rest</td>
</tr>
<tr>
<td>6 set exercises</td>
<td>6 x (stepping on the stair per 0.5 second for 20 seconds + 10 seconds rest)</td>
</tr>
<tr>
<td>Rest</td>
<td>1 minute rest</td>
</tr>
<tr>
<td>Cool down</td>
<td>stepping on the stair per 0.75 second for 5 minutes</td>
</tr>
</tbody>
</table>

Table 3.2. Uncontrolled Exercise Protocol for 10 data in Phase 1.

With these datasets, the knowledge discovery and data mining techniques provides the extraction and classification of information that we need. Using data along with data mining techniques, a J48 decision tree is generated by Weka, a powerful data mining tool. The J48 Decision tree classifier follows the following simple algorithm. First, in order to classify a new item, creating a decision tree based on the attribute values of the available training data is considered. Thus, whenever it encounters a training set it identifies the attribute that discriminates the various instances most clearly. Among the possible values of this feature, if there is any value for which there is no ambiguity, that is, for which the data instances falling within its category have the same value for the target variable, then it terminates that branch and assign to it the target value that we have obtained. For the other cases, it looks for another attribute that gives the highest information gain. Hence it continues in this manner until it either gets a clear decision of what combination of attributes gives a particular target value, or it runs out of attributes. In the event that it runs out of attributes, or if it cannot get an unambiguous result from the available information, it assigns this branch a target value that the majority of the items under this branch possess. In addition, if more data sets are accumulated in the database, the system will be adopted and can achieve a more accurate classification.

B. Phase 2: Modified interval training based on the constructed J48 decision tree

According to the constructed J48 decision tree in Phase 1, conditions which extend exercise periods and shorten rest periods related to the time constants were obtained. For example, if the user has slept for more than 6 hours the previous night or has eaten within the last 2 hours, the system can tell the participant to exercise more. Or in other words, the system can tell user to exercise for a longer period of time in order to maximize the exercise within the 22 minute duration.

Exercise information such as the heart rate value of when a person starts exercising and stops exercising can also be obtained from Phase 1. In phase 2, people should exercise until the maximal heart rate reaches the optimal rising curve in the exercise period. They should also start exercising when the minimal heart rate reaches the optimal falling curve during the rest period. Participants should satisfy conditions in decision tree information obtained from phase 1. By combining conditions related to time constants and increase in the amount of exercise, we can obtain better result which makes users exercise more without fatigue.

As data is accumulated in the data base, the decision tree obtained from phase 2 also can be updated. As mentioned, the decision tree will become more accurate as more data is accumulated in the database. This means the decision tree will be modified and adapted, as the number of data increases.
Fig. 3.1. Interval Training Command Program. This system uses accelerometer functionality and graphical, sound, vibration functionalities on iPhone.

IV. EXPERIMENTAL RESULTS

A. 5 minutes 30 seconds Interval Training for seven individuals

Seven different individuals in Table 3.1 participated in five exercises for the 1st phase which uses the modified Tabata Protocol (Warm up - stepping on the stair every 0.75 second for 2 minutes 30 seconds, 6 x (stepping on the stair every 0.5 second for 20 seconds + 10 seconds rest), 1 minute rest). Afterwards they completed one 2nd phase of the exercise satisfying the conditions of the decision tree obtained from the 1st phase. We observed an improvement of up to 22.73% in the amount of calories burned within 5 minutes and 30 seconds. This proves our methods help people to exercise more within a given heart rate threshold range.

<table>
<thead>
<tr>
<th>Person</th>
<th>Person</th>
<th>Person</th>
<th>Person</th>
<th>Person</th>
<th>Person</th>
<th>Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>male</td>
<td>female</td>
<td>male</td>
<td>male</td>
<td>female</td>
<td>male</td>
</tr>
<tr>
<td>Age</td>
<td>29</td>
<td>30</td>
<td>27</td>
<td>26</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>170</td>
<td>170</td>
<td>170.8</td>
<td>167.7</td>
<td>158</td>
<td>164</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>62</td>
<td>58</td>
<td>62</td>
<td>79.3</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Activity</td>
<td>medium</td>
<td>low</td>
<td>low</td>
<td>high</td>
<td>medium</td>
<td>low</td>
</tr>
<tr>
<td>Disease</td>
<td>none</td>
<td>none</td>
<td>none</td>
<td>none</td>
<td>none</td>
<td>none</td>
</tr>
</tbody>
</table>

Table 4.1 Information about individuals who participated in the modified Tabata interval training exercise.

Studies have shown that an individual's heart rate will return to a resting rate after 3 minutes. And as such, in our experiments, individuals were asked to take a 3 minute break between the 5 different exercise sessions.

B. 22 minutes Interval Training for an individual

Individual #7 in Table 4.1 participated in the original 22 minute modified Tabata interval training (Table 3.2.). She finished 10 sets of 22 minutes Tabata exercise sessions with different conditions on different days. Information related to one’s conditions in Table 3.1 is stored and used to generate the J48 tree after completing 10 interval trainings. The generated J48 tree after finishing 10 Tabata interval trainings is shown below. This gives the conditions which maximize the time constant of exercise periods and minimize the time constants of rest periods. Conditions mentioned in Fig. 4.2 can be updated and modified as data is accumulated.

Fig 4.2. J48 decision tree which shows conditions which make the time constant of 4 rising curves which have a characteristic of exponential sine hyperbolic curves maximize. J48 decision tree which makes the time constant of 3 falling curves which have a characteristic of exponential curves minimize are also same. Sleep: the amount of hours to take a sleep before exercising. Inaccuracy: the inaccuracy between the pedometer data and scheduled data. Meal: the amount of hours after having a meal.

Figure 4.1. Result of the 2nd phase. (The amount of work in the 1st phase is the standard.) This shows when the person follows conditions determined by J48 tree obtained from the 1st phase, the increase of the amount of exercise is from 1.36 to 22.73 percent.

B. 22 minutes Interval Training for an individual

Individual #7 in table 4.1. participated in the original 22 minute modified Tabata interval training (table 3.2.). She finished 10 sets of 22 minutes Tabata exercise sessions with different conditions on different days. Information related to one’s conditions in Table 3.1 is stored and used to generate the J48 tree after completing 10 interval trainings. The generated J48 tree after finishing 10 Tabata interval trainings is shown below. This gives the conditions which maximize the time constant of exercise periods and minimize the time constants of rest periods. Conditions mentioned in Fig.4.2 can be updated and modified as data is accumulated.

Fig 4.2. J48 decision tree which shows conditions which make the time constant of 4 rising curves which have a characteristic of exponential sine hyperbolic curves maximize. J48 decision tree which makes the time constant of 3 falling curves which have a characteristic of exponential curves minimize are also same. Sleep: the amount of hours to take a sleep before exercising. Inaccuracy: the inaccuracy between the pedometer data and scheduled data. Meal: the amount of hours after having a meal.

IV. EXPERIMENTAL RESULTS

A. 5 minutes 30 seconds Interval Training for seven individuals

Seven different individuals in table 3.1 participated in five exercises for the 1st phase which uses the modified Tabata Protocol (Warm up - stepping on the stair every 0.75 second for 2 minutes 30 seconds, 6 x (stepping on the stair every 0.5 second for 20 seconds + 10 seconds rest), 1 minute rest). Afterwards they completed one 2nd phase of the exercise satisfying the conditions of the decision tree obtained from the 1st phase. We observed an improvement of up to 22.73 % in the amount of calories burned within 5 minutes and 30 seconds. This proves our methods help people to exercise more within a given heart rate threshold range.

<table>
<thead>
<tr>
<th>Person</th>
<th>Person</th>
<th>Person</th>
<th>Person</th>
<th>Person</th>
<th>Person</th>
<th>Person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>male</td>
<td>female</td>
<td>male</td>
<td>male</td>
<td>female</td>
<td>male</td>
</tr>
<tr>
<td>Age</td>
<td>29</td>
<td>30</td>
<td>27</td>
<td>26</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>170</td>
<td>170</td>
<td>170.8</td>
<td>167.7</td>
<td>158</td>
<td>164</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>62</td>
<td>58</td>
<td>62</td>
<td>79.3</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>Activity</td>
<td>medium</td>
<td>low</td>
<td>low</td>
<td>high</td>
<td>medium</td>
<td>low</td>
</tr>
<tr>
<td>Disease</td>
<td>none</td>
<td>none</td>
<td>none</td>
<td>none</td>
<td>none</td>
<td>none</td>
</tr>
</tbody>
</table>

Table 4.1 Information about individuals who participated in the modified Tabata interval training exercise.

Studies have shown that an individual's heart rate will return to a resting rate after 3 minutes. And as such, in our experiments, individuals were asked to take a 3 minute break between the 5 different exercise sessions.
Tabata interval training done in the second phase shows the increase of the amount of work up to 24.09 percent (Fig 4.3)

Fig 4.3. The percentage of increase in the amount of work within 22 minutes.

The original schedule of modified Tabata interval training is shown in Fig 4.4. When the participant follows conditions in Fig 4.5 and exercises until the current heart rate reaches the level in the rising curve which has the maximum time constant among the individual's previous rising curves. The individual resumed exercising when the heart rate recovered to the lowest level of the heart rate curve with the minimum time constant. The changed schedule of exercise is shown on Fig.4.6. We observed that the rest period of the 2nd phase exercise is sparser and shorter than the original schedule.

Fig 4.4. The original schedule of Tabata Protocol. Intensity 120: 1 step per 0.5 sec. Intensity 75: 1 step per 0.75 sec.

Fig 4.5. The changed schedule of Tabata Protocol when following the constructed J48 decision tree. Intensity 120: 1 step per 0.5 sec. Intensity 75: 1 step per 0.75 sec.

After finishing the second 10 sets of Tabata interval training which fulfills conditions in Fig 4.2, new J48 tree was generated with 20 accumulated interval training data. The conditions which maximize the amount of work are shown in Fig.4.6. Also other decision trees which satisfy conditions in Fig 4.7 are obtained by backtracking. 10 sets of

Fig 4.6. J48 tree which shows conditions which make the amount of work maximize. Within, X_Y : X to Y percent increase of the amount of exercise. 
TimeX: Heart rate at X time unit. (Each time unit: 0.1 sec)

Fig 4.7. J48 tree which shows conditions which make heart rate at 1090.4 sec more than 115.

The results of 3 sets of 22 minutes exercise data satisfying the J48 decision trees in Fig.4.2 and Fig.4.6 shows more improvement in the amount of exercise. We see an improvement of up to 29.54 percent compared with the results the original protocol.

Fig 4.8. The percentage of increase in the amount of work within 22 minutes satisfying two J48 trees in Fig 4.2 and Fig.4.6

C. Improvement in cardiovascular build-up

For an individual, person 7 on table 4.1, the amount of time required to reach the same heart rate (137) increases as we repeat the interval training protocol. This means person 7 has adapted to the Tabata interval training protocol, and as a result, increases their endurance to complete the exercise and benefits the effect of cardiovascular build-up.
interval training, which originally helps cardiovascular function, heart rate variability, endothelial function and in many patho-physiological aspects of heart failure have shown improvements in symptoms within a relatively short time period. Several studies of exercise training younger and older patients with heart failure have shown improvements in symptoms and exercise capacity and in many patho-physiological aspects of heart failure. This includes skeletal myopathy, ergoreceptor function, heart rate variability, endothelial function, and cytokine expression (Witham [19]). Compared with studies done before which usually shows improvement within 10 weeks etc., signs of improvement in the amount of work appeared within 2 weeks, since the decision tree and the interval training which originally helps cardiovascular build-up were combined together.

V. CONCLUSION

Interval training is a well known exercise protocol which is beneficial for weight loss, rehabilitation, and cardiovascular build-up. Computer-based analysis and classification, such as data mining, can be very helpful for developing optimized interval training protocols individually tailored for users. In this paper, a J48 tree in Weka was used to formulate the conditions which increase the amount of high-intensity activity within the 22 minute program compared with original exercise program. Based on this information we obtained up to 24.09% increase in the amount of work compared with the modified Tabata interval training protocol. Also, by adding conditions which maximize the improvement in the amount of exercise, we see an increase of up to 29.54% improvement compared with the original one.

To construct the optimized interval training based on the J48 decision trees, heart rate data also should be considered. However, the method that we used in this experiment is unable to process the data stream coming from the pulse oximeter sensor device in real time. To improve the processing speed and make this real time system possible, using a data stream management systems such as Aurora (di, DJ. [26]) can be considered.

REFERENCES


Fig 4.9: The required time to reach the heart rate 137. This shows that for the first exercise period in the 22 minutes Tabata protocol with the same heart rate threshold increased due to the adaptation and cardiovascular build-up.

CARDIOVASCULAR BUILD-UP

<table>
<thead>
<tr>
<th>Seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Number of Exercise</td>
</tr>
</tbody>
</table>

Cardiovascular build-up