ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
UDC/UDK 004.891.032.26:796.015

MODELLING OF THE ATHLETE'S TRAINING DECISION SUPPORT

Ladislav Havaš, Zoran Skočir, Vladimir Medved

Original scientific paper
In every training process it is of great importance to generate a quality program for athletic preparations. Systematic monitoring of an athlete's sport performance does not end when the program is created. This study shows modelling of athlete's training decision support. Based on created data model and with the application of ICT technologies, in this paper is presented the development of an expert telematic system, which creates an individual program for the training of athletes considering his/her current abilities and potentials, together with his/her maximal aerobic capacity, $VO_{2\max}$ factor. Inputting proper training and feedback information in database and in Data Warehouse of this system allows the program to conduct OLAP and Data Mining of the achieved results. Prompt divulgation of symptoms that indicate over(under)training of an individual enables the system to modify the training process in real time. In that way athletes can perform better and the chances for injury/trauma and for inadequate results are reduced.

Keywords: *artificial intelligence, athletic training program, Data Mining, Data Warehouse, expert system, ICT technologies, neural networks, OLAP, ORT, Running Log, training process,* $VO_{2\max}$*, WEB*

Modeliranje sustava za potporu odlučivanja u pripremi sportaša

Izvorni znanstveni članak
U svakom trenažnom procesu je od iznimne važnosti generiranje što kvalitetnijeg programa za pripreme. Samom izradom programa, ne završava proces sustavnog praćenja izvedbe sportaša. U ovom se radu prikazuje modeliranje sustava za potporu odlučivanja u pripremi sportaša. Na temelju kreiranog podatkovnog modela, primjenom ICT tehnologija prikazuje se razvoj jednog ekspertnog telematičkog sustava, koji na osnovu trenutnog potencijala pojedinog sportaša i njegovog maksimalnog aerobnog kapaciteta, $VO_{2\max}$ faktora, kreira njegov program za pripreme. Unosom odgovarajućih treninga i povratnih informacija u bazu i skladište podataka ovog sustava, omogućava se dubinska i OLAP analiza ostvarenih rezultata. Pravovremenim otkrivanjem simptoma pod(pre) treniranosti moguće je modificirati trenažni proces u realnom vremenu. Na taj način se omogućuje bolja izvedba i pravovremeno se smanjuje vjerojatnost pojave povrede ili neadekvatnog rezultata.

Ključne riječi: *dnevnik trčanja, dubinska analiza, ekspertni sustav, ICT tehnologije, neuronske mreže, OLAP, program priprema, skladište podataka, trenažni proces, umjetna inteligencija,* $VO_{2\max}$*, WEB*

1 Introduction

The paper gives an overview of the development of one expert WEB system that would serve long-distance runners as "Online Coach" and as a system for storing, recording and analysis of successfully completed training. It is very challenging to build a good system for generating programs for the preparation of athletes, since every person, including every athlete, is a unique subsystem that requires a specific individual approach and specific methods to develop their potential. The realization of the system that will support a large range of services leads to the generation of a large number of diverse data, which is conceptual, logical and physical modelling implemented within the Database Management System (DBMS) and Data Warehouse. Quick access to the required information and services, providing minimal use of communication resources, is the main demand in the construction of such optimal architecture of telecommunications information system. Generating high-quality programs to target athletic discipline, or race, is only the first step in the systematic monitoring of athletes in their training process. For the analysis of successfully completed training, it is necessary to develop a database and a Data Warehouse, which will run in a form of a Running Log, via the WEB, E-mail and SMS to collect data from users. Special design of Data Warehouses and periodic migration of data from Running Log, would provide a Data Mining and OLAP imagined and actual results of each competitor.

It would recognize the symptoms of overtraining in running or inadequate/ineffective training elements that could be swapped in real time with better and more appropriate solutions. In this way it builds a dynamic and flexible program for preparations, with a reduced likelihood of injury or illness during the training process. They are also looking for better methods, tailored to the respective athlete, in order to improve their results. The flow chart of this WEB expert system is shown in Fig. 1.

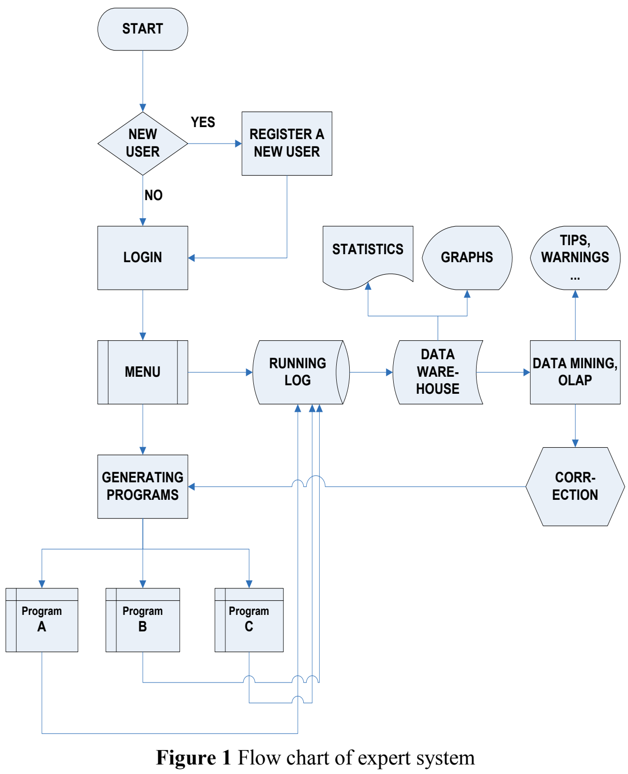

Figure 1 Flow chart of expert system

The second section of this paper presents the possibility of using modern ICT technologies in the development of sports expert system. The third section of the paper shows a brief overview of the existing systems and methods that are used for similar purposes. The fourth section of the paper presents the telecommunications subsystem that is used to communicate with customers. The fifth section of the work shows the determination of the current potential of long distance runners, calculates their maximum aerobic capacity $VO_{2\max}$, and generates programs for preparations. The sixth section shows the creation of dimensional data models and building Data Warehouses. The seventh section describes the methods used in Data Mining and dynamic redesign of preparation programs. In the eighth section are described the results of the verification of the users of the system, and the final one, the ninth section of this paper comes to a conclusion, followed by a list of references.

2 Using ICT technologies in sport

Contemporary way of life, doing business and service providing is nowdays unthinkable without the application of systems based on information and communication technologies. The speed of development of new ICT technologies demands their continuous monitoring in order to upgrade old systems or to introduce new ones so one could promptly and optimally use the possibilities of upgrading certain services. Since mobile data services enable the availability of information from any place anytime, and the development and introduction of second generation of WEB opens the possibility for quick interactive bidirectional communication among users or a user and the computer service.

This paper describes the development of one sport expert system, which is based on treating the telecommunication network as a platform for accelerated flow of information from trainer to athlete. Openness, liberty and collective intelligence of this particular system grows over time and also with number of users, which, unlike other classic training systems, decreases the possibility of making mistakes and allows greater chance of scoring better results.

By using Internet and mobile communications, there are drastic changes in the structure of access to each athlete, regardless of their location or availability. In that way we enable:

- Enhanced flexibility and practicality
- Increased transparency
- Showing the athlete's readiness in real time
- Immediate bidirectional exchange of information with athletes
- Direct and individual approach to ultimate users
- Decreased expanses of training
- Encouraging the disappearance of the mediator
- Accelerated development of software agents which take care of athletes during the training period.

ICT technology nowdays has the possibility of monitoring the user/athlete in all phases of the training period, from the decision concerning the creation of the training program, consultations with the system during the training period, dynamically modifying training program and after the worked off period allows the analysis of achieved results. Some systems are even more sophisticated, so they by using statistical methods process previous queries and interests of users in order to offer provide the best possible service which would completely satisfy users. Moreover, very significant is the range of services that are provided in the training of athletes on mobile devices. Cell phones have become appropriately good in size, price, design, connectivity and memory strength, and other features which allow their excellent supporting usage in the sport, training and data transfer on open space. Of course, when designing that kind of services for mobile devices one should bear in mind many interdisciplinary factors; psychological (design of user interface), technological (permeability of connection to Internet, the strength of processor, memory, speech recognition, and speech orders) as well as security (privacy of data).

By using above-mentioned ICT technologies we should aim at achieving the greater localization and personalization of applications and services in order to achieve contextually directed services tailored by cognitive and psychological aspects of athletes.

That kind of systems needs to be flexible and modular and to be able to adapt to new needs and demands of users, with the preservation of integrity of database. Design of those services has to aim at allowing the user access to all necessary information regarding the circumstances in which he/she is in present moment, which creates important preconditions for accepting those applications from the user and upgrading their sport achievements with decreasing the probability of injury or illness during their career.

3 Related work

There are systems which are developed in the way to provide the users with services for creating training programs, based on input parameters and users' wishes. One well-known that type of system is "Smart Coach" [10], which, based on input parameters, generates WEB programs for training of runners. For the same input parameters, the system always generates an identical program. The program is displayed on the desktop and it can be printed out, where the interaction with users ends.

A bit older methods of creating training programs were based on the usage of publications of experts [3, 7], where they offered developed programs or knowledge supposed to help the readers in creating their own training programs. At this point, as in the first example, this was the time when the "care" for users stops as the channels for feedback information regarding trainings and the effect they produce were not created. With time, as the appropriate programs and methods were used in sport preparations, software for collecting the data on PC or WEB, concerning the features of training, started developing. That kind of database was called "Running Log". Offered analysis of input data is usually based on run miles or kilometers, the type and the amount of training, on weekly, monthly and annually basis. There is no analysis of the quality of finished trainings, as there is no method for measuring the value of each particular training, nor is there a systematic monitoring of the progress in the training process.

There are systems which, by using some of the methods of Data Mining, evaluate the ability of a particular athlete for good performance in chosen sport [5]. It is evident that there is a need for the system which would use all of above-mentioned methods, add new ones, evaluate done trainings and would, by using ICT technologies and channels, communicate with its users in real time. This paper describes one such system which managed to establish a bidirectional communication with its users in real time with dynamical redesign of their training processes.

4 Presentation of telecommunication subsystem

Telecommunication subsystem consists of servers which actuate software solution and telecommunication elements by which "Online Running Trainer" (ORT) system achieves mobile interaction with users. Due to increased availability, it is derived in redundant form. Datacenter is the place which is located at Internet service provider and it is made in order to satisfy all the conditions for data storage. Servers are linked on permanent Internet connection of high availability and guaranteed stable permeability (Fig. 2).

For the immediate implementation three static and public IP addresses are allocated. Each server has one, previosely given, public IP address. Third IP address always has the server which has the role of "active server" by which users can access the resources. Redundancy of a system is realized in standby mode. In that mode active server provides the users with necessities, while "passive server" monitors the active one with heartbeat protocol and in a case of unavailability of the active server the passive one can take its role and all its network parameters.

Among all the roles stated above, active server also has the role of primary E-mail server.

It is allocated for reading and sending SMS via Clickatel SMS Gateway which runs by HTTP protocol.

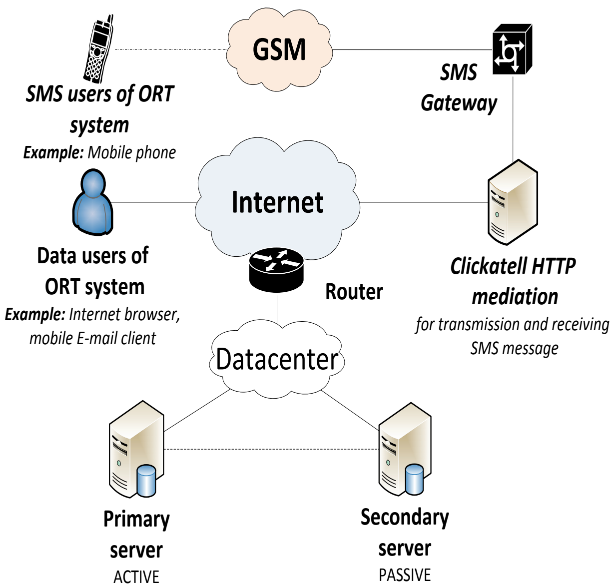

Figure 2 Connecting of telecomunication servers

MySQL database, which contains all the data and software parameters, has been synchronizing between active and passive server via "MySQL cross server replication" technology. What is more, "MySQL cross server replication" technology assures that data, which are entered on one server, are always synchronized with another server and vice versa. In the case of server malfunction, upon its return all the changes in redundant server base will be synchronized with the database of primary server. If the need for more servers occurs due to increased activity of users, it is possible to upgrade the system with more active servers which would form a load balancing group within, whose Internet traffic would be equally arranged. Router which is superior to redundant system will, on the ground of "Round-robin" algorithm, distribute user's traffic to different WEB servers whose content is generated on the ground of mutual database. Databases will, depending on load, be distributed to servers and synchronized in MySQL cluster. Primary E-mail server performs SMTP and POP3 functions. FQDN (Fully Qualified Domain Name) is hostname of server together with assigned domain and it is written in DNS zones for this domain as Mail eXchanger (MX). When one of the users sends E-mail to E-mail address of this system, in order to store and register particular training, E-mail servers inside the Internet will take care that E-mail reaches the server in POP3 where it will be processed by ORT server. By the end of processing, ORT system can, via SMTP protocol, give the acknowledgment that E-mail has been received and processed. SMTP protocol will take care that E-mail reaches the E-mail server which is allocated for domain on which is the E-mail address of user. Via HTTPS protocol SMSs have been exchanging along with SMS gateway.

Information about incoming SMSs is reached through periodic queries and those are directed to further processing. Results of process of incoming information are periodically or in real time being sent to telephone number of users. This telematic solution is divided on frontend and beckend operations. Frontend applications allow the users interface for input and show of data. Beckend applications are taking care of periodic executions of operations by reading incoming E-mails and SMSs and execution of other operations in the system which is requested by the users via frontend. Beckend applications are also allocated for critical, pro-active monitoring of system condition, with the purpose that users would not see eventual system breakdowns. Mediation between beckend and frontend is database. That split enables the possibility of technology mix in which the system is programed.

5 Calculation of $VO2_{max}$ factor and generation of the program

This module is designed in such a way that through the form of our expert system it creates the desired program for preparation of athletes, long runners for one of the following disciplines:

- 5000 meters
- 10 000 meters
- 21 097 meters (half marathon)
- 42 195 meters (marathon).

For each of the listed disciplines you can choose 3 difficulty generated programs, Program A (beginners), Program B (advanced runners), Program C (semi-elite and elite runners). The user can also choose the length of his preparation, between 12 ÷ 24 weeks, and the first day of the preparation. Current weekly mileage is recorded as an input parameter and the users 'starting point', and his current potential are calculated. We decided to determine $VO_{2\max}$ factor, that is, to determine the maximal aerobic capacity.

5.1 Calculation of maximum aerobic capacity $VO_{2\max}$

When talking about the ability of the body to perform aerobic work, we refer to the ability of taking oxygen, transfer of oxygen from the lungs into the blood and its utilization in muscles. All together it was an attempt to convert the formulas and figures that this ability could be measured, monitored and changed. That is how is devised the concept of maximum oxygen uptake or $VO_{2\max}$ receipt as a measure of fitness.

Maximal aerobic capacity $VO_{2\max}$ could be defined as the maximum amount of oxygen that the heart muscle can send and which could be used to produce energy.

$VO_{2\max}$ is expressed relative to the weight of the athlete. The units in which it is stated are milliliter of oxygen per kilogram of body weight per minute (ml/kg /min).

David E. Martin and Peter N. Coe [4] developed the following expression:

$$VO_{2\max} = (f_{c\max} \times Q_{s\max}) \times \max(a - VO_2) \text{ difference}, \quad (1)$$

where:

- $f_{c\max}$ – maximum heart rate
- $Q_{s\max}$ – maximum stroke volume
- $(a - VO_2)$ = difference of transmitted and consumed oxygen of the arterial and venous blood

An average man in his twenties who lives a sedentary lifestyle has $VO_{2\max}$ between 44 ÷ 51 (ml/(kg/min)), and a woman of the same age 35 ÷ 43 (ml/(kg/min)). People who train for long-distance running, at levels from recreational to elite have values of 43 ÷ 73 (ml/(kg/min)) for women and of 51 ÷ 84 (ml/(kg/min)) for men. Although accurate measurement and determination of aerobic capacity ($VO_{2\max}$) is performed in the laboratory, due to the impossibility that all users of the system undergo this test, we had to use other ways of determining $VO_{2\max}$ factors. To calculate the current maximum aerobic capacity, there is a more available (and better) method, "Balke test", "Cooper Test", "Astrand test", etc. [13]. We chose the method of calculating $VO_{2\max}$ factors proposed by Daniels and Gilbert of the year 1979 [2].

From all well-known (and recognized) method, this is the most difficult to calculate, but it provides the most accurate data. Maximal aerobic capacity $VO_{2\max}$ is calculated based on the recently finished race, using the following formula:

$$p_{\max} = 0{,}8 + 0{,}1894393 \cdot \mathrm{e}^{-0{,}012778 \cdot time} + \\ + 0{,}2989558 \cdot \mathrm{e}^{-0{,}1932605 \cdot time}. \quad (2)$$

$$VO_2 = -4{,}60 + 0{,}182258 \cdot velocity + 0{,}000104 \cdot velocity^2. \quad (3)$$

$$VO_{2\max} = \frac{VO_2}{p_{\max}}, \quad (4)$$

where:

- *time* – running time in minutes
- *velocity* – running speed in meters per minute
- $p_{\max}$ – percent_max.

and then with the help of "Newton-Raphson" method [12] are calculated the potential achievements at other distances from 800 meters to the marathon (42,195 meters), as follows:

With known $VO_{2\max}$, now we predict the "time" in (2) and calculate $p_{\max}$. From this information and known $VO_{2\max}$ we calculate VO_2. Now that we know VO_2, we can solve quadratic equations in (3). It takes only positive solution of the two possible solutions which are given by quadratic equation. From the known solutions for velocity (meters per minute) and the desired distance, 3000 meters, we calculate the *time* = *distance*/*velocity*. Now we compare the calculated time and what we have planned in the expression (2).

The "time" is corrected inside the loop until it gets to a deviation of less than 0,001 % respectively, to roughly match the time. Now one takes another distance, for example, 5000 meters, and repeats the iterative procedure.

Example: For run distance of 5000 meters in the time of 17'30", calculated $VO_{2\max}$ = 58,20 (ml/(kg/min)). With above-mentioned Newton-Raphson method one is able to calculate the potential performance on the runner's other distances (Tab. 1).

Table 1 Prediction of the results

Distance / m	Running time
800	2:23.12
1500	4:42.81
3000	10:05.95
5000	**17:30.00**
10 000	36:17.86
21 097	1:20:14.37
42 195	2:47:40.52

In order not to have to repeat the iterative process for each new user or for any change in maximal aerobic capacity, which is a common occurrence in the training process, they are pre-calculated and stored in the database, the data for $VO_{2\max}$ from 30 to 90.

Part of such table is shown in Fig. 3.

On the basis of these data users are allowed to enter their own recently realized results on some of the proposed distances:

- 5000 meters
- 10 000 meters
- 21 097 meters (half marathon)
- 42 195 meters (marathon).

By selecting and entering a distance and a running time, the system automatically displays to them their potential at other distances. By that help, they could determine the quality of the running speed in appropriate trainings or racings.

id	VO2max	800m	1500m	3000m	5000m	half_marathon	marathon
1	30	04:16.2	08:30	17:56	30:40	02:21:04	04:49:17
2	31	04:08.4	08:15	17:27	29:51	02:17:21	04:41:57
3	32	04:02.4	08:02	16:59	29:05	02:13:49	04:34:59
4	33	03:55.6	07:49	16:33	28:21	02:10:27	04:28:22
5	35	03:43.2	07:25	15:45	27:00	02:04:13	04:16:03
6	36	03:36.3	07:14	15:23	26:22	02:01:19	04:10:19
7	37	03:32.3	07:04	15:01	25:46	01:58:34	04:04:50
8	38	03:27.7	06:54	14:41	25:12	01:55:55	03:59:35
9	39	03:23.7	06:44	14:21	24:39	01:53:24	03:54:34
10	40	03:18.7	06:35	14:03	24:08	01:50:59	03:49:45
11	41	03:14.1	06:27	13:45	23:38	01:48:40	03:45:09
12	42	03:10.1	06:19	13:28	23:09	01:46:27	03:40:43
13	43	03:06	06:11	13:11	22:41	01:44:20	03:36:28
14	44	03:02	06:03	12:55	22:15	01:42:17	03:32:23
15	45	02:59.4	05:56	12:40	21:50	01:40:20	03:28:26
16	46	02:56.2	05:49	12:26	21:25	01:38:27	03:24:39
17	47	02:52.1	05:42	12:12	21:02	01:36:38	03:21:00

Figure 3 Display of the part of $VO_{2\max}$ table

Subsystem for forecasting results based on known or calculated aerobic capacity $VO_{2\max}$ is shown in Fig. 4 (used terms are written in Croatian language).

Online Running
Trener

Online Running Korisnik Administrator
Prijavljeni ste kao **admin** Odjavi se

Prognoziranje rezultata na osnovu aerobnog kapaciteta

Maksimalni aerobni kapacitet: VO2max = 58 (ml/kg/min)

VO2max	800 m	1500 m	3000 m	5000 m	10000 m	Polumaraton	Maraton
58	02:24	04:44	10:08	17:33	36:24	01:20:30	02:48:14

Figure 4 Forecasting results

5.2 Generating programs for the preparation

Once we have calculated the maximum aerobic capacity, the user has to select the target discipline for which he/she wants to prepare, the length of the program (between 12 and 24 weeks), set the upper limit of acceptable weekly mileage (time slot), and decide whether to train as a novice (Program A), as advanced runner (Program B) or as an elite runner (Program C).

Based on the created base of training elements for particular types of the long running trainings, the expert system formes the program of the preparation, taking into account that the offered trainings are adequatly selected for the particular discipline, that they are arranged in the proper sequence within someone's week training period, that the suggested running speeds of individual training are aligned with one's maximal aerobic capacity, that the mileage (total minutes) of individual training is in accordance with the capabilities and preferences of an individual, and the desired progress and timing for the day (week) is formed at the end of the training period. In this way the formed program is printed out on the computer screen and is stored in the database and Data Warehouse. The user is allowed to print it out when he needs an active generated program, or any of the pre-generated programs that are stored in the Data Warehouse.

Part one of this program prepares for the marathon (first 3 weeks). Adapted for printing on paper it is shown in Fig. 5 (used terms are written in Croatian language).

Program A-početni za 42.195m, VO2max=49.81(ml/kg/min)

Korisnik: John Doe

Tjedan	Datum	Ponedjeljak	Utorak	Srijeda	Četvrtak	Petak	Subota	Nedjelja
1	01/06/2009 07/06/2009	Lagano trčanje 25min @5:16	3 km UG 2 x (400 m + 1000 m + 2400 m) 1:38 min / 4:20 min / 10:51 min 240 sek odmor 3 km HL	Lagano trčanje 25min @5:16	3 km UG 4 x (200 m + 400 m + 0 m) 0:49 min / 1:44 min / 0:0 min 90 sek odmor 3 km HL	Lagano trčanje 25min @5:16	Odmor (Off)	Dugo trčanje 90min @5:16
2	08/06/2009 14/06/2009	Alternativni trening (XT)	4 km UG 8 x 400m / 1:38 min 120 sek odmor 4 km HL	Lagano trčanje 25min @5:11	4 km UG 7 x 1000m / 4:31 min 60 sek odmor 4 km HL	Lagano trčanje 25min @5:11	Lagano trčanje 25min @5:11	Trčanje u tempu 3 km UG 20min @4:31 3 km HL
3	15/06/2009 21/06/2009	Lagano trčanje 25min @5:6	4 km UG 8 x 400m / 1:38 min 120 sek odmor 4 km HL	Lagano trčanje 25min @5:6	4 km UG 8 x 200m / 0:44 min 120 sek odmor 4 km HL	Lagano trčanje 25min @5:6	Odmor (Off)	Dugo trčanje 90min @5:11

Figure 5 Generated program (first 3 weeks)

It is necessary to check if the data already exist in the warehouse and if there are active programs for the same user for the specified period of time. If there is an overlap, it is necessary to disable all existing programs, but without the last one generated.

6 Dimensional data model

Since it is necessary to preserve all generated programs for a great number of users, a classic, transactional database, which was developed by MySQL technology, is not sufficient. The system for Data Warehousing was created, which was built on the factual basis, on which are radially attached appropriate dimensions [11].

In Fig. 6 are shown some facts of this dimensional data model.

This model of Data Warehousing is specifically adjusted to OLAP and Data Mining analysis, which is in more details described hereafter.

7 Used data mining methods

7.1 Running log

Second important module of this system is Running Log. Along with usual data of every training (when, how far, how fast, type of training, etc). We would like to enable the possibility of valorisation of percieved rate of feeling of the training based on experienced fatigue, sleep deprivation, temperature, wind and other factors, using Visual Analogue Scale (VAS), often used in medicine.

This scale enables to pick a value between 0 and 10; 0 being ideal factors and 10 the worst imaginable. Choosing an appropriate day of training in the Log and filling in the fields is possible also backwards in time with the purpose to fill as many correct data as possible. In that way better filling of Data Warehouse is possible and better and more precise results analysis.

By entering data in the Log database is filling and by Real-Time or Near-Real Time ETL (Extract, Transform, Load) Data Warehouse is filled too. In that way preconditions for OLAP and Data Mining are made.

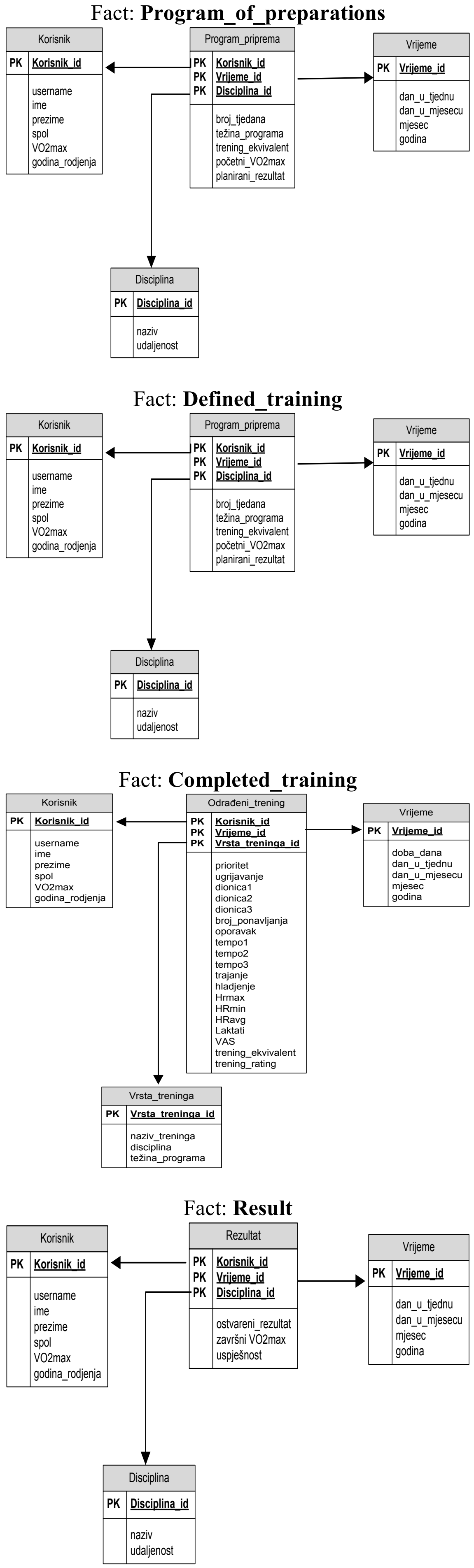

Figure 6 Dimensional data model

7.2 Data mining

Our idea is that the system autonomously (or on the request of the user), by Data Mining, valorises anticipated and realised training and predicts overtraining, undertraining and injury possibilities. On the basis of that, the system then makes a new (corrected) training plan for the rest of the training plan.

For Data Mining we used Meta-Learning [6], EDA (Exploratory Data Analysis) and Neural Networks [8, 9]. We developed programme modules that permanently analyse runners with "active" programs and periodically offer corrections after signing in or user demand. Combination of the mentioned Data Mining concepts and methods developed a new method suited for long distance runners. Besides all the trainings in the original training plan, occasional control trainings are incorporated, by which one can calculate and extract new maximal aerobic capacity values ($VO_{2\max}$). That positive (or negative) change is dynamically incorporated in the rest of the training plan, making faster progress in training. Our method is not based on the analysis of one user data but of all users and it tries to find patterns in training of similar runner profiles. The longer the Data Warehouse is filled, the more precise the system gets in corrections

7.2.1 Neural networks

Neural networks training is iterrative method where on the entrance the entered data is present and the expected exit is defined, where neural network is making connections between neurons more difficult (W_0 ...W_n) (Fig. 7).

Learning can be:

- On-line: for every single example complex connections between neurons are made
- Batch: all the entered data are present to the network and after that complexity of neuron connection is made

Learning can be supervised or unsupervised [1].

In supervised learning every entrance network gets expected exit and in unsupervised learning exits are unknown.

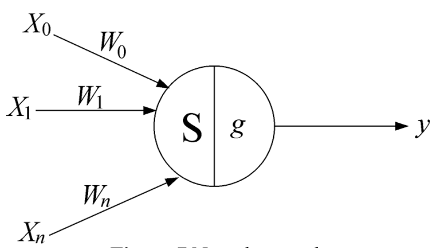

Figure 7 Neural network

In our system (ORT) complexity of neural connections are training effects and expected exit is racing result made in artificial neural network (ANN) calculated by this equation [9]:

$$y(x) = g \cdot \left(\sum_{i=0}^{n} w_i \cdot x_i \right). \tag{5}$$

After training data fom the user are returned to our ORT system via E-mail, WEB, SMS or Smart Phone applications (Fig. 8) or using sports watch (Polar, Garmin, Timex, … Fig. 9) training equivalents of every single training are calculated and iterrative exits from every layer (neuron) are calculated and tracked to the goal (race results).

Figure 8 Communication in real-time

Figure 9 Polar RS800CX

The main goal in that learning phase is to attenuate mean quadratic error and to get as close as possible to planned race result. The result of learning is to correct complexity between neurons e.g. training equivalent betwen trainings.

This results in changing of trainings and dynamic changing of the rest of the training plan and informing the user about it (WEB, SMS, E-mail).

The second phase of ANN is data analysis which cannot be separeted from learning phase.

Learning phase (in one user group) results in correction of complexion factors in training (in that group) but also accumulates knowledge in the system, which enables new group to enter analysis phase with better trainings which would enable smaller mistakes and better training results.

This phase can also be considered as learning phase for next user groups.

7.2.2 Clustering

Clustering is the method of grouping data set in similar groups or clusters and this is an example of one of the descriptive modelling. Clustering and classification are classification methods but clustering is unsupervised and classification is not.

Here are some limits of clustering:

- Flexibility - many ways of method implementation
- Hierarchic and non-hierarchic approach: data sets can be hierarchically organized or grouped on the basis of in advance defined numbers
- Subjectivity-different problems, different approach
- Interpretativity - observations are grouped by the same measure
- Speed - some clustering methods are time consuming, especially for big data sets
- Limited analysing quantity - some of the clustering methods are limited to the number of observations.

7.2.3 Hierarchical agglomerative clustering

Hierarchical agglomerative clustering is one of two most known methods of clustering of supervised subjects of experiment. Bottom-up method is used where initially every object is put in different cluster.

Later, these groups are clustered to bigger clusters on the basis of similarity. This method is limited to aprox. 10 000 objects but this is not too big in ORT system, maybe later.

Every runner is compared to all the registered users and similar runners are found based on gender, age, disciplines, etc.

When found all the data is checked for all the users of the same cluster. In the case of repeating training sequence of unsuccesfull runner, the method is changed and the system uses different method previously evaluated by ANN or OLAP.

8 Verification of results

This system has been verified on the selected sample of athletes, in order to evaluate the quality of methods and algorithms implemented in the system, with the purpose of some of the methods being eventually corrected or supplemented. The selected athletes of both genders were from 18 to 65 years of age. Among this group there were beginners, average marathon runners, as well as experienced elite marathon runners, winners of marathon and ultra marathon races.

They were divided into two control groups, the first one that had started training earlier and the second one that started three months later.

In the first group of examinees the departures from assumed and achieved results were in ±7 % interval, with $VO_{2\max}$ factor alterations of ±3 %. In order to achieve these results, for more than 50 % of examines it was necessary to dynamically correct the training process. Most of the corrections were applied after the first few of micro cycles and only two corrections were made in the last third of the training process.

In the second group of examinees the departures from assumed and achieved results were in ±5 % interval, with $VO_{2\max}$ factor alterations of ±2 %.

All necessary corrections of implemented algorithms and evaluation of certain training elements were automatically implemented in the system. It has been confirmed that "self learning" ability of this expert system after longer usage enables even better convergence of assumed and achieved results, with a minimum chance of injury during the training process.

9 Conclusion

This paper presents the approach for generating the training program for long-distance runners. It encompasses the programs for all popular long-distance disciplines, from 5000 meters up to the marathon. The procedure is automated, but in a way that maintains the flexibility and broadness to the end-user. The user himself decides about the volume, beginning and intensity of the training.

After determination of the starting point, which is maximal aerobic capacity at the beginning of the training process ($VO_{2\max}$), the user is offered a preparation program for the targeted discipline or race. This does not designate the end of the user's interaction with the system. Another important subsystem, developed with the monitoring purpose of achieved results within the training cycle, is the Running Log.

The researches of this paper were directed towards combining of both segments, the generated programme and the Running Log in the mutual Data Warehouse, as well towards using the modern telecommunication platform for real-time interaction between the user and the system.

Along with the existing methods of Data Mining and OLAP analyses, we have developed the data model adapted to the specific needs of athletes, which serves as a user's support in a real-time frame. Open source technologies and modern telecommunication channels have been used. We have also developed new methods for preparing and loading of the data in the Data Warehouse, which enable, by Data Mining, identifying of aberrations from prescribed program, dynamically corrected according to the observed requirements. Verification of the results of the real athletes is another way of implementing valuable information to the Data Warehouse, which enables the system to learn and improve its own algorithms for program generation and evaluation of results. In the past, the lack of ability of classical operational systems to present behaviour of observed model was one of the key promoters in the development of the Data Warehouse.

This kind of Warehouse, as well as the implemented methods of Data Mining and OLAP analyses, gives us the opportunity to analyse records of the state of a given athlete in the past and to search for distinctions and similarities of behaving in comparison to the other athletes in a certain time span. That kind of analysis would be impossible only by using the classical OLTP systems. In that way the time dimension of Data Warehouse becomes an important element in improvement of quality and quantity of this expert system.

This developed system has been implemented for training and monitoring of athletes (distance runners), but with adequate modifications it could be adjusted for some other sports or business fields.

10 References

[1] Berry, M. J. A.; Linoff, G. S. Mastering Data Mining, Wiley: New York, 2000.

[2] Daniels, J. T.; Gilbert, J. Oxygen power: Performance tables for distance runners. Tempe, AZ: Oxygen Power., 1979.

[3] Daniels, J. T. Daniels' Running Formula, Champaign, IL: Human Kinetics, 1998.

[4] David, E. M.; Coe, N. P. Better Training for Distance Runners, Champaign, IL: Human Kinetics, 1997.

[5] Fieltz, L.; Scott, D. Prediction of Physical Performance Using Data Mining, Research Quarterly for Exercise and Sport, March 2003, v74 i1 pA-25.

[6] Han, J.; Kamber, M. Data Mining: Concepts and Techniques, 2001.

[7] Havaš, L.; Vlahek, P. Road running (in Croatian), TK Međimurje, 953-95603-0-6, Čakovec, 2006.

[8] Neural_network, http://en.wikipedia.org/wiki/Neural_network, 2012.

[9] Neural_network_theory, http://fann.sourceforge.net/report/, 2012.

[10] Runner's World SmartCoach, http://www.runnersworld.co.uk/racing/runners-world/, 2012.

[11] Kimball, R.; Ross, M. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, (Second Edition), John Wiley & Sons, Inc., 2002.

[12] Tjalling J. Ypma, Historical development of the Newton-Raphson method, SIAM Review 37(4), 531-551, 1995.

[13] $VO2_{\max}$_evaluation_tests, http://www.brianmac.co.uk/vo2max.htm/, 2012.

Authors' addresses

Ladislav Havaš, M.Sc.
Polytechnic of Varaždin
Križanićeva 33
42000 Varaždin, Croatia
E-mail: ladislav.havas@velv.hr

Zoran Skočir, PhD., Full Professor
University of Zagreb
Faculty of Electrical Engineering and Computing
Unska 3
10000 Zagreb, Croatia
E-mail: zoran.skocir@fer.hr

Vladimir Medved, PhD., Full Professor
University of Zagreb
Faculty of Kinesiology
Horvaćanski zavoj 15
10000 Zagreb, Croatia
E-mail: vladimir.medved@kif.hr