Noname manuscript No.
(will be inserted by the editor)

Topology-based generation of sport training sessions

Iztok Fister Jr., Dušan Fister, Iztok Fister

Received: date / Accepted: date

Abstract Recently, sports training sessions have been generated automatically according to the TRIMP load quantifier that can be calculated easily using data obtained from mobile devices worn by an athlete during the session. This paper focuses on generating a sport training session in cycling, and bases on data obtained from power-meters that, nowadays, present unavoidable tools for cyclists. In line with this, the TSS load quantifier, based on power-meter data, was applied, while the training plan was constructed from a topology of already realized training sessions represented as a topological graph, where the edges in the graph are equipped with the real length, absolute ascent and average power needed for overcoming the path between incident nodes. The problem is defined as an optimization, where the optimal path between two user selected nodes is searched for, and solved with an Evolutionary Algorithm using variable length representation of individuals, an evaluation function inspired by the TSS quantifier, while the variation operators must be adjusted to work with the representation. The results, performed on an archive of sports training sessions by an amateur cyclist showed the suitability of the method also in practice.

Keywords optimization · topology · sport training sessions · metaheuristics

Citation details: Iztok Fister Jr., Dušan Fister, Iztok Fister. Topology-based generation of sport training sessions. Journal of ambient intelligence & humanized computing. 2020, DOI: 10.1007/s12652-020-02048-1

Iztok Fister Jr., Iztok Fister
Faculty of Electrical Engineering and Computer Science, University of Maribor, Koroška cesta 34, 2000 Maribor, Slovenia, E-mail: iztok.fister1@um.si

Dušan Fister
Faculty of Economics and Business, University of Maribor, Razlagova 14, 2000 Maribor, Slovenia

1 Introduction

To have an appropriate and quickly generated plan of sport training sessions from the armchair is a desire of many amateur athletes all over the world (Fister et al. 2019). A sport training plan is an unmissable diary that guides athletes in achieving the desired goal, either in serious competition, or just as a joy in beating friends on the track. Usually, it is constructed by human sport trainers who have a lot of knowledge in the Sport domain. However, hiring a professional sport trainer results in huge costs on the one hand, while some of them do not achieve the best harmony with the athletes (e.g., either controversial trainers or athletes) on the other. Additionally, some athletes, especially from rural areas, suffer from a lack of the advantages affordable in big cities (Fister et al. 2019).

Nowadays, the scientific community is experienced with a new Artificial Intelligence (AI) spring fostered by integration of traditional logic based AI (ontologies) and statistical Machine Learning (ML), which may lead to context-adaptive systems (Holzinger et al. 2018). On the other hand, the AI spring has fostered emerging of concept Interactive Machine Learning (iML) that looks for algorithms interacting with agents, where the agent can be human (also human-in-the-loop) (Holzinger 2016). In such algorithms, human intelligence can influence on the machine intelligence (Holzinger et al. 2019). Obviously, a part of the AI spring presents also growing a connectivity between sport science and computer science, where among others, the new ways have been explored that should allow athletes the generation of sport training sessions for a particular period of time automatically (Fister et al. 2019).

In order to generate training sessions automatically, a lot of test data are necessary (Zuo 2019). In the era of pervasive computing (Weiser 1991; Saha and Mukherjee 2003), we have several possibilities how to collect data during the sport training sessions. Modern sport trackers are just one of the many options, besides numerous wearable sport watches and heart rate monitors that are also equipped with GPS sensors. Let us mention that this technology is, nowadays, very cheap and affordable for almost every athlete. However, the more expensive devices offer more additional features (Žemgulys et al. 2019) that are not included in the cheaper equipment. For instance, the former afford a better analysis of training sessions, a step counter, sleep monitoring, etc. Sport trackers and sport watches allow us to get the whole information of training sessions. They are able to track training load indicators as: Heart rate, position, power, temperature, and so on.

Most of the current efforts in the domain of Automatic Generation of Training Sessions work on basis of classical training load indicators: The average heart rate, and the training duration. These can easily be obtained from heart rate monitors worn by athletes during the sports training session. Typically, the TRaining IMPulse (TRIMP), proposed by Banister (Banister 1991), was used for quantifying the training load, that is expressed as a simple product of both the aforementioned indicators. The TRIMP is used widely in practice, but its main drawback is that it is insensitive to different levels of training.

This means that longer training sessions of lower intensity (low average heart rate) have similar TRIMP values as the shorter high intensity training (high average heart rate). This drawback might affect the quality of the generated training sessions.

On the other hand, the heart rate monitors have some weaknesses, especially in cycling; for instance, slow response during short acceleration, dependence on weather, mental stress, and diet. Therefore, power-meters have been devised that are capable of gauging intensity during a ride (Allen and Coggan 2019). Although they have had a long history (e.g., they had already appeared in the late 1800s), unsuitable technology and higher cost prevented their wide usage. The price of the power-meters has been decreasing in recent years. This means that they have also become affordable for amateur athletes.

Nowadays, the equipment is also used in sport training. The main advantages of the technology can be summarized in three facts, precisely, that the power-meter (Allen and Coggan 2019):

- allows tracking in athlete's fitness changes,
- defines athlete's weaknesses easily,
- refocus training to eliminate the weaknesses.

All three facts can be mastered because the power-meter gives access to a huge amount of data. These data are captured in second-by-second intervals, and allow users to obtain the complex information.

This study focuses on the generation of sport training sessions based on data produced by heart rate monitors equipped with a Global Positioning System (GPS) and power-meters. Together, these data allow the new momentum in the generation. In line with this, a complex method is proposed consisting of more steps: At first, data from more sports activities are assembled into a collection of base training maps, which serve as a source for constructing a topological graph with nodes and edges. Each edge, representing a real path where one of the sport activities takes place, is equipped with training load indicators, like: The path length, absolute ascent, and average power. While the first two indicators are calculated from GPS data, the latter originate from the power-meter. Then, the problem is defined as constraint optimization, and solved using an evolutionary algorithm. Because the algorithm uses a variable length of solution, the special variation operators need to be defined for working on the representation. The evaluation function is inspired by a training load quantifier, i.e., Training Stress Score (TSS), and needs to be minimized by determining the boundary limits of training length and selecting a starting to ending point. Finally, the optimal path is visualized.

A case-study was performed in order to test the proposed method. The case study based on a collection of sport activities realized by an amateur cyclist. This collection served as a foundation for constructing the base training map from which the topological graph was built. Two scenarios were performed on this topological graph, while the obtained results were visualized.

The main contributions in this study are:

– a new approach is proposed for planning the sport training sessions, based on topology,
– the evaluation function is developed, inspired by the training load quantifier TSS,
– the new method is constructed and an evolutionary algorithm is defined for optimization,
– the obtained results are analyzed in detail.

The structure is as follows: Section 2 deals with the basic information needed for understanding the subjects that follow. In Section 3, the proposed method is described in detail. The experiments and results are the subject of Section 4. The paper concludes with Section 5, where the performed work is summarized and directions outlined for the future work.

2 Basics of evolutionary algorithms

Evolutionary Algorithms (EAs) are stochastic nature-inspired population-based algorithms mastering the Darwininan struggle of existence (Darwin 1859) in a search process. Similar as in natural evolution, where those individuals that are better adapted to environmental conditions have more chances to survive and pass their genetic material to the next generation using reproduction, also the better solutions have more chances to survive and reproduce in conditions of simulated evolution. In this programed environment, a problem to be solved plays the role of environment. Although there are many types of EAs, they share the same characteristics. Actually, they are distinguished among each other according to the representation of individuals. In general, there are four types of EAs, as follows:

– Genetic Algorithms (GA) (Goldberg 1989),
– Genetic Programming (GP) (Koza 1994),
– Evolution Strategies (ES) (Bäck 2010),
– Evolutionary Programming (EP) (Fogel 1999).

Typically, the EA consists of the following components (Eiben and Smith 2015):

– representation of individuals,
– evaluation function,
– population,
– parent selection,
– variation operators (i.e., crossover and mutation),
– survivor selection.

Initialization and termination conditions also need to be included In order to complete the component list. As a matter of fact, the pseudo-code of the EA is illustrated in the Algorithm 1,

from which it can be seen that an evolutionary cycle (lines 4-10) consists of the following steps: The parent selection (line 5) selects parents entering into

Algorithm 1 Pseudo-code of an evolutionary algorithm

```
1: procedure EVOLUTIONARY_ALGORITHM
2:     INITIALIZE_population_randomly;
3:     EVALUATE_each_candidate_solution;
4:     while TERMINATION_CONDITION_not_met do
5:         SELECT_PARENTS;
6:         RECOMBINE_pairs_of_parents;
7:         MUTATE_the_resulting_offspring;
8:         EVALUATE_each_candidate_solution;
9:         SELECT_SURVIVOR_solution_for_the_next_generation;
10:    end while
11:    return best_solution
12: end procedure
```

a crossover operation (line 6). However, the crossover is executed with regard to the probability of crossover p_c. Then, mutation (line 7) is launched, that modifies the corresponding individuals according to the probability of mutation p_m. The evaluation function (line 8) assesses the quality of the individual. Finally, survivor selection (line 9) determines the best individual for the next generation.

3 The proposed method for topology-based generation of sports training sessions

The proposed method for topology-based generation of sports sessions consists of four steps:

1. composition of base training map,
2. preprocessing,
3. optimization,
4. visualization.

The goal of the first step is to assemble a collection of sport training activities realized by some athlete and compose them into a base training map. From the base training map, a topological graph is built in the second step. The graph consists of nodes representing intersections of edges, while edges are incidental to the nodes. Interestingly, two nodes can be joined with more edges. Moreover, some nodes can even be joined with themselves by more edges. The length, absolute ascent, and average power are attached to each edge in the graph. The third step is devoted to topology-based generation of sport training sessions using an EA. Finally, the optimum path proposed by the EA is visualized.

3.1 Composition of base training map

In the first step, the sport activities obtained from the specific athlete realized in the same training cycle are composed into a base training map. Thus, a wider area is obtained, depicting the geographical region in which the cyclist mostly carries out the training sessions. Indeed, the non-existent paths in one activity can supplement the graphical region on the one hand, while, on the other, the paths already present in this region cause overlapping.

As a result, the sport activities in the first case widen the geographical region in which the training sessions are usually conducted, while the second leave the size of the region unchanged, but increase the volume of measured track points identifying this path. Consequently, the unreliability of data is increased, because the athlete can cover the observed path in a different direction, in different weather or psycho-physical conditions.

Typically, the base training map is visualized using OpenStreetMaps in order to verify that the selected sport activities really form a connected region that is appropriate for representation in a topological graph.

3.2 Preprocessing

The aim of this step is to build the constructed topological training graph from the base training map. Undoubtedly, this is the hardest step, because, here, the GPS data are taken into consideration, and these data are typically not consistent. Actually, this step is divided into two phases:

1. building the topological training map,
2. building the topological training graph.

The first preprocessing phase is devoted for building the topological map from the base training map. This phase is more quantitative, because of processing the large amount of track points (i.e., GPS positions) that represent the smallest part of the path, which must be synthesized into the whole. However, there are several track points referring to the same paths. In order to reduce the number of track points before synthesizing the path, some track points are either merged with the nearest ones, or even eliminated by rounding their positions to three decimals. The result of this phase is a connected topology map presenting an outline of the training region. This means that the map consists only of edges, i.e., without any vital information characterizing their nature.

On the other hand, the second phase is more qualitative, and includes processing raw data in the term of .tcx files, from which training load indicators, like latitude-longitude points, heart rate data, power-meter values, and altitude are extracted. In this phase, the nodes for the topological graph are determined by the intersection of two or more edges, while all the other vital information is assigned to each edge incident by two nodes.

Let us mention that nodes give the optimization algorithm an opportunity to find alternative paths from two nodes and, thus, the topological graph represents the search space of potential solutions for the optimization process. Indeed, this graph is managed by the EA for topology-based generation of a training plan in the optimization step.

3.3 Optimization

Topology-based generation of sport training sessions is defined as an optimization problem, where the topology is defined as graph $G = (V, E)$ consisting of a set of vertices $V(G) = \{v_1, \ldots, v_N\}$ and a set of edges $E(G) = \{e_1, \ldots, e_M\}$, together with an incidence function ψ_G that associates an unordered pair $\psi_G(e_k) = \{v_i, v_j\}$ with each edge of G. Thus, variables N and M denote the maximum number of vertices and the maximum number of edges, respectively.

The topology graph G has two main characteristics:

- a vertex can be joined by loop,
- vertices can be joined by multiple edges.

All the mentioned characteristics are incorporated in the adjacent matrix $\mathbf{A}_G = (a_{i,j})$ of dimension $N \times N$, where $a_{i,j}$ is the number of edges joining vertices v_i and v_j. If $a_{i,i} > 0$, the vertex $v_{i,i}$ is connected with a loop, while the degree of the vertex is almost two.

Additionally, three functions are assigned to each edge in the topology graph: The first measures the absolute ascent necessary to overcome the sub-path between both nodes joined by the k-th edge, the second determines the average power needed by overcoming the same sub-path, while the third its length. In other words, the absolute ascent $Asc_G(e_k)$ is expressed as:

$$Asc_G(e_k) = \sum_{l=1}^{n_T} Alt^{+}(TP_l), \tag{1}$$

where n_T denotes the maximum number of track points constituting the edge e_k, and the function $Alt^{+}(TP_l)$ is a positive altitude (i.e., uphill) of the track point TP_l. The function approximates the intensity factor, with which the stress, put on the athlete's body during a ride, is measured in the following sense: The higher/lower the altitude difference between vertices joined by an edge, the bigger/smaller the intensity factor. Actually, the absolute ascent function complies with the following relation:

$$\begin{cases} Asc_G(e_k) > 0, & \text{climbing upwards,} \\ Asc_G(e_k) = 0, & \text{no climbing.} \end{cases} \tag{2}$$

Obviously, the value of this function cannot be lower than zero due to the sum operation used in Eq. (1).

The average power $\overline{W}_G(e_k)$ is expressed as:

$$\overline{W}_G(e_k) = \frac{1}{n_T} \sum_{l=1}^{n_T} W(TP_l), \tag{3}$$

where the function $W(TP_l)$ is the power indicated by overcoming the track point TP_l.

The length of an edge $L_G(e_k)$ is expressed as:

$$L_G(e_k) = \sum_{l=1}^{n_T} L(TP_l), \tag{4}$$

where the function $L(TP_l)$ is the length of the track point TP_l.

The proposed EA algorithm demands modification of the following components:

- representation of solutions,
- variation operators,
- evaluation function,
- initialization.

Let us emphasize that the parent selection operator selects two parents, entering into crossover, randomly. As a survivor selection, the one-to-one selection was implemented, borrowed from the Differential Evolution (DE) (Storn and Price 1997). Thus, the better between the parent and offspring according to the value of the fitness function will survive into the next generation. The maximum number of generations is used as a termination condition. In the remainder of the paper, the components of the modified EA are described in detail.

3.3.1 Representation of solutions

Each solution $\mathbf{x}_i$ in population of Np individuals is defined as variable length vector expressed mathematically as follows:

$$\mathbf{x}_i = \{n_i, e_{i,1}, e_{i,2}, \ldots, e_{i,n_i}\}, \quad \text{for } i = 1, \ldots, Np, \tag{5}$$

where the variable n_i denotes the number of edges $e_{i,k}$ for $k = 1, \ldots, n_i$ that connects the starting vertex of the first edge $e_{i,1} = \{v_A, v_{i,1}\}$ with the ending vertex of the last edge $e_{i,n} = \{v_{i,n}, v_B\}$, and it holds that each ending vertex of the edge $e_{i,j} = \{v_{i,j-1}, v_{i,j}\}$ is the starting vertex of the edge $e_{i,j+1} = \{v_{i,j}, v_{i,j+1}\}$.

Actually, each solution describes a path from vertex v_A to vertex v_B that is selected by the user. In the case $v_A = v_B$, the starting point A is equal to ending point B, and we have to deal with a cycle.

3.3.2 Crossover

A crossover operator operates on two randomly selected individuals $\mathbf{x}_i$ and $\mathbf{x}_j$. Let us assume, the first solution describes the path as follows:

$$v_A \rightarrow v_{i,1} \rightarrow \ldots \rightarrow \boldsymbol{v_{i,k}} \rightarrow \boldsymbol{v_{i,k+1}} \rightarrow \ldots \rightarrow \boldsymbol{v_{i,n_i-1}} \rightarrow v_B,$$

while the second as follows:

$$v_A \to v_{j,1} \to \ldots \to \boldsymbol{v_{j,l}} \to \boldsymbol{v_{j,l+1}} \to \ldots \to \boldsymbol{v_{j,n_j-1}} \to v_B.$$

At first, the crossover point is selected randomly in the first parent $\mathbf{x}_i$. Let us assume that the crossover point represents vertex $v_{i,k}$. Then, the corresponding vertex in the second parent is found such that $v_{i,k} = v_{j,l}$ holds. Finally, the remainder of the paths are exchanged in both parents, in other words:

$$v_A \to v_{i,1} \to \ldots \to \boldsymbol{v_{j,l}} \to \boldsymbol{v_{j,l+1}} \to \ldots \to \boldsymbol{v_{j,n_j-1}} \to v_B,$$

and

$$v_A \to v_{j,1} \to \ldots \to \boldsymbol{v_{i,k}} \to \boldsymbol{v_{i,k+1}} \to \ldots \to \boldsymbol{v_{i,n_i-1}} \to v_B.$$

Let us mention that if a matching vertex is not found in the second parent, the closest common vertex in the adjacent matrix is selected and inserted into the path.

3.3.3 Mutation

A mutation operator selects one individual $\mathbf{x}_i$ in the population randomly. Then, the vertex $v_{i,j}$ is chosen randomly, such that $v_{i,j} \neq v_A$ and $v_{i,j} \neq v_B$, and the length of the mutated sub-path l. Let us assume an individual is given that defines the following path from point A to point B:

$$v_A \to v_{i,1} \to \ldots \to \underbrace{\boldsymbol{v_{i,k}} \to \ldots \to \boldsymbol{v_{i,k+l}}}_{\text{Mutated Sub-Path}} \to v_{i,k+l+1} \to \ldots \to v_{k,n_i-1} \to v_B,$$

and the v_i, k is selected as the starting vertex, while the variable l determines the ending vertex $v_i, k+l$ of the mutated sub-path. Let us assume that the alternative sub-path $v'_0 \to v'_1 \to \ldots \to v'_{n_l-1} \to v'_{n_l}$ is given, where $v'_0 = v_{i,k}$ and $v'_{n_l} = v_{i,k+l}$. Then, the alternative sub-path replaces the mutated one, as follows:

$$v_A \to v_{i,1} \to \ldots \to \underbrace{\boldsymbol{v'_0} \to \ldots \to \boldsymbol{v'_{n_l}}}_{\text{Alternate Sub-Path}} \to v_{i,k+l+1} \to \ldots \to v_{k,n_i-1} \to v_B,$$

Interestingly, if the length $l = 0$, the loop of the vertex $v_{i,k}$ is added to the path. However, if the loop is not permitted according to the adjacent matrix, the length l is incremented, and the alternate sub-path of the new length is searched for. The searching is repeated by incrementing the length until a new alternative path is found, or no mutation is performed. To avoid already visited vertices in the sub-path, Tabu search heuristics are applied (Glover and Laguna 1997).

$= 0, \ldots, n_i - 1$ is incremented. For each l, the alternative new path from $v_{i,k}$ to $v_{i,k+l}$ is searched for. If the path is not found, the sub-path length l is incremented.

3.3.4 Evaluation function

The inspiration for the evaluation function is presented by the so-called Training Stress Score (TSS) training load quantifier defined by Coggan (Allen and Coggan 2019), based on data obtained from power-meters. Actually, introducing the power-meters into cycling has launched a revolution in the process of sport training.

The quantifier TSS is calculated using the following equation (Allen and Coggan 2019):

$$TSS = \frac{TD \cdot NP \cdot IF}{FTP \cdot 36}, \tag{6}$$

where TD means a training duration, NP is the normalized power used for quantifying training intensity with power data, IF is an intensity factor, used for determining the power intensity zone, and FTP is a functional threshold power. While the FTP depends strongly on the characteristics of the athlete's body, IF is the ratio of the NP and the athlete's FTP, that determines the intensity with which the prescribed training session must be performed.

Unfortunately, the TSS cannot be used in the evaluation function in its complete form, because data for constructing the topological graph were obtained by merging data from more sport activities realized at different times. This means that these were not conducted under the same conditions. For instance, the part of a path represented by an edge e_k can be overcome by the rider at the beginning of the training session in the first case, and at the end of this in the second, where he/she is already tired. On the other hand, the direction of overcoming can be opposite in different sport sessions. Also, weather conditions are not the same when comparing two different sport activities with overlapping parts of routes.

All these issues caused that we introduced the average power $\overline{W}_G(.)$ instead of normalized power NP, while the intensity factor IF was replaced by the training load indication $1/Asc_G(.)$. The duration of a training activity TD is very unreliable and, therefore, we eliminated from the equation. As a result, the evaluation function is expressed as:

$$f(\mathbf{x}_i) = \sum_{j=1}^{n_i-1} \frac{\overline{W}_G(\psi_G^{-1}(v_{i,j}, v_{i,j+1}))}{FTP \cdot Asc_G(\psi_G^{-1}(v_{i,j}, v_{i,j+1}))}, \tag{7}$$

subject to

$$MIN\_LEN \leq \sum_{j=1}^{n_i} L_G(\psi_G^{-1}(v_{i,j}, v_{i,j+1})) \leq MAX\_LEN, \tag{8}$$

where $v_{i,0} = v_A$ and $v_{i,n_i+1} = v_B$, respectively, the function ψ_G^{-1} is an inversion function assigning to each ordered pair of vertices $(v_{i,j}, v_{i,j+1})$ the corresponding edge e_k, and MIN_LEN and MAX_LEN determines the minimum and maximum length of the generated sport training, respectively.

The goal of the optimization is to minimize the value of the evaluation function. Actually, the proposed function prefers the solutions of the higher average power with more climbing upwards.

3.3.5 Initialization

The topology-based generation of sports training sessions demands the presence of feasible solutions in a population. This means that each of the individuals must prescribe a valid path from point A to point B. Therefore, individuals in the population of solutions need to be generated heuristically. One of the generated solutions, however, is generated using Dijkstra Shortest Path First (SPF) (Dijkstra 1959), while the other uses the Tabu search algorithm to avoid selecting the already visited vertices by creation of the new individuals.

3.4 Visualization

The last step is devoted for presenting the numerical data obtained after the topology-based generation of a training plan in a graphical presentation. The graphical presentation enable athletes to discover insights into the relationships and patterns in the numerical data, and, thus, avoid deception and confusion in their interpretation. The natural way of presenting these data is a topological outline of the path that needs to be overcome by the athlete. This path is put onto the topological training graph that represents the context of the drawing.

Although the more natural way could be presentation using OpenStreetMap, the advantage of this approach is that this is embedded into the topological graph with already marked edges and nodes, and, therefore, is closer to the subordinate numerical data.

4 Experiments and results

The goal of the experimental work was to show that the method of topology-based generation of training sessions is capable of building training plans and, thus, is appropriate for application in practice. In line with this, a case-study was prepared with which the proposed method was tested. The experiments were applied to a collection of three sport activities realized by an amateur cyclist during one training cycle (i.e., season). The selected sport activities covered a geographical region of moderate size, where the corresponding topological graph was awkward enough for an evolutionary search process on the one hand, and was not too complex for visual representation on the other.

Actually, two scenarios were tested, as follows:

- Scenario A: short-distance training,
- Scenario B: endurance training.

The section is organized in such a way that the results of all four steps of the proposed method can be tracked in detail. At the end of the section, the performed work is taken into deep consideration.

4.1 Step 1: Composition of the base training map

In summary, the collection of the sport activities must be selected very carefully, because they must supplement each other on the map. This means that each activity must have some mutual contact points, because, only in this way, can the graphical region covered by the sport activities increase. In order to ensure that the user verifies the regularity of his/her selection, this area is visualized using OpenStreetMap.

The sample of the base topological map used in our case-study is illustrated in Fig. 1, from which it can be seen that the collection of three sport activities

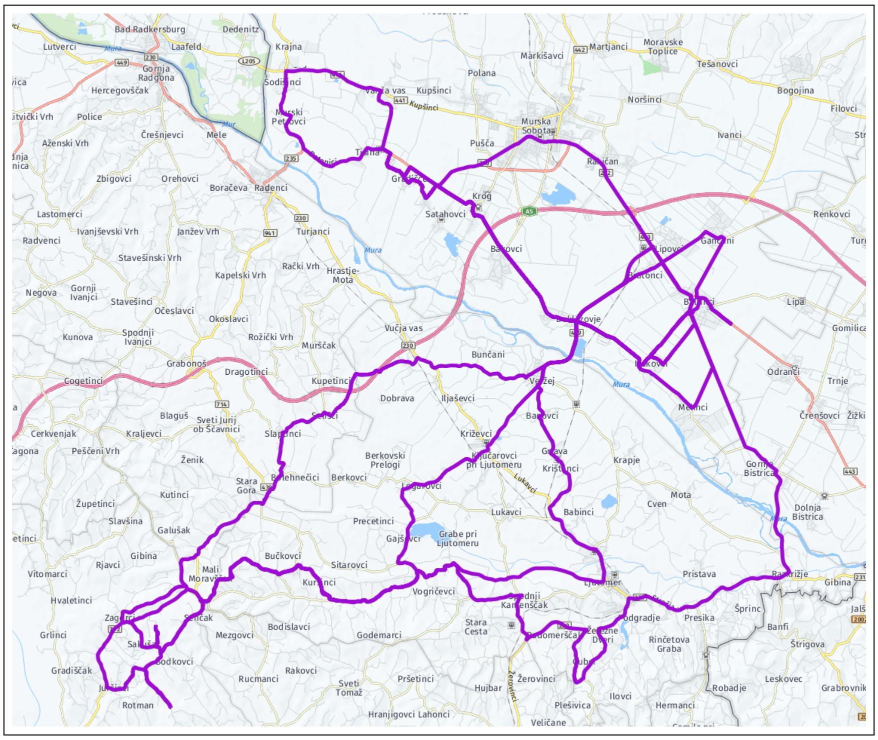


Fig. 1: Base training map: This map has been composed of three different cycling training sessions that were realized by an amateur athlete.

really builds a grid of paths and intersections between two or more paths. Each path represents a sequence of track points, with their GPS locations, altitudes, duration and intensity. In summary, the observed base topological map covers the geographical region consisting of connected nodes and there are no islands detected.

4.2 Step 2: Preprocessing

The result of the first preprocessing phase is a topological training map represented as an undirected graph with edges, where no nodes are outlined. The resulting topological training map in the case-study is depicted in Fig. 2. In

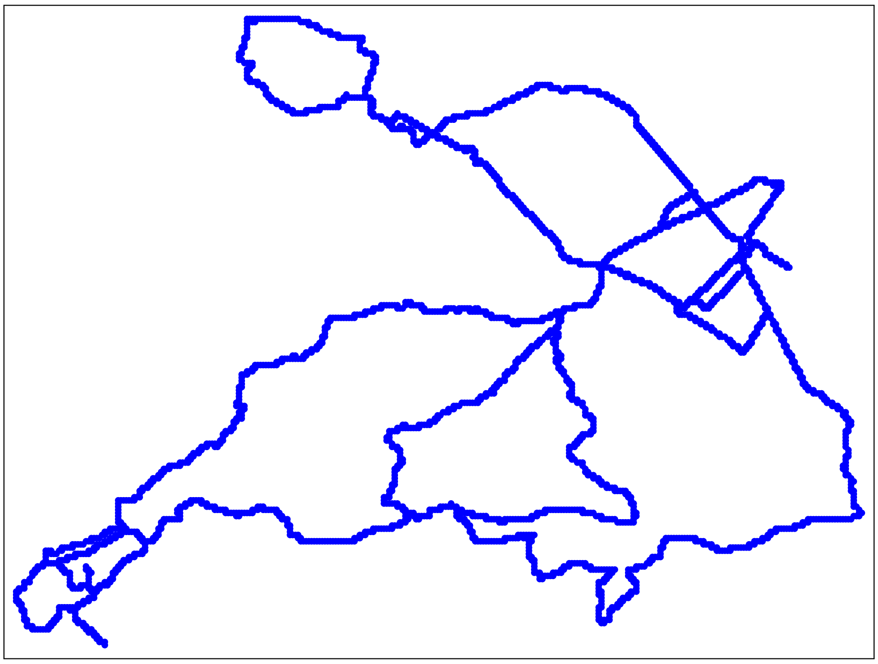


Fig. 2: Topological training map: Obtained after the first preprocessing phase without vital information, like nodes, lengths of edges, and altitudes.

order to determine the nodes, the second preprocessing phase is executed, where a node is identified as an intersection of two or more edges. The results of this phase for our case-study is a topological training graph illustrated in Fig. 3, from which it can be seen that the topological training graph in question consists of 24 nodes and 40 edges. As a result, the nodes constitute a connected graph without islands, while each edge is equipped with all the vital information needed for the optimization step.

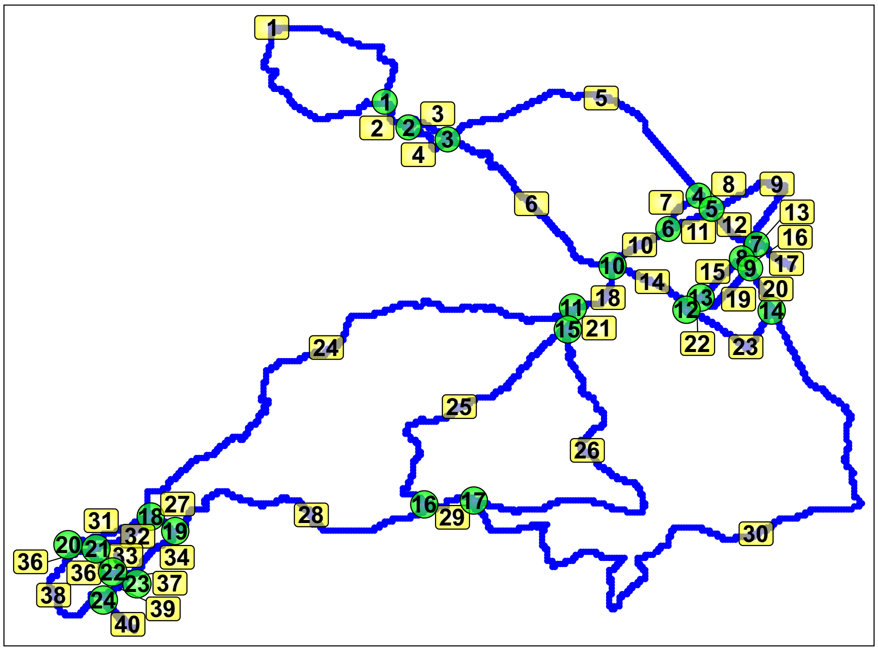


Fig. 3: Topological training graph: Obtained after the second preprocessing phase, represented as a graph equipped with all vital information needed by optimization.

4.3 Step 3: Optimization

The purpose of this step is to generate the sport training sessions based on the topological training graph using the EA. The parameter setup during the experimental work was set as presented in Table 1.

Table 1: Parameter setup of the EA for topology-based generation of sport training sessions.

Nr.	Parameter	Abbreviation	Value
1	Population size	Np	100
2	Maximum number of generations	MAX_GEN	500
3	Probability of crossover	p_c	1.00
4	Probability of mutation	p_m	0.01
5	Length of individual	n	$\|e_k\| + 1$
6	Minimum length of training plan	MIN_LEN	$\{40, 90\}$
7	Maximum length of training plan	MAX_LEN	$\{50, 110\}$
8	Functional threshold power	FTP	220

As can be seen from Table 1, the parameters are divided into two groups: (1) The algorithm's ("Nr." 1-4 in the table), and (2) problem's parameters ("Nr." 5-8 in the Table). The former controls the behavior of the EA, while the latter determines the characteristics of the problem. For instance, the EA works with a population of 100 individuals, each individual in the population undergoes the operation of crossover, while the crossover operator is launched in each generation, on average once. As a termination condition, the maximum number of generations $MAX\_GEN = 500$ is employed.

On the other hand, the length of the individual is unknown in advance, and depends on the potential generated solution. Constraints MIN_LEN and MAX_LEN were set as depicted in the Table, where the first value for both variables was applied in Scenario A and the second in Scenario B. The functional threshold power depends on the characteristics of the athlete in question, and was set to $FTP = 220$ in our case-study.

The best solution according to the fitness value obtained after 25 runs by the EA for topology-based generation of sport training sessions in Scenario A is presented in Table 2 that is divided into six columns representing the step

Table 2: The best training plan for Scenario A.

Step	Edge	Nodes	$\overline{W}_G$ [W]	Asc_G [km]	L_G [km]
1	e_9	(v_7, v_5)	204	4.0	6.610
2	e_{11}	(v_5, v_6)	194	4.0	1.893
3	e_7	(v_6, v_4)	177	8.0	2.142
4	e_5	(v_4, v_3)	198	11.0	12.831
5	e_6	(v_3, v_{10})	223	6.0	9.575
6	e_{14}	(v_{10}, v_{12})	247	0.1	3.480
7	e_{23}	(v_{12}, v_{14})	212	2.0	5.391
8	e_{20}	(v_{14}, v_9)	219	2.0	2.064
9	e_{16}	(v_9, v_8)	189	0.1	0.598
$\sum$			209	4.1	44.584

number ("Step"), edges ("Edge"), incident nodes ("Nodes"), average power ($\overline{W}_G$ [W]), absolute ascent (Asc_G), and length of the edge (L_G). In the last row, the summary values of the corresponding columns are added. Let us mention that the sum of average power values was expressed as a weighted average, where the lengths of the edges were adopted as weights. That is, the longer the path, the more effect the adopted intensity has on the weighted average.

In summary, the best path in Scenario A consists of nine edges, describing a path of length 44.584 km, almost without uphills (i.e., $Asc_G = 4.1$ m) that was overcome with an intensity close to the athletes FTP (precisely 95 % of FTP). This means that he/she rode the bike in the so-called "Lactate threshold" power intensity zone (Allen and Coggan 2019), where the average power is between $91 - 105$ % of the personal FTP. The heart rate corresponding to this power intensity zone is between $95 - 105$ % of the personal VO2max (Allen and Coggan 2019).

The best training plan in Scenario B presented in Table 3, shows that a path of the optimal training session consists of 15 edges. Thus, the optimal

Table 3: The best training plan for Scenario V.

Step	Edge	Nodes	$\overline{W}_G$ [W]	Asc_G [km]	L_G [km]
1	e_9	(v_7, v_5)	204	4.0	6.610
2	e_{11}	(v_5, v_6)	194	4.0	1.893
3	e_{10}	(v_6, v_{10})	225	2.0	3.078
4	e_{18}	(v_{10}, v_{11})	209	2.0	2.706
5	e_{24}	(v_{11}, v_{18})	178	97.0	23.641
6	e_{32}	(v_{18}, v_{21})	168	1.0	2.988
7	e_{33}	(v_{21}, v_{22})	218	6.0	1.845
8	e_{37}	(v_{22}, v_{23})	142	0.1	0.376
9	e_{34}	(v_{23}, v_{19})	168	22.0	3.184
10	e_{28}	(v_{19}, v_{16})	174	23.0	12.073
11	e_{29}	(v_{16}, v_{17})	187	8.0	2.276
12	e_{30}	(v_{17}, v_{14})	198	104.0	33.557
13	e_{20}	(v_{14}, v_9)	219	2.0	2.064
14	e_{19}	(v_9, v_{13})	39	7.0	3.147
15	e_{15}	(v_{13}, v_8)	209	0.1	2.772
$\sum$			186	18.8	102.209

path has a length of 102 km that goes uphill for almost 19 m. This course was overcome with the average intensity of 186 W, thich is substantially lower than the athletes FTP (precisely 84 %). Actually, this value corresponded to the so-called "Tempo" power intensity zone, where the intensity is between $76 - 90$ %, which coincides with the heart rate zone between $84 - 94$ % (Allen and Coggan 2019).

Interestingly, a closer look at the column average power shows that the worst result was achieved in step 14, where the average power $\overline{W}_G = 39$ was held by overcoming even 7 km. As a matter of fact, this part of the course has a great potential to be improved in the future.

4.4 Step 4: Visualization

In this subsection, the numerical data presenting the optimal paths for both scenarios in our case-study as obtained by the EA for topology-based generation of sport training sessions are visualized using the existing topological training graph. The results of the visualization are depicted in Fig. 4 that is divided into two diagrams. The former illustrates the optimal path obtained in Scenario A, while the latter in Scenario B.

In both diagrams, the bubble A describes the starting node and the bubble B the ending node. As can be seen from Fig. 4a, the course is relatively short, and, therefore, suitable for high-intensive training sessions. This course is also flat, as can be viewed from the last test.

Contrarily, the course displayed in Fig. 4b is much longer, and, thus, more appropriate for endurance training sessions, where the optimal course is longer (> 100 km) and suitable for moderately-intensive training sessions.

4.5 Discussion

The accomplished experimental work using the method for automatic generation of sport training sessions showed the advantages, as well as disadvantages, of the proposed approach. The advantages are supported with evidence that this kind of generation is possible, while the results are appropriate for real-world application. On the other hand, these are represented visually, together with all the information needed by the athlete for whom the sport training session is planned.

However, the experiments also exposed the weaknesses of the method. The majority of these refer to the data and data handling. Obviously, the hardest problem was presented by the unreliability of data obtained from wearable tracking devices equipped with GPS receivers. The data are problematic by their nature, because their acquisition depends on several conditions, including weather conditions, GPS accuracy, and noise.

Weather conditions affect the psycho-physical state of an athlete, which, consequently, has an impact on his/her performance. Typically, athletes achieve better results in sunny rather than in rainy weather. GPS accuracy is dependent on satellite geometry, atmospheric conditions, signal blocking, and receiver design. The satellite geometry refers to the arrangement of the satellites in orbits which influences the quality of the GPS signal on Earth. The atmospheric conditions can weaken the GPS signal, while some obstacles on the Earth can also affect the GPS accuracy. Usually, the quality of a GPS receiver is proportional to its cost. That is, the more expensive the GPS receiver, the better the GPS signal.

Noise refers to capturing some track points using wearable mobile devices of different tracking quality, at different tracking times, and in different tracking directions. An athlete can track the same parts of the paths in more sport activities. Thus, different devices can be used for tracking the points onto the same courses. As a result, the quality of these devices can incorporate the noisy into regularly captured data. The time of capturing data refers to the training phase in which the data are produced. Obviously, it is not equal to ride the same part of the course at the start as to ride the same part at the end of the training session. Normally, the athlete is more powerful at the start than at the end of the training session. Also the tracking direction has a big influence on the noise, because performances of the athlete riding uphill are lesser than riding the same part of the course downhill.

In general, information about average power acquired from wearable tracking devices must be taken carefully. At the moment, due to unreliability, these data can be used as a hint in planning the sport training session, which serves the sport trainer by planning the real training session. Amore complex preprocessor of the sport activities must be developed if we want to build a pure automatic solution of the noise problems.

5 Conclusions

Recently, planning the sport training sessions was based on measuring the heart rate, that represents the training load indicator for estimating the intensity load. With the advent of power-meters, the sport training theory of cycling has achieved a new momentum. Using this new equipment mounted on the bicycle, it is possible to monitor the intensity load during the training session. Consequently, the new TSS qualifier was employed for calculating intensity load that, nowadays, has been changing the proven practices of sport training in cycling, based mostly on the TRIMP load quantifier.

This paper focuses on the generation of sport training sessions based on the topology of a geographical region where training sessions normally take place. The method consists of four steps: (1) Composition of the base training map, (2) Preprocessing, (3) Optimization, and (4) Visualization.

In the first step, the base training map is composed from an archive of sport activities obtained with a mobile tracking device worn by an athlete during the realization of a training session. On basis of the GPS data found in the collection, the preprocessor builds a topological training graph. The topological graph consists of edges representing a sequence of GPS track points, and nodes intersecting edges. Thus, each edge is equipped with vital information, such as: Path length, absolute ascent, and average power. The topological graph is manipulated in the optimization step by the EA for the topology-based generation of sport training sessions. This EA implements proprietary crossover and mutation adjusted to manipulate topological graphs. The evaluation function is inspired by the TSS quantifier. The optimal path is presented as a sequence of edges constituting the optimal path from the starting point A to the the finishing point B under prescribed constraints. The result is presented in numerical form that is inappropriate for application in practice. Therefore, the visual representation of the numerical data completes the proposed method.

A case-study was prepared for testing the proposed method, in which a collection of three sport activities was taken into consideration. In this case-study, two scenarios were tested: (1) Short-distance, and (2) Endurance training. The results showed that the method found the optimal courses for both scenarios in the sense of course length, but the proposed intensities must be observed carefully by a real sport trainer due to noise in the input data.

There are a lot of directions in which the proposed method can be developed in the future. Mainly, the topological training graph should be changed from an undirected to a directed graph. However, this change affects mostly two steps, i.e., preprocessing, and optimization. Additionally, the method could also be applied in other sports, where power-meters are part of the sport training, i.e., triathlon. Finally, the biggest challenge presents also idea to include the cyclist's sport trainer in concept human-in-the-loop, where he/she could actively contribute with his/her athlete during realization of training (i.e., online) in the sense of own expert knowledge.

Conflict of interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgements

Iztok Fister Jr. thanks the financial support from the Slovenian Research Agency (Research Core Funding No. P2-0057). Iztok Fister thanks the financial support from the Slovenian Research Agency (Research Core Funding No. P2-0042 - Digital twin).

References

H. Allen and A. Coggan. *Training and Racing with a Power Meter.* VeloPress, 2019. ISBN 1937715930.

T. Bäck. Evolution strategies: Basic introduction. In *Proceedings of the 12th Annual Conference Companion on Genetic and Evolutionary Computation*, GECCO '10, pages 2263–2288, New York, NY, USA, 2010. ACM. ISBN 978-1-4503-0073-5. doi: 10.1145/1830761.1830895.

E. Banister. Modeling elite athletic performance. *Physiological testing of elite athletes*, pages 403–424, 1991.

C. Darwin. *On the Origin of Species by Means of Natural Selection.* Murray, London, 1859. or the Preservation of Favored Races in the Struggle for Life.

E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–271, 1959. URL `http://gdzdoc.sub.uni-goettingen.de/sub/digbib/loader?did=D196313`.

A. E. Eiben and J. E. Smith. *Introduction to Evolutionary Computing.* Springer Publishing Company, Incorporated, 2nd edition, 2015. ISBN 3662448734, 9783662448731.

I. Fister, I. Fister, and D. Fister. *Computational Intelligence in Sports.* Springer, 2019.

D. B. Fogel. An overview of evolutionary programming. In L. D. Davis, K. De Jong, M. D. Vose, and L. D. Whitley, editors, *Evolutionary Algorithms*, pages 89–109. Springer New York, New York, NY, 1999. ISBN 978-1-4612-1542-4. doi: 10.1007/978-1-4612-1542-4_5.

F. Glover and M. Laguna. *Tabu Search*. Kluwer Academic Publishers, Norwell, MA, USA, 1997. ISBN 079239965X.

D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison Wesley, Reading, MA, 1989.

A. Holzinger. Interactive machine learning (iml). *Informatik-Spektrum*, 39(1): 64–68, 2016.

A. Holzinger, P. Kieseberg, E. Weippl, and A. M. Tjoa. Current Advances , Trends and Challenges of Machine Learning and Knowledge Extraction : From Machine Learning to Explainable AI. In *2018 International Cross Domain Conference for Machine Learning & Knowledge*, volume 2, pages 1–8, Hamburg, Germany, 2018. Springer International Publishing. ISBN 9783319997407. doi: 10.1007/978-3-319-99740-7.

A. Holzinger, M. Plass, M. Kickmeier-Rust, K. Holzinger, G. C. Crişan, C.-M. Pintea, and V. Palade. Interactive machine learning: experimental evidence for the human in the algorithmic loop. *Applied Intelligence*, 49(7):2401–2414, 2019.

J. R. Koza. Genetic programming as a means for programming computers by natural selection. *Statistics and Computing*, 4(2):87–112, Jun 1994. ISSN 1573-1375. doi: 10.1007/BF00175355.

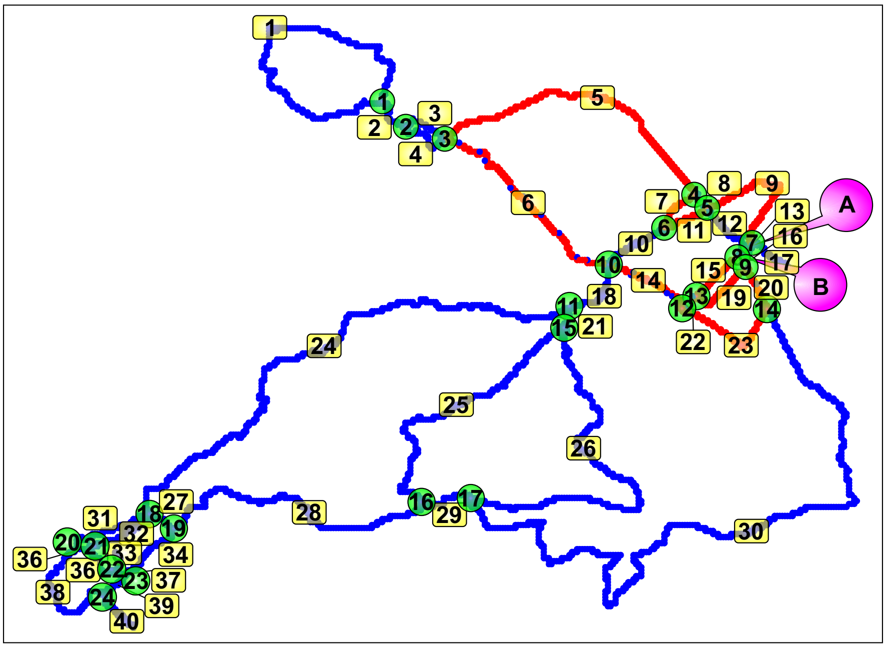
D. Saha and A. Mukherjee. Pervasive computing: a paradigm for the 21st century. *Computer*, 36(3):25–31, 2003.

R. Storn and K. Price. Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359, 1997.

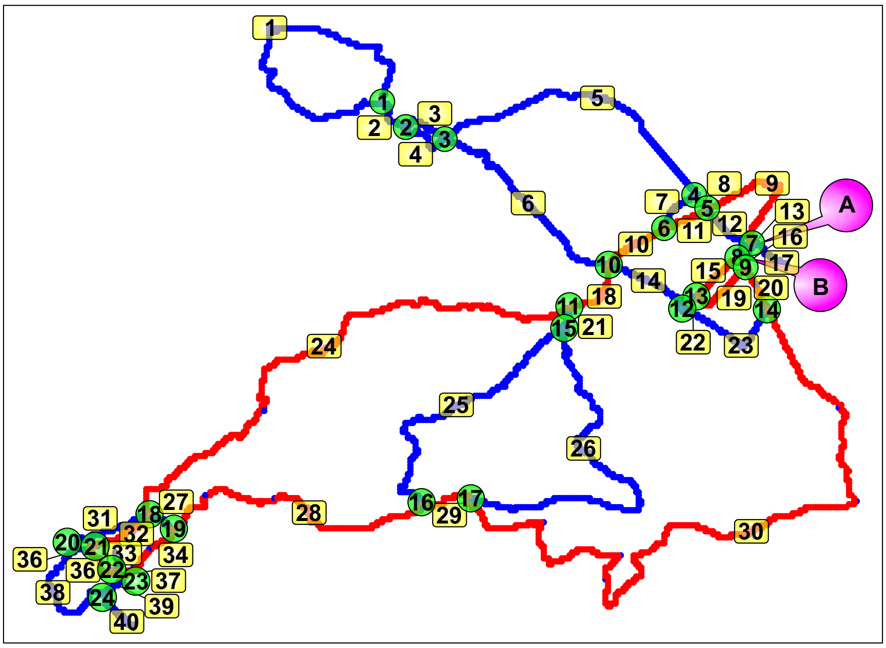
M. Weiser. The computer for the 21 st century. *Scientific american*, 265(3): 94–105, 1991.

J. Žemgulys, V. Raudonis, R. Maskeliūnas, and R. Damaševičius. Recognition of basketball referee signals from real-time videos. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–13, 2019.

Y. Zuo. Research and implementation of human-autonomous devices for sports training management decision making based on wavelet neural network. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–7, 2019.



a: Scenario A.



b: Scenario B.

Fig. 4: Visualization of generated sport training sessions.