



A minimalistic toolbox for extracting features from sport activity files

Iztok Fister Jr., Luka Lukač, Alen Rajšp, Iztok Fister, Luka
Pečnik, Dušan Fister
Presenter: Luka Lukač

Motivation

- Nowadays, Machine Learning (ML) is causing a revolution in many research areas, and sports training is no exception.
- In line with this, professional, as well as, amateur athletes are monitoring their sport activities/training using modern sport trackers.
- Until recently, not enough devotion was given to those indicators that are not visible directly, but can be obtained as the result of extensive data analysis, e.g. information extracted from topographic maps, weather conditions, and interval data.

Contributions of this study

- An overview of indicators that can be extracted from sport activity datasets as features is outlined.
- A new tool for detection of topographic features in sport activity datasets is proposed.
- A new method for extracting historical weather data is developed.
- A new method for detection of interval training sessions from datasets is developed.

Overview of indicators in sports activities

The problem of overall (integral) load indicators found in sports activity datasets, such as total duration, total distance, average heart rate, etc., is commonly associated with the following biases:

- details are not expressed sufficiently,
- only a general/integral outlook of the race/training is captured,
- the intensity indicator of the realized race/training may be fallacious and
- different stages/phases during the sport race/training, i.e. warming-up, endurance, intervals, etc., are not recognized directly.

Proposed sources for extracting features

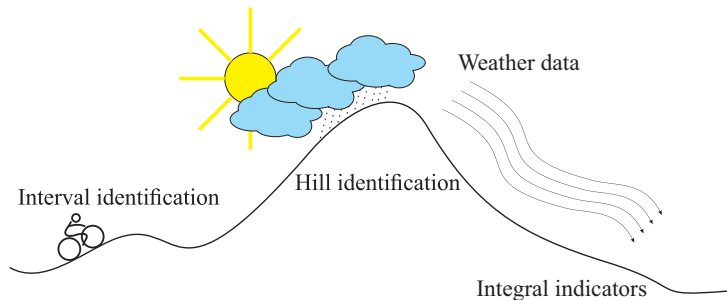


Figure: Sources for extracting features hidden in cycling sports activity datasets.

Topographic maps' extraction

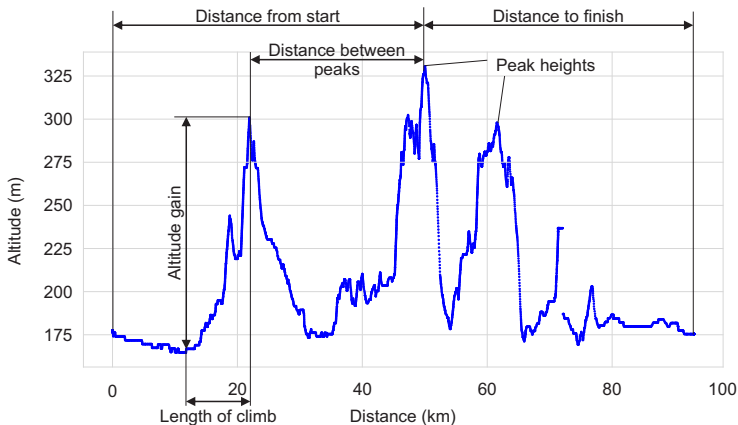


Figure: An example of topographic map and its basic labelled features.

Detection of intervals

```
1 function DETECT-INTERVALS (segments,  
   minimumTime)  
2   intervals  $\leftarrow$  {};  
3   avg  $\leftarrow$  getAverageHeartRate(segments);  
4   foreach segment  $\in$  segments do  
5     if segment.hearttrate > avg then  
6       | intervals  $\leftarrow$  intervals + {segment};  
7     end  
8   end  
9   for i  $\leftarrow$  1 to intervals.length do  
10    | if getAverageHearttrateBetween(intervals[i - 1],  
    | intervals[i]) < 10 then  
11    |   merge(intervals[i - 1], intervals[i]);  
12    | end  
13  end  
14  foreach interval  $\in$  intervals do  
15    | if interval.time < minimumTime then  
16    |   intervals  $\leftarrow$  intervals - {interval};  
17    | end  
18  end  
19  return intervals;
```

Figure: Detection of intervals by heart rate.

Detection of intervals

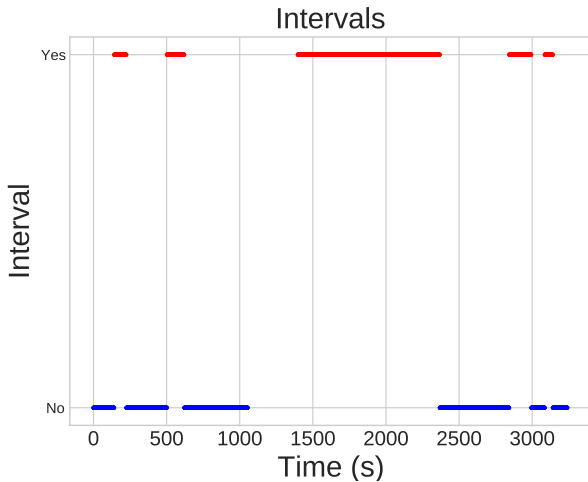


Figure: Detected intervals in sport activity.

Fork me on GitHub!

The screenshot shows the GitHub repository page for `firefly-cpp/sport-activities-features`. The repository is a fork of `firefly-cpp` with 1 branch and 4 tags. It has 67 commits and was last updated 4 days ago. The repository contains a minimalistic toolbox for extracting features from sport activity files written in Python. The file list includes `.github/templates`, `datasets`, `docs`, `examples`, `sport_activities_features`, `tests`, `.gitignore`, `.readthedocs.yml`, `CHANGELOG.md`, `CODE_OF_CONDUCT.md`, `LICENSE`, and `README.md`. The repository is licensed under the MIT License and has 3 releases, with the latest being 0.2.1. There are no packages published.

firefly-cpp / sport-activities-features

Unwatch 1 Star 5 Fork 3

<> Code Issues 3 Pull requests Actions Projects Wiki Security Insights Settings

main 1 branch 4 tags Go to file Add file Code

firefly-cpp Merge pull request #25 from luckyLukac/main 67 commits 4 days ago

.github/templates	add templates	2 months ago
datasets	Add test csv file	last month
docs	tests added, docs fixed	4 months ago
examples	Multiple data identified inside area	5 days ago
sport_activities_features	Added test for area_identification	4 days ago
tests	testing framework and docs established	4 months ago
.gitignore	potential code fixes	4 months ago
.readthedocs.yml	testing framework and docs established	4 months ago
CHANGELOG.md	update CHANGELOG	4 months ago
CODE_OF_CONDUCT.md	new changelog	4 months ago
LICENSE	Initial commit	7 months ago
README.md	Multiple data identified inside area	5 days ago

About

A minimalistic toolbox for extracting features from sport activity files written in Python

machine-learning data-mining sport-analytics ice-parser

Readme

MIT License

Releases 4

0.2.1 Latest 5 days ago

3 releases

Packages

No packages published

Figure: Main software repository.

Conclusion

- The domain of sports training has been connecting more and more with the computer science that offers tools for analysis data obtained by mobile devices worn during the training session with modern ML methods.
- Unfortunately, a lot of data obtained from mobile devices are hidden, and can be explained with sophisticated algorithms before entering into pure ML methods.
- The paper is focused on the development of extracting hidden features within sports activities.
- The following preprocessing methods were proposed, constituting the minimalistic toolbox: detailed analysis of topological maps, weather conditions, and identification of interval training sessions in cycling.
- In the future, the method should be included into Artificial Sport Trainer (AST).