# The Winning Solution to AAIA'15 Data Mining Competition: Tagging Firefighters Activities at a Fire Scene

**Jan Lasek**[1] Marek Gagolewski[2]

[1]Interdisciplinary PhD Studies Program
Institute of Computer Science, Polish Academy of Sciences
janek.lasek@gmail.com

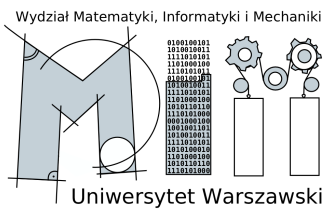
[2]Systems Research Institute, Polish Academy of Sciences
and
Faculty of Mathematics and Information Science, Warsaw University of Technology
gagolews@ibspan.waw.pl

FedCSIS
16 September 2015, Łódź, Poland

# Competition's task

The goal of the competition is to **recognize activities performed by firemen** at an emergency scene.

The competition was organised by Faculty of Mathematics, Informatics and Mechanics of University in Warsaw and Main School of Fire Service in Warsaw.





`https://knowledgepit.fedcsis.org/`

# Data

Two datasets are provided – training and test set – of the same size 20K instances:

- each row contains 42 time series of length 400 (i.e., $\approx$ 1.8 sec; sampled every 4-5 milisec.)
- additionally, a set of 42 summary statistics are provided for monitoring a fireman's vital functions
- each instance in the training set is tagged with a pair of labels describing main **posture** and **action** of a fireman
- datasets are of size approximately 2.4 GB

| Posture | Action | avg_ecg | . | acc_left_leg_x0 | . | gyr_torso_z399 |
|---|---|---|---|---|---|---|
| standing | no_action | -0.027 | . | -6.98 | . | 28.49 |
| stooping | manipulating | -0.042 | . | -9.40 | . | 63.84 |
| moving | running | -0.034 | . | -36.60 | . | -134.26 |
| crawling | searching | -0.040 | . | -2.99 | . | -7.21 |
| ... | ... | ... | . | ... | . | ... |

# Data - sensors

# Target attributes

There are two class attributes for dataset: main **posture** and specific **action** performed by a fireman.

| | crawl | crouch | move | stand | stoop |
|---|---|---|---|---|---|
| ladder_down | 0 | 0 | 465 | 0 | 0 |
| ladder_up | 0 | 0 | 476 | 0 | 0 |
| manipulating | 0 | 1764 | 331 | 2356 | 1898 |
| no_action | 0 | 87 | 0 | 491 | 0 |
| nozzle_usage | 0 | 492 | 0 | 443 | 0 |
| running | 0 | 0 | 4324 | 0 | 0 |
| searching | 459 | 0 | 0 | 0 | 0 |
| sig_hose_pb | 0 | 0 | 0 | 98 | 0 |
| sig_wat_first | 0 | 0 | 41 | 496 | 0 |
| sig_wat_main | 0 | 46 | 0 | 405 | 0 |
| sig_wat_stop | 0 | 0 | 0 | 277 | 0 |
| stairs_down | 0 | 0 | 644 | 0 | 0 |
| stairs_up | 0 | 0 | 1157 | 0 | 0 |
| striking | 0 | 0 | 0 | 1022 | 0 |
| throwing_hose | 0 | 0 | 0 | 234 | 930 |
| walking | 0 | 0 | 1064 | 0 | 0 |

# Sample plot of data

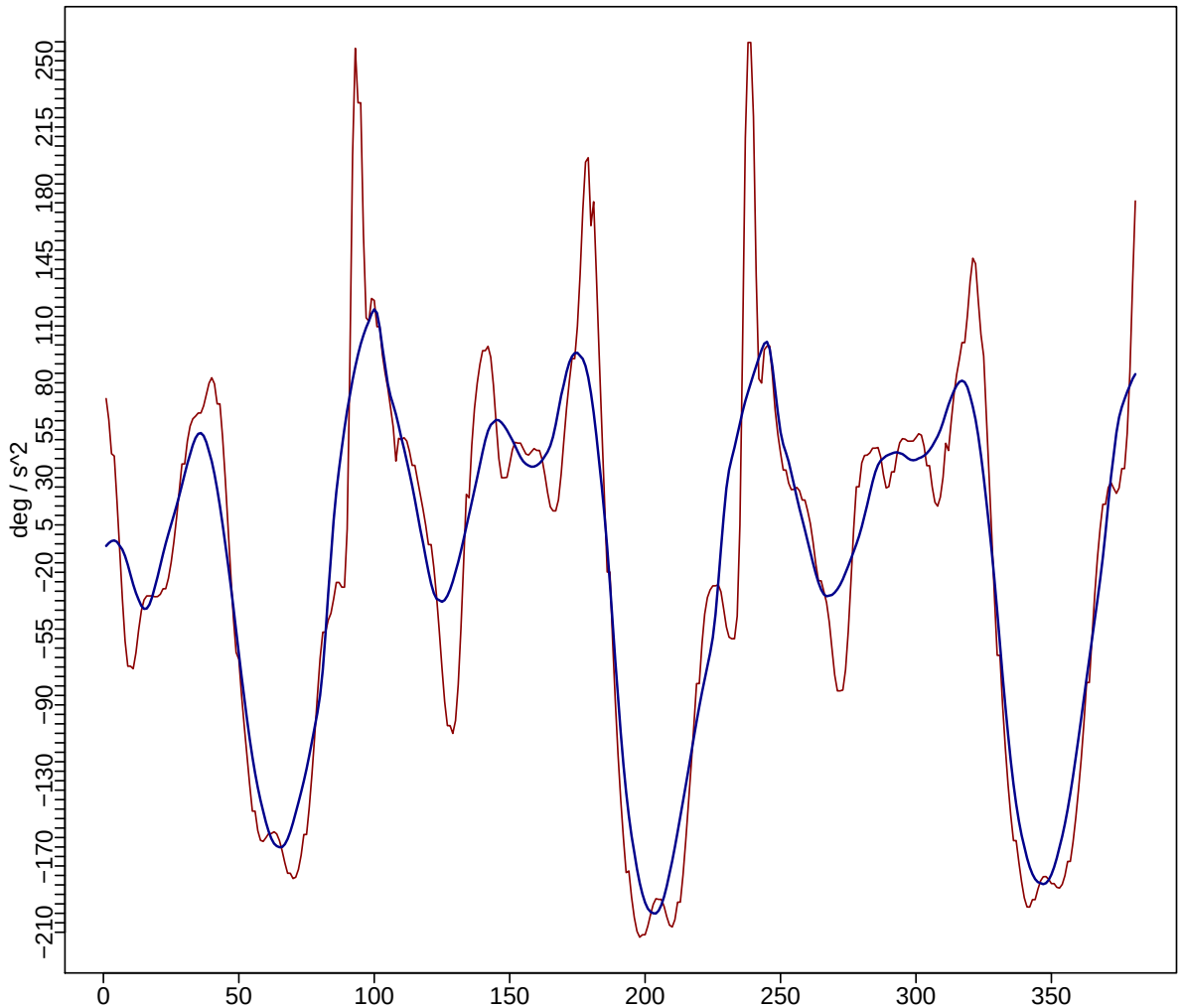



Figure 1 : Plot of raw series (red) and MA filtered (blue) for pair of labels (moving, running) of gyroscope recordings at left hand along $x$-axis.

# Competition evaluation metric – balanced accuracy

Classification accuracy (precision) for a given label:

$$\mathrm{acc}(l_i) = \frac{|\{j : I(x_j) = l_i \wedge p(x_j) = l_i\}|}{|\{j : I(x_j) = l_i\}|}$$

Balanced accuracy score for class $C$ with $L$ labels:

$$\mathrm{BAC}(C) = \frac{1}{L}\sum_{i=1}^{L} \mathrm{acc}(l_i)$$

**Evaluation metric** is the weighted average of balanced accuracy scores for *posture* and *action* classes:

$$\mathrm{EvaluationMetric} = \frac{1}{3}\mathrm{BAC}(\texttt{posture}) + \frac{2}{3}\mathrm{BAC}(\texttt{action}).$$

During the competition the solutions were evaluated against 10% of test data. Final evaluation was done using the other 90% of data.

# Our approach

There are a couple of issues to address:

- How to deal with a classifier in a problem with two (dependent) class attributes?
- What to do with the time series data?
- How to tailor a model to a given evaluation metric – balanced accuracy – given an imbalanced labels distribution within each class?
- How to evaluate model locally?
- Finally, which classifier to use?

# Two–class classification task

To tackle the problem of two (dependent) classes in our solution we decided to make predictions in a stepwise model: first predict **posture** and then **action** given posture.

$$\mathbb{P}(posture,\ action) = \mathbb{P}(action \mid posture) \cdot \mathbb{P}(posture)$$

Some other approaches were also tested:

- two independent classifiers
- one–vs–all setup
- labels concatenation

$$posture + action = posture\_action$$

# Feature extraction - main part of the solution

We constantly added new features to the training data.

- Basic summary stats: quantiles (**!**), mean, sd, skewness, kurthosis
- some transformations of features: amplitude, "derivative", max(abs(·)) to median ratio
- quantiles, sd of Fourier–transformed data and periodogram of series, first 5 Fourier coefficients
- correlations (**!**)
- We also did some experiments with peaks identification (i.e., the number of sub–chunks of a series where the observations exceeds mean by one or two sd)
- counts how many times a series crosses 0 and its mean

Overall, almost **5000** features were extracted. The dataset size compressed to 1.4 GB (for both training and test set).

Moreover, MA filter for raw series was applied to average them in a window of 20 observation (roughly over 0.1 sec).

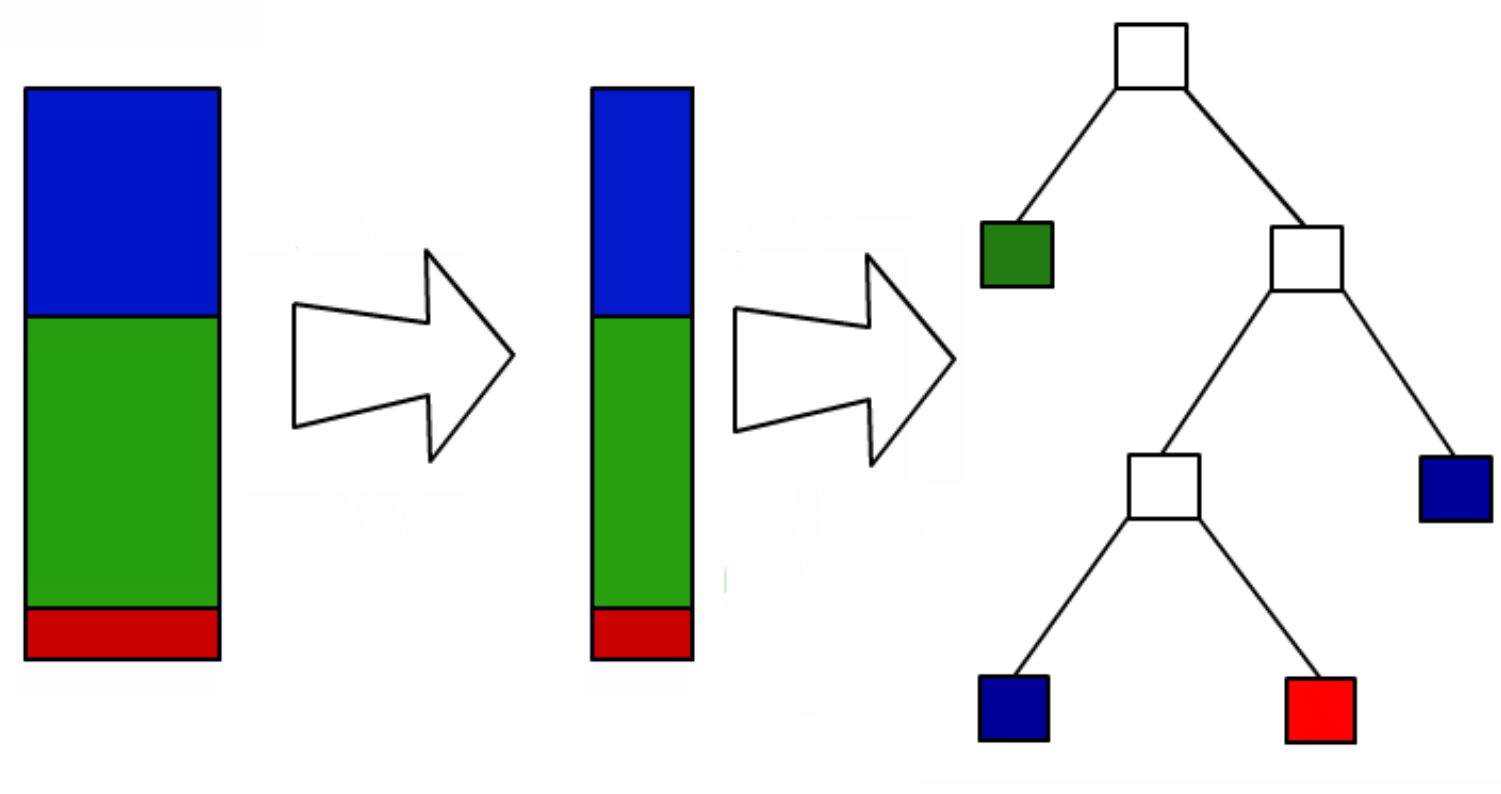
# Classifier

To classify both firefighter posture and action we used **Random Forest** model (Breiman, 2001).

- Random Forest is an **ensemble** of decision trees
- The model's prediction are derived by majority voting between weak learners
- This only works if the votes are (approximately) uncorrelated - this is achieved by subsampling both the instances as well as features in the training data
- The model needs to be tailored to evaluation metric proposed in the contest

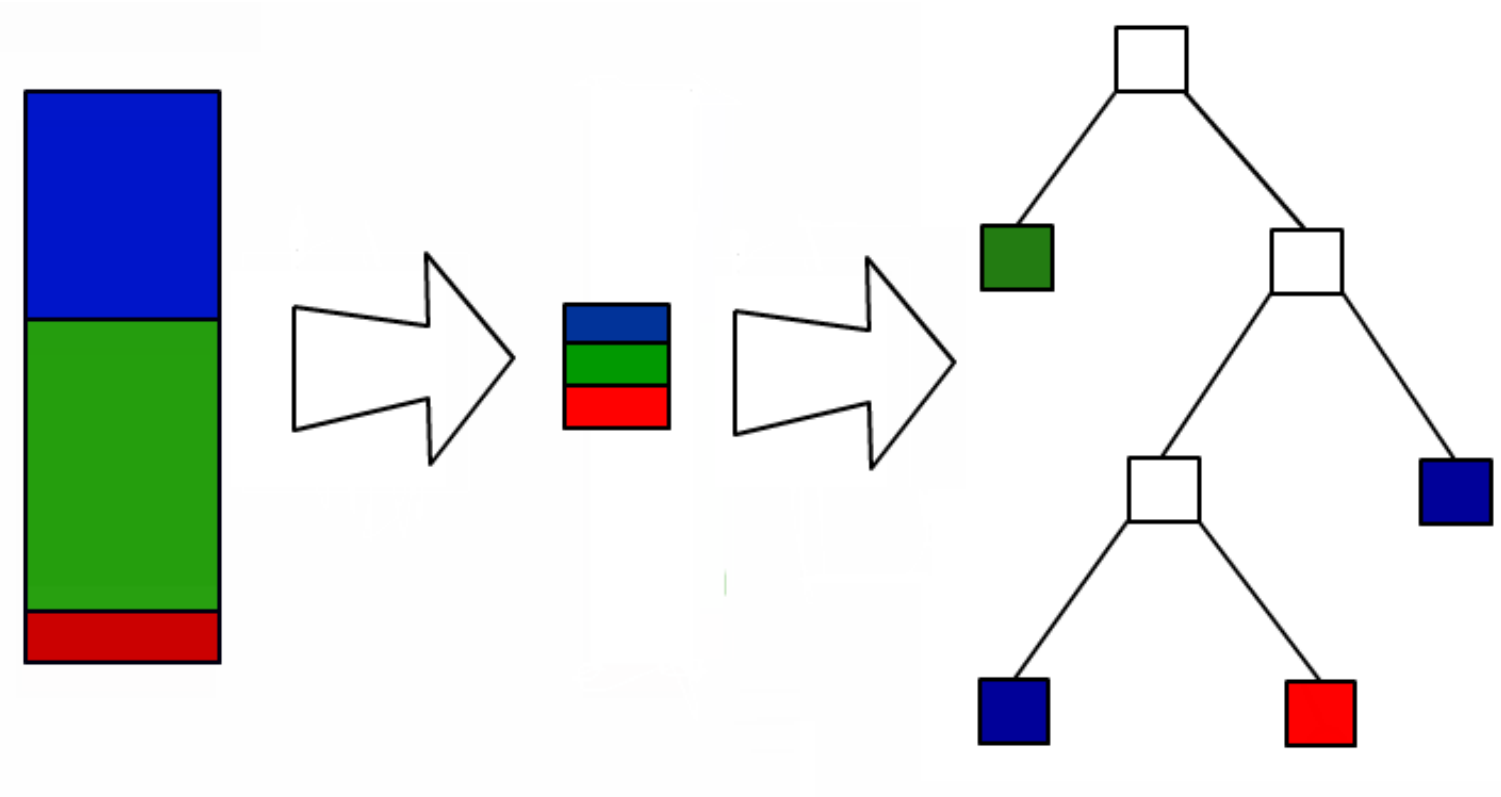
# Classifier – Random Forest: usual setup

700 $\times$



# Classifier – Random Forest: balanced

700 $\times$



# Model optimization

In order to compete you should be able to evaluate your ideas locally: this posed a real challenge.

- the activities in the training set and the test set are performed by different firemen (firemen's identifiers were not provided)
- local evaluation (by cross–validation/on hold–out test set/out–of–bag error) yielded scores as high as 99% accuracy
- my decisions were predominantly based on the results of preliminary evaluation scores on the leader–board

# Predictions of the model (paired)

| | crawl | crouch | move | stand | stoop |
|---|---|---|---|---|---|
| ladder_down | 0 | **1** | 459 | **209** | 0 |
| ladder_up | 0 | **2** | 452 | **118** | 0 |
| manipulating | 0 | 1576 | 12 | 1639 | 2438 |
| no_action | 0 | 71 | 0 | 467 | **31** |
| nozzle_usage | 0 | 454 | 0 | 1060 | 0 |
| running | 0 | **13** | 3974 | 0 | **2** |
| searching | 513 | **42** | 0 | 0 | 0 |
| signal_hose_pullback | 0 | 0 | 0 | 96 | 0 |
| signal_water_first | 0 | **3** | 10 | 580 | 0 |
| signal_water_main | 0 | 55 | 0 | 174 | 0 |
| stairs_down | 0 | 0 | 533 | 0 | 0 |
| stairs_up | 0 | 0 | 1442 | 0 | 0 |
| striking | 0 | **13** | **7** | 1026 | **49** |
| throwing_hose | 0 | 0 | 0 | 196 | 982 |
| walking | 0 | **2** | 1251 | **46** | **2** |

# Unsolved puzzles a.k.a. future work

The main problems with the proposed solution are (still)

- some contradictory pairs of labels – this problem was only limited by the described chaining method
- there are huge discrepancies between local evaluation scores and leader–board results: 99% v.s. 84% of balanced accuracy scores
- the model still does not generalize between different people – perhaps more generic set of features could help

# Conclusions

The competition was a very exciting event!

- Thanks to the Organizers and all other Participants
- The source code for the solution is available at GitHub
  `https://github.com/janekl/AAIA15_Data_Mining_Contest`
- If you are interested in this kind of contests visit one of the competition hosting sites
  1. `https://knowledgepit.fedcsis.org/`
  2. `https://www.kaggle.com/`
  3. `http://www.drivendata.org/`
  4. ...

# Thank you!