Metamodels for complex structured objects classification

ROY team: Ilya Zharikov,
Roman Isachenko,
Artem Bochkarev

Skolkovo Institute of Science and Technology
Machine Learning course

March 20, 2017

Project goal

Aim

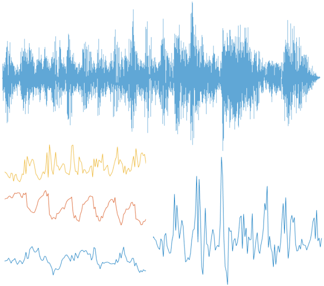
Classification model for complex-structured objects.

Problem

Initial object has no appropriate feature description.

Applications:

- image processing;
- signal classification;
- topic modelling;
- *time series analysis*.



Related work

1. Wang W. et al. Human activity recognition using smart phone embedded sensors // *International Joint Conference on Neural Networks*. 2014.
2. Kwapisz, Jennifer R., Gary M. Weiss, and Samuel A. Moore. Activity recognition using cell phone accelerometers. // *ACM SigKDD Explorations Newsletter*. 12(2): 74-82. 2011.
3. Kuznetsov M. P., and Ivkin N. P. Time series classification algorithm using combined feature description. // *Journal of Machine Learning and Data Analysis*. 2015.

Problem Statement

Let: $s \in \mathcal{S}$ — complex structured object;
$y \in Y$ - class label.

Task

Suppose to be given the set of labeled data $\mathfrak{D} = \{(s_i, y_i)\}_{i=1}^{m}$.
Our goal is to determine function f^* such that

$$f^* = \arg\min_f L(f, \mathfrak{D}),$$

where $L(\cdot, \cdot)$ is an error function and $f : \mathcal{S} \to Y$.

Approach

Suppose $f = g \circ h$, where

1. $h(s) : \mathcal{S} \to H \subset \mathbb{R}^n$ is map from $\mathcal{S}$ into feature space H;
2. $g(\boldsymbol{h}, \boldsymbol{\theta}) : H \to Y$ is parametric map (classification model).

Optimal parameters

$h(s)$

Choice of feature map $h(s)$ by

- prior (expert) knowledge;
- minimizing error functional.

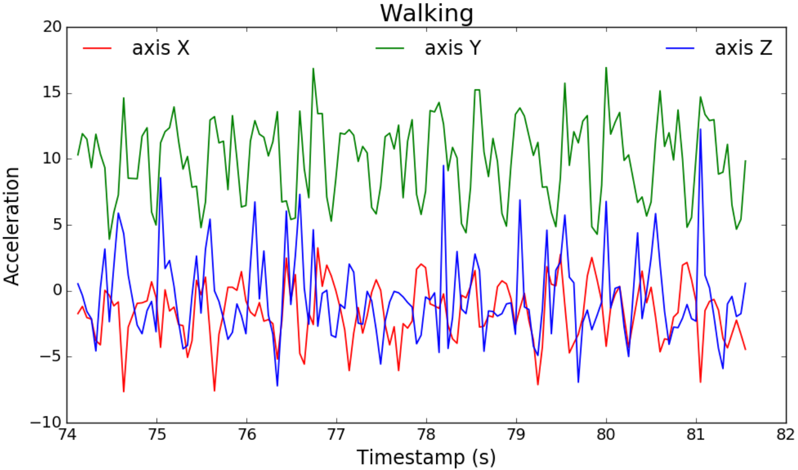
$g(\boldsymbol{h}, \theta)$

Classification for $\{(\boldsymbol{h}_i, y_i)\}_{i=1}^m$, $\boldsymbol{h}_i = h(s_i)$:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} L(g, \boldsymbol{\theta}, \mathfrak{D}).$$

E.g.: $g(\mathbf{h}, \boldsymbol{\theta})$ - classification model;
$\boldsymbol{\theta}$ - model parameters;
$L(g, \boldsymbol{\theta}, \mathfrak{D})$ - classification error function.

Time series example



Expert functions

Prior knowledge about the objects allows to choose the features.

Feature description

$\boldsymbol{h}_i = h(s_i) \in \mathbb{R}^{40}$ — different statistics:

- $\boxed{3}$ average accelerations;
- $\boxed{3}$ standard deviations;
- $\boxed{3}$ mean absolute deviations;
- $\boxed{1}$ average acceleration;
- $\boxed{30}$ values of histogram with 10 equal parts.

Autoregressive model

Data generation hypothesis

Let's assume that time series $s = (x_1, \dots, x_T)$ is generated by the following autoregressive model:

$$\widehat{x}_t = w_0 + \sum_{j=1}^{n} w_j x_{t-j}.$$

Feature description

$$h(s) = \mathbf{w}^* = \arg \min_{\mathbf{w} \in \mathbb{R}^{n+1}} \sum_{j=n+1}^{T} \|x_j - \hat{x}_j\|^2 .$$

Singular Spectrum Analysis (SSA)

Let's consider a **trajectory matrix** for time series $s = (x_1, \ldots, x_T)$:

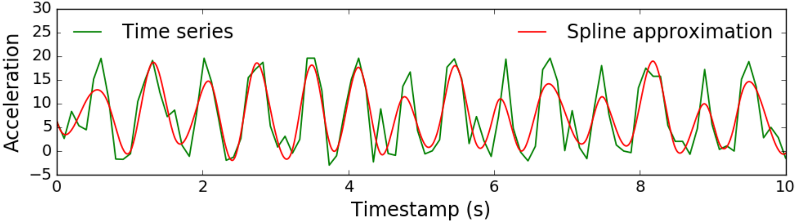
$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ x_2 & x_3 & \cdots & x_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T-n+1} & x_{T-n+2} & \cdots & x_T \end{pmatrix}$$

Feature description

$$h(s) = (\lambda_1, \ldots, \lambda_n),$$

where $\{\lambda_i\}_{i=1}^n$ are eigenvalues of the matrix $\mathbf{X}^\top \mathbf{X}$, obtained by Singular Value Decomposition: $\mathbf{X}^\top \mathbf{X} = \mathbf{V} \cdot \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \cdot \mathbf{V}^\top$.

Splines



Spline is defined by $\{\xi_\ell\}_{\ell=1}^{L}$ — the set of knots and
$\{\mathbf{w}_\ell\}_{\ell=1}^{L}$ — parameters of the models are built on the interval $[\xi_\ell; \xi_{\ell+1}]$.

Feature description

$$\boldsymbol{h} = h(s) = (\xi_1, \dots, \xi_L, \mathbf{w}_1, \dots, \mathbf{w}_L).$$

Data

WISDM[1]

Activity	# objects	
Standing	229	5.3%
Walking	1917	44.4%
Upstairs	466	10.8%
Sitting	277	6.4%
Jogging	1075	24.9%
Downstairs	357	8.3%
Total	4321	

USC-HAD[2]

Activity	# objects	
Walking-downstairs	951	7%
Walking-upstairs	1018	7.4%
Walking-forward	1874	13.8%
Walking-right	1305	9.6%
Walking-left	1280	9.4%
Elevator-up	764	5.6%
Elevator-down	763	5.6%
Standing	1167	8.6%
Sitting	1294	9.5%
Sleeping	1860	13.7%
Jumping	495	3.6%
Running	849	6.2%
Total	13620	

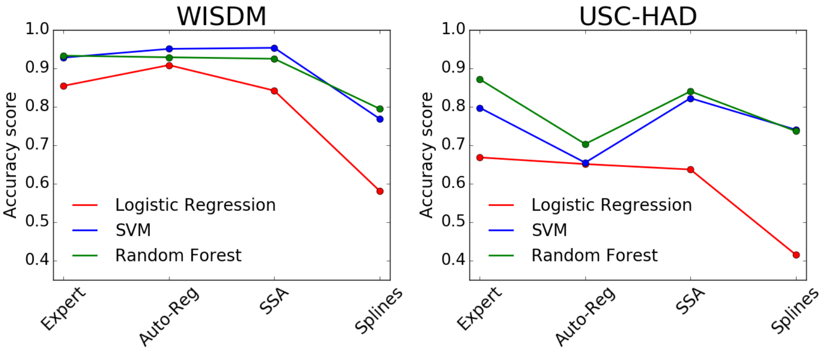
[1] http://www.cis.fordham.edu/wisdm/

[2] http://sipi.usc.edu/HAD/

Experiment

Datasets:	WISDM, USC-HAD.
Feature extraction methods:	expert functions;
	autoregressive models;
	singular spectral analysis;
	splines.
Classification models:	logistic regression;
	support vector machine;
	random forest.
Tuning parameters:	cross-validation.
Quality measure:	accuracy score.

Results



Feature union approach. Best models:

SVM: **0.983**

Random Forest: **0.882**

Conclusion

Done

- different approaches to classification of complex structured objects were studied;
- the experiments on the smart phone accelerometer data were conducted;
- the results of experiments on human activities datasets outperform many previous methods.

Future work

- new approaches to feature extraction (e.g. modified splines);
- implementing of structured learning methods.