Parameter Grid Search for Random Forest Classifier on Fall Detection Data

Vladislav Belov

Project Overview

This small project takes a look at the Fall Detection Data from China. The data contains 6 variables:

- **TIME** - the total time of patient's monitoring;
- **SL** - the level of sugar in the organism;
- **EEG** - electroencephalography monitoring rate;
- **BP** - blood pressure;
- **HR** - heart beat rate;
- **CIRCULATION** - blood circulation.

The response variable **ACTIVITY** classifies the type of activity patients were doing during the period of taking measurements of variables presented above:

ACTIVITY	Type of the Activity
0	Standing
1	Walking
2	Sitting
3	Falling
4	Cramps
5	Running

Here is a quick look at the head of the dataset we will be dealing with (it contains 16382 rows in total):

Table 2: Fall Detection Data from China

ACTIVITY	TIME	SL	EEG	BP	HR	CIRCLUATION
3	4722.92	4019.64	-1600.00	13	79	317
2	4059.12	2191.03	-1146.08	20	54	165
2	4773.56	2787.99	-1263.38	46	67	224
4	8271.27	9545.98	-2848.93	26	138	554
4	7102.16	14148.80	-2381.15	85	120	809

In this project we will use the Random Forest classifier to train the machine to classify activities according to basic inner body measurements. The classifier will be trained on a pre-prepaired dataset, as it is going to be cleaned and, moreover, all of the explanatory variables will be normalized.

There will be multiple training sessions depending on the initial parameters of the Random Forest model. This set of parameters will be called the grid of parameters. After each training session the prediction accuracy will be evaluated given the fixed test set. Aforementioned procedures will provide us with a new dataset with the parameter grid as explanatory variables and model accuracy as the response variable. We will analyze significance of parameters and their interactions and look for the most efficient model given their range which were used in the grid.

The grid of parameters we will be focusing on are the following ones:

Parameter Name	Description	Considered Values
`bootstrap`	whether bootstrap samples are used when building trees	
`max_depth`	the maximum depth of the tree	
`max_features`	the number of features to consider when looking for the best split	
`min_samples_split`	the minimum number of samples required to split an internal node	
`criterion`	the function to measure the quality of a split	
`n_estimators`	the number of trees in the forest	

Results of the Random Forest Classifier Training

TBD