AlvsPK Challenge: FACT SHEET

Title:
Gradient Boosted Trees with embedded FS and Particle Filtering optimization.

Name, address, email:
ASML team, INTEL Corporation,
vladimir.martyanov @intel.com

Acronym of your best entry:
Intel_GBT_PF_3 (Intel final 4)
Intel_GBT_PF_4 (Intel final 5)

Reference:

A. Borisov, V. Eruhimov, and E. Tuv. Dynamic soft feature selection for tree-based ensembles. In I. Guyon, S. Gunn, M. Nikravesh, and L. Zadeh, editors, *Feature Extraction, Foundations and Applications.* Springer, New York, 2006.

V.Eruhimov, V.Martyanov, E.Tuv, "Signal Classi¯cation Through Massive Feature Extraction From Warped Signals", to appear in Proceedings of PKDD'2007

Method:

No preprocessing was done. Five Gradient Boosting Trees (GBT) models were built on cross-validation subsets of input data and the final result was predicted by the majority vote. Features with importance (calculated by a GBT model) lower than a predefined threshold were removed from the model before training. Parameters of GBT (the number of trees, shrinkage, importance adjust rate for dynamic feature selection, importance threshold, stratified sampling parameters and probability threshold for final classification) were chosen by minimizing cross-validation error. Optimization was done with a Particle Filtering algorithm combined with simulated annealing.

Results:

Table 1: Our methods best results

Dataset	Entry name	Entry ID	Test BER	Test AUC	Score	Track
ADA	intel-GBT-PF-3 (Intel final 4)	1056	0.1745	0.8258	0.0184	Agnos
GINA	intel-GBT-PF-3 (Intel final 4)	1056	0.038	0.9618	0.2479	Agnos
HIVA	Intel-GBT-PF-1 (Intel final 4)	1054	0.2908	0.7067	0.1245	Agnos
NOVA	intel-GBT-PF-3 (Intel final 4)	1056	0.0727	0.9279	0.5	Agnos
SYLVA	GBT + PF	1033	0.0079	0.9921	0.2111	Agnos
Overall	GBT + PF	1033	0.1193	0.8805	0.2432	Agnos

Table 2: Winning entries of the AlvsPK challenge

Best results agnostic learning track						
Dataset	**Entrant name**	**Entry name**	**Entry ID**	**Test BER**	**Test AUC**	**Score**
ADA	Roman Lutz	LogitBoost with trees	13, 18	0.166	0.9168	0.002
GINA	Roman Lutz	LogitBoost/Doubleboost	892, 893	0.0339	0.9668	0.2308
HIVA	Vojtech Franc	RBF SVM	734, 933, 934	0.2827	0.7707	0.0763
NOVA	Mehreen Saeed	Submit E final	1038	0.0456	0.9552	0.0385
SYLVA	Roman Lutz	LogitBoost with trees	892	0.0062	0.9938	0.0302
Overall	Roman Lutz	LogitBoost with trees	892	0.1117	0.8892	0.1431
Best results prior knowledge track						
Dataset	**Entrant name**	**Entry name**	**Entry ID**	**Test BER**	**Test AUC**	**Score**
ADA	Marc Boulle	Data Grid	920, 921, 1047	0.1756	0.8464	0.0245
GINA	Vladimir Nikulin	vn2	1023	0.0226	0.9777	0.0385
HIVA	Chloe Azencott	SVM	992	0.2693	0.7643	0.008
NOVA	Jorge Sueiras	Boost mix	915	0.0659	0.9712	0.3974
SYLVA	Roman Lutz	Doubleboost	893	0.0043	0.9957	0.005
Overall	Vladimir Nikulin	vn3	1024	0.1095	0.8949	0.095967

- quantitative advantages: GBT with dynamic feature selection is very fast to learn in extremely high dimensional spaces, handles both numeric and categorical variables, robust against outliers
- qualitative advantages: we get an estimate of posterior probability and an estimate of feature importance as byproducts of the model. The novelty is in stochastic optimization of GBT parameters and feature filtering.

Keywords:

- Preprocessing or feature construction: ----
- Feature selection approach: embedded feature selection, importance thresholding
- Feature selection engine: Gradient boosting trees.
- Feature selection search: linear search in sorted array of features
- Feature selection criterion: 5-fold cross-validation.
- Classifier: Gradient boosting trees.
- Hyper-parameter selection: Particle filtering with simulated annealing
- Other: ------.