AlvsPK Challenge: FACT SHEET

Title: **Ensemble of ensemble of tree and neural network**

Louis Duclos-Gosselin, 205 Gosselin Street, St-Agapit, Quebec, Canada, g0s 1z0 louis.gosselin@hotmail.com

Neural Network

Reference: Ensemble of ensemble of tree and neural network presented at IJCNN; More details will be fill on demand

Method:

The 2007 Agnostic Learning v.s. Prior Knowledge Challenge permits me to illustrate one of my personal algorithms on different datasets. With this kind of algorithms I did a great score on PAKDD 2007 (30th). I propose to use a special case of mixed ensemble of boosting tree and neural network. In brief, a single tree is used to adjust the setting of my boosting tree and neural network. First, I propose to use a combination of Gini, Entropy and Misclassification algorithms to construct a single tree. This single tree, in conjunction to Genetic algorithm permit to set parameters for ensemble method (Category weights, Misclassification costs, Variable weights, Max. categories for continuous predictors, Minimum size node to split, Use surrogate splitters for missing, Tree pruning and validation method, Tree pruning criterion). Second, Genetic algorithms, Wrapper techniques, Link analysis, SOM, clustering technique and filter techniques allow me to chose the best predictors for ensemble methods. Third, a Special case of Gradient-boosting is constructed with the single tree's setting. In addition, annealing techniques are used to choose the best neural network architecture (S.V.M., R.B.F., Bayes networks, Cascade correlation, Projection pursuit). The parameters of those neural networks are set with Genetic algorithm (Learning algorithm and parameter, Number of neuron and hidden layer, activation function). Finally, the ensemble method is constructed. This is the important part of the process. In function of the goal of managers (classification goal or ranking goal) a minimisation criteria is choose and various techniques are used to aggregate the ensemble of tree and the neural network. In conclusion, there are many facts which are interesting with this kind of algorithm: it doesn't over fitting because k-folds-validation and genetic algorithms are used during all the process to keep the over learning as low as possible and this process is particular powerful on small category problem.

- Preprocessing or feature construction: Optimal binning, Standardize, Maximize normality
- Feature selection approach: Filter, Wrapper, Link analysis, SOM, Clustering technique
- Feature selection engine: Relief, Information theory, Mutual information, X2, Single tree
- Feature selection search: Annealing, Genetic algorithm
- Feature selection criterion: K-fold cross-validation
- Classifier: Neural networks, Tree classifier, Ensemble of tree, S.V.M., R.B.F., Bayes networks, Cascade correlation, Projection pursuit
- Hyper-parameter selection: grid-search, pattern search, cross-validation, K-fold, Genetic algorithm.

Results: The strength of my method is this kind of algorithm doesn't over fitting because k-folds-validation and genetic algorithms are used during all the process to keep the over learning as low as possible and this process is particular powerful on small category problem.

The model performs well on ADA:

Table 1: Our methods best results

Dataset	Entry name	Entry ID	Test BER	Test AUC	Score	Track
ADA	Neural Network13	969	0.1776	0.8216	0.0429	Prior
SYLVA	Neural Network3	974	0.0113	0.9887	0.3769	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 and qualitative advantages :The error is minimized in keeping over fitting as low as possible; The multiple use of genetic techniques make it really long to compute (However, it gives you the best model by finding the *real global minimum*); The process is powerful on small category problem and can handle different managerial goal (which is really important in the business world); The final model is simple to present to marketer and can be partially explain with the single tree constructed at the beginning; Unlike other proposed model in the literature this one can be view as a whole process: all the possibility are explored; all the architecture are visited; all the parameters are tested); In my opinion, we should always test all possibility, all category of model and all new ideas available in the literature to provide The Best Solution to the manager. It involves that as applied mathematician it's primordial to always keep us informed about The Best Technique. Keep in mind that my process doesn't take more time than other, it's a little more time consuming on the modeling phase, which represent not the big part of the resolution process.

Keywords:

- Preprocessing or feature construction: Optimal binning, Standardize, Maximize normality
- Feature selection approach: Filter, Wrapper, Link analysis, SOM, Clustering technique
- Feature selection engine: Relief, Information theory, Mutual information, X2, Single tree
- Feature selection search: Annealing, Genetic algorithm
- Feature selection criterion: K-fold cross-validation
- Classifier: Neural networks, Tree classifier, Ensemble of tree, S.V.M., R.B.F., Bayes networks, Cascade correlation, Projection pursuit
- Hyper-parameter selection: Grid-Search, Pattern search, Cross-validation, K-fold, Genetic algorithm.