Progress Report

Tomáš Frýda

Task

Project name: Scalable predictive models and hierarchical ensembles
Supervisor: Ing. Pavel Kordík, Ph.D.
Sponsor: FIT ČVUT

Weak learners such as linear, polynomial, sigmoid or gaussian regression are fast and efficient when well trained. Training needs to be scalable and ensembles of weak learners is a good choice. We implemented many week learners and ensemble methods into the open source H2O.ai framework. Now it is time to find a way, how to build scalable ensembles that would complement existing ensembling approaches such as Super Learner.

Goals

- Build a benchmarking environment in H2O for both regression and classification problems. Consider the time dimension (anytime learning).
- Tune existing weak learners and experiment with heuristics how to build scalable anytime ensemble.
- Compare with existing ensembling strategies such as SuperLearner.
- Write draft of a paper.

Progress so far

So far I've been able to implement all required features in the benchmarking environment for benchmarking anytime learning.

The benchmarking environment consists of python script. This python script is responsible for running H2O models and H2O ensemble (which is the Super Learner ensemble written in R). It can be configured via python or yaml. The benchmarking environment can be found at https://gitlab.fit.cvut.cz/frydatom/h2o-benchmarker.

Now I am in the process of merging FAKEGAME with H2O and implementing some weak learners to FAKEGAME. My efforts can be followed in https://gitlab.fit.cvut.cz/frydatom/h2o-fakegame and https://gitlab.fit.cvut.cz/frydatom/fakegame repositories mostly in "fakegame-stable" branch.

Future plans

After I conclude the merge of FAKEGAME into H2O and finish implementing weak learners I'll benchmark the models, tune the ensembling process and write a paper draft.

Other Remarks

All related code can be found in https://gitlab.fit.cvut.cz/frydatom/frydatom-vylet-2016 and linked submodules.