

Ensembles for predicting structured outputs

SUMMARIZED CONCLUSIONS FOR REGRESSION AND HMLC

Dragi Kocev, Celine Vens, Jan Struyf, Sašo Džeroski

July 19, 2010

1 Multi-target Regression

1.1 Statistical tests

We consider RRMSE as evaluation measure. Similar conclusions can be drawn for both Correlation Coefficient and RMSE.

- MTBag, MTRF, STBag, STRF saturate @50 iterations (after 50 iterations there is no more statistically significant difference!)
- @50 iterations there is no stat. sign. difference between the algorithms. They are ordered as follows: MRF50, STRF50, MTBag50, STBag50.
- @100 iterations there is no stat. sign. difference between the algorithms. They are ordered as follows: STRF100 \approx MTRF100, MTBag50, STBag50.
- The multi target ensembles are more dominant when smaller ensembles are used!
- The multi target ensembles are much faster to learn and in total have smaller model size (stat.sign.).

1.2 Saturation Curves

- The smaller datasets (by number of examples ($<^? 1000$) and number of features $<^? 10$) exhibit more chaotic behaviour - the learning curve is not smooth.
- On the big datasets (number of examples ($>^? 10000$), ex. Forestry_Kras, Vegetation Clustering, Vegetation Condition), the random forests (both ST and MT) outperform bagging (ST and MT).
- On the medium datasets (number of examples between 1000 and 10000), the bagging methods (MT and ST) perform better than the random forests.
- On the overall curve (averaged accross all datasets), we can note that the MT methods perform better across the complete learning curve (except for 100 iterations).

2 Multi-Target Classification

These are present in a separate documents!!!

2.1 Statistical tests

2.2 Saturation curves

3 Hierarchical Multi-Label Classification

We consider $AU\overline{PRC}$ as evaluation measure.

3.1 Statistical tests

- HMLCBag, HMLCRF, and HSLCRF saturate @50 iterations and HSLCBag @25.
- @50 iterations there is no stat. sign. difference between the algorithms. They are ordered as follows: HSLCRF50, HMLCBag50, HMLCRF50, HSLCBag50
- @100 iterations there is no stat. sign. difference between the algorithms. They are ordered as follows: HMLCBag100, HSLCRF100, HMLCRF100, HSLCBag100
- The multi label ensembles are much faster to learn (stat.sign.).

3.2 Saturation curves

- Functional genomics: when the target classes are organized in DAG, Multi-label methods are better than single label, while HMLC RF is the best performing method.
- Image Classification: Multi-label methods are better than single label, while HMLC RF is the best performing method.
- Text Classification: Single-label approaches show better performance. Note the extremely large number of features for Reuters and WIPO datasets.
- The overall curve shows that the multi label approaches perform better than the single label ones: HMLCRF is best performing.

3.3 Dataset stats

These are desired for more in-depth discussion of the results.

- Size of the hierarchy
- Average number of labels per example
- Average number of labels that are leafs per example