



October 18, 2017

F O R S C H U N G S P R A X I S
for
Jianxiang Feng
Student ID 03679926, Degree EI

Hierarchical classification with Random Forests

Problem description:

Hierarchical structuring of classes of data items and of the corresponding classification problem has been investigated in the past mostly for the purpose of an efficient decomposition of multi-class problems into a cascade of binary classification problems [1][2]. On the other hand, Decision Trees and their ensembles, the Random Forests[3], are also hierarchical approaches to classification but capable of efficient multi-classification by treating discrimination of all the classes simultaneously.

In this research project, we investigate the gain in accuracy that can be made by hierarchical reorganization of multi-class problems, using Random Forests at each level of classification for the compound classes. An adequate class hierarchy has to be constructed automatically. We here consider more directly the complexity of the Random Forest trained to discriminate the classes (quantified as number of leaf nodes) and the errors made in class prediction on validation datasets. These will be the measures to guide the merging into or partitioning of the compound classes, hence the generation of the classification hierarchy. The classification hierarchy may be built bottom-up by recursively merging groups of classes, or top-down by recursively splitting compound classes. Several variants of constructing the hierarchy will be compared on synthetic datasets of varying complexity and on at least one real-world dataset.

Work schedule:

- Literature review on hierarchical classification and random forest
- build hierarchical structure with random forest
- test on synthetic dataset and modify algorithm
- test on at least one real-world dataset

Bibliography:

- [1] Marcin Marszałek and Cordelia Schmid. Constructing category hierarchies for visual recognition. In *European Conference on Computer Vision*, pages 479–491. Springer, 2008.
- [2] Babak Shahbaba, Radford M Neal, et al. Improving classification when a class hierarchy is available using a hierarchy-based prior. *Bayesian Analysis*, 2(1):221–237, 2007.
- [3] Danhang Tang, Tsz-Ho Yu, and Tae-Kyun Kim. Real-time articulated hand pose estimation using semi-supervised transductive regression forests. In *Proceedings of the IEEE international conference on computer vision*, pages 3224–3231, 2013.

Supervisor: Dr. Ulrich Hillenbrand
Start: 19.10.2017
Delivery: 31.12.2017

(D. Lee)
Univ.-Professor