

ML: Reliable Calibrated Probability Estimation in Classification

Written by Marinka

Monday, 16 January 2012 23:12 - Last Updated Sunday, 25 August 2013 21:31

I followed a Machine Learning course this semester at the university (along with the [one](http://www.ml-class.org/course/auth/welcome), offered by Stanford Uni). I have been working on a seminar studying **Reliable probability estimation in classification using calibration**. The final report is in the form of the scientific article, which is attached below.

Here is the abstract:

Estimating reliable class membership probabilities is of vital importance for many applications in data mining in which classification results are combined with other sources of information to produce decisions. Other sources include domain knowledge, outputs of other classifiers or example-dependent misclassification costs. We revisit the problem of classification calibration motivated by the issues of the isotonic regression and binning calibration. These methods can behave badly on small or noisy calibration sets, producing inappropriate intervals or boundary generalization. We propose an improvement of the calibration with isotonic regression and binning method by using bootstrapping technique, named boot-isotonic regression and boot-binning, respectively. Confidence intervals obtained by repeatedly calibrating the set sampled with replacement from the original training set are used for merging unreliable or too narrow calibration intervals. This method has been experimentally evaluated with respect to two calibration measures, several classification methods and several problem domains. The results show that the new method outperforms the basic isotonic regression and binning methods in most configurations.

Short presentation about work and final report:

- [Presentation \(English\) \(.pdf\)](../ISRM/ML_seminar.pdf)
- [Article \(English\) \(.pdf\)](../ISRM/zitnik12a_final.pdf)