Open Problems/Emerging Directions in Conformal Prediction

Vladimir Vovk

Department of Computer Science
Royal Holloway, University of London

Workshop on Conformal Prediction for Reliable Machine Learning
IIT Hyderabad, 18 December 2015

Outline

The Burnaev–Wasserman programme

- Applicable to both conformal and Venn prediction (although first stated for conformal prediction).
- Suppose we know the true probability distribution generating the data.
- In this case we know the optimal algorithm (the Bayes algorithm).
- Burnaev: how much do we lose when we use the conformal/Venn machine of top of the Bayes algorithm. (Hopefully, not much.)
- Wasserman's complement: if we do lose much, the Bayes algorithm is in trouble (very fragile).

Implementation for conformal prediction

- Implemented only in one case: ridge regression (Burnaev and Vovk, COLT 2014).
- Under way: kernel ridge regression.
- To do: other regression algorithms.
- To do: classification (e.g., LDA).

Implementation for Venn prediction (not yet)

- One possible precise open problem: based on Naive Bayes. Labels: $y_i \in \{-1, 1\}$ with equal probabilities. Independent attributes $x_i^k := y_i + N(0, \sigma^2)$, $k = 1, \dots, K$.
- When the model is violated, we can easily produce empirical miscalibration results for Naive Bayes; theoretical results: further assumptions required.
- But if the model happens to be true, we are in the game: show theoretically how much the Venn–Abers predictor (build on top of Bayes) suffers as compared to the Bayes algorithm.
- And this can also be done empirically: how much VAPs or IVAPs (build on top of Bayes) lose as compared to Bayes? how do they compare to standard ML algorithms? (such as Neural Nets)

Validity in the batch mode

- In practice prediction algorithms are often applied in the batch mode.
- The validity of conformal predictors in the batch mode is an empirical fact.
- Proved theoretically for inductive conformal predictors (my ACML 2013 paper).
- To do for conformal predictors (in various special cases, perhaps there are no universal results).
- To do for inductive Venn predictors and Venn predictors.

Outline

Competitions

- Part of the excitement of standard machine learning: formal and informal “races” (such as Le Cunn vs Vapnik on the USPS and NIST datasets of hand-written images).
- We need to come up with rules of the game.
- The rules are different for conformal prediction [inference] and Venn prediction [decision making].

Rules for conformal prediction

- Some criteria of efficiency traditionally used for conformal prediction are “improper” (such as the percentage of uncertain predictions).
- They encourage strange conformity measures.
- An example of a proper criterion: the average size of the prediction set.
- Decide on a small set (ideally of size 1) of proper criteria.

Rules for Venn prediction

- For probabilistic prediction, there is a well-developed theory of proper scoring rules (“encouraging honesty”).
- It can be used after multiprobability predictions have been merged into probability predictions.
- How can we evaluate the quality of multiprobability predictions? (Encouraging precision but not excessively.)

Outline

Example of application fields

- Medicine: a very suitable area for conformal prediction (important individual predictions).
- Applications for active learning (e.g., in pharmaceutical industry).
- Large-scale decision making: CVAPs.

What is needed (1)

- Good packages in, say, R and MATLAB.
- The work has started, on and off.
- We need to have a big library of (non)conformity measures and taxonomies.
- Imbalanced data sets becoming important, which requires implementing Mondrian (label-conditional) CP.

What is needed (2)

We should be able to deal with really big data sets.

- Adapting conformal and Venn predictors (and their inductive and cross-versions) to the map/reduce framework.

Other open problems or important directions?