Review: Combination of Conformal Predictors for Classification

Norio Kosaka

December 16, 2018

1 Paper Profile

- Title: Combination of Conformal Predictors for Classification
- Author: Paolo Toccaceli, Alexander Gammerman
- Organisation: Royal Holloway, University of London, CS Research Centre
- Publish Year: 2017
- URL: http://proceedings.mlr.press/v60/yanovich17a/yanovich17a.pdf

2 Prerequisites

- Conformal Prediction: Intuitively it is the methods that uses past experience to determine precise level of confidence in new predictions, and it employs statistical indicators to evaluate the validity of the prediction Good PPT: http://www.clrc.rhul.ac.uk/copa2017/presentations/CP_Tutorial_2017.pdf

3 Contents in the paper

1. Introduction
2. Results Description
   (a) Manifold Learning Data Model
   (b) Statistics Form
   (c) Main Results
3. Data Model
4. Main Results
5. Proof of Main Theorems
6. Conclusion

4 Abstract

The author proposed some possible approaches to the combination of *Conformal Predictors* in the binary classification case.

- p-value combination techniques
- Calibration of p-values into Bayes factors

And the result shows that the **P-value combined with Fisher's Method** worked fine, when ranking compounds by strength of evidence.

5 Introduction

Conformal Predictors(CP)(Vovk et al., 2005 [2]; Gammerman and Vovk, 2007[1]) was proposed to provide a validity property on the prediction stage.The efficiency of CP hugely relies on underlying *Machine Learning algorithms*. The objective of this paper is to explore ways to improve CP by some forms *ensembling*, which in general refers to the approaches, for example, Random Forests or Bagging. But the author claimed that he has differentiate the proposition from other approaches in a way that it does not explicitly aim at combating overfitting and correlation per se. So, the challenge which he has encountered was to find a method of wide applicability that combines the predictions in a synergistic way.

6 Conformal Predictors

Key points:

- the training set is made up of l independent identically distributed samples
- CP assigns a p-value to a prediction y
- **Non-Conformity Measure(NCM)** is a real-valued function α expressing how odd the sampled example is among the training set.
- Applying one of the following methods to a test set
  - Given a significance level ϵ, a *region predictor* outputs for each test object the set of labels, such that the actual label is not in the set no more than a fraction ϵ of the times.
  - Or, one can pick the largest p-value for a given test object alongside with its credibility and confidence.

Two forms of CP

- Transductive CP: computationally expensive, requires the computation from scratch for each object
- Inductive CP: requires just one training of the underlying, yet also it requires that the training data set which is split into a proper training set and a calibration set.

References

[1] Alexander Gammerman and Vladimir Vovk. "Hedging predictions in machine learning". In: *The Computer Journal* 50.2 (2007), pp. 151–163.

[2] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer. *Conformal prediction*. Springer, 2005.