Conformal Predictors

Alexander Gammerman and Vladimir Vovk
Computer Learning Research Centre
Royal Holloway, University of London

{alex,vovk}@cs.rhul.ac.uk
http://clrc.rhul.ac.uk/people/

The application of traditional machine-learning algorithms to modern high-throughput and high-dimensional (with many thousands of features) data sets often leads to serious computational difficulties. Several new algorithms, foremost support vector machines (SVM) and other kernel methods, have been developed recently with the goal of tackling high-dimensional problems. However, a typical drawback of the new algorithms is that they usually do not provide any useful measures of confidence in their predictions for new unclassified examples.

This talk will describe a technique for "hedging" predictions produced by the new and traditional machine-learning algorithms, including, for example, SVM, (kernel) ridge regression, (kernel) nearest neighbors, (kernel) logistic regression, bootstrap, decision trees, boosting, and neural networks. We call our algorithms for producing hedged predictions *conformal predictors* [1, 2].

The problem of assigning confidence to predictions is closely connected to the problem of defining random sequences, and the ideal conformal predictor can be constructed using the universal notion of algorithmic randomness; this notion was defined by A. Kolmogorov, P. Martin-Löf, and L. Levin based on the existence of universal Turing machines. Because of its universality, algorithmic randomness is not computable, and conformal predictors are based on computable approximations to it. It has been shown that such approximations can be developed using typical machine-learning algorithms; for example, for SVM the Lagrange multipliers of the support vectors can be used to approximate the randomness level of the data set.

The talk will introduce the main ideas behind this approach, paying particular attention to the following advantages of the hedging technique:

- it gives provably *valid* measures of confidence, in the sense that they never overrate the accuracy and reliability of the predictions;
- it does not make any additional assumptions about the data beyond the IID assumption (the examples are independent and identically distributed);

- it allows one to estimate the confidence in the prediction of individual examples;
- conformal predictors can be used as *region predictors*, allowed to output a range of labels as their prediction, so that one can control the number of erroneous predictions by selecting a suitable confidence level;
- the well-calibrated prediction regions produced by conformal predictors can be used in both on-line and off-line modes of learning, as well as in several intermediate modes, allowing, for example, “slow” and “lazy” teachers.

Applications of conformal predictors to a number of real-world problems, including promoters recognition and medical diagnosis, will also be considered.

References

[1] **Vovk, V., Gammerman, A., Shafer, G.:** *Algorithmic Learning in a Random World*, Springer, New York, 2005.

[2] **Gammerman, A., Vovk, V.:** “Hedging Predictions in Machine Learning”. To be published in the *Computer Journal*; the paper is also available at http://clrc.rhul.ac.uk/events/TCJ.pdf.

Topic: learning algorithms
Presentation: oral/poster