Simultaneous Learning and Prediction

Loizos Michael
Open University of Cyprus
loizos@ouc.ac.cy

In a typical supervised learning scenario, an agent trains on labeled examples, constructs a model of the data, and applies it to predict the missing labels of new examples. In many real-world settings, however, neither such explicit labels nor the examples are fully observed, as assumed. Yet, one would hope that the existing, and extensive, work on supervised learning could be still rigorously applied and exploited in these settings.

Call autodidactic learning the setting where an agent trains on partially-observable examples, constructs a model of the data, and applies it to predict the missing values of those attributes that are masked in new partially-observable examples. A physician, for instance, with access to partial patient records may autodidactically learn rules governing the human physiology, and use them to predict unobserved (N.B. not unobservable) medical conditions of new patients. Earlier work investigated how autodidactic learning algorithms can be obtained from supervised algorithms for concept learning [2, 4].

Consider, then, an algorithm that autodidactically learns a rule to predict any single given attribute. How should this algorithm be used so that the performance of reliably completing missing information in new partially-observable examples is maximized?

Assume that a rule for each attribute is available. The application of the rules on an example is: flat if no rule has access to the predictions of other rules; chained if at least one rule is applied after another, so that the former can use the prediction of the latter as if it were part of the example. It can be shown that there exist situations where the chained application of rules strictly outperforms the flat application of any set of rules.

How can the provable benefits of rule chaining be attained? Learning the rules first and then chaining them increases the completed information compared to their flat application, but at the expense of the reliability of the predictions. Simultaneous learning and prediction achieves the best of both worlds: learn a rule for each attribute; use a flat application of these rules on the training examples to complete some missing information; repeat t times on the resulting examples. Chain all available rules in the order they were learned, and apply the chained set of rules to make predictions on future examples. By appropriate choice of the learning parameters, it can be shown that SLAP-ing outperforms a flat application of these rules while retaining the reliability of the predictions (cf. experiments and theory [1, 3]), and it is fixed-parameter tractable for parameter t.

References