Classifier Chains, Trees, and Graphs for Multi-Target Learning

Jesse Read

Aalto University, Department of Information and Computer Science,
PO Box 15400, FI-00076 Aalto, Espoo, Finland, and
Helsinki Institute for Information Technology (HIIT), Finland
E-mail: jesse.read@aalto.fi

Classifier Chains is a method for multi-label (or multi-target) classification, where label predictions are cascaded along a chain of classifiers; the output label variable of each classifier is used in the input space of the following classifiers in the chain. There are now several varieties of this approach in the multi-label literature, with varying strategies for inference, and deriving a chain order and structure (including those which may be more aptly called a classifier tree or graph, rather than a chain).

It is now firmly established that classifier chain methods are able to outperform independent classifiers (that learn each label independently), and perform among the state of the art in multi-label classification, under a range of evaluation measures. The main challenges involved are

1. performing efficient inference (for L labels, there are already 2^L possible classifications only in the binary case, for a given test instance); and
2. finding an appropriate order or structure with which to model the labels (even only considering chain order yields factorial possibilities).

The performance advantage of a classifier chain clearly comes from learning labels together, although the exact mechanism behind it is still being unraveled. Most recent approaches focus on measures of dependence among labels to determine a suitable structure, yet there is still no efficient dependable recipe for best predictive performance across datasets in general; ensembles of random chains are still performing very competitively, and even in examples of conditional independence, a chain among labels can provide benefit.

Classifier chain methods do not stand out alone in the literature. Strong connections can be made with established areas such as neural networks and probabilistic graphical models, which help explain its performance advantages and limitations, and offer promise for future direction.

Although originally devised for multi-label prediction, they can also be applied successfully to other scenarios not traditionally seen as 'multi-label'. For example, segmentation, localization, and similar structured-output tasks, of which we give several examples.

Some issues are still undergoing investigation, such as error propagation due to missing and noisy labels, and the way the extra input attributes (which are outputs of other classifiers) are handled by different learners.