

Keynote

Robustness through Prior Knowledge: Using Explanation-Based Learning to Distinguish Handwritten Chinese Characters

Gerald DeJong

Professor of Computer Science and the Beckman Institute
University of Illinois, Urbana-Champaign
mrebl@uiuc.edu

Abstract

Offline handwritten digit and character recognition has been a popular research vehicle in the machine learning community. Such image patterns are unmistakable to humans and yet reflect complex variabilities including noise, writer idiosyncrasy, variations both systematic and unsystematic from the writing implement, etc. Classification of individual isolated glyphs precludes classification based on the surrounding text. In this unstructured task, digit recognition accuracy of state-of-the-art learners is 99.5%. But digit shapes are relatively simple and these accuracies are achieved with thousands of training instances for each digit. Humans exhibit greater accuracies on more complex shapes with far fewer training examples.

We believe that a key component, missing from statistical machine learning, is prior domain knowledge. Explanation-Based Learning (EBL) holds a natural place for domain knowledge. In EBL, a training example is treated as an illustration of some deeper pattern. This pattern emerges from an explanation or justification (constructed by inference over the prior domain knowledge) for why the example's features should merit the teacher-assigned class label.

Our systems learn to distinguish handwritten Chinese characters. Chinese characters can be much more complex than digits or western characters, necessitating higher resolution images. In the database we use, each Chinese character image has a resolution of 63x63 pixels (compared to 20x20 pixels in the NIST database). Furthermore, with thousands of common Chinese characters, the best available databases contain only hundreds of labeled handwritten images for each (rather than the thousands available for digits and western characters).

Our domain knowledge consists of the ideal glyph shapes, in terms of geometry among strokes, and knowledge of how strokes of a writing implement leave pixels in an image. Explanations over the training set suggest which pixel combinations can be expected to carry the bulk of the classification information. The EBL learner then biases its classifier to emphasize these corresponding high expected information pixels.

In one approach the system automatically constructs specialized Feature Kernel Functions for support vector machines. Like other kernel functions these define a distance metric between example pairs. But Feature Kernel Functions magnify the contribution to the distance metric of image pixels that have high expected class-relevant information. Another approach extends the basic SVM procedure with an additional bias penalty. This approach, which we call the Explanation-Augmented Support Vector Machine, employs a conventional kernel function but penalizes classifiers that are less consistent with confirmed explanations.

We present and evaluate these approaches, discuss some limitations, and suggest directions by which these limitations might be overcome.