massachusetts institute of technology — artificial intelligence laboratory

An Empirical Comparison of SNoW and SVMs for Face Detection

Mariano Alvira and Ryan Rifkin

AI Memo 2001-004 January 2001
CBCL Memo 193

© 2001 massachusetts institute of technology, cambridge, ma 02139 usa — www.ai.mit.edu

Abstract

Impressive claims have been made for the performance of the SNoW algorithm on face detection tasks by Yang et. al. [7]. In particular, by looking at both their results and those of Heisele et. al. [3], one could infer that the SNoW system performed substantially better than an SVM-based system, even when the SVM used a polynomial kernel and the SNoW system used a particularly simplistic "primitive" linear representation. We evaluated the two approaches in a controlled experiment, looking directly at performance on a simple, fixed-sized test set, isolating out "infrastructure" issues related to detecting faces at various scales in large images. We found that SNoW performed about as well as linear SVMs, and substantially worse than polynomial SVMs.

This report describes research done within the Center for Biological & Computational Learning in the Department of Brain & Cognitive Sciences and in the Artificial Intelligence Laboratory at the Massachusetts Institute of Technology.

This research was sponsored by grants from: Office of Naval Research under contract No. N00014-93-1-3085, Office of Naval Research (DARPA) under contract No. N00014-00-1-0907, National Science Foundation (ITR) under contract No. IIS-0085836, National Science Foundation (KDI) under contract No. DMS-9872936, and National Science Foundation under contract No. IIS-9800032.

Additional support was provided by: Central Research Institute of Electric Power Industry, Center for e-Business (MIT), Eastman Kodak Company, DaimlerChrysler AG, Compaq, Honda R&D Co., Ltd., Komatsu Ltd., Merrill-Lynch, NEC Fund, Nippon Telegraph & Telephone, Siemens Corporate Research, Inc., and The Whitaker Foundation.

1 Introduction

Face detection systems have been used to evaluate learning algorithms and feature selection. The present study focuses on the experiments performed by Heisele, Poggio, and Pontil [3] using Support Vector Machines [6], and by Yang, Roth and Ahuja [7] using the Sparse Network of Winnows [1], or SNoW, algorithm.

On a particular face recognition dataset containing 125 images with a total of 483 faces, Yang et. al. claimed a dection rate of 93.6% with only 3 false positives, using the so-called "primitive" feature representation. This representation contained one feature for every possible gray-scale value of every pixel. Each feature had value zero or one.

This result indicated a better precision-recall tradeoff than any other published result. Using SVMs on the same testing set, with a polynomial kernel, Heisele et. al. were only able to achieve a detection rate of 85.6%, with 9 false positives, or a detection rate of 89.9%, with 75 false positives. We were quite surprised that SNoW could perform so well using such a simple representation scheme, and decided to do our own controlled experiments to better evaluate the SNoW algorithm.

2 Data Sets and Software

One important possible source of variation that we wanted to control for was the methodology used to detect faces in large images. In general, windows at several scales are placed at all possible positions in the image, and the underlying classifier is invoked. The extent to which recognitions suppress other nearby recognitions could have a potentially large effect on the total accuracy. To control for this, we decided to use a test set containing images of the same size that the classifiers were trained on. This allowed us to sidestep any differences resulting from the underlying infrastructure systems and compare the algorithms directly.

The data set was similar to and derived from the one used by Heisele et. al [3], although not identical. It consisted of a training set of 6977 images (2429 face and 4548 non-face) and a test set of 24045 images (472 face and 23573 non-face). The images were 19x19 grayscale and histogram normalized. The data is available on the CBCL webpage [2].

To train and test SVMs, we used SvmFu version 2.001 [2]. We used SNoW version 2.0.3 [1].

3 Experimental Results

We trained SNoW using the primitive features, described above. The images were 8-bit grayscale, with 361 pixels each, resulting in $361 * 256 = 92416$ features, with exactly 361 features active per image. We trained linear SVMs using these binary features. We also trained linear SVMs using the original grayscale values as features. Finally, we trained SVMs using the grayscale values and a polynomial kernel of degree 2. Instead of computing just a single point on the ROC curve [4], we generate the entire ROC curve for each method.

Looking at Figure 1 we see that SNoW with binary features performs approximately as well as the linear SVMs, but substantially worse than the polynomial SVM.

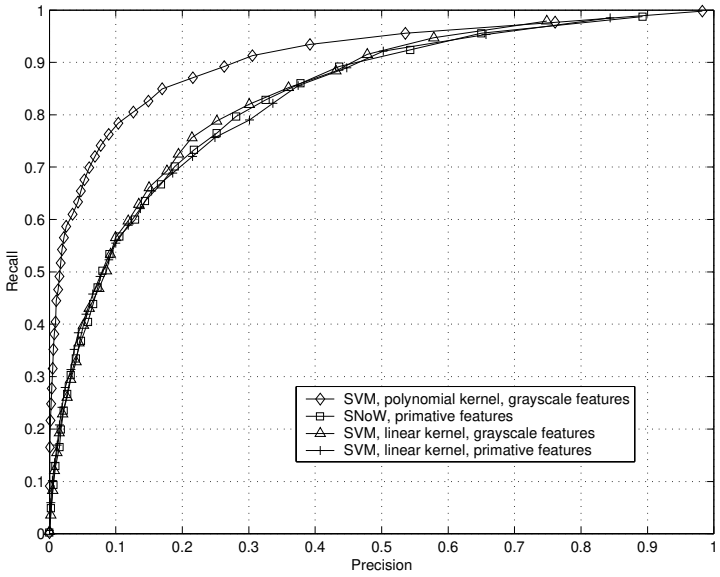

Figure 1: Face Detection ROC Curves

4 Discussion

Impressive claims have been made for the performance of the SNoW system on face classification tasks. However, when restricted to the pure "classification" component of the face detection task, we found that SNoW did not perform particularly well. One possible source of discrepancy is the infrastructure systems used by the various algorithms to find faces in large images. For instance, if SNoW's system were substantially better at suppressing closely occurring false positives, this could explain the published results [5]. However, this would have nothing to do with the underlying classifiers, and we would then expect a polynomial SVM using SNoW's infrastructure system to perform even better than SNoW did.

Additionally, this study points out some of the difficulties involved in comparing algorithms for face detection. Firstly, we suggest that displaying entire ROC curves is more appropriate than simply giving a single point on that curve in the form of a recognition rate along with a number of false positives [4]. More importantly, we suggest that for the comparison to be fair, the two algorithms should be trained and tested on precisely the same data, and that this comparison should be separated from the infrastructure needed to detect faces at various scales in large images. Because infrastructure differences can give rise to substantial differences in system performance, it is difficult to impossible to accurately compare classification algorithms by comparing the outputs of complete systems. Detecting faces in large images is certainly important for real-world systems, but ideally, this task should be separate from the face detection algorithm per se: we should use the best possible face detection algorithm *and* the best possible infrastructure system. To help address this issue, we have made the data used in this study available on the CBCL webpage [2].

References

[1] Andrew J. Carlson, Chad M. Cumby, Jeff L. Rosen, and Dan Roth. Snow user's guide. Technical report, UIUC, 1999.

[2] http://www.ai.mit.edu/projects/cbcl/software-datasets/index.html.

[3] Bernd Heisele, Tomaso Poggio, and Massimiliano Pontil. Face detection in still gray images. *CBCL Paper#187/AI Memo #1687*, 2000.

[4] Foster Provost, Tom Fawcett, and Ron Kohavi. The case against accuracy estimation for comparing induction algorithms. *IMLC-98*, 1998.

[5] Dan Roth. Personal communication., 2000.

[6] Vladimir N. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, 1998.

[7] Ming-Hsuan Yang, Dan Roth, and Narendra Ahuja. A snow-based face detector. *NIPS-12*, 1999.