Encoding Reusable Perceptual Features Enables Learning Future Categories from Few Examples

Michael Fink
Interdisciplinary Center for Neural Computation
The Hebrew University of Jerusalem
Jerusalem 91904, Israel
fink@huji.ac.il

Kobi Levi
School of Computer Science and Engineering
The Hebrew University of Jerusalem
Jerusalem 91904, Israel
kobilevi@cs.huji.ac.il

September 29, 2004

Abstract

A perceptual system coping with a dynamic environment must be able to learn to detect new object categories from a few examples. However, learning from a small sample is restricted by the hindering effects of model overfitting. We present an algorithm aimed at circumventing the effects of overfitting by utilizing a set of reusable features, learned from several previously trained categories. We show that when applying a feature reuse strategy the algorithm learns complex real world objects from few examples.

Introduction

Learning from few examples minimizes training cost. Two difficulties arise when one attempts to learn a perceptual category from few examples. First, a small sample might have an infinite number of incidental characteristics that are not representative of the entire class, which might lead to many false alarms. Second, the structure of the category might not be fully represented in a small sample, thus leading to many misses. Both effects are a result of an overfitting between the learned model and the small sample.

In the extreme case of learning from a single example it seems that the crippling effects of overfitting might prevent any robust category generalization. Nevertheless, the human perceptual system possesses the capacity of learning to detect a new visual object category from a brief exposure to a single example (see Fig. 2). We will term this as the *efficient detection learning* characteristic.

In training to detect a new object the features that characterize the new category must be selected from a large set of potential features. Detection schemes based on a selected set of simple features have been demonstrated to exhibit robustness to noise and category variation in many qualitatively different perceptual categories like human faces (Viola & Jones, 2002), moving cars (Levin, Viola, & Freund, 2003) and walking pedestrians (Viola, Jones, & Snow, 2003). These methods require many labeled training examples (typically in the range of 1,000-10,000 positive examples and usually many more negative ones). Like many other models, feature selection frameworks are prone to the over fitting effects induced by small samples. When viewing the human *efficient detection learning* characteristic (demonstrated in Fig. 2) in the context of a feature selection paradigm, it remains unclear how the effect of overfitting is avoided.

We claim that in many natural detection settings a perceptual system might identify a restricted set of features that captures the regularities of many perceptual categories. The existence of such a visual alphabet could explain why the *efficient detection learning* capacity is not hindered by the effects of overfitting. Following Thrun (1996) we state that a rich feature set, emerging after many years of perceptual learning and feature selection processes, might enable *efficient detection learning*.

A Feature Reuse Algorithm

We have previously hypothesized that a feature set echoing the regularities of many objects in a given environment might overcome the effects of overfitting and enable detection learning from small samples. This section describes an algorithm aimed at implementing the feature reuse principle. We consider a scenario composed of two consecutive stages. First, in the *prolonged multiclass stage*, the characteristics of many objects ($N > 1$) are learned. This initial learning might require many examples of each category. Next, in the constrained stage, a novel category ($N + 1$)

Figure 1: Stimulus used in demonstration 1

Figure 2: When asked which of the five creatures is related to the creature in Fig. 1, most participants point out the correct one without requiring an additional look, thus demonstrating object detection capabilities after a single-trial learning phase.

appears in the already familiarized environment. This novel category poses limited training conditions. Our algorithm postulates two types of perceptual learning:

1. Find a non-incidental feature set, by selecting features that characterize many of the initial N categories.
2. Restrict training of the $N+1$ category to the previously selected non-incidental feature set.

Although many mechanisms might be used for implementing these two types of perceptual learning (see for example (Ullman, Vidal-Naquet, & Sali, 2002)), we chose an iterative procedure called AdaBoost as our feature selection mechanism (see (Freund & Schapire, 1999)). In this feature selection scheme an importance weight (initialized as $w_i = 1/M$) is associated with each positive $y_i = 1$ and negative $y_i = -1$ example $x_{i,i=1,..,M}$ of the trained category n. At each round t of the training process one new feature h_t is selected. The criterion for selection is based on finding a feature that discriminates well between the *weighted* positive and negative examples thus minimizing the error $\sum_{i:h(x_i^n)\neq y_i^n} w_i$

Next the relative weights of all of the examples that were misclassified using the selected feature are increased, thus the next feature to be selected at time $t+1$ will focus on the sections of the sample that were difficult for the previous features. Detection of the learned object category is performed by a voting scheme between the selected features. This procedure could be continued until iteration T when the error on the labeled training sample is reduced to zero.

A naive application of AdaBoost will perform this procedure independently for each category and select a different feature set for each target category (Fig. 3a). In order to implement a feature reuse strategy the goal of the feature selection procedure was modified to find at each round a non-incidental feature that characterizes well many of the N categories, minimizing $< \sum_{i:h_k(x_i^n)\neq y_i^n} w_i^n >_n$. We adopt the relaxing assumption that a feature performing slightly better than chance on all categories could always be found. Once the initial learning stage is concluded a new object category is efficiently learned by restricting the training to already provably non-incidental features (Fig. 3b).

	a. **Independent Feature Selection**	b. **Maximize Feature Reuse**

prolonged multiclass stage	**independentTrain** of classes $1 \ldots N$
	input: $(x_i^n, y_i^n)_{i=1}^{M_n}$ examples of classes $n = 1, .., N$
	initialize: $w_i^n = \frac{1}{M_n}$
	for $n = 1, ..., N$
	  for $t = 1, ..., T$
	  $h_t^n = weakLearner$
	  $E_t^n = \sum_{i:h_t^n(x_i^n) \neq y_i^n} w_i^n$
	  $\alpha_t^n = \frac{log(1-E_t^n)}{2E_t^n}$
	  update: $w_i^n = w_i^n e^{-\alpha_t^n h_t(x_i^n) y_i^n}$
	  normalize: $w_i^n = \frac{w_i^n}{\sum_i w_i^n}$
	return $H^n = \{h_t^n\}, A^n = \{\alpha_t^n\}$ $n = 1, ..., N$
	classify new x as $sgn(\sum_t \alpha_t^n h_t^n(x))$
		featureReuseSelection of classes $1 \ldots N$
		input: $(x_i^n, y_i^n)_{i=1}^{M_n}$ examples of classes $n = 1, .., N$
		initialize: $w_i^n = \frac{1}{M_n}$
		for $t = 1, ..., T$
		  $h_t = commonWeakLearner$
		  for $n = 1, ..., N$
		  $E_t^n = \sum_{i:h_t^n(x_i^n) \neq y_i^n} w_i^n$
		  $\alpha_t^n = \frac{log(1-E_t^n)}{2E_t^n}$
		  update: $w_i^n = w_i^n e^{-\alpha_t^n h_t(x_i^n) y_i^n}$
		  normalize: $w_i^n = \frac{w_i^n}{\sum_i w_i^n}$
		return $H = \{h_t\}, A^n = \{\alpha_t^n\}$ $n = 1, ..., N$
		classify new x as $sgn(\sum_t \alpha_t^n h_t(x))$

constrained stage	**independentTrain** of class $N+1$
	input: $(x_i^{N+1}, y_i^{N+1})_{i=1}^{M}$ examples of class $N+1$
	initialize: $w_i = \frac{1}{M}$
	for $t = 1, ..., T$
	  $h_t^N + 1 = weakLearner$
	  $E_t = \sum_{i:h_t^{N+1}(x_i^{N+1}) \neq y_i^{N+1}} w_i$
	  $\alpha_t = \frac{log(1-E_t)}{2E_t}$
	  update: $w_i = w_i e^{-\alpha_t h_t^{N+1}(Xni) Yni}$
	  normalize: $w_i = \frac{w_i}{\sum_i w_i}$
	return $H^{N+1} = \{h_t^{N+1}\}, A^{N+1} = \{\alpha_t\}$
	classify new x as $sgn(\sum_t \alpha_t^{N+1} h_t^{N+1}(x))$
		efficientTrain of class $N+1$
		input: $(x_i^{N+1}, y_i^{N+1})_{i=1}^{M}$ examples of class $N+1$
		initialize: $w_i = \frac{1}{M}$
		for $t = 1, ..., T$
		  $h_t = H_t$
		  $E_t = \sum_{i:h_t^{N+1}(x_i^{N+1}) \neq y_i^{N+1}} w_i$
		  $\alpha_t = \frac{log(1-E_t)}{2E_t}$
		  update: $w_i = w_i e^{-\alpha_t h_t(Xni) Yni}$
		  normalize: $w_i = \frac{w_i}{\sum_i w_i}$
		return $H = \{h_t\}, A^{N+1} = \{\alpha_t^{N+1}\}$
		classify new x as $sgn(\sum_t \alpha_t^{N+1} h_t(x))$

	weakLearner $E_k^n = \sum_{i:h_k^n(x_i^n) \neq y_i^n} w_i^n$ $h^n = argmin_{h^k}(E_k^n)$ **return** h^n	**commonWeakLearner** $E_k = < \sum_{i:h_k(x_i^n) \neq y_i^n} w_i^n >_n$ $h = argmin_{h^k}(E_k)$ **return** h

Figure 3: (a) Multiclass pseudo code for an independent feature selection strategy (b) Multiclass pseudo code for feature reuse strategy.

Efficient Detection Learning in Office Scenes

We would like to test whether the proposed feature reuse mechanism can facilitate detection of real world objects. Our application details (including the candidate feature set and the training and testing data) are identical to those described at Levi, Fink, and Weiss (2004). We will therefore constrain ourselves to describing the performance results. As in Levi et al. (2004) we selected computer monitors as our target object category for the constrained training stage. Ten images of computer monitors were randomly selected for training and 100 pictures containing other monitors were set aside for later evaluating the detection performance. Object images were taken from the Caltech office scene database described at (Fink & Perona, 2003). Negative examples of the object category were generated by cutting sections of the images known not to contain computer monitors. The AdaBoost feature selection scheme was then applied until reaching perfect classification of the training examples. Under these conditions learning to detect a monitor from ten examples leads to a very poor detector with a 9% detection rate ($3x10^{-5}$ false alarms).

We now focus on examining the feature reuse model by first performing a prolonged training stage of four other categories frequently appearing in a similar office environment. Doors, picture frames, book shelves and light switches were used for this initial training stage. We restrict the training of the monitor detector from the initial 1.3 million candidate features to 100 features selected by the feature reuse model. Under these conditions the detection rate was significantly improved to 44% ($1.6x10^{-5}$ false alarms).

Summary

We have presented a simple algorithm for incorporating the feature reuse policy into a Boosting framework. By restricting training to features that characterize many objects in a *prolonged multiclass stage* we are able to avoid overfitting when learning future categories from few examples in the *constrained stage*. A similar feature reuse strategy has been proposed by Torralba, Murphy, and Freeman (2004) while aiming at multiclass detection rather than learning from few examples. Our proposed method relies on the hypothesis that the features encoding the classes in the prolonged stage are both adequately restricted (in order to avoid overfitting) and nevertheless sufficiently relevant for training the novel object detector in the *constrained stage*. When the number of initial classes is very large, including many classes that are not relevant for the target class, we remain with the challenge of how to automatically prune down the set of classes to be used in the prolonged stage. This is the focus of our current research.

References

Fink, M., & Perona, P. (2003). Mutual boosting for contextual inference. *Proceedings of the Neural Information Processing System conference (NIPS) 2003*.

Freund, Y., & Schapire, R. E. (1999). A short introduction to boosting. *Journal of Japanese Society for Artificial Intelligence*, *14*(5), 771–780.

Levi, K., Fink, M., & Weiss, Y. (2004). Learning from a small number of training examples by exploiting object categories. *Proceedings of the Computer Vision and Pattern Recognition conference: Workshop of Learning in Computer Vision (LCVPR) 2004*.

Levin, A., Viola, P., & Freund, Y. (2003). Unsupervised improvement of visual detectors using co-training. *Proceedings of the International Conference on Computer Vision (ICCV) 2003*.

Thrun, S. (1996). *Learning to learn: Introduction*. Kluwer Academic Publishers.

Torralba, A., Murphy, K., & Freeman, W. (2004). Sharing visual features for multiclass and multiview object detection. *Proceedings of the Computer Vision and Pattern Recognition conference (CVPR) 2004*.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, *5*(7), 682–687.

Viola, P., & Jones, M. (2002). Robust real-time object detection. *International Journal of Computer Vision*.

Viola, P., Jones, M., & Snow, D. (2003). Detecting pedestrians using patterns of motion and appearance. *International Journal of Computer Vision*, *2*, 735–741.