Features for Nearly-Infallible (Name Subject to Change) Classification

Jonathan Krause
Computer Science Department, Stanford University
Stanford, CA
jkrause@cs.stanford.edu

1. Introduction

Within the field of object classification, the goal of nearly-infallible classification is to maximize reward over some set of classes, while maintaining a very high, pre-specified accuracy. For any dataset of even a reasonable size, current state-of-the-art algorithms have no chance of being able to achieve an arbitrary level of accuracy. However, if the possible labels associated with each image form a semantic hierarchy, then one can achieve any arbitrary accuracy by predicting that certain examples are members of more general classes, such as "Dog" instead of "German Shepherd" and "Object" instead of "Animal". There is an algorithm that achieves this, but it currently uses as input features that were chosen because they perform well with regards to the classic multiclass clasification task. However, it has yet to be seen how well the algorithm does under a varying choices of features, and whether there are other state-of-the-art classification algorithms that are more suitable to the infallible classication problem. That is the goal of this work.

2. Problem Statement

Formally, the objective for infallibile classification is

$$\begin{array}{ll} \underset{f(x)}{\text{maximize}} & \mathbb{E}\left[r_{f(x)}[f(x) \in \pi(y)]\right] \\ \text{subject to} & \mathbb{E}[f(x) \in \pi(y)] \geq 1 - \epsilon \end{array}$$

where x is an input image, y is its correct class as a leaf node in the semantic hierarchy, $r_{f(x)}$ is a reward associated with class $f(x)$, and $\pi(y)$ is the path from y to the root node in the hierarchy. That is, the goal is to maximize the expected reward while maintaining an accuracy of at least $1 - \epsilon$.

2.1. Data

The data that will be used, at least initially, is a subset of 57 leaf nodes in the ImageNet [1] dataset, together with 8 internal nodes (Object, Animal, Vehicle, Bird, Cat, Dog, Boat, Car). There are also datasets consisting of 1,000 and 10,000 nodes in the ImageNet hierarchy which can be used for more promising features, though these will not be the primary datasets involved since creating very discriminative features on the millions of images in the 10,000-node dataset takes more than a day on a powerful cluster. The ultimate goal, though, would be to use the best-performing features on all of the 10,000-node dataset. As the ImageNet dataset is semantically very rich, this could conceivably become a system capable of classifying any object with high accuracy.

3. Technical Approach and Current Progress

Currently most work has been done on the algorithmic side of infallibile classification, which is not the focus of this work. On the plus side, there is a definitive algorithm that will be used (called DARTS), which involves doing a binary search on a dual variable to obtain the optimal reward under the infallibility contraint. Posterior probabilities are obtained at leaf nodes using an SVM and Platt scaling, then are summed to get probabilities for internal nodes.

The current approach to nearly-infallible classification uses Locality-constrained Linear Coding [3] for features. The very first step will be to reproduce those results independently, which shouldn't take very long once work on features is underway. At the same time, one or two weaker features will be tried to get a sense of how DARTS does under more punitive conditions. These should result in poor performance, but we will still satisfy the $1 - \epsilon$ constraint on accuracy.

Another state-of-the-art classifier that will be tested is Object Bank [2]. Due to existing code provided by the creators of Object Bank, this should also not take long to test. It is worth noting that while Object Bank is primarily concerned with scene classification, where having responses to multiple object filters would seem to be more useful than in the single-object case, since it does not require too much effort to try and might perform well,

References

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[2] L. Li, H. Su, E. Xing, and L. Fei-Fei. Object bank: A high-level image representation for scene classification & semantic feature sparsification. In *Neural Information Processing Systems (NIPS)*.

[3] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. 2010.