MANIFOLD-BASED APPROACHES FOR IMPROVED CLASSIFICATION

Mark A. Davenport,[r] Chinmay Hegde,[r] Michael B. Wakin[m] and Richard G. Baraniuk[r]

[r] Department of Electrical and Computer Engineering, Rice University

[m] Department of Electrical Engineering and Computer Science, The University of Michigan at Ann Arbor

ABSTRACT

While manifold structure is often exploited for dimensionality reduction or feature extraction, this structure is rarely used by classification algorithms. We present a class of algorithms that utilize the low-dimensional manifold nature of signal ensembles and result in improved classification performance. The algorithms are built within theoretical frameworks that take into consideration prior knowledge of geometric structure in both labeled and unlabeled data points. Additionally, these frameworks can exploit recent results on random projections of smooth manifolds to ensure computational feasibility on extremely high-dimensional problems.

1. INTRODUCTION

In many classification problems, the fundamental assumption is that distinct classes are well-separated in the feature space (often represented by a subset of $\mathbb{R}^N$). Thus, the goal is to construct a hypersurface in $\mathbb{R}^N$ that divides the feature space into distinct sub-regions. Algorithms typically try to find a surface that is in some sense "optimal", for instance, one that minimizes the probability of classification error. However, there is a catch: often, this approach fails to fully exploit the intricate geometric relationships that may exist both among the training data as well as the unlabeled samples.

In an attempt to address this shortcoming, one option is to assume that the available data resides on a manifold $\mathcal{M}$ in the ambient space. This has been successfully exploited to improve the accuracy of semi-supervised classification in [1]. Typically, the dimension K of the manifold is many orders of magnitude smaller than the native dimension N of the data. While this assumption is not universally applicable, it can been rigorously shown that smooth manifolds are good models for a variety of naturally occurring signal classes. Of particular interest to us are image appearance manifolds (IAMs), which in many cases can be shown to be isometric to Euclidean space [2]; . A simple example is the set of all images generated by translating a simple object in a 2D plane.

While manifold-related algorithms have generated considerable attention in the machine learning literature, most of them do not address practical issues encountered in classification problems, such as the presence or absence of noise, the possibility of having insufficient training data, and the extent of topological knowledge about the different signal classes. To address these issues, we require algorithms which are based on a strong theoretical foundation, and which are particularly tailored to take into account the intrinsic structure of both the available training data as well as the incoming unlabeled examples.

2. TOPOLOGY-AWARE CLASSIFICATION

We now describe a succession of manifold-based classification algorithms that can help enhance performance in the following senses:

1. The manifold assumption leads to *lower probability of error*, i.e., better classification rates.
2. The manifold model helps increase the *speed* of the classification process; this is particularly relevant in high-dimensional problems.

Generalized Maximum Likelihood Classification (GMLC)
A manifold-based framework for performing optimal classification (in the maximum likelihood sense) is developed in [3]. In this setting, we wish to classify an observed signal $y \in \mathbb{R}^N$ into one of P possible classes. We assume that $y = x + \omega$, where ω is an N-dimensional noise vector corrupting the "true" signal x. We then model each class by a K-dimensional manifold $\mathcal{M}_i$, i.e., $x = f_i(\theta_i)$ for some $i = 1, \ldots, P$, where θ_i is a K-dimensional parameter vector. Let $\mathcal{H}_i$ be the hypothesis that the image x is of class C_i for some $i = 1, \ldots, P$. The GMLC classifier is given by:

$$C(y) = \arg \max_{i=1,\ldots,P} p(y|\widehat{\theta}_i, \mathcal{H}_i),$$

where $p(y|\widehat{\theta}_i, \mathcal{H}_i)$ denotes the probability distribution of y conditioned on θ_i and $\mathcal{H}_i$, and

$$\widehat{\theta}_i = \arg \max_{\theta} p(y|\theta, \mathcal{H}_i)$$

is the maximum likelihood estimate of the parameter vector θ_i under hypothesis $\mathcal{H}_i$. Under a spherically symmetric noise model for ω, this can be shown to be equivalent to a two step procedure where we first find the closest point on each manifold to the observation y, and then classify according to which manifold is closest.

It is worth noting that in the case where the manifold is parameterized by an unknown shift of a known signal and the noise is spherically symmetric, the GMLC classifier is equivalent to the classical matched filter classifier (used in digital communications). Thus, the GMLC framework simply generalizes the matched filter to a arbitrary signal manifolds.

Note also that under the assumption that our training data consists of the union of sufficiently dense samplings of each of the underlying classes, the GMLC classifier reduces to the familiar nearest-neighbor (NN) labeling scheme. This is usually easy to implement for moderate problem sizes. An important additional benefit of this approach is as follows: suppose we possess explicit generative models[1] for our signal manifolds $\mathcal{M}_i, i = 1, \ldots, P$.

[1] Such models are frequently encountered in image rendering and computer graphics applications.

Then, we may *directly* compute the signals for the maximum likelihood parameters $\widehat{\theta}_i$ corresponding to each $\mathcal{M}_i$, and pick the class based on which signal is closest to our observation y.

Classification using multiscale manifold navigation

For higher-dimensional data or larger datasets, it can be computationally prohibitive to find the maximum likelihood parameter estimates for each class since doing so will involve calculating the distance between y and a large number of candidate signals belonging to each particular class. To speed up the classification, the following alternative has been proposed [4]: the procedure for obtaining the maximum likelihood estimate $\widehat{\theta}_i$ is to minimize

$$D(\theta_i) = ||y - f_i(\theta_i)||_2^2.$$

by starting with an initial estimate of θ_i and iteratively computing the maximum likelihood estimate using Newton's method. Unfortunately, in many important cases the manifolds may not be smooth. In particular, the IAMs induced by images that contain sharp edges are nowhere differentiable [5]. In this event, instead of directly trying to minimize $D(\theta)$, we minimize a sequence of functions

$$D_n(\theta_i) = ||G_n y - G_n f_i(\theta_i)||_2^2,$$

where $G_1, G_2, \ldots$ is a sequence of nested regularization kernels at progressively finer scales as advocated in [5]. To handle the possibility of local minima, we can repeat the minimization algorithms with different initial estimates of θ_i.

Classification using nonlinear manifold learning

The algorithms described above rely on either a generative model for the different classes or the presence of a sufficiently dense sampling of the underlying manifolds. In the event that the sampling is not very dense, it may happen that the nearest neighbor for any given unlabeled point y among the available training data accidentally belongs to a class *different* from its true class. At first glance, this seems to be an insurmountable problem. However, if we realize that there is also a manifold structure *in the unlabeled samples*, we can exploit implicit geometric relationships between the training and test data points to improve our classification performance.

As a first step, suppose that the set of test (unlabeled) points Y consists of samples exclusively drawn from one of the P classes. For each i, append Y to the set of training points from class $\mathcal{M}_i$ (which is of intrinsic dimension K). Perform a nonlinear manifold learning operation (for instance, Isomap[2] [6]) and project the data into K-dimensional Euclidean space. The performance of this nonlinear mapping is typically quantified in terms of a metric that describes how well the lower-dimensional representation "fits" the original data (in case of Isomap, this is known as the *residual variance*). The class that yields the lowest value according to our metric is declared as the class of the set of unlabeled points.

This scheme can be used in classification of slowly-varying video sequences, where it is known that all the unlabeled images arise from sampling a single candidate manifold among several test cases. Thus, this type of algorithm exploits topological structure among both training and test examples.

In the case where Y consists of samples drawn from a number of different possible classes, we may proceed similarly, but in a greedy fashion: we separately classify each sample and sequentially append the classified elements of Y to the set of labeled points, thus increasing the overall size of our training dataset.

[2]Isomap can be proven to yield good learning performance for certain types of image appearance manifolds [2].

3. RANDOM PROJECTIONS FOR HIGH-DIMENSIONAL PROBLEMS

Most manifold-based methods involve performing a considerable number of computations with the available training and test data; this may prove prohibitively expensive if the dimension N of the data is very large. Fortunately, new theory [3, 7, 8] examining the method of *random linear projections* of the original data enables the above classification algorithms to become computationally feasible. The central theme of these papers is as follows: if the original data is of dimension N, but lies on a K-dimensional manifold, then if the data is projected into a *random* subspace of dimension $O(K \log N)$, the classification performance for each of the above algorithms remains *virtually the same* (this statement is made quantitatively precise in the indicated references).

Thus, manifold models do not merely aid in improved classification performance; they also ensure stable embedding of the available data under a small set of random projections and pave the way for the development of feasible, possibly real-time classification algorithms.

4. OPEN QUESTIONS

The algorithms described above can be shown to lead to significant gains in performance. However, there are several open problems. The precise effect of noise on manifold-based algorithms has not been fully explored from a theoretical perspective. Additionally, the performance of classification algorithms can sharply degrade under improper (e.g., non-uniform) sampling of the underlying manifolds. Also, the bounds obtained in some of the results pertaining to the method of random projections are imprecise. Research is underway to address these issues.

5. REFERENCES

[1] M. Belkin and P. Niyogi, "Using manifold structure for partially labelled classification," in *Advances in NIPS*, vol. 15. MIT Press, 2003.

[2] C. Grimes and D.L. Donoho, "Image manifolds isometric to euclidean space," *Journal of Mathematical Imaging and Vision*, 2005.

[3] M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Proc. IS&T/SPIE Symposium on Electronic Imaging: Computational Imaging*, 2007, vol. 6498.

[4] M. F. Duarte, M. A. Davenport, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "Multiscale measurements for compressive classification," in *ICIP*, 2007.

[5] M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk, "The multiscale structure of non-differentiable image manifolds," in *SPIE Wavelets XI*, 2005.

[6] J. B. Tenenbaum, V.de Silva, and J. C. Landford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.

[7] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," 2007, To appear in *Foundations of Computational Mathematics*.

[8] C. Hegde, M.B. Wakin, and R.G. Baraniuk, "Random projections for manifold learning," in *Neural Information Processing Systems (NIPS)*, 2007.