TARGET DETECTION WITH SPATIO-SPECTRAL DATA VIA CONCORDANCE LEARNING

M. Murat Dundar

Computer & Information Science Department
Indiana University - Purdue University (IUPUI)
Indianapolis, IN 46022

ABSTRACT

In challenging environments, in order to uniquely define a sample as a target, multiple representations of the samples might be required. As a case study, we consider cars in the parking lots of an urban imagery as targets. What makes this problem challenging is the copresence of several parking garages and parking lots in the same imagery. Both the cars in the parking lots and in the parking garages present with similar spectral characteristics. Spectral representation alone is not sufficient to uniquely define a pixel as a car in the parking lot. In this example, before a pixel is confirmed as a target or rejected as not being a target, classifiers corresponding to spectral and spatial representations of the samples has to concord. The current study discusses some possible ways these classifiers can be trained so that the rate of true concordance is maximized. We consider independent training and feature concatenation first and then propose a joint optimization scheme. The proposed approach aims to optimize multiple classifiers at once so as to maximize concordance among the classifiers while minimizing the classification error.

***Index Terms*—** concordance learning, multiple representation, heterogeneous data, target detection

1. INTRODUCTION

Given a dataset where each sample is characterized by multiple feature vectors each extracted from a different representation of that sample, we define concordance learning as jointly optimizing one classifier for each of the underlying representations of the data such that the number of samples classified correctly by concordance is maximized. The key point here is that we seek concordance before a sample is assigned to one of the classes. When no concordance can be achieved the existing evidence is considered conflicting in which case additional evidence might be required. Collecting additional evidence may be very costly. Therefore to minimize cost, it is important to exploit the evidence already extracted from different aspects of the problem so as to better correlate one aspect with the other. This becomes a challenging problem especially when each representation is only vaguely characterized with several redundant and perhaps irrelevant features available and each class data may contain multiple subgroups.

To avoid confusions it is worthwhile to link concordance learning with co-training [1]. Like concordance learning co-training is a learning algorithm proposed to deal with multiple representations of the data. Co-training assumes that each representation of the data would be sufficient for learning if there was enough labeled data. Even though this is not required in concordance learning, problems with known correlations across different representations are the main target of this work. The main motivation for the co-training algorithm is to use unlabeled data to boost the performance of a learning algorithm in the presence of limited labeled data. As such unlike concordance learning, co-training algorithm does not require concordance among classifier outputs and thus joint optimization of the classifiers are not necessary.

Another confusion could stem from misinterpreting concordance learning as a purely data fusion technique. Even though we believe concordance learning can also be used as a data fusion technique under special circumstances, it fits better in applications where feature sets extracted from different representations of the problem are correlated as opposed to being complementary.

In this study we propose a learning algorithm for target detection in hyperspectral imagery. More specifically we aim to detect cars in the parking lots of an urban imagery. What makes this problem challenging is the copresence of several parking garages and parking lots in the same imagery. Both the cars in the parking lots and in the parking garages present with similar spectral characteristics. Therefore a learning algorithm based only on the spectral data does not uniquely identify cars in the parking lots. Even though the spectral characteristics are quite similar, the immediate spatio-spectral background surrounding the cars differs significantly between parking lots and parking garages, i.e. parking garages are made of concrete whereas parking lots are made of asphalt.

2. PROPOSED APPROACH

To deal with this problem, we first characterize each pixel in the hyperspectral imagery with its spectral and spatio-spectral representations and then design two binary classifiers one for each of the representations. During real-time classification, a pixel is confirmed as target when outputs of both classifiers concord as positive, i.e. when spectral classifier indicates a pixel for a car *and* spatio-spectral classifier indicates a pixel for a parking lot.

This study mainly focuses on the training of these two classifiers. Our approach proposes to jointly optimize these two classifiers by minimizing the total cost of discordance between the output of the classifiers while making sure both classifiers are sufficiently well-regularized to prevent overfitting. Unlike traditional learning algorithms where the cost of misclassification rate is minimized during training, the proposed concordance scheme minimizes the total cost of discordance between the outputs of the classifiers while also minimizing the misclassification rate. Since during online execution concordance is sought between the outputs of the classifiers before a pixel is confirmed as target, by imposing the cost of discordance on the objective function during training, the offline training and online execution objectives are better aligned for the underlying problem.

The proposed approach is based on the extension of the hyperplane classifiers with multi-variable hinge loss functions.

2.1. Hyperplane Classifiers

We are given a training dataset $\{(x_i, y_i)\}_{i=1}^{\ell}$, where each pixel s_i is characterized by a feature vector $x_i \in \Re^d$ and $y_i \in \{-1, 1\}$ is the corresponding ground truth and ℓ is the number of samples. We consider a class of models of the form $f(x) = \alpha^T x$, with the sign of $f(x)$ predicting the label associated with the point x. An hyperplane classifier with hinge loss can be designed by minimizing the following cost function.

$$\mathcal{J}(\alpha) = \Phi(\alpha) + \sum_{i=1}^{\ell} w_i max\left(0, 1 - \alpha^T y_i x_i\right) \tag{1}$$

where the function $\Phi : \Re^{(d)} \Rightarrow \Re$ is a regularization function or regularizer on the hyperplane coefficients and $max(0, 1 - \alpha^T y_i x_i)$ represents the hinge loss, and $\{w_i : w_i \geq 0, \forall i\}$ is the weight preassigned to the loss associated with x_i. For balanced data usually $w_i = w$, but for unbalanced data it is a common practice to weight positive and negative classes differently, i.e. $\{w_i = w_+, \, \forall i \in C^+\}$ and $\{w_i = w_-, \, \forall i \in C^-\}$ where C^+ and C^- are the corresponding sets of indices for the positive and negative classes respectively.

The function $max(0, 1 - \alpha^T y_i x_i)$ is a convex function. The weighted sum of convex functions is also convex. Therefore for a convex function $\Phi(\alpha)$ (1) is also convex. The problem in (1) can be formulated as a mathematical programming problem as follows:

$$\begin{array}{rrcl} \min\limits_{(\alpha,\xi) \in R^{d+\ell}} & \Phi(\alpha) & + & \sum_{i=1}^{\ell} w_i \xi_i \\ \text{s.t.} & \xi_i & \geq & 1 - \alpha^T y_i x_i \\ & \xi_i & \geq & 0, \, \forall i \end{array} \tag{2}$$

For $\Phi(\alpha) = \|\alpha\|_2^2$, where $\|.\|_2$ is the 2-norm, (1) results in the conventional Quadratic-Programming-SVM, and for $\Phi(\alpha) = |\alpha|$, where $|.|$ is the 1-norm it yields the sparse Linear-Programming-SVM.

2.2. Concordance via Multivariable Hinge Loss

This time each pixel is characterized by two feature vectors $x_i = (\bar{x}_{i1}, \bar{x}_{i2})$, with $\bar{x}_{ik} \in \Re^{d_k}$ for $k = \{1, 2\}$ are feature vectors extracted for the spectral and spatio-spectral representations of the pixel s_i. In this framework concordance among classifiers is imposed on the objective function in (1) by using a multi-variable hinge loss function as follows.

We aim to optimize the following cost function

$$\mathcal{J}(\alpha_1, \alpha_2) = \sum_{k=1}^{2} \Phi_k(\alpha_k) + \sum_{i=1}^{\ell} w_i \max\left(0, e_{i1}, e_{i2}\right) \tag{3}$$

where $e_{ik} = 1 - \alpha_k^T y_i \bar{x}_{ik}$ defines the margin error due to the k^{th} representation of pixel s_i committed by the classifier f_k. The loss induced by pixel s_i is zero only if $\forall k : 1 - \alpha_k^T y_i \bar{x}_{ik} \leq 0$, i.e. margin error for both classifiers are zero. In other words before a pixel is assigned to one of the classes concordance among classifier outputs is sought. Classifiers are considered concordant for pixel s_i when they all have zero margin error on s_i.

Multi-variable hinge loss functions of this form were used earlier in [2], [3]. In [2] an offline-training algorithm for a cascaded classifier was proposed. Multi variable hingle-loss was imposed on the objective function to ensure that positive samples are correctly classified by all of the sub-classifiers in the cascade. In [3] a similar form of the multivariable hinge loss is used to train a polyhedral classifier when dealing with the multi-modality nature of the negative samples in a computer-aided detection application for colorectal cancer.

The function $\max\left(0, e_{i1}, e_{i2}\right)$ is a convex function. The weighted sum of convex functions is also convex. Therefore for a convex function $\Phi_k(\alpha_k)$ (1) is also convex. We formulate this problem as a mathematical programming problem with inequality constraints and solve for α_1 and α_2.

The problem (3) can be formulated as follows

$$\begin{array}{rrcl} \min\limits_{(\alpha,\xi) \in R^{Kd+\ell}} & \sum_{k=1}^{K} \Phi_k(\alpha_k) & + & \sum_{i==1}^{\ell} w_i \xi_i \\ \text{s.t.} & \xi_i & \geq & e_{i1} \\ & \xi_i & \geq & e_{i2} \\ & \xi_i & \geq & 0 \end{array} \tag{4}$$

for $i = 1, ..., \ell$. Note that for a convex function $\Phi(\alpha)$ the problem in (4) is convex. In a nutshell we designed two classifiers, one for each of the different representations and construct a learning algorithm to jointly optimize these classifiers such that the cost induced by a sample is zero if and only if both classifiers classifies this sample correctly, i.e. $\forall k : 1 - \alpha_k^T \bar{x}_{ik} \leq 0$.

3. EXPERIMENTAL RESULTS

The hyperspectral imagery used in this study is collected by the airborne HYMAP system on September 30, 1999 over the Purdue University West Lafayette campus. It contains 126 bands covering 0.40-2.40 μm region of the spectrum. Pixel size is about 5 meters. The spatio-spectral representation for each pixel is obtained by averaging out the spectral values of the pixels over the $3x3$ neighborhood region.

Three different training approaches for the training of the spectral and spatio-spectral classifiers are compared. First, classifiers are trained independently. Second, feature vectors for each of the spectral and spatio-spectral representations are concatenated to train a single classifier. Third, the two classifiers are trained jointly with concordance imposed on the objective function. Varying numbers of training sample sizes are considered. Classifier paramaters w_+ and w_- are tuned by 10-fold cross validation with the training data. For each experiment the Receiver Operating Characteristics (ROC) curves are plotted on the test data and areas under the curves are recorded. $\Phi_k(\alpha_k)$ is chosen as the 1-norm regularizer, i.e. $\Phi_k(\alpha_k) = |\alpha_k|$. Apart from regularizing the classifier coefficients, one norm regularizer yields sparse solutions and acts as a built-in feature selection algorithm.

sequence order	1	2	3	4	5
ℓ	5/44	10/88	20/176	20/410	50/440
concatenation	0.82	0.81	0.91	0.93	0.95
independent	0.74	0.86	0.93	0.96	0.96
concordance	0.86	0.89	0.94	0.96	0.96

Table 1. Areas under the ROC curves obtained for three different training approaches. ℓ indicates the number of training samples used for each experiment displayed in the following format: *(number of samples from the target class)/(total number of training samples)*

Results favor concordance approach for smaller sample sizes over the others. For larger sample sizes we don't observe any statistically significant difference among the three approaches. It is well known that, when there is limited number of training samples available and the dimensionality is high, classifiers suffer from the *curse of dimensionality* [4]. However, intuitively speaking, this problem can be alleviated, by training classifiers jointly to achieve concordance among them. Even though further verification is necessary, imposing concordance among the classifiers outputs seems to act as an effective regularizer over the classifier coefficients.

4. CONCLUSIONS

In this study, mainly motivated by a challenging target detection problem, we define a new machine learning concept, called *concordance learning*. Target detection involves a critical decision-making process. A prediction error might have serious consequences. We argue that for target detection problems where multiple representations of the objects are available, verifying the target through each representation independently increases the confidence in the final decision being made. We further argue that in order to exploit correlations between different representations and maximize concordance among the outputs of the classifiers, a joint training scheme involving all classifiers is required. We propose a solution based on linear hyperplanes and validate our approach using a hyperspectral dataset with cars in the parking lots as our targets. Although further validation with other datasets and possibly in different domains are necessary, we believe the preliminary results obtained are intriguing.

5. REFERENCES

[1] Avrim Blum and Tom Mitchell, "Combining labeled and unlabeled data with co-training," in *COLT' 98: Proceedings of the eleventh annual conference on Computational learning theory*, New York, NY, USA, 1998, pp. 92–100, ACM.

[2] Murat Dundar and Jinbo Bi, "Joint optimization of cascaded classifiers for computer aided detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–8.

[3] Murat Dundar, Matthias Wolf, Sarang Lakare, Marcos Salganicoff, and Vikas Raykar, "Polyhedral classifiers for target detection: A case study: Colorectal cancer," in *Proceedings of ICML'08*, Helsinki, Finland, 2008.

[4] R E Bellman, *Dynamic Programming*, Princeton University Press, 1957.