

Learning Using Group Information

Vladimir Cherkassky and Feng Cai
Department of Electrical & Computer Engineering
University of Minnesota, Minneapolis MN 55455
Email cherk001@umn.edu, caixx043@umn.edu

Abstract:

Sparse heterogeneous data is common in many machine learning applications. For example, estimation of predictive (diagnostic) models using patients' data from clinical studies requires effective integration of genetic, clinical and demographic data. Typically all heterogeneous inputs are properly encoded and mapped onto a single feature vector, used for estimating (training) a predictive model. This approach, known as standard inductive learning, is commonly used in application studies. More recently, several new learning methodologies have emerged. In particular, when training data can be naturally separated into *several groups* (or *structured*), we can view model estimation for each group as a separate task, leading to Multi-Task Learning (MTL) framework. Similar setting where training data is structured, but the goal is to estimate a single predictive model (for all groups), leads to Learning with Structured Data and SVM+ methodology recently proposed by Vapnik [1].

We review several approaches for learning using group information in the training data, currently under investigation, such as SVM+ and the new methodology for MTL based on SVM+ technology [2]. Implementations for classification supervised-learning problems will be presented. Advantages and limitations of these new data modeling methods (relative to standard inductive SVM) are demonstrated via empirical comparisons using several biomedical data sets.

New approaches for modeling heterogeneous data:

Suppose that training data are the union of $t > 1$ groups. Let us denote the indices of samples from group r by $T_r = \{r_{n1}, \dots, r_{nr}\}$, $r = 1, \dots, t$. Then the total training data set is a union of t groups: $\{\{\mathbf{X}_r, \mathbf{Y}_r\}, r = 1, \dots, t\}$, $\{\mathbf{X}_r, \mathbf{Y}_r\} = \{\{\mathbf{x}_{r_{n1}}, y_{r_{n1}}\}, \dots, \{\mathbf{x}_{r_{nr}}, y_{r_{nr}}\}\}$. However, the group label of test data is unknown. Vapnik [1] proposed to map the input vectors $\mathbf{x}_i, i \in T_r$ simultaneously into two different Hilbert spaces: into the decision space Z ($\mathbf{z}_i = \Phi_z(\mathbf{x}_i) \in Z$) which defines the decision function and into correcting space Z_r ($\mathbf{z}_i^r = \Phi_{z_r}(\mathbf{x}_i) \in Z_r$) which defines the set of correcting functions for a given group r . The correcting functions are specified as: $\xi^r(\mathbf{x}_i) = (\mathbf{x}_i, \mathbf{w}_r) + d_r, r = \{1, \dots, t\}$. The goal is to find the decision function in decision space Z , $f(\mathbf{x}) = (\mathbf{w}, \Phi_z(\mathbf{x})) + b$. SVM+ tries to solve the following optimization problem:

$$\min_{w, w_1, \dots, w_t, b, d_1, \dots, d_t} \frac{1}{2} (\mathbf{w}, \mathbf{w}) + \frac{\gamma}{2} \sum_{r=1}^t (\mathbf{w}_r, \mathbf{w}_r) + C \sum_{r=1}^t \sum_{i \in T_r} \xi_i^r \quad (\text{OP1})$$

subject to:

$$y_i((\mathbf{w}, \mathbf{z}_i) + b) \geq 1 - \xi_i^r, i \in T_r, r = 1, \dots, t$$

$$\xi_i^r \geq 0, i \in T_r, r = 1, \dots, t$$

$$\xi_i^r = (\mathbf{z}_i^r, \mathbf{w}_r) + d_r, i \in T_r, r = 1, \dots, t$$

When the group label of test data is known, we have Multi-Task Learning method called SVM+MTL[2]. Similar to SVM+, we map the input vectors $\mathbf{x}_i, i \in T_r$ simultaneously into two different Hilbert spaces: into the decision space $Z (\mathbf{z}_i = \Phi_z(\mathbf{x}_i) \in Z)$ and into correcting space $Z_r (\mathbf{z}_i^r = \Phi_{z_r}(\mathbf{x}_i) \in Z_r)$ for a given group r . The goal is to find the t decision functions $f_r(\mathbf{x}) = (\Phi_z(\mathbf{x}), \mathbf{w}) + b + (\Phi_{z_r}(\mathbf{x}), \mathbf{w}_r) + d_r, r = 1, \dots, t$. SVM+MTL solves the following optimization problem:

$$\min_{\mathbf{w}, b} \frac{1}{2} (\mathbf{w}, \mathbf{w}) + \frac{\gamma}{2} \sum_{r=1}^t (\mathbf{w}_r, \mathbf{w}_r) + C \sum_{r=1}^t \sum_{i \in T_r} \xi_i^r \quad (\text{OP2})$$

subject to:

$$y_i^r ((\mathbf{w}, \mathbf{z}_i) + b + (\mathbf{w}_r, \mathbf{z}_i^r) + d_r) \geq 1 - \xi_i^r, i \in T_r, r = 1, \dots, t$$

$$\xi_i^r \geq 0, i \in T_r, r = 1, \dots, t$$

Empirical Comparisons:

This section describes empirical comparisons of various modeling approaches for classification with heterogeneous biomedical data, such as single SVM (sSVM), multiple independent SVM (mSVM), SVM+ and SVM+MTL. All comparisons use linear and RBF kernels for sSVM and mSVM, Common decision space for SVM+ and SVM+MTL use linear kernel while the unique correction space uses RBF kernel. These modeling methods are compared using several data sets from UCI repository. For SVM+ and SVM+MTL, each dataset was separated into several groups by some variable.

Table 1: Test error rate for various learning approaches

	sSVM(linear)	sSVM(rbf)	SVM+	mSVM	mSVM(rbf)	SVM+MTL
Statlog heart	19.3 ± 7.5	18.2 ± 6.5	16.3 ± 6.1	16.6 ± 4.3	21.5 ± 5.3	15.2 ± 4.0
Ljubljana breast cancer	29.3 ± 6.2	25.7 ± 4.5	24.9 ± 4.8	29.6 ± 1.6	24.2 ± 2.5	23.5 ± 3.4
Wisconsin breast cancer	3.4 ± 1.3	3.8 ± 0.8	3.1 ± 1.0	3.4 ± 0.8	3.1 ± 1.0	2.9 ± 0.9
Hepatitis	16.3 ± 8.4	17.5 ± 5.2	16.3 ± 8.4	16.3 ± 8.4	16.3 ± 8.4	15.0 ± 7.1

Table 1 shows comparisons for standard inductive SVM, multiple SVMs, SVM+ and SVM+MTL. These comparisons indicate that for the single-model setting, SVM+ is consistently better than standard SVM classifier, and that for multiple-model setting, SVM+MTL is consistently better than several independent SVMs.

Acknowledgement: This work was supported by the BICB grant from the University of Minnesota, Rochester.

References

1. Vapnik, V. Empirical Inference Science Afterword of 2006, Springer, 2006
2. Liang, L. and Cherkassky, V. Connection between SVM+ and Multi-Task Learning, Proc. IJCNN, 2008

Topic: Learning algorithms

Preference: Any