

Multi-Task Support Vector Machine for Data Classification

Yunyan Song¹ and Wenxin Zhu²

¹College of Science, Tianjin University of Technology, Tianjin, 300384, China

²College of Basic Science, Tianjin Agricultural University, Tianjin, 300384, China
songyunyanlg@163.com¹, zhuwenxinyan@163.com²

Abstract

Multi-task Learning (MTL) algorithms aim to improve the performance of several learning methods through shared information among all tasks. One particularly successful instance of multi-task learning is its adaptation to support vector machine (SVM). Recently advances in large-margin learning have shown that their solutions may be misled by the spread of data and preferentially separate classes along large spread directions. In this paper, we propose a novel formulation for multi-task learning by extending the recently published relative margin machine algorithm to the multi-task learning paradigm. The new method is an extension of support vector machine for single task learning. The objective of our algorithm is to obtain a different predictor for each task while taking into account the fact that the tasks are related as well as the spread of the data. We test the proposed method experimentally using real data. The experiments show that the proposed method performs better than existing multi-task learning with SVM and single-task learning with SVM.

Keywords: Support vector machine, Multi-task learning, Kernel methods, Relative margin

1. Introduction

The aim of classification problems is to learn a classifier that generalizes well on future or testing data from a finite training data set. SVM and maximum margin classifiers [1-3] have been a particularly successful approach both in theory and in practice [4-6]. In many practical problems a number of prediction models need to be estimated from data. Multi-task Learning (MTL) algorithms [7] aim to improve the performance of several learning methods through shared information among all tasks. The objective of MTL is to obtain a different predictor for each task while taking into account the fact that the tasks are related. If the different problems are sufficiently related, MTL can lead to better generalization and benefit all of the tasks. Note that tasks can be considered as mutually related if the data distributions are similar enough and likely refer to the same problem. There has been extensive research on MTL in the past decade starting with the early works of [7-8]. Multi-task learning has been explored by various authors in different frameworks. This has led to multi-task learning using kernel-methods [9-12], probabilistic approaches [13-15], maximum-entropy discrimination [16], large margin multi-task metric [17]. One particularly successful instance of multi-task learning is its adaptation to support vector machine [11-12] and [18]. As mentioned above, support vector machines are arguably amongst the most successful classification algorithms of all times. These multi-task learning formulations based on support vector machines aim to deal with the problem of considering different task relations. However, the multi-task learning adaptation of support vector machine regardless of how spread of the data. Their solutions may be

misled by the spread of data and preferentially separate classes along large spread directions. This is a serious limitation when looking for the different predictor for each task.

Recently advances in large-margin learning have shown that the generalization ability of these learners can be improved by utilizing second order information, such as second order perceptron [19], Gaussian margin machines [20], confidence weighted learning [21], and adaptive regularization of weights (AROW) [22-23]. Recently, Shivaswamy and Jebara [24] proposed an effective and less computationally expensive way to incorporate the spread of the data—second order information about the distance between hypotheses when projected onto the line defined by the weight vector w . The method is called Relative Margin Machine (RMM). RMM which can deal with the presence of arbitrary affine transformations was introduced in [25]. In that work, the distance of the data from the separating hyperplane is bounded from above by a scalar R . RMM maximizes in this way the relative margin (relative to that upper bound) of the data from the separating hyperplane. The motivation behind this line of research is the fact that large margin on its own is not a meaningful quantity, a better way to measure margin is in relation to the spread of the data. RMM has shown significant improvements over the large margin counterparts on real-world problems. Recently, Zhu and Zhong proposed an improved RMM based on additional information hidden in the data [26], and obtained surprising results. V. Eidelman presented an online gradient-based algorithm for relative margin maximization, which bounds the spread of the projected data while maximizing the margin [27]. Some other extensions to RMM include structured RMM [28], and so on. All of these suggest that the RMM and maximum relative margin approach obtained significant improvements in classification accuracy.

In this paper, we propose a novel formulation for multi-task learning by extending the recently published relative margin machine algorithm to the multi-task learning paradigm. The new method is an extension of support vector machine for single task learning. The objective of our algorithm is to obtain a different predictor for each task while taking into account the fact that the tasks are related as well as the spread of the data. We test the proposed method experimentally using real data. The experiments show that the proposed method perform better than existing multi- task learning with SVM and single-task learning with SVM.

The paper is organized as follows. Section 2 briefly focuses on the relative margin support vector machine. We propose a novel multi-task learning approach (relative margin multitask learning, RMMTL) in Section 3. The experiments of the RMMTL on benchmark datasets are described in Section 4. Moreover, we compare it with the single SVM (sSVM), multiple SVM (mSVM) and multi-task learning (MTL). Section 5 concludes this paper.

2. Relative Margin Machine

In [25], the SVM was modified such that the projections on the training examples remain bounded. The new method was called Relative Margin Machine (RMM). RMM was introduced to deal with a possible bad scaling of the data across the different dimensions within the large margin classification framework. The method proposes finding the large margin separating hyperplane such that the distance of the data from it is bounded from above by a scalar R . In this way, a large relative margin, rather than the absolute (classic) large margin solution is reached. One important aspect is that RMM might be advantageous in comparison to SVM in cases of drifts in particular directions in feature space. Data drifts in applications are quite common, *e.g.*, drifts in sensor data due to noise or spatial shifts. The optimization problem in RMM is defined as

$$\begin{aligned}
 \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \\
 \text{s.t.} \quad & y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad i = 1, \dots, N. \\
 & \frac{1}{2} (w \cdot \phi(x_i) + b)^2 \leq \frac{R^2}{2} \quad i = 1, \dots, N.
 \end{aligned} \tag{1}$$

This formulation is similar to the SVM primal except for the additional constraints $\frac{1}{2} (w \cdot \phi(x_i) + b)^2 \leq \frac{R^2}{2}$. The formulation has one extra parameter R in addition to the SVM parameter C . When R is large enough, the above RMM gives the same solution as the SVM. Also note that only settings of $R > 1$ are meaningful since a value of R less than one would prevent any training examples from clearing the margin. As R is decreased, the RMM solution increasingly differs from the SVM solution. Specifically, with a smaller R , the RMM still finds a large margin solution but with a smaller projection of the training examples. Relative margin machine was shown to outperform the support vector machine on several datasets (both synthetic and real world) [25].

3. Relative Margin Multi-Task Learning (RMMTL)

This section introduces the proposed formulations for multi-task SVM data classification. The proposed formulations combine relative margin SVM classification with the multi-task learning paradigm. Suppose we have t learning tasks. Let us denote the indices from task r by $T_r = \{i_{n_1}, \dots, i_{n_r}\}, r = 1, \dots, t$. Then all training samples can be represented as:

$$\left\{ \{X_r, Y_r\}, r = 1, \dots, t \right\}, \{X_r, Y_r\} = \left\{ \{x_{n_1}, y_{n_1}\}, \dots, \{x_{n_r}, y_{n_r}\} \right\}.$$

We follow the intuition of Hierarchical Bates [29, 30, 31]. In particular we assume that all w_r can be written, for every $r \in \{1, \dots, t\}$, as $w_r = w + v_r$, where the vectors v_r are "small" when the different tasks are similar to each other. We then estimate all v_r as well as the w simultaneously.

3.1. Linear Relative Margin Multi-Task Learning

Formally, the proposed linear RMMTL method needs to solve the following optimization problem:

$$\min \quad \frac{1}{2} w^T w + \frac{\lambda}{2t} \sum_{r=1}^t v_r^T v_r + C \sum_{i \in T_r} \xi_i^r \tag{2}$$

$$\begin{aligned}
 \text{s.t.} \quad & y_i^r (w^T x_i + b + v_r^T x_i^r + d_r) \geq 1 - \xi_i^r, \quad r = 1, \dots, t, \quad i \in T_r \\
 & w^T x_i + b + v_r^T x_i^r + d_r \leq R, \quad r = 1, \dots, t, \quad i \in T_r \\
 & -(w^T x_i + b + v_r^T x_i^r + d_r) \leq R, \quad r = 1, \dots, t, i \in T_r \\
 & \xi_i^r \geq 0, \quad r = 1, \dots, t, i \in T_r
 \end{aligned} \tag{3}$$

In this problem, λ is positive regularization parameters and ξ_i^r are slack variables measuring the error that each of the final models w_r makes on the data. The last two linear constraints were added such that the projections on the training examples remain bounded. When R is large enough, the above model gives the similar solution as the

MTL. In order to solve this problem, we construct the Lagrangian function for problem (2)-(3)

$$\begin{aligned}
 L = & \frac{1}{2} w^T w + \frac{\lambda}{2t} \sum_{r=1}^t v_r^T v_r + C \sum_{i \in T_r} \xi_i^r - \sum_{r=1}^t \sum_{i \in T_r} \alpha_i^r [y_i^r (w^T x_i + b + v_r^T x_i^r + d_r) - 1 + \xi_i^r] \\
 & + \sum_{r=1}^t \sum_{i \in T_r} \beta_i^r [w^T x_i + b + v_r^T x_i^r + d_r - R] - \sum_{r=1}^t \sum_{i \in T_r} \mu_i^r [w^T x_i + b + v_r^T x_i^r + d_r + R] \\
 & - \sum_{r=1}^t \sum_{i \in T_r} \gamma_i^r \xi_i^r
 \end{aligned} \quad (4)$$

Minimize it with respect to $w_r, v_r, b, d_r, \xi_i^r$ and maximize with respect to Lagrange multipliers α, β, μ . According to the KKT necessary and sufficient optimality conditions, we differentiate the primal variables w, v_r, b and d_r , equate them to zero for L resulting in

$$\frac{\partial L}{\partial w} = 0 \Rightarrow w = \sum_{r=1}^t \sum_{i \in T_r} (\alpha_i^r y_i^r - \beta_i^r + \mu_i^r) x_i \quad (5)$$

$$\frac{\partial L}{\partial v_r} = 0 \Rightarrow v_r = \frac{t}{\lambda} \sum_{i \in T_r} (\alpha_i^r y_i^r - \beta_i^r + \mu_i^r) x_i \quad (6)$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow w = \sum_{r=1}^t \sum_{i \in T_r} (\alpha_i^r y_i^r - \beta_i^r + \mu_i^r) = 0 \quad (7)$$

$$\frac{\partial L}{\partial d_r} = 0 \Rightarrow \sum_{i \in T_r} (\alpha_i^r y_i^r - \beta_i^r + \mu_i^r) = 0 \quad (8)$$

$$\frac{\partial L}{\partial \xi_i^r} = 0 \Rightarrow C - \alpha_i^r - \gamma_i^r = 0 \quad (9)$$

The dual problem of (2)-(3) is given by:

$$\min \frac{1}{2} \sum_{s=1}^t \sum_{j \in T_s} \sum_{r=1}^t \sum_{i \in T_r} (\alpha_i^r y_i^r - \beta_i^r + \mu_i^r) (\alpha_j^s y_j^s - \beta_j^s + \mu_j^s) (x_i^r \cdot x_j^s) \quad (10)$$

$$\begin{aligned}
 & + \frac{t}{2\lambda} \sum_{r=1}^t \sum_{i \in T_r} \sum_{j \in T_r} (\alpha_i^r y_i^r - \beta_i^r + \mu_i^r) (\alpha_j^r y_j^r - \beta_j^r + \mu_j^r) \\
 & + R * \sum_{r=1}^t \sum_{i \in T_r} (\beta_i^r + \mu_i^r) - \sum_{r=1}^t \sum_{i \in T_r} \alpha_i^r \\
 \text{s.t.} \quad & \sum_{i \in T_r} (\alpha_i^r y_i^r - \beta_i^r + \mu_i^r) = 0, \quad r = 1, \dots, t \\
 & 0 \leq \alpha_i^r \leq C, \quad r = 1, \dots, t, \quad i \in T_r \\
 & \alpha_i, \beta_i, \mu_i \geq 0
 \end{aligned} \quad (11)$$

Let the solution of (10)-(11) be $\alpha^*, \beta^*, \mu^*, b^*$ and d_r^* . Then the corresponding decision function for task r is

$$f_r(x) = w^T x + b + v_r^T x + d_r \quad (12)$$

$$= \sum_{r=1}^t \sum_{i \in T_r} (\alpha_i^{*r} y_i^r - \beta_i^{*r} + \mu_i^{*r}) k(x_i, x) + b^* + \frac{t}{\lambda} \sum_{i \in T_r} (\alpha_i^{*r} y_i^r - \beta_i^{*r} + \mu_i^{*r}) k(x_i^r, x) + d_r^* \quad (13)$$

3.2. Nonlinear Relative Margin Multi-Task Learning

We generalize the linear multi-task learning method outlined above to non-linear case using kernels as is done for SVM.

$$\begin{aligned} \min \quad & \frac{1}{2} w^T w + \frac{\lambda}{2t} \sum_{r=1}^t v_r^T v_r + C \sum_{i \in T_r} \xi_i^r \\ \text{s.t.} \quad & y_i^r (w^T z_i + b + v_r^T z_i^r + d_r) \geq 1 - \xi_i^r, \quad r = 1, \dots, t, \quad i \in T_r \\ & w^T z_i + b + v_r^T z_i^r + d_r \leq R, \quad r = 1, \dots, t, \quad i \in T_r \\ & -(w^T z_i + b + v_r^T z_i^r + d_r) \leq R, \quad r = 1, \dots, t, i \in T_r \\ & \xi_i^r \geq 0, \quad r = 1, \dots, t, i \in T_r \end{aligned} \quad (14)$$

In this problem $z_i = \phi(x_i)$, $z_i^r = \phi(x_i^r)$, where $k(x_i, x_j) = (\phi(x_i), \phi(x_j))$ is kernel function, and $\phi: X \rightarrow F$ is a nonlinear mapping transforming the examples in the input space into the feature space. It is not hard to derive the separating hyperplane of nonlinear RMMTL:

$$f_r(x) = w^T \phi(x) + b + v_r^T \phi(x) + d_r \quad (15)$$

$$= \sum_{r=1}^t \sum_{i \in T_r} (\alpha_i^{*r} y_i^r - \beta_i^{*r} + \mu_i^{*r}) k(x_i, x) + b^* + \frac{t}{\lambda} \sum_{i \in T_r} (\alpha_i^{*r} y_i^r - \beta_i^{*r} + \mu_i^{*r}) k(x_i^r, x) + d_r^* \quad (16)$$

4. Experiments

In order to examine the generalization performance of the proposed RMMTL, we conduct experiments on several datasets, including Dermatology dataset (<http://www.ics.uci.edu/mlearn/MLRespository.html>), Isolet spoken alphabet recognition and Monk taken from the UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/>), and Radar Landmine detection data (<http://www.ee.duke.edu/lcarin/LandmineData.zip>). For Dermatology dataset, isolet spoken alphabet recognition dataset and Monk dataset, we compare RMMTL with the single SVM (sSVM), multiple SVM (mSVM) and multi-task learning (MTL). In addition, in order to explore the effect of the R parameter, the average test error rate were computed across different R. For radar landmine detection dataset, we compare RMMTL with sSVM, mSVM and MTL. The linear kernel and Gaussian kernel functions are adopted in the experiments.

4.1. Dermatology Dataset

This data set consists of 366 differential diagnosis of erythemato squamous in dermatology. The goal is to diagnose one of six dermatological diseases (psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis, and pityriasis rubra pilaris) based on 33 clinical and histopathological attributes. That is to say, this is a multi-class (6-class) problem. As in [10], we convert this problem to 6 binary one-versus-rest classification problems, each of which is considered to be "one task". Hence we have six tasks in total. This dataset is divided into ten random splits of 200 training and 166 testing samples.

4.2. Isolet Dataset

The Isolet dataset is collected from 150 subjects speaking each letter of the alphabet twice. Hence, we have 52 training examples from each speaker. Due to the lack of three examples, there are 7797 examples in total. These speakers are grouped into five sets of 30 speakers each. These groups are referred to as isolet1-isolet5. Each of these datasets has 26 classes. We treat each of the subsets as its own classification task. Therefore, there are five tasks that are highly related with each other because they are taken from the same utterances. They are different from each other because they come from different groups that vary largely in the way of speaking the English alphabets. The attribute information include spectral coefficients, contour features, sonorant features, pre-sonorant features and post-sonorant features. To remove low variance noise and to speed up computation time we preprocess the Isolet data with PCA and project it onto its leading principal components that capture 95% the dimensionality from 617 to 169.

4.3. Monk Dataset

The Monk dataset is the basis of a first international comparison of learning algorithms. There are a total of three problems corresponding to three tasks. There are 432 instances, each of which has 7 attributes.

4.4. Radar Landmine Detection Dataset

In this set of experiments, we evaluate the performance of the proposed method in a landmine detection problem. This dataset is collected from 29 different landmine fields (groups). Each sample in each of the groups consists of a 9 dimensional vector extracted from airborne radar images. A binary label is assigned to each sample, indicating landmine presence or clutter. We note that groups are not uniformly related. Actually, groups 1 to 15 correspond to regions highly foliated, whilst the ones from 16 to 29 correspond to deserted areas [18].

Data from 4 different groups (tasks 1, 15, 20, 25) are selected and number of samples in each group is varied from 40 to 120. For each group, one quarter of training samples correspond to presence, and the rest to clutters. For estimating the prediction error of a learning method, we use fivefold cross validation. In the experiment, samples from one landmine field are considered one task.

Tables 1-3 report the testing results of sSVM, mSVM, MTL and the proposed RMMTL on dermatology dataset, isolet dataset and monk dataset. We observe that the results of RMMTL under linear and gaussian kernels are obviously better than sSVM, mSVM and MTL on all datasets, which show that the proposed algorithm achieves better generalization performance. Moreover, to explore the effect of the R parameter, the test error rate were computed. Figures 1-3 report the results on all datasets. For Dermatology dataset, the error rate decreases as R decreases from the right. For isolet dataset and monk dataset, As the R value is decreased, the error rate decreases to a reasonably wide minimum before starting to increase.

Figure.4 reports the average test error rate of sSVM, mSVM, MTL and RMMTL with linear kernel and gaussian kernel. Because this is an unbalanced dataset, we measured the performance of the different methods as a function of the percentage of samples from the minority class in the training set. Hence, we have considered a fixed training size of 200 samples and 10, 15, 20 and 25% of samples from the positive class. We can find that the proposed RMMTL has advantages over these three methods both in the linear and nonlinear case.

Table 1. Dermatology Dataset

Method	sSVM	mSVM	MTL	RMMTL
Accuracy (linear)	0.8167 ± 0.054	0.8275 ± 0.054	0.8563 ± 0.054	0.9367 ± 0.046
Accuracy (nonlinear)	0.8250 ± 0.054	0.8275 ± 0.054	0.8668 ± 0.054	0.9600 ± 0.034

Table 2. ISOLET Dataset

Method	sSVM	mSVM	MTL	RMMTL
Accuracy (linear)	0.9575 ± 0.026	0.9425 ± 0.024	0.7550 ± 0.069	0.9750 ± 0.015
Accuracy (nonlinear)	0.8800 ± 0.207	0.8350 ± 0.078	0.8475 ± 0.027	0.9175 ± 0.02

Table 3. MONK Dataset

Method	sSVM	mSVM	MTL	RMMTL
Accuracy (linear)	0.6317 ± 0.054	0.7417 ± 0.048	0.5850 ± 0.045	0.7967 ± 0.027
Accuracy (nonlinear)	0.5517 ± 0.014	0.5717 ± 0.052	0.6017 ± 0.039	0.7733 ± 0.016

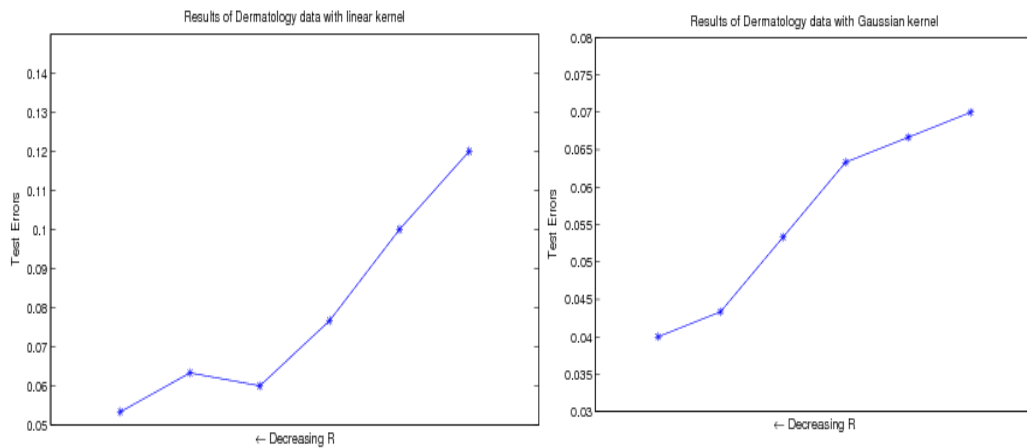


Figure 1. Behavior on the Dermatology Dataset: Linear Case and Nonlinear Case

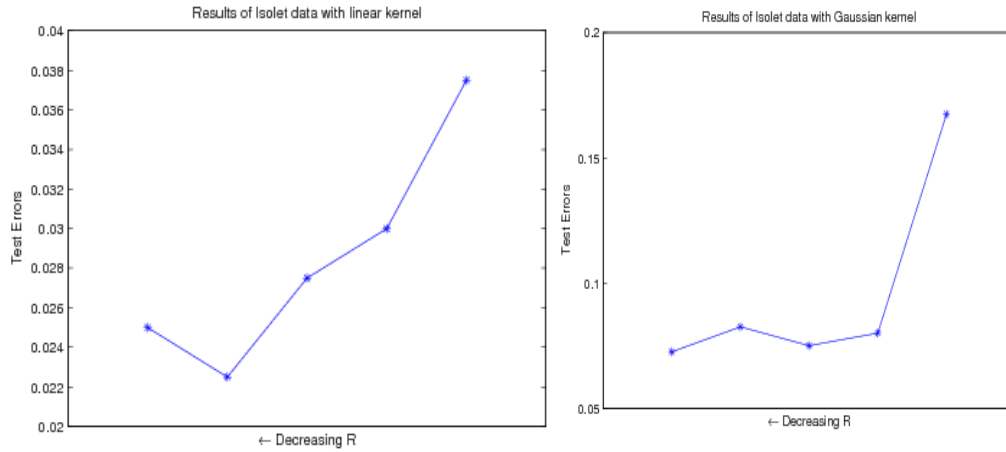


Figure 2. Performance on the Isolet Spoken Alphabet Recognition Dataset: Linear Case and Nonlinear Case

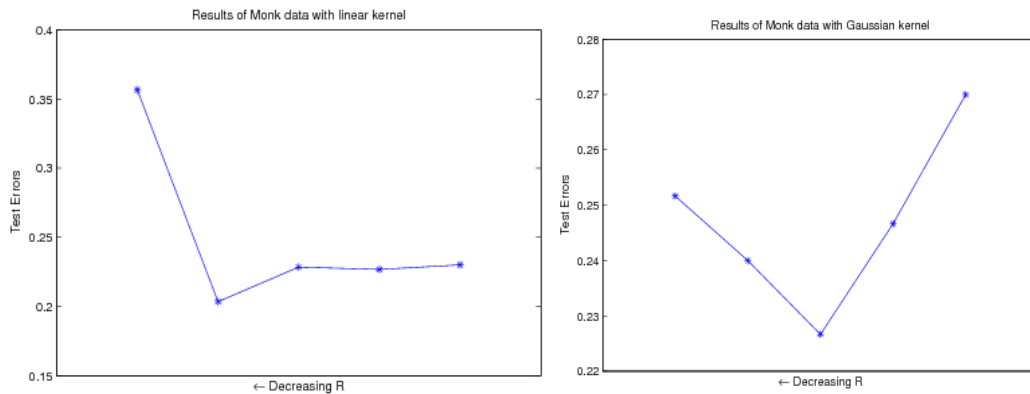


Figure 3. Performance on the Monk Dataset: Linear Case and Nonlinear Case

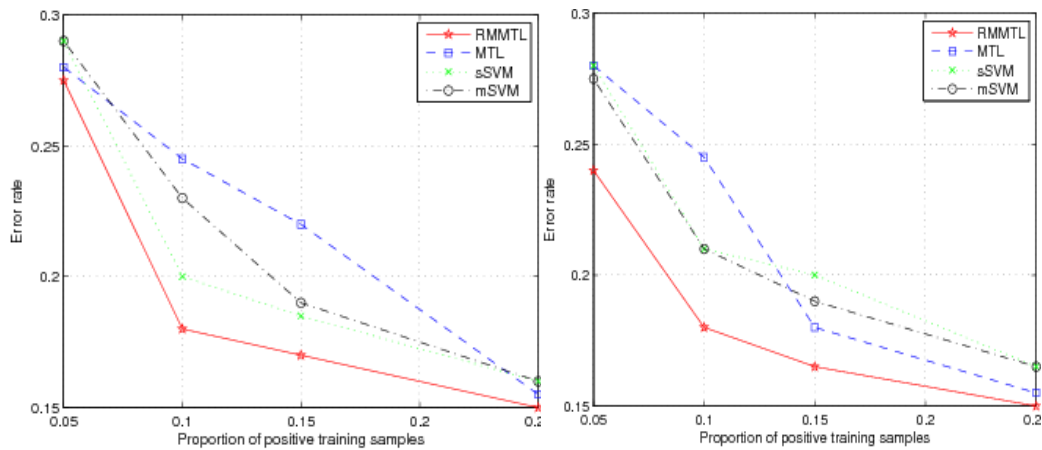


Figure 4. Radar Landmine Dataset. Results of Methods Under Linear Kernel (Left) and Nonlinear Case (Right)

5. Conclusions and Discussion

In this paper, we propose a novel formulation for multi-task learning by extending the recently published relative margin machine algorithm to the multi-task learning paradigm. The new method is an extension of support vector machine for single task learning. The objective of our algorithm is to obtain a different predictor for each task while taking into account the fact that the tasks are related as well as the spread of the data. We test the proposed method experimentally using real data. The experiments show that the proposed method performs better than the existing multi-task learning with SVM, relative margin support vector machine and single-task learning with SVM.

Acknowledgments

The work is supported by the National Science Foundation of China (Grant No. 11171346) and Tianjin University of Technology Young Teachers' Scientific Research Fund (No. LGYM201235).

References

- [1] V. Vapnik, "The Nature of Statistical Learning Theory", Springer-Verlag, New York, (1995).
- [2] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers", Neural Processing Letters, vol. 9, (1999), pp. 293-300.
- [3] G. Fung and O. Mangasarian, "Proximal support vector machines", Proc. of the 7th International Conference Knowledge Discovery and Data Mining, (2001), pp. 77-86.
- [4] Y. H. Zhao and P. Zhong, "A feature selection method for twin support vector regression", ICIC Express Letters, Part B: Applications, vol. 3, (2012), pp. 91-98.
- [5] Z. Y. Zhao, P. Zhong and Y. H. Zhao, "Reduced least squares one-class SVM in empirical feature space for imbalanced data", ICIC Express Letters, vol. 11, (2011), pp. 4115-4121.
- [6] S. S. Liang and D. J. He, "Image classification using compound image transformations, multi-class SVM", ICIC Express Letters, vol. 3, (2012), pp. 685-691.
- [7] R. Caruana, "Multi-task learning". Machine Learning, vol. 28, (1997), pp. 41-75.
- [8] S. Thrun, "Is learning the n-th thing any easier than learning the first?", Advances in neural information processing systems, (1996), pp. 640-646.
- [9] A. Evgeniou and M. Pontil, "Multi-task feature learning", Advances in neural information processing systems, vol. 19, no. 41, (2007).
- [10] A. Argyriou, T. Evgeniou and M. Pontil, "Convex multi-task feature learning", Machine Learning, vol. 73, (2008), pp. 243-272.
- [11] T. Evgeniou and M. Pontil, "Regularized multi-task learning", Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, (2004), pp. 109-117.
- [12] T. Evgeniou and C. A. Micchelli and M. Pontil, "Learning multiple tasks with kernel methods", Journal of Machine Learning Research, (2005), pp. 615-637.
- [13] B. Bakker and T. Heskes, "Task clustering and gating for bayesian multitask learning", The Journal of Machine Learning Research, vol. 4, (2003), pp. 83-99.
- [14] Y. Xue, X. Liao, L. Carin, "Multi-task learning for classification with Dirichlet process priors", The Journal of Machine Learning Research, vol. 8, (2007), pp. 35-63.
- [15] K. Yu, V. Tresp and A. Schwaighofer, "Learning Gaussian processes from multiple tasks", Proceedings of the 22nd international conference on Machine learning, ACM, (2005), pp. 1012-1019.
- [16] T. Jebara, "Multi-task feature and kernel selection for SVMs", Proceedings of the twenty-first international conference on Machine learning, ACM, vol. 55, (2004).
- [17] S. Parameswaran and K. Q. Weinberger, "Large margin multi-task metric learning", Advances in neural information processing systems", (2010), pp. 1867-1875.
- [18] J. M. L. Murillo, L. G. Chova and G. C. Valls, "Multitask remote sensing data classification", Geoscience and Remote Sensing", IEEE Transactions on, vol. 51, (2013), pp. 151-161.
- [19] N. Cesa-Bianchi and A. Conconi, "A second-order perceptron algorithm", SIAM Journal on Computing, vol. 34, no. 3, (2005), pp. 640-668.
- [20] K. Crammer, M. Mohri and F. Pereira, "Gaussian margin machines", Journal of Machine Learning Research - Proceedings Track, vol. 5, (2009), pp. 105-112.
- [21] M. Dredze and K. Crammer, "Confidence-weighted linear classification", In Proceedings of the 25th international conference on Machine learning, ACM, (2008), pp. 264-271.
- [22] D. Chiang, "Hope and fear for discriminative training of statistical translation models", Machine Learning Research, vol.13, (2012), pp. 1159-1187.

- [23] K. Crammer, A. Kulesza and M. Dredze, "Adaptive regularization of weight vectors", In Advances in Neural Information Processing Systems, vol. 22, (2009), pp. 414-422.
- [24] P. K. Shivaswamy and T. Jebara, "Relative Margin Machines", In Advances in Neural Information Processing Systems, MIT Press, (2008), pp. 1481-1488.
- [25] P. K. Shivaswamy and T. Jebara, "Maximum relative margin and data-dependent regularization", The Journal of Machine Learning Research, vol. 11, (2010), pp. 747-788.
- [26] W. X. Zhu, K. N. Wang and P. Zhong, "Improving support vector classification by learning group information hidden in the data", ICIC Express Letters, Part B: Applications, vol. 5, no. 3, (2014), pp. 781-786.
- [27] V. Eidelman and Y. Marton and P. ResnikOnli, "The relative margin maximization for statistical machine translation", Proceedings of ACL, (2013).
- [28] P. K. Shivaswamy and T. Jebara, "Structured prediction with relative margin", Machine Learning and Applications, ICMLA'09. International Conference on, IEEE, (2009), pp. 281-287.
- [29] G. M. Allenby and P. E. Rossi, "Marketing models of consumer heterogeneity", Journal of Econometrics, vol. 89, (1999), pp. 57-58.
- [30] N. Arora, G. M. Allenby and J. L. Ginter, "A hierarchical Bayes model of primary and secondary demand", Marketing Science, vol. 17, (1998), pp. 29-44.
- [31] T. Heskes, "Empirical Bayes for learning to learn", In: Proceedings of the 17th international conference on machine learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, (2000), pp. 367-374.

Authors

Yunyan Song, received the BS degree in mathematics from Hebei Normal University in 2003. She received the MS degree in applied mathematics from the Tianjin Polytechnic University in 2006. Now, she is also a mathematical teacher of Tianjin University of Technology. Her current interests are in the area of pattern recognition, optimization method and machine learning.

Wenxin Zhu, (Corresponding author) received his Ph.D degree on mathematics in 2014 from China Agricultural University. Currently is a professor in College of Science, Tianjin Agricultural University. His research topics are: the mathematics, operational research and pattern recognition.