Multi-view Multi-task Support Vector Machine

Jiashuai Zhang[1(✉)], Yiwei He[2], and Jingjing Tang[1]

[1] School of Mathematical Sciences, University of Chinese Academy of Science,
Beijing 100049, China
`zhangjiashuai16@mails.ucas.ac.cn`
[2] School of Computer and Control Engineering,
University of Chinese Academy of Science, Beijing 101408, China

Abstract. Multi-view Multi-task (MVMT) Learning, a novel learning paradigm, can be used in extensive applications such as pattern recognition and natural language processing. Therefore, researchers come up with several methods from different perspectives including graph model, regularization techniques and feature learning. SVMs have been acknowledged as powerful tools in machine learning. However, there is no SVM-based method for MVMT learning. In order to build up an excellent MVMT learner, we extend PSVM-2V model, an excellent SVM-based learner for MVL, to the multi-task framework. Through experiments we demonstrate the effectiveness of the proposed method.

Keywords: SVM-based · MVMT learning · PSVM-2V
Regularization method

1 Introduction

With the promotion of diversified information acquisition technology, many samples are characterized in many ways, and thus there are a variety of multi-view learning theories and algorithms. Those works have already been extensively used in the practical applications such as pattern recognition [1] and natural language processing [2]. However, multi-view learning merely solves a single learning task.

In many real-world applications, problems exhibit dual-heterogeneity. To state it clearly, a single task has features due to multiple views (i.e., feature heterogeneity); different tasks are related with one another through several shared views (i.e., task heterogeneity) [3]. Confronted with this problem, neither multi-task learning nor multi-view learning is suitable to model. Aiming at settling this complex problem, a novel learning paradigm (i.e. multi-view multi-task learning, or MVMT Learning) has been proposed, which deals with multiple tasks with multi-view data.

He and Lawrence [3] firstly proposed a graph-based framework ($GraM^2$) to figure out MVMT problems. Correspondingly, an effective algorithm ($IteM^2$) was designed to solve the problem. Zhang and Huan [4] developed a regularized

method to settle MVMT learning based on co-regularization. Algorithm based on share structure to deal with multi-task multi-view learning [5]was also proposed afterwards. Besides classification problem, Zhang et al. [6] introduced a novel problem named Multi-task Multi-view Cluster Learning. In order to deal with this special cluster problem, the author presented an algorithm based on graph model to handle nonnegative data at first [6]. Then an improved algorithm [7] was introduced to solve the negative data set.

For decades, SVMs have been acknowledged as powerful tools in machine learning [8,9]. Therefore, many SVM-based algorithms have been proposed for MVL and MTL separately. Although there are several methods dealing with the MVMT learning, models based on SVM have not yet to be established. In order to make use of the excellent performance of SVM, we incorporate multi-task learning into the existing SVM-based multi-view model.

From the perspective of MVL, both consensus principle and complementarity principle are essential for MVL. While the consensus principle emphasizes the agreement among multiple distinct views, the complementary principle suggests that different views share complementary information. Most MVL algorithms achieve either consensus principle or complementary principle. However, a novel MVL model PSVM-2V under the framework of Privileged SVM satisfies both consensus and complementary through combining the LUPI and MVL [10].

In this paper, we construct a new model PSVM-2VMT by extending the PSVM-2V model to the multi-task learning framework. In a single task, we take advantage of PSVM-2V to learn from multiple distinct views; among different tasks, we add regularized terms to ensure the parameters of the same view are similar to each other. Hence, we establish a SVM-based model to solve the MVMT learning. According to the conventional solution of SVM problem, we derive the dual problem of the primal problem and then adopt the classical quadratic programming (QP) solver. We conduct experiments to demonstrate the effectiveness of our model.

To sum up, there are two main contributions of this paper. Firstly, we extend the PSVM-2V model to the multi-task learning framework. Secondly, we conduct experiments on multi-view multi-task data sets, and the results validate the effectiveness of our method.

The rest of this paper is organized as follows. In Sect. 2, we survey related work. Concrete model and corresponding optimization method are presented in Sects. 3 and 4. In Sect. 4, we carry on experiments to demonstrate the effectiveness of our model. At last, we conclude our work in Sect. 5.

2   Related Work

2.1   Multi-task Learning

Multi-task learning (MTL) is a learning paradigm with the help of other tasks to improve the generalization performance of original task [11]. Specifically, characterizing the relationships among tasks is the core of MTL.

In the early study of MTL, we assume that different tasks are closely related. Multi-Task feature learning is a classical method based on this assumption.

According to the relationship between the original feature space and learned feature space, there are two distinctive methods, i.e. feature transformation methods and feature selection methods. Multi-Task feature learning (MTFL) [12] transformed original feature space into low-dimension common feature space. Multi-Task feature selection (MTFS) [13] was the first method to select feature from the original feature space in multi-task learning by adding $l_{2,1}$ norm of the weight matrix to the objective function. There were other developments in feature selection by substituting different norms such as $l_{\infty,1}$ [14], capped-$l_{p,1}$ [15].

Besides MTFL, there were others methods brought up based on the positive relation correlation. The regularized multi-task support vector machine [16] extended SVM into the multi-task learning framework by confining parameters for all tasks as similar as possible. Parameswaran and Weinberger [17] extended large margin nearest neighbor (lmnn) algorithm to the MTL paradigm.

However, the assumption of positive tasks correlation is too strong to conform the practical situation. Therefore, researchers come up with distinct models to figure out the outlier tasks and negative task correlation. Thrun and O'Sullivan [18] firstly came up with the task clustering method by introducing a weighted nearest neighbor classifier for each task. Bakker and Heskes [19] developed a multi-task Bayesian neural network model. The work by Jacob et al. [20] explored task clusters under the regularization framework using three orthogonal terms.

Learning the task relationships automatically from data is an advanced learning method. In [21], the covariance matrix of tasks relationships was learned by assuming the data samples conforming to Gauss distribution. Multi-task relationship learning (MTRL) [22] also learned the covariance matrix of tasks relationship but through a more direct way, assuming parameter matrix conforming to the matrix normal distribution. [23] was similar to MTRL, but the model construct the covariance matrix of tasks relationship as well we feature.

2.2 Multi-view Learning

Multi-view learning (MVL) makes use of the data coming from multiple sources to explore the latent knowledge. For MVL models both consensus principle and complementary principle are crucial principles to obey [10]. According to different application, existing multi-view learning is mainly divided into tree categories: co-training, multiple kernel learning and subspace learning [24].

Co-training utilizes the complementary information among multiple views to learn alternatively, minimizing the disagreement and thus improving the model generalization. Multiple kernel learning explores the connection among multiple views by integrating distinctive kernel functions corresponding to distinctive feature spaces. Subspace learning assumes multiple views share common latent space. Although these three learning methods are seemingly diverse, they all follow consensus principle and complementary principle.

With the extensive study of MVL, there are a variety of SVM-based MVL models. Brefeld and Scheffer [25] developed the Co-EM SVM to exploit the unlabeled data. SVM-2K [26] was proposed to take advantage of two views by combining SVM and the distance minimization version of KCCA. In [27],

Li et al. linked co-training to random sampling building up a new model MTSVM. The work by Xu et al. [28] introduced the theory of the information bottleneck to multi-view learning. Rakotomamonjy et al. suggested a multi-view intact space learning algorithm [29] by incorporating the encoded complementary information to MVL.

2.3   Multi-view Multi-task Learning

Many real-world problems are so complicated that they usually require to learn several tasks at the same time with diverse data sources. Because this kind of problems own task heterogeneity as well as feature heterogeneity, multi-task learning or multi-view learning cannot provide solution for these kind of problems. Existing multi-task learning merely takes advantage of the relatedness among different tasks ignoring the consistency within distinct views; however, existing multi-view learning have not yet to take the information from other tasks into consideration. Therefore, multi-view multi-task learning (MVMTL) comes into being recently.

A graph-based framework ($GraM^2$) to deal with multi-task multi-view problem was proposed in [3]. He and Lawrence assumed that in a single task each of the view keep consistency with other views, and the shared views among different tasks own the similar predictions. Under this situation, shared views became the bridge to connect distinct tasks. Correspondingly, an effective algorithm ($IteM^2$) was designed to solve the problem. However, the $GraM^2$ framework only aimed at nonnegative data set. In order to expand the range of data set to the negative data, a regularized framework was proposed. Based on the co-regularization in a single task, Zhang and Huan [4] added regularized multi-task learning method into the co-regularization model. Algorithm based on share structure to deal with multi-view multi-task learning [5] was also proposed afterwards. Save for aiming at classification problem, in [6] Zhang et al. introduced a novel problem named Multi-view Multi-task Cluster Learning. In order to deal with this special cluster problem, they presented an algorithm based on graph model to handle nonnegative data at first [6]. Then an improved algorithm [7] was introduced to solve more general data set including negative data.

3   PSVM-2VMT Model

There are several multi-view multi-task learning methods based on different perspective such as graph models and co-regularized methods. However, models based on SVM have not yet been studied. SVMs, as traditional powerful machine learning models, outperformance most other learning methods. Hence, we propose a SVM-based model to deal with the MVMT learning. We firstly apply an advanced multi-view learning method PSVM-2V within each task and then learn multiple related tasks simultaneously using regularization techniques. Through extending PSVM-2V model to multi-task learning framework, we establish a powerful model based on SVM to solve MVMT problem.

3.1   Notation and Problem Overview

Consider a multi-view multi-task learning problem with T tasks. In each task, there is a supervised multi-view learning problem with data set (X_t, Y_t), where X_t comes from multiple sources. In order to make use of all tasks simultaneously with all views, an unified model is needed to learn the decision function $f(x)$ for every view in every task. In this paper, our proposed model is based on PSVM-2V. As a result, there are only two views have been taking into considerations. The scripts of A and B represent the certain two views. Suppose we use low-ercase letter t to present the serial number of tasks, then there are l_t samples for task t and the ith training point in task t is presented as $(x_{i_t}^A, x_{i_t}^B, y_{j_t})$. In proposed model, w_A^t, w_B^t denote weight vectors for views A and B in task t. C, C^A, C^B, γ, θ are hyperparameters remain to be chosen.

3.2   PSVM-2V

PSVM-2V model is a novel MVL method which incorporates Learning Using Privileged Information (LUPI) into MVL [10]. This model takes views A and B into consideration, regarding each view as the other view's privileged informa-tion. The concrete formulation of PSVM-2V is presented as follow:

$$
\begin{aligned}
\min_{w_A, w_B} \quad & \frac{1}{2}(\|w_A\|^2 + \gamma\|w_B\|^2) + C^A\sum_{i=1}^{l}\xi_i^{A^*} + C^B\sum_{i=1}^{l}\xi_i^{B^*} + C\sum_{i=1}^{l}\eta_i \\
\text{s. t.} \quad & |(w_A \cdot \phi_A(x_i^A)) - (w_B \cdot \phi_B(x_i^B))| \leqslant \varepsilon + \eta_i, \\
& y_i(w_A \cdot \phi_A(x_i^A)) \geqslant 1 - \xi_i^{A^*}, \\
& y_i(w_B \cdot \phi_B(x_i^B)) \geqslant 1 - \xi_i^{B^*}, \\
& \xi_i^{A^*} \geqslant y_i(w_B \cdot \phi_B(x_i^B)), \ \xi_i^{A^*} \geqslant 0, \\
& \xi_i^{B^*} \geqslant y_i(w_A \cdot \phi_A(x_i^A)), \ \xi_i^{B^*} \geqslant 0, \\
& \eta_i \geqslant 0, \ i = 1, \cdots, l.
\end{aligned}
\tag{1}
$$

3.3   PSVM-2VMT

Existing PSVM-2V only aims at single task with two views. When we are con-fronted with multiple tasks, one direct way to extend the PSVM-2V is to learn each of the multiple task individually, the optimization goal is presented below:

$$
\min_{w_A^t, w_B^t} \sum_{t=1}^{T}\left[\frac{1}{2}(\|w_A^t\|^2 + \gamma\|w_B^t\|^2) + C^A\sum_{i_t=1}^{l_t}\xi_{i_t}^{A^*} + C^B\sum_{i_t=1}^{l_t}\xi_{i_t}^{B^*} + C\sum_{i_t=1}^{l_t}\eta_{i_t}\right]
\tag{2}
$$

Apparently Eq. (2) has not utilize the relationship among different tasks. To use the relationship among multiple tasks, we add a regularized term in the objective function. We chose the least square loss as the formulation of the regularized term, on one hand this regularization term limits the change of weight among

tasks, on the other hand it is easy to optimize by calculating the gradient. At last, we gain the following model:

$$
\min_{w_A^t, w_B^t} \sum_{t=1}^{T} \left[\frac{1}{2}(\|w_A^t\|^2 + \gamma\|w_B^t\|^2) + C^A \sum_{i_t=1}^{l_t} \xi_{i_t}^{A^*} + C^B \sum_{i_t=1}^{l_t} \xi_{i_t}^{B^*} + C \sum_{i_t=1}^{l_t} \eta_{i_t} \right]
$$

$$
+ \frac{\theta}{2} \sum_{t \neq t'} (\|w_A^t - w_A^{t'}\|^2 + \|w_B^t - w_B^{t'}\|^2)
$$

$$
\text{s. t. } |(w_A^t \cdot \phi_A(x_{i_t}^A)) - (w_B^t \cdot \phi_B(x_{i_t}^B))| \leqslant \varepsilon + \eta_{i_t},
$$

$$
y_{i_t}(w_A^t \cdot \phi_A(x_{i_t}^A)) \geqslant 1 - \xi_{i_t}^{A^*},
$$

$$
y_{i_t}(w_B \cdot \phi_B(x_{i_t}^B)) \geqslant 1 - \xi_{i_t}^{B^*},
$$

$$
\xi_{i_t}^{A^*} \geqslant y_{i_t}(w_B^t \cdot \phi_B(x_{i_t}^B)), \ \xi_{i_t}^{A^*} \geqslant 0,
$$

$$
\xi_{i_t}^{B^*} \geqslant y_{i_t}(w_A^t \cdot \phi_A(x_{i_t}^A)), \ \xi_{i_t}^{B^*} \geqslant 0,
$$

$$
\eta_{i_t} \geqslant 0, \ i_t = 1, \cdots, l_t. \tag{3}
$$

According to the traditional method to settle the SVM problem, deriving the corresponding dual problem is an effective way to simplify the primal problem. Hence, we take Eq. (3) as primal problem and derive the dual problem. On the basis of the dual theory, we calculate the derivative of the Lagrangian function, gain the KKT conditions and obtain the dual problem as shown in Eq. (4).

$$
\min \sum_{t=1}^{T} [(\theta + \frac{1}{2} - \theta T) \sum_{i_t, j_t=1}^{l_t} (\alpha_{i_t}^A y_{i_t} - \beta_{i_t}^+ + \beta_{i_t}^- - \lambda_{i_t}^B y_{i_t})(\alpha_{j_t}^A y_{j_t} - \beta_{j_t}^+ + \beta_{j_t}^- - \lambda_{j_t}^B y_{j_t}) \kappa_A(x_{i_t}^A, x_{j_t}^A)
$$

$$
+ (\theta + \frac{1}{2\gamma} - \theta T) \sum_{i_t, j_t=1}^{l_t} (\alpha_{i_t}^B y_{i_t} + \beta_{i_t}^+ - \beta_{i_t}^- - \lambda_{i_t}^A y_{i_t})(\alpha_{j_t}^B y_{j_t} + \beta_{j_t}^+ - \beta_{j_t}^- - \lambda_{j_t}^A y_{j_t}) \kappa_B(x_{i_t}^B, x_{j_t}^B)]
$$

$$
+ \theta \sum_{t \neq t'} [\sum_{i_t=1}^{l_t} \sum_{j_{t'}=1}^{l_{t'}} (\alpha_{i_t}^A y_{i_t} - \beta_{i_t}^+ + \beta_{i_t}^- - \lambda_{i_t}^B y_{i_t})(\alpha_{j_{t'}}^A y_{j_{t'}} - \beta_{j_{t'}}^+ + \beta_{j_{t'}}^- - \lambda_{j_{t'}}^B y_{j_{t'}}) \kappa_A(x_{i_t}^A, x_{j_{t'}}^A)
$$

$$
+ \sum_{i_t=1}^{l_t} \sum_{j_{t'}=1}^{l_{t'}} (\alpha_{i_t}^B y_{i_t} + \beta_{i_t}^+ - \beta_{i_t}^- - \lambda_{i_t}^A y_{i_t})(\alpha_{j_{t'}}^B y_{j_{t'}} + \beta_{j_{t'}}^+ - \beta_{j_{t'}}^- - \lambda_{j_{t'}}^A y_{j_{t'}}) \kappa_B(x_{i_t}^B, x_{j_{t'}}^B)]
$$

$$
+ \sum_{t=1}^{T} [\varepsilon \sum_{i_t=1}^{l_t} (\beta_{i_t}^+ + \beta_{i_t}^-) - \sum_{i_t=1}^{l_t} (\alpha_{i_t}^A + \alpha_{i_t}^B)]
$$

$$
\text{s. t. } \alpha_{i_t}^A + \lambda_{i_t}^A \leqslant C^A, \ \alpha_{i_t}^B + \lambda_{i_t}^B \leqslant C^B, \ \beta_{i_t}^+ + \beta_{i_t}^- \leqslant C,
$$

$$
\alpha_{i_t}^A, \alpha_{i_t}^B, \beta_{i_t}^+, \beta_{i_t}^-, \lambda_{i_t}^A, \lambda_{i_t}^B \geqslant 0. \tag{4}
$$

Because the formulation of dual problem in Eq. (4) is a classical convex QPP, we can solve the problem using QP solver. Moreover, using the KKT conditions we have the following conclusions without proof, which is similar to the conclusions in [30]. Suppose that $\alpha_A^1\top, \alpha_B^1\top, \beta_+^1\top, \beta_-^1\top, \lambda_A^1\top,$ $\lambda_B^1\top, \ldots, \alpha_A^T\top, \alpha_B^T\top, \beta_+^T\top, \beta_-^T\top, \lambda_A^T\top, \lambda_B^T\top$ is a solution of Eq. (4), then the solutions w_A^t and w_B^t of Eq. (3) can be formulated as follows.

$$
w_A^t = \sum_{i_t=1}^{l_t} (\alpha_{i_t}^A y_{i_t} - \beta_{i_t}^+ + \beta_{i_t}^- - \lambda_{i_t}^B y_{i_t}) \phi_A(x_{i_t}^A), \tag{5}
$$

$$w_B^t = \frac{1}{\gamma} \sum_{i_t=1}^{l_t} (\alpha_{i_t}^B y_{i_t} + \beta_{i_t}^+ - \beta_{i_t}^- - \lambda_{i_t}^A y_{i_t}) \phi_B(x_{i_t}^B). \tag{6}$$

Since in PSVM-2V there is a assumption that each view has sufficient information to learn a classifier, we assume that in PSVT-2VMT two discriminative classifiers learning from different feature views are equally important. Hence, we have the following prediction function to predict the label of a new sample (x_t^A, x_t^B) for task t:

$$f_t = sign(f_t(x_t^A, x_t^B)) = sign(0.5(w_A^t *^\top \phi_A(x_t^A) + w_B^t *^\top \phi_B(x_t^B))). \tag{7}$$

where w_A^t* and w_B^t* are the optima of Eq. (3)

In summary, we can predict using Eq. (7) when both the two views of a new sample are available.

4   Numerical Experiment

In this section, we demonstrate the effectiveness of proposed model for binary classification based on 10 data sets obtained from *Animals with Attributes (AwA)*. We carry out experiments on a Windows workstation with Inter Core CPU(i7-6700K@4.00 GHz) and 32-GB RAM. In order to measure the performance of different models, we take the accuracy as a criterion. Through using fivefold cross validation, we gain the best parameter for each model. The details of experiments are as follow.

4.1   Experimental Setup

Data Sets. *Animals with Attributes:* The *Animals with Attributes (AwA)*[1] contains 30475 images of 50 animals classes with six pre-extracted feature representations for each image. In our experiments, we take the 252-dimensional HOG features and the 2000-dimensional L_1 normalized SURF descriptors as views A and B. Moreover, we take out ten classes as train and test data sets and construct nine binary classifications regarding as nine tasks. There are 200 samples selected randomly for each task to train. Table 1 shows the details of these nine tasks.

Parameters. In PSVM-2VMT, there are several hyperparameters which influence the performance of model. In order to obtain the best parameters for all models, we implement fivefold cross validation. Empirically, the smaller the parameter ϵ in SVM is, the performance of SVM is better. Hence, we set ϵ to be 0.001. For convenience, we set $C = C^A = C^B$. Under this situation, there are still four hyperparameters including kernel parameter σ, penalty parameter C,θ and nonnegative parameter γ need to be chosen. We adopt grid search as a means of choosing hyperparameters. Since a grid search usually picks values approximately on a logarithmic scale, we select those four hyperparameter from $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3\}$.

---

[1] Available at http://attributes.kyb.tuebingen.mpg.de.

Table 1. Details of multiple tasks

Task number	Classification problem
Task 1	Chimpanzee vs Giant panda
Task 2	Chimpanzee vs Leopard
Task 3	Chimpanzee vs Persian cat
Task 4	Chimpanzee vs Pig
Task 5	Chimpanzee vs Hippopotamus
Task 6	Chimpanzee vs Humpback whale
Task 7	Chimpanzee vs Raccoon
Task 8	Chimpanzee vs Rat
Task 9	Chimpanzee vs Seal

4.2   Experimental Results

We use PSVM-2VMT to settle MVMT learning aiming at the aforementioned tasks. Due to the limitation of QP solver for large-scale data set, we choose two tasks as the input of PSVM-2VMT. Hence, we obtain 80 results for each task pair combination, as shown in Table 2. Select the optimal accuracy for each task, we draw the histogram as shown in Fig. 1.

Table 2. Performance on PSVM-2VMT based on 2 tasks

| Training task	1:75.28	1:76.3	1:75.44	1:76.42	1:76.56	1:76.46	1:75.78	1:75.27
	2:84.34	3:82.4	4:75.15	5:79.82	6:95.52	7:76.45	8:68.31	9:83.72
Training task	1:75.28	1:76.3	1:75.44	1:76.42	1:76.56	1:76.46	1:75.78	1:75.27
	2:84.34	3:82.4	4:75.15	5:79.82	6:95.52	7:76.45	8:68.31	9:83.72
Training task	2:83.82	2:83.99	2:80.38	2:**86.86**	2:83.5	2:82.95	2:83.87	2:84.54
	1:76.64	3:82.22	4:71.4	5:78.41	6:96.1	7:77.8	8:68.89	9:83.37
Training task	3:80.95	3:80.95	3:80.95	3:82.13	3:81.8	3:**82.57**	3:81.89	3:81.41
	1:**77.69**	2:83.19	4:72.68	5:80.4	6:**97.13**	7:78.66	8:68.91	9:83.34
Training task	4:72.33	4:73.15	4:72.57	4:71.99	4:72.12	4:72.66	4:72.12	4:71.58
	1:76.81	2:84.04	3:81.3	5:**81.59**	6:96.7	7:76.28	8:65.86	9:84.45
Training task	5:79.22	5:78.86	5:80.04	5:79.19	5:78.6	5:78.6	5:78.6	5:79.75
	1:75.74	2:85.14	3:81.49	4:71.81	6:95.77	7:76.07	8:72.11	9:84.3
Training task	6:96.3	6:96.3	6:96.3	6:96.3	6:96.3	6:96.3	6:95.59	6:96.71
	1:77.06	2:82.38	3:81.7	4:72.18	5:78.92	7:77.48	8:71.04	9:83.72
Training task	7:75.84	7:76.23	7:**78.89**	7:76.55	7:77.13	7:77.11	7:76.13	7:76.13
	1:76.11	2:86.3	3:81.84	4:**75.81**	5:79.64	6:96.21	8:65.76	9:83.66
Training task	8:65.6	8:65.6	8:65.6	8:65.6	8:65.6	8:69.91	8:65.44	8:68.91
	1:76.94	2:84.46	3:81.79	4:72.63	5:79.46	6:95.72	7:77.09	9:84.48
Training task	9:83.64	9:84.46	9:84.11	9:84.39	9:84.7	9:84.78	9:84.78	9:**84.78**
	1:75.98	2:86.08	3:81.03	4:71.78	5:79.13	6:96.35	7:76.27	8:**75.11**
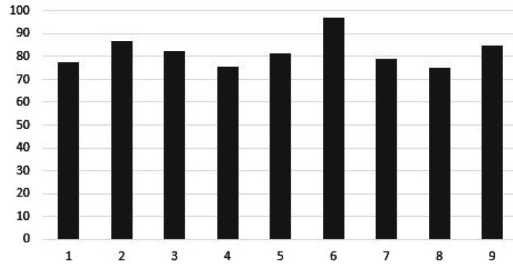
Fig. 1. Best accuracy of 9 tasks

5   Conclusion

In this paper, we proposed a novel model based on SVM to settle the MVMT learning. The existing model PSVM-2V is an effective model for MVL achieving both consensus and complementary principle. Based on PSVM-2V, we construct PSVM-2VMT to settle the MVMT learning. We have derived the corresponding dual problem and adopted the classical QP to solve it. Experimental results demonstrated the effectiveness of our models. In the future, we will design correspond speedup algorithm to solve our problems. Furthermore, because we assume all tasks are related in PSVM-2VMT, we will explore more complicated task relationship in the future study.

References

1. Su, H., Maji, S., Kalogerakis, E., Learned-Miller, E.: Multi-view convolutional neural networks for 3D shape recognition. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 945–953 (2015)
2. Dhillon, P., Foster, D.P., Ungar, L.H.: Multi-view learning of word embeddings via CCA. In: Advances in Neural Information Processing Systems, pp. 199–207 (2011)
3. He, J., Lawrence, R.: A graph-based framework for multi-task multi-view learning. In: ICML, pp. 25–32 (2011)
4. Zhang, J., Huan, J.: Inductive multi-task learning with multiple view data. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 543–551. ACM (2012)
5. Jin, X., Zhuang, F., Wang, S., He, Q., Shi, Z.: Shared structure learning for multiple tasks with multiple views. In: Blockeel, H., Kersting, K., Nijssen, S., Železný, F. (eds.) ECML PKDD 2013. LNCS (LNAI), vol. 8189, pp. 353–368. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40991-2_23
6. Zhang, X., Zhang, X., Liu, H.: Multi-task multi-view clustering for non-negative data. In: IJCAI, pp. 4055–4061 (2015)

7. Zhang, X., Zhang, X., Liu, H., Liu, X.: Multi-task multi-view clustering. IEEE Trans. Knowl. Data Eng. **28**(12), 3324–3338 (2016)
8. Tian, Y., Qi, Z., Ju, X., Shi, Y., Liu, X.: Nonparallel support vector machines for pattern classification. IEEE Trans. Cybern. **44**(7), 1067–1079 (2014)
9. Tian, Y., Ju, X., Qi, Z., Shi, Y.: Improved twin support vector machine. Sci. China Math. **57**(2), 417–432 (2014)
10. Tang, J., Tian, Y., Zhang, P., Liu, X.: Multiview privileged support vector machines. IEEE Trans. Neural Netw. Learn. Syst. (2017)
11. Zhang, Y., Yang, Q.: A survey on multi-task learning. arXiv preprint arXiv:1707.08114 (2017)
12. Argyriou, A., Evgeniou, T., Pontil, M.: Multi-task feature learning. In: Advances in Neural Information Processing Systems, pp. 41–48 (2007)
13. Obozinski, G., Taskar, B., Jordan, M.: Multi-task feature selection. Statistics Department, UC Berkeley, Technival report 2 (2006)
14. Liu, H., Palatucci, M., Zhang, J.: Blockwise coordinate descent procedures for the multi-task lasso, with applications to neural semantic basis discovery. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 649–656. ACM (2009)
15. Gong, P., Ye, J., Zhang, C.: Multi-stage multi-task feature learning. In: Advances in Neural Information Processing Systems, pp. 1988–1996 (2012)
16. Evgeniou, T., Pontil, M.: Regularized multi-task learning. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 109–117. ACM (2004)
17. Parameswaran, S., Weinberger, K.Q.: Large margin multi-task metric learning. In: Advances in Neural Information Processing Systems, pp. 1867–1875 (2010)
18. Thrun, S., O'Sullivan, J.: Discovering structure in multiple learning tasks: the TC algorithm. In: ICML, vol. 96, pp. 489–497 (1996)
19. Bakker, B., Heskes, T.: Task clustering and gating for Bayesian multitask learning. J. Mach. Learn. Res. **4**(May), 83–99 (2003)
20. Jacob, L., Vert, J.P., Bach, F.R.: Clustered multi-task learning: a convex formulation. In: Advances in Neural Information Processing Systems, pp. 745–752 (2009)
21. Bonilla, E.V., Chai, K.M., Williams, C.: Multi-task Gaussian process prediction. In: Advances in Neural Information Processing Systems, pp. 153–160 (2008)
22. Zhang, Y., Yeung, D.Y.: A convex formulation for learning task relationships in multi-task learning. arXiv preprint arXiv:1203.3536 (2012)
23. Zhang, Y., Schneider, J.G.: Learning multiple tasks with a sparse matrix-normal penalty. In: Advances in Neural Information Processing Systems, pp. 2550–2558 (2010)
24. Xu, C., Tao, D., Xu, C.: A survey on multi-view learning. arXiv preprint arXiv:1304.5634 (2013)
25. Brefeld, U., Scheffer, T.: Co-EM support vector learning. In: Proceedings of the Twenty-first International Conference on Machine learning, p. 16. ACM (2004)
26. Sonnenburg, S., Rätsch, G., Schäfer, C., Schölkopf, B.: Large scale multiple kernel learning. J. Mach. Learn. Res. **7**(Jul), 1531–1565 (2006)
27. Muslea, I., Minton, S., Knoblock, C.A.: Active + semi-supervised learning = robust multi-view learning. In: ICML, vol. 2, pp. 435–442 (2002)
28. Xu, C., Tao, D., Xu, C.: Large-margin multi-viewinformation bottleneck. IEEE Trans. Pattern Anal. Mach. Intell. **36**(8), 1559–1572 (2014)
29. Suzuki, T., Tomioka, R.: SpicyMKL. arXiv preprint arXiv:0909.5026 (2009)
30. Deng, N., Tian, Y., Zhang, C.: Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions. CRC Press, Boca Raton (2012)