Multi-view Support Vector Machines with Sub-view Learning

Qi Hao
Tianjin University of Technology
Wenguang Zheng (wenguangz@tjut.edu.cn)
Tianjin University of Technology
Yingyuan Xiao
Tianjin University of Technology
Wenxin Zhu
Tianjin Agricultural University

Research Article

Keywords: Multi-view learning, Support vector machines, Sub-views, Privileged information

Posted Date: September 27th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1779269/v1

Multi-view Support Vector Machines with Sub-view Learning

Qi Hao[1,2], Wenguang Zheng[1,2*], Yingyuan Xiao[1,2*] and Wenxin Zhu[3]

[1]School of Computer Science and Engineering, Tianjin University of Technology, Tianjin, 300384, China.

[2]Engineering Research Center of Learning-Based Intelligent System, Ministry of Education, Tianjin, 300384, China.

[3]College of Basic Science, Tianjin Agricultural University, Tianjin, 300384, China.

*Corresponding author(s). E-mail(s): wenguangz@tjut.edu.cn; yyxiao@tjut.edu.cn;
Contributing authors: haoqi@stud.tjut.edu.cn; zhuwenxin@tjau.edu.cn;

Abstract

Multi-view learning improves the performance of existing learning tasks by using complementary information between multiple feature sets. In the latest research, multi-view learning model using privilege information is proposed, specific models such as PSVM-2V and MCPK. In these models, views complement each other by acting as privileged information policies, However, a single view contains privilege information that can guide the classifier, and the existing framework does not consider it. In order to use this information to correct multi-view support vector machine classifier, we propose a framework for generating a series of small-scale views based on information hidden in a single view, which extends the original multi-view parallel structure to a hierarchical structure with sub-view mechanism. In this paper, two sub-view learning structures SL-PSVM-2V and SL-MCPK are constructed. The two new models fully exploit the data features in the view. Similarly, they follow the principles of consistency and complementarity. We use the standard quadratic programming solver to solve the new model. In 55 groups of classification experiments and noise sensitivity tests, the new model has better performance than the benchmark model. Statistical comparison shows that the new method is significantly different from the existing methods.

Keywords: Multi-view learning, Support vector machines, Sub-views, Privileged information

1 Introduction

In recent years, multi-view learning has become an active research direction of machine learning, and has been applied to learning problems in different fields. Such as image classification (Han et al, 2018; Sun et al, 2019; Zhang et al, 2020), brain network analysis (Ahmed et al, 2017; Appice and Malerba, 2016a), treatment research (Chao et al, 2019).

Multi-view data are directly collected in the real world or extracted by different feature extraction methods. Studies have shown that various viewpoints are interrelated and complementary. Compared with single-view training and direct connection of different view data, multi-view learning can mine more information.

The classification models of multi-view learning research are mainly divided into three

categories: co-training style algorithms, co-regularization style algorithms, and margin-consistency. In the co-training style algorithm, the machine learning method trains alternately on different views. For example, multi-view collaborative clustering algorithm (Appice and Malerba, 2016b) and Kim et al (2019) uses multiple collaborative training for document classification. The co-regularization style algorithm takes the divergence between different viewpoints as a new regularization term or constraint in the learning objective function. The typical methods are SVM-2K (Farquhar et al, 2005), multi-view LS-TSVMs (Xie, 2018), multi-view LSSVMs (Houthuys et al, 2018), multi-view RSVMs (Li et al, 2018), etc. Margin-consistency algorithm has the edge consistency style, they model the marginal variables of multiple views as close as possible, so that the machine learning model can use the potential consistency of the classification results from multiple views, and the representative algorithms include MVMED (Li et al, 2018), SMVMED (Sun and Chao, 2013) and MED-2C (Chao and Sun, 2016).

Most SVM-based multi-view models only consider the consensus principle, while ignoring the complementarity principle. The consensus principle is to maximize the consistency between multiple views. The complementary principle emphasizes that each view contains some knowledge that other views do not have, and the complementary information is shared between the views to comprehensively describe the data. In summary, the principle of consistency and complementarity is an important basis for multi-view learning.

Learning using privileged information(LUPI) is proposed for accompanying or hidden information in the learning model (Vapnik et al, 2007). Using privileged information for learning has been widely used in many tasks, such as text clustering (Sinoara et al, 2014), image recognition (Guo et al, 2018; Yan et al, 2016), etc. In classification tasks, this information can provide an effective supplementary strategy for classification. Vapnik and Vashist first proposed a SVM-based model under LUPI: SVM+ (Vapnik et al, 2007). In order to reduce the calculation time, L.Niu constructed a new LUPI model using the improved L2 norm (Niu and Wu, 2012). M. Lapin assigns different weights to samples with privileged information (Lapin et al, 2014). For support vector machines with different tasks, the version of privileged information learning is introduced. Xu et al (2019) proposed an autonomous learning model using privileged information. Liang and Cherkassky (2007) used group information as privileged information. The training data can be grouped according to a feature attribute, and a formal optimization formula is sorted out. Che et al (2021) proposed a twin support vector machine model for privileged information learning based on LUPI paradigm . Some other models have been improved in the SVM with privileged information to make the model more robust (Li et al, 2021).

LUPI can be used for information complementarity and information interaction. Under the principles of consensus and consistency, A series of classification models applying privilege information to multi-view learning are proposed. The initial model is PSVM-2V (Tang et al, 2018b), and then a version that can realize multiple views (more than two views) is proposed: IPSVM-MV (Tang et al, 2018a). PSVM-2V and IPSVM-MV give full play to the performance of complementary information between views, but the consistency constraint of regularization makes the model solving too complex and time consuming in application. In the latest research, Tang proposed the MCPK (coupled privileged kernel method for multi-view learning) model. MCPK uses coupling terms in the original target to minimize error combinations in all views, thus ensuring consistency. At the same time, due to the existence of coupling terms, the complexity of model optimization is greatly reduced (Tang et al, 2019).

PSVM-2V and MCPK use privileged information as complementary constraints between views. However, each view also has its own privileged or structural information, which will guide the classifier to work better. We can get inspiration from using group information as privileged information and propose the concept of sub-view learning. Sub-views are generated from the original view and trained in the LUPI paradigm, which can form multiple sub-view correction planes to correct multi-view learning. Based on the existing multi-view learning methods, this paper expands the view framework and proposes a classification model based on sub-view learning.

The following is a brief description of our contribution:

- In this paper, a new multi-view learning method is proposed by using the sub-view mechanism. This new method is easy to implement and can effectively use the hidden features in multi-view data.
- Two new classification methods with sub-view paradigm, SL-PSVM-2V and SL-MCK, are constructed by using PSVM-2V and MCPK multi-view models, respectively. The solution strategy based on quadratic programming is given.
- Compared with multiple multi-view learning, analyze their compliance with the consistency principle and complementary principle, and sort out the transformation method between each model.
- The effectiveness of the new method and the classification performance of each data set are verified through multiple sets of experiments, and the performance of each classification method in the noise data set is compared. At the same time, non-parametric tests are carried out to verify the significant differences between models.

The paper is organized as follows. Section 2 describes the main related work, including learning using privileged information principle, PSVM-2V and MCPK. Section 3 introduces two new classification models: SL-PSVM-2V and SL-MCPK. Section 4 analyzes and compares several algorithms, and shows the transformation method. Section 5 shows the experimental results and analysis. Section 6 is the summary of this paper and the prospect of future research.

2 Related works

2.1 Learning Using Privileged Information

Privileged information can be used as an additional feature to help specific classifiers work better. Training data are used according to privilege information, which contains additional information provided only in the training process rather than in the test process. privileged information is ubiquitous and useful.

Vapnik et al (2007) developed the earliest LUPI model: SVM+ by incorporating privileged knowledge into SVM. Many experiments have proved a comprehensive theoretical explanation and algorithm description of privileged information learning are carried out to ensure and improve the prediction performance .

SVM+ appears in the form of privilege features. These features are used not only to estimate the relaxation of constraints, but also to establish an upper bound for the risk of decision functions. The purpose of LUPI is to use privileged information to learn a model, so as to further constrain the solution in the original space. The SVM model that realizes LUPI at the training stage can be expressed as follows:

$$\begin{aligned} \underset{w,w^*,b,d}{min} \quad & \frac{1}{2}\|w\|^2 + \frac{\gamma}{2}\|w^*\|^2 + C\sum_{i=0}^{n}(w^* \cdot x_i^* + d), \\ s.t. \quad & y_i(w \cdot x_i) \geq 1 - (w^* \cdot x_i^* + d), \\ & w^* \cdot x_i^* + d \geq 0, i = 1, ..., n. \end{aligned} \tag{1}$$

where w and w^* represent the weight vector, b and d represent the bias term, C is used to balance the loss, and γ balances the weight of privileged information in the privileged space. The optimal parameter w and b are solved for classification prediction, and this w is the result of correcting the standard classification plane through the correction plane. The final function for decision making is:

$$h(x) = sign(w \cdot x + b) \tag{2}$$

2.2 PSVM-2V

PSVM-2V introduces LUPI paradigm into multi-view learning. The basic idea of PSVM-2V is that two views complement each other as privileged information. The learning structure uses the regularization term to bridge the gap between the two classifiers and correct the classification hyperplane.

Considering a multi-view classification problem, the training data is:

$$S = \{x_i^A, x_i^B, y_i\}_{i=1}^{l} = \{(x_i^A;\ 1), (x_i^B;\ 1)\}_{i=1}^{l},$$

where the label $y_i \in \{-1, +1\}$, and each training sample is independently distributed. In data x^A and x^B, there is a constant term (1) connection at the end of each feature data to express the classifier without clear bias term. PSVM-2V is formally

built as Eq.(3):

$$\begin{aligned}
&\min_{w_A,w_B,\xi_A,\xi_B} \quad \frac{1}{2}(\|w_A\|^2+\gamma\|w_B\|^2)\\
&\quad + C_A\sum_{i=1}^{l}\xi_i^A + C_B\sum_{i=1}^{l}\xi_i^B + C\sum_{i=1}^{l}\eta_i,\\
&s.t. \quad |w_A\cdot\phi_A(x_i^A) - w_A\cdot\phi_B(x_i^B)| \le \varepsilon+\eta_i,\\
&\quad y_i(w_A\cdot\phi_A(x_i^A)) \ge 1-\xi_i^A,\\
&\quad y_i(w_B\cdot\phi_B(x_i^B)) \ge 1-\xi_i^B,\\
&\quad \xi_i^A \ge y_i(w_B\cdot\phi_B(x_i^B)),\\
&\quad \xi_i^B \ge y_i(w_A\cdot\phi_A(x_i^A)),\\
&\quad \xi_i^A\ge 0, \xi_i^B\ge 0, \eta_i\ge 0, i=1,...l.
\end{aligned} \tag{3}$$

where w_A and w_B are the weight vectors of view A and view B respectively. Under the excitation of LUPI paradigm, PSVM-2V limits the non-negative slack variables ξ_i^A and ξ_i^B of view A and view B through the unknown non-negative correction function determined by view B and view A respectively. Thus the complementary principle is realized. The first constraint $|w_A\cdot\phi_A(x_i^A) - w_A\cdot\phi_B(x_i^B)| \le \varepsilon+\eta_i$ realizes the consistency between the two views, and uses the slack variable η_i to weigh the number of points that fail to meet the ε similarity. C_A,C_B and C are non-negative penalty parameters. γ is a nonnegative tradeoff parameter that weighs two views.

2.3 MCPK

MCPK is a simple and effective multi-view learning privileged coupling kernel method (Tang et al, 2018a). To inherit the advantages of PSVM-2V directly, MCPK takes one view as main information and another view as privileged information. This pair view, namely the main view and privileged view, shares complementary information. Similarly, MCPK models the complementary point of view by drawing on the idea of LUPI to achieve the complementary principle. In order to complete the consensus principle, MCPK introduces a coupling term to establish a bridge for two different viewpoints. This term forces the prediction error combination of the two views to be small. Therefore, information from both views is merged into the model, and the high error variable of a sample in one view can be compensated by the corresponding low error variable in another view. Formally, MCPK classification optimization problem can be built as follows:

$$\begin{aligned}
&\min_{w_A,w_B,\xi_A,\xi_B} \quad \frac{1}{2}(\|w_A\|^2+\gamma\|w_B\|^2)\\
&\quad + C_A\sum_{i=1}^{l}\xi_i^A + C_B\sum_{i=1}^{l}\xi_i^B + C\sum_{i=1}^{l}\xi_i^A\xi_i^B,\\
&s.t. \quad y_i(w_A\cdot\phi_A(x_i^A)) \ge 1-\xi_i^A,\\
&\quad y_i(w_B\cdot\phi_B(x_i^B)) \ge 1-\xi_i^B,\\
&\quad \xi_i^A \ge y_i(w_B\cdot\phi_B(x_i^B)),\\
&\quad \xi_i^B \ge y_i(w_A\cdot\phi_A(x_i^A)),\\
&\quad \xi_i^A\ge 0, \xi_i^B\ge 0, i=1,...,l.
\end{aligned} \tag{4}$$

where w_A and w_B are the weight vectors of view A and view B, respectively, and the two views are weighed by the non-negative trade-off parameter γ. As slack variables, ξ_i^A and ξ_i^A are constrained by the correction functions determined by the two views. The coupling term $C\sum_{i=1}^{l}\xi_i^A\xi_i^B$ makes the product of error variables of the two views as small as possible. When classifiers constructed from different views are more consistent, errors from both views are small, resulting in smaller couplings. Therefore, its consistency can be fully ensured. C is a non-negative coupling parameter that controls the influence of the coupling term. C_A and C_B are non-negative penalty parameters.

3 Our proposed method

3.1 Primal problem

In multi-view learning base on LUPI, views are located in parallel and corrected as privileged information. However, each view also contains information that helps to achieve classification. Such information may come from the practical significance of a feature, or may come from the feature distribution of the data. Either way, it can be used as privileged information. We use this hidden information to divide the original view data into multiple groups. Since this subset is part of the view, we call it a sub-view. Sub-views be divided by the privileged information of the original view itself (illustrated in Fig.1), or dividing foundation based on other views (illustrated in Fig.2). Sub-views can form a classification space different from

the main view (Source view of sub-view), providing more accurate guidance for classification. The sub-views formed by the mutual guidance of different views can allow different views to share this privileged information and more strictly follow the principle of complementarity.

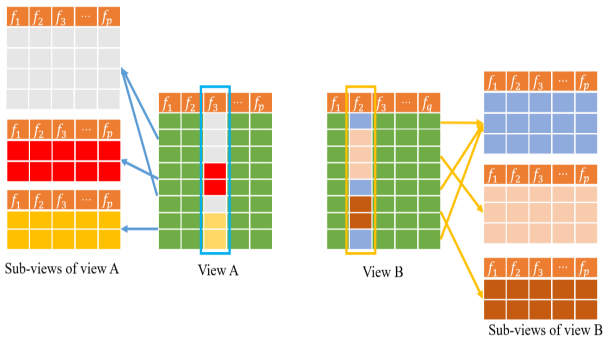


Fig. 1 Dividing sub-views by their own privileged information

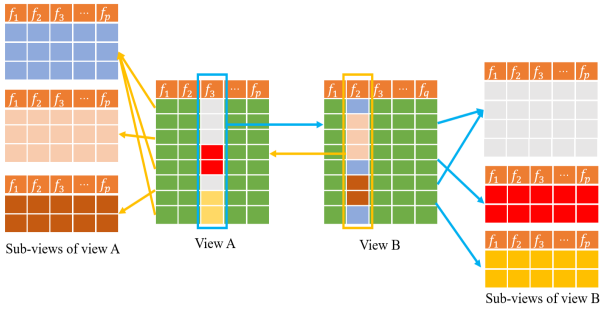


Fig. 2 Dividing sub-views through privileged information from different views

The sub-view learning considers multi-view training learning with privileged information. The training data are $S = \{x_i^A, x_i^B, y_i\}_{i=1}^l = \{(x_i^A;\ 1), (x_i^B;\ 1)\}_{i=1}^l$, where $y_i \in \{-1, +1\}$, sub-views stored by T, $T = \{T_A, T_B\}$, Sub-view data indexes in view A and view B in T_A and T_B, respectively. Sub-view regularization term and constraints on slack variables are added to transfer sub-view information, and kernel technique maps sub-view data to different feature spaces. As the sub-view learning framework shown in Fig.3, The final classification hyperplane is not only composed of information from two views, but also the correction plane formed by sub-views is corrected by LUPI paradigm. Sub-views are only used in the training classifier stage, and do not exist in the prediction classification result stage. Two multi-view sub-view learning version classification models are introduced in 3.2 and 3.3 respectively.

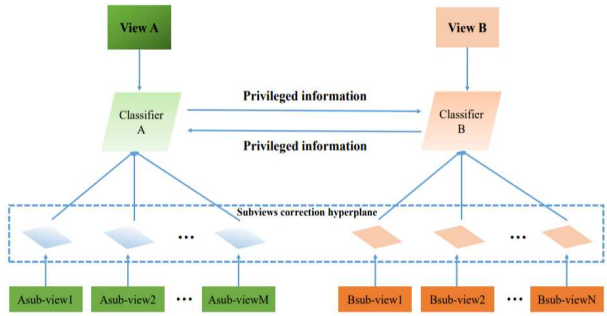


Fig. 3 The framework of sub-view learning

In order to clearly explain the model in the paper, the main symbols and their meanings are Table 1:

Table 1 Summary of notation

Notation	Description
(x_i^A, x_i^B, y)	$i-th$ training samples
w_A, w_B	Weight vectors for views A and B
w_{Asub}^r, w_{Bsub}^r	Weight vectors for the r-th sub-view of view A and view B
$\phi_A(\cdot), \phi_B(\cdot)$	Mapping to high-dimensional feature spaces (view A and view B)
$\phi_{Asub}^r(\cdot), \phi_{Bsub}^r(\cdot)$	Mapping to high-dimensional feature spaces (r-th sub-view of view A and B)
$K_A(x_i^A, x_i^A)$	Kernel function$(\phi_A(x_i^A) \cdot \phi_A(x_i^A))$
$K_B(x_i^B, x_i^B)$	Kernel function$(\phi_B(x_i^A) \cdot \phi_B(x_i^B))$
$K_{Asub}^r(x_i^A, x_i^A)$	Kernel function$(\phi_{Asub}^r(x_i^A) \cdot \phi_{Asub}^r(x_i^A))$
$K_{Bsub}^r(x_i^A, x_i^A)$	Kernel function$(\phi_{Bsub}^r(x_i^B) \cdot \phi_{Bsub}^r(x_i^B))$
T_A^r, T_B^r	The sets of indexes for storing samples contained in sub-views

3.2 Sub-views learning for PSVM-2V

Sub-views learning for PSVM-2V(SL-PSVM-2V) adds a regularization term of sub-view under the framework of PSVM-2V, and adds new constraints to the slack variables of the original model by

using the privileged information learning strategy. SL-PSVM-2V model can be built as follows:

$$
\begin{aligned}
\min_{w_A,w_B,\xi^A,\xi^B,w^r_{Asub},w^r_{Bsub}} \quad & \frac{1}{2}(\|w_A\|^2 \\
& + \gamma_A \sum_{r=1}^{M} \|w^r_{Asub}\|^2) \\
& + \frac{1}{2\gamma}(\|w_B\|^2 + \gamma_B \sum_{r=1}^{N} \|w^r_{Bsub}\|^2) \\
& + C_A \sum_{i=1}^{l} \xi_i^A + C_B \sum_{i=1}^{l} \xi_i^B + C \sum_{i=1}^{l} \eta_i, \\
s.t. \quad & |w_A \cdot \phi_A(x_i^A) - w_B \cdot \phi_B(x_i^B)| \le \varepsilon + \eta_i, \\
& y_i(w_A \cdot \phi_A(x_i^A)) \ge 1 - \xi_i^A, \\
& y_i(w_B \cdot \phi_B(x_i^B)) \ge 1 - \xi_i^B, \\
& \xi_i^A \ge y_i(w_B \cdot \phi_B(x_i^B)), \\
& \xi_i^B \ge y_i(w_A \cdot \phi_A(x_i^A)), \\
& y_i(w^r_{Asub} \cdot \phi^r_{Asub}(x_i^A)) \ge 1 - \xi_i^A, r = 1...M, i \in T_A^r, \\
& y_i(w^r_{Bsub} \cdot \phi^r_{Bsub}(x_i^B)) \ge 1 - \xi_i^B, r = 1...N, i \in T_B^r, \\
& \xi_i^A \ge 0, \xi_i^B \ge 0, \eta_i \ge 0, i = 1, ..., l.
\end{aligned} \tag{5}
$$

In this model, $\|w_A\|^2$ and $\|w_B\|^2$ are regularization terms of view A and view B, respectively. $\|w^r_{Asub}\|^2$ denotes the r-th sub-view regularization term of view A, and $\|w^r_{Bsub}\|^2$ is the r-th sub-view regularization term of view B. and are non-negative slack parameters. γ is the balance parameter to balance the weight of two views. γ_A and γ_B are the balance parameters of balance view A data and its sub-view and balance view B data and its sub-view, respectively.

$\phi_A(x_i^A)$ and $\phi_B(x_i^B)$ represent the mapping of two view data. $\phi^r_{Asub}(x_i^A)$ in the constraint is the mapping transformation of the $r-th$ sub-view data of view A,and $\phi^r_{Bsub}(x_i^B)$in the constraint is the mapping transformation of the $r-th$ sub-view data of view B.

In the constraint $y_i(w^r_{Asub} \cdot \phi^r_{Asub}(x_i^A)) \ge 1-\xi_i^A$ and $y_i(w^r_{Bsub} \cdot \phi^r_{Bsub}(x_i^B)) \ge 1-\xi_i^B$ denote that the slack variables are constrained by the correction hyperplane formed by sub-views, so the slack variables ξ_i^A and ξ_i^B can not only achieve the purpose of correlation correction between two views, but also achieve the purpose of correction of the original view by the sub-view under the background of sub-view learning. C_A and C_B are non-negative penalty parameters.

The SL-PSVM-2V non-negative slack variable η_i is used to control the gap between the classifiers associated with the two views, so as to ensure their consistency principle. C is a nonnegative penalty parameter, ε is an error parameter that allows violation of constraints.

In order to obtain the solution of Eq.(5), we conduct dual optimization. The Lagrange function is:

$$
\begin{aligned}
L = & \frac{1}{2}(\|w_A\|^2 + \gamma_A \sum_{r=1}^{M} \|w^r_{Asub}\|^2 \\
& + \gamma(\|w_B\|^2 + \gamma_B \sum_{r=1}^{N} \|w^r_{Bsub}\|^2)) \\
& + C_A \sum_{i=1}^{l} \xi_i^A + C_B \sum_{i=1}^{l} \xi_i^B + C \sum_{i=1}^{l} \eta_i \\
& + \sum_{i=1}^{l} \beta_i^A((w_A \cdot \phi_A(x_i^A)) - (w_B \cdot \phi_B(x_i^B)) - \varepsilon - \eta_i) \\
& + \sum_{i=1}^{l} \beta_i^B((w_B \cdot \phi_B(x_i^B)) - (w_A \cdot \phi_A(x_i^A)) - \varepsilon - \eta_i) \\
& + \sum_{i=1}^{l} \alpha_i^A(1 - \xi_i^A - y_i(w_A \cdot \phi_A(x_i^A))) \\
& + \sum_{i=1}^{l} \alpha_i^B(1 - \xi_i^B - y_i(w_B \cdot \phi_B(x_i^B))) \\
& + \sum_{i=1}^{l} \lambda_i^A(y_i(w_B \cdot \phi_B(x_i^B)) - \xi_i^A) \\
& + \sum_{i=1}^{l} \lambda_i^B(y_i(w_A \cdot \phi_A(x_i^A)) - \xi_i^B) \\
& + \sum_{r=1}^{N} \sum_{i \in T_A^r} \mu_i^A(1 - \xi_i^A - y_i(w^r_{Asub} \cdot \phi^r_{Asub}(x_i^A))) \\
& + \sum_{r=1}^{M} \sum_{i \in T_B^r} \mu_i^B(1 - \xi_i^B - y_i(w^r_{Bsub} \cdot \phi^r_{Bsub}(x_i^A))) \\
& - \sum_{i=1}^{l} \xi_i^A \beta_i^A - \sum_{i=1}^{l} \xi_i^B \beta_i^B + \eta_i \tau_i.
\end{aligned} \tag{6}
$$

where $\alpha_i^A, \alpha_i^B, \beta_i^A, \beta_i^B, \lambda_i^A, \lambda_i^B, \mu_i^A, \mu_i^B$ are non-negative Lagrange multiplier vectors. According

to the KKT(Karush-Kuhn-Tucker) principle, we can get the following equation:

$$\frac{\partial L}{\partial w_A} = w_A - \sum_{i=1}^{l}(\alpha_i^A y_i - \beta_i^A + \beta_i^B - \lambda_i^B y_i)\phi_A(x_i^A) = 0. \tag{7}$$

$$\frac{\partial L}{\partial w_B} = \gamma w_B - \sum_{i=1}^{l}(\alpha_i^B y_i + \beta_i^A - \beta_i^B - \lambda_i^A y_i)\phi_B(x_i^B) = 0. \tag{8}$$

$$\frac{\partial L}{\partial w_{Asub}^r} = \gamma_A w_{Asub}^r - \sum_{T_A^r}\sum_{i\in T_A^r}\mu_i^A y_i \phi_r^A(x_i^A) = 0, \quad i \in T_A^r, r = 1, ..., M. \tag{9}$$

$$\frac{\partial L}{\partial w_{Bsub}^r} = \gamma\gamma_B w_{Bsub}^r - \sum_{T_B^r}\sum_{i\in T_B^r}\mu_i^B y_i \phi_r^B(x_i^B) = 0, \quad i \in T_B^r, r = 1, ..., N. \tag{10}$$

$$\frac{\partial L}{\partial \xi_i^A} = C_A - (\alpha_i^A + \lambda_i^A + \mu_i^A + \beta_i^A) = 0, i = 1, ..., l. \tag{11}$$

$$\frac{\partial L}{\partial \xi_i^B} = C_B - (\alpha_i^B + \lambda_i^B + \mu_i^B + \beta_i^B) = 0, i = 1, ..., l. \tag{12}$$

$$\frac{\partial L}{\partial \eta_i} = C - (\beta_i^A + \beta_i^B + \tau_i) = 0, i = 1, ..., l. \tag{13}$$

$$\alpha_i^A(1 - \xi_i^A - y_i(w_A \cdot \phi_A(x_i^A))) = 0, i = 1, ..., l. \tag{14}$$

$$\alpha_i^B(1 - \xi_i^B - y_i(w_B \cdot \phi_B(x_i^B))) = 0, i = 1, ..., l. \tag{15}$$

$$\lambda_i^A(y_i(w_B \cdot \phi_B(x_i^B)) - \xi_i^A) = 0, i = 1, ..., l. \tag{16}$$

$$\lambda_i^B(y_i(w_A \cdot \phi_A(x_i^A)) - \xi_i^B) = 0, i = 1, ..., l. \tag{17}$$

Substitute formula Eq.(7)-(17) to Eq.(6), The objective function of the dual problem can be transformed. Since $\beta_i^A, \beta_i^B, \tau_i \geq 0$, we can conclude that $\alpha_i^A + \lambda_i^A + \mu_i^A \leq C_A$, $\alpha_i^B + \lambda_i^B + \mu_i^B \leq C_B$, $\beta_i^A + \beta_i^B \leq C$ are constraints. By substitution, the original dual problem can be reorganized as Eq.(18):

$$\begin{aligned}
\min \quad & \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}((\alpha_i^A y_i - \beta_i^B + \beta_i^A - \lambda_i^B y_i) \\
& K_A(x_i^A, x_j^A)(\alpha_i^A y_i - \beta_i^B + \beta_i^A - \lambda_i^B y_i) \\
& + \frac{1}{\gamma}(\alpha_i^A y_i + \beta_i^B - \beta_i^A - \lambda_i^B y_i) \\
& K_B(x_i^B, x_j^B)(\alpha_i^A y_i + \beta_i^B - \beta_i^A - \lambda_i^B y_i)) \\
& + \frac{1}{2\gamma_A}(\sum_{r=1}^{M}\sum_{j\in T_A^r}\mu_i^A y_i K_{Asub}^r(x_i^A, x_j^A)\mu_i^A y_i) \\
& + \frac{1}{2\gamma\gamma_B}(\sum_{r=1}^{N}\sum_{j\in T_B^r}\mu_i^A y_i K_{Asub}^r(x_i^A, x_j^A)\mu_i^A y_i), \\
s.t. \quad & \alpha_i^A + \lambda_i^A + \mu_i^A \leq C_A, \\
& \alpha_i^B + \lambda_i^B + \mu_i^B \leq C_B, \\
& \beta_i^A + \beta_i^B \leq C, \\
& \alpha_i^A, \alpha_i^B, \beta_i^A, \beta_i^B, \lambda_i^A, \lambda_i^B, \mu_i^A, \mu_i^B \geq 0.
\end{aligned} \tag{18}$$

This is a quadratic convex programming problem, which can be solved by the quadratic convex programming method. Solving the optimal parameters $\alpha_i^{A*}, \alpha_i^{B*}, \beta_i^{A*}, \beta_i^{B*}, \lambda_i^{A*}, \lambda_i^{B*}, \mu_i^{A*}, \mu_i^{B*}$,We use the KKT condition to get the optimal result w_A^* and w_B^*.

$$w_A^* = \sum_{i=1}^{l}(\alpha_i^{A*} y_i - \lambda_i^{B*} y_i)\phi_A(x_i^A), \tag{19}$$

$$w_B^* = \sum_{i=1}^{l}(\alpha_i^{B*} y_i - \lambda_i^{A*} y_i)\phi_B(x_i^B). \tag{20}$$

After getting the optimal w_A^* and w_B^*, use the following formula to predict the labels of the new samples (x^A, x^B) from view A and view B:

$$f_A = sign(f_A(x^A)) = sign({w_A^*}^\top \phi_A(x_i^A)), \tag{21}$$

$$f_B = sign(f_B(x^B)) = sign({w_B^*}^\top \phi_B(x_i^B)). \tag{22}$$

The final predictor of multi-views can be constructed as the average prediction factor of each

view:

$$
\begin{aligned}
f &= sign(\frac{1}{2}f_A(x^A) + \frac{1}{2}f_B(x^B)) \\
&= sign(\frac{1}{2}{w_A^*}^\top \phi_A(x^A) + \frac{1}{2}{w_B^*}^\top \phi_B(x^B)).
\end{aligned} \tag{23}
$$

Taking into account the absence of views in the test phase, if the absence of views can be used Eq.(21) or Eq.(22) to obtain results, otherwise use the multi-view classification decision function Eq.(23).

In order to express clearly, the process of SL-PSVM-2V is given in Algorithm 1:

Algorithm 1 QP Algorithm for SL-PSVM-2V

Require: Data set: $S = \left\{x_i^A, x_i^B, y_i\right\}_{y=1}^{l} = \left\{\left(x_i^A;\ 1\right), \left(x_i^B;\ 1\right)\right\}_{i=1}^{l}, y_i \in \{+1, -1\}$; Sub-views element index: $T = \{(T_A^r, T_B^r)\}$; Initial parameters: $\gamma, \gamma_A, \gamma_B, C_A, C_B, C \geq 0$.

Ensure: Decision functions f_A,f_B,f.

1: Select kernel function: kernels function of view A and view B: $K_A(x_i^A, x_j^A)$, $K_B(x_i^B, x_j^B)$, The kernel of the sub-views of A:$K_{Asub}^r(x_i^A, x_i^A)$, kernel of sub-views of B:$K_{Bsub}^r(x_i^B, x_i^B)$, And initializing kernel parameters.

2: Create and solve quadratic programming problem Eq.(18) and using cross validation to determine the optimal parameters.

3: Solving quadratic programming Eq.(18) and retaining optimal result parameters $\alpha_i^{A*}, \alpha_i^{B*}, \beta_i^{A*}, \beta_i^{B*}, \lambda_i^{A*}, \lambda_i^{B*}, \mu_i^{A*}, \mu_i^{B*}$, Get the optimal weight w_A^* and w_B^* by substituting formula Eq.(19), Eq.(20).

4: The final decision function is solved by parameters w_A^* and w_B^*:

$$
\begin{aligned}
f_A &= sign(f_A(x^A)) = sign({w_A^*}^\top \phi_A(x_i^A)), \\
f_B &= sign(f_B(x^B)) = sign({w_B^*}^\top \phi_B(x_i^B)), \\
f &= sign(\frac{1}{2}f_A(x^A) + \frac{1}{2}f_B(x^B)).
\end{aligned}
$$

3.3 Sub-views learning for MCPK

Sub-views learning for MCPK(SL-MCPK) is a sub-view learning version of MCPK, and the coupling term is retained in the framework to achieve consensus principle and complementary principle. The model can be established as Eq.(24):

$$
\begin{aligned}
\min_{w_A, w_B, \xi^A, \xi^B, w_{Asub}^r, w_{Bsub}^r} \quad & \frac{1}{2}(\|w_A\|^2 \\
& + \gamma_A \sum_{r=1}^{M} \|w_{Asub}^r\|^2) \\
& + \gamma(\|w_B\|^2 + \gamma_B \sum_{r=1}^{N} \|w_{Bsub}^r\|^2) \\
& + C_A \sum_{i=1}^{l} \xi_i^A + C_B \sum_{i=1}^{l} \xi_i^B + C \sum_{i=1}^{l} \xi_i^A \xi_i^B, \\
s.t. \quad & y_i(w_A \cdot \phi_A(x_i^A)) \geq 1 - \xi_i^A, \\
& y_i(w_B \cdot \phi_B(x_i^B)) \geq 1 - \xi_i^B, \\
& \xi_i^A \geq y_i(w_B \cdot \phi_B(x_i^B)), \\
& \xi_i^B \geq y_i(w_A \cdot \phi_A(x_i^A)), \\
& y_i(w_{Asub}^r \cdot \phi_{Asub}^r(x_i^A)) \geq 1 - \xi_i^A, r = 1...M, i \in T_A^r, \\
& y_i(w_{Bsub}^r \cdot \phi_{Bsub}^r(x_i^B)) \geq 1 - \xi_i^B, r = 1...N, i \in T_B^r, \\
& \xi_i^A \geq 0, \xi_i^B \geq 0, i = 1, ..., l.
\end{aligned} \tag{24}
$$

Similar to SL-PSVM-2V, $\|w_A\|^2$ and $\|w_B\|^2$ are regularization terms for view A and view B, w_{Asub}^r represents the $r-th$ sub-view regularization of view A, and w_{Bsub}^r represents the $r-th$ sub-view regularization of view B.

Parameters $\xi^A = [\xi_1^A, ..., \xi_l^A]$ and $\xi^B = [\xi_1^B, ..., \xi_l^B]$ are non-negative slack parameter. Under the background of sub-view learning, slack variables ξ^A and ξ^B can not only achieve the purpose of correction of the two views, but also realize the correction of the original view by the sub-view. As a coupling term, contains the sub-view information on the basis of the original balance of the errors of the two views.

Parameters C_A and C_B are non-negative penalty parameters. In the constraint formula, $\phi_A({x_i}^A)$, $\phi_B({x_i}^B)$,are different mappings for two views, $\phi_{Asub}^r\left(x_i^A\right)$is a mapping transformation of the $r-th$ sub-view data of view A,$\phi_{Bsub}^r\left(x_i^B\right)$ is the mapping transformation of the $r-th$ sub-view data of view B. $y_i(w_{Asub}^r \cdot \phi_{Asub}^r(x_i^A)) \geq 1 - \xi_i^A$ and $y_i(w_{Bsub}^r \cdot \phi_{Bsub}^r(x_i^B)) \geq 1 - \xi_i^B$ represents that the correction plane generated by the privileged information of the sub-view constrains the slack

variables to achieve the purpose of correcting the classification hyperplane.

Solving the optimal w_A^* and w_B^* in the above problem to construct multi-view classifier. The dual optimization of Eq.(24) is established. The Lagrangian function is as Eq.(25):

$$\begin{aligned}
L =&\frac{1}{2}(\|w_A\|^2+\gamma_A\sum_{r=1}^{M}\|w_{Asub}^r\|^2 \\
&+\gamma(\|w_B\|^2+\gamma_B\sum_{r=1}^{N}\|w_{Bsub}^r\|^2)) \\
&+C_A\sum_{i=1}^{l}\xi_i^A+C_B\sum_{i=1}^{l}\xi_i^B+C\sum_{i=1}^{l}\xi_i^A\xi_i^B \\
&+\sum_{i=1}^{l}\alpha_i^A(1-\xi_i^A-y_i(w_A\cdot\phi_A(x_i^A))) \\
&+\sum_{i=1}^{l}\alpha_i^B(1-\xi_i^B-y_i(w_B\cdot\phi_B(x_i^B))) \\
&+\sum_{i=1}^{l}\lambda_i^A(y_i(w_B\cdot\phi_B(x_i^B))-\xi_i^A) \\
&+\sum_{i=1}^{l}\lambda_i^B(y_i(w_A\cdot\phi_A(x_i^A))-\xi_i^B) \\
&+\sum_{r=1}^{N}\sum_{i\in T_A^r}\mu_i^A(1-\xi_i^A-y_i(w_{Asub}^r\cdot\phi_{Asub}^r(x_i^A))) \\
&+\sum_{r=1}^{M}\sum_{i\in T_B^r}\mu_i^B(1-\xi_i^B-y_i(w_{Bsub}^r\cdot\phi_{Bsub}^r(x_i^B))) \\
&-\sum_{i=1}^{l}\xi_i^A\beta_i^A-\sum_{i=1}^{l}\xi_i^B\beta_i^B.
\end{aligned} \tag{25}$$

where $\alpha_i^A,\lambda_i^A,\mu_i^A,\beta_i^A,\alpha_i^B,\lambda_i^B,\mu_i^B,\beta_i^B$ as nonnegative Lagrange multiplier vectors. According to KKT principle, we can get:

$$\frac{\partial L}{\partial w_A}=w_A-\sum_{i=1}^{l}(\alpha_i^A y_i-\lambda_i^B y_i)\phi_A(x_i^A)=0. \tag{26}$$

$$\frac{\partial L}{\partial w_B}=\gamma w_B-\sum_{i=1}^{l}(\alpha_i^B y_i-\lambda_i^A y_i)\phi_B(x_i^B)=0. \tag{27}$$

$$\begin{aligned}
\frac{\partial L}{\partial w_{Asub}^r}=&\gamma_A w_{Asub}^r-\sum_{T_A^r}^{N}\sum_{i\in T_A^r}\mu_i^A y_i\phi_r^A(x_i^A)=0, \\
&i\in T_A^r, r=1,...,M.
\end{aligned} \tag{28}$$

$$\begin{aligned}
\frac{\partial L}{\partial w_{Bsub}^r}=&\gamma\gamma_B w_{Bsub}^r-\sum_{T_B^r}^{N}\sum_{i\in T_B^r}\mu_i^B y_i\phi_r^B(x_i^B)=0, \\
&i\in T_B^r, r=1,...,N.
\end{aligned} \tag{29}$$

$$\begin{aligned}
\frac{\partial L}{\partial \xi_i^A}=&C_A+C\xi_i^B-(\alpha_i^A+\lambda_i^A+\mu_i^A+\beta_i^A)=0, \\
&i=1,...,l.
\end{aligned} \tag{30}$$

$$\begin{aligned}
\frac{\partial L}{\partial \xi_i^B}=&C_B+C\xi_i^A-(\alpha_i^B+\lambda_i^B+\mu_i^B+\beta_i^B)=0, \\
&i=1,...,l.
\end{aligned} \tag{31}$$

$$\alpha_i^A(1-\xi_i^A-y_i(w_A\cdot\phi_A(x_i^A)))=0, i=1,...,l. \tag{32}$$

$$\alpha_i^B(1-\xi_i^B-y_i(w_B\cdot\phi_B(x_i^B)))=0, i=1,...,l. \tag{33}$$

$$\lambda_i^A(y_i(w_B\cdot\phi_B(x_i^B))-\xi_i^A)=0, i=1,...,l. \tag{34}$$

$$\lambda_i^B(y_i(w_A\cdot\phi_A(x_i^A))-\xi_i^B)=0, i=1,...,l. \tag{35}$$

$$\xi_i^A\beta_i^A=0, \xi_i^B\beta_i^B=0, i=1,...,l. \tag{36}$$

Substituting the results into Formula Eq.(25), the following optimization form can be obtained :

$$\begin{aligned}
\min\quad &\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}((\alpha_j^A-\lambda_j^B)y_iK_A(x_i^A,x_j^A)(\alpha_i^A-\lambda_i^B)y_j \\
&+\frac{1}{\gamma}(\alpha_i^B-\lambda_i^A)y_iK_B(x_i^B,x_i^B)(\alpha_j^B-\lambda_j^B)) \\
&+\frac{1}{2}(\frac{1}{\gamma_A}\sum_{r=1}^{M}\sum_{j\in T_A^r}\mu_i^A y_iK_{Asub}^r(x_i^A,x_j^A)\mu_j^A y_j \\
&+\frac{1}{\gamma\gamma_B}\sum_{r=1}^{M}\sum_{j\in T_B^r}\mu_i^B y_iK_{Bsub}^r(x_i^B,x_j^B)\mu_j^B y_j) \\
&+\frac{1}{C}\sum_{i=1}^{l}(\alpha_i^A+\lambda_i^A+\mu_i^A+\beta_i^A-C_A) \\
&(\alpha_i^B+\lambda_i^B+\mu_i^B+\beta_i^B-C_B) \\
&-\sum_{i=1}^{l}(\alpha_i^A+\alpha_i^B) \\
s.t.\quad &\alpha_i^A,\alpha_i^B,\lambda_i^A,\lambda_i^B,\mu_i^A,\mu_i^B,\beta_i^A,\beta_i^B\geq 0.
\end{aligned} \tag{37}$$

In Eq.(37), according to the conclusion $\xi_i^A, \xi_i^B \geq 0$, $\alpha_i^A + \lambda_i^A + \mu_i^A - C\xi_i^B \leq C_A$ and $\alpha_i^B + \lambda_i^B + \mu_i^B - C\xi_i^A \leq C_B$ participate as constraints. Therefore, the optimization problem can be transformed into Eq.(38):

$$\begin{aligned}
\min \quad & \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}((\alpha_i^A - \lambda_i^B)y_i K_A(x_i^A, x_j^A)(\alpha_j^A - \lambda_j^B)y_j \\
& + \frac{1}{\gamma}(\alpha_i^B - \lambda_i^A)y_i K_B(x_i^B, x_j^B)(\alpha_j^B - \lambda_j^A)y_j) \\
& + \frac{1}{2}(\frac{1}{\gamma_A}\sum_{r=1}^{M}\sum_{j\in T_A^r}\mu_i^A y_i K_{Asub}^r(x_i^A, x_j^A)\mu_j^A y_j \\
& + \frac{1}{\gamma\gamma_B}\sum_{r=1}^{M}\sum_{j\in T_B^r}\mu_i^B y_i K_{Bsub}^r(x_i^B, x_j^B)\mu_j^B y_j) \\
& - \sum_{i=1}^{l}(\alpha_i^A + \alpha_i^B) + \frac{1}{C}\sum_{i=1}^{l}\xi_i^A\xi_i^B, \\
s.t. \quad & \alpha_i^A + \lambda_i^A + \mu_i^A - C\xi_i^B \leq C_A, \\
& \alpha_i^B + \lambda_i^B + \mu_i^B - C\xi_i^A \leq C_B, \\
& \alpha_i^A, \alpha_i^B, \lambda_i^A, \lambda_i^B, \mu_i^A, \mu_i^B, \xi_i^A, \xi_i^B \geq 0.
\end{aligned} \tag{38}$$

Like SL-PSVM-2V, quadratic convex programming is used to solve the problem. Through the quadratic programming problem Eq.(38), solving the optimal parameters $\alpha_i^{A*}, \alpha_i^{B*}, \beta_i^{A*}, \beta_i^{B*}, \lambda_i^{A*}, \lambda_i^{B*}, \mu_i^{A*}, \mu_i^{B*}$ in the best case w_A^* and w_B^*:

$$w_A^* = \sum_{i=1}^{l}(\alpha_i^{A*}y_i - \lambda_i^{B*}y_i)\phi_A(x_i^A), \tag{39}$$

$$w_B^* = \sum_{i=1}^{l}(\alpha_i^{B*}y_i - \lambda_i^{A*}y_i)\phi_B(x_i^B). \tag{40}$$

After getting the optimal w_A^* and w_B^*, the labels for predicting new samples (x^A, x^B) from view A and view can be derived from the following formula:

$$f_A = sign(f_A(x^A)) = sign({w_A^*}^\top \phi_A(x_i^A)), \tag{41}$$

$$f_B = sign(f_B(x^B)) = sign({w_B^*}^\top \phi_B(x_i^B)). \tag{42}$$

For multi-view final predictions:

$$\begin{aligned}
f &= sign(\frac{1}{2}f_A(x^A) + \frac{1}{2}f_B(x^B)) \\
&= sign(\frac{1}{2}{w_A^*}^\top \phi_A(x^A) + \frac{1}{2}{w_B^*}^\top \phi_B(x^B)).
\end{aligned} \tag{43}$$

The process of SL-MCPK is given in algorithm 2:

Algorithm 2 QP Algorithm for SL-MCPK

Require: Data set: $S = \{x_i^A, x_i^B, y_i\}_{y=1}^{l} = \{(x_i^A;\ 1), (x_i^B;\ 1)\}_{i=1}^{l}, y_i \in \{+1, -1\}$; Sub-views element index: $T = \{(T_A^r, T_B^r)\}$; Initial parameters: $\gamma, \gamma_A, \gamma_B, C_A, C_B, C \geq 0$.

Ensure: Decision functions f_A,f_B,f.

1: Select kernel function: kernels function of view A and view B: $K_A(x_i^A, x_j^A)$, $K_B(x_i^B, x_j^B)$, The kernel of the sub-views of A:$K_{Asub}^r(x_i^A, x_i^A)$, kernel of sub-views of B:$K_{Bsub}^r(x_i^B, x_i^B)$, And initializing kernel parameters.

2: Create and solve quadratic programming problem Eq.(38) and using cross validation to determine the optimal parameters.

3: Solving quadratic programming Eq.(38) and retaining optimal result parameters $\alpha_i^{A*}, \alpha_i^{B*}, \beta_i^{A*}, \beta_i^{B*}, \lambda_i^{A*}, \lambda_i^{B*}, \mu_i^{A*}, \mu_i^{B*}$, Get the optimal weight w_A^* and w_B^* by substituting formula Eq.(40) and Eq.(41).

4: The final decision function is solved by parameters w_A^* and w_B^*:

$$\begin{aligned}
f_A &= sign(f_A(x^A)) = sign({w_A^*}^\top \phi_A(x_i^A)), \\
f_B &= sign(f_B(x^B)) = sign({w_B^*}^\top \phi_B(x_i^B)), \\
f &= sign(\frac{1}{2}f_A(x^A) + \frac{1}{2}f_B(x^B)).
\end{aligned}$$

4 Model comparison and transformation method

In this section, the classification models SL-PSVM-2V and SL-MCPK with sub-view structure are compared with three related multi-view classification models: PSVM-2V, MCPK and SVM-2K.

As an extension of the multi-view learning structure, the sub-view learning structure transforms the original multi-view model in the objective function and constraint term. The expansion of the structure does not affect the compliance of the original model with the principles of consistency and complementarity. PSVM-2V integrates LUPI learning framework and KCCA style consistency constraints into a unified framework. MCPK can be used as an improved and more effective version of PSVM-2V and SVM-2K in multi-view learning. Therefore, SL-MCPK can be considered as an evolutionary version of SL-PSVM-2V. For the complementarity principle, SVM-2K ignores it. In contrast, PSVM-2V and SL-PSVM-2V implement it by connecting multiple views and privileged information, and MCPK and SL-MCPK-2V also leverage the LUPI concept. For each individual view, it can receive complementary information from other views. Several models can be transformed into each other through the transformation of the model. These transformations can also show that the new model can complete the tasks that the old model cannot consider while having the characteristics of the old model. The transformation method is shown in Fig.4.

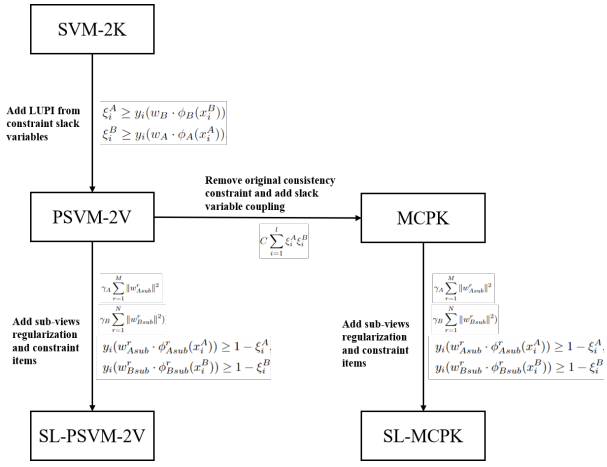


Fig. 4 The transformation methods among the five multi-view models

5 Experiments

In this section, we show the experimental results of the new model. The experiment is carried out on multiple data sets, including 15 sets of data set constructed from **Digits**, 40 sets of data set constructed from **Corel**, and **Ionosphere** noise data set constructed to explore the model noise sensitivity. The experimental environment is a computer with i7-6500 CPU and 8 GB memory. The program runs in MATLAB 2018b in Windows 10 operating system. All the comparison algorithms are solved by CVX convex optimization toolbox (Grant and Boyd, 2014).

5.1 Datasets and experimental design

5.1.1 Datasets

Digits dataset extracted from Dutch utility maps. It consists of (0-9) handwritten digits, each of which consists of 200 examples digitizing binary images (Sun et al, 2015). These data are represented in the following six views: 1) mfeat-fou: 76 Fourier coefficients of the character shapes. 2) mfeat-fac: 216 profile correlations. 3) mfeat-kar: 64 Karhunen-Love coecients. 4) mfeat-pix: 240 pixels averages in a 2 by 3 window. 5) mfeat-zer: 47 Zernike moments. 6) mfeat-mor: 6 morphological features. we construct two classes (positive class and negative class) from the data set, with the original label of 0–4 as the positive class and 5–9 as the negative class. Each class randomly selected 100 samples, each group of experiments a total of 200 samples. The privileged information used to guide the partition of sub-views is the different numeric labels in each class (positive and negative): for example, referring to $\{\{1,7\},\{2,4,6\}\{0,3,5,7\},\{8,9\}\}$, the two views are divided into four sub-views.

Corel consists of 599 classes with 97-100 images representing semantic topics such as elephants, roses, horses, etc. To be exact, Category 238 contains 97 samples, Categories 342 and 376 contain 99 samples, and the rest contain 100 samples. Each image has eight pre-extracted feature representations that can be tested as eight different views. Three different pre-extracted features (Color Structure, Color Layout, Dominant Color) are selected from this data as different views for experiments (Eidenberger, 2004). Using the data of the first 200 classes in 599 classes (each class is 100 samples). 20 experimental groups are composed of 10 classes, and the first 5 classes in

each group are set as positive classes, and the last 5 classes are set as negative classes. We used the strategy of Color Structure vs. Color Layout, Dominant Color vs. Color Layout for experiments. The sub-views of another view are divided by the instruction of the privileged information of different views. For example, the information division serial number group $\{\{1,2,7\},\{3,4,5,6\},\{8,9\}\}$ in view A guides view B to divide three sub-views, and the privileged information serial number group $\{\{1,7\},\{2,3,6,8\},\{4,9\},\{5\}\}$ in view B guides view A to divide four sub-views.

Ionosphere contains 351 samples (225 positive samples and 126 negative samples). The positive example is the example of the radar return, which shows a certain structure in the ionosphere, the signal through the ionosphere, and the negative example is not. Each method can perform well in noise-free Ionosphere data. Therefore, this data is suitable for exploring the anti-noise ability of existing methods and new methods. Two types of experiments are carried out on the Ionosphere data set. Gaussian distribution and standard deviation are used to generate noisy samples, namely, the standard deviation σ_i of the i-th feature in the training data is calculated, and the Gaussian noise with a range of $[0,\sigma_i]$ and a size of $(l,1)$ .The Gaussian noise is randomly added to the sample data as an increment in proportion $[0,0.05,0.1,0.15,0.2,0.25,0.3]$, Therefore, seven groups of training sets with different proportions of noise can be generated. In addition, we randomly change the label symbol in proportion to compare the performance of different models in the experiment with noise labels.

5.1.2 Benchmark methods

SL-PSVM-2V and SL-MCPK are compared with the following benchmark methods.

SVM+_A and **SVM+_B**: SVM+ method uses the non-negative correction function determined by privileged information to replace the slack variable of standard SVM. View B is used as privileged information in SVM+_A, and view A is used as privileged information in SVM+_B.

SVM-2K : The SVM-2K method combines KCCA (Kernel Canonical Correlation Analysis) with two SVM models for two-view classification. It is the earliest multi-view SVM model and the basis of this series of classification methods.

PSVM-2V: PSVM-2V model uses privileged information to meet two principles of multi-view learning.

MCPK: MCPK method uses coupling term and LUPI framework for two-view classification

5.1.3 Standard of comparison

Average accuracy(Acc.) and average standard deviation (Std.) are derived from 10 replicate experiments to measure the performance of different methods and give average rankings. Furthermore, the receiver operating characteristic(ROC) curve is used to demonstrate the performance improvement of the new method (Beck and Shultz, 1986). The average CPU running time of quadratic programming on CVX is selected to compare the computational complexity of different methods.

5.1.4 Parameter Setting

Grid search strategy and five-fold cross validation method are used to select the best parameters. Gaussian RBF kernel function $K(x_i,x_j) = \exp(-\frac{\|x_i-x_j\|^2}{2\sigma})$ as the kernel function of each model. The kernel parameter σ for the Gaussian RBF kernel function is selected from $[10^{-3},10^{-2},10^{-1},1,10,10^2,10^3]$, set $C_A = C_B = C$ and select over $[10^{-3},10^{-2},10^{-1},1,10,10^2,10^3]$. The tradeoff parameter γ, γ_A and γ_B tuned in the range $[10^{-3},10^{-2},10^{-1},1,10,10^2,10^3]$.

5.2 Experimental results

5.2.1 Performance on Digits

The performance on the Digits dataset is shown in Table 2. The sub-view partition method in these 15 groups of experiments is based on the features of the view itself. The sub-view learning versions of SL-MCPK and SL-PSVM-2V are more competitive than the original MCPK and PSVM-2V. SL-MCPK model has the highest average accuracy. In 15 groups of experiments, the average ranking was 1.1538, only one ranking is 3, and its ranking is the first with absolute advantage. SL-PSVM-2V of the sub-view learning version is also better than PSVM-2V in accuracy. The experimental results are shown in Fig.5. in the form of the line graph, which can more intuitively compare the differences of each model.

Table 2 Performance on Digits (Acc.% (Std.%))

	SVM+_A	SVM+_B	SVM-2K	PSVM-2V	SL-PSVM-2V	MCPK	SL-MCPK
fac vs. fou	91.46(2.38)	87.57(7.23)	95.93(3.03)	95.98(3.89)	97.48(1.69)	97.51(1.81)	97.51(1.87)
fac vs. kar	96.56(2.75)	95.43(2.28)	96.40(3.01)	97.95(2.11)	98.50(1.37)	98.89(2.21)	98.96(1.45)
fac vs. mor	95.11(5.03)	82.98(4.60)	94.53(4.03)	95.50(3.68)	96.57(3.70)	96.99(3.27)	97.00(1.12)
fac vs. pix	95.49(3.30)	96.55(2.79)	95.44(2.19)	97.04(2.05)	98.01(3.26)	97.54(1.65)	98.54(1.34)
fac vs. zer	94.04(2.14)	91.48(2.95)	96.45(2.35)	94.13(3.54)	98.43(1.43)	97.91(3.21)	97.97(2.10)
fou vs. kar	95.92(3.23)	96.40(3.15)	94.00(1.21)	96.97(1.01)	96.65(2.62)	97.99(2.07)	98.44(2.22)
fou vs. mor	87.89(4.08)	87.69(7.78)	90.79(5.05)	91.17(4.66)	93.15(5.41)	95.00(1.75)	95.35(3.32)
fou vs. pix	93.77(4.22)	95.51(1.79)	94.28(1.90)	97.41(1.57)	95.87(3.27)	96.70(2.16)	97.85(2.98)
fou vs. zer	91.54(2.03)	91.97(5.74)	93.99(1.26)	94.66(2.95)	96.59(2.57)	96.01(1.16)	97.01(2.36)
kar vs. mor	92.56(1.19)	84.32(5.64)	93.28(3.62)	92.16(2.29)	97.07(2.01)	96.98(2.15)	97.08(1.85)
kar vs. pix	95.56(3.82)	97.06(2.96)	95.75(2.61)	96.57(2.18)	98.56(1.32)	98.17(1.70)	98.59(2.09)
kar vs. zer	95.07(1.50)	93.66(2.31)	95.96(1.45)	97.01(2.11)	97.87(2.72)	97.81(3.34)	97.94(2.15)
mor vs. pix	88.94(5.62)	94.64(5.98)	92.57(3.86)	95.28(3.47)	96.61(1.22)	96.77(3.02)	97.97(2.42)
mor vs. zer	89.51(6.50)	9 87(4.13)	94.24(1.94)	94.40(2.02)	94.76(2.36)	94.57(6.19)	97.45(0.38)
pix vs. zer	97.50(1.77)	94.50(1.13)	95.47(2.75)	96.01(2.83)	98.02(2.08)	98.01(1.39)	98.51(1.36)
Avg. Acc.	93.39	92.04	94.61	95.48	96.95	97.08	97.78
Avg. Rank.	6.1538	5.9231	5.6154	4.1538	2.6923	2.3077	1.1538

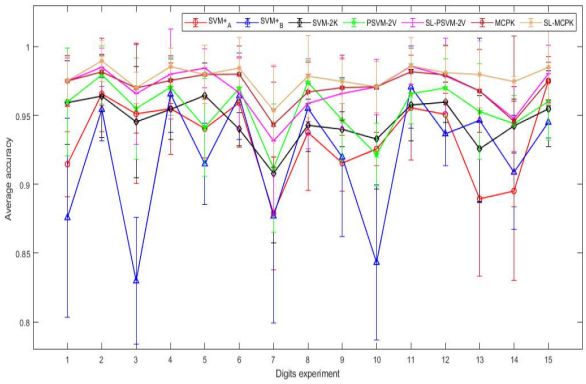


Fig. 5 The performance of 15 groups of experiments of various methods on Digits

5.2.2 Performance on Corel

On the Corel dataset, the division of sub-views is formed by the mutual guidance of two views. We have done two sets of multi-view experiments: Color structure vs. Color Layout and Dominant Color vs. Color Layout, respectively. The experimental results of Color structure vs. Color Layout are shown in Table 3. In 20 sets of experiments, SL-PSVM-2V and SL-MCPK using sub-view learning have achieved good results in the average accuracy index. SL-MCPK ranks first and SL-PSVM-2V ranks second. Their average rankings are better than other models. SL-MCPK does not have the highest ranking in Ex6, Ex11 and Ex13, but it is close to the best model performance. The experimental results of Dominant Color vs. Color Layout are shown in Table 4. The average accuracy of SL-MCPK on this dataset is 81.04%, with an average ranking of 1.3846, which is a leading result in many methods. The second is MCPK, with an average ranking of 2.3077. This is because the coupling structure has achieved good results in the experiment. In the experiment, SVM-2K does not take into account the complementarity of views, and SVM+ considered the view more simply, so their classification performance could not achieve the desired results. For the original versions of PSVM-2V and MCPK, the corresponding multi-view classification model of sub-view versions has achieved good performance scores.In order to facilitate comparison, the experiment results are displayed visually in Fig.6 and Fig.7.

5.3 ROC curve and AUC

A good classifier will try to minimize two types of errors: false positive rate and true positive

Table 3 Performance on experiments of Color structure vs. Color Layout. (Acc.% (Std.%))

	SVM+_A	SVM+_B	SVM-2K	PSVM-2V	SL-PSVM-2V	MCPK	SL-MCPK
Ex.1	59.84(4.31)	58.61(2.91)	68.42(4.29)	68.65(6.78)	72.50(1.22)	68.42(3.95)	72.98(1.96)
Ex.2	59.57(4.83)	61.07(3.48)	70.27(8.92)	74.89(7.58)	84.32(1.30)	78.38(4.50)	85.37(1.19)
Ex.3	70.59(1.22)	57.02(1.32)	67.46(2.25)	75.61(7.48)	76.32(2.13)	73.17(6.47)	76.79(1.22)
Ex.4	71.41(4.97)	62.69(4.14)	71.72(6.22)	73.41(3.13)	74.36(4.56)	80.56(9.74)	74.36(1.98)
Ex.5	63.10(2.66)	62.71(6.01)	71.83(5.21)	73.68(8.32)	80.83(1.99)	79.55(9.73)	81.82(1.11)
Ex.6	66.54(2.82)	64.29(5.42)	71.94(6.98)	76.19(4.54)	77.42(4.95)	80.00(5.87)	78.57(1.16)
Ex.7	59.49(0.89)	67.69(4.70)	75.68(9.31)	76.61(2.90)	80.20(4.39)	77.54(5.93)	81.41(5.16)
Ex.8	60.61(0.91)	59.09(4.26)	74.72(5.99)	77.51(9.03)	70.11(2.65)	77.52(8.40)	78.58(1.55)
Ex.9	64.27(1.53)	57.84(0.58)	72.97(2.30)	74.36(4.14)	81.58(2.57)	76.32(5.19)	82.05(2.03)
Ex.10	65.63(3.31)	60.94(1.54)	74.81(5.71)	78.05(8.74)	80.03(2.98)	80.03(2.03)	80.56(2.81)
Ex.11	70.03(4.17)	72.31(2.14)	76.48(9.93)	77.27(9.19)	77.92(1.11)	79.55(4.94)	77.27(2.18)
Ex.12	59.68(1.95)	57.36(2.73)	67.63(6.93)	70.27(6.68)	72.97(1.02)	74.37(1.20)	78.38(9.29)
Ex.13	61.95(6.01)	65.71(0.71)	72.42(7.24)	71.43(7.02)	82.58(3.71)	71.79(7.80)	82.23(1.19)
Ex.14	60.01(0.13)	72.06(4.77)	79.49(9.10)	80.66(6.49)	81.47(1.02)	79.49(9.28)	81.85(1.32)
Ex.15	71.43(1.12)	62.45(5.56)	72.50(2.25)	81.08(2.83)	73.68(1.34)	75.04(2.07)	85.37(3.69)
Ex.16	66.67(3.62)	64.25(4.71)	76.92(1.74)	77.01(5.33)	81.68(5.99)	76.92(5.62)	82.93(2.94)
Ex.17	72.73(3.62)	61.62(3.50)	67.77(8.99)	72.95(2.15)	71.79(2.99)	71.43(2.19)	74.29(1.01)
Ex.18	69.26(0.26)	71.64(9.91)	79.53(2.56)	81.29(2.55)	83.33(2.55)	83.33(4.19)	83.33(2.55)
Ex.19	81.08(2.15)	74.91(5.56)	89.90(3.27)	94.74(4.70)	95.55(2.42)	97.37(5.90)	97.37(2.50)
Ex.20	69.70(6.14)	70.18(6.08)	76.32(4.61)	78.02(3.05)	84.68(1.36)	76.32(4.88)	85.41(2.34)
Avg. Acc.	63.12	69.36	70.48	73.24	78.38	75.53	79.48
Avg. Rank.	6.6923	5.6154	4.7692	3.8462	2.6154	3.3077	1.1538

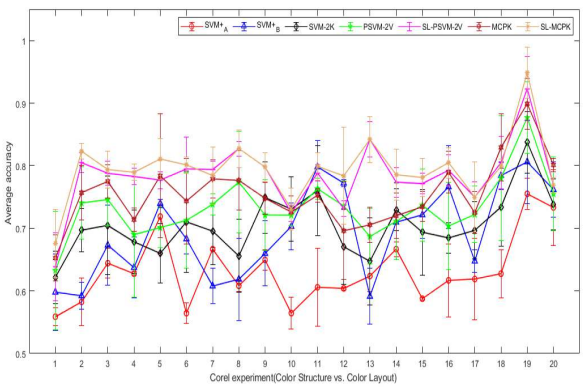


Fig. 6 20 groups of experimental performance of various methods on Corel (Color structure vs. Color Layout)

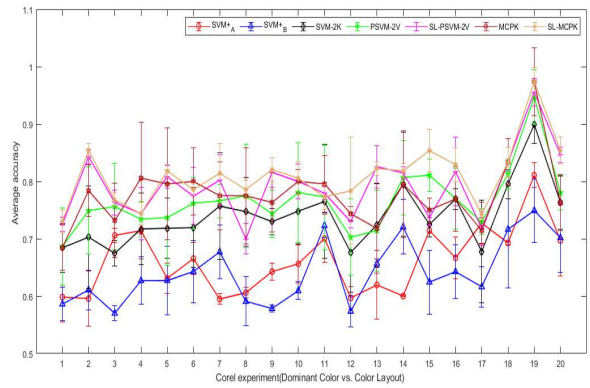


Fig. 7 20 groups of experimental performance of various methods on Corel (Dominant Color vs. Color Layout)

rate. Two corresponding errors can be obtained by changing the threshold, and then a ROC curve can be obtained. In order to comprehensively compare and show the performance improvement of the sub-view learning model based on the original model, we compare the ROC curves of PSVM-2V, SL-PSVM-2V, MCPK and SL-MCPK. Area Under Curve(AUC) under ROC curve can measure the performance of a classifier. The higher the AUC value is, the better the classification performance is. In Fig.8, the results of the first 8 groups of digits experiments are shown. In Fig.9 and Fig.10, the ROC and AUC of the first 8 groups of Corel dataset multi-view experiment are shown.

Table 4 Performance on experiments of Dominant Color vs. Color Layout. (Acc.% (Std.%))

	SVM+_A	SVM+_B	SVM-2K	PSVM-2V	SL-PSVM-2V	MCPK	SL-MCPK
Ex.1	55.88(1.46)	59.80(6.14)	62.16(4.18)	63.27(9.46)	63.83(5.41)	65.24(3.66)	67.57(5.41)
Ex.2	58.25(3.79)	59.22(3.79)	69.70(3.46)	74.04(5.80)	80.49(1.56)	75.70(4.32)	82.32(1.22)
Ex.3	64.42(2.35)	67.31(2.35)	70.48(7.82)	74.60(4.15)	78.79(1.90)	77.50(2.68)	79.39(2.91)
Ex.4	62.75(0.29)	63.73(0.29)	67.84(5.11)	69.01(9.96)	78.29(1.32)	71.38(1.88)	78.95(1.32)
Ex.5	71.88(6.02)	73.96(6.02)	66.01(4.78)	70.11(3.17)	77.70(1.35)	78.38(9.93)	81.08(3.32)
Ex.6	56.44(1.66)	68.32(1.66)	71.03(8.08)	71.27(7.64)	79.49(5.11)	74.36(6.83)	80.13(1.28)
Ex.7	66.67(0.23)	60.78(0.23)	69.55(5.33)	73.84(1.67)	79.40(1.28)	77.89(3.03)	78.44(4.58)
Ex.8	60.82(0.92)	61.86(0.92)	65.57(5.83)	77.38(8.13)	82.74(1.19)	77.66(4.76)	82.74(3.01)
Ex.9	64.95(1.67)	65.98(1.67)	74.91(5.68)	72.12(5.63)	79.79(2.32)	74.79(2.55)	79.79(2.32)
Ex.10	56.44(2.53)	70.30(2.53)	73.05(5.16)	72.10(1.63)	72.44(1.28)	72.59(2.56)	73.08(3.31)
Ex.11	60.58(6.25)	79.81(6.25)	76.04(7.15)	76.33(1.73)	78.85(1.28)	75.27(1.06)	79.89(2.21)
Ex.12	60.40(0.23)	77.23(0.23)	67.06(6.01)	73.58(3.93)	73.13(1.25)	69.61(7.78)	78.38(7.78)
Ex.13	62.37(2.52)	59.14(2.52)	64.73(6.94)	68.70(4.44)	84.21(2.86)	70.56(2.82)	84.21(3.68)
Ex.14	66.67(1.19)	70.97(1.19)	72.97(3.37)	71.11(6.14)	77.38(2.38)	72.02(3.59)	78.57(4.12)
Ex.15	58.76(0.23)	72.16(0.23)	69.38(6.86)	73.70(5.76)	77.16(1.63)	73.44(2.38)	78.12(3.07)
Ex.16	61.68(5.84)	76.64(5.84)	68.46(2.41)	70.36(6.92)	79.27(2.44)	78.95(3.36)	80.49(2.44)
Ex.17	61.90(6.47)	64.76(6.47)	69.68(4.58)	72.22(4.54)	75.02(2.41)	72.50(3.37)	75.01(5.61)
Ex.18	62.75(3.87)	78.43(3.87)	73.42(6.22)	78.05(9.96)	80.49(4.14)	82.93(5.41)	79.88(1.22)
Ex.19	75.51(2.46)	80.61(2.46)	83.75(4.91)	87.68(5.68)	92.31(5.17)	89.90(4.03)	94.87(4.03)
Ex.20	73.33(6.09)	76.19(6.09)	73.85(4.10)	75.52(5.93)	76.88(1.25)	80.13(1.09)	76.88(1.25)
Avg. Acc.	66.17	64.22	73.93	76.68	79.15	77.84	81.04
Avg. Rank.	6.3077	6.5385	4.9231	3.6154	2.9231	2.3077	1.3846

The model parameters of the prediction category are the best results in the quintuple. It can be seen in the diagram that the model with sub-view learning structure performs better than the model without sub-view learning structure.

5.4 Average computer time

In this section, we give the average training time and average prediction time and solve all models by CVX tools solver. The experimental data is constructed based on Ionosphere, the view A dimension is 54 and the view B dimension is 25, the data length is 200, and the test set length is 40. All models adopt RBF Gaussian kernel function and run fairly in the same solution environment. According to Fig.11, since the coupling term replaces the consistency term, MCPK consumes less training time than other SVM-based multi-view benchmark test methods, and the data dimension used by SVM+ method is the size of a single view, so it is not included. MCPK is an effective learning model with less time consumption. In the corresponding sub-view learning method, SL-MCPK also requires less solving time than SL-PSVM-2V. However, due to the addition of constraints and regularization terms, the sub-view method will linearly increase the scale of quadratic programming. However, the solving time of SL-MCPK is second only to MCPK method, and it is still in the front position compared with SVM-2K and PSVM-2V. In the category prediction time consumption, the sub-view structure is not used in the classification decision stage, so the solution time is consistent with other multi-view methods.

5.5 Performance on noisy data

The experimental results on the noise data are shown in Fig.12. Whether the noise exists in the feature data or in the label, the sub-view version model SL-MCPK and SL-PSVM-2K are less sensitive to noise than other multi-view support vector machine models without sub-view structure. In the absence of noise, all models have good performance. With the increase of noise ratio, the accuracy of the multi-view support vector machine model of the sub-view learning

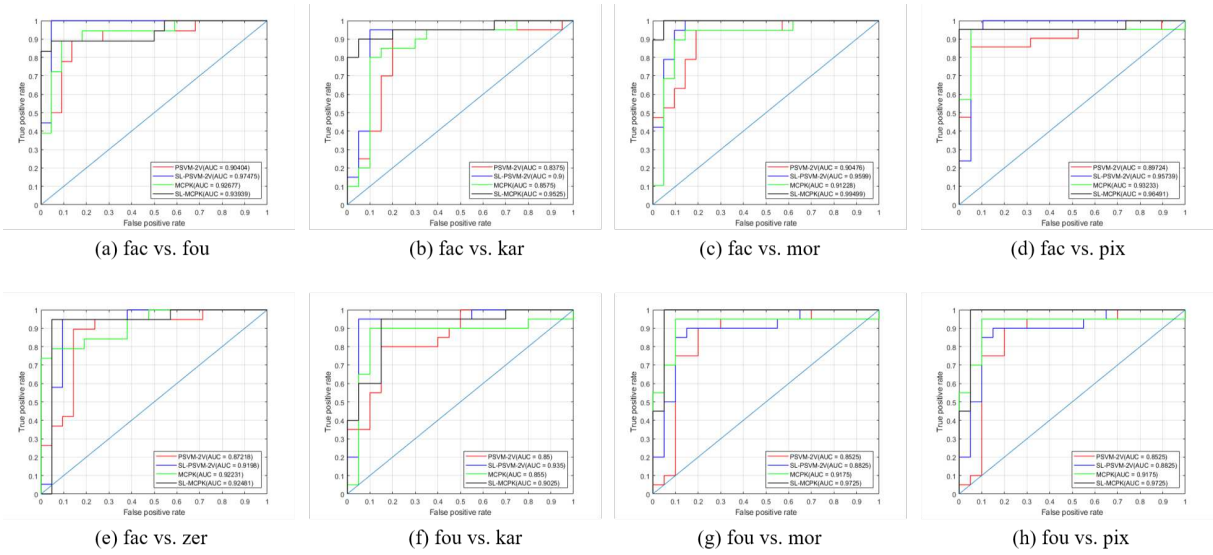


Fig. 8 ROC curves of AUC values corresponding to PSVM-2V, SL-PSVM-2V, MCPK and SL-MCPK on the first 8 datasets from Digits

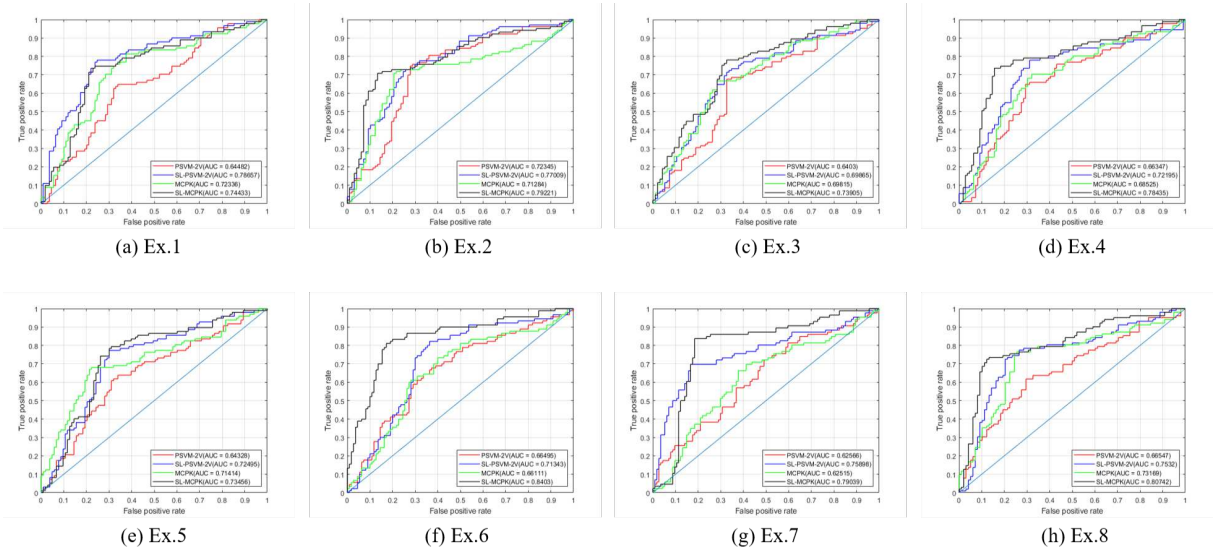
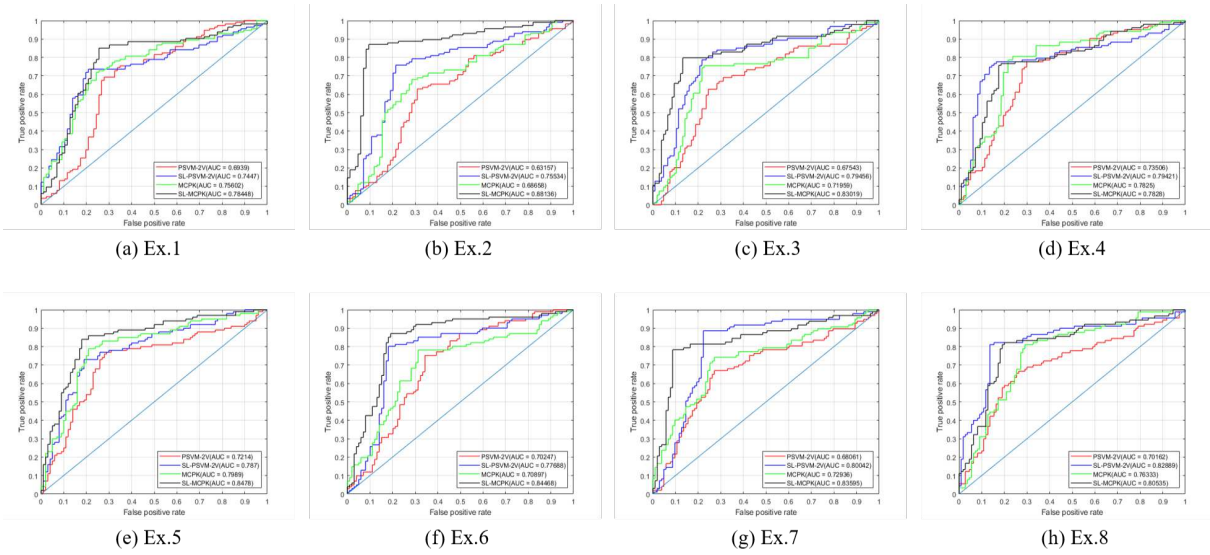


Fig. 9 ROC curves of AUC values corresponding to PSVM-2V, SL-PSVM-2V, MCPK and SL-MCPK on the first 8 datasets from Corel (Color structure vs. Color Layout)

version decreases slightly. It is worth noting that the PSVM-2V model has better anti-noise ability than the MCPK model. SL-PSVM-2V inherits this character as a sub-view learning version of PSVM-2V, and the SL-MCPK model has the learning strategy of the sub-view structure, which to some extent makes up for the noise sensitivity of MCPK.

5.6 Non-parametric statistical test

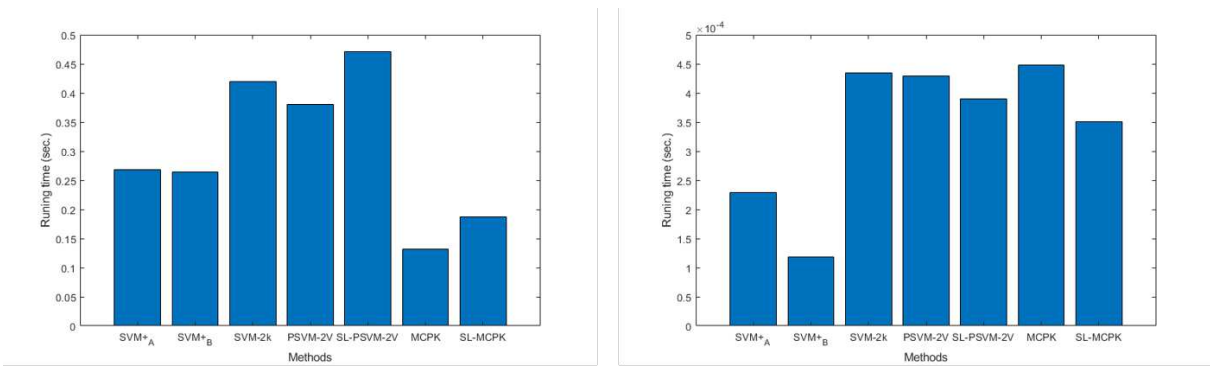
We use the nonparametric statistical Wilcoxon test to further study the performance differences between the proposed sub-view learning method and other methods. The test ranks the performance differences between the two classifiers of each data set, ignores the positive and negative symbols, and compares the rankings of positive



(a) Ex.1 (b) Ex.2 (c) Ex.3 (d) Ex.4

(e) Ex.5 (f) Ex.6 (g) Ex.7 (h) Ex.8

Fig. 10 ROC curves of AUC values corresponding to PSVM-2V, SL-PSVM-2V, MCPK and SL-MCPK on the first 8 datasets from Corel (Dominant Color vs. Color Layout)



(a) average running time of model training (b) average running time of model prediction

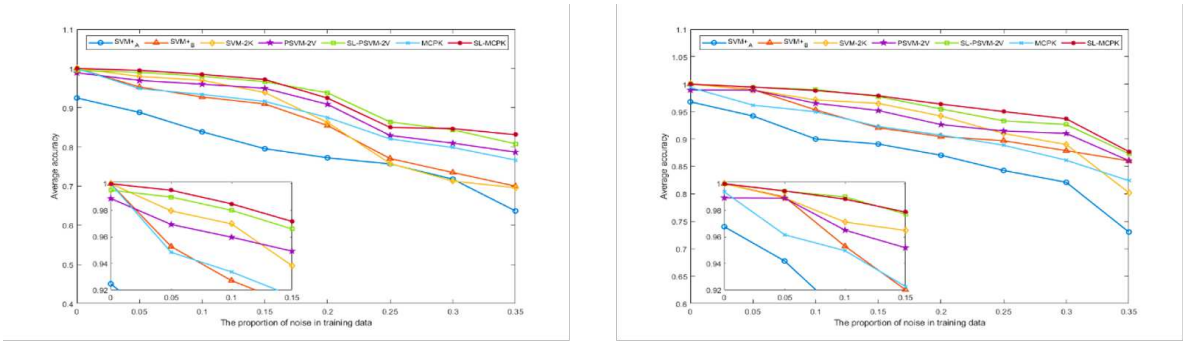
Fig. 11 Average running time of proposed method and the baselines

and negative differences (Wilcoxon, 1992; Demšar, 2006). Let d_i be the difference between the performances (Acc.) of the two classifiers on i-th out of N datasets. R^+ be the sum of the ranks of the data sets of one algorithm over another, and R^- be the sum of the ranks of the opposite algorithms. Ranks of are $d_i = 0$ split evenly among the sums.

$$
\begin{aligned}
R^+ &= \sum_{d_i>0} rank(d_i) + \frac{1}{2}\sum_{d_i=0} rand(d_i), \\
R^- &= \sum_{d_i<0} rank(d_i) + \frac{1}{2}\sum_{d_i=0} rand(d_i).
\end{aligned}
\tag{44}
$$

Let T be the smaller of the sums, $T = \min(R^+, R^-)$, The wilcoxon test value p-value obtained by z-value ($z\text{-}value = \frac{T-\frac{1}{4}N(N+1)}{\frac{1}{24}N(N+1)(2n+1)}$) between methods.

We all know that if the test value is less than the confidence level of $\alpha = 0.05$, there is a significant difference between the proposed method and the baseline. From the information in Table 5, it can be seen that the performance of the proposed method and the comparison method are significantly improved. Compared with two models with sub-view structure, SL-MCPK has better performance.



(a) average running time of model training

(b) average running time of model prediction

Fig. 12 Average running time of proposed method and the baselines

Table 5 Wilcoxon signed ranks test result($\alpha = 0.05$)

Comparison	R^+	R^-	p-value
SL-PSVM-2V vs. SVM+	450	0	1.3328e-12
SL-PSVM-2V vs. SVM-2K	435	5	1.4933e-09
SL-PSVM-2V vs. PSVM-2V	421	9	1.8570e-07
SL-PSVM-2V vs. MCPK	230	220	0.036
SL-PSVM-2V vs. SL-MCPK	121	329	1.0928e-05
SL-MCPK vs. SVM+	450	0	1.3328e-12
SL-MCPK vs. SVM-2K	450	0	1.8189e-12
SL-MCPK vs. PSVM-2V	419	31	3.6379e-12
SL-MCPK vs. MCPK	422	28	4.2559e-06

6 Conclusion and future works

This paper presents a multi-view support vector machine strategy called sub-view learning. This method extends the model structure of multi-view support vector machine based on privileged information learning. The sub-views are divided by considering the privileged information in the view, which can be solved by the dual problem of quadratic programming. Through the given model transformation method, the specific models SL-PSVM-2V and SL-MCPK are the sub-view learning versions of PSVM-2V and MCPK, respectively. We conducted experiments on 55 multi-view datasets to verify that the new method has better performance and can effectively use more comprehensive data feature information to guide the generation of classifier. At the same time, the experimental results on noisy data sets show that the new method has better anti-noise ability. In the future work, we plan to consider more basis for dividing sub-views and try to apply the multi-view learning containing sub-views to the regression task.

Acknowledgments. This research is supported by Tianjin "Project + Team" Key Training Project under Grant No. XC202022 and Tianjin Research Innovation Project for Postgraduate Students under Grant No. 2021YJSS095.

Declarations

Conflict of interest

The authors declare that they have no conflict of interest.

Human participants or animals

This article does not contain any studies with human participants or animals performed by any of the authors.

Authorship contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Qi Hao, Wenguang Zheng and Yingyuan Xiao. The first draft of the manuscript was written by Qi Hao and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

References

Ahmed OB, Lecellier F, Paccalin M, et al (2017) Multi-view visual saliency-based MRI classification for alzheimer's disease diagnosis. In: Seventh International Conference on Image Processing Theory, Tools and Applications, IPTA

2017, Montreal, QC, Canada, November 28 - December 1, 2017. IEEE, pp 1–6, https://doi.org/10.1109/IPTA.2017.8310118, URL https://doi.org/10.1109/IPTA.2017.8310118

Appice A, Malerba D (2016a) A co-training strategy for multiple view clustering in process mining. IEEE Trans Serv Comput 9(6):832–845. https://doi.org/10.1109/TSC.2015.2430327, URL https://doi.org/10.1109/TSC.2015.2430327

Appice A, Malerba D (2016b) A co-training strategy for multiple view clustering in process mining. IEEE Trans Serv Comput 9(6):832–845. https://doi.org/10.1109/TSC.2015.2430327, URL https://doi.org/10.1109/TSC.2015.2430327

Beck JR, Shultz EK (1986) The use of relative operating characteristic (roc) curves in test performance evaluation. Archives of pathology & laboratory medicine 110(1):13–20

Chao G, Sun S (2016) Consensus and complementarity based maximum entropy discrimination for multi-view classification. Inf Sci 367-368:296–310. https://doi.org/10.1016/j.ins.2016.06.004, URL https://doi.org/10.1016/j.ins.2016.06.004

Chao G, Sun J, Lu J, et al (2019) Multi-view cluster analysis with incomplete data to understand treatment effects. Inf Sci 494:278–293. https://doi.org/10.1016/j.ins.2019.04.039, URL https://doi.org/10.1016/j.ins.2019.04.039

Che Z, Liu B, Xiao Y, et al (2021) Twin support vector machines with privileged information. Inf Sci 573:141–153. https://doi.org/10.1016/j.ins.2021.05.069, URL https://doi.org/10.1016/j.ins.2021.05.069

Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research 7:1–30

Eidenberger H (2004) Statistical analysis of content-based mpeg-7 descriptors for image retrieval. Multimedia Systems 10(2):84–97

Farquhar JDR, Hardoon DR, Meng H, et al (2005) Two view learning: Svm-2k, theory and practice. In: Advances in Neural Information Processing Systems 18 [Neural Information Processing Systems, NIPS 2005, December 5-8, 2005, Vancouver, British Columbia, Canada], pp 355–362, URL https://proceedings.neurips.cc/paper/2005/hash/46b2644cbdf489fac0e2d192212d206d-Abstract.html

Grant M, Boyd S (2014) Cvx: Matlab software for disciplined convex programming, version 2.1

Guo Y, Xiao H, Kan Y, et al (2018) Learning using privileged information for hrrp-based radar target recognition. IET Signal Process 12(2):188–197. https://doi.org/10.1049/iet-spr.2016.0625, URL https://doi.org/10.1049/iet-spr.2016.0625

Han L, Jing X, Wu F (2018) Multi-view local discrimination and canonical correlation analysis for image classification. Neurocomputing 275:1087–1098

Houthuys L, Langone R, Suykens JAK (2018) Multi-view least squares support vector machines classification. Neurocomputing 282:78–88. https://doi.org/10.1016/j.neucom.2017.12.029, URL https://doi.org/10.1016/j.neucom.2017.12.029

Kim D, Seo D, Cho S, et al (2019) Multi-co-training for document classification using various document representations: Tf-idf, lda, and doc2vec. Inf Sci 477:15–29. https://doi.org/10.1016/j.ins.2018.10.006, URL https://doi.org/10.1016/j.ins.2018.10.006

Lapin M, Hein M, Schiele B (2014) Learning using privileged information: SV M+ and weighted SVM. Neural Networks 53:95–108. https://doi.org/10.1016/j.neunet.2014.02.002, URL https://doi.org/10.1016/j.neunet.2014.02.002

Li Y, Wang Y, Zhou J, et al (2018) Robust transductive support vector machine for multi-view classification. J Circuits Syst Comput 27(12):1850,185:1–1850,185:22. https://doi.org/10.1142/S0218126618501852, URL https://doi.org/10.1142/S0218126618501852

Li Y, Sun H, Yan W, et al (2021) R-CTSVM+: robust capped l_1-norm twin support vector machine with privileged information. Inf Sci 574:12–32. https://doi.org/10.1016/j.ins.2021.06.003, URL https://doi.org/10.1016/j.ins.2021.06.003

Liang L, Cherkassky V (2007) Learning using structured data: Application to fmri data analysis. In: Proceedings of the International Joint Conference on Neural Networks, IJCNN 2007, Celebrating 20 years of neural networks, Orlando, Florida, USA, August 12-17, 2007. IEEE, pp 495–499, https://doi.org/10.1109/IJCNN.2007.4371006, URL https://doi.org/10.1109/IJCNN.2007.4371006

Niu L, Wu J (2012) Nonlinear L-1 support vector machines for learning using privileged information. In: Vreeken J, Ling C, Zaki MJ, et al (eds) 12th IEEE International Conference on Data Mining Workshops, ICDM Workshops, Brussels, Belgium, December 10, 2012. IEEE Computer Society, pp 495–499, https://doi.org/10.1109/ICDMW.2012.79, URL https://doi.org/10.1109/ICDMW.2012.79

Sinoara RA, Sundermann CV, Marcacini RM, et al (2014) Named entities as privileged information for hierarchical text clustering. In: Desai BC, Almeida AM, Bernardino J, et al (eds) 18th International Database Engineering & Applications Symposium, IDEAS 2014, Porto, Portugal, July 7-9, 2014. ACM, pp 57–66, https://doi.org/10.1145/2628194.2628225, URL https://doi.org/10.1145/2628194.2628225

Sun S, Chao G (2013) Multi-view maximum entropy discrimination. In: Rossi F (ed) IJCAI 2013, Proceedings of the 23rd International Joint Conference on Artificial Intelligence, Beijing, China, August 3-9, 2013. IJCAI/AAAI, pp 1706–1712, URL http://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/view/6271

Sun S, Xie X, Yang M (2015) Multiview uncorrelated discriminant analysis. IEEE transactions on cybernetics 46(12):3272–3284

Sun Y, Li L, Zheng L, et al (2019) Image classification base on PCA of multi-view deep representation. J Vis Commun Image Represent 62:253–258. https://doi.org/10.1016/j.jvcir.2019.05.016, URL https://doi.org/10.1016/j.jvcir.2019.05.016

Tang J, Tian Y, Liu X, et al (2018a) Improved multi-view privileged support vector machine. Neural Networks 106:96–109. https://doi.org/10.1016/j.neunet.2018.06.017, URL https://doi.org/10.1016/j.neunet.2018.06.017

Tang J, Tian Y, Zhang P, et al (2018b) Multiview privileged support vector machines. IEEE Trans Neural Networks Learn Syst 29(8):3463–3477. https://doi.org/10.1109/TNNLS.2017.2728139, URL https://doi.org/10.1109/TNNLS.2017.2728139

Tang J, Tian Y, Liu D, et al (2019) Coupling privileged kernel method for multi-view learning. Inf Sci 481:110–127. https://doi.org/10.1016/j.ins.2018.12.058, URL https://doi.org/10.1016/j.ins.2018.12.058

Vapnik V, Vashist A, Pavlovitch N (2007) Learning using hidden information: Master-class learning. In: Fogelman-Soulié F, Perrotta D, Piskorski J, et al (eds) Mining Massive Data Sets for Security - Advances in Data Mining, Search, Social Networks and Text Mining, and their Applications to Security, Proceedings of the NATO Advanced Study Institute on Mining Massive Data Sets for Security, Gazzada (Varese), Italy, 10-21 September 2007, NATO Science for Peace and Security Series - D: Information and Communication Security, vol 19. IOS Press, pp 3–14, https://doi.org/10.3233/978-1-58603-898-4-3, URL https://doi.org/10.3233/978-1-58603-898-4-3

Wilcoxon F (1992) Individual comparisons by ranking methods. In: Breakthroughs in statistics. Springer, p 196–202

Xie X (2018) Regularized multi-view least squares twin support vector machines. Appl Intell 48(9):3108–3115. https://doi.org/10.1007/s10489-017-1129-3, URL https://doi.org/10.1007/s10489-017-1129-3

Xu W, Liu W, Chi H, et al (2019) Self-paced learning with privileged information. Neurocomputing 362:147–155. https://doi.org/10.1016/j.

neucom.2019.06.072, URL https://doi.org/10.1016/j.neucom.2019.06.072

Yan Y, Nie F, Li W, et al (2016) Image classification by cross-media active learning with privileged information. IEEE Trans Multim 18(12):2494–2502. https://doi.org/10.1109/TMM.2016.2602938, URL https://doi.org/10.1109/TMM.2016.2602938

Zhang C, Cheng J, Tian Q (2020) Multi-view image classification with visual, semantic and view consistency. IEEE Trans Image Process 29:617–627. https://doi.org/10.1109/TIP.2019.2934576, URL https://doi.org/10.1109/TIP.2019.2934576