GSPL: A Succinct Kernel Model for Group-Sparse Projections Learning of Multiview Data

Danyang Wu, Jin Xu, Xia Dong, Meng Liao, Rong Wang, Feiping Nie, and Xuelong Li

Abstract

This paper explores a succinct kernel model for Group-Sparse Projections Learning (GSPL), to handle multiview feature selection task completely. Compared to previous works, our model has the following useful properties: 1) Strictness: GSPL innovatively learns group-sparse projections strictly on multiview data via $\ell_{2,0}$-norm constraint, which is different with previous works that encourage group-sparse projections softly. 2) Adaptivity: In GSPL model, when the total number of selected features is given, the numbers of selected features of different views can be determined adaptively, which avoids artificial settings. Besides, GSPL can capture the differences among multiple views adaptively, which handles the inconsistent problem among different views. 3) Succinctness: Except for the intrinsic parameters of projection-based feature selection task, GSPL does not bring extra parameters, which guarantees the applicability in practice. To solve the optimization problem involved in GSPL, a novel iterative algorithm is proposed with rigorously theoretical guarantees. Experimental results demonstrate the superb performance of GSPL on synthetic and real datasets.

1 Introduction

Recently, a large amount of multiview data have emerged in real scenarios, in which the features are characterized from different sources, extractors and external environments. For example, a person can be described via iris, fingerprint and face features, a document can be narrated via multiple languages, an image can be described by different feature extractors, etc [1, 2, 3]. Considering high dimensionality of multiview data and expensive cost of label acquisition, how to select discriminative features on multiview data with unsupervised paradigm has attracted more attentions in recent years.

Related works. The previous works can be simply divided into two categories, including serial models and parallel models. 1) The serial models directly concatenate multiview feature vectors into a long vector, and explore a whole sparse Projection Matrix (PM) via employing single view unsupervised feature selection models (SOCFS [4], SOGFS [5], URAFS [6], models are frequently used), in which the discriminative features can be selected from sparse PM. Except for directly utilizing single-view models, [7] Li *et al.* proposed a ACSL model which learns both a unified projection and a collaborative graph to mine more discriminative information. 2) The parallel models explore a sparse PM for each view and integrate the selected features of different views artificially. The most exemplary works contain: ASVW [8], which collectively learns sparse PM for each view via $\ell_{2,p}$-norm regularization and a common graph from multiple views; MVUFS [9], which performs joint local learning regularized

orthogonal nonnegative matrix factorization and $\ell_{2,1}$-norm minimization; RMFS [10], which employs K-means to efficiently obtain labels for guiding $\ell_{2,1}$-norm regularized feature selection; CGMVUFS [11], which incorporates feature selection with $\ell_{2,1}$-norm constraint into non-negative matrix factorization (NMF) based clustering model.

Confronting problems. However, previous works suffer from several open flaws. In serial models, the inconsistency of multiple views cannot be captured completely, especially when concatenating multiview feature vectors. In parallel models, the number of selected features for each view depends on artificial settings, which is hard to apply in practice. Furthermore, both series and parallel works conventionally utilize $\ell_{2,1}$ or $\ell_{2,p}$-norm regularization to encourage the sparsity of PM, which cannot guarantee the group-sparsity of PM and leads to mismatched problem between the group-sparsity of PM and the number of selected features. More importantly, previous works contain several hyper-parameters for exploring the adaptive combination of different regularizations, including $\ell_{2,1}$ or $\ell_{2,p}$-norm regularization, which affects the applicability of model. Through the above analysis, Wu et al. can conclude that a complete multiview feature selection model should handle the following 4 problems:

- **IC Problem:** How to capture the **InConsistency** among multiple views?
- **SF Problem:** How to allocate the number of **Selected Features** for multiple views based on a total number?
- **GS Problem:** How to guarantee the **Group-Sparsity** of learnt projection matrix (matirces)?
- **SC Problem:** How to guarantee the **SuCcinctness** of model?

Our proposals. To this goal, Wu et al. propose a succinct kernel model for Group-Sparse Projections Learning (GSPL), to handle **IC**, **SF**, **GS** and **SC** problems simultaneously. The workflow is plotted into Fig. 1. Specifically, GSPL first learns a fused feature matrix, mapped and fused by sparse projection matrices and adaptive weights, and then encloses it to mutiview spectral embeddings based on Hilbert-Schmidt Independence Criterion (HSIC) on Reproducing Kernel Hilbert Spaces (RKHSs) adaptively. As aforementioned in Abstract, the proposed GSPL contains three useful functionalities, including **Adaptivity**, **Strictness** and **Succinctness**, wherein **Adaptivity** solves problems **IC** and **SF** via learning $\mathbf{p}$ and $\mathbf{z}$ in Fig. 1. **Strictness** solves problem **GS** via $\ell_{2,0}$-norm constraint of $\widehat{W}$ in Fig. 1, **Succinctness** means that that our model does not add any hyper-parameter except for the number of selected features and the number of reduced dimensionality brought by feature selection task, which handles the **SC** problem effectively. Except for the above three functionalities, the HSIC metric on RKHSs helps the learnt fused feature matrix to capture the high-order information from multiview spectral embeddings. To solve the optimization problem involved in GSPL, Wu et al. propose a novel iterative algorithm with theoretical guarantees. Eventually, the effectiveness of GSPL and the convergent speed of proposed algorithm are evaluated. The experimental results demonstrate that the proposed GSPL model achieves SOTA performance and the proposed solver can converge rapidly. In a word, Wu et al. concisely summarize the main contributions of this paper as follows:

- Wu et al. propose a GSPL model to solve the **IC**, **SF**, **GS** and **SC** vital problems simultaneously and effectively
- Wu et al. propose a novel iterative algorithm to solve the optimization problem involved in GSPL.
- The proposed GSPL model has **no extra hyperparameters**, but achieves **SOTA** performance in several real datasets compared to previous works.

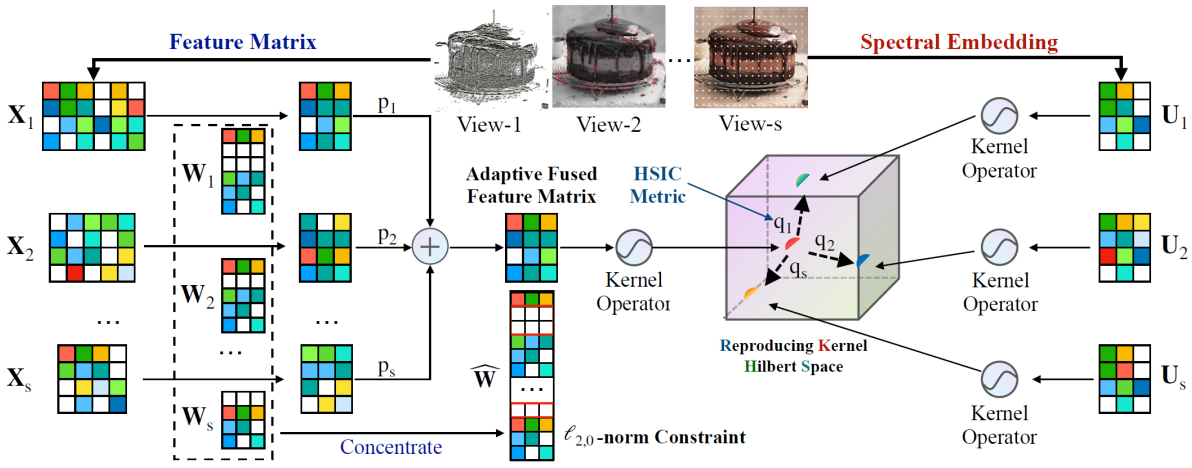


Figure 1: The workflow of GSPL Model.

2 Methodology

2.1 Proposed GSPL Model

In this paper, Wu et al. proposed GSPL model as follows:

$$\begin{aligned} &\max_{\mathbf{p},\mathbf{z},\{\mathbf{W}_\mathrm{v}\}_{\mathrm{v}=1}^{\mathrm{s}}} \sum_{\mathrm{g}=1}^{\mathrm{s}} \mathrm{z_g} \cdot \left\|\mathbf{U}_\mathrm{g}^{\mathrm{T}}\mathbf{H}\sum_{\mathrm{v}=1}^{\mathrm{s}} \mathrm{p_v}\mathbf{X}_\mathrm{v}\mathbf{W}_\mathrm{v}\right\|_{\mathrm{F}}^2 \\ &\text{s.t. } \widehat{\mathbf{W}} = [\mathbf{W}_1;\ldots;\mathbf{W}_\mathrm{s}]^{\mathrm{T}}, \widehat{\mathbf{W}}^{\mathrm{T}}\widehat{\mathbf{W}} = \mathbf{I}_{\mathrm{m}\times\mathrm{m}}, \\ &\|\widehat{\mathbf{W}}\|_{2,0} = \mathrm{k}, \|\mathbf{p}\|_2 = 1, \|\mathbf{z}\|_2 = 1, \mathbf{z} \geqslant \mathbf{0}, \end{aligned} \tag{1}$$

where $\mathbf{X}_\mathrm{v} \in \mathcal{R}^{n\times d_v}$ is the feature matrix of v-th view, n and d_v are the numbers of samples and dimensions of $\mathbf{X}_\mathrm{v}$. Besides, the given multi-view data $\mathcal{X} = \{\mathbf{X}_1, \ldots, \mathbf{X_s}\}$. $\mathbf{W}_\mathrm{v} \in \mathbb{R}^{\mathrm{d_v}\times\mathrm{m}}$ is the projection matrix. The v-th spectral embedding denoted as $\mathbf{U}_\mathrm{v} \in \mathbb{R}^{\mathrm{n}\times\mathrm{c}}$ generated via Laplacian eigenmap [12], i.e., $\min_{\mathbf{U}_\mathrm{v}^{\mathrm{T}}\mathbf{U}_\mathrm{v}=\mathbf{I}_{\mathrm{c}\times\mathrm{c}}} \mathrm{Tr}\left(\mathbf{U}_\mathrm{v}^{\mathrm{T}}\mathbf{L_S}\mathbf{U}_\mathrm{v}\right)$, on the similarity matrix $\mathbf{S}_\mathrm{v} \in \mathbb{R}^{n\times n}$ constructed from $\mathbf{X}_v$, where c is the number of clusters and $\mathbf{L_S} \in \mathbb{R}^{\mathrm{n}\times\mathrm{n}}$ is the Laplacian matrix of $\mathbf{S}$.

2.2 Optimization for GSPL Model

The problem (12) can be optimized iteratively.

Update p and fix others. When $\mathbf{W}$ and $\mathbf{z}$ are fixed, problem (12) becomes

$$\max_{\|\mathbf{p}\|_2=1} \mathrm{Tr}\left(\widehat{\mathbf{U}}\left(\sum_{\mathrm{v}=1}^{\mathrm{s}} \mathrm{p_v}\mathbf{X}_\mathrm{v}\mathbf{W}_\mathrm{v}\right)\left(\sum_{\mathrm{v}=1}^{\mathrm{s}} \mathrm{p_v}\mathbf{X}_\mathrm{v}\mathbf{W}_\mathrm{v}\right)^{\mathrm{T}}\right) \tag{2}$$

where $\widehat{\mathbf{U}} = \sum_{\mathrm{g}=1}^{\mathrm{s}} \mathrm{z_g}\mathbf{H}^{\mathrm{T}}\mathbf{U}_\mathrm{g}\mathbf{U}_\mathrm{g}^{\mathrm{T}}\mathbf{H}$. Let $\mathbf{A}_\mathrm{v} = \widehat{\mathbf{U}}\mathbf{X}_\mathrm{v}\mathbf{W}_\mathrm{v}$, $\mathbf{B}_\mathrm{v} = \mathbf{X}_\mathrm{v}\mathbf{W}_\mathrm{v}$, then problem (2) becomes the following problem: $\max_{\|\mathbf{p}\|_2=1} \mathrm{Tr}\left(\left(\sum_{\mathrm{v}=1}^{\mathrm{s}} \mathrm{p_v}\mathbf{A}_\mathrm{v}\right)\left(\sum_{\mathrm{v}=1}^{\mathrm{s}} \mathrm{p_v}\mathbf{B}_\mathrm{v}\right)^{\mathrm{T}}\right)$. For convenience, Wu et al. denote $\widehat{\mathbf{A}} = [\mathrm{Vec}\left(\mathbf{A}_1\right), \ldots, \mathrm{Vec}\left(\mathbf{A}_\mathrm{s}\right)]$, $\widehat{\mathbf{B}} = [\mathrm{Vec}\left(\mathbf{B}_1\right), \ldots, \mathrm{Vec}\left(\mathbf{B}_\mathrm{s}\right)] \in \mathbb{R}^{\mathrm{nm}\times\mathrm{s}}$, where $\mathrm{Vec}(\cdot)$ is the column-based

matrix vectorization operator. Then the problem can be rewritten as the following form:

$$\max_{\|\mathbf{p}\|_2=1} \mathrm{Tr}\left(\mathbf{p}^{\mathrm{T}}\widehat{\mathbf{B}}^{\mathrm{T}}\widehat{\mathbf{A}}\mathbf{p}\right) \tag{3}$$

The solution $\mathbf{p}$ of problem (3) can be formed as the eigenvectors corresponding to the s largest eigenvalues of $\widehat{\mathbf{B}}^{\mathrm{T}}\widehat{\mathbf{A}}$.

Update z and fix others. When $\mathbf{W}$ and $\mathbf{p}$ are fixed, problem (12) becomes

$$\max_{\|\mathbf{z}\|_2=1,\mathbf{z}\geqslant 0} \sum_{\mathrm{g}=1}^{\mathrm{s}} \mathrm{z_g} \cdot \left\| \mathbf{U}_{\mathrm{g}}^{\mathrm{T}}\mathbf{H} \sum_{\mathrm{v}=1}^{\mathrm{s}} \mathrm{p_v}\mathbf{X_v}\mathbf{W_v} \right\|_{\mathrm{F}}^2 \tag{4}$$

Considering the term $\left\|\mathbf{U}_{\mathrm{g}}^{\mathrm{T}}\mathbf{H}\sum_{\mathrm{v}=1}^{\mathrm{s}} \mathrm{p_v}\mathbf{X_v}\mathbf{W_v}\right\|_{\mathrm{F}}^2 \geq 0$, Wu et al. have the following derivations according to Cauchy-Schwarz inequality [13]:

$$\begin{aligned} &\sum_{\mathrm{g}=1}^{\mathrm{s}} \mathrm{z_g} \cdot \left\| \mathbf{U}_{\mathrm{g}}^{\mathrm{T}}\mathbf{H} \sum_{\mathrm{v}=1}^{\mathrm{s}} \mathrm{p_v}\mathbf{X_v}\mathbf{W_v} \right\|_{\mathrm{F}}^2 \\ &\overset{(a)}{\leq} \sqrt{\left(\sum_{\mathrm{g}=1}^{\mathrm{s}} \left\| \mathbf{U}_{\mathrm{g}}^{\mathrm{T}}\mathbf{H} \sum_{\mathrm{v}=1}^{\mathrm{s}} \mathrm{p_v}\mathbf{X_v}\mathbf{W_v} \right\|_{\mathrm{F}}^4\right)\left(\sum_{\mathrm{g}=1}^{\mathrm{s}} \mathrm{z_g^2}\right)} \\ &= \sqrt{\sum_{\mathrm{g}=1}^{\mathrm{s}} \left\| \mathbf{U}_{\mathrm{g}}^{\mathrm{T}}\mathbf{H} \sum_{\mathrm{v}=1}^{\mathrm{s}} \mathrm{p_v}\mathbf{X_v}\mathbf{W_v} \right\|_{\mathrm{F}}^4} \end{aligned} \tag{5}$$

Considering $\|\mathbf{z}\|_2 = 1$, the equality holds when

$$\mathrm{z_g} = \frac{\left\|\mathbf{U}_{\mathrm{g}}^{\mathrm{T}}\mathbf{H}\sum_{\mathrm{v}=1}^{\mathrm{s}} \mathrm{p_v}\mathbf{X_v}\mathbf{W_v}\right\|_{\mathrm{F}}^2}{\sqrt{\sum_{\mathrm{g}=1}^{\mathrm{s}} \left\|\mathbf{U}_{\mathrm{g}}^{\mathrm{T}}\mathbf{H}\sum_{\mathrm{v}=1}^{\mathrm{s}} \mathrm{p_v}\mathbf{X_v}\mathbf{W_v}\right\|_{\mathrm{F}}^4}} \tag{6}$$

which is the closet solution of problem (4).

Update $\{\mathbf{W_v}\}_{\mathrm{v}=1}^{\mathrm{s}}$ and fix others. When $\mathbf{p}$ and $\mathbf{z}$ are fixed, problem (12) can be written as

$$\max_{\widehat{\mathbf{W}}^T\widehat{\mathbf{W}}=\mathbf{I}_{\mathrm{m}\times\mathrm{m}},\|\widehat{\mathbf{W}}\|_{2,0}=\mathrm{k}} \sum_{\mathrm{g}=1}^{\mathrm{s}} \mathrm{z_g} \cdot \left\| \mathbf{U}_{\mathrm{g}}^{\mathrm{T}}\mathbf{H}\widehat{\mathbf{X}}\tilde{\mathbf{P}}\widehat{\mathbf{W}} \right\|_{\mathrm{F}}^2 \tag{7}$$

where $\widehat{\mathbf{X}} = [\mathbf{X}_1, \ldots \mathbf{X}_{\mathrm{s}}] \in \mathbb{R}^{\mathrm{n}\times\widehat{\mathrm{d}}}$, $\tilde{\mathbf{P}} = \mathrm{diag}(\tilde{\mathbf{p}}) \in \mathbb{R}^{\widehat{\mathrm{d}}\times\widehat{\mathrm{d}}}$ is a diagonal matrix and $\tilde{\mathbf{p}} = [\mathrm{p}_1\mathbf{1}_{\mathrm{d}_1}; \ldots; \mathrm{p}_{\mathrm{s}}\mathbf{1}_{\mathrm{d}_{\mathrm{s}}}] \in \mathbb{R}^{\widehat{\mathrm{d}}}$ is the diagonal vector of it, $\mathrm{diag}(\cdot)$ is the vector-diagonal operator. Denote $\mathbf{E}_{\mathrm{g}} = \mathbf{U}_{\mathrm{g}}^{\mathrm{T}}\mathbf{H}\widehat{\mathbf{X}}\tilde{\mathbf{P}}$, the problem (7) can be written as

$$\max_{\widehat{\mathbf{W}}^{\mathrm{T}}\widehat{\mathbf{W}}=\mathbf{I}_{\mathrm{m}\times\mathrm{m}},\|\widehat{\mathbf{W}}\|_{2,0}=\mathrm{k}} \sum_{\mathrm{g}=1}^{\mathrm{s}} \mathrm{z_g} \cdot \mathrm{Tr}\left(\widehat{\mathbf{W}}^{\mathrm{T}}\mathbf{E}_{\mathrm{g}}^{\mathrm{T}}\mathbf{E}_{\mathrm{g}}\widehat{\mathbf{W}}\right). \tag{8}$$

Further, Wu et al. denote $\mathbf{S_E} = \sum_{\mathrm{g}=1}^{\mathrm{s}} \mathrm{z_g}\mathbf{E}_{\mathrm{g}}^{\mathrm{T}}\mathbf{E}_{\mathrm{g}} + \lambda\mathbf{I}_{\mathrm{m}\times\mathrm{m}}$, where λ guarantees that $\mathbf{S}$ is positive semi-definite, then problem (8) can be written as

$$\max_{\widehat{\mathbf{W}}^{\mathrm{T}}\widehat{\mathbf{W}}=\mathbf{I}_{\mathrm{m}\times\mathrm{m}},\|\widehat{\mathbf{W}}\|_{2,0}=\mathrm{k}} \mathrm{Tr}\left(\widehat{\mathbf{W}}^{\mathrm{T}}\mathbf{S}_{\mathrm{E}}\widehat{\mathbf{W}}\right) \tag{9}$$

which is NP-hard, then Wu et al. consider to solve it into two cases. At first, Wu et al. consider the case $\mathrm{rank}\,(\mathbf{S_E}) \leqslant m$. Since $\|\widehat{\mathbf{W}}\|_{2,0} = \mathrm{k}$, suppose the full rank decomposition of $\widehat{\mathbf{W}}$ is $\widehat{\mathbf{W}} = \mathbf{BD}$. $\mathbf{B} \in \{\mathbf{0}, \mathbf{1}\}_{\widehat{\mathrm{d}}\times\mathrm{k}}$ is the selection matrix and $\mathbf{D} \in \mathbb{R}^{\mathrm{k}\times\mathrm{m}}$ is composed by the non-zero row of $\widehat{\mathbf{W}}$. Naturally $\mathbf{D}^{\mathrm{T}}\mathbf{D} = \mathbf{I}_{\mathrm{m}\times\mathrm{m}}$. Then problem (9) can be written as the following problem w.r.t $\mathbf{B}$ and $\mathbf{D}$:

$$\max_{\mathbf{B},\mathbf{D}=\mathbf{I}_{\mathrm{m}\times\mathrm{m}}} \mathrm{Tr}\left(\mathbf{D}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}\mathbf{S_E}\mathbf{B}\mathbf{D}\right) \tag{10}$$

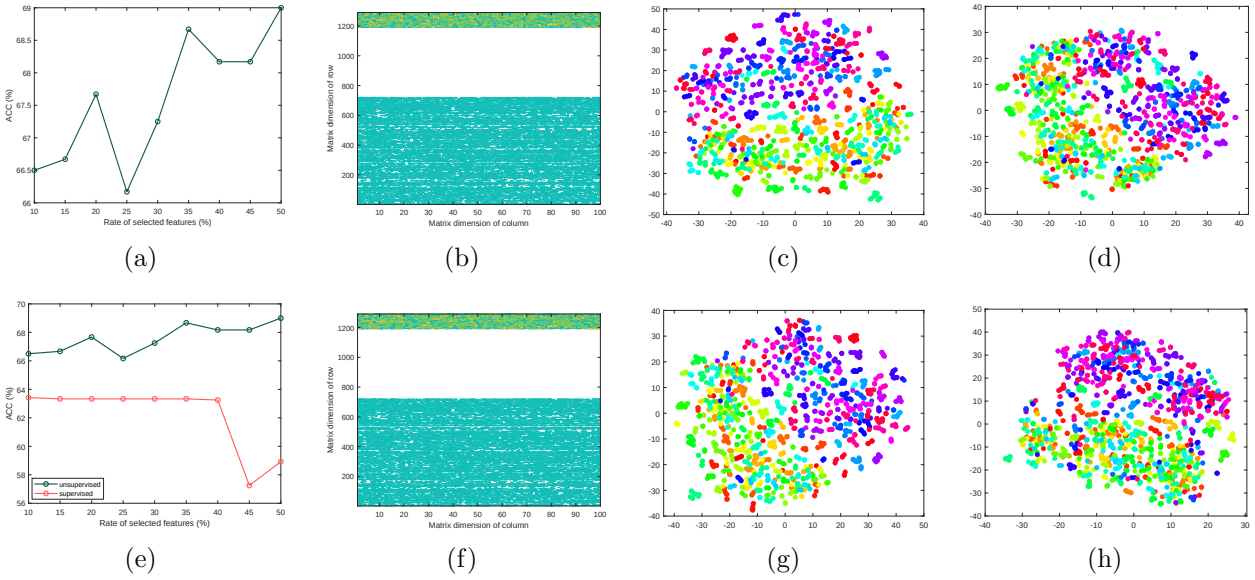


Figure 2: Experimental results on the ORL database. (a) The accuracy (%) with different rates of selected features. (b) The projection matrix. (c) T-SNE visulization of test data. (d) T-SNE visulization of projected test data. (e) The accuracy (%) with different rates of selected features. (f) The projection matrix. (g) T-SNE visulization of test data. (h) T-SNE visulization of projected test data.

Then Wu et al. consider the case $\operatorname{rank}(\mathbf{S_E}) > \mathrm{m}$. In this case, they consider utilizing the famous Majorize-Minimization (MM) framework. $\widehat{\mathbf{W}}_0^{\mathrm{T}}\mathbf{S_E}\widehat{\mathbf{W}}_0 = \widehat{\mathbf{W}}_0^{\mathrm{T}}\left(\mathbf{S_E}\widehat{\mathbf{W}}_0\left(\widehat{\mathbf{W}}_0^{\mathrm{T}}\mathbf{S_E}\widehat{\mathbf{W}}_0\right)^{\dagger}\widehat{\mathbf{W}}_0^{\mathrm{T}}\mathbf{S_E}\right)\widehat{\mathbf{W}}_0$ is the current solution. Then the problem (10) can be

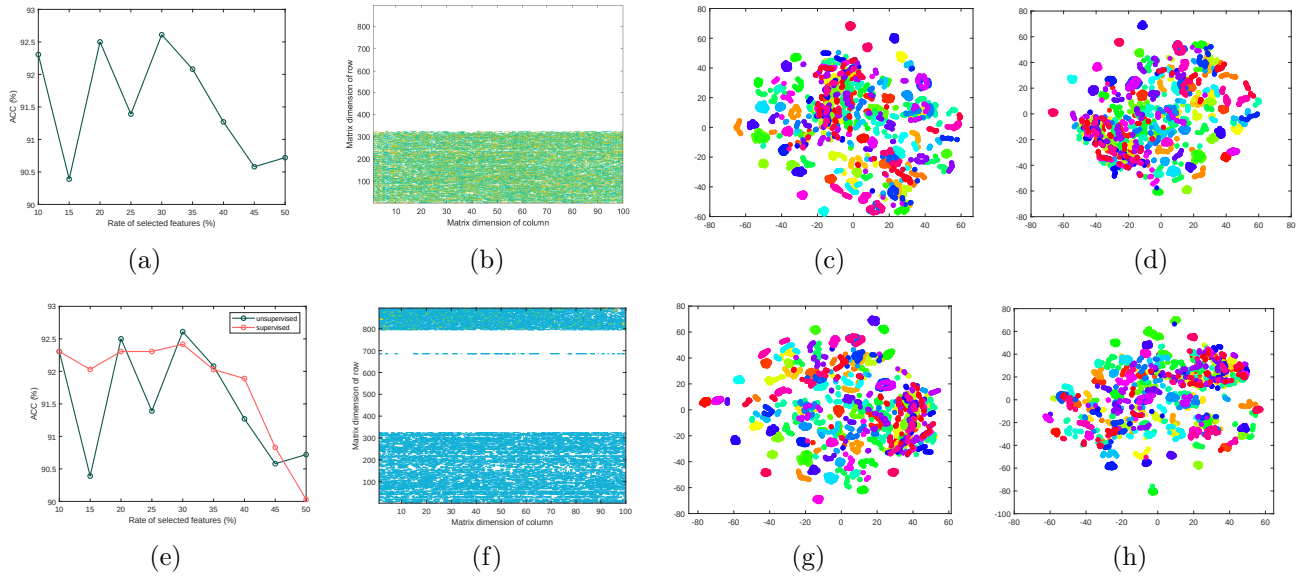
$$\begin{aligned}&\max_{\widehat{\mathbf{W}}} \operatorname{Tr}\left(\widehat{\mathbf{W}}^{\mathrm{T}}\left(\mathbf{S_E}\widehat{\mathbf{W}}_0\left(\widehat{\mathbf{W}}_0^{\mathrm{T}}\mathbf{S_E}\widehat{\mathbf{W}}_0\right)^{\dagger}\widehat{\mathbf{W}}_0^{\mathrm{T}}\mathbf{S_E}\right)\widehat{\mathbf{W}}\right)\\ &\text{s.t. } \widehat{\mathbf{W}}^{T}\widehat{\mathbf{W}} = \mathbf{I}_{\mathrm{m}\times\mathrm{m}}, \|\widehat{\mathbf{W}}\|_{2,0} = \mathrm{k}.\end{aligned} \tag{11}$$

Algorithm 1 The Algorithm for Solving problem (12).

Input: $\mathcal{X} \in \mathcal{X} \in \mathbb{R}^{n\times\hat{\mathrm{d}}}, \{\mathbf{U}_{\mathrm{g}}\}_{\mathrm{g}=1}^{\mathrm{s}}, \widehat{\mathrm{d}}, \mathrm{k}, \mathrm{m}$.

Output: $\mathbf{p}, \mathbf{z}, \left\{\mathbf{W}_1, \ldots, \widehat{\mathbf{W}_{\mathbf{s}}}\right\}$.

Initialize $\mathbf{p}, \mathbf{q}, \widehat{\mathbf{W}} = [\mathbf{W}_1; \ldots; \mathbf{W}_{\mathrm{s}}]^{\mathrm{T}}$.
while not converge **do**
  Update $\mathbf{p}$ via solving problem (3)
  Update $\mathbf{z}$ via Eq. (6).
  if $\operatorname{rank}(\mathbf{S_E}) \leqslant m$ **then**
    Update $\mathbf{B}$ and $\mathbf{D}$ via solving problem (10).
    Update $\widehat{\mathbf{W}}$ via $\widehat{\mathbf{W}} = \mathbf{BD}$.
  else
    while not converge **do**
      Set $\widehat{\mathbf{W}}_0$ by current $\widehat{\mathbf{W}}$.
      Update $\widehat{\mathbf{W}}$ via solving problem (11).
    end while
  end if
end while



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 3: Experimental results on the COIL100 database. (a) The accuracy (%) with different rates of selected features. (b) The projection matrix. (c) T-SNE visulization of test data. (d) T-SNE visulization of projected test data. (e) The accuracy (%) with different rates of selected features. (f) The projection matrix. (g) T-SNE visulization of test data. (h) T-SNE visulization of projected test data.

2.3 My Model

To enhance the discriminability of GSPL, I proposed GSPL-II as follows:

$$\begin{aligned} &\max_{\mathbf{p},\mathbf{z},\{\mathbf{W}_v\}_{v=1}^{s}} \sum_{g=1}^{s} z_g \cdot \left\| \mathbf{U}_g^{T}\mathbf{H} \sum_{v=1}^{s} p_v \mathbf{X}_v \mathbf{W}_v \right\|_F^2 \\ &\text{s.t. } \widehat{\mathbf{W}} = [\mathbf{W}_1; \ldots; \mathbf{W}_s]^{T}, \widehat{\mathbf{W}}^{T}\widehat{\mathbf{W}} = \mathbf{I}_{m\times m}, \\ &\|\widehat{\mathbf{W}}\|_{2,0} = k, \|\mathbf{p}\|_2 = 1, \|\mathbf{z}\|_2 = 1, \mathbf{z} \geqslant \mathbf{0}, \\ &\mathrm{Tr}(\widehat{\mathbf{W}}^{T}\widehat{\mathbf{X}}^{T}\widehat{\mathbf{X}}\widehat{\mathbf{W}}) = cons \end{aligned} \tag{12}$$

where *cons* means the constant.

3 Experiments

In this section, I conduct several experiments to evaluate the method.

3.1 Experiments on the ORL and COIL100 database

There are ten different images of each of 40 distinct subjects in ORL database. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement).

To extend the original database to a multi-view database, I extract the features with several feature extraction methods, i.e., LBP, HOG and GIST. The same kind of features are the samples in the same view. I set the rate of selected features in the range of $\{10, 15, \cdots, 50\}$ to get the k. As for the reduced dimension of feature, i.e., m, I set it 100. Fig. 2 and FIG. 3 show the result of the experiments.

4 Conclusion and future work

In this report, I learn the GSPL model and reproduce the method. Moreover, I proposed a novel method called GSPL-II which is a supervised method. I conduct several experiments to demonstrate GSPL and GSPL-II work well on multi-view human face database and rotational object image database. They can select the multi-view features from the weighted data. There is no hyper-parameter to initialize in the experiments. Hence, the optimal solution is easy to get. GSPL-II works better than GSPL sometimes. It means GSPL-II is a robust method. In the future, I can introduce the discriminant information into the graph to improve the discriminability and locality. On the other hand, I can transfer the model into the neural networks version to get a interpretable networks.

References

[1] Z. Zhang, Z. Zhai, and L. Li, "Uniform projection for multi-view learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 8, pp. 1675–1689, 2017.

[2] F. Nie, L. Tian, and X. Li, "Multiview clustering via adaptively weighted procrustes," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery amp; Data Mining*, KDD '18, (New York, NY, USA), p. 2022–2030, Association for Computing Machinery, 2018.

[3] D. Wu, F. Nie, R. Wang, and X. Li, "Multi-view clustering via mixed embedding approximation," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3977–3981, 2020.

[4] D. Han and J. Kim, "Unsupervised simultaneous orthogonal basis clustering feature selection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[5] F. Nie, W. Zhu, and X. Li, "Unsupervised feature selection with structured graph optimization," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30, Feb. 2016.

[6] X. Li, H. Zhang, R. Zhang, Y. Liu, and F. Nie, "Generalized uncorrelated regression with adaptive graph for unsupervised feature selection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 5, pp. 1587–1595, 2019.

[7] X. Dong, L. Zhu, X. Song, J. Li, and Z. Cheng, "Adaptive collaborative similarity learning for unsupervised multi-view feature selection," *arXiv preprint arXiv:1904.11228*, 2019.

[8] C. Hou, F. Nie, H. Tao, and D. Yi, "Multi-view unsupervised feature selection with adaptive similarity and view weight," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 9, pp. 1998–2011, 2017.

[9] M. Qian and C. Zhai, "Unsupervised feature selection for multi-view clustering on text-image web news data," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, (New York, NY, USA), p. 1963–1966, Association for Computing Machinery, 2014.

[10] H. Liu, H. Mao, and Y. Fu, "Robust multi-view feature selection," in *2016 IEEE 16th International Conference on Data Mining (ICDM)*, pp. 281–290, 2016.

[11] C. Tang, J. Chen, X. Liu, M. Li, P. Wang, M. Wang, and P. Lu, "Consensus learning guided multi-view unsupervised feature selection," *Knowledge-Based Systems*, vol. 160, pp. 49–60, 2018.

[12] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing Systems* (T. Dietterich, S. Becker, and Z. Ghahramani, eds.), vol. 14, MIT Press, 2001.

[13] J. M. Steele, *The Cauchy-Schwarz master class: an introduction to the art of mathematical inequalities*. Cambridge University Press, 2004.