Semi-Supervised Classification via Simultaneous Label and Discriminant Embedding Estimation

F. Dornaika, A. Baradaaji, Y. El Traboulsi
University of the Basque Country UPV/EHU, San Sebastian, SPAIN
IKERBASQUE, Basque Foundation for Science, Bilbao, SPAIN}

Introduction

In this paper, we introduce a novel semi-supervised framework for linear metric learning (i.e., feature extraction) and soft label inference that integrates two types of graph-based various smoothness constraints. The proposed framework simultaneously calculates the class indicator matrix and the linear projection. This kind of novel criterion is expected to learn more discriminative semi-supervised models. The discrimination adopted in our work is inspired by the supervised Local Discriminant Embedding, which aims at maximizing a local margin of the projection space.

The main contributions to the paper are as follows.

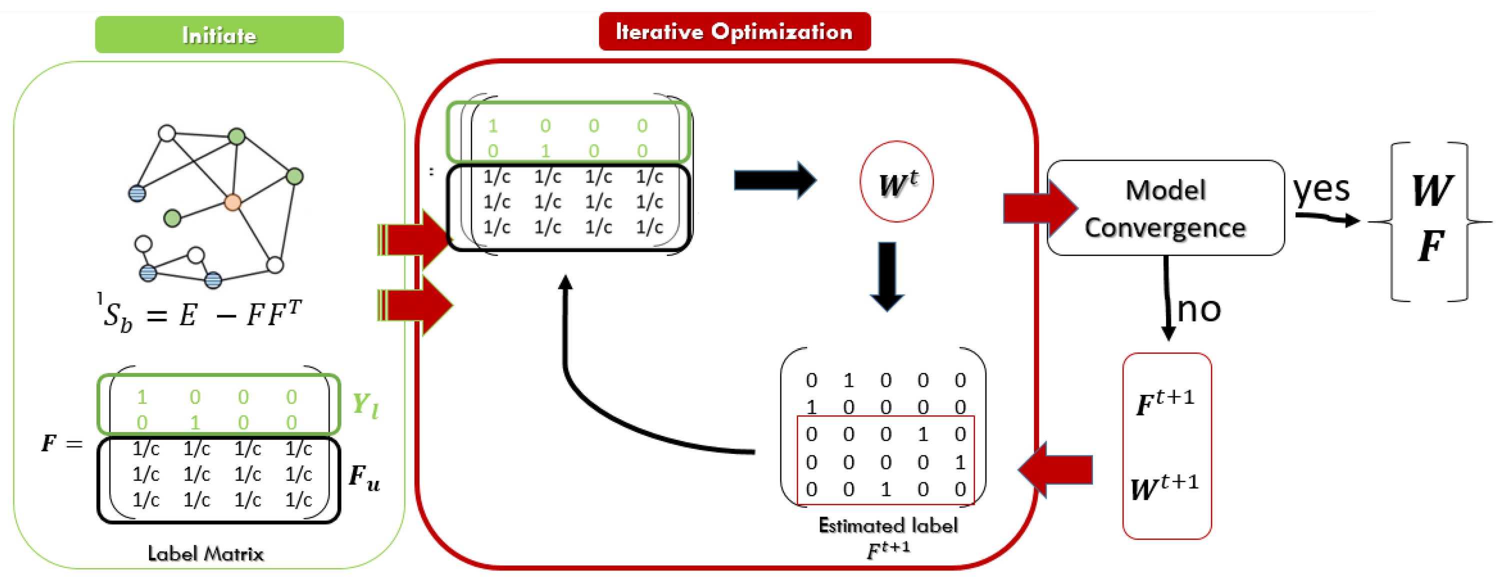
- Linking the soft label matrix to the local margin criterion (via approximating the between-class graph), our proposed method can fully exploit the label and unlabeled data samples in generating a discriminant embedding that overcomes the Linear Discriminant Analysis (LDA) limitations.
- We enforce graph regularization on both the soft labels and the projected data to get a more accurate solution.
- We derive an efficient algorithm for the joint estimation of the soft labels and the linear transform. Due to the iterative process of the algorithm, the current update of the soft labels would progressively increase the amount of supervision information needed for extracting a good discriminant linear embedding.

Proposed Methods

A proposed semi-supervised framework for linear metric learning (i.e., feature extraction) and soft label inference that integrates two types of graph-based various smoothness constraints.

$$\begin{aligned} g(\mathbf{W},\mathbf{F}) &= tr\left(\mathbf{W}^T\mathbf{X}\,(\mathbf{E}-\mathbf{F}\,\mathbf{F}^T)\,\mathbf{X}^T\mathbf{W}\right)+\beta\, tr\left(\mathbf{F}^T\mathbf{L}\,\mathbf{F}\right) \\ &+ \gamma\, tr\left(\mathbf{W}^T\mathbf{X}\,\mathbf{L}\,\mathbf{X}^T\mathbf{W}\right)+\alpha\, tr\left((\mathbf{F}^T\mathbf{F}-\mathbf{I})^T(\mathbf{F}^T\mathbf{F}-\mathbf{I})\right) \\ & s.t.\ \mathbf{W}^T\mathbf{X}^T\mathbf{D}_b\,\mathbf{X}\,\mathbf{W}=\mathbf{I}\ \text{ and }\ F_{ij}\geq 0, \end{aligned}$$

We minimize the above criterion using an alternating minimization scheme



(1) S_b is between-class graph, E is a matrix of ones.

Table 1. Best average recognition rates using several competing semi-supervised learning methods

EXT - Yale	1 Sample		2 Samples		3 Samples	
Method	*T*	*U*	*T*	*U*	*T*	*U*
FME	36.8%	40.7%	51.2%	54.7%	56.2%	59.9%
KFME	27.5%	29.3%	45.3%	46.5%	52.2%	56.3%
SDA	32.8%	35.0%	48.5%	50.7%	54.6%	58.2%
SDE	43.8%	46.0%	61.4%	60.5%	65.4%	67.1%
TR-FSDA	41.8%	45.2%	59.4%	58.6%	63.4%	64.0%
GLPP	25.5%	28.1%	41.0%	42.6%	49.7%	54.4%
SLDA	**47.9%**	**51.1%**	63.4%	64.3%	67.1%	69.7%
ISDA	**47.9%**	**51.1%**	63.3%	63.9%	67.2%	69.4%
JSLDE	46.4%	49.4%	**65.4%**	**67.0%**	**68.8%**	**71.9%**

Results

Table 1 depict the best average recognition rates obtained by nine semi-supervised methods over EXT - Yale dataset. The average was computed over the ten random splits. The results obtained on the unlabeled, and test data are referring to U and T, respectively. The classification was performed using the Nearest Neighbor classifier in the projected space. The bold numbers indicate the best result among all competing methods.

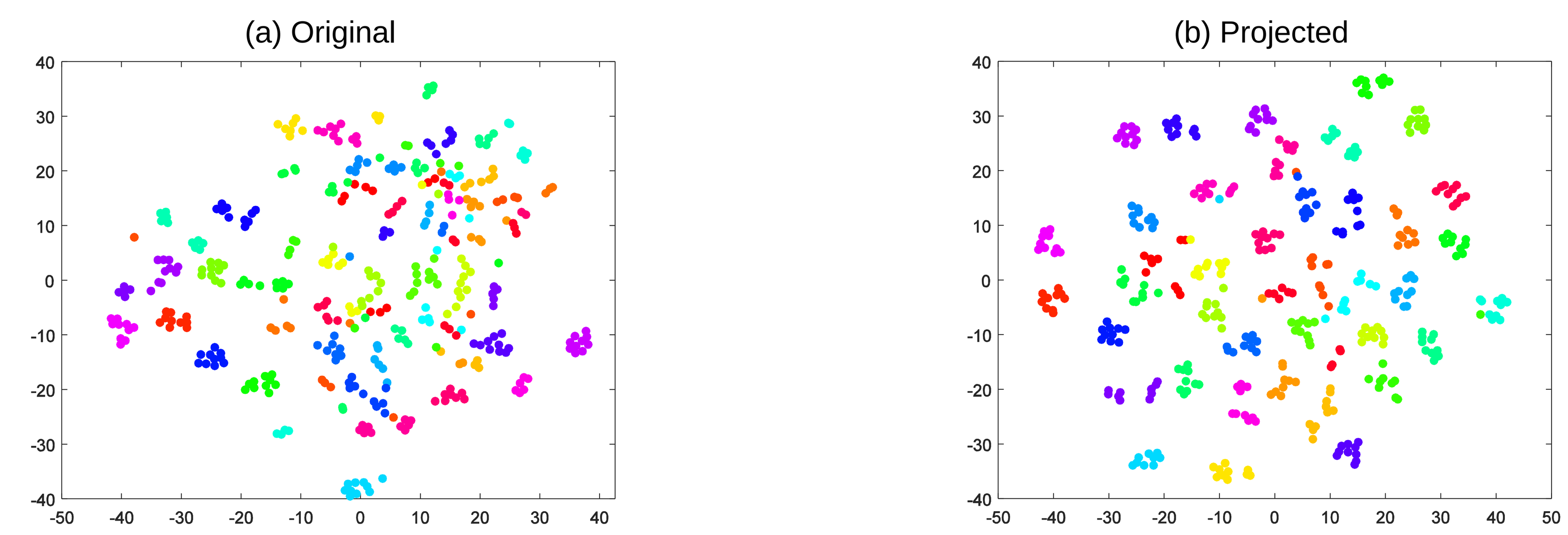


Figure 1. t-SNE visualization of the ORL dataset. (a) Face images using their original features. (b) Face images using their projection obtained by JSLDE.

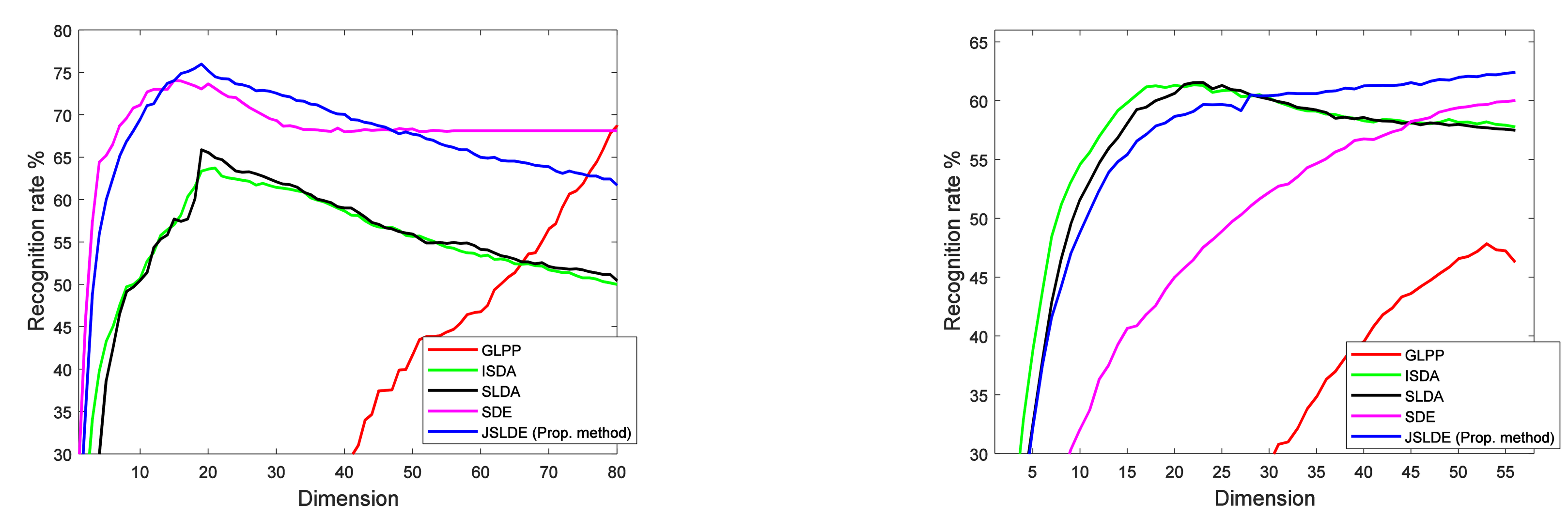


Figure 2. Recognition accuracy vs. feature dimension obtained with the EXT Yale and Coil-20 datasets (test evaluation). Three labeled samples per class were used. The classifier used was the Nearest Neighbor classifier.

Discussion

We have generated a visual representation of the image samples distribution after projecting them by JSLDE using t-SNE. Figure 1.a and Figure 1.b illustrated the obtained distribution of the ORL images (400 images) when t-SNE used the original features and JLSDE features, respectively. We set the number of labeled images per class to seven. From these figures, it seems that by using the projection of the proposed JLSDE method, the images of the same class become close to each other, and the samples of different classes are pushed far away as much as possible. This helps in obtaining a good classification of samples.

Figure 2 illustrates the average performance of the competing methods GLPP, ISDA, SLDA, SDE, and our proposed JSLDE method, as a function of the number of features obtained after data projection. This figure corresponds to the Extended Yale and Coil-20 datasets for which the performance obtained on the test part and with three labeled images per class.

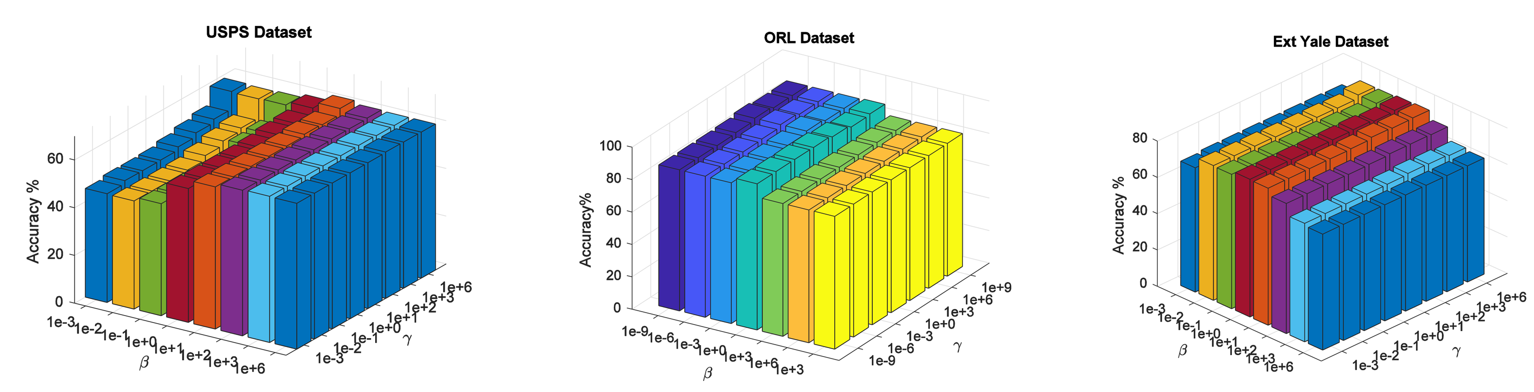


Figure 3. Recognition accuracy as a function of α and β .

Conclusions

We introduced a framework that is able to jointly estimate the soft labels and linear embedding for a semi-supervised context. Unlike the existing methods for the joint estimation, the between-class matrix is considered to reflect local margins among classes. Furthermore, the proposed criterion enforces the smoothness of both the predicted labels and the linear projection of the data samples. Due to the iterative process of the algorithm, the current update of the soft labels would progressively increase the amount of supervision information needed for extracting a good discriminant linear embedding.

Contact

Abdullah BARADAAJi
University of the Basque Country
San Sebastian, Spain
E-mail: baradaaji.abdu@gmail.com

Fadi Dornaika
University of the Basque Country
IKERBASQUE Research Foundation
San Sebastian, Spain
E-mail: fadi.dornaika@ehu.es

References

1. J. Xu, B. Tang, H. He, and H. Man, “Semi-supervised feature selection based on relevance and redundancy criteria,” IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 9, pp. 1974–1984, Sep. 2017.
2. Y. E. Traboulsi and F. Dornaika, “Flexible semi-supervised embedding based on adaptive loss regression: Application to image categorization,” Information Sciences, vol. 444, pp. 1–9, 2018.
3. F. Dornaika and Y. El Traboulsi, “Learning flexible graph-based semi -supervised embedding,” IEEE Transactions on Cybernetics, vol. 46, no. 1, pp. 206–218, 2016.
4. D. Cai, X. He, and J. Han, “Semi-supervised discriminant analysis,” in IEEE International Conference on Computer Vision, 2007.