SEMI-SUPERVISED DICTIONARY LEARNING WITH GRAPH REGULARIZED AND ACTIVE POINTS

Khanh-Hung TRAN, Fred-Maurice NGOLE-MBOULA, Jean-Luc STARCK and Vincent PROST
khanh-hung.tran@cea.fr

Abstract

Dictionary learning is a feature learning method which aims at learning a basis from input data, then a sample can be represented by sparse code, which is the form of a linear combination of several elements in the basis. Many works have shown the discriminating power of sparse code, then supervised dictionary learning (SDL) methods have been firstly developed for classification problem. However, in general, supervised learning needs a large number of labelled samples per class to achieve an acceptable result. In order to deal with databases which have just several labelled samples per class, semi-supervised learning, which uses also unlabelled samples in training phase, is employed. In our work, we try to generalize all semi-supervised dictionary learning (SSDL) methods and propose a new one based on combining two pillars: on one hand, we enforce the manifold structure preservation from the original data into the sparse code space by Locally Linear Embedding (LLE), which can be considered as sparse code regularization; on the other hand, we learn jointly with the dictionary, a semi-supervised classifier in sparse code space to enhance the discriminating power. We show that our approach provides an improvement over state-of-the-art in semi-supervised dictionary learning family.

Performance with only 20 labelled samples per class in training

	Method	Reconstruction for data $\mathcal{R}$	Discrimination $\mathcal{D}$	Preservation $\mathcal{F}$	USPS	MNIST
SSDL methods	OSSDL	✓	supervised		80.8 ± 2.8	73.2 ± 1.8
	SD2D	✓	semi-supervised		86.6 ± 1.6	77.6 ± 0.8
	SSR-D	✓			87.2 ± 0.5	83.8 ± 1.2
	SSP-DL	✓		✓	87.8 ± 1.1	85.8 ± 1.2
	USSDL	✓	semi-supervised		91.6 ± 1.2	84.8 ± 1.7
	PSSDL*	✓	supervised	✓	86.9 ± 1.0	87.4 ± 1.2
	SSD-LP	✓	semi-supervised		90.3 ± 1.3	87.8 ± 1.6
	SSDL-GA	✓	semi-supervised	✓	$\mathbf{93.6 \pm 1.0}$	$\mathbf{90.0 \pm 0.8}$
	CNN				89.28 ± 1.4	88.4 ± 1.1

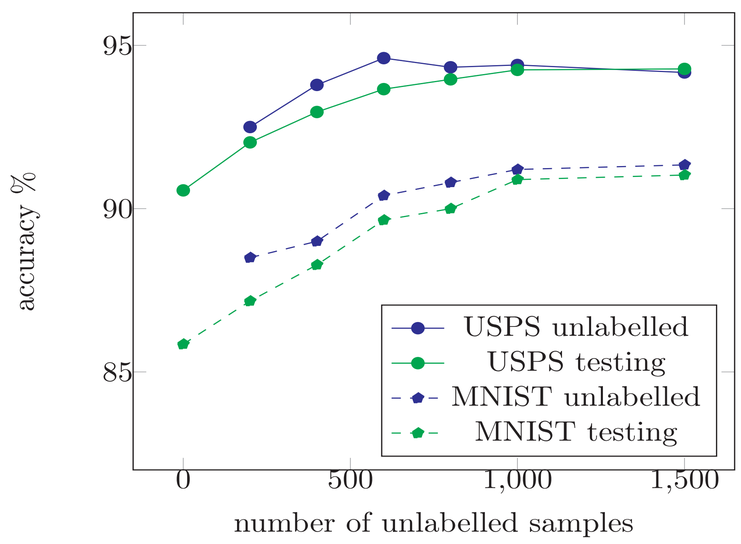
Optimization

Algorithm 1 SSDL-GA

Input: $\mathbf{X}, \mathbf{Y}^l, \beta$, k nearest neighbors, γ, λ, μ.
1: **Initialize :** L, $\mathbf{D}, \mathbf{A}, \mathbf{W}, \mathbf{b}$
2: **while** not converged **do**
3: Update active points
4: Update the probability matrix $\mathbf{P}$
5: Update sparse code $\mathbf{A}$ (FISTA)
6: Update dictionary $\mathbf{D}$ (FISTA)
7: Update classifier $\mathbf{W}$, $\mathbf{b}$
8: **end while**
Output: $\mathbf{D}, \mathbf{A}, \mathbf{W}, \mathbf{b}, \mathbf{P}$

In an iteration, actives points are only updated at the beginning to guarantee the convexity for next optimization problems.

Usage of unlabelled samples



In training phase, while fixing number of labelled samples and increasing number of unlabelled samples :

* The accuracy rates increase and converge to a stable value.

* The accuracy rate for testing samples reaches up the one for training unlabelled samples.

From these two remarks, we infer that no need to use the whole unlabelled data in training to obtain the optimal accuracy rate for testing data.

SSDL Model Formulation

Almost SSDL models can be presented with three main objectives : $\mathcal{R}$ Reconstruction, $\mathcal{D}$ Discrimination and $\mathcal{F}$ Preservation :

$$\min_{\Theta} \left[\mathcal{R}(\mathbf{A}, \mathbf{D}) + \mathcal{D}(\mathbf{W}, \mathbf{b}, \mathbf{A}, \mathbf{D}, \mathbf{P}) + \mathcal{F}(\mathbf{D}, \mathbf{A})\right],$$
$$\text{where } \Theta = \{\mathbf{W}, \mathbf{b}, \mathbf{A}, \mathbf{D}, \mathbf{P}\},$$

Input : $\mathbf{X} = [\mathbf{X}^l, \mathbf{X}^u]$ labelled and unlabelled samples and $\mathbf{Y}^l$ label matrix for $\mathbf{X}^l$.

Output: $\mathbf{D}$ dictionary, $\mathbf{A} = [\mathbf{A}^l, \mathbf{A}^u]$ sparse code, $[\mathbf{W}, \mathbf{b}]$ linear classifier, $\mathbf{P}$ class estimating probability of unlabelled sample in training.

Sparse code of testing sample

A testing sample $\mathbf{x}$ is sparse coded by taking into account also manifold structure preservation :

$$\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \beta \mathcal{F}'(\mathbf{a}) + \lambda \|\mathbf{a}\|_1 ,$$

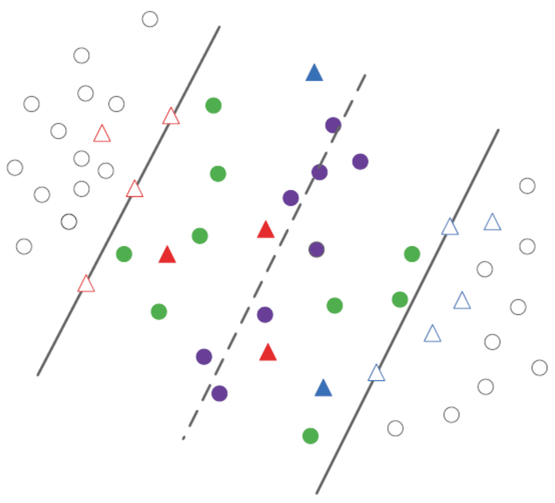
where $\mathcal{F}'$ preserves locally linear relationship between $\mathbf{x}$ and its k nearest neighbor among training samples into sparse code $\mathbf{a}$.

Our approach SSDL-GA

* $\mathcal{R}(\mathbf{A}, \mathbf{D}) = \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + \lambda \|\mathbf{A}\|_1$: reconstruction error for both labelled and unlabelled samples.

* $\mathcal{F}(\mathbf{A}) = \beta \operatorname{tr}(\mathbf{A}L\mathbf{A}^\top)$, where L is a matrix that learns the nature of manifold structure from original presentation $\mathbf{X}$. In this case, L is performed according to LLE strategy [1]: maintaining locally linear relationships between each sample and its k nearest neighbors into sparse code.

* $\mathcal{D}(\mathbf{W}, \mathbf{b}, \mathbf{A}, \mathbf{P}) = \gamma\Big(\sum_{i \in I_c} \sum_{c} \left\|y_i^c(\mathbf{w}_c^\top \mathbf{a}_i^l + b_c) - 1\right\|_2^2 + \sum_{k} \sum_{j \in J_c} (\mathbf{P}[k,j])^r \sum_{c} \left\|y_j^c(k)(\mathbf{w}_c^\top \mathbf{a}_j^u + b_c) - 1\right\|_2^2\Big)$
$+ \mu(\|\mathbf{W}\|_F^2 + \|\mathbf{b}\|_2^2)$, a semi-supervised classifier learnt by active points (sparse codes), with "one vs all" strategy [2]. Here an illustrative for two-class (red and blue) active points example, which can be considered as a decision boundary $\mathbf{w}_c, b_c$ for class c :



The black dotted line depicts the decision boundary, the margin is determined by two solid lines. The triangles and circles denote the labeled and unlabeled sparse codes, respectively. The solid shapes denote active sparse codes which are within the margin. All unlabelled sparse codes are associated with probability $\mathbf{P}$ based on distance between them and decision boundary. The green solid circles are active unlabelled sparse codes of high probability, while the purple circles are ones of low probability. Then only active points are used to rectify more sophisticatedly decision boundary in next iteration.

Conclusion

SSDL methods in general can be considered as classifier that outperforms other ones in case of low number of labelled samples. SSDL-GA, in particular, by integrating at the same time manifold structure preservation and internal semi-supervised classifier, improves significantly performance compared other SSDL methods.

References

[1] B. Babagholami-Mohamadabadi, A. Zarghami, M. Zolfaghari, and M. Baghshah. Pssdl: Probabilistic semi-supervised dictionary learning. In *Machine Learning and Knowledge Discovery in Databases*, pages 192–207, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[2] X. Wang, X. Guo, and S. Z. Li. Adaptively unified semi-supervised dictionary learning with active points. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1787–1795, Dec 2015.

Acknowledgements

This research is supported by the European Community through the grant DEDALE (contract no. 665044) and the Cross-Disciplinary Program on Numerical Simulation of CEA, the French Alternative Energies and Atomic Energy Commission.