Greedy structured dictionaries for fast sparse coding.

Y-Lan Boureau[1,2], Jean Ponce[2,3], and Yann LeCun[1]

[1] The Courant Institute of Mathematical Sciences - New York University
715 Broadway, 12th Floor, New York, NY 10003
[2] INRIA - WILLOW Project-Team
Laboratoire d'Informatique de l'Ecole Normale Supérieure (INRIA/ENS/CNRS UMR 8548)
23, avenue d'Italie 75214 Paris CEDEX 13, France
[3] Ecole Normale Supérieure

Many feature extraction techniques for image recognition in recent years implement some variant of sparse coding [6] within a processing pipeline that alternates coding and pooling operations (e.g., [1, 2, 4, 7, 8]). The resulting feature vectors can then be fed to a linear classifier such as a support vector machine.

Despite its popularity, sparse coding suffers from several shortcomings. When increasing dictionary size, the computational cost of sparse coding often becomes impractical before performance saturates for realistic datasets, with dictionaries of tens of thousands of atoms performing better than smaller ones. Restricting the set of atoms that can be activated to encode a given input reduces the size of the effective dictionary over which the optimization is performed. This general idea underlies several recent proposals such as locally linear coding (LLC) [7], which selects the active set among the k nearest neighbors of the input to be encoded, or work by Yang et al. [9], which first clusters the data, then learns a separate smaller dictionary within each cluster, with the full underlying dictionary having more than 250,000 atoms.

Another limitation of vanilla sparse coding is that atoms in a dictionary are all treated as equals, with no structure to define a hierarchy, or excitatory/inhibitory interactions between atoms. Jenatton et al. [5] and Gregor et al. [3] impose structure on the dictionary by using specific regularization penalties, e.g, allowing activation of a given atom to gate that of other atoms.

Here, we combine these two lines of thought by decomposing sparse encoding into stages of increased refinement:

- Each input is decomposed over a small dictionary with a regular sparse coding procedure.
- The resulting code is fed to a linear classifier that selects a latent class.
- Each latent class is assigned a given active set chosen among the atoms of a secondary dictionary.
- The residual after the first coarse decomposition is decomposed over that active set.

The classifier and dictionaries are trained concurrently. This process is fast given the small sizes of the classifiers involved, and can be reiterated several times to yield dictionaries of increasing specialization. We show how this framework relates to previous approaches and gives them more flexibility, while using the same computations to select the active set and perform the coding, and present experiments on several classification tasks.

References

[1] Y. Boureau, J. Ponce, and Y. LeCun. A theoretical analysis of feature pooling in vision algorithms. In *ICML*, 2010.

[2] S. Gao, I. Tsang, L. Chia, and P. Zhao. Local features are not lonely–Laplacian sparse coding for image classification. In *CVPR*, 2010.

[3] K. Gregor, A. Szlam, and Y. LeCun. Structured sparse coding via lateral inhibition. In *Advances in Neural Information Processing Systems (NIPS 2011)*, volume 24, 2011.

[4] K. Jarrett, K. Kavukcuoglu, M. Ranzato, and Y. LeCun. What is the best multi-stage architecture for object recognition? In *ICCV*, 2009.

[5] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach. Proximal methods for sparse hierarchical dictionary learning. In *International Conference on Machine Learning (ICML)*, 2010.

[6] B. A. Olshausen and D. J. Field. Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision Research*, 37:3311–3325, 1997.

[7] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong. Locality-constrained linear coding for image classification. In *CVPR*, 2010.

[8] J. Yang, K. Yu, Y. Gong, and T. Huang. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification. In *CVPR*, 2009.

[9] J. Yang, K. Yu, and T. Huang. Efficient Highly Over-Complete Sparse Coding using a Mixture Model. *ECCV*, 2010.

Topic: vision
Preference: oral