Non-parametric Bayesian Dictionary Learning with Landmark-Dependent Hierarchical Beta Process

[1]Mingyuan Zhou, [2]Hongxia Yang, [3]Guillermo Sapiro, [2]David Dunson and [1]Lawrence Carin

[1]Department of Electrical and Computer Engineering
Duke University, Durham, NC 27708, USA
{mz1,lcarin}@ee.duke.edu
[2]Department of Statistical Science
Duke University, Durham, NC 27708, USA
{hy35,dunson}@stat.duke.edu
[3]Department of Electrical and Computer Engineering
University of Minnesota, Minneapolis, MN 55455, USA
{guille}@umn.edu

Abstract

There has been significant recent interest in dictionary learning and sparse coding, with applications in denoising, interpolation, feature extraction, and classification [1]–[3]. Increasingly it has been recognized that these models may be improved by imposing additional prior information, beyond sparseness. For example, a locality constraint has been used successfully in the context of feature learning and image classification [4]. Structured sparsity has been used for compressive sensing [5]. Other examples include hierarchical tree-based dictionary learning [6], submodular dictionary selection [7], and exploitation of self-similarity in images [8].

We propose a landmark-dependent hierarchical beta process to address dictionary learning for data that are endowed with an associated covariate. We explore the idea that data nearby in the covariate space ($e.g.$, nearby in terms of temporal, spatial or cosine distances) are likely to share similar sparseness properties, and hence we employ "landmarks" to guide the usage probabilities of dictionary atoms. We address covariate-dependent dictionary learning from a Bayesian perspective. Compared to many optimization-based approaches, which assume the noise is Gaussian and its variance is known [1], our model has the advantage of automatically inferring an appropriate dictionary size and the underlying noise statistics.

Our model is connected to recent research devoted to removing the exchangeability assumption of the IBP [9], [10] and BP [11]. The exchangeability assumption ignores relational information provided by covariates. A dependent IBP (dIBP) model has been introduced recently, with a hierarchical Gaussian process (GP) used to account for covariate dependence [12]. In the proposed model, rather than imposing relational information via a parametric covariance matrix, as in GP, we do so by employing a kernel-based construction. We introduce "landmarks" in the covariate space, whose positions are learned, defining local regions where the dictionary usages are likely to be similar. The normalized kernels are localized via learned "landmarks," establishing links between data and landmark-dependent sparseness properties. The proposed model is related to the kernel stick breaking process (KSBP) [13] and Bayesian density regression (BDR) [14], although it is distinct from both. For example, the original KSBP construction focused primarily on covariate-dependent mixture models, and here we extend such ideas to a sparse factor analysis (SFA) model for covariate-dependent dictionary learning and sparse coding; we also

employ the kernel in a distinct manner. The proposed model may be viewed as a landmark-based extension of BDR, which is considerably more efficient than the original BDR. The proposed model removes the exchangeability assumption in the hierarchical beta process (HBP) [11], and we refer to it as a landmark-based dependent HBP, or Landmark-dHBP.

The proposed model employs a robustness term to model sparse spiky noise or localized data anomalies, related to robust principal component analysis (RPCA) [15], [16]. However, our model differs from RPCA in that the low-rank assumption on data is replaced with the richer covariate-dependent union-of-subspace assumption, which is realized with a sparse factor analysis model using the Landmark-dHBP as the prior. Therefore, our model might better handle data that violate the low-rank assumption, and the use of covariates imposes a locality constraint that is important when handling spiky or anomalous data components.

We present efficient inference using hybrid Gibbs, Metropolis-Hastings and slice sampling. The unique aspects of the proposed model, including covariate-dependent local latent feature discovery and non-local sparse spiky components modeling, are demonstrated on image-processing and document-analysis applications.

References

[1] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.*, 2006.

[2] M. Zhou, H. Chen, J. Paisley, L. Ren, G. Sapiro, and L. Carin. Non-parametric bayesian dictionary learning for sparse image representations. In *NIPS*, 2009.

[3] Y.-L. Boureau, F. Bach, Y. LeCun, and J. Ponce. Learning mid-level features for recognition. In *CVPR*, 2010.

[4] K. Yu, T. Zhang, and Y. Gong. Nonlinear learning using local coordinate coding. In *NIPS*, 2009.

[5] J. Huang, T. Zhang, and D. Metaxas. Learning with structured sparsity. In *ICML*, 2009.

[6] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach. Proximal methods for sparse hierarchical dictionary learning. In *ICML*, 2010.

[7] A. Krause and V. Cevher. Submodular dictionary selection for sparse representation. In *ICML*, 2010.

[8] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *ICCV*, 2009.

[9] T.L. Griffiths and Z. Ghahramani. Infinite latent feature models and the indian buffet process. In *NIPS*, 2005.

[10] D. Knowles and Z. Ghahramani. Infinite sparse factor analysis and infinite independent components analysis. In *Independent Component Analysis and Signal Separation*, 2007.

[11] R. Thibaux and M. I. Jordan. Hierarchical beta processes and the Indian buffet process. In *AISTATS*, 2007.

[12] S. Williamson, P. Orbanz, and Z. Ghahramani. Dependent Indian buffet processes. In *AISTATS*, 2010.

[13] D.B. Dunson and J.-H. Park. Kernel stick-breaking processes. *Biometrika*, 2008.

[14] D.B. Dunson, N.S. Pillai, and J.-H. Park. Bayesian density regression. *J. Royal. Statist. Soc B.*, 2007.

[15] J. Wright, Y. Peng, Y. Ma, A. Ganesh, and S. Rao. Robust principal component analysis: Exact recovery of corrupted low-rank matrices by convex optimization. In *NIPS*, 2009.

[16] E.J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Submitted to Journal of the ACM*, 2009.

Topic: learning algorithms
Preference: oral/poster