

Sparse Kernel Dictionary Learning

Cian O’Brien, Mark D. Plumbley
Centre for Vision, Speech and Signal Processing
University of Surrey, UK.
{cj.obrien, m.plumbley}@surrey.ac.uk

Abstract

Dictionary Learning (DL) has seen widespread use in signal processing and machine learning. Given a data set, DL seeks to find a so-called ‘dictionary’ such that the data can be well represented by a sparse linear combination of dictionary elements. The representational power of DL may be extended by the use of kernel mappings, which implicitly map the data to some high dimensional feature space. In Kernel DL we wish to learn a dictionary in this underlying high-dimensional feature space, which can often model the data more accurately than learning in the original space. Kernel DL is more challenging than the linear case however since we no longer have access to the dictionary atoms directly – only their relationship to the data via the kernel matrix. One strategy is therefore to represent the dictionary as a linear combination of the input data whose coefficients can be learned during training [1], relying on the fact that any optimal dictionary lies in the span of the data.

A difficulty in Kernel DL is that given a data set of size N , the full $(N \times N)$ kernel matrix needs to be manipulated at each iteration and dealing with such a large dense matrix can be extremely slow for big datasets. Here, we impose an additional constraint of sparsity on the coefficients so that the learned dictionary is given by a *sparse* linear combination of the input data. This greatly speeds up learning, and furthermore the speed-up is greater for larger datasets and can be tuned via a dictionary-sparsity parameter. The proposed approach thus combines Kernel DL with the ‘double sparse’ DL model [2] in which the learned dictionary is given by a sparse reconstruction over some base dictionary (in this case, the data itself). We investigate the use of sparse Kernel DL as a feature learning step for a music transcription task and compare it to another Kernel DL approach based on the K-SVD algorithm [1] (which doesn’t lead to sparse dictionaries in general), in terms of computation-time and performance. Initial experiments show that Sparse Kernel DL is significantly faster than the non-sparse Kernel DL approach ($6\times$ to $8\times$ speed-up depending on the size of the training data and the sparsity level) while leading to similar performance.

References

- [1] Hien Nguyen, Vishal M. Patel, Nasser M. Nasrabadi, and Rama Chellappa. Kernel dictionary learning. In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2021–2024, 2012.
- [2] Ron Rubinfeld, Michael Zibulevsky, and Michael Elad. Double sparsity: Learning sparse dictionaries for sparse signal approximation. *IEEE Transactions on Signal Processing*, 58(3):1553–1564, 2010.