
A spectral approximation to the Indian Buffet Process

J Lloyd, Z Ghahramani, C Reed

Department of Engineering
Cambridge University

Abstract

Literature review of spectral clustering for maximum-likelihood IBP formulation.

1 Paper Summaries

1.1 Fast Approximate Spectral Clustering: Yan *et al.* [8]

Yan *et al.* [8] discuss fast approximate spectral clustering framework based on minimizing the distortion (misclustering) of the resulting clusters as compared to non-approximate spectral clustering. Specifically, they introduce a K-means and RP-trees preprocessing step that forms a set of representative points and then performs spectral clustering on the representative points. These algorithms are referred to as *KASP* and *RASP*, respectively. The complexity of *KASP* is $O(knt) + O(k^3)$ while *RASP* is $O(hn) + O(k^3)$, where k is the number of representative points, n is the number of data points, t is the number of iterations of the preprocessing algorithms, and h is the depth of the RP tree.

Yan *et al.* [8] use a perturbation analysis of the graph Laplacian to show that the misclustering of (K—R)ASP converges to 0 as the number of representative points increases. Furthermore, they show that their techniques perform much better than K-means and comparable to a Nyström approximation but with smaller memory requirements and runtime.

1.2 Non-redundant Multi-view Clustering via Orthogonalization: Cui *et al.* [1]

Cui *et al.* [1] propose general methods for forming multiple clustering solutions for a given dataset, where each clustering is in a different subspace. Their method operates iteratively by clustering the data, orthogonalizing the data to a new subspace that is not covered by the previous clusters, clustering the data in the new subspace, et cetera until reaching some form of convergence. The authors discuss two approaches to performing this clustering: orthogonal clustering and clustering in orthogonal spaces.

Orthogonal clustering: given current data $X^{(t)}$ and the clustering solution of $X^{(t)}$, $M^{(t)} = [\mu_1^{(t)} \mu_2^{(t)} \dots \mu_k^{(t)}]$, Cui *et al.* [1] view orthogonal clustering as projecting the data points from the original data space to a compressed space (subspace) spanned by the mean vectors. They describe two variations for performing clustering in the space that is orthogonal to the compressed data space based on the hard and soft clustering views:

- hard clustering: each data point $x_i^{(t)}$ belonging to cluster j is projected onto the cluster center $\mu_j^{(t)}$ and $x_i^{(t+1)}$ is determined as $x_i^{(t)}$ projected onto the subspace orthogonal to the cluster centroids:

$$x_i^{(t+1)} = (I - \mu_j^{(t)} \mu_j^{(t)T} / \mu_j^{(t)T} \mu_j^{(t)}) x_i^{(t)}$$

- soft clustering: since each data point can partially belong to each cluster, we can project $X^{(t)}$ onto all cluster means and compute $X^{(t+1)}$ as $X^{(t)}$ projected onto the subspace or-

thogonal to all the cluster centroids:

$$X^{(t+1)} = (I - M^{(t)}(M^{(t)T}M^{(t)})^{-1}M^{(t)T})X^{(t)}$$

Clustering in orthogonal subspaces: Cui *et al.* [1] view clustering in orthogonal spaces as projecting the data into a reduced dimensional space that discriminates the given classes: i.e. via latent discriminant analysis or PCA. In particular, given current data $X^{(t)}$ and the clustering solution of $X^{(t)}$, $M^{(t)} = [\mu_1^{(t)} \mu_2^{(t)} \cdots \mu_k^{(t)}]$, the authors find the PCA solution of $M^{(t)}$, keep the $k^{(t)} - 1$ eigenvectors to obtain the subspace $A^{(t)}$ that captures the current clustering, and then projects $X^{(t)}$ to the space orthogonal to $A^{(t)}$:

$$X^{(t+1)} = (I - A^{(t)}(A^{(t)T}A^{(t)})^{-1}A^{(t)T})X^{(t)}$$

2 Literature to Investigate

[2] [4] [1] [5] [6] [7] [3] [8]

References

- [1] Cui, Y., Fern, X. Z., and Dy, J. G. (2007). Non-redundant Multi-view Clustering via Orthogonalization. *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, **3**, 133–142.
- [2] Griffiths, T. L. and Ghahramani, Z. (2011). The Indian Buffet Process : An Introduction and Review. *Journal of Machine Learning Research (JMLR)*, **12**, 1185–1224.
- [3] Higham, D. and Kibble, M. (2004). A unified view of spectral clustering. *University of Strathclyde mathematics research ...*, (53441), 1–17.
- [4] Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, **17**(4), 395–416.
- [5] Niu, D., Dy, J. G., and Jordan, M. I. (2010). Multiple Non-Redundant Spectral Clustering Views. In *Proceedings of the International Conference on Machine Learning (ICML)*.
- [6] Niu, D., Dy, J. G., and Jordan, M. I. (2011). Dimensionality Reduction for Spectral Clustering. In *Proceeding of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 15.
- [7] Niu, D., Dy, J. G., and Ghahramani, Z. (2012). A Nonparametric Bayesian Model for Multiple Clustering with Overlapping Feature Views. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*.
- [8] Yan, D., Huang, L., and Jordan, M. (2009). Fast approximate spectral clustering. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 907–916. ACM.