From Clustering to Co-clustering: Generative Modeling Approach

D. Lashkari, P. Golland
Computer Science and Artificial Intelligence Lab, MIT
32 Vassar Street
Cambridge, MA 02139
danial@mit.edu

Co-clustering is the problem of simultaneously clustering rows and columns of a matrix of data points. There has been a new interest in the problem due to the newly emerging applications, e.g., in biological data analysis [1]. In traditional clustering, generative models provide algorithms that can be understood as statistical generalizations of the k-means algorithm. By turning the original combinatorial clustering into a continuous problem, this approach enables efficient search of the space of solutions and expresses uncertainty in cluster memberships. In this work, we develop a general generative model for co-clustering which is similarly related to basic hard-assignment co-clustering algorithms such as Bregman co-clustering [2]. We explicitly demonstrate that common modeling assumptions are shared between our model and the hard co-clustering algorithms. More specifically, our approach leads to algorithms that find MAP solutions for the model while the hard co-clustering solves the corresponding ML problem. Making the connection to the mixture-modeling, we also show how this MAP vs. ML distinction corresponds to the soft vs. hard clustering assignments. We derive algorithms for both Gaussian and co-occurrence data and show empirically that they offer better performance than their hard-assignment counterparts especially with increasing problem complexity.

The simplicity of our model makes it possible to use mean field theory to derive co-clustering algorithms, maintaining close connection to mixture-model clustering. This is in contrast to the previous work on model-based co-clustering that depart from the basic clustering framework of the mixture-model clustering by building sophisticated models that are inspired by other relevant learning algorithms. For instance, employing Dirichlet processes for the membership priors leads to an MCMC-based inference in [3]; the LDA-like Dirichlet priors add an extra level of hierarchy to the problem [4]. The membership priors in our model are simple multinomial distributions similar to the basic mixture-models.

Topic: learning algorithms
Preference: oral/poster

Our motivation for co-clustering specifically comes from the analysis of functional brain imaging data where the data matrix represents the brain response of different brain locations (voxels) to the set of experimental stimuli. While clustering functional data has been traditionally utilized to group the voxels together based on their response, we aim to generalize this approach to simultaneously find a clustering of the presented stimuli as well. For instance, in the studies of object recognition with functional MRI, our method enables us to discover the categories of visual stimuli intrinsic to the brain visual system along with the clusters of voxels with similar patterns of selectivity. The simplicity of our model allows addition of relevant structures to the model that are specific to the fMRI data.

References

[1] S. C. Madeira, A. L. Oliveira, "Biclustering Algorithms for Biological Data Analysis: A Survey", IEEE/ACM Trans. Comput. Biol. Bioinf., Vol. 1, No. 1, pp. 24–44 (2004).

[2] A. Banerjee, I. Dhillon, and J. Ghosh, S. Merugu, D. S. Modha, "A Generalized Maximum Entropy Approach to Bregman Co-Clustering and Matrix Approximation," JMLR, No. 8, pp. 1919–1986, (2007).

[3] C. Kemp, J.. B. Tenenbaum, T. L. Griffiths, T. Yamada, and N. Ueda, "Learning systems of concepts with an infinte relational model," Proceedings of AAAI-06 (2006).

[4] H. Shan and A. Banerjee, "Bayesian Co-clustering", IEEE International Conference on Data Mining (ICDM) (2008).