POTTS MODEL CLUSTERING

A. Murua[†1], L. Stanberry[2], W. Stuetzle[2], D. Cordes[2]

[1] *Université de Montréal, Montréal, Canada*
[2] *University of Washington, Seattle, USA*

[†] E-mail: *murua@dms.umontreal.ca*

In this work we re-introduce Potts model clustering. We built on the work of Blatt, Wiseman and Domany (1996) who, borrowing from known algorithms in physics, used the Potts model as a general tool for data clustering. We show that Potts model clustering is linked to kernel K-means and the MNCut methods. Moreover, we embed all these methods in a single framework within which Potts model arises as a simple modification of kernel K-means. We also show that a slightly modified version of both Potts model clustering and kernel K-means (a penalized Potts model clustering and a weighted kernel K-means, respectively), solve the same clustering problem. We introduce an algorithm, a penalized version of the Wolff algorithm, to uncover the cluster structure suggested by the penalized model. We also show that kernel-based methods are linked to non-parametric kernel density estimation. We use this link to propose several estimates of the kernel bandwidths in order to improve the performance of the algorithms. We present some applications to gene expression and fMRI images. For the latter, we extend the voxel intensity data with the spatial neighborhood structure inherent in the 3D images. Our findings indicate that by so-doing Potts model clustering induces large clusters to attract neighboring loosely connected small clusters, and hence it tends to yield more coherent clustering structures.