



POLITECNICO
MILANO 1863

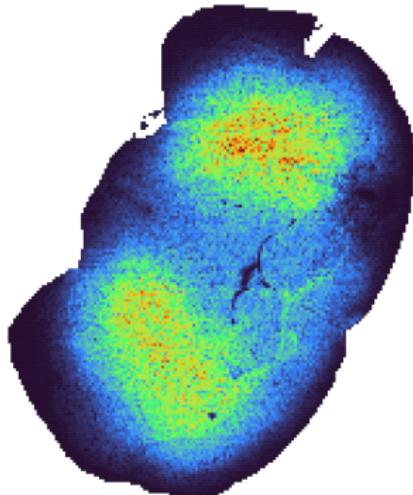
Scalable Bayesian Image Segmentation of Brain Spectroscopy Data

**Simone Colombara, Alessia Cotroneo, Francesco De Caro,
Riccardo Morandi, Chiara Schembri, and Alfredo G. Zapiola**

Tutors: Giulia Capitoli and Francesco Denti

14 February 2023

The goal of the project is to perform image segmentation.



Recall that:

- we have 3 datasets (mass spectrometry measurements about lipids, glycans and peptides) with about 18k pixels;
- for each pixel we have a signal (spectrum) that we preprocessed using fPCA;
- the aim is to cluster the pixels considering the spatial correlation.

- As our first Bayesian model, we fitted a **univariate** Gaussian mixture model using the first functional principal component score to obtain a benchmark result. This model **does not** take into account any spatial information.
- To include the spatial dependencies we fitted a **univariate** Hidden Potts Model, obtaining less noisy results.
- The natural development is to extend the models to the **multidimensional** case, in order to consider more than one fPC or the three first fPCs of the three different datasets.

To investigate if the clustering can be improved by taking into account **more than one principal component**, we want to focus on multivariate likelihoods.

As the first step, we extended the previous model to a multidimensional GMM:

$$\begin{aligned} \mathbf{y}_i \mid z_i = k, \boldsymbol{\mu}, \boldsymbol{\Sigma} &\stackrel{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \\ z_i \mid \mathbf{w} &\sim \text{Categorical}(\mathbf{w}) \\ \boldsymbol{\mu}_k &\sim \mathcal{N}(\mathbf{b}_{k,0}, \mathbf{B}_{k,0}) \\ \boldsymbol{\Sigma}_k &\sim \text{Inv-Wishart}(V_{k,0}, n_{k,0}) \\ \mathbf{w} &\sim \text{Dirichlet}(\boldsymbol{\lambda}) \end{aligned} \tag{1}$$

This Model **does not** take into account any spatial information.

Multivariate GMM: Gibbs sampler

4/23

To sample from the posterior distributions we used a **Gibbs sampler**. The full conditionals of the model are:

$$\boldsymbol{\mu}_k \mid \Sigma, \mathbf{y}, \mathbf{z} \sim \mathcal{N}(\mathbf{b}_{k,p}, B_{k,p})$$

$$\Sigma_k \mid \boldsymbol{\mu}, \mathbf{y}, \mathbf{z} \sim \text{Inv-Wishart}(V_{k,p}, n_{k,p})$$

$$\mathbf{w} \mid \mathbf{z} \sim \text{Dirichlet}(\lambda_1 + N_1, \dots, \lambda_K + N_K)$$

$$z_i \mid \mathbf{w}, \boldsymbol{\mu}, \Sigma, \mathbf{y}, \sim \text{Cat} \left(\frac{w_1 \phi(\mathbf{y}_i \mid \boldsymbol{\mu}_1, \Sigma_1)}{\sum_{k=1}^K w_k \phi(\mathbf{y}_i \mid \boldsymbol{\mu}_k, \Sigma_k)}, \dots, \frac{w_K \phi(\mathbf{y}_i \mid \boldsymbol{\mu}_K, \Sigma_K)}{\sum_{k=1}^K w_k \phi(\mathbf{y}_i \mid \boldsymbol{\mu}_k, \Sigma_k)} \right)$$

$$B_{k,p} = (N_k \Sigma_k^{-1} + B_{k,0}^{-1})^{-1}$$

$$\mathbf{b}_{k,p} = B_{k,p} (N_k \Sigma_k^{-1} \bar{\mathbf{y}}_k + B_{k,0}^{-1} \mathbf{b}_{k,0})$$

$$N_k = \sum_{i=k}^N \mathbb{1}_{\{z_i=k\}}$$

$$\bar{\mathbf{y}}_k = \frac{\sum_{i:z_i=k} \mathbf{y}_i}{N_k}$$

(2)

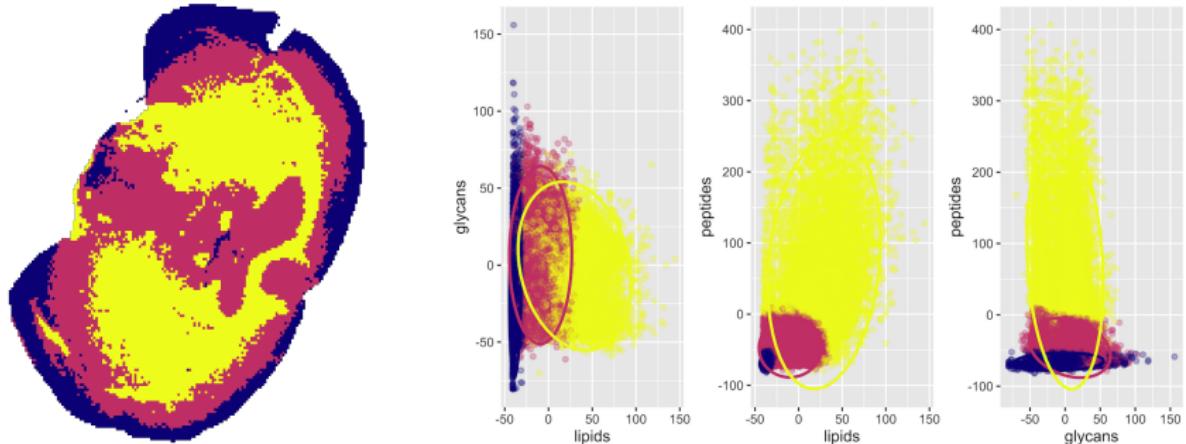
$$V_{k,p} = \left(V_{k,0}^{-1} + \sum_{i:z_i=k} (\mathbf{y}_i - \boldsymbol{\mu}_k)(\mathbf{y}_i - \boldsymbol{\mu}_k)^T \right)^{-1}$$

$$n_{k,p} = n_{k,0} + N_k$$

Multivariate GMM: results on combined datasets

5/23

First we jointly modeled the first fPC for each of the three molecular type.

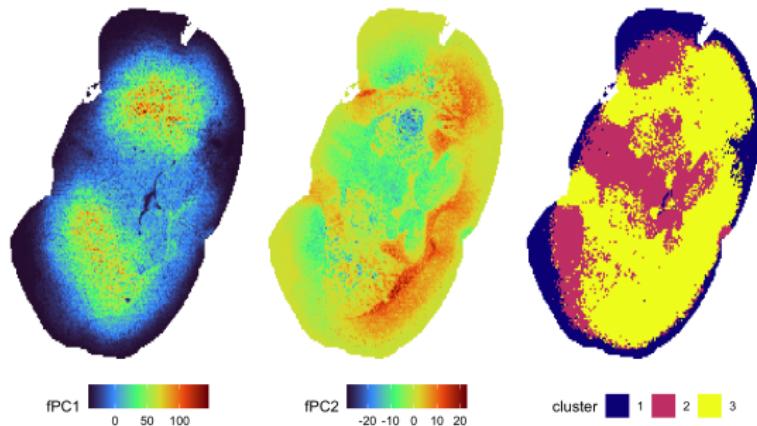


Combined dataset. Left-most panel: clustering obtained via GMM. Second, third and fourth panel: distribution of the fPC for each couple of molecular type and respective clustering allocation.

Multivariate GMM: results on lipids dataset

6/23

Then we jointly model more than one fPC for each dataset.



Lipids dataset. Left panel: first fPC distribution. Center panel: second fPC distribution. Right panel: the obtained clustering.

Note that the different components in a multivariate GMM all share the same “importance” when performing clustering, while usually the first fPC is much more representative (in terms of explained variance) than the second one.

We expect neighbouring pixels to have a higher probability of being clustered together. Therefore we specify a **hidden Markov random field** as the distribution of the membership labels, introducing a dependence of pixel i on its neighbouring pixels \mathcal{N}_i , via a Gibbs distribution:

$$p(z_i | \mathbf{z}_{\setminus i}, \beta) = \frac{\exp(\beta \sum_{l \in \mathcal{N}_i} \delta(z_i, z_l))}{\sum_{j=1}^k \exp(\beta \sum_{l \in \mathcal{N}_i} \delta(j, z_l))},$$

where β (*inverse temperature*) governs the degree of interaction between pixels.

The Potts model undergoes a **phase transition** at a critical value of β , switching from a disordered to an ordered state with a dominating cluster.

$$\beta_{critic} = \log(1 + \sqrt{K}).$$

We first start by fixing β a priori, choosing its best value by trial and error. Later on, we introduce it as part of the parameters to be estimated from the data.

The model becomes:

$$\begin{aligned} \mathbf{y}_i | z_i = k, \boldsymbol{\mu}, \boldsymbol{\Sigma} &\stackrel{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \\ \boldsymbol{\mu}_k &\sim \mathcal{N}(\mathbf{b}_{k,0}, B_{k,0}), \\ \boldsymbol{\Sigma}_k &\sim \text{Inv-Wishart}(V_{k,0}, n_{k,0}), \\ z_i | \mathbf{z}_{\setminus i} &\sim \text{Gibbs}(\beta). \end{aligned} \tag{3}$$

The Potts model can be viewed as a **spatially-correlated generalisation** of the GMM.

Note that the joint distribution of the pixel labels can be expressed in the form of an exponential family:

$$p(\mathbf{z} | \beta) = \exp\{\beta S(\mathbf{z}) - \log \mathcal{C}(\beta)\}.$$

The sufficient statistic $S(\mathbf{z}) = \sum_{i \sim l \in \mathcal{E}} \delta(z_i, z_l)$ represents the total number of like neighbour pairs in the image, while $\mathcal{C}(\beta)$ is a normalizing constant.

Potts Model: Gibbs sampler

9/23

In order to sample from the posterior we used a **Gibbs sampler** with the following full conditionals:

$$\begin{aligned}\mu_k \mid \Sigma, \mathbf{y}, \mathbf{z} &\sim \mathcal{N}(\mathbf{b}_{k,p}, B_{k,p}) \\ \Sigma_k \mid \mu, \mathbf{y}, \mathbf{z} &\sim \text{Inv-Wishart}(V_{k,p}, n_{k,p})\end{aligned}\tag{4}$$

where $\mathbf{b}_{k,p}, B_{k,p}, V_{k,p}, n_{k,p}$ have the same expression as in the GMM (2).

The cluster allocation variables differ:

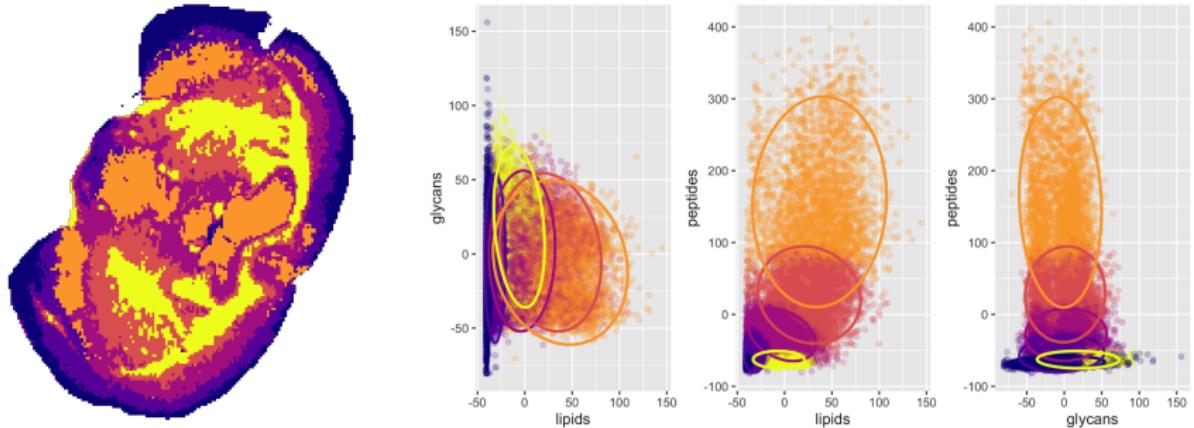
$$\begin{aligned}z_i \mid \mathbf{z}_{\setminus i}, \mu, \sigma^2, \mathbf{y} &\sim \text{Categorical}(\mathbf{p}_i) \\ (\mathbf{p}_i)_k = \frac{\phi(\mathbf{y}_i \mid \mu_k, \Sigma_k) \exp\{\beta \sum_{l \sim I} \delta(k, z_l)\}}{\sum_{k=1}^K \phi(\mathbf{y}_i \mid \mu_k, \Sigma_k) \exp\{\beta \sum_{l \sim I} \delta(k, z_l)\}}\end{aligned}\tag{5}$$

where $\phi(\mathbf{x} \mid \mu, \Sigma)$ represents the probability density function of a multivariate normal distribution with mean μ and covariance matrix Σ evaluated in \mathbf{x} .

Potts Model: results on combined dataset

10/23

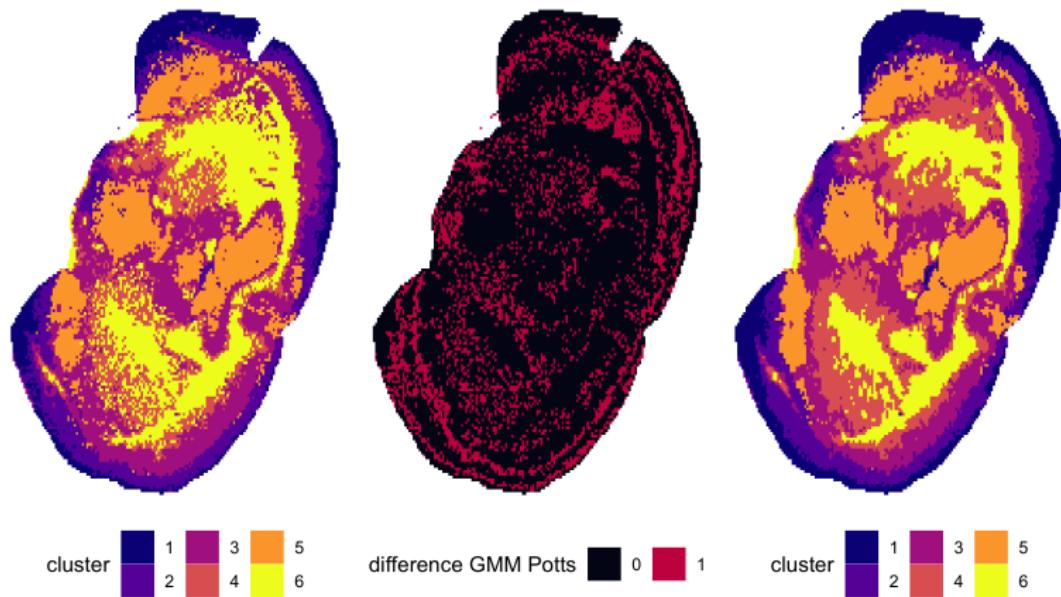
The multivariate Potts model allowed us to improve our clustering on the combined dataset. We found the best value of β to be 0.6, which is about half the value of β_{critic} for $K = 6$.



Combined dataset. Left-most panel: clustering obtained via multivariate HPM. Second, third and fourth panel: distribution of the fPC for each couple of molecular type and respective clustering.

Comparison GMM vs. Potts model

11/23



Combined dataset. Left panel: clustering via GMM. Right panel: clustering via HPM. Center Panel: the pixels for which the cluster allocation differs are highlighted in red.

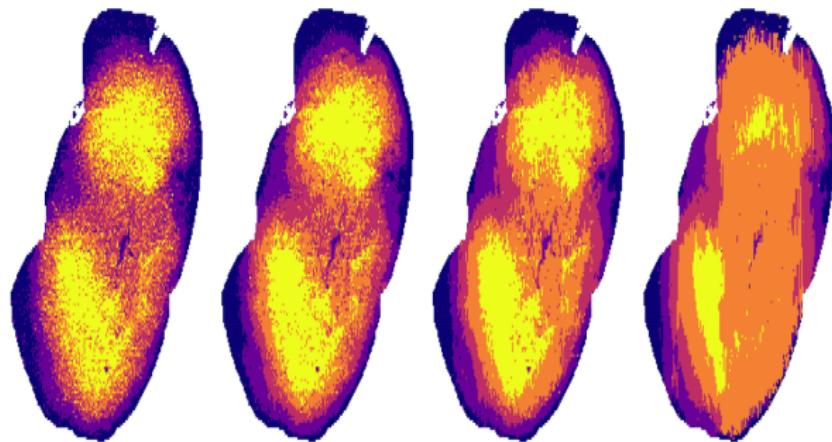
Comparing the clustering obtained via multidimensional GMM and via HPM on the combined dataset, using the same priors for the parameters, the differences are remarkable:

- The Potts model gives us a much more **uniform clustering**, as the isolated points are significantly reduced;
- In the Potts model the cluster have **sharper edges**;
- In this case around **25%** of the pixels have differing allocations under the two models.

The importance of β

13/23

Choosing an appropriate value for β is key to our problem: when $\beta = 0$ we have just a GMM. For low values of β , the cluster boundaries are not well defined and we have a noisier cluster. As we increase the beta, we get more homogeneous clusters but we loose detail about their shapes and we tend to have a dominant cluster in the image.



Lipids dataset: Potts model for different values of the inverse temperature. From left to right $\beta = 0, 0.5\beta_{critic}, \beta_{critic}, 1.5\beta_{critic}$.

The difficulties in fixing a correct value for beta lead us to incorporate the beta in the model as a **parameter to be estimated**. Since β takes real values and we know the value of the phase transition β_{critic} , we chose as prior:

$$\beta \sim \mathcal{U}([0, \beta_{max}]), \quad \beta_{max} > \beta_{critic}.$$

We opt for the uniform distribution since we want it to be non-informative.

We impose β larger than zero to make neighbouring pixels more likely to be clustered together.

We set the value of β_{max} to be higher than the critical temperature since we don't exclude, a priori, the presence of a dominating class. Being the phase transition quite abrupt, we are able to cover even cases with a highly dominant class, using a value of β_{max} of the order of 2 times β_{critic} .

Posterior sampling with β as a parameter

15/23

We can sample from the posterior distribution using a Gibbs sampler, in particular $\pi(\mu | \Sigma, \mathbf{y}, \mathbf{z}, \beta)$, $\pi(\Sigma | \mu, \mathbf{y}, \mathbf{z}, \beta)$ and $\pi(\mathbf{z} | \mu, \Sigma, \mathbf{y}, \beta)$ follow the same full conditional of equations (4) and (5).

On the other hand, $p(\beta | \mathbf{z})$ involves an intractable normalizing constant since

$$p(\beta | \mathbf{z}) = \frac{p(\mathbf{z} | \beta) \pi(\beta)}{\int p(\mathbf{z} | \beta) \pi(\beta) d\beta} \propto \frac{\exp\{\beta S(\mathbf{z})\}}{\mathcal{C}(\beta)} \pi(\beta).$$

The normalising constant $\mathcal{C}(\beta)$ has a computational complexity of $\mathcal{O}(nK^n)$, since it involves a sum over all possible configurations of the labels $\mathbf{z} \in \mathcal{Z}$:

$$\mathcal{C}(\beta) = \sum_{\mathbf{z} \in \mathcal{Z}} \exp\{\beta S(\mathbf{z})\}.$$

It is unfeasible to calculate this value exactly for large images: a **computational approximation** is required.

A possible approximation consists in replacing $p(\mathbf{z} \mid \beta)$ with the **pseudolikelihood**:

$$p(\mathbf{z} \mid \beta) \approx \prod_{i=1}^n p(z_i \mid \mathbf{z}_{\setminus i}, \beta) = \prod_{i=1}^n p(z_i \mid \mathbf{z}_{\delta_i}, \beta),$$

where z_{δ_i} indicates the z_j such that $j \in \mathcal{N}_i$.

The normalising constants for the factors of the pseudolikelihood are now computable.

Pseudolikelihood is exact when $\beta = 0$ and provides a reasonable approximation for small values of the inverse temperature.

The approximation error increases rapidly for $\beta \geq \beta_{critic}$, due to long-range dependence between the labels, which is inadequately modelled by the local approximation.

This approximation enables updates for the inverse temperature at a certain iteration of the Gibbs sampler to be simulated using a Metropolis-Hastings step.

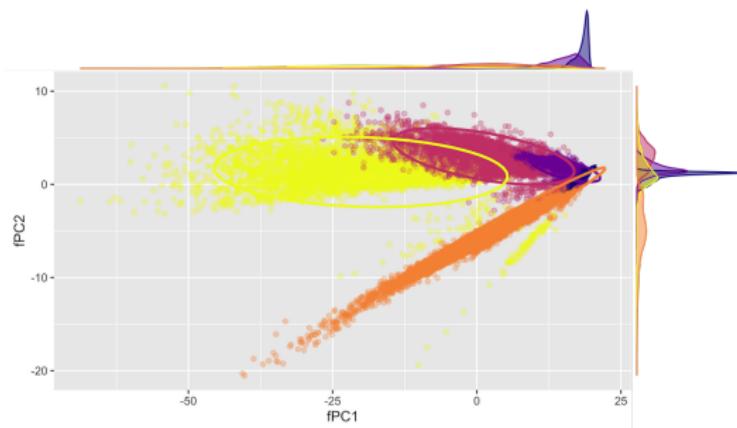
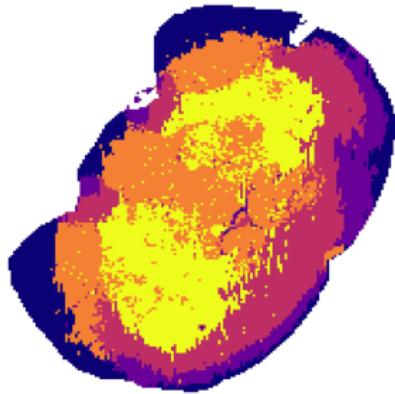
The proposal density $q(\beta' \mid \beta_{t-1})$ can, in theory, be any probability distribution.

In order to find the correct balance between exploration and acceptance rate, we used an **adaptive random walk Metropolis Hastings algorithm with Gaussian proposal**, which automatically log-linearly tunes the bandwidth (the variance of the proposal distribution) to target a specific Metropolis Hastings acceptance rate, set at 0.44.

Numerical results on lipids dataset

18/23

We are able to estimate a satisfactory clustering even when estimating β as a parameter. In this case the estimated value for β was around 1.4, which is higher than the critical value.



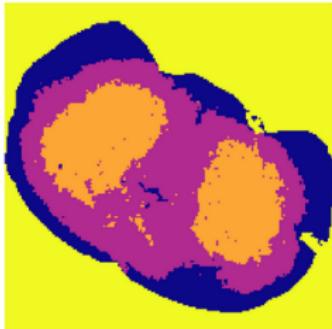
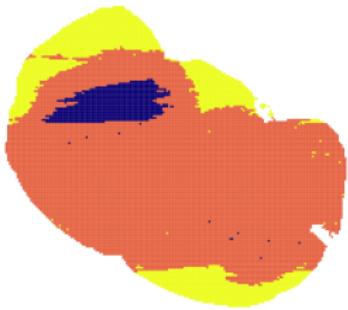
Lipids dataset: Left panel: clustering obtained via HPM with β as parameter. Right panel: distribution of the fPC and respective clustering allocation.

Numerical results: overestimating β

19/23

The tendency of the model to estimate β **larger** than β_{critic} was quite common. In some cases the estimate went considerably above the critical value and produced a clustering having a dominating class.

To partially fix this problem we considered a **square image** with an added background class, which helped the algorithm in estimating more suitable clusters, but it did not reduce the tendency of the estimate of β from being much higher than the critical value.



Lipids dataset: Left panel: an extreme case of overestimating β , resulting in a meaningless clustering. Right panel: clustering the square image using an auxiliary background class and the same prior parameters.

multiPotts: an Rcpp Library

20/23

multiPotts-package (multiPotts) R Documentation

Multivariate Potts model

Description

A library that fits the potts model utilising a mixture of multivariate normal distributions. The MCMC is run with Rcpp.

Package Content

Index of help topics:

GibbPotts	wrapper function for the gibbs sampler for the potts model with fixed beta
GibbsGMM	wrapper function for the gibbs sampler for the GMM
MCMCPotts	wrapper function for the gibbs sampler for the potts model with beta approximated by pseudolikelihood
mcmcPottsmd	multidimensional Potts model gibbs sampler with beta sampled via Pseudolikelihood
multiPotts-package	Multivariate Potts model

Maintainer

Riccardo Morandi <riccardo3.morandi@mail.polimi.it>, Alfredo G. Zapiola <agimenezzapiola@gmail.com>

Author(s)

R. Morandi, A. G.Zapiola, S. Colombara, F. De Caro, A. Cotroneo, C. Shembri



The **multiPotts R package** Left panel: a snip of the documentation. Right panel: installation instructions

In this project, we addressed the problem of image segmentation.

- We started from the standard **GMM**. Such model, however, does not take into account the spatial dependence between different pixels.
- We therefore moved to the **HPM**, introducing such spatial relation via a random Markov field on the cluster allocation variables.
 - ▶ This model, which can be seen as a spatially correlated generalization of a GMM, introduces a new parameter, the hidden temperature, which is quite difficult to estimate a priori.
- Therefore, we incorporated it as part of the parameters to be estimated. This choice introduced some complications since, in order to compute the posterior of the inverse temperature, we had to compute an intractable constant, which lead us to the approximate procedure of **pseudolikelihood** to obtain a computable approximation of the constant.
- We implemented in **Rcpp** an efficient Gibbs Sampler for all of the methods above, capable of handling images of arbitrary shape and number of channels.

There's still room for improvement, addressing the challenges that arose.
Some possible future developments include:

- Finding a way of **weighting** the fPCs, in order to give more importance to the ones explaining most of the variance, to obtain more meaningful clusters;
- Reducing the computational cost: **Variational Inference** could be used as a different model estimation technique, as opposed to Gibbs sampling;
- Exploring a **non-parametric** framework, in order to avoid fixing the number of clusters a priori.

-  **Matthew Moores, Geoff K. Nicholls, Anthony N. Pettitt, Kerrie Mengersen**
Scalable Bayesian Inference for the Inverse Temperature of a Hidden Potts Model
-  **Matthew Moores, Anthony N. Pettitt, Kerrie Mengersen**
R Package bayesImageS: Bayesian Methods for Image Segmentation using a Hidden Potts Model
-  **Giulia Capitoli**
New diagnostic frontiers in thyroid cytopathology: the role of MALDI-MS Imaging
-  **Julian Besag**
Statistical Analysis of Non-Lattice Data
-  **Tobias Rydén and D. M. Titterington**
Computational Bayesian Analysis of Hidden Markov Models