Mixture of multivariate multiple-scaled Student distributions : application to the characterization of brain tumors by multiparametric MRI.

Alexis Arnaud

Supervisors : F. Forbes, E. Barbier, B. Lemasson
Laboratories : LJK, INRIA & GIN
Funding : UJF

1st year-PhD
26 June 2015
Motivation:

- MRI data
- MMSD
- Tumor characterization
- Future work

Outline

1. Motivation: brain tumor characterization
2. Clustering of MRI data
3. Mixture of multivariate multiple-scaled Student distributions
4. Tumor characterization from multiparametric MRI
5. Work in progress
Outline

1. **Motivation**: brain tumor characterization

2. Clustering of MRI data

3. Mixture of multivariate multiple-scaled Student distributions

4. Tumor characterization from multiparametric MRI

5. Work in progress
How to characterize brain tumor?
Histology vs multiparametric MRI

Histology section

+ gold standard: provides the ground truth
+ precise information
  - local information: only parts of the tumor may be sampled
  - invasive operation (biopsy), not always feasible
How to characterize brain tumor?

Histology vs multiparametric MRI

2 MRI maps

- under development method
- global information: whole tumor visible
- non-invasive operation

Goal: find the voxels inside the MRI maps which belong to the tumor, in order to characterize a tumor, to avoid invasive biopsies.
How to characterize brain tumor?

Problem of multiparametric MRI

How to extract information from all of the parametric maps?

Approach: multivariate clustering with mixture models.
Outline

1. Motivation: brain tumor characterization
2. Clustering of MRI data
3. Mixture of multivariate multiple-scaled Student distributions
4. Tumor characterization from multiparametric MRI
5. Work in progress
Multiparametric MRI data

- 5 physiological parameters:
  - **ADC**: apparent diffusion coefficient
  - **CBV**: cerebral blood volume
  - **CBF**: cerebral blood flow
  - **AUC**: blood vessel permeability
  - **StO\textsubscript{2}**: oxygen saturation

- 5 dimensional data set:
  \[ Y = \{Y_1, \ldots, Y_N\} \] the set of all voxels, of size \( N \)
  with \( Y_n = \{Y_{n,\text{ADC}}, \ldots, Y_{n,\text{StO}_2}\} \) the measures on the \( n \)th voxel
Clustering of MRI data

We suppose that data arise from $K$ different classes (for $K$ different tissues), and we want to recover those classes.

Simulated data  |  Latent classification
Clustering of MRI data via mixture modeling

Let $Z$ be the latent variable which links one observation to one class:

\[
\begin{align*}
\{ (Y_n \mid Z_n = k) \sim f_k(\theta_k) \\
Z_n \sim \mathcal{M}(\pi_1, \ldots, \pi_K) \}
\end{align*}
\]

(1)

\[
\Leftrightarrow
\]

\[
\Pr(y_n \mid \theta) = \sum_{k=1}^{K} \pi_k f_k(y_n \mid \theta_k)
\]

(2)
Clustering of MRI data via mixture modeling

In a previous study, \( f_k(\theta_k) = \mathcal{N}_5(\mu_k, \Sigma_k) \) with \( \mu_k \in \mathbb{R}^5 \) and \( \Sigma_k \in \mathcal{S}_5^{+}(\mathbb{R}) \) : lack of flexibility in cluster shape modeling (Coquery et al. - 2014).

In this study, \( f_k(\theta_k) \) is a heavy-tailed distribution : a multivariate multiple-scaled Student distribution (MMSD).
Outline

1. Motivation: brain tumor characterization
2. Clustering of MRI data
3. Mixture of multivariate multiple-scaled Student distributions
4. Tumor characterization from multiparametric MRI
5. Work in progress
Standard multivariate Student distribution

One possible form of a $M$-dimensional distribution

$$p_{MS}(y; \mu, \Sigma, \nu) = \frac{\Gamma\left(\frac{\nu+M}{2}\right)}{|\Sigma|^{1/2} \Gamma\left(\frac{\nu}{2}\right) \left(\pi \nu\right)^{M/2}} \left[1 + \frac{\delta^2(y, \mu, \Sigma)}{\nu}\right]^{-\frac{\nu+M}{2}}$$

with: $$\delta^2(y, \mu, \Sigma) = (y - \mu)^t \Sigma^{-1} (y - \mu)$$

the Mahalanobis distance

But the degree of freedom remains scalar.
Standard multivariate Student distribution

Useful representation: infinite mixture of scaled Gaussians

\[ p_{MS}(y; \mu, \Sigma, \nu) = \frac{\Gamma \left( \frac{\nu+M}{2} \right)}{|\Sigma|^{\frac{1}{2}} \Gamma \left( \frac{\nu}{2} \right) (\pi \nu)^{\frac{M}{2}}} \left[ 1 + \frac{\delta^2(y, \mu, \Sigma)}{\nu} \right]^{-\frac{\nu+M}{2}} \]

\[ = \int_0^\infty \mathcal{N}_M \left( y; \mu, \frac{1}{w} \Sigma \right) \mathcal{G} \left( w; \frac{\nu}{2}, \frac{\nu}{2} \right) dw \]

with:

- \( \mathcal{N}_M \) the \( M \)-multivariate Gaussian distribution
- \( \mathcal{G} \) the Gamma distribution
- the real latent variable \( W \) is called the weight
Multivariate multiple-scaled Student distribution
Use of the eigenvalue decomposition to get a multidimensional degree of freedom (Forbes and Wraith - 2014)

Let \(U\) and \(D\) the eigenvalue decomposition of covariance matrix : \(\Sigma = UDU^t\), with :
\[
U \in \mathcal{O}(M) \text{ the orthogonal matrix of eigenvectors,}
\]
\[
D \in \mathcal{D}(M) \text{ the diagonal matrix of eigenvalues.}
\]

\[
p_{\text{MS}}(y ; \mu, U, D, \nu) = \int_{\mathbb{R}_+^M} \mathcal{N}_M \left( y ; \mu, \frac{1}{w}UDU^t \right) \mathcal{G} \left( w ; \frac{\nu}{2}, \frac{\nu'}{2} \right) \, dw
\]
Multivariate multiple-scaled Student distribution

Use of the eigenvalue decomposition to get a multidimensional degree of freedom (Forbes and Wraith - 2014)

Let $W \in \mathbb{R}_{+}^{M}$ a $M$-dimensional weight (one scalar weight for each dimension):

$$p_{\text{MMSD}}(y; \mu, U, D, \nu) =$$

$$\int_{\mathbb{R}_{+}^{M}} \mathcal{N}_{M}(y; \mu, U \text{diag}(w)^{-1} DU^{t}) \prod_{m=1}^{M} G(w_{m}; \frac{\nu_{m}}{2}, \frac{\nu_{m}}{2}) \, dw$$
Outline

1. Motivation: brain tumor characterization
2. Clustering of MRI data
3. Mixture of multivariate multiple-scaled Student distributions
4. Tumor characterization from multiparametric MRI
5. Work in progress
Multiparametric MRI data

- 26 parameters per class

- estimation by EM and Flury & Gautschi or Stiefel manifold optimization algorithms:
  - written in C++
  - use inside R with the Rcpp package

- choice of the number of classes according with BIC and ICL criterions:
  - 10 to 50 repetitions in parallel
  - use of the snow package

- data in dimension 5

- 8 healthy rats (49 000 voxels)

- 37 rats with tumors (290 000 voxels)

- 4 tumor models: 9L, C6, F98, RG2

Alexis Arnaud
Processing pipeline
Healthy voxels classification

Healthy MRI data
Healthy tissue classification
Tumor localization
Tumor dictionary
Tumor MRI data
Voxels classification
Processing pipeline
Tumor localization (ROI)
Processing pipeline

Voxels classification

Healthy MRI data → Healthy tissue classification → Tumor localization

Tumor MRI data

Voxels classification ➔ Tumor dictionary
Classification rate of discriminant analysis, by tumor: 50, 62.5, 75, 85.7%.

- Except for the 9L rats, the classification rates are as good as a previous study with a Gaussian model and a manual tumor delimitation.

Alexis Arnaud

Supervisors: F. Forbes, E. Barbier, B. Lemasson
Laboratories: LJK, INRIA & GIN

Rencontres R 2015 - Mixture of MMSD and tumors characterization
Outline

1. Motivation: brain tumor characterization
2. Clustering of MRI data
3. Mixture of multivariate multiple-scaled Student distributions
4. Tumor characterization from multiparametric MRI
5. Work in progress
Work in progress

- Validation of the protocol.
- Taking into account spatial dependences (Markov field).
- Automatic selection of the number of classes (Bayesian extension).
- Parameters sensitivity analysis.
- Link between histology and automatic tissue characterization.
Bibliography


The end

Thank you!

Image credit: dianephotos.deviantart.com