SpaCEM^3: a software for the spatial clustering of incomplete, high dimensional data

Florence Forbes[1], Matthieu Vignes[2]
http://spacem3.gforge.inria.fr/

[1]Mistis Project, INRIA Rhône-Alpes

[2]BIA Unit, INRA Toulouse

MSTGA - INRA Toulouse - 4 décembre 2009

Outline

Introduction

Outline

Outline

Introduction

Models included in the software for classifying objects

- Hidden Markov Random Fields
- Gaussian model for high-dimensional data
- Supervised classification with Triplet Markov fields

Algorithms available with the software

- Classical algorithms
- Variational (mean field-like) EM approximations for the Markovian modelling
- Practical use of the algorithms

Outline

Outline

Outline

Introduction

- Goal : classifying objects of interest (image pixels, genes . . .) from *complex* datasets *i.e.* grouping them into homogeneous groups as regards carried out measurements and possibly some prior knowledge.

Introduction

- Goal : classifying objects of interest (image pixels, genes . . .) from *complex* datasets *i.e.* grouping them into homogeneous groups as regards carried out measurements and possibly some prior knowledge.
- Peculiar features of data we focused on :
  - there are dependencies between objects,
  - data are high-dimensional and
  - some measures can be missing.

Introduction

- Goal : classifying objects of interest (image pixels, genes . . .) from *complex* datasets *i.e.* grouping them into homogeneous groups as regards carried out measurements and possibly some prior knowledge.
- Peculiar features of data we focused on :
  - there are dependencies between objects,
  - data are high-dimensional and
  - some measures can be missing.
- SpaCEM^3 tackles these requirements in a Markovian setting; dependencies are encoded in neighbourhood relationships.

Introduction

- Goal : classifying objects of interest (image pixels, genes . . .) from *complex* datasets *i.e.* grouping them into homogeneous groups as regards carried out measurements and possibly some prior knowledge.
- Peculiar features of data we focused on :
  - there are dependencies between objects,
  - data are high-dimensional and
  - some measures can be missing.
- SpaCEM^3 tackles these requirements in a Markovian setting; dependencies are encoded in neighbourhood relationships.
- Applications: Image analysis (biomedical, satellite surveys. . . More generally computer vision), genomics datasets. . . .

Included fonctionalities

- Unsupervised clustering based on Hidden Markov Random Fields (HMRF); can be seen as a generalization of Independent Mixture Models (IMM) with dependencies encoded in a graph (regular grid or general neighbourhood setting). Allows data to be high-dimensional, variables to be correlated and some observations to be missing.

Included fonctionalities

- Unsupervised clustering based on Hidden Markov Random Fields (HMRF); can be seen as a generalization of Independent Mixture Models (IMM) with dependencies encoded in a graph (regular grid or general neighbourhood setting). Allows data to be high-dimensional, variables to be correlated and some observations to be missing.
- Supervised classification when noise modelling is neither independent nor unimodal: learning and test steps based on Triplet Markov models.

Included fonctionalities

- Unsupervised clustering based on Hidden Markov Random Fields (HMRF); can be seen as a generalization of Independent Mixture Models (IMM) with dependencies encoded in a graph (regular grid or general neighbourhood setting). Allows data to be high-dimensional, variables to be correlated and some observations to be missing.
- Supervised classification when noise modelling is neither independent nor unimodal: learning and test steps based on Triplet Markov models.
- Model selection is performed with some criterion that selects the *best* model given the data. BIC, ICL and their approximations in a variational setting are included.

Included fonctionalities

- Unsupervised clustering based on Hidden Markov Random Fields (HMRF); can be seen as a generalization of Independent Mixture Models (IMM) with dependencies encoded in a graph (regular grid or general neighbourhood setting). Allows data to be high-dimensional, variables to be correlated and some observations to be missing.
- Supervised classification when noise modelling is neither independent nor unimodal: learning and test steps based on Triplet Markov models.
- Model selection is performed with some criterion that selects the *best* model given the data. BIC, ICL and their approximations in a variational setting are included.
- Simulation of the different models: MRF, HMRF and Triplet Markov models. Classical graphs (Delaunay, Gabriel, relative neighbours, ϵ neighbours, k-nearest neighbours) can be generated.

Technical characteristics

- Written in C++: 52 classes, $30,000$ lines of code.
- Present version (2.0) includes a GUI (QT library; + 20,000 lines of code) in addition to the (more flexible) line command software.
- Freely downloadable (CeciLL-B licence) at `http://spacem3.gforge.inria.fr/`. Works on Linux (Fedora/Red Hat and Debian/Ubuntu packages), MacOS and Windows environements.
- Data in text or binary formats: individual on rows and variables in columns (measurements); specifying the graph: Image-like grid or irregular graph (neighbour list to be given). Program I/O in XML format.
- Documentation.

Markov Random Field (MRF)

Definition

$\mathbf{Z} = (Z_1 \ldots Z_n)$ is a Markov Random Field iif:
(i) $P(Z_i|\mathbf{Z}) = P(Z_i|\mathbf{Z}_{N_i})$ and
(ii) $P(\mathbf{Z} = \mathbf{z}) > 0$.

Markov Random Field (MRF)

Definition

$\mathbf{Z} = (Z_1 \dots Z_n)$ is a Markov Random Field iif:
(i) $P(Z_i|\mathbf{Z}) = P(Z_i|\mathbf{Z}_{N_i})$ and
(ii) $P(\mathbf{Z} = \mathbf{z}) > 0$.

Consequence: (Hamersley-Clifford Theorem) $\mathbf{Z}$ has a Gibbs distribution: $\frac{\exp(-H(\mathbf{z}))}{W}$ where the energy function is decomposed on clique potentials: $H(\mathbf{z}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{z}_c)$.

Markov Random Field (MRF)

Definition

$\mathbf{Z} = (Z_1 \ldots Z_n)$ is a Markov Random Field iif:
(i) $P(Z_i|\mathbf{Z}) = P(Z_i|\mathbf{Z}_{N_i})$ and
(ii) $P(\mathbf{Z} = \mathbf{z}) > 0$.

Consequence: (Hamersley-Clifford Theorem) $\mathbf{Z}$ has a Gibbs distribution: $\frac{\exp(-H(\mathbf{z}))}{W}$ where the energy function is decomposed on clique potentials: $H(\mathbf{z}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{z}_c)$.

Potts model and extensions

Potentials on singletons (external field) & pairs (dependencies):

$$H(\mathbf{z}) = \sum_i \underbrace{V_i(z_i)}_{=_{(\text{if not dep. site } i)} - z_i' \alpha} + \sum_{j \in N_i} \underbrace{V_{ij}(z_i, z_j)}_{=_{(\text{if not dep. sites } i,j)} - z_i' \beta z_j}$$

Hidden Markov Random Fields (HMRF)

...With independent noise (seen as a generalization of mixture models):

$$\mathbf{Z} \text{ MRF} + P(X|Z) = \prod_i P(X_i|Z_i) \; (\Rightarrow (\mathbf{X}, \mathbf{Z}) \text{ MRF}).$$

Hidden Markov Random Fields (HMRF)

...With independent noise (seen as a generalization of mixture models):

$$\mathbf{Z} \text{ MRF} + P(X|Z) = \prod_i P(X_i|Z_i) \; (\Rightarrow (\mathbf{X}, \mathbf{Z}) \text{ MRF}).$$

Hence (but not equivalent to) $\mathbf{Z}|\mathbf{x}$, *a posteriori* distribution is a MRF as well with energy function: $H(\mathbf{z}, \alpha, \beta) - \sum_i \log f(x_i|\theta_{z_i})$; classical Bayesian methods for parameter estimation and clustering can be used.

Extension to pairwise and Triplet Markov fields...See slides to come.

Gaussian model for high-dimensional data

Idea from 14 models in Banfield & Raftery, 1993 (orientation, size and shape of the distribution around the mean).

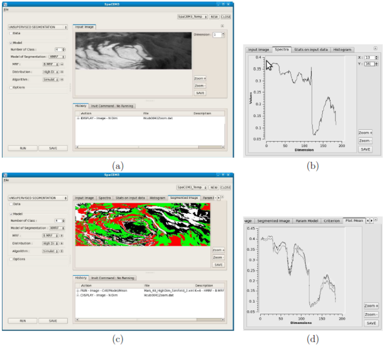
Gaussian model for high-dimensional data

Idea from 14 models in Banfield & Raftery, 1993 (orientation, size and shape of the distribution around the mean).

Models from Bouveyron et al. 2007: Spectral decomposition of the covariance matrix $\Sigma_k = Q_k \Delta_k Q_k'$:

$$\Delta_k = \left(\begin{array}{cc} \begin{array}{|ccc|} \hline a_{k1} & & 0 \\ & \ddots & \\ 0 & & a_{kD_k} \\ \hline \end{array} & (\mathbf{0}) \\ (\mathbf{0}) & \begin{array}{|ccc|} \hline b_k & & 0 \\ & \ddots & \\ 0 & & b_k \\ \hline \end{array} \end{array}\right) \begin{array}{l} \left.\right\} D_k \\ \left.\right\} (D - D_k) \end{array}$$

High-D segmentation of an image of Mars.

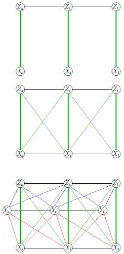


(a) Image to be clustered, (b) A pixel spectrum, (c) Segmented image and (d) average spectra for the 4 classes.

Triplet Markov model for supervised classification

The { independent/unimodal } noise hypothesis can be too restrictivre (*e.g.* modelling textures).

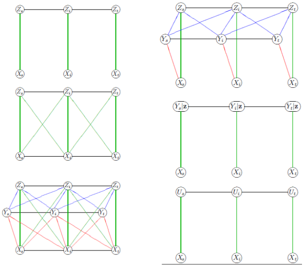
$$P_G(\mathbf{x},\mathbf{y},\mathbf{z}) \propto \exp\left(-\sum_{i\sim j} \underbrace{V_i j(y_i, z_i, y_j, z_j)}_{-y_i' \mathbb{B}_{z_i z_j} y_j - z_i' \mathbb{C} z_j} + \sum_i \log f(x_i|\theta_{y_i,z_i}) \right)$$



Triplet Markov model for supervised classification

The { independent/unimodal } noise hypothesis can be too restrictivre (*e.g.* modelling textures).

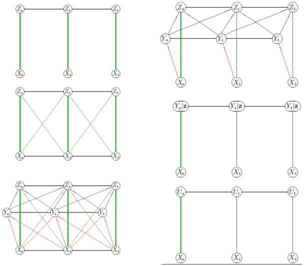
$$P_G(\mathbf{x}, \mathbf{y}, \mathbf{z}) \propto \exp\left(- \sum_{i\sim j} \underbrace{V_ij(y_i, z_i, y_j, z_j)}_{-y_i' \mathbb{B}_{z_i z_j} y_j - z_i' \mathbb{C} z_j} + \sum_i \log f(x_i | \theta_{y_i, z_i}) \right)$$

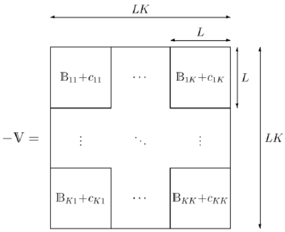


Triplet Markov model for supervised classification

The { independent/unimodal } noise hypothesis can be too restrictivre (*e.g.* modelling textures).

$$P_G(\mathbf{x},\mathbf{y},\mathbf{z}) \propto \exp\left(-\sum_{i\sim j}\underbrace{V_ij(y_i,z_i,y_j,z_j)}_{-y_i'\mathbb{B}_{z_iz_j}y_j-z_i'\mathbb{C}z_j} + \sum_i \log f(x_i|\theta_{y_i,z_i})\right)$$



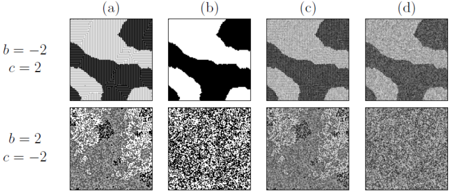


Triplet Markov model simulation

Learning: $(X, Y|Z) \rightarrow \theta_{lk}$ and $B_{kk'}$ estimated
Test: $(X, (Y, Z)) \rightarrow C$ to be estimated (θ and B fixed) and then clustering.

Triplet Markov model simulation

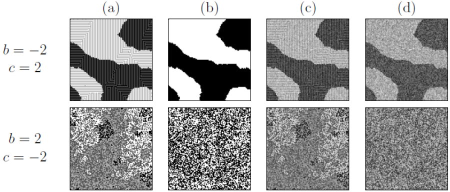
Learning: $(X, Y|Z) \rightarrow \theta_{lk}$ and $B_{kk'}$ estimated
Test: $(X, (Y, Z)) \rightarrow C$ to be estimated (θ and B fixed) and then clustering.



Simulations with L=K=2; each of the 4 different (y_i, z_i)'s is associated to a different grey level.
(a) $(\mathbf{Y}, \mathbf{Z})$ realization, (b) $\mathbf{Z}$) realization, (c) $\mathbf{X}$) realization and (d) realization of an HMRF adding on independent noise $\mathcal{N}(0, 0.3)$.

Triplet Markov model simulation

Learning: $(X, Y|Z) \rightarrow \theta_{lk}$ and $B_{kk'}$ estimated
Test: $(X, (Y, Z)) \rightarrow C$ to be estimated (θ and B fixed) and then clustering.



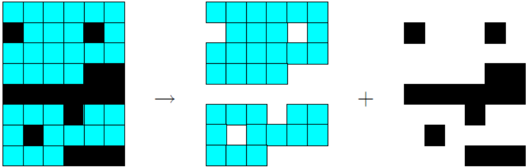
Simulations with L=K=2; each of the 4 different (y_i, z_i)'s is associated to a different grey level.
(a) $(\mathbf{Y}, \mathbf{Z})$ realization, (b) $\mathbf{Z}$) realization, (c) $\mathbf{X}$) realization and (d) realization of an HMRF adding on independent noise $\mathcal{N}(0, 0.3)$.

Drawback: supervised framework needed (identifiability).

Classical algorithms

Iterated Conditional Modes, k-means, EM (Dempster et al.; J. Roy. Statist. Soc. Ser. B 1977) and extensions (Clustering EM, Neighbour EM and NCEM).

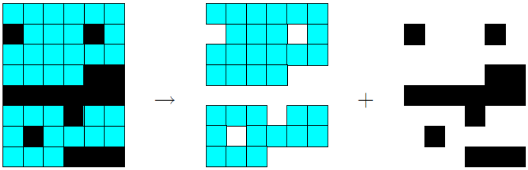
EM with spatial dependencies and missing observations ?



$\mathbf{x} = (\mathbf{x^o}, \mathbf{x^m})$ $\quad$ $\mathbf{x^o} = \{x_i^{o_i}\}$ $\quad$ $\mathbf{x^m} = \{x_i^{m_i}\}$

MAR hypothesis ($P(\mathbf{m}|\mathbf{x}, \mathbf{z}) = P(\mathbf{m}|\mathbf{x}^o)$).

EM with spatial dependencies and missing observations ?



$\mathbf{x} = (\mathbf{x^o}, \mathbf{x^m})$ → $\mathbf{x^o} = \{x_i^{o_i}\}$ + $\mathbf{x^m} = \{x_i^{m_i}\}$

MAR hypothesis ($P(\mathbf{m}|\mathbf{x},\mathbf{z}) = P(\mathbf{m}|\mathbf{x}^o)$).

EM aims at maximizing he completed likelihood:

$$\psi^{(q+1)} = \arg\max \mathbb{E}\left[\log P(\mathbf{x}^o, \mathbf{X}^m, \mathbf{Z}|\psi)|\mathbf{x}^o, \psi^{(q)}\right]$$

...Intractable when **Z** MRF but ok when factorized distribution → Celeux et al., 2003.

Neighbour Recovery EM (NREM) with missing observations

$$P_G(\mathbf{Z}) \approx \prod_i Q_i(Z_i)$$

Neighbour Recovery EM (NREM) with missing observations

$$P_G(\mathbf{Z}) \approx \prod_i P(Z_i | \tilde{Z}_{N_i})$$

(MF-like approximation)

Neighbour Recovery EM (NREM) with missing observations

$$P_G(\mathbf{Z}) \approx \prod_i P(Z_i|\tilde{Z}_{N_i})$$

(MF-like approximation)

Iteratif EM-like procedure:

- **NR** Fix a $\tilde{z}$ configuration from x^o and $\psi^{(q)}$. In particular $\tilde{z}$ can be simulated according to $P(\mathbf{Z}|\mathbf{x}^o, \psi^{(q)})$ (Gibbs sampler): Simuilated Field (SF) algorithm.
- **EM** Apply EM on factorized model to update $\psi^{(q+1)}$.
  - Then MAP (or MPM) to reconstruct $\mathbf{z}$. But also $\mathbf{x}^m$.

Neighbour Recovery EM (NREM) with missing observations

Iteratif EM-like procedure:

NR Fix a $\tilde{z}$ configuration from x^o and $\psi^{(q)}$. In particular $\tilde{z}$ can be simulated according to $P(\mathbf{Z}|\mathbf{x}^o, \psi^{(q)})$ (Gibbs sampler): Simuilated Field (SF) algorithm.

EM Apply EM on factorized model to update $\psi^{(q+1)}$.

- Then MAP (or MPM) to reconstruct $\mathbf{z}$. But also $\mathbf{x}^m$.

L'étape (EM)

(E) Calcul des probabilités *a posteriori* $\tilde{t}_{ik}^{(q)} = P(Z_i = k|\tilde{\mathbf{z}}_{N_i}, x_i^{o_i}, \psi^{(q)})$:

$$\tilde{t}_{ik}^{(q)} = \begin{cases} \dfrac{\tilde{\pi}_{ik}^{(q)} f(x_i^{o_i}|\theta_k^{(q)})}{\sum_{k' \in \mathcal{K}} \tilde{\pi}_{ik'}^{(q)} f(x_i^{o_i}|\theta_{k'}^{(q)})} & \text{si } o_i \neq \varnothing \\ \tilde{\pi}_{ik}^{(q)} & \text{si } o_i = \varnothing \end{cases}$$

où $\tilde{\pi}_{ik}^{(q)} = P(Z_i = k|\tilde{\mathbf{z}}_{N_i}, \phi^{(q)})$

(M) Mise à jour des paramètres ϕ de $P_G(\mathbf{z}|\phi)$ et $\boldsymbol{\theta} = (\theta_k)_{k \in \mathcal{K}}$ des densités $f(.|\theta_k)$:

$$\phi^{(q+1)} = \underset{\phi}{\text{argmax}} \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}} \tilde{t}_{ik}^{(q)} \log \tilde{\pi}_{ik} \quad (1)$$

$$\theta_k^{(q+1)} = \underset{\theta_k}{\text{argmax}} \sum_{i \in \mathcal{I}} \tilde{t}_{ik}^{(q)} \mathbb{E}(\log f(x_i^{o_i}, X_i^{m_i}|\theta_k)|x_i^{o_i}, \theta_k^{(q)}) \quad (2)$$

L'équation (1) est identique au cas complet (descente de gradient)
L'équation (2) est identique au cas du mélange indépendant [Litt86]

Neighbour Recovery EM (NREM) with missing observations

Iteratif EM-like procedure:

NR Fix a $\tilde{z}$ configuration from x^o and $\psi^{(q)}$. In particular $\tilde{z}$ can be simulated according to $P(\mathbf{Z}|\mathbf{x}^o, \psi^{(q)})$ (Gibbs sampler): Simuilated Field (SF) algorithm.

EM Apply EM on factorized model to update $\psi^{(q+1)}$.

- Then MAP (or MPM) to reconstruct $\mathbf{z}$. But also $\mathbf{x}^m$.

Cas gaussien - moyenne

On s'intéresse au cas où $f(.|\theta_k)$ est gaussienne, avec $\theta_k = (\mu_k, \Sigma_k)$.

$$\begin{aligned} f(x_i^{o_i}|\theta_k) &= \mathcal{N}(x_i^{o_i}|\mu_k^{o_i}, \Sigma_k^{o_i o_i}) \\ f(x_i^{m_i}|x_i^{o_i}, \theta_k) &= \mathcal{N}(x_i^{m_i}|\eta_{ik}, \Gamma_{ik}) \\ \text{où } \eta_{ik} &= \mu_k^{m_i} + \Sigma_k^{m_i o_i}(\Sigma_k^{o_i o_i})^{-1}(x_i^{o_i} - \mu_k^{o_i}) \\ \text{et } \Gamma_{ik} &= \Sigma_k^{m_i m_i} - \Sigma_k^{m_i o_i}(\Sigma_k^{o_i o_i})^{-1}\Sigma_k^{o_i m_i} \end{aligned}$$

Mise à jour de la composante $d \in [\![1, D]\!]$ de la moyenne μ_k :

$$\mu_k^d = \frac{\sum_i \tilde{t}_{ik} l_{ik}^d}{\sum_i \tilde{t}_{ik}} \quad \text{avec} \quad l_{ik}^d = \begin{cases} x_i^d & \text{si } d \in o_i \\ \eta_{ik}^d & \text{sinon} \end{cases}$$

→ consiste à remplacer les valeurs manquantes $x_i^{m_i}$ par leur moyenne η_{ik} conditionnellement aux observations $x_i^{o_i}$.

Neighbour Recovery EM (NREM) with missing observations

Iteratif EM-like procedure:

NR Fix a $\tilde{z}$ configuration from x^o and $\psi^{(q)}$. In particular $\tilde{z}$ can be simulated according to $P(\mathbf{Z}|\mathbf{x}^o, \psi^{(q)})$ (Gibbs sampler): Simuilated Field (SF) algorithm.

EM Apply EM on factorized model to update $\psi^{(q+1)}$.

- Then MAP (or MPM) to reconstruct $\mathbf{z}$. But also $\mathbf{x}^m$.

Cas gaussien - covariance

On s'intéresse au cas où $f(.|\theta_k)$ est gaussienne, avec $\theta_k = (\mu_k, \Sigma_k)$.

$$\begin{aligned} f(x_i^{o_i}|\theta_k) &= \mathcal{N}(x_i^{o_i}|\mu_k^{o_i}, \Sigma_k^{o_i o_i}) \\ f(x_i^{m_i}|x_i^{o_i}, \theta_k) &= \mathcal{N}(x_i^{m_i}|\eta_{ik}, \Gamma_{ik}) \\ \text{où } \eta_{ik} &= \mu_k^{m_i} + \Sigma_k^{m_i o_i}(\Sigma_k^{o_i o_i})^{-1}(x_i^{o_i} - \mu_k^{o_i}) \\ \text{et } \Gamma_{ik} &= \Sigma_k^{m_i m_i} - \Sigma_k^{m_i o_i}(\Sigma_k^{o_i o_i})^{-1}\Sigma_k^{o_i m_i} \end{aligned}$$

Mise à jour de la composante $d, d' \in [\![1, D]\!]$ de covariance Σ_k :

$$\Sigma_k^{dd'} = \frac{\sum_i \tilde{t}_{ik} S_{ik}^{dd'}}{\sum_i \tilde{t}_{ik}} \text{ avec}$$

$$(S_{ik}^{dd'}) = \begin{cases} (x_i^d - \mu_k^d)(x_i^{d'} - \mu_k^{d'}) & \text{si } d \in o_i \text{ et } d' \in o_i \\ (x_i^d - \mu_k^d)(\eta_{ik}^{d'} - \mu_k^{d'}) & \text{si } d \in o_i \text{ et } d' \in m_i \\ (\eta_{ik}^d - \mu_k^d)(x_i^{d'} - \mu_k^{d'}) & \text{si } d \in m_i \text{ et } d' \in o_i \\ (\eta_{ik}^d - \mu_k^d)(\eta_{ik}^{d'} - \mu_k^{d'}) + \Gamma_{ik}^{dd'} & \text{si } d \in m_i \text{ et } d' \in m_i \end{cases}$$

$\rightarrow$ n'est pas équivalent à remplacer les valeurs manquantes $x_i^{m_i}$ par leur moyenne η_{ik} conditionnellement aux observations $x_i^{o_i}$.

INRIA Mistis · INRA BIA Toulouse

Algorithms in practice

- Initialization: random , k-means or fixed by the user.

Algorithms in practice

- Initialization: random , k-means or fixed by the user.

- Stopping criterion:
  - Limit on the maximum difference between values for the completed likelihood (estimate).
  - Limit on the maximum difference of conditional probabilities for an individual.
  - Limit on the proportion of individual that are assigned different classes.
  - Number of iterations.

Model selection: selection criteria

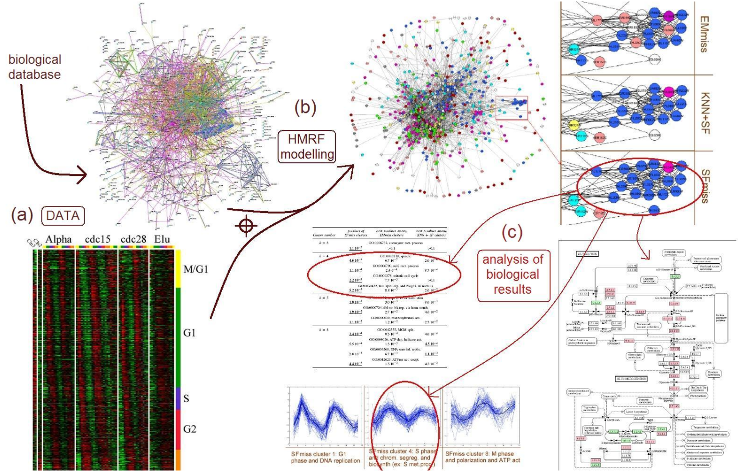
The "best" model should be a compromise between a fit to the data (adequacy to what is observed) and allowed complexity (!!Overfitting!!). Among the many existing criteria we used the Bayes Information Criterion (BIC, Schwarz, Ann. Stat. 1978) and the Integrated Completed Likelihood (ICL, Biernacki et al., IEEE PAMI 2000, designed for classification purpose).

Model selection: selection criteria

The "best" model should be a compromise between a fit to the data (adequacy to what is observed) and allowed complexity (!!Overfitting!!). Among the many existing criteria we used the Bayes Information Criterion (BIC, Schwarz, Ann. Stat. 1978) and the Integrated Completed Likelihood (ICL, Biernacki et al., IEEE PAMI 2000, designed for classification purpose).

Approximations are needed when the model is Markovian. 2 approximations in Forbes and Peyrard, 2003: BIC^p that approximates P_G with a mean field approach while BIC^w approximates the partition function W. Proves $BIC^p \leq BIC^w \leq BIC^{true}$ in theory. Verified empirically.

Summary of the data analysis workflow



(Blanchet and Vignes, 2009)

Summary and perspectives

Wrap up

SpaCEM^3 is wonderful ;). Did I tell you *Spatial Clustering with EM Markov Models* ??

Summary and perspectives

Wrap up

SpaCEM^3 is wonderful ;). Did I tell you *Spatial Clustering with EM Markov Models* ??

Prospects (or my Xmas wish list)

- Promote the use of the software (*e.g.* on varied molecular biology datasets): Present collaborations at the INRA in Toulouse and Application Note in *Bioinformatics* to be submitted soon.
- Graph not totally fixed ? incomplete ? Treat edges as missing in a similar manner to observations (theoretical work needed).
- Include different distribution: multinomial useful for ecological data (theoretical work needed).
- Triplet models for unsupervised clustering (theoretical work needed).

Some references

Jeffrey D. Banfield and Adrian E. Raftery
Model-based Gaussian and non-Gaussian clustering.
Biometrics, 49(3):803-821, 1993.

Gilles Celeux, Florence Forbes and Nathalie Peyrard.
EM procedures using mean field-like approximations for Markov model-based image segmentation.
Pat. Rec., 36(1):131-144, 2003.

Florence Forbes and Nathalie Peyrard.
Hidden Markov random field model selection criteria based on mean field-like approximations.
IEEE Trans. PAMI, 25(8): 1089-1101, 2003.

Charles Bouveyron, Stéphane Girard and Cordelai Schmidt.
High dimensional data clustering.
Comput. Statist. Data Analysis, 52(1):502-519, 2007.

Juliette Blanchet and Florence Forbes.
Triplet Markov fields for the supervised classification of complex structure data.
IEEE Trans. PAMI, 30(6):1055-1067, 2008.

Matthieu Vignes and Florence Forbes.
Gene clustering via integrated Markov models combining individual and pairwise features.
IEEE/ACM Trans. Comput. Biol. Bioinform., 6(2):260-270, 2009.

Juliette Blanchet and Matthieu Vignes.
A model-based approach to gene clustering with missing observations reconstruction in a Markov random field framework.
J. Comput. Biol., 16(3), 475-486 (2009).

Merci à tous les collègues

Nathalie Peyrard, Lemine Abdallah, Sophie Choppart, Lamia Azizi, Juliette Blanchet. . .

Merci à tous les collègues

Nathalie Peyrard, Lemine Abdallah, Sophie Choppart, Lamia Azizi, Juliette Blanchet. . .

et vous

pour votre attention.

Questions, critiques, remarques. . . bienvenues ?!