A MAP-MRF Approach for Multispectral Image Contextual Classification using Combination of Suboptimal Iterative Algorithms

Alexandre L. M. Levada
Nelson D. A. Mascarenhas
Alberto Tannús

Physics Institute of São Carlos (IFSC) – University of São Paulo (USP)
Computer Department (DC) – Federal University of São Carlos (UFSCar)

Overview

- Introduction
- MAP-MRF Contextual Classification
- MRF Parameter Estimation
- Statistical Inference on Markov Random Fields
- Metrics for Performance Evaluation
- Experiments and Results
- Conclusions

Introduction

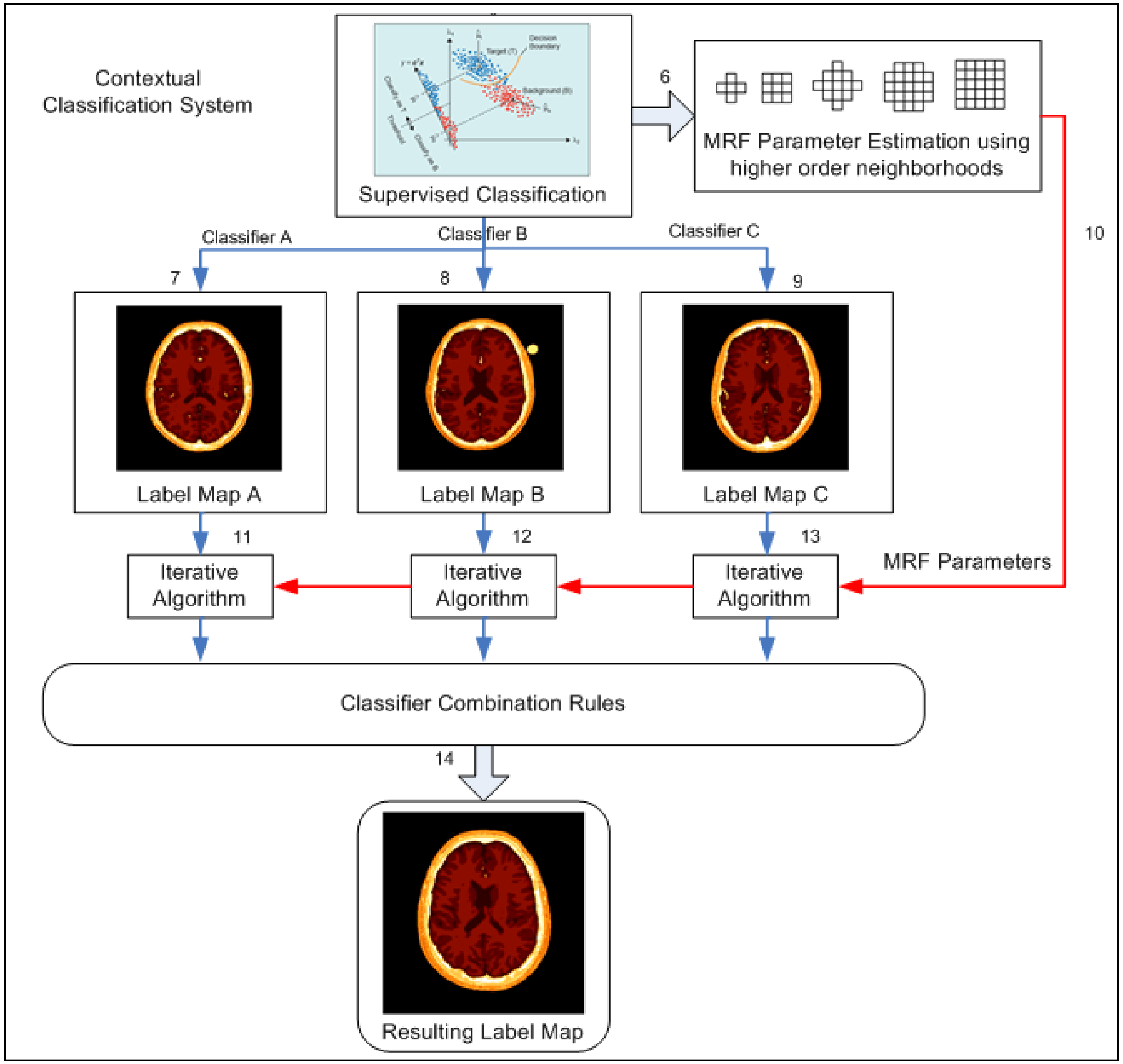
- Simulated Annealing
  - “Best” algorithm for MAP-MRF estimation
  - Obtained solutions are close to the the optimum one
- However, it has some drawbacks
  - High computational cost
  - Extremelly slow convergence
- Unfeasible for several real applications
- Tipical alternatives
  - Suboptimal combinatorial optimization algorithms
    - Iterated Conditional Modes (ICM), Graduated Non-Convexity (GNC), Highest Confidenve First (HCF), etc

Motivation

- Limitations of suboptimal algorithms
  - Strong dependence on initial conditions
  - Convergence to local maxima solutions

- Our solution
  - Decision rule that combines two MRF's
    - Multispectral Gaussian Markov Random Field (GMRF) (observations)
    - Potts MRF model (smooth prior)
  - Combination of multiple initial conditions
  - Maximum pseudo-likelihood estimation (for the MRF parameters)

Proposed MAP-MRF Framework



MAIN GOAL

Incorporate more information in the decision rule by using:

- Multiple initializations
- Higher-order neighborhoods

Here, we adopt second order systems and 7 different initial conditions

MAP-MRF Contextual Classification

- Combination of two MRF models
  - Multispectral Gaussian MRF (Likelihood)
  - Potts MRF model (Prior knowledge)

$$x_{ij}^{(p+1)} = \arg\left[\max_{m} Q\left(x_{ij} = m | x_{S/(i,j)}^{(p)}, \mathrm{y}, \hat{\Phi}, \hat{\Theta}, \hat{\beta}\right)\right]$$

$$Q\left(x_{ij} = m | x_{S/(i,j)}^{(p)}, \mathrm{y}, \hat{\Phi}, \hat{\Theta}, \hat{\beta}\right) = -\frac{1}{2}\ln\left|\hat{\Sigma}_m\right| - \frac{1}{2}\left[\vec{y}_{ij} - \hat{\vec{\mu}}_m - \left(\hat{\theta}^T y_{n_{ij}} - 2\left(\sum_{ct}\hat{\theta}^{ct}\right)\hat{\vec{\mu}}_m\right)\right]^T$$
$$\times\hat{\Sigma}_m^{-1}\left[\vec{y}_{ij} - \hat{\vec{\mu}}_m - \left(\hat{\theta}^T y_{n_{ij}} - 2\left(\sum_{ct}\hat{\theta}^{ct}\right)\hat{\vec{\mu}}_m\right)\right] + \hat{\beta}U_{ij}(m)$$

Spectral Model

* *"Regularization" Parameter*

Spatial Model

* Controls the tradeoff between data fidelity and prior knowledge

Maximum Pseudo-Likelihood Estimation

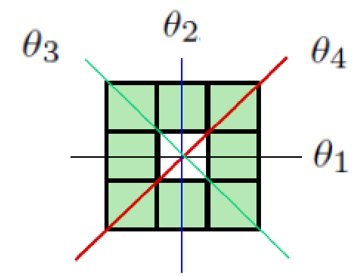
- For MRF models, maximum likelihood is not feasible
  - The joint Gibbs distribution is intractable
    - A solution is to use maximum pseudo-likelihood (MPL)

- Our motivation
  - MPL estimatores have good statistical properties
    - Consistency (asymptotic unbiased)
    - Asymptotic normal

  - It is possible to completely characterize its behavior in the limiting case (when $N \rightarrow \infty$)

MPL Estimation on GMRF Model

- Assuming that the model parameters are uncorrelated
  - The estimation of the Gaussian MRF model parameters is straightforward
    - Covariance matrix is diagonal
    - Estimation is performed in each band separately



$$\vec{\theta}^T = [\theta_1, \theta_2, \theta_3, \theta_4]$$

  - Pseudo-likelihood equation

$$\log PL(\theta, \mu, \sigma^2) = \sum_{(i,j)\in W} \left\{ -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left[y_{ij} - \theta\psi_{ij} - \mu(1 - 2\theta\mathbf{u})\right]^2 \right\}$$

where $\psi_{ij} = [(y_{i+1,j} + y_{i-1,j}), (y_{i,j+1} + y_{i,j-1}), (y_{i+1,j-1} + y_{i-1,j+1}), (y_{i-1,j-1} + y_{i+1,j+1})]^T$

  - Closed-form solution

$$\hat{\theta} = \left\{ \left[\sum_{(i,j)\in W} (y_{ij} - \hat{\mu})\tilde{\psi}_{ij}^T \right] \left[\sum_{(i,j)\in W} \tilde{\psi}_{ij}\tilde{\psi}_{ij}^T \right]^{-1} \right\}$$

where

$\hat{\mu}$: Sample mean

$\tilde{\psi}_{ij} = \psi_{ij} - \frac{1}{N}\sum_{(k,l)\in W}\psi_{ij}$

MPL Estimation on Potts Model

- The Potts model is suitable for discrete random variables
  - Local Conditional Density Function (LCDF)

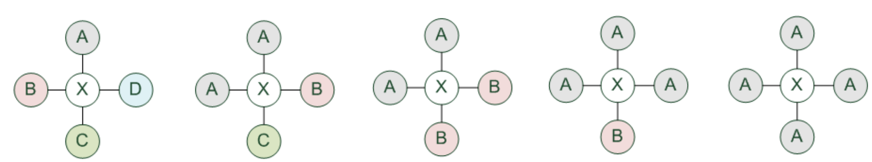
$$p(x_{ij} = m_{ij} | \eta^s_{ij}) = \frac{\exp\{\beta U_{ij}(m_{ij})\}}{\sum_{\ell=1}^{M} \exp\{\beta U_{ij}(\ell)\}}$$

probability of a given value is proportional to the number of occurences in the neighborhood

  - Pseudo-likelihood equation

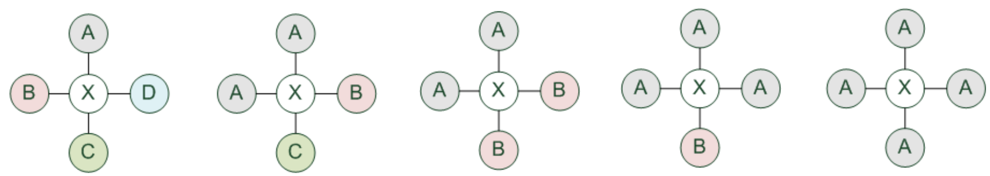
$$\sum_{(i,j)\in W} U_{ij}(m_{ij}) - \sum_{(i,j)\in W} \left[\frac{\sum_{\ell=1}^{M} U_{ij}(\ell) \exp\{\beta U_{ij}(\ell)\}}{\sum_{\ell=1}^{M} \exp\{\beta U_{ij}(\ell)\}} \right] = 0$$

    - <u>Methodology:</u> Identify the spatial configuration patters that offer distinct contributions to the pseudo-likelihood equation, defining a <u>*dictionary of patterns*</u>



MPL Estimation on Potts Model

- First order systems
  - Only five possible configuration patterns
  - From zero-agreement to total-agreement



$\vec{v}_0 = [1,1,1,1]$ $\vec{v}_1 = [2,1,1,0]$ $\vec{v}_2 = [2,2,0,0]$ $\vec{v}_3 = [3,1,0,0]$ $\vec{v}_4 = [4,0,0,0]$

First order

5 patterns

- Second order systems

[1, 1, 1, 1, 1, 1, 1, 1]	[2, 1, 1, 1, 1, 1, 1, 0]	[3, 1, 1, 1, 1, 1, 0, 0]	[2, 2, 1, 1, 1, 1, 0, 0]
[4, 1, 1, 1, 1, 0, 0, 0]	[3, 2, 1, 1, 1, 0, 0, 0]	[2, 2, 2, 1, 1, 0, 0, 0]	[5, 1, 1, 1, 0, 0, 0, 0]
[4, 2, 1, 1, 0, 0, 0, 0]	[3, 3, 1, 1, 0, 0, 0, 0]	[3, 2, 2, 1, 0, 0, 0, 0]	[2, 2, 2, 2, 0, 0, 0, 0]
[6, 1, 1, 0, 0, 0, 0, 0]	[5, 2, 1, 0, 0, 0, 0, 0]	[4, 3, 1, 0, 0, 0, 0, 0]	[4, 2, 2, 0, 0, 0, 0, 0]
[3, 3, 2, 0, 0, 0, 0, 0]	[4, 4, 0, 0, 0, 0, 0, 0]	[5, 3, 0, 0, 0, 0, 0, 0]	[6, 2, 0, 0, 0, 0, 0, 0]
	[7, 1, 0, 0, 0, 0, 0, 0]	[8, 0, 0, 0, 0, 0, 0, 0]	

Second order

22 patterns

MPL Estimation on Potts Model

- Pseudo-likelihood (PL) equation (2ª order)

$$
\begin{aligned}
\frac{\partial}{\partial\beta}\log PL(\beta) = & \sum_{s\in\Omega} U_s\left(m_s\right) - \frac{8e^{8\hat{\beta}}}{e^{8\hat{\beta}}+M-1}K_1 - \frac{7e^{7\hat{\beta}}+e^{\hat{\beta}}}{e^{7\hat{\beta}}+e^{\hat{\beta}}+M-2}K_2 - \frac{6e^{6\hat{\beta}}+2e^{2\hat{\beta}}}{e^{6\hat{\beta}}+e^{2\hat{\beta}}+M-2}K_3 \\
& - \frac{6e^{6\hat{\beta}}+2e^{\hat{\beta}}}{e^{6\hat{\beta}}+2e^{\hat{\beta}}+M-3}K_4 - \frac{5e^{5\hat{\beta}}+3e^{3\hat{\beta}}}{e^{5\hat{\beta}}+e^{3\hat{\beta}}+M-2}K_5 - \frac{5e^{5\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{5\hat{\beta}}+e^{2\hat{\beta}}+e^{\hat{\beta}}+M-3}K_6 \\
& - \frac{5e^{5\hat{\beta}}+3e^{\hat{\beta}}}{e^{5\hat{\beta}}+3e^{\hat{\beta}}+M-4}K_7 - \frac{8e^{4\hat{\beta}}}{2e^{4\hat{\beta}}+M-2}K_8 - \frac{4e^{4\hat{\beta}}+3e^{3\hat{\beta}}+e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{3\hat{\beta}}+e^{\hat{\beta}}+M-3}K_9 \\
& - \frac{4e^{4\hat{\beta}}+4e^{2\hat{\beta}}}{e^{4\hat{\beta}}+2e^{2\hat{\beta}}+M-3}K_{10} - \frac{4e^{4\hat{\beta}}+2e^{2\hat{\beta}}+2e^{\hat{\beta}}}{e^{4\hat{\beta}}+e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{11} \\
& - \frac{4e^{4\hat{\beta}}+4e^{\hat{\beta}}}{e^{4\hat{\beta}}+4e^{\hat{\beta}}+M-5}K_{12} - \frac{6e^{3\hat{\beta}}+2e^{2\hat{\beta}}}{2e^{3\hat{\beta}}+e^{2\hat{\beta}}+M-3}K_{13} - \frac{6e^{3\hat{\beta}}+2e^{\hat{\beta}}}{2e^{3\hat{\beta}}+2e^{\hat{\beta}}+M-4}K_{14} \\
& - \frac{3e^{3\hat{\beta}}+4e^{2\hat{\beta}}+e^{\hat{\beta}}}{e^{3\hat{\beta}}+2e^{2\hat{\beta}}+e^{\hat{\beta}}+M-4}K_{15} - \frac{3e^{3\hat{\beta}}+2e^{2\hat{\beta}}+3e^{\hat{\beta}}}{e^{3\hat{\beta}}+e^{2\hat{\beta}}+3e^{\hat{\beta}}+M-5}K_{16} \\
& - \frac{3e^{3\hat{\beta}}+5e^{\hat{\beta}}}{e^{3\hat{\beta}}+5e^{\hat{\beta}}+M-6}K_{17} - \frac{8e^{2\hat{\beta}}}{4e^{2\hat{\beta}}+M-4}K_{18} - \frac{6e^{2\hat{\beta}}+2e^{\hat{\beta}}}{3e^{2\hat{\beta}}+2e^{\hat{\beta}}+M-5}K_{19} \\
& - \frac{4e^{2\hat{\beta}}+4e^{\hat{\beta}}}{2e^{2\hat{\beta}}+4e^{\hat{\beta}}+M-6}K_{20} - \frac{2e^{2\hat{\beta}}+6e^{\hat{\beta}}}{e^{2\hat{\beta}}+6e^{\hat{\beta}}+M-7}K_{21} - \frac{8e^{\hat{\beta}}}{8e^{\hat{\beta}}+M-8}K_{22} = 0
\end{aligned}
$$

Important Remarks about the PL Equation

- The number of terms in the equation is always equal to the number of possible configuration patterns
  - Each term = product of 2 factors
    - A fraction: contribution of each pattern to the equation
    - K_i : number of occurences of the *i-th* pattern along the field
      - Histogram of contextual patterns

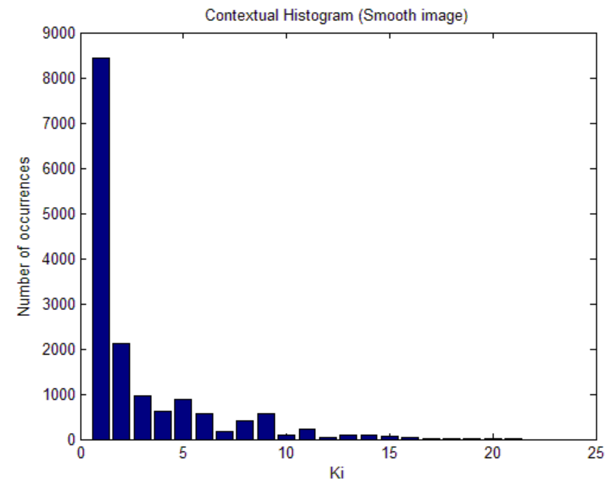
- Transcendental equation
  - Does not have closed-form solution

- It is valid for na arbitrary number of states/classes (M)
  - When M is small, the proposed PL equation is further simplified

Example of a Practical Application

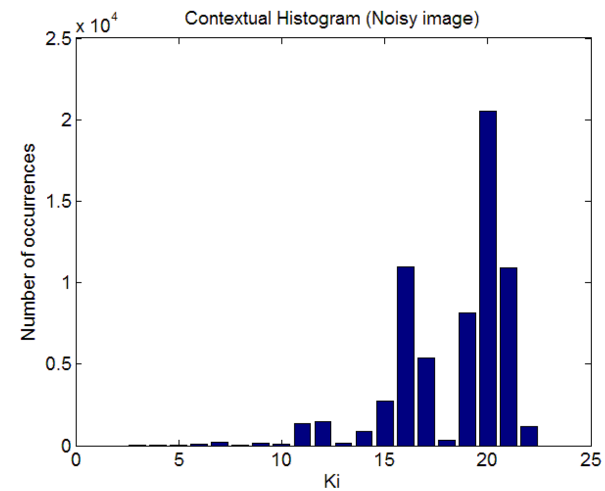
- Smooth image

83	82	80	81	78	80	80	74	73	80	77	89	95	104	107	114	115	116
82	78	81	81	78	82	74	79	72	81	78	89	94	106	112	109	115	118
80	77	76	78	75	76	71	69	73	77	80	82	92	102	108	114	118	117
80	79	80	78	76	76	75	68	68	74	78	81	95	100	112	114	116	116
79	77	79	75	74	75	70	69	64	74	75	85	93	104	107	112	118	118
81	77	79	77	79	70	75	70	69	69	75	83	95	102	108	115	119	121
76	79	74	75	75	74	71	69	65	70	78	86	94	105	109	117	119	120
78	76	79	78	72	73	74	71	69	70	78	80	95	105	112	117	123	123
77	77	76	75	74	73	74	72	65	70	74	81	89	105	109	119	122	121
76	73	71	74	77	79	73	69	65	69	73	83	94	105	114	117	125	122
74	74	74	71	69	74	67	67	65	66	75	78	94	110	118	122	124	123
76	72	68	71	71	73	70	70	65	69	75	83	97	109	114	123	125	127



- Noisy image

52	104	59	82	59	41	88	68	61	120	84	83	98	121	133	117	135	92
79	134	130	113	95	49	91	118	48	31	17	151	73	73	77	73	123	120
145	80	80	82	29	95	58	33	102	91	81	95	82	100	86	89	159	162
70	67	74	127	65	84	123	56	22	83	109	114	103	120	101	69	167	115
92	71	47	118	79	90	57	73	106	56	106	79	108	85	31	168	120	123
123	106	65	74	0	91	69	79	94	81	52	100	32	123	64	119	86	112
104	100	60	93	96	85	87	96	57	48	108	110	131	92	81	151	105	113
87	82	63	68	74	67	90	87	52	108	85	170	75	149	113	168	141	99
79	44	96	61	100	57	118	80	90	99	45	85	67	90	94	89	112	126
137	15	88	63	85	56	87	65	50	40	64	101	101	80	87	132	121	124
137	89	92	112	67	99	79	103	99	85	76	110	127	99	98	81	99	149
58	77	96	120	80	34	102	60	55	88	85	76	133	57	118	145	94	121



Asymptotic Evaluations on MRF's

- Little is known about the accuracy of MPL estimation
  - Approximation for the asymptotic variance of MPL estimators
    - Using the observed Fisher Information
  - Asymptotic covariance matrix of MPL estimators

$$C(\theta) = H^{-1}(\theta) J(\theta) H^{-1}(\theta)$$

$$H(\theta) = E[\nabla^2 logPL(\theta)]$$
$$J(\theta) = Var[\nabla logPL(\theta)]$$

  - After some manipulations, the asymptotic variances for the GMRF model are given by:

$$c_{kk}(\theta) = \frac{\widehat{I}^1_{obs}(\theta)}{\left[\widehat{I}^2_{obs}(\theta)\right]^2}, \quad k = 1, \ldots, 4$$

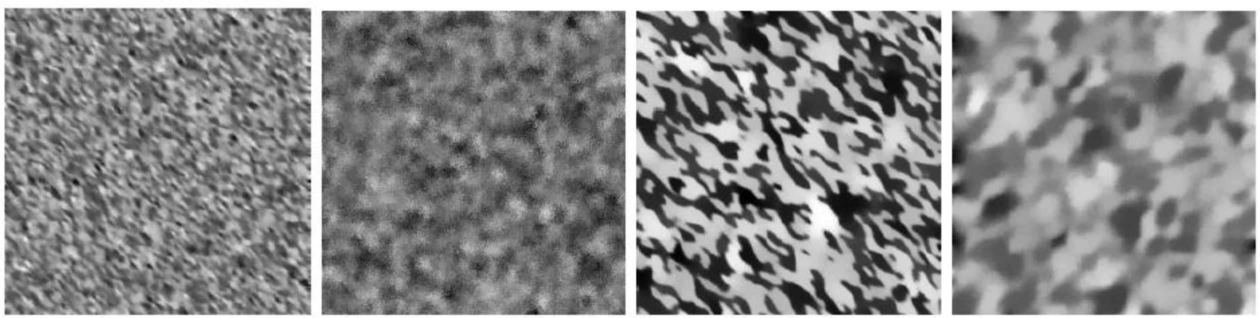
$$\widehat{I}^1_{obs}(\theta) = \frac{1}{N\sigma^4} \sum_{i=1}^{N} \left\{ \left[y_{ij} - \theta\psi_{ij} - \mu(1 - 2\theta\mathbf{u}) \right] \left[\psi^k_{ij} - 2\mu \right] \right\}^2$$

$$\widehat{I}^2_{obs}(\theta) = -\frac{1}{N\sigma^2} \sum_{i=1}^{N} \left[\psi^k_{ij} - 2\mu \right]^2$$

Observed Fisher Information calculated by first and second derivatives

Asymptotic Evaluations on MRF's

- Experiments with synthetic images generated by MCMC simulation algorithms
  - Interval estimation



Non-isotropic Isotropic

K	θ_k	$\hat{\theta}_k$	$\hat{\sigma}_k$	90% IC
1	0.25	0.2217	0.0390	[0.1799 0.3077]
2	0.3	0.2758	0.0387	[0.2398 0.3667]
3	−0.1	−0.1145	0.0394	[−0.1771 −0.0479]
4	0.2	0.1743	0.0386	[0.1150 0.2416]

Non-isotropic

K	θ_k	$\hat{\theta}_k$	$\hat{\sigma}_k$	90% IC
1	0.2	0.1908	0.0506	[0.1079 0.2738]
2	0.15	0.1605	0.0524	[0.0746 0.2464]
3	0.07	0.0716	0.0482	[−0.0074 0.1506]
4	0.05	0.0523	0.0418	[−0.0146 0.1192]

Isotropic

In all cases, the estimated intervals contain the real parameter values

Asymptotic Evaluations on MRF's

- The asymptotic variance for the Potts model is given by

$$c(\beta) = \frac{\widehat{I}^1_{obs}(\beta)}{\left[\widehat{I}^2_{obs}(\beta)\right]^2}$$

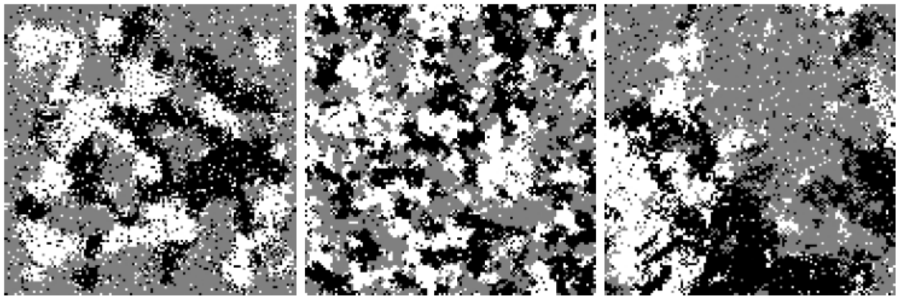
Fisher Information calculated using the first derivate ←

$$\widehat{I}^1_{obs}(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\sum_{\ell=1}^{M}\left[\sum_{k=1}^{M}(U_i(m_i) - U_i(\ell))(U_i(m_i) - U_i(k))e^{\beta(U_i(\ell)+U_i(k))}\right]}{\left[\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}\right]^2}\right\}$$

Fisher Information calculated using the second derivate ←

$$\widehat{I}^2_{obs}(\beta) = \frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\sum_{\ell=1}^{M-1}\left[\sum_{k=\ell+1}^{M}(U_i(\ell) - U_i(k))^2 e^{\beta(U_i(\ell)+U_i(k))}\right]}{\left[\sum_{\ell=1}^{M} e^{\beta U_i(\ell)}\right]^2}\right\}$$

Asymptotic Evaluations on MRF's



In order to check the accuracy of the proposed MPL estimation method, we tested the following hypothesis:

- H: the proposed pseudo-likelihood equations provide results that are statistically equivalent to the real parameter values, that is:

$$H : \beta = \hat{\beta}_{MPL}.$$

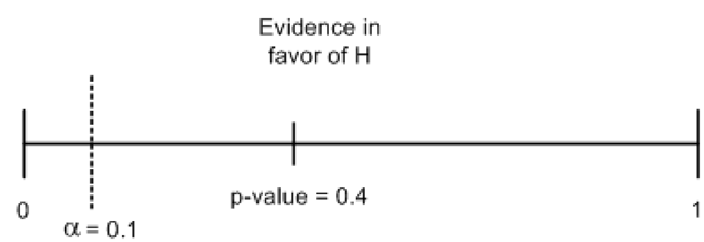
Test statistic: $Z = \frac{\hat{\beta}_{MPL} - \mu(\hat{\beta}_{MPL})}{\sqrt{Var(\hat{\beta}_{MPL})}} = \frac{\hat{\beta}_{MPL} - \beta}{\sqrt{Var_n(\hat{\beta}_{MPL})}} \approx N(0,1)$

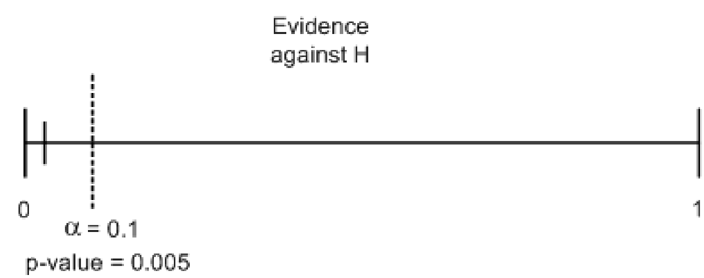
The test statisitc, together with the the p-values, indicated that the derived PL equation provide estimates that are statistically equivalent to the real parameters

MPL estimators, observed Fisher information, asymptotic variances, test statistics and *p*-values for synthetic MCMC images using second-order systems.

	Swendsen–Wang		Gibbs Sampler		Metropolis	
M	3	4	3	4	3	4
β	0.4	0.4	0.45	0.45	0.5	0.5
$\hat{\beta}_{MPL}$	0.4460	0.4878	0.3849	0.4064	0.4814	0.4889
$\lvert\beta - \hat{\beta}_{MPL}\rvert$	0.0460	0.0878	0.0651	0.0436	0.0186	0.0111
$\hat{I}^1_{obs}$	0.4694	0.6825	0.8450	1.3106	0.3908	0.8258
$\hat{I}^2_{obs}$	3.0080	3.3181	3.8248	4.5387	2.2935	2.6436
$\hat{Var}_n(\hat{\beta}_{MPL})$	0.0519	0.0620	0.0578	0.0636	0.0743	0.1182
Z_n	0.2458	0.3571	0.2707	0.1729	0.0682	0.0322
p-values	0.8104	0.7264	0.7872	0.8650	0.9520	0.9760

Interpreting the results





Metrics for Performance Evaluation

- Cohen's Kappa coefficient
  - Agreement between ground truth and classifier output
  - Calculated directly from the confusion matrix

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1c} \\ c_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \\ c_{c1} & \cdots & & c_{cc} \end{bmatrix}$$

$$\hat{K} = \frac{N \sum_{i=1}^{C} c_{ii} - \sum_{i=1}^{C} x_{i+}x_{+i}}{N^2 - \sum_{i=1}^{C} x_{i+}x_{+i}}$$

<u>Asymptotic variance</u>

$$\hat{\sigma}_k^2 = \frac{1}{N}\left[\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2} + \frac{2(1-\theta_1)(2\theta_1\theta_2-\theta_3)}{(1-\theta_2)^3} + \frac{(1-\theta_1)^2(\theta_4-4\theta_2^2)}{(1-\theta_2)^4}\right]$$

$$\theta_1 = \frac{1}{N}\sum_{i=1}^{C} x_{ii} \qquad \theta_2 = \frac{1}{N^2}\sum_{i=1}^{C} x_{i+}x_{+i}$$

$$\theta_3 = \frac{1}{N^2}\sum_{i=1}^{C} x_{ii}(x_{i+}+x_{+i}) \qquad \theta_4 = \frac{1}{N^3}\sum_{i=1}^{C}\sum_{j=1}^{C} x_{ij}(x_{j+}+x_{+i})^2$$

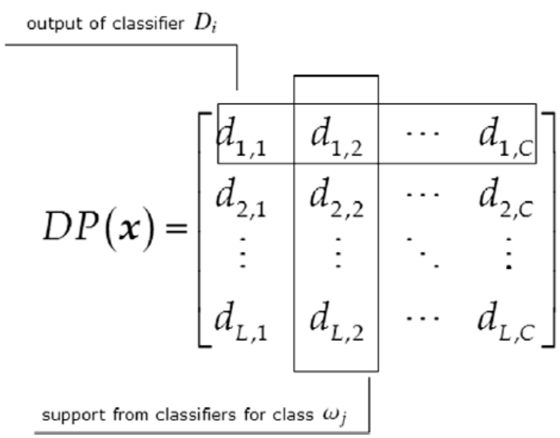
<u>Significance test</u>

$$Z_n = \frac{\left|\hat{k}_1 - \hat{k}_2\right|}{\sqrt{\hat{\sigma}_{k_1}^2 + \hat{\sigma}_{k_2}^2}}$$

$$Z_n > 1.96$$

Combining Contextual Classifiers

- For contextual information fusion
  - Decision profile



Combining function

$$\mu_j(\vec{x}) = \Im\left[d_{i,j}, \ldots, d_{L,j}\right]$$

The final classification label is given by the index of the maximum of $\mu_j(\vec{x})$

Classifier combination functions for soft decision fusion.

Sum	Product	Minimum	Maximum	Median
$\mu_j(\mathbf{x}) = \sum_{i=1}^{L} d_{i,j}(\mathbf{x})$	$\mu_j(\mathbf{x}) = \prod_{i=1}^{L} d_{i,j}(\mathbf{x})$	$\mu_j(\mathbf{x}) = \min_i\{d_{i,j}(\mathbf{x})\}$	$\mu_j(\mathbf{x}) = \max_i\{d_{i,j}(\mathbf{x})\}$	$\mu_j(\mathbf{x}) = \mathrm{median}_i\{d_{i,j}(\mathbf{x})\}$

Experiments and Results

- Experiments with noisy MRI brain images (marmosets)
  - CInAPCe project
    - Brazilian research project that has as main purpose the establishment of a scientific network seeking the development of neuroscience research through multidisciplinary approaches
  - 3 classes (300 samples per class)*
    - White matter
    - Gray matter
    - Background

* It represents only 1% of the image pixels

Experimental Setup

- The experiments were performed to compare the effect of using single and multiple initializations simultaneously.
- To generate initial conditions, seven pointwise statistical classifiers were employed.
  - Linear (LDC) and Quadratic (QDC) Bayesian classifiers under Gaussian hypothesis
  - Logistic classifier (LOGLC)
  - K-Nearest-Neighbor (KNNC)
  - Parzen Classifier (PARZENC)
  - Nearest Mean Classifier (NMC)
  - Decision Trees (TREEC)

Experimental Setup

- Iterative algorithms
  - ICM (*Iterated Conditional Modes*)
  - GSA (*Game Strategy Approach*)
  - MPM (*Maximizer of the Posterior Marginals*)
- Convergence criterion
  - For ICM and GSA algorithms
    - Less than 0.1% of the pixels are modified
    - The maximum of 5 iterations is reached
  - For the MPM algorithm
    - MCMC simulation parameters:
      - *Burn-in* window: k = 10
      - Number of samples: n = 50

Comparison of the Best Results

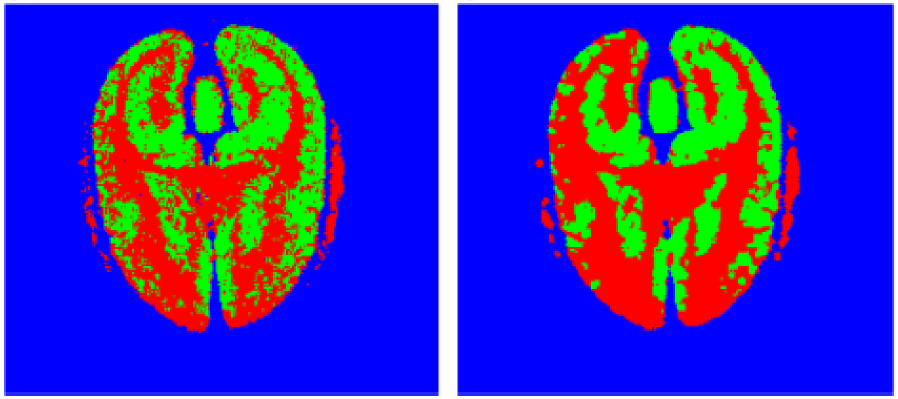


Fig. 11. Classification maps for the best general individual ICM contextual classifier and for the best result obtained by combining all the initial conditions in ICM

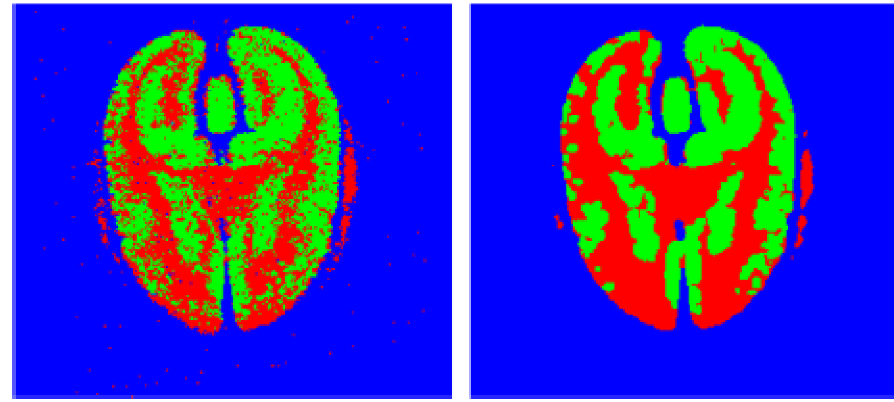


Fig. 12. Classification maps for the best general individual GSA contextual classifier and for the best result obtained by combining all the initial conditions in GSA

Comparison between the best general individual performance and the best result obtained by combining all the initial conditions for ICM, GSA and MPM combinatorial optimization algorithms.

Optimization algorithm	ICM	GSA	MPM
	Kappa	Kappa	Kappa
Best individual	0.9617(LDC)	0.9367(LDC)	0.9833(LDC)
Best combination	0.9850(Max)	0.9950(Max)	0.9983(Sum)
Z statistic	2.4988(>1.96)	5.5745(>1.96)	2.7278(> 1.96)

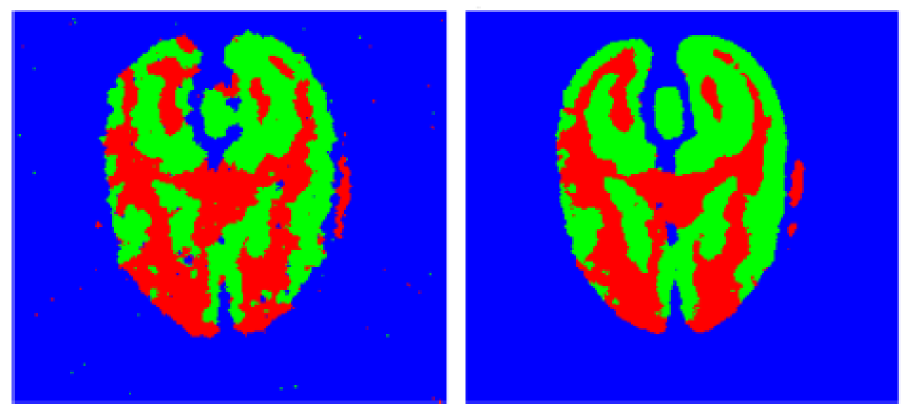


Fig. 13. Classification maps for the best general individual MPM contextual classifier and for the best result obtained by combining all the initial conditions in MPM

Average Performances and Elapsed Times

- The average performances are significantly improved when using multiple initializations.

Algorithm	ICM Average Kappa	GSA Average Kappa	MPM Average Kappa
Single Initialization	0.9021	0.8922	0.9269
Multiple Initializations	0.9767	0.9601	0.9897
T statistic	3.8971 (> 1.943)	3.3873 (> 1.894)	2.5361 (> 1.943)

- In general, the GSA algorithm is the fastest one.
  - With 7 initial conditions
    - MPM: 3991 s
    - ICM: 370 s
    - GSA: 298 s
      - All algorithms were implemented using MATLAB

Conclusions and Final Remarks

- Statistical analysis showed that the proposed method is valid, and more, it is capable of significantly improving the classification performance
  - The combination scheme tries to avoid convergence to "poor" local maxima solutions
- Maximum Pseudo-Likelihood estimation allowed automatic determination of MRF parameters.
- Future works include
  - Investigation of the best tradeoff between classification performance and computational cost
  - Incorporation of additional information through higher-order neighborhood systems

Acknowledgments

- Hilde Buzzá por for the acquisition of the MRI brain images.
- We would like to thank FAPESP for the financial support.

Thanks !!!

Alexandre L. M. Levada – alexandre.levada@gmail.com
Nelson D. A. Mascarenhas – nelson@dc.ufscar.br
Alberto Tannús – goiano@ifsc.usp.br