Jugowice, 11th–15th September 2017

MAXIMUM ENTROPY PRINCIPLE CLUSTERING IN SEGMENTATION OF MULTIMODAL MEDICAL IMAGES

Mateusz Cieszyński[1], Karol Górski[2], Piotr Boguś[3]

[1.3]Medical University of Gdańsk, Department of Physics and Biophysics
Dębinki 1, 80-822 Gdańsk ,
[2]University of Copernicus Hospital, Nowe Ogrody 1-6, 80-803 Gdańsk
[3]piotr.bogus@gumed.edu.pl

ABSTRACT

The paper presents the idea of applying maximum entropy principle MEP and statistical physics methodology in clustering for segmentation of multimodal medical images. The MEP clustering methods are compared with clustering algorithms which base on fuzzy and possibilistic approach. The paper shows that the fuzzy and possibilistic approach is equivalent to MEP for some special conditions e.g. without normalization constraint. The considerations in this field allow better understanding of the application of MEP in clustering and are exemplified by results of MRI images segmentation using the all mentioned techniques.

INTRODUCTION

The clustering is still a developing technique which is applied in image segmentation. There are many clustering methods and many of them are used in the area of medical images segmentation "[2], [5], [7], [10], [11], [12], [13], [16], [18]". The leading role in clustering is playing by algorithms based on statistical physics with using maximum entropy principle MEP ("[17], [20], [21]"). Many researches are still performed to solve theoretical and practical aspects of applying MEP in clustering (see e.g. "[20], [21]"). In medicine at present there is a big development of new imaging technologies like CT, MRI, SPECT or PET. The physician who interprets a medical image must recognize all anatomical structures and find any abnormality. The problem is that physicians are very efficient in interpretation of one image but their ability fall down significantly while they have to interpret two or more images represented the same section at the same time. This problem can be solved using multimodal imaging techniques "[2], [4], [8]".

FORMULATION OF CLUSTERING PROBLEM

The classical clustering problem "[1], [6]" consider the set $\boldsymbol{\Omega}$ consists of n date vectors and a set $\boldsymbol{Y}$ of c clusters represented by their centroids

$$\boldsymbol{\Omega} = \{\boldsymbol{x}_i; i = 1,...,n\} \qquad \boldsymbol{Y} = \{\boldsymbol{y}_j; j = 1,...,m\} \tag{1}$$

For each data point i the normalization condition is usually demand

$$\sum_{j=1}^{c} p_{ij} = 1 \tag{2}$$

where $p_{ij} = p(\mathbf{x}_i \in cluster_j)$ is the probability that a data point $\mathbf{x}_i$ belongs to a cluster j. The local energy (cost function, similarity measure) is introduced as a function $E_{ij} = E(\mathbf{x}_i \in cluster_j)$ of association data point x_i to the cluster j. Then the total averaged energy is

$$< E >= \sum_{i=1}^{n}\sum_{j=1}^{c} p_{ij} E_{ij} \tag{3}$$

The classical "hard" c-means HCM algorithm can be found as optimization of the objective function "[1], [6]" in the form

$$J = \sum_{i}^{n}\sum_{j=1}^{c} v_{ij} \left| \boldsymbol{x}_i - \boldsymbol{y}_j \right|^2 = \sum_{i=1}^{n}\sum_{j=1}^{c} v_{ij} d_{ij} \tag{4}$$

where

$$v_{ij} = \begin{cases} 1 & \text{if } x_i \text{ belongs to cluster } j, \\ 0 & \text{if } x_i \text{ does not belong to cluster } j. \end{cases} \tag{5}$$

In turn the fuzzy clustering methods "[1], [3], [6]" base on a concept of fuzzy partition of the data set. For fuzzy approach, an objective function was defined as "[3]":

$$J_m(v,Y) = \sum_{i=1}^{n}\sum_{j=1}^{c} \left(v_{ij}\right)^m d_{ij} \tag{6}$$

where m is a weighting exponent, $m \in [1,\infty)$, d_{ij} is a distance of the given vector $\mathbf{x}_i$ to the centroid $\mathbf{y}_j$ of cluster j and $\mathbf{v} \in \mathbf{M}_{fc}$ is the fuzzy c-partition of the data set $\mathbf{\Omega}$ which have following properties

- $$\forall_{ij}\, v_{ij} \in <0,1>, \tag{7}$$
- $$\forall_i \sum_{j=1}^{c} v_{ij} = 1, \tag{8}$$
- $$\forall_i\, 0 < \sum_{i=1}^{n} v_{ij} < n. \tag{9}$$

The essence of the fuzzy c-mean algorithm FCM is the theorem of Bezdek "[3]" which formulates the FCM algorithm as iterations of the fuzzy cluster centroids by formula

$$\boldsymbol{y}_j^t = \frac{\sum_{i=1}^{n} (v_{ij}^{t-1})^m \boldsymbol{x}_i}{\sum_{i=1}^{n} (v_{ij}^{t-1})^m} \tag{10}$$

for j=1,.....,c with new matrix v_{ij} calculated by

$$v_{ij} = \frac{1}{\sum_{s=1}^{c} \left(\frac{d_{ij}}{d_{is}} \right)^{\frac{2}{m-1}}} \tag{11}$$

The FCM algorithm reduces to HCM algorithm for m=1 and d_{ij} taken as Euclidean distance. The FCM algorithm was criticized because of the condition "Eq. (8)" that makes membership coefficients similar to probabilities. Only possibilistic approach can be considered as a fully fuzzy. The essence of the possibilistic approach "[9], [10]" lays in ignoring the constraint Eq. (8). The objective function for a possibilistic approach in clustering was formulated as "[9]"

$$J_m = \sum_{j=1}^{c}\sum_{i=1}^{n} (v_{ji})^m E_{ji} + \sum_{j=1}^{c} \eta_j \sum_{i=1}^{n} \left(1 - v_{ji}\right)^m \tag{12}$$

where η_j are positive numbers. Optimizing over "Eq. (12)" leads to the following equation for v_{ij}

$$v_{ji} = \frac{1}{1+\left(\frac{E_{ij}}{\eta_j}\right)^{\frac{1}{m-1}}} \tag{13}$$

In "[10]" there was an alternative formulation of the possibilistic objective function

$$J_m = \sum_{j=1}^{c}\sum_{i=1}^{n} v_{ji}E_{ji} + \sum_{j=1}^{c}\eta_j \sum_{i=1}^{n}\left(v_{ji}\,ln\,v_{ji} - v_{ji}\right) \tag{14}$$

Performing the optimization of Eq. (14) is giving the following solution

$$v_{ji} = exp\left(-\frac{E_{ji}}{\eta_j}\right) \tag{15}$$

MAXIMUM ENTROPY PRINCIPLE IN CLUSTERING

For classically formulated clustering problem "[17], [19], [20]" the entropy is

$$H(p_{11},p_{12},.....,p_{nc}) = -\sum_{i=1}^{n}\sum_{j=1}^{c} p_{ij}\ln p_{ij} \tag{16}$$

where the summation is performed over all the clusters and all the data points. The entropy "Eq. (16)" can be maximized under constraints: normalization conditions Eq. (2) and the given expectation value of energy "Eq. (3)". The full entropy "Eq. (16)" can be written as the sum of partial entropies considered for different data point $\mathbf{x}_i$ (where i=1,.....,n). The corresponding Lagrangian function of this problem can be written also in this way

$$L = \sum_{i=1}^{n} L_i \qquad L_i = -\sum_{j=1}^{c} p_{ij}\ln\, p_{ji} - \lambda_i \sum_{j=1}^{c} p_{ij} - \beta\sum_{j=1}^{c} p_{ij}E_{ji} \tag{17}$$

where L_i are partial Lagrangian functions. Usually the assumption *that the probabilities relating different* $\mathbf{x}_i$ *to their clusters are independent* is taken. Then the full optimization can be done by optimization of each partial Lagrangian function separately. The result of this optimization gives the Gibbs distribution

$$p_{ij} = \frac{\exp\left(-\beta E_{ij}\right)}{Z_i} \tag{18}$$

where Z_i is a partial partition function. In statistical mechanics, the Lagrange multiplier β is interpreted as the inverse of temperature T (β=1/T). The total free energy is then

$$F = \sum_{i=1}^{n} F_i = -\sum_{i=1}^{n}\frac{1}{\beta}ln\sum_{j=1}^{c} exp\left(-\beta d_{ij}\right) = -\frac{1}{\beta}\sum_{i=1}^{n} ln\sum_{j=1}^{c} exp\left(-\beta d_{ij}\right) \tag{19}$$

It can be demonstrated that $\lim_{\beta\to\infty} F = \langle E \rangle$ and $\lim_{\beta\to\infty}\left(\min_Y F\right) = \min_Y \langle E \rangle$. These limits allow us to find the solution of the constrained minimization of <E> by performing so called *Deterministic Annealing* on F as first proposed by Rose, Gurewitz and Fox "[17]". The general concept of annealing is to track the global minimum of *F* while increasing β. In the classical approach, the energy is described as Euclidean distance between data points $\mathbf{x}_i$ and the cluster centroids $\mathbf{y}_j$ ($E_{ij} = d_{ij} = d(\mathbf{x}_i,\mathbf{y}_j) = |\mathbf{x}_i - \mathbf{y}_j|^2$) and then from "Eq. (18)" the association probabilities are

$$p_{ij} = \frac{\exp\left(-\beta\left|\boldsymbol{x}_i - \boldsymbol{y}_j\right|^2\right)}{\sum_{k=1}^{c}\exp\left(-\beta\left|\boldsymbol{x}_i - \boldsymbol{y}_k\right|^2\right)} \tag{20}$$

and corresponding free energy is

$$F(Y) = -\frac{1}{\beta}\sum_{i=1}^{n} ln \sum_{j=1}^{c} exp\left(-\beta\left|\boldsymbol{x}_i - \boldsymbol{y}_j\right|\right) \tag{21}$$

Then the variational free energy minimization method can be used “[12], [14]”. It takes the set of cluster representatives **Y** (see Eq. (1)) as parameters. It is obtained eventually

$$\boldsymbol{y}_j = \frac{\sum_{i=1}^{n} p_{ij}\,\boldsymbol{x}_i}{\sum_{i=1}^{n} p_{ij}} \quad \text{for all j} \tag{22}$$

what naturally leads to fixed-point iterations based on “Eqs. (20), (22)”.

MAXIMUM ENTROPY PRINCIPLE IN CLUSTERING WITHOUT NORMALIZATION CONSTRAINT

For the case without a normalization constraint the Lagrangian function will be as following

$$L = \sum_{i=1}^{n} L_i = \sum_{i}^{n}\left(-\sum_{j=1}^{c} p_{ij}\ln\, p_{ji} + \beta\left(< E > -\sum_{j=1}^{c} p_{ij}E_{ji}\right)\right) \tag{23}$$

Because there are no normalization constraints instead of symbol p_{ij} the symbol of generalized probability u_{ij} will be used. Optimization over partial Lagrangian gives

$$u_{ij} = \frac{exp\left(-\beta\, E_{ij}\right)}{e} \quad \text{where} \quad Z_{ij} = e \quad \text{and} \quad F_i = -\frac{1}{\beta} \tag{24}$$

Then a full free energy is

$$F = -\frac{1}{\beta}\ln Z = -\frac{1}{\beta}\ln\left(\prod_{i=1}^{n} Z_i\right) = -\frac{1}{\beta}\ln\left(\prod_{i=1}^{n} e\right) = -\frac{1}{\beta}\ln e^n = -\frac{n}{\beta} \tag{25}$$

The full free energy is going to zero while $\beta \to \infty$. Taking into account “Eq.(25)” one obtains

$$\lim_{\beta\to\infty} < E > = \lim_{\beta\to\infty} F + \lim_{\beta\to\infty}\left(\frac{1}{\beta}H_{max}\right) = 0 \tag{26}$$

In this place the question appears: how to interpret the above results? Considering the general solution of maximum entropy principle gives “Eq. (20)” in the form

$$u_j(x_i) = \frac{exp\left(-\beta\, E_j(\boldsymbol{x}_i)\right)}{\sum_{k=1}^{c} exp\left(-\beta\, E_k(\boldsymbol{x}_i)\right)} \tag{27}$$

Let us express an energy in logarithm way $\mathrm{E}_j(\boldsymbol{x}_i) \equiv \log\left(\mathrm{E}_j^{\beta}(\boldsymbol{x}_i)\right)$ and introduce the parameter m by

$$\beta \equiv \frac{2}{m-1} \tag{28}$$

Now the final form of $u_j(x_i)$ appears to be equivalent to this which was obtain in a FCM algorithm

$$u_j(x_i) = \frac{exp\left(-\frac{2}{m-1}\, log\, E_j^{\beta}(\mathbf{x}_i)\right)}{\sum_{k=1}^{c} exp\left(-\frac{2}{m-1}\, log\, E_k^{\beta}(\mathbf{x}_i)\right)} == \frac{1}{\sum_{k=1}^{c}\left(\frac{E_j^{\beta}(\mathbf{x}_i)}{E_k^{\beta}(\mathbf{x}_i)}\right)^{\frac{2}{m-1}}} \tag{29}$$

Using regularization theory in optimization problems is nowadays the classical approach to ill-posed problems (see e.g. “[15]”). In this approach, the standard FCM and MEP methods can be considered as different types of regularization for the classic HCM algorithm. Let us now consider the MEP optimization problem from the regularization theory point of view. First of all, find that now in “Eq. (17)” values of <E> and 1 are constant hence after some transformations Lagrangian function “Eq. (17)” can be formulated as

$$L = \sum_{i=1}^{n} u_i E_i + \frac{1}{\beta}\sum_{i=1}^{n} u_i \ln u_i + \lambda \sum_{i=1}^{n} u_i \tag{30}$$

where a new constant λ is a result of dividing an old value by β. It leads us to a conclusion that there is a possibility of such interpretation where a main term of the Lagrangian function is an energy, not an entropy. From the mathematical point of view, it is the same to optimize the entropy with normalization and energy constraint and to optimize the energy with normalization constraint and constraint for entropy. The second approach is not easy to explain. The entropy is a function of the probability distribution and this approach can be interpreted as the minimization of the energy with an entropy constraint on probability distribution. From the regularization theory point of view in MEP approach the energy should correspond to classical „error” term and the entropy (and other normalization term) should correspond to „regularization” terms. Find also that the Lagrangian function for the clustering problem where entropy is given by “Eq. (17)” can also be written in the form where entropy is one of regularization terms

$$L = \sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij} E_{ij} + \frac{1}{\beta}\sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij} \ln u_{ij} + \sum_{i=1}^{n} \lambda_j \Big(\sum_{j=1}^{c} u_{ij}\Big) \tag{31}$$

Taking into account the above considerations the following schema of the clustering algorithms is proposed.

Classical HCM algorithm of clustering can be considered as the optimization of the objective function in the form

$$J = \sum_{i}^{n}\sum_{j=1}^{c} u_{ij}\left|\mathbf{x}_i - \mathbf{y}_j\right|^2 = \sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij} d_{ij} \tag{32}$$

The MEP algorithm of clustering with constraints can be considered as the optimization of the objective function

$$J = \sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij} E_{ij} + \frac{1}{\beta}\sum_{i=1}^{n}\sum_{j=1}^{c} u_{ij} \ln u_{ij} + \sum_{i=1}^{n} \lambda_j \Big(\sum_{j=1}^{c} u_{ij}\Big) \tag{33}$$

With an assumption that the probabilities relating different $\mathbf{x}_i$ to their clusters are independent, the solution gives here the Gibbs distributions and finally one obtains a deterministic schema.

The FCM algorithm of clustering can be considered as the optimization of the objective function in the form

$$J = \sum_{i=1}^{n}\sum_{j=1}^{c} \left(u_{ij}\right)^m d_{ij} \tag{34}$$

This is an example of nonlinear regularization. It introduces nonlinearity in the „error” term of objective function. Taking into account the equivalence of MEP and FCM one can find here the possibility of preparing the deterministic annealing. Find that from “Eq. (28)” we have m = 2/β +

1 = 2T + 1. When $\beta \to \infty$ then $m \to 1$ and when $\beta \to 0$ then $m \to \infty$. This gives the possibility of the deterministic annealing for FCM algorithm by going with m from ∞ to 1 (see e.g. “[3]”).

The possibilistic PCM algorithm can be considered as the optimizing of the objective function

$$J = \sum_{j=1}^{c}\sum_{i=1}^{n} u_{ji} E_{ji} + \eta_j \sum_{j=1}^{c}\sum_{i=1}^{n}\left(u_{ji}\ ln\ u_{ji} - u_{ji}\right) \tag{35}$$

and taking into account that *each cluster is independent of the other clusters* the optimization is performing over each objective function corresponding to the given cluster j

$$J_j = \sum_{i=1}^{n} u_{ji} E_{ji} + \eta_j \sum_{i=1}^{n}\left(u_{ji}\ ln\ v_{ji} - u_{ji}\right) \text{ where } J = \sum_{j=1}^{c} J_j \tag{36}$$

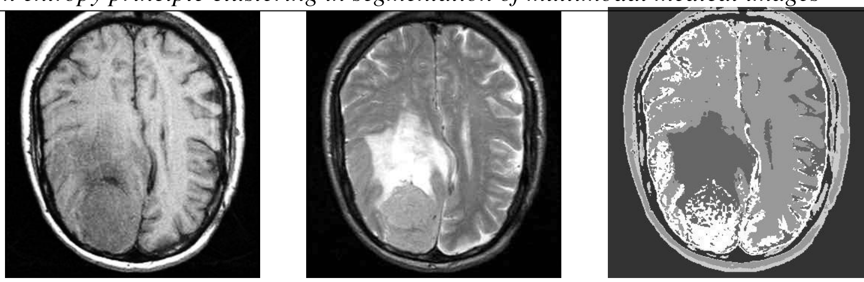
Find that in the MEP Lagrangian function “Eq. (33)” a term corresponding to the normalization constraint has its own separate Lagrangian multiplier. In possibilistic approach there are the same multipliers for the given energy and normalization terms. Considering a function “Eq. (35)” and making some transformations the possibilistic objective function can be expressed as

$$J = \sum_{j=1}^{c}\sum_{i=1}^{n} u_{ji} E_{ji} + \sum_{j=1}^{c}\eta_j \sum_{i=1}^{n} u_{ji}\ ln\ u_{ji} \tag{37}$$

what corresponds to MEP Lagrangian function “Eq. (35)” without a normalization constraint.

RESULTS

Using p different imaging methods (e.g. PET and CT in scanner PET/CT or three MRI modes PD, T1 and T2) of the same given structure give us the possibility of using multimodal image techniques. Multimodal image is the set of p images representing different physical and chemical properties of the same cross-section. In multimodal image each pixel has p corresponding values of it intensity in successive image modes. The multimodal image creates p-dimensional feature space considering each mode as one feature “[4], [8]” where clustering is performed. After clustering the labelling of successive clusters give a final segmentation. After segmentation, a physician instead of two or more images of the same cross-section obtain one segmented image which contain all information from all modes. Figure 1 presents the results the two-modal MRI image segmentation with visible tumor for four clustering methods: HCM, FCM, PCM and MEP. Comparing the segmentation results the classic c-means methods give us the simply picture while the FCM, MEP and PCM approaches show structural segments and more complicated pictures of a tumor. Generally, figures are showing the similar results for fuzzy and MEP clustering algorithms.



A B C

D E F

Fig. 1. The results of segmentation of two-modal MRI image (A – T1 image, B – T2 image) for algorithms HCM (C), FCM (D), MEP (E) and PCM (F).

CONCLUSION

The comparison of segmentation results of multimodal medical images shows the big variability depending different clustering methods, both in the speed of algorithms and in final image qualities. The example results (see fig. 1) are not perfect but this problem is strongly depended on algorithms convergence and their sensitivity to initial conditions in cluster centers. There is also a question: how to explain the MEP approach without normalization constraint. From a practical point of view the possibilistic approach can agree with the MEP. Consider the set of elements with an assumption that the number of clusters is c=2 (“Fig. 2”). The one outstanding point is usually considered as an effect of the noise. But it can be taken as a third little, unknown and unexpected cluster. Then the normalization condition should be “for all i: $v_{i1} + v_{i2} + v_{i3} = 1$”. But this way the sum over two clusters will fulfil “for all i: $v_{i1} + v_{i2} \neq 1$ and $v_{i1} + v_{i2} < 1$”. In conditions of incomplete information, it can happen that the number of clusters is bigger than a prior assumption. From this point of view the maximization of entropy without the normalization constraint have a sense for clustering with established number of clusters.

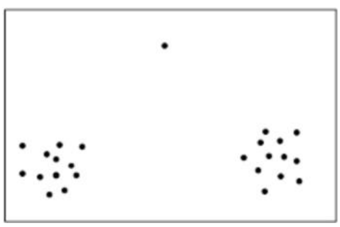


Figure 2. Example of a data set with one outstanding point.

REFERENCES

[1] J.C. Bezdek and S.K. Pal: *Fuzzy models for pattern recognition: methods that search for structures in data*, Institute of Electrical and Electronics Engineers, New York, 1992.

[2] J.C. Bezdek, I.O. Hall and L.P. Clarke: *Review of MR image segmentation techniques using pattern recognition*, Medical Physics, **20** (1993), 1033-1048.

[3] J.C. Bezdek: *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York, Second Edition, 1987.

[4] P. Boguś, A.M. Massone, F. Masulli and A. Schenone: *Interactive graphical system for segmentation of multimodal medical volumes using fuzzy clustering*, Machine Graphics & Vision **7** (1998), 781-791.

[5] M. K. Date and S. P. Akarte: *Brain Image Segmentation Algorithm using K-Means Clustering*, International Journal Of Computer Science And Applications **6** (2013), 285-289.

[6] R. Duda, P.E. Hart and D.G. Stork: *Pattern Classification*, Second Edition, John Wiley & Sons, New York, 2001.

[7] H. Kaur and H. Kaur: *Segmentation Techniques Comparison in Medical Image Processing*, International Journal of Advanced Research in Computer Science & Technology **4** (2016), 52-54.

[8] M.K. Kessler, S. Potluck, P. Petti and J.R. Castro, J.R: *Integration of multimodality imaging data for radiotherapy treatment planning*, Journal of Radiation Oncology, Biology and Physics, 21, 6, (1991), 1653-1667.

[9] R. Krishnapuram and J.M. Keller: *A Possibilistic Approach to Clustering*, IEEE Transactions on Fuzzy Systems **1** (1993), 98-110.

[10] R. Krishnapuram and J.M. Keller: *The Possibilistic C-Means Algorithm: Insights and Recommendations*, IEEE Transactions on Fuzzy Systems **4** (1996), 385-393.

[11] G. S. Mahajan and K. S. Bhagat: *Survey On Medical Image Segmentation Using Enhanced K-Means And Kernelized Fuzzy C- Means*, International Journal of Advances in Engineering & Technology **6** (2014), 2531-2536.

[12] D.J.C. MacKay: *Information Theory, Inference and Learning Algorithms*, Cambridge University Press, 2003.

[13] A. Norouzi, M. S. M. Rahim, A. Altameem, T. Saba, A. E. Rad, A. Rehman and M. Uddin: *Medical Image Segmentation Methods, Algorithms, and Applications*, IETE Technical Review **31** (2014), 199-213.

[14] P. M. Parekh, D. Katselis, C. L. Beck and S. M. Salapaka: *Deterministic Annealing for Clustering: Tutorial and Computational Aspects*, Proceedings of „2015 American Control Conference, July 1-3, 2015. Chicago, IL, USA, 2906-2911.

[15] T. Poggio and F. Girosi: *Networks for Approximation and Learning*, Proceedings of the IEEE **78** (1990), 1481-1497.

[16] A. Rodriguez and A. Laio: *Clustering by fast search and find of density peaks*, Science **344** (2014), 1492-1496.

[17] K. Rose, E. Gurewitz and G. Fox: *A deterministic annealing approach to clustering*, Pattern Recognition Letters **11** (1990), 589-594.

[18] K. Shrivastava, N. Gupta and N. Sharma: *Medical Image Segmentation using Modified K Means Clustering*, International Journal of Computer Applications 108 (2014), 12-15.

[19] M. Yasuda: *Entropy Maximization and Deterministic Annealing Approach to Fuzzy C-means Clustering*, Journal of Innovative Computing, Information and Control **5** (2009), 1349-4198.

[20] M. Yasuda: *Fuzzy c-Means Clustering, Entropy Maximization, and Deterministic and Simulated Annealing*. Chapter 13 in: *Computer and Information Science – Numerical Analysis and Scientific Computing, "Simulated Annealing - Advances, Applications and Hybridizations"* , M. de Sales G. Tsuzuki (eds.), INTECH, 2012.