On the Estimation of Mass Functions Using Self Organizing Maps

Imen Hammami[1], Jean Dezert[2], Grégoire Mercier[1], and Atef Hamouda[3]

[1] Telecom Bretagne, Brest 29238, France
imen.hammammi@gmail.com, gregoire.mercier@telecom-bretagne.eu
[2] The French Aerospace Lab, F-91761 Palaiseau, France
jean.dezert@onera.fr
[3] Faculty of Sciences of Tunis, El Manar Tunis 2092, Tunisie
atef.hamouda@fst.rnu.tn

Abstract. In this paper, an innovative method for estimating mass functions using Kohonen's Self Organizing Map is proposed. Our approach allows a smart mass belief assignment, not only for simple hypotheses, but also for disjunctions and conjunctions of hypotheses. This new method is of interest for solving estimation mass functions problems where a large quantity of multi-variate data is available. Indeed, the use of Kohonen map that allows to approximate the feature space dimension into a projected 2D space (so called map) simplifies the process of assigning mass functions. Experimentation on a benchmark database shows that our approach gives similar or better results than other methods presented in the literature so far, with an ability to handle large amount of data.

Keywords: Evidence Theory; Belief assignment; Kohonen map; Estimation.

1 Introduction

When it comes to exploit the redundancy and the complementarity of information stemming from very varied sources to give a unique representative information, the belief function theory, introduced by Dempster [1] and formalized by Shafer [2], is considered as an appealing formalism in information fusion domain. Indeed, it offers a mathematical framework that allows the processing of both imprecise and uncertain information.

The initial theory was improved in different directions, for example through the work of Dezert-Smarandache [3], a paradoxical reasoning has been proposed. Despite, the fact that belief function theory performs well in extracting the most truthful proposition from a multisource context, it nevertheless presents a major difficulty that is the estimation of basic belief assignments.

If we take a look at the various works that deal with this problem of belief function estimation, we distinguish two main family approaches. Likelihood based approaches [2, 4], require the knowledge, or the estimation, of the conditional probability density for each class. The second family is the distance-based approaches [5, 6]. However, these two types of estimation present some limits: among them we can mention the need of the *a priori* knowledge on the hypotheses which is not always easy to know, especially, for compound hypotheses.

An original method is presented in this paper to estimate mass functions when a large quantity of multi-variate data is available. In the feature space (in $\mathbb{R}^p$), operations on Basic Belief Assignment (BBA) can be much more complex and may not be feasible due to computing time or accuracy consideration. The proposed method to overcome this limitation is based on Kohonen's Map that allows to approximate the feature space dimension into a projected 2D space (so called map). Then, the use of Kohonen's map simplifies the process of assigning mass functions on conjunctions and disjunctions of hypotheses when considering relative distance of an observation to the map. Thus, it can model at the same time ignorance, imprecision, paradox as result exploits all the conceptual contribution of the theory.

2 Evidence Theory

Dempster-Shafer Theory (DST) is used for representing belief on imperfect observation through the Basic Belief Assignment (BBA), defined on all the subsets of the frame of discernment Θ, noted 2^Θ. In our context, $m(\cdot)$ will have to be built from the observation provided by a sensor, that is from a sample $\boldsymbol{x} \in \mathbb{R}^p$. A BBA $m(\cdot)$ is the mapping from elements of the power set 2^Θ onto $[0, 1]$ under constraints:

$$\begin{cases} m(\boldsymbol{x} \in \emptyset) = 0 \\ \sum_{A \subseteq 2^\Theta} m(\boldsymbol{x} \in A) = 1. \end{cases} \tag{1}$$

The frame of discernment Θ is the set of possible answers of the problem under concern. It is composed of exhaustive and exclusive hypotheses: $\Theta = \{\theta_1, \theta_2, \ldots, \theta_N\}$. From this frame of discernment, the power set noted 2^Θ can be built, including all the disjunctions of hypotheses θ_i such as $\theta_i \cup \theta_j$ or $\theta_i \cup \theta_j \cup \theta_k$... As discussed above, Dezert-Smarandache theory (DSmT) [3] is considered as amelioration of beliefs theory. The main idea of DSmT is to work on the hyperpowerset of the frame of discernment. The hyper-power set D^Θ is defined as the Dedekind's lattice built from Θ whith $\cap$ and $\cup$ operators. For decision making from mass function, the Generalized Pignistic Transformation [3] noted $BetP_g$ is frequently used:

$$BetP_g(\boldsymbol{x} \in A) = \sum_{B \in D^\Theta} \frac{C_M(B \cap A)}{C_M(B)} m(\boldsymbol{x} \in B), \quad \forall A \in D^\Theta \tag{2}$$

where C_M is the cardinality within DSmT. The decision is taken by the maximum of pignistic probability function $BetP(\cdot)$. Similarly, the Pignistic Transformation $BetP$ can be used within DST framework for decision making.

3 Overview on Kohonen's Map

There exist many versions of the Self Organizing Maps (SOM). However, the basic philosophy is very simple and already effective [7]. A SOM defines a mapping from the input space (say $\mathbb{R}^p$) onto a regular array of $M \times N$ nodes (see Fig. 1) [8].

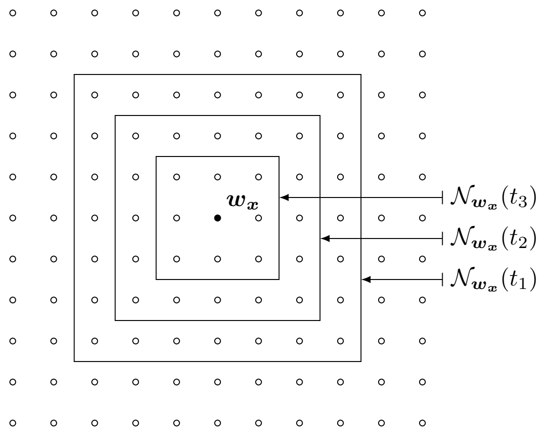

Fig. 1: A schematic view of a 11×11 Kohonen's Self Organizing Map. Several topological neighborhoods $\mathcal{N}_{\boldsymbol{w_x}}(t_i)$ of the winning neuron $\boldsymbol{w_x}$ are drawn. The size is decreasing with the number of iterations ($t_1 < t_2 < t_3$) during the training phase.

A reference vector, also called weighting vector, $\boldsymbol{w}(i,j) \in \mathbb{R}^p$ is associated to the node at each position (i,j) with $1 \leqslant i \leqslant N$ and $1 \leqslant j \leqslant M$. An input vector $\boldsymbol{x} \in \mathbb{R}^p$ is to be compared to each $\boldsymbol{w}(i,j)$. The best match is defined as output of the SOM: thus, the input data $\boldsymbol{x}$ is mapped onto the SOM at location $(i_{\boldsymbol{x}}, j_{\boldsymbol{x}})$ where $\boldsymbol{w}(i_{\boldsymbol{x}}, j_{\boldsymbol{x}})$ is the neuron the most similar to $\boldsymbol{x}$ according to a given metric. SOM performs a non linear projection of the probability density function $p(\boldsymbol{x})$ from the high-dimensional input data onto the two-dimensional array.

In practical applications, the Euclidean distance is usually used to compare $\boldsymbol{x}$ and $\boldsymbol{w}(i,j)$. The node that minimizes the distance between $\boldsymbol{x}$ and $\boldsymbol{w}(i,j)$ defines the best-matching node (or the so-called winning neuron), and is denoted by the subscript $\boldsymbol{w_x}$:

$$\|\boldsymbol{x} - \boldsymbol{w_x}\| = \min_{\substack{1 \leqslant i \leqslant M \\ 1 \leqslant j \leqslant N}} \|\boldsymbol{x} - \boldsymbol{w}(i,j)\|. \tag{3}$$

An optimal mapping would be the one that maps the probability density function $p(\boldsymbol{x})$ in the most faithful fashion, preserving at least the local structures of $p(\boldsymbol{x})$.

It can be considered also that the SOM achieves a non-uniform quantization that transforms $\boldsymbol{x}$ to $\boldsymbol{w_x}$ by minimizing the given metric. Nevertheless, thanks to the training phase (detailed below) the neurons $\boldsymbol{w}$ are located on the map according to their similarity. Then, when considering neurons $\boldsymbol{w}(i,j)$ located not *too far* from the winning neuron $\boldsymbol{w}_{\boldsymbol{x}}$, the distance in $\mathbb{R}^p$ between $\boldsymbol{x}$ and $\boldsymbol{w}(i,j)$ is not dramatically different from the one between $\boldsymbol{x}$ and $\boldsymbol{w}_{\boldsymbol{x}}$. That means that in the neighborhood of $\boldsymbol{w}_{\boldsymbol{x}}$ on the map, are located the wining neurons of the neighbors (in $\mathbb{R}^p$) of $\boldsymbol{x}$. Hence, a class in $\mathbb{R}^p$ is projected into the map at the same area, remaining homogeneous. Moreover, whatever the initial shape of the class in the $\mathbb{R}^p$ feature space is, the projected class is highly likely to be of isotropic shape in the map.

4 Feature Space for Smart BBA

The proposed smart BBA intends to evaluate the mass of each class in 2^{Θ} or 2^{D} according to the topology of the observed manifold. Then, two sets of data may be handled (see Fig. 2): on the first hand the initial observations $\boldsymbol{x}$ and class centers $\{C_1, C_2, \dots, C_K\}$ in $\mathbb{R}^p$ and, on the other hand the so-called *winning neurons* $\boldsymbol{w}_{\boldsymbol{x}}$ and the projected class centers $\boldsymbol{w}_{C_k}$.

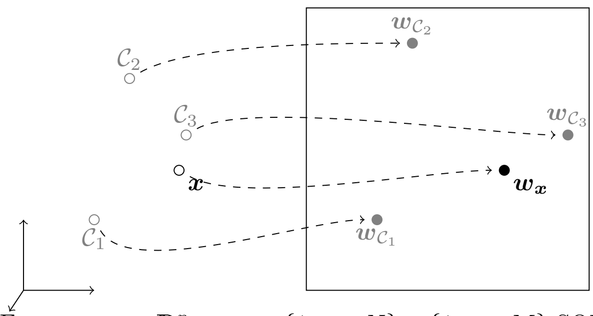

Fig. 2: Observations in the feature space and their projections into Kohonen's map. Note that the neurons $\boldsymbol{w}_{\boldsymbol{x}}$ and $\boldsymbol{w}_{C_k}$ can be located on the map through their location index (n,m) or in $\mathbb{R}^p$ with their p component value.

Then, Kohonen's map can be used to build easily BBA and to balance between conjunction and disjunction when considering relative distance of an observation to the map. Moreover, the use of Kohonen's map simplifies the evaluation of the masses since operations on the maps require calculation on index only, while operations on the feature space (in $\mathbb{R}^p$) may be much more complex (when dealing with stochastic divergence for instance). So two kinds of distances will be considered and their related difference will induce uncertainty:

1. $d_{\mathbb{R}^p}(\cdot,\cdot)$ which is the distance in $\mathbb{R}^p$. It can be defined through the Euclidean norm $\mathcal{L}^2(\mathbb{R}^p)$ but also through a spectral point of view such as the spectral angle mapper or the spectral information divergence [9]. It may also be based

on the Kullback-Leibler divergence or the mutual information when dealing with Synthetic Aperture Radar (SAR) [10].

2. $d_{\text{map}}(\cdot,\cdot)$ which is the distance along Kohonen's map. It is mainly based on the Euclidean norm and uses the index that locates the 2 vectors on the map: $d_{\text{map}}(\boldsymbol{w}_1,\boldsymbol{w}_2) = \sqrt{(n_1-n_2)^2+(m_1-m_2)^2}$ if $\boldsymbol{w}_1$ (resp. $\boldsymbol{w}_2$) is located at position (n_1,m_1) (resp. (n_2,m_2)) on the map.

5 A New Method to Build BBA

This section details a strategy for building a BBA by using Kohonen's map and an initial classifier on $\mathbb{R}^p$.

Mass of simple hypotheses: The definition of masses of focal elements could be based on the distance on the feature space. Nevertheless, an appropriated definition should take into account the variance of the classes to weight each of them, as it is the case in a likelihood point of view. This weighting is already performed by the projection onto Kohonen's map so that, the mass of focal class is defined as:

$$\begin{cases} m(\boldsymbol{x}\in\theta_k) & \sim 1 \qquad \text{if } \boldsymbol{w_x}{=}\boldsymbol{w}_{\mathcal{C}_k} \\ m(\boldsymbol{x}\in\theta_k) & \sim \dfrac{d_{\text{map}}(\boldsymbol{w_x},\boldsymbol{w}_{\mathcal{C}_k})^{-1}}{\sum_{\ell=1}^{K} d_{\text{map}}(\boldsymbol{w_x},\boldsymbol{w}_{\mathcal{C}_\ell})^{-1}} \quad \text{otherwise} \end{cases} \tag{4}$$

where $k=1,2,\ldots,K$, $\boldsymbol{w}_{\mathcal{C}_k}$ is the projected class, $\boldsymbol{w_x}$ is the winning neurons.

According to eq. (4), we consider that the more the distance $d_{\text{map}}(\boldsymbol{w_x},\boldsymbol{w}_{\mathcal{C}_k})$ (relatively to the other distances between $\boldsymbol{x}$ and $\mathcal{C}_\ell$ on the map) the less the mass $m(\boldsymbol{x}\in\theta_k)$.

Total ignorance case: From the feature space, we may consider that the mass evaluation of an observation falls into ignorance if its distance to the map is much more important that the distance of its related class center to the map. Then, it can be expressed as follows:

$$m(\boldsymbol{x}\in\Theta) \sim 1-\min\left(\frac{d_{\mathbb{R}^p}(\boldsymbol{x},\boldsymbol{w_x})}{d_{\mathbb{R}^p}(\mathcal{C}_{\boldsymbol{x}},\boldsymbol{w}_{\mathcal{C}_{\boldsymbol{x}}})},\frac{d_{\mathbb{R}^p}(\mathcal{C}_{\boldsymbol{x}},\boldsymbol{w}_{\mathcal{C}_{\boldsymbol{x}}})}{d_{\mathbb{R}^p}(\boldsymbol{x},\boldsymbol{w_x})}\right) \tag{5}$$

where $\mathcal{C}_{\boldsymbol{x}}$ is the class center of $\boldsymbol{x}$, $\boldsymbol{w}_{\mathcal{C}_{\boldsymbol{x}}}$ is its projection on the map.

Mass of the conjunction between two classes: The conjunction between two classes may be defined into the feature space as the space in-between the two classes. But, one has to account for the variance of each classes that increases the complexity of this measure. Once again, it is much more convenient to define the $\phi_k\cap\phi_\ell$ mass into Kohonen's map, as:

$$m(\boldsymbol{x}\in\theta_k\cap\theta_\ell) \sim e^{-\gamma(z-1)^2} \tag{6}$$

where $z = d_{\text{map}}(\boldsymbol{w_x}, \frac{\boldsymbol{w}_{\mathcal{C}_k}+\boldsymbol{w}_{\mathcal{C}_\ell}}{2}), 0 < k, \ell \leqslant K, \ell \neq k$. By adopting eq. (6) we consider that the value of $m(\boldsymbol{x} \in \theta_k \cap \theta_\ell)$ becomes maximal when $\boldsymbol{x}$ reaches the middle of $[\boldsymbol{w}_{\mathcal{C}_k}, \boldsymbol{w}_{\mathcal{C}_\ell}]$ segment. Eq. (6) yields a value of $m(\boldsymbol{x} \in \phi_k \cap \phi_\ell)$ closed to 1 in the middle. Moreover, $m(\boldsymbol{x} \in \phi_k \cap \phi_\ell)$ vanishes when $\boldsymbol{x}$ is far away from the $[\boldsymbol{w}_{\mathcal{C}_\ell}, \boldsymbol{w}_{\mathcal{C}_k}]$ segment. The γ parameter tunes this vanishing behavior. For example, if we want eq. (6) be over $\frac{1}{2}$ between the 1$^{\text{st}}$ and the 3$^{\text{rd}}$ quartile of $[\boldsymbol{w}_{\mathcal{C}_k}, \boldsymbol{w}_{\mathcal{C}_\ell}]$ segment, then γ should be equal to $2\sqrt{2}$. For a smaller domain around the median of $[\boldsymbol{w}_{\mathcal{C}_k}, \boldsymbol{w}_{\mathcal{C}_\ell}]$ segment, γ should be greater.

Mass of disjunction between two classes: The ignorance in the decision making between two classes $\mathcal{C}_k$ and $\mathcal{C}_\ell$ may be considered to a dual of eq. (6), but here by considering distances in the feature space. When a sample $\boldsymbol{x}$ is not *too far* from class $\mathcal{C}_k$ or $\mathcal{C}_\ell$, it is not *too difficult* to decide if it has too be associated to the class k or ℓ. But if $\boldsymbol{x}$ is *far* from $\mathcal{C}_k$ and $\mathcal{C}_\ell$, it comes the disjunction. Then, disjunction mass may be related to:

$$m(\boldsymbol{x} \in \theta_k \cup \theta_\ell) \sim 1 - \tanh(\beta h) \tag{7}$$

where $h = \frac{d_{\mathbb{R}^p}(\mathcal{C}_k, \mathcal{C}_\ell)}{d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_k) + d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_\ell)}, 0 < k, \ell \leqslant K, k \neq \ell$. Here, the β parameter stands for the level of ambiguity. When $\boldsymbol{x}$ is close, in $\mathbb{R}^p$, to the segment $[\mathcal{C}_k, \mathcal{C}_\ell]$, $d(\mathcal{C}_k, \mathcal{C}_\ell) \simeq d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_k) + d_{\mathbb{R}^p}(\boldsymbol{x}, \mathcal{C}_\ell)$ so that z is close to 1, and $m(\boldsymbol{x} \in \theta_k \cup \theta_\ell)$ has to vanish. The more the β, the less the ambiguous mass.

Conjunction and disjunction for more than 2 classes: This construction that takes into consideration the ratio of distance between 2 classes or the distance to the middle of 2 classes can be extended to more than 2 classes. For instance, eq. (6) can be based on the centroid of more than 2 class. Eq. (7) can be generalized by the composition of one against one class from a set of K classes, divided by the sum of distance of $\boldsymbol{x}$ to each of the K class centers. Nevertheless, this part has not been deeper investigated since those compositions should not have significative impact on the fusion or the classification results.

Final mass belief function: The complete BBA has to respect eq. (1) constraint so that is it necessary to apply a normalization step to the unnormalized BBA obtained by separately calculates the belief masses on simple and compound hypotheses.

6 Experiments on Benchmark Dataset

In order to highlight some advantages and possible drawbacks of the proposed SOM-based BBA, the performance of the SOM-based BBA is compared to EV-CLUS [11] and ECM [6] ones by using dataset provided by the University of California - Irvine (UCI) Machine Learning Repository[4]. Seven numerical data

[4]The dataset is available at http://archive.ics.uci.edu/ml

sets out of 270 have been taken into consideration with various amounts of features (that correspond to the feature space dimension $\mathbb{R}^p$) and number of classes (from 2 to 7) as detailed in Table 1.

Kohonen's map has been trained with the following parameters: a size of 20×20 neurons, trained with 200 iterations. An initial neighborhood size $\mathcal{N}_{\boldsymbol{w}}(t_0)$ of 10 neurons and a learning rate $\alpha(t_0)$ of 0.9. It appears that the SOM-based BBA yields most of the time the highest classifications results (put in boldface in the results of Table 2). In each row of Table 2, the first line corresponds to the number of correctly classified samples, the second line corresponds to the proportion of samples correctly classified, and the last line shows the computation time. It is worth noting that when ECM performs better, the SOM-based approach is close to the best accuracy (73.52 % versus 74.11 % for the benefit of ECM with the Wine database, and 69.24 % versus 69.62 % with the Statlog Landsat satellite images database). Equivalent results prove that SOM-based BBA is just a simplified (*i.e.* quantized) version of the feature space ECM work with. Better results are due to the fact that distances on the map (in 2D) are more appropriated for complex (or non isotropic) class (in pD). EVCLUS is always below. It seems that the performance ranking between ECM and SOM-based BBA is not depending on the feature space dimension nor the number of classes since the Wine and Statlog Landsat satellite image data bases are very different to each other. Since the SOM-based approach considers a projected feature space of dimension 2, it may induce on those cases a too coarse approximation of the manifold in comparison to ECM. Nevertheless, it is worth noting that the benefit in using a SOM-based approach for BBA is related to the number of samples to be handled. Fig. 3 shows that the more the number of sample the fastest the SOM-based approach in comparison to the ECM while yielding the same level of accuracy. Then the SOM-based approach appears to be a valuable alternative to handle large data set such as real images for classification purpose. In fact, distance in $\mathbb{R}^p$ is more computational demanding than in $\mathbb{R}^2$. Indeed, the form of the class in the SOM is more isotropic, so that no consideration on the shape of the manifold is to be considered. On the contrary, ECM has to care of the standard deviation of the classes to build the mass distribution.

Table 1: Characteristics of the UCI datasets used for comparison.

Dataset	**Features**	**classes**	**samples**
Banknote authentication	4	2	1372
Pima Indians Diabetes	8	2	768
Seeds	7	3	210
Wine	13	3	170
Statlog (Landsat Satellite)	36	6	6435
Statlog (Image Segmentation)	19	7	2130
Synthetic control chart time series	60	6	600

Table 2: Classification results of EVCLUS, ECM with decision by $BetP$ and SOM-based BBA with decision by $BetP_g$.

Dataset	Banknote authentication	Pima Indians Diabetes	Seeds	Wine	Statlog (Landsat Satellite)	Statlog (Image Segmentation)	Synthetic control chart time series
EVCLUS	843 61.44 % 1172.2sec	475 61.84 % 181.7sec	157 74.76 % 34.3sec	103 60.58 % 6.7 sec	3027 47.03 % 5857 sec	895 42.01 % 3657 sec	384 64.0 % 370 sec
ECM	848 61.80 % 3.4sec	506 65.88 % 3.2sec	189 90.0 % 0.3sec	126 **74.11 %** 0.9sec	4480 **69.62 %** 480sec	1282 55.49 % 161sec	453 72.5 % 6.9sec
SOM-based	1090 **79.44 %** 8.6sec	549 **71.48 %** 6.7sec	191 **90.95 %** 5.8sec	125 73.52 % 5.9sec	4456 69.24 % 163sec	1431 **67.18 %** 84sec	501 **83.5 %** 8.0sec

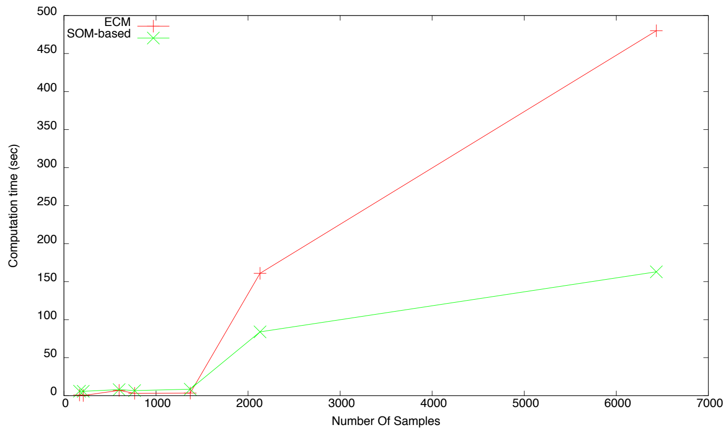

Fig. 3: Computation time depending on the feature space dimension.

7 Conclusion

In this article, a new method for mass function construction through Kohonen's map has been proposed. Our method performs the assignment of belief masses on simple, conjunctive and disjunctive hypotheses. So, unlike the other approaches, it exploits all the conceptual contribution of the theory thanks to its ability to deal with uncertain and paradoxical data through the proposed BBA. Experiments on a set of benchmark database showed that our approach yields better accuracy as stated by the confusion matrices. Moreover, the overall SOM-based BBA algorithm is much less demanding in term of computation so that it is possible to handle large data set.

References

1. Dempster, A.P.: Upper and Lower Probabilities Induced by a Multivalued Mapping. Annals of Mathematical Statistics 38, 325–339 (1967)
2. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, NJ, U.S.A (1976)
3. Smarandache, F., Dezert, J.: Advances and Applications of DSmT for Information Fusion: (Collected works). Volumes 1, 2 & 3. American Research Press, Rehoboth, U.S.A (2004–2009) `http://www.gallup.unm.edu/~smarandache/DSmT.htm`
4. Smets, P.: Belief functions: The Disjunctive Rule of Combination and the Generalized Bayesian Theorem. In: R.R. Yager, L. Liu (eds.) Classic Works of the Dempster-Shafer Theory of Belief Functions, Studies in Fuzziness and Soft Computing, vol. 219, 633–664. Springer (2008)
5. Zouhal, L.M., Denœux, T.: An Evidence-Theoretic K-NN Rule with Parameter Optimization. IEEE Transactions on Systems, Man, and Cybernetics, Part C 28(2), 263–271 (1998)
6. Masson, M.H., Denœux, T.: ECM: An Evidential Version of the Fuzzy C-means Algorithm. Pattern Recogn 41(4), 1384–1397 (2008)
7. Kohonen, T.: The Self-Organizing Map. Proceedings of the IEEE 78(9), 1464–1480 (1990)
8. Kraaijveld, M.A., Mao, J., Jain, A.K.: A Nonlinear Projection Method Based on Kohonens Topology Preserving Maps. IEEE Trans. Neural Networks 6(3), 548–559 (1995)
9. Chang, C.: An Information Theoretic-based Measure for Spectral Similarity and Discriminability. IEEE Trans. on Information Theory 46(5), 1927–1932 (2000)
10. Inglada, J., Mercier, G.: A New Statistical Similarity Measure for Change Detection in Multitemporal SAR Images and its Extension to Multiscale Change Analysis. IEEE Trans. on Geosci. Remote Sensing 45(5), 1432–1446 (2007)
11. Denœux, T., Masson, M.H.: EVCLUS: EVidential CLUSstering of Proximity Data. IEEE Trans. on Systems, Man, and Cybernetics, Part B 34(1), 95–109 (2004)