Constructing rule-based models using the belief functions framework

Rui Jorge Almeida[1], Thierry Denoeux[2], and Uzay Kaymak[3]

[1] Erasmus University Rotterdam, Erasmus School of Economics,
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands.
rjalmeida@ese.eur.nl
[2] Université de Technologie de Compiègne, U.M.R. C.N.R.S. 6599 Heudiasyc
B.P. 20529 F-60205 Compiègne, France.
thierry.denoeux@hds.utc.fr
[3] School of Industrial Engineering, Eindhoven University of Technology,
P.O. Box 513, 5600 MB Eindhoven, The Netherlands.
u.kaymak@ieee.org

Abstract. We study a new approach to regression analysis. We propose a new rule-based regression model using the theoretical framework of belief functions. For this purpose we use the recently proposed Evidential c-means (ECM) to derive rule-based models solely from data. ECM allocates, for each object, a mass of belief to any subsets of possible clusters, which allows to gain a deeper insight on the data while being robust with respect to outliers. The proposed rule-based models convey this added information as the examples illustrate.

1 Introduction

Dempster-Shafer theory of evidence is a theoretical framework for reasoning with partial and unreliable information [1]. It provides a very powerful tool to deal with epistemic uncertainty, by allowing experts to express partial beliefs, such as partial or even total ignorance in a very flexible way, and can be easily extended to deal with objective probabilities. It provides tools to combine several pieces of evidence, such as the conjunctive and the disjunctive rules of combination. Furthermore, using for instance the pignistic transformation, it is possible to solve decision making problems.

Real-world problems can be solved using the Dempster-Shafer theory of evidence by modeling information pieces using a belief function on a specific domain, and manipulating the resulting belief functions using available operations of this framework. The two conventional sources of partial or unreliable knowledge are human experts and observation data. In this work, we consider the latter. Observation data often contain partial and unreliable information. The framework of belief functions is well suited to deal with such data.

Rule-based models are simple, yet powerful tools that can be used for a great variety of problems, such as regression, classification, decision making and control. An example of such a model, defined in the framework of fuzzy sets, is the

Takagi-Sugeno fuzzy model [2]. Usually these types of models translate the domain knowledge and relation between the variables defined for the system model in the form of if-then rules. The if-then rules provide a transparent description of the system, which may reflect the possible nonlinearity of the system. One way of obtaining rule-based models from data is using product-space clustering. A clustering algorithm finds a partition matrix which best explains and represents the unknown structure of the data with respect to the model that defines it [3]. Different clustering algorithms can be used, which will yield different information and insights about the underlying structure of the data.

In this paper we study how to construct, within the framework of belief functions, as understood in the transferable belief model [4], a rule-based model, by means of product-space clustering for regression analysis. We use the Evidential c-means (ECM) algorithm [5] to derive rule-based models solely from data.

Regression analysis is a technique for modeling and analyzing relationships between variables. More specifically, regression analysis helps to ascertain the causal effect of one variable upon another. In classical statistics, it is assumed that the variables are measured in a precise and certain manner. In reality, observation data often contain partial and unreliable information both on the dependent and independent variables. Several approaches have been proposed to deal with different origins of uncertainty in the data, such as fuzzy linear regression [6, 7], fuzzy rule-base models [2], fuzzy rule-base models with a belief structure as output [8], nonparametric belief functions [9] and function regression using neural network with adaptive weights [10]. The approach proposed in this work combines the formalism to handle imprecise and partially conflicting data, given by the belief function theory, with a transparent description of the system in the form of if-then rules.

This paper is organized as follows. Section 2 reviews briefly the main concepts underlying the theory of belief functions, and explains the clustering algorithm used for deriving a credal partition from object data. Section 3 proposes a rule-based model for regression analysis and the identification of the model parameters. An example is shown in Section 4 and finally the conclusions and future work are given in Section 5.

2 Belief Functions

2.1 Basic Concepts

Dempster-Shafer theory of evidence is a theoretical framework for reasoning with partial and unreliable information. In the following, we briefly recall some of the basics of the belief function theory. More details can be found in [1, 4, 11]. In this work, we adopt the subjectivist, nonprobabilistic view of Smet's transferable belief model (TBM) [4, 11]

Let Ω be a finite set of elementary values ω called the frame of discernment. A basic belief assignment (bba)[1] is defined as a function m from 2^{Ω} to $[0, 1]$, satisfying:

$$\sum_{A \subseteq \Omega} m(A) = 1 \,, \tag{1}$$

which represents the partial knowledge regarding the actual value taken by ω. The subsets A of $\Omega = \{\omega_1, \ldots, \omega_c\}$ such that $m(A) > 0$ are the focal sets of m. Each focal set A is a set of possible values for ω and the value $m(A)$ can be interpreted as the part of belief supporting exactly that the actual event belongs to A. Perfect knowledge of the value of ω is represented by the allocation of the whole mass of belief to a unique singleton of Ω and m is called a certain bba. Complete ignorance corresponds to $m(\Omega) = 1$, and is represented by the vacuous belief function [1]. When all focal sets of m are singletons, m is equivalent to a probability function and is called a Bayesian bba.

A bba m such that $m(\emptyset) = 0$ is said to be normal [1]. This condition may be relaxed by assuming that ω might take its value outside Ω, which means that Ω might be incomplete [12]. The quantity $m(\emptyset)$ is then interpreted as a mass of belief given to the hypothesis that ω might not lie in Ω.

The information provided by a bba m can be represented by a belief function $bel : 2^{\omega} \mapsto [0, 1]$, defined as

$$\mathrm{bel}(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \quad \forall A \subseteq \Omega. \tag{2}$$

The quantity $\mathrm{bel}(A)$ represents the amount of support given to A. A bba m can be equivalently represented by a plausibility function $pl : 2^{\Omega} \mapsto [0, 1]$, defined as

$$\mathrm{pl}(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad \forall A, B \subseteq \Omega. \tag{3}$$

The plausibility $\mathrm{pl}(A)$ represents the potential amount of support given to A.

Given two bba's m_1 and m_2 defined over the same frame of discernment Ω and induced by two distinct pieces of evidence, we can combine them using a binary set operation $\triangledown$, which can be defined as [11]:

$$m_1 \textcircled{\triangledown} m_2(A) = \sum_{B \triangledown C = A} m_1(B) m_2(C), \quad \forall A \in \Omega. \tag{4}$$

The conjunctive and disjunctive rule can be obtained by choosing $\triangledown = \cap$, and $\triangledown = \cup$, respectively. For the case of the conjunctive rule $\textcircled{\cap}$, the normality condition $m(\emptyset) = 0$ may be recovered, by using the Dempster normalization procedure, by converting the subnormal BBA $(m_1 \textcircled{\cap} m_2)$ into a normal one $(m_1 \oplus m_2)^*(A)$, defined as follows:

$$(m_1 \oplus m_2)^*(A) = \frac{(m_1 \textcircled{\cap} m_2)(A)}{1 - (m_1 \textcircled{\cap} m_2)(\emptyset)}, \quad \forall A \neq \emptyset, (m_1 \oplus m_2)^*(\emptyset) = 0. \tag{5}$$

The Dempster's rule of combination [1], noted as $\oplus$ corresponds to the conjunctive sum combined by the Dempster's normalization. The choice of the combination rule depends on the reliability of the two sources. The conjunctive rule should be used when all sources of information are fully reliable and distinct. Otherwise, if there are doubts over the reliability of at least one sources then the disjunctive rule of combination should be chosen.

The decision making problem regarding the selection of one single hypothesis in Ω, is solved in the transferable belief model framework, by using a pignistic probability, BetP, defined, for a normal bba, by [11]:

$$\mathrm{BetP}(\omega) \triangleq \sum_{\omega \in A} \frac{m(A)}{|A|} \quad \forall \omega \in \Omega, \tag{6}$$

where $|A|$ denotes the cardinality of $A \subseteq \Omega$. It is shown, that this is the only transformation between belief function and a probability function satisfying elementary rationality requirements, in which each mass of belief $m(A)$ is equally distributed among the elements of A [13].

2.2 Evidential c-Means Algorithm

In [5], the Evidential c-Means (ECM) algorithm was proposed to derive a credal partition from object data. In this algorithm the partial knowledge regarding the class membership of an object i is represented by a bba m_i on the set Ω. This representation makes it possible to model all situations ranging from complete ignorance to full certainty concerning the class label of the object. This idea was also applied to proximity data [14].

A credal partition is defined as the n-tuple $M = (m_1, m_2, \ldots, m_n)$. It can be seen as a general model of partitioning, where:

- when each m_i is a certain bba, then M defines a conventional, crisp partition of the set of objects; this corresponds to a situation of complete knowledge;
- when each m_i is a Bayesian bba, then M specifies a fuzzy partition, as defined by Bezdek [15];

Determining a credal partition M from object data, using ECM, implies determining, for each object i, the quantities $m_{ij} = m_i(A_j),\ A_j \neq \emptyset, A_j \subseteq \Omega)$ in such a way that the mass of belief m_{ij} is low (high) when the distance d_{ij} between i and the focal set A_j is high (low). The distance between an object and any non empty subset of Ω is defined by associating to each subset A_j of Ω the barycenter $\bar{\mathbf{v}}$ of the centers associated to the classes composing A_j. Each cluster ω_k is represented by a center $\mathbf{v}_k \in \mathbb{R}^p$. Specifically,

$$s_{kj} = \begin{cases} 1, \text{ if } \omega_k \in A_j \\ 0 \ \text{ otherwise} \end{cases} . \tag{7}$$

The barycenter $\bar{\mathbf{v}}_j$ associated to A_j is:

$$\mathbf{v}_j = \frac{1}{|A_j|} \sum_{k=1}^{c} s_{kj} \mathbf{v}_k \ , \tag{8}$$

The distance d_{ij} is then defined as $d_{ij}^2 \triangleq ||\mathbf{x}_i - \bar{\mathbf{v}}_j||$. The proposed objective function for ECM, used to derive the credal partition M and the matrix V containing the cluster centers, is given by:

$$J_{ECM}(M,V,A) = \sum_{i=1}^{n} \sum_{\{j/A_j \subseteq \Omega, A_j \neq \emptyset\}} \tau_j^{\alpha} m_{ij}^{\beta} d_{ij}^2 + \sum_{i=1}^{n} \delta^2 m_{i\emptyset}^{\beta}, \tag{9}$$

subject to

$$\sum_{\{j/A_j \subseteq \Omega, A_j \neq \emptyset\}} m_{ij} + m_{i\emptyset} = 1, \; \forall i = 1, \ldots, n, \tag{10}$$

where $\beta > 1$ is a weighting exponent that controls the fuzziness of the partition, δ controls the amount of data considered as outliers and $m_{i\emptyset}$ denotes $m_i(\emptyset)$,the amount of evidence that the class of object i does not lie in Ω. The weighting coefficient τ_j^{α} was introduced to penalize the subsets in Ω of high cardinality and the exponent α allows to control the degree of penalization. The second term of (10) is used to give a separate treatment term for the empty set. This focal element is in fact associated to a noise cluster, which allows to detect atypical data. The minimization of (10) can be done using the Lagrangian method, with the following update equation for the credal partition:

$$m_{ij} = \frac{c_j^{-\alpha/(\beta-1)} d_{ij}^{-2/(\beta-1)}}{\sum_{A_j \neq \emptyset} c_j^{-\alpha/(\beta-1)} d_{ij}^{-2/(\beta-1)} + \delta^{-2/(\beta-1)}}, \; \forall i = 1, \ldots, n, \; \forall j/A_j \subseteq \Omega, A_j \neq \emptyset, \tag{11}$$

and

$$m_{i\emptyset} = 1 - \sum_{A_j \subseteq \Omega, A_j \neq \emptyset} m_{ij}, \; \forall i = 1, \ldots, n. \tag{12}$$

The credal partition provides different structures, which can give different types of information about the data. A possibilistic partition could be obtained by computing from each bba m_i the plausibilities $\mathrm{pl}_i(\{\omega_k\})$ of the different clusters, using (3). The value $\mathrm{pl}_i(\{\omega_k\})$ represents the plausibility that object i belongs to cluster k. In the same way, a probabilistic fuzzy partition may be obtained by calculating the pignistic probability $\mathrm{BetP}_i(\{\omega_k\})$, using (6) induced by each bba m_i. Furthermore, other approximations such as a hard credal partition and lower and upper approximations of each cluster can be retrieved from the credal partition [5]. The information obtained from the credal partition and its approximations can be considered intuitive and simple to interpret. In this work, we try to incorporate the additional degrees of freedom and information obtained from the credal partition, in the rule-based classification systems.

3 Rule-Based Model

3.1 Regression Problem

Supervised learning is concerned with the prediction of an quantitative measure of the output variable y, based on a vector $\mathbf{x} = (x_1, \ldots, x_p)$ of n observed input variables. Let $\mathbf{x}$ be an arbitrary vector, and y the corresponding unknown output. In classical regression literature, the objective is to determine the best mathematical expression describing the functional relationship between one response and one or more independent variables. Following the nomenclature used, the problem is to obtain some information on y from the training set $\mathcal{L} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ of n observations of the input and output variables. Classically, it is assumed that the variables are measured in a precise and certain manner. In reality, observation data often contain partial and unreliable information both on the dependent and independent variables. For this case, it necessary to use a formalism to handle such imprecise and partially conflicting data, such as the belief function framework.

3.2 Model Structure

Given a data set with n data samples, given by $X = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$, $\mathbf{y} = \{y_1, y_2, \ldots, y_n\}$ where each data sample has a dimension of p ($n \gg p$), following a structure similar to a Takagi-Sugeno fuzzy model [2], the objective is to obtain, directly from the data, rule-based models according to

$$R_j : \text{ If } x_1 \text{ is } M_{j1} \text{and } x_2 \text{ is } M_{j2} \text{ and } \ldots \text{ and } x_p \text{ is } M_{jp} \text{ then } y_j = f_j(\mathbf{x}), \quad (13)$$

where R_j denotes the j-th rule, $j = 1, 2, \ldots, K$ is the number of rules, $\mathbf{x} \in \mathbb{R}^p$ is the antecedent variable, M_j is the (multidimensional) antecedent set M_{jq} of the j-th rule for $q = 1, \ldots, p$. Each rule j has a different function f_j yielding a different value y_j for the output. This type of system consists of a set of if-then rules combined with an inference mechanism and logical connectives to establish relations between the variables defined for the model of the system. This type of model can be identified, by product space clustering. A possibility for the output function, is to use an affine function for the output function as:

$$y_j = \mathbf{a}_j^T \mathbf{x} + b_j.$$

The sets M_j are ill-known regions of $\mathbb{R}^p$. For a given $\mathbf{x}$ it is not possible to define exactly to which region M_j belongs, but instead, it is possible to compute a mass function based on the credal partition M. From this mass function it is possible to compute the pignistic expectation of y given by:

$$y_{bet}(\mathbf{x}) = \sum_{A_j \subseteq \Omega, A_j \neq \emptyset} m^{\Omega}(A_j) \frac{1}{|A_j|} y_j \quad (14)$$

where

$$m^{\Omega}(A_j) = \frac{\beta_j(\mathbf{x})}{\sum_{B_j \subseteq \Theta, B_j \neq \emptyset} \beta_j(\mathbf{x})}, \quad (15)$$

$\beta_j(\mathbf{x})$ is the degree of activation, and $|A_j|$ is the cardinality of $A_j \subseteq \Omega$.

3.3 Model Parameters

To form the rule-base model from the available learning set $\mathcal{L}$, the structure of the model is first determined and afterward the parameters of the structure are identified. Clustering is applied to the learning set $\mathcal{L}$, using ECM. The number of rules characterizes the structure of a rule-based system and in our case corresponds to the number of non-empty subsets of partitions obtained from the clustering algorithm.

In this work, we use ECM to partition the space using the framework of belief function. The Evidential c-Means algorithm was proposed to derive a credal partition from object data. In this algorithm the partial knowledge regarding the observation data of an object is represented by a basic belief assignment on the finite set of frame of discernment. This representation makes it possible to model all situations ranging from complete ignorance to full certainty. Using the credal partition it is possible to highlight the points that unambiguously belong to one cluster, and the points that lie at the boundary of two or more clusters. For this research we try to incorporate the added degrees of freedom and information obtained from the credal partition, in the rule-based systems. This type of model will provide a rich description of the data and its underlying structure, while making it robust to partial and unreliable data.

Antecedent Belief Functions The antecedent functions can be obtained by projecting the credal partition onto the antecedent variables. The principle of generating antecedent functions by projection is to project the multidimensional sets defined point wise in the rows of the credal partition matrix $M = (m_1, \ldots, m_n)$ onto the individual antecedent variables of the rules. This method projects the credal partition matrix onto the axes of the antecedent variables x_q, $1 \leq q \leq p$.

In order to obtain the mass of belief functions for the antecedent sets M_{jq}, the multidimensional set defined pointwise in the j-th row of the partition matrix M are projected onto the axes of the antecedent variables x_q, by:

$$m_{M_{jq}}(x_{qi}) = \mathrm{proj}_q(m_{ij}). \tag{16}$$

where m_{ij} is given by (11), defined on frame of discernment Ω. In order to obtain a model, the point-wise defined sets M_{jq} can be approximated by appropriate parametric functions. Several types of functions such as triangular, gaussian or trapezoidal, can be used. In this work we choose a combination of gaussian functions of the form

$$M_{jq} \approx f(x_q; \sigma_{1jq}, c_{1jq}, \sigma_{2jq}, c_{2jq}) = e^{\left(\frac{-(x_q - c_{1jq})^2}{2\sigma_{1jq}^2} + \frac{-(x_q - c_{2jq})^2}{2\sigma_{2jq}^2}\right)}. \tag{17}$$

When computing the degree of activation $\beta_j(\mathbf{x})$ of the j-th rule, the original cluster in the antecedent product space is reconstructed by applying the intersection operator in the cartesian product space of the antecedent variables:

$$\beta_j(\mathbf{x}) = m_{Mj1}(x_1) \wedge m_{Mj2}(x_2) \wedge \ldots \wedge m_{Mjp}(x_p), \tag{18}$$

where $\wedge$ denotes a t-norm. Suitable possible t-norms are the product or the minimum operator. In this work, we consider the p features to be uncorrelated, hence we use the product operator. Other possibilities [16] include combination operators which generalize Dempster rule and the cautious rule [17], based on a generalized discounting process or alternatively a parameterized family of t-norms containing both the product and the minimum as special cases, such as Frank's family of t-norms [18]

Consequents Parameters The consequent parameters for each rule can be estimated by the least-squares method. A set of optimal parameters with respect to the model output can be estimated from the identification data set by ordinary least-squares methods. This approach is formulated as a minimization of the total prediction error of the model. Let $V_j^T = [a_j^T, b_j]$ be the vector of consequent parameters, let X_e denote the regressor matrix $[X; 1]$ and let W_j denote a diagonal matrix with the normalized degree of activation $\beta_j(\mathbf{x})$ in its i-th diagonal element. Denote $X' = [W_1 X_e, \ldots, W_K X_e]$. Assuming that the columns of X_e are linearly independent, the resulting solution of the least-squares problem $y = X'V + \varepsilon$ becomes

$$V = [X'^T X']^{-1} X'^T \mathbf{y}. \tag{19}$$

The determination of the consequent parameters concludes the identification of the rule-based system.

4 Examples

In this section, two examples are presented to verify the validity of the proposed strategy. One is a univariate function, while the other is the gas furnace data of Box and Jenkins [19]. To assess model performance, the mean squares error (MSE) will be used. We compare the results obtained using belief rule–base models proposed in this paper to Takagi–Sugeno fuzzy rule–base models and also to the results presented in [8]. The Takagi–Sugeno fuzzy rule–base models antecedent membership functions are obtained using Fuzzy C-Means [15] while the consequent parameters are obtained using by least-square estimation.

For the first example, let us consider the approximation of a nonlinear static univariate function:

$$y(x) = 3 \exp^{-x^2} \sin(\pi x) + \eta \tag{20}$$

where η is a normal distribution with zero mean and variance 0.15 and x is an input random sample vector of size $n = 30$, uniformly distributed in the domain $[-3, 3]$.

The Box and Jenkins gas furnace data is a well known and frequently used benchmark data for modeling and identification. The data consist of 296 measurements. The input $u(t)$ is the gas flow rate and the output is CO_2 concentration in outlet gas $y(t)$. A possible way of modeling this process is to consider that the output $y(t)$ is a function of the input variables $x = [y(t-1), u(t-4)]^T$.

For both examples a belief rule–base model was derived with $c = 3$ clusters. The resulting model will have $K = 2^c - 1$ rules, as we do not consider a rule for the empty–set. This model was compared to a FCM T–S fuzzy model with $c = 3, 7$ clusters, resulting in a model with 3 or 7 rules. The results for this model can be found in Table 1.

Table 1. MSE for the univariate static function and the gas furnace example.

Model	Univariate	Gas furnace
Ramdani [8]	0.018	0.045
FCM $c = 3$	0.432	0.070
FCM $c = 7$	0.013	0.064
ECM $c = 3$	0.017	0.063

Table 1 shows that the MSE obtained with the belief rule–base model proposed in this paper are in line with previous studies. Notice that both FCM and ECM use Euclidean distance as the inner product norm. A more suitable choice for modeling this type of systems is to employ an adaptive distance norm, such as the Mahalanobis distance, to detect clusters of different shape and orientation [20]. This explains the poor results in the univariate case for the fuzzy–rule base models with $c = 3$ clusters. Notice that in the case of the belief rule–base models, a rule is derived for each possible subset of clusters. Thus local models are identified for objects which are clearly identified as belonging to one cluster, but also to objects in overlapping cluster regions. This is an advantage of the proposed method as it helps to improve the results using a low number of clusters. Adding more clusters may increase the number of overlapping cluster regions and consequently the number of rules. This may result in a system which will overfit the data. Furthermore, we note that the proposed model is developed in an automated manner solely from data. This model combines the capability to handle imprecise and partially conflicting data, given by the belief function theory, with a transparent description of the system in the form of if-then rules.

5 Conclusions and Future Work

This paper proposes the use of the credal partition obtained from the Evidential C-Means based on the theoretical framework of belief functions, to derive rule-based models. This type of model provides a rich description of the data and its underlying structure, which can be successfully translated into rules, while making it robust to partial and unreliable data. Future research will focus on assessing properties and characteristics of the proposed model.

Acknowledgments This work was supported by the European Science Foundation through COST Action IC0702, under the grant ECOST-STSM-IC0702-180411-006041.

References

1. Shafer, G.: A mathematical theory of evidence. Princeton university press (1976)
2. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its aplication to modeling and control. IEEE Transactions on Systems, Man and Cybernetics **15** (1985) 116–132
3. Hartigan, J.A.: Clustering Algorithms. Wiley-Interscience, New York, London, Sydney, Toronto (1974)
4. Smets, P., Kennes, R.: The transferable belief model. Artif. Intell. **66** (1994) 191–234
5. Masson, M.H., Denoeux, T.: ECM: An evidential version of the fuzzy c-means algorithm. Pattern Recognition **41** (2008) 1384–1397
6. Tanaka, H., Uejima, S., Asai, K.: Linear regression analysis with fuzzy model. Systems, Man and Cybernetics, IEEE Transactions on **12** (1982) 903–907
7. Chang, Y.H.O., Ayyub, B.M.: Fuzzy regression methods - a comparative assessment. Fuzzy Sets and Systems **119** (2001) 187–03
8. Ramdani, M., Mourot, G., Ragot, J.: A multi-modeling strategy based on belief function theory. In: Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05. 44th IEEE Conference on. (2005) 3261–3266
9. Petit-Renaud, S., Denoeux, T.: Nonparametric regression analysis of uncertain and imprecise data using belief functions. International Journal of Approximate Reasoning **35** (2004) 1–28
10. Denoeux, T.: Function approximation in the framework of evidence theory: A connectionist approach. In: In Proceedings of the 1997 International Conference on Neural Networks, Houston, USA (1997) pages
11. Smets, P.: The transferable belief model for quantified belief representation. In Gabbay, D., Smets, P., eds.: Handbook of Defeasible Reasoning and Uncertainty Management Systems. Kluwer, Doordrecht (1998)
12. Smets, P.: The combination of evidence in the transferable belief model. IEEE Trans. Pattern Anal. Mach. Intell. **12** (1990) 447–458
13. Smets, P.: Decision making in the TBM: the necessity of the pignistic transformation. Int. J. Approx. Reasoning **38** (2005) 133–147
14. Masson, M.H., Denux, T.: RECM: Relational evidential c-means algorithm. Pattern Recogn. Lett. **30** (2009) 1015–1026
15. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
16. Quost, B., Masson, M.H., Denoeux, T.: Classifier fusion in the dempster-shafer framework using optimized t-norm based combination rules. International Journal of Approximate Reasoning **52** (2011) 353–374 Dependence Issues in Knowledge-Based Systems.
17. Kallel, A., Hgarat-Mascle, S.L.: Combination of partially non-distinct beliefs: The cautious-adaptive rule. International Journal of Approximate Reasoning **50** (2009) 1000–1021 Special Section on Graphical Models and Information Retrieval.
18. Klement, E., Mesiar, R., Pap, E.: Triangular Norms. Kluwer Academic Publishers, Dordrecht (2000)
19. Box, G.E.P., Jenkins, G.M.: Time Series Analysis, Forecasting, and Control. Holden-Day, San Francisco (1976)
20. Gustfson, D., Kessel, W.: Fuzzy clustering with a fuzzy covariance matrix. In: Proceedings of IEEE Conference on Decision and Control, CDC, San Diego, CA, USA (1979) 761–766