EINDHOVEN UNIVERSITY OF TECHNOLOGY

TS-models from evidential clustering

Citation for published version (APA):
Almeida, R. J., & Kaymak, U. (2010). TS-models from evidential clustering. In E. Hüllermeier, R. Kruse, & F. Hoffmann (Eds.), *Information Processing and Management of Uncertainty in Knowledge-Based Systems : Theory and Methods (13th International Conference, IPMU 2010 Dortmund, Germany, June 28- July 2, 2010, Proceedings, Part I)* (pp. 228-237). (Communications in Computer and Information Science; Vol. 80). Springer. https://doi.org/10.1007/978-3-642-14055-6_24

DOI:
10.1007/978-3-642-14055-6_24

Document status and date:
Published: 01/01/2010

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Download date: 17. Nov. 2023

TS-Models from Evidential Clustering

Rui Jorge Almeida[1] and Uzay Kaymak[1,2]

[1] Erasmus University Rotterdam, Erasmus School of Economics, Rotterdam, The Netherlands
[2] Eindhoven University of Technology, School of Industrial Engineering, Eindhoven, The Netherlands
rjalmeida@ese.eur.nl, u.kaymak@ieee.org

Abstract. We study how to derive a fuzzy rule-based classification model using the theoretical framework of belief functions. For this purpose we use the recently proposed Evidential c-means (ECM) to derive Takagi-Sugeno (TS) 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 in the data while being robust with respect to outliers. Some classification examples are discussed, which show the advantages and disadvantages of the proposed algorithm.

1 Introduction

Classification problem is an important subject for a variety of fields, including pattern recognition, artificial intelligence, credit risk and direct marketing. In a classification problem the aim is to assign class labels to a set of data instances described by multiple features. A possible method to solve classification problems is to use a fuzzy rule based model, built from data [1,2,3]. Fuzzy models use if-then rules and logical connectives to establish relations between the variables defined for the model of the system. The fuzzy sets in the rules serve as an interface amongst qualitative conceptualization in the model, and the numerical input and output variables. The if-then rules provide a transparent description of the system, that may reflect a possible nonlinearity of the system. The rule-based nature of the model allows for a linguistic description of the knowledge.

One way of obtaining Takagi-Sugeno fuzzy models is product-space fuzzy 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 [4]. Different clustering algorithms can be used, which will yield different information and insights about the underlying structure of the data.

Uncertainty in the data is a challenge for classification [5]. Several approaches have been proposed to deal with this problem, using the framework of belief functions. In [6,7,8], it was proposed to represent the partial knowledge regarding the class membership of an object using a basic belief assignment. A classification method based on the decision tree approach that takes into consideration the uncertainty characterized by the classes of the training examples, as well as the uncertainty of their attribute values was proposed in [9].

E. Hüllermeier, R. Kruse, and F. Hoffmann (Eds.): IPMU 2010, Part I, CCIS 80, pp. 228–237, 2010.
© Springer-Verlag Berlin Heidelberg 2010

In this paper we study the use of Evidential C-Means (ECM) for system identification. For this it is necessary to relate information represented in the framework of the theory of beliefs, as understood in the transferable belief model [10], and fuzzy sets. Previous studies have shown that consonant beliefs and fuzzy sets are related [11]. In our approach, we consider that the obtained credal partition captures relevant information for correct interpretation of data substructure, and we discuss the possibility to map the obtained credal partition to a fuzzy set, providing linguistic interpretation and labels to the obtained structure. 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. We try to convene this added information into the rule based classification system.

The paper is organized as follows. Section 2 reviews briefly the main concepts underlying the theory of belief functions. Section 3 presents its use for deriving a credal partition from object data. Section 4 presents the method used for classification and system identification in this work. The experimental setup and the results are presented in Section 5 and Section 6, respectively. A brief discussion of the results is in Section 7. Finally the conclusions are given in Section 8.

2 Belief Functions

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 [12,10,13].

Let Ω be a finite set of elementary values ω called the frame of discernment. The basic belief assignment (bba)[12] 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 Ω 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.

A bba m such that $m(\emptyset) = 0$ is said to be normal [12]. This condition may be relaxed by assuming that ω might take its value outside Ω, which means that Ω might be incomplete [14]. The quantity $m(\emptyset)$ is then interpreted as a mass of belief given to the hypothesis that ω might not lie in Ω. A bba m can be equivalently represented by a plausibility function $pl : 2^{\omega} \mapsto [0, 1]$, defined as

$$\mathrm{pl}(A) \triangleq \sum_{B \cap A \neq \emptyset} m(B) \quad \forall A, B \subseteq \Omega \ . \tag{2}$$

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

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 [13]:

$$\mathrm{BetP}(\omega) \triangleq \sum_{\omega \in A} \frac{m(A)}{|A|} \quad \forall \omega \in \Omega \ , \tag{3}$$

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 [15].

3 Evidential c-Means

In [6], 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 relational data in [8] and proximity data [7].

Determining a credal partition $M = (m_1, m_2, \ldots, m_n)$ 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 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. It is assumed that each class ω_k is represented by a center $\mathbf{v}_k \in \mathbb{R}^p$. Specifically,

$$s_{kj} = \begin{cases} 1, \text{ if } \omega_k \in A_j \\ 0 \ \text{ otherwise} \end{cases} . \tag{4}$$

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

$$\bar{v}_j = \frac{1}{\tau_j} \sum_{k=1}^{c} s_{kj} \mathbf{v}_k \ , \tag{5}$$

where $\tau_j = |A_j|$ denotes the cardinality of A_j. 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:

$$J_{ECM}(M, V, A) = \sum_{i=1}^{n} \sum_{\{j/A_j \subseteq \Omega, A_j \neq \emptyset\}} \tau_j^{\alpha} m_{ij}^{\beta} d_{ij}^2 + \sum_{i=1}^{n} \delta^2 m_{i\emptyset}^{\beta} \ , \tag{6}$$

subject to

$$\sum_{\{j/A_j \subseteq \Omega, A_j \neq \emptyset\}} m_{ij} + m_{i\emptyset} = 1 \ \forall i = 1, n \ , \tag{7}$$

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 aims at penalizing the subsets in Ω of high cardinality and the exponent α allows to control the degree of penalization. The second term of (7) 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 (7), can be done using the Lagrangian method.

The credal partition provides different structures, that can give different types of information about the data. A possibilistic partition could be obtained by computing from each bba m_i the plausibilities $\text{pl}_i(\{w_k\})$ of the different clusters, using (2). The value $\text{pl}_i(\{w_k\})$ 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 $\text{BetP}_i(\{w_k\})$, using (3) induced by each bba m_i.

Furthermore, a hard credal partition can be obtained, by assigning each object to the set of clusters with the highest mass. This allows to divide the partition space into a maximum of 2^c groups. Formally, the $X(A_j)$ for $j = 1, \ldots, 2^c$ defines a hard credal partition of the n objects [6]:

$$X(A_j) = \{i / m_i(A_j) = \max_k m_i(A_k)\} . \tag{8}$$

Finally, it is possible to characterize each cluster ω_k by a set of objects which can be classified in ω_k without any ambiguity and the set of objects which could possibly be assigned to ω_k. These two sets ω_k^L and ω_k^U, are defined as the lower and upper approximations of ω_k respectively [6], and they are defined as:

$$\omega_k^L = X\{\omega_k\}, \qquad \text{and} \qquad \omega_k^U = \underset{j/\omega_k \in A_j}{U} X(A_j) . \tag{9}$$

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 added degrees of freedom and information obtained from the credal partition, in the rule based classification systems.

4 Rule Based Classification

4.1 Model Structure

A fuzzy classification system consists of a set of fuzzy IF-THEN rules combined with a fuzzy inference mechanism. This type of rules can be viewed as an extension of the Takagi-Sugeno fuzzy model [16], and it can be described by N rules of the following type [17]:

$$R_k^q : \text{ If } x_1 \text{ is } F_{k1} \text{ and } \ldots \text{ and } x_n \text{ is } F_{kn} \text{then } d_q(\mathbf{x}) = g_k^q(\mathbf{x}) , \tag{10}$$

where $g_k, k = 1, 2, \ldots, N$ is the consequent function for rule R_k, and $d_q(\mathbf{x})$ is a discriminant function associated with each class $\omega_q, q = 1, \ldots, Q$. Note that the index q indicates that the rule is associated with the output of class q. The output of the classifier assigns the class label corresponding to the maximum value of the discriminant functions. The antecedent parts of the rules are the same for different discriminants, but the consequents may be different. The output of each discriminant function $d_q(\mathbf{x})$ can be interpreted as a score for the associated class q given the input feature vector $\mathbf{x}$. The degree of fulfillment $\beta_i(x)$ of the ith rule, is computed using the intersection operator in the cartesian product space of the antecedent variables as $\beta_i(x) = \mu_{F_{i1}}(x_1) \wedge \mu_{F_{i2}}(x_2) \wedge \ldots \wedge \mu_{F_{ip}}(x_p)$. Other t-norms, such as the product, can be used instead of the minimum operator.

4.2 Model Parameters

To form the fuzzy system model from the data set with n data samples, given by $X = [x_1, x_2, \ldots, x_n]^T$, $Y = [y_1, y_2, \ldots, y_n]^T$ where each data sample has a dimension of p ($n \gg p$), the structure is first determined and afterwards the parameters of the structure are identified. The number of rules characterizes the structure of a fuzzy system and in our case corresponds to the number of partitions obtained from the clustering algorithm.

In this work, we use ECM to partition the space using the framework of belief function and map the obtained credal partition as a fuzzy set. When clustering the object data with ECM, several clustering structures can be obtained, as explained in Section 3. In this work we focus on the partitioning structure obtained from the credal partition, which we expect to better describe the data and its underlying structure.

Using c clusters, the credal partition obtained from ECM partitions the space in at most 2^c intervals, with a center associated with each interval. In contrast, using the well-known FCM [18], the space is partition in, at most, c intervals. The added information from the credal partition allows to reveal objects that unambiguously belong to a given cluster and the set of objects that lie at the boundaries of each cluster. Since the values of the credal partition m are in $[0, 1]^p$, this value can be perceived as an assignment to each subset of the partition. Thus we can obtain the following mapping $\varphi : m \in [0, 1]^p \mapsto A_{ij}$, using one the following functions:

$$\varphi_1 : \mu_{F_{ij}}(x_{jk}) = \text{proj}_j(m_{ik}) \tag{11a}$$

$$\varphi_2 : \mu_{F_{ij}}(x_{jk}) = X(Aj) \tag{11b}$$

$$\varphi_3 : \text{supp}(F) = \omega_k^L,\ \text{core}(F) = \omega_k^U \tag{11c}$$

where $\text{supp}(F) = \{x \in X | \mu_F(x) > 0\}$, $\text{core}(F) = \{x \in X | \mu_F(x) = 1\}$ and proj_j is a pointwise projection of the partition matrix M onto the axes of the antecedent variables x_j.

The obtained point-wise fuzzy sets F_{ij} can now be approximated by appropriate parametric functions, such as Gaussian functions, resulting in the antecedent membership functions. Although, the obtained antecedent membership functions from each one of the mappings presented in (11) will be different from one another, there are situations where they can be very similar. For example, the mappings (11b) and (11c) will be equivalent for cases where the obtained credal partition does not indicate that some points lie in the boundaries of two or more clusters. The number of obtained rules using the proposed method is, at most $2^c - 1$, as we exclude, $m(\emptyset)$. The consequent parameters for each rule are obtained by means of linear least square estimation, which concludes the identification of the classification system.

5 Experimental Setup

We used several data sets to test the proposed modelling approach. The Wisconsin Breast Cancer (WBC) database [19] is composed of 699 objects, 10 features and has missing values and an uneven distribution of classes [20]. Also, two databases related with bankruptcy prediction were used. The Altman data set [21] has 66 objects and 5 features, and the CR data set [22] which contains extreme values, missing values and a very skewed class distribution. The CR data set is composed of 1817 companies and has 51 features.

In our experiments we used minimum amount of improvement $\varepsilon = 0.0001$, maximum number of iterations 100, $\beta = 3$ for both ECM and FCM, $\alpha = 1$ and $\delta = 2000$. The initialization for the cluster prototypes centers in ECM were obtained with FCM, as suggested in [6]. Since all cases are binary classification, two clusters were used for all models. In [6] an index for choosing the number of clusters to be used with ECM is discussed. All trials terminated with the convergence criteria after a few iterations.

In all tests we used a simple holdout method for validation. 50 trials were made and only the results obtained with the testing set are reported below. The overall performance of the models is measured by the classification accuracy.

As already stated, the Wisconsin Breast Cancer and the CR database have missing values. It was considered that these missing values are missing completely at random (MCAR) [23], and thus imputation of values is the usual course of action. The missing values were inferred using the expectation-maximization (EM) algorithm, as explained in [22].

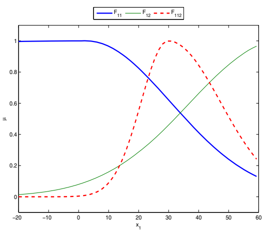
6 Examples

Table 1 exhibits the obtained mean accuracy and respective variance, for each class and the global accuracy of different fuzzy models, for all the data sets under study. These models were obtained using the three methods explained in (11) for mapping the antecedent fuzzy sets. Also, for comparison purposes, the results obtained for a fuzzy rule-based model that uses the FCM to derive the antecedent fuzzy sets is shown.

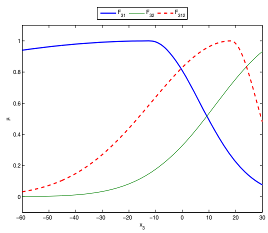
The obtained global accuracy of the model obtained using ECM, is comparable to the results obtained using FCM algorithm. Depending on the type of mapping used, different accuracies for each class are obtained. For the case of the CR database, the mapping φ_1 and φ_2 give better results in terms of the class Bankrupt, which is arguably, the more relevant class in study. In general,

Table 1. Classification accuracy for the databases in study- mean (variance)

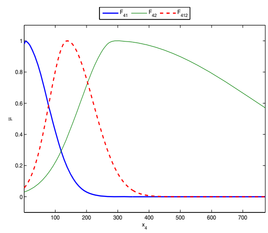
Database	Method	φ_1	φ_2	φ_3	FCM
WBC	Class 0	0.978 (0.016)	0.971 (0.014)	0.976 (0.014)	0.972 (0.018)
	Class 1	0.936 (0.032)	0.961 (0.023)	0.946 (0.026)	0.956 (0.027)
	Global	0.964 (0.016)	0.968 (0.013)	0.966 (0.013)	0.966 (0.013)
Altman	Bankrupt	0.960 (0.067)	0.990 (0.032)	1.000 (0.000)	0.980 (0.042)
	Not Bankrupt	1.000 (0.189)	0.730 (0.048)	0.620 (0.079)	0.600 (0.094)
	Global	0.980 (0.111)	0.860 (0.021)	0.810 (0.039)	0.790 (0.046)
CR	Bankrupt	0.465 (0.068)	0.478 (0.105)	0.326 (0.132)	0.361 (0.082)
	Not Bankrupt	0.988 (0.008)	0.985 (0.003)	0.989 (0.005)	0.989 (0.005)
	Global	0.966 (0.009)	0.964 (0.006)	0.961 (0.006)	0.962 (0.005)



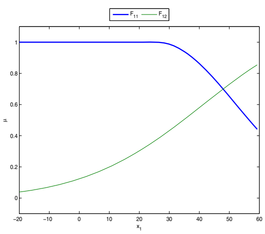
(a) ECM-feature 1



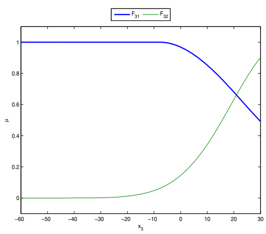
(b) ECM-feature 3



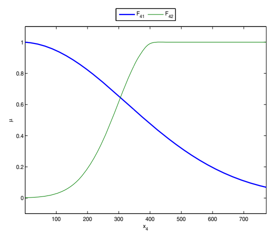
(c) ECM-feature 4



(d) FCM-feature 1



(e) FCM-feature 3



(f) FCM-feature 4

Fig. 1. Antecedent Membership Functions for ECM with mapping φ_2 and FCM

the mapping φ_1 and φ_2 give better results. Also note that using the mapping φ_3, only two rules are obtained, while for the other mappings, three rules are obtained.

Figure 1 shows the obtained antecedent membership functions using ECM and mapping φ_2, as well as using FCM, for three features of the Altman database. It is interesting to note that although only two clusters are used, three membership functions are obtained for ECM, corresponding to the focal sets $\{\omega_1, \omega_2, \Omega\}$. It is our opinion that this conveys more information than using, for instance, the FCM algorithm. Linguistic terms can easily be assigned on the cases of for the membership functions obtained with ECM, e.g. *High* for F_{k1}, *Low* for F_{k2} and *Medium* for F_{k12}. This proposed method seems to provide more information regarding the underlying structure of the model of the model, using the same number of clusters. Specifically for the case of bankruptcy prediction it is very appealing to be able to have a model that identifies the cases which are in between classes, as in general, these cases are the ones that require further assessment [22].

7 Discussion

During the experiments performed in all data sets, it was noticed that if the parameters of ECM were not chosen carefully, no object would be assigned as belonging to the boundaries of two or more clusters. This would result in results that are not as good as the ones obtained with FCM. Furthermore, it was noticed that ECM works very good in databases that contain many points in between clusters, i.e. noisy data. This is due to the fact that the credal partition gives relevant information about the points that unambiguously belong to one cluster, and the points that lie at the boundary of two or more clusters. Since these points can be seen as noise, FCM may have problems clustering them.

The proposed method seems to provide more information regarding the underlying structure of the model, using the same number of clusters. In the case of the (noisy) CR database, better results were obtained with the use of ECM, specially for the bankruptcy class while the overall accuracy remains comparable. In the general case of bankruptcy prediction it is very appealing to derive a model that identifies the cases which are in between classes, as in general, these cases are the ones that require further assessment [22]. More research is needed into transforming the information about more difficult cases (i.e. boundary cases) in specific rules.

The proposed models, obtained by extracting rules from the credal partition obtained with ECM, are computationally only slightly more complex compared to models using FCM, because a small number of clusters is used. This increase is due to the fact that in the case of FCM c rules are derived, while in the case of ECM, at most 2^c rules are derived, for the cases of the mapping φ_1 and φ_2. Note that the added rules refer to the points that lie at the boundary of two or more clusters. The assignment of linguistic terms to the obtained credal partitions, can be seen as an quantification of the degree of belief, and can be easily done by inspecting the obtained membership functions. It was noted that in some cases, due to the overlap of membership functions, the linguistic interpretation may be difficult, and further optimization of the models is required [24].

8 Conclusions

This paper discusses the use of the credal partition obtained from the Evidential C-Means based on the theoretical framework of belief functions, in deriving rule based classification models. We compare the performance of the proposed methodology with the models obtained by using the fuzzy partition matrix derived from the well know Fuzzy C-Means algorithm. The results with the proposed methodology are similar to the results obtained using FCM, but have the advantage that the use of ECM seems to provide more information about the system, which is successfully translated into rules. However, more future research is needed to assess all the characteristics of the proposed method. We will also concentrate on comparing the partitions obtained with ECM with the partitions obtained from FCM using 2^c clusters.

Acknowledgments. The authors thank the anonymous reviewers for their valuable comments. This work was (partially) supported by the European Science Foundation through COST Action IC0702.

References

1. Wang, L., Mendel, J.: Generating fuzzy rules by learning from examples. IEEE Transactions on Systems, Man and Cybernetics 22(6), 1414–1427 (1992)
2. Wu, T., Chen, S.: A new method for constructing membership functions and fuzzy rules from training examples. IEEE Transactions on Systems, Man and Cybernetics, Part B 29(1), 1083–4419 (1999)
3. Rojas, I., Pomares, H., Ortega, J., Prieto, A.: Self-organized fuzzy system generation from training examples. IEEE Transactions on Fuzzy Systems 8(1), 23–36 (2000)
4. Hartigan, J.A.: Clustering Algorithms. Wiley-Interscience, New York (1974)
5. Elouedi, Z., Mellouli, K., Smets, P.: Classification with belief decision trees. In: AIMSA 2000: Proceedings of the 9th International Conference on Artificial Intelligence, London, UK, pp. 80–90. Springer, Heidelberg (2000)
6. Masson, M.H., Denoeux, T.: ECM: An evidential version of the fuzzy c-means algorithm. Pattern Recognition 41(4), 1384–1397 (2008)
7. Masson, M.H., Denux, T.: RECM: Relational evidential c-means algorithm. Pattern Recogn. Lett. 30(11), 1015–1026 (2009)
8. Denoeux, T., Masson, M.H.: EVCLUS: evidential clustering of proximity data. IEEE Transactions on Systems, Man, and Cybernetics, Part B 34(1), 95–109 (2004)
9. Elouedi, Z., Mellouli, K., Smets, P.: Belief decision trees: theoretical foundations. Int. J. Approximate Reasoning 28, 91–124 (2001)
10. Smets, P., Kennes, R.: The transferable belief model. Artif. Intell. 66(2), 191–234 (1994)
11. Smets, P.: The transferable belief model and possibility theory. In: NAFIPS 1990, pp. 215–218 (1990)
12. Shafer, G.: A mathematical theory of evidence. Princeton University Press, Princeton (1976)
13. 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, Dordrecht (1998)
14. Smets, P.: The combination of evidence in the transferable belief model. IEEE Trans. Pattern Anal. Mach. Intell. 12(5), 447–458 (1990)
15. Smets, P.: Decision making in the TBM: the necessity of the pignistic transformation. Int. J. Approx. Reasoning 38(2), 133–147 (2005)

16. Takagi, T., Sugeno, M.: Fuzzy identification of systems and its aplication to modeling and control. IEEE Transactions on Systems, Man and Cybernetics 15(1), 116–132 (1985)
17. van den Berg, J., Kaymak, U., van den Bergh, W.M.: Fuzzy classification using probability-based rule weighting. In: Proceedings of 2002 IEEE International Conference on Fuzzy Systems, Honolulu, Hawaii, May 2002, pp. 991–996 (2002)
18. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York (1981)
19. Mangasarian, O.L., Wolberg, W.H.: Cancer diagnosis via linear programming. SIAM News 23(5), 1–18 (1990)
20. Murphy, P.M.: UCI repository of machine learning databases and domain theories (1985),
`http://www.ics.uci.edu/-mlearn/MLRepository.html`
`http://www.ftp.ics.uci.edu:pub/machine-learning-databases`
21. Altman, E.I.: Financial ratios, discriminate analysis and the prediction of corporate bankruptcy. Journal of Finance 23(4), 589–609 (1968)
22. Almeida, R.J., Vieira, S., Milea, V., Kaymak, U., Sousa, J.M.C.: Knowledge discovery in the prediction of bankruptcy. In: IFSA World Congress / EUSFLAT Conference (IFSA-EUSFLAT 2009), Lisbon, Portugal, July 2009, pp. 1785–1790 (2009)
23. Timm, H., Döring, C., Kruse, R.: Differentiated treatment of missing values in fuzzy clustering. In: De Baets, B., Kaynak, O., Bilgiç, T. (eds.) IFSA 2003. LNCS, vol. 2715, pp. 354–361. Springer, Heidelberg (2003)
24. Setnes, M., Roubos, H.: GA fuzzy modeling and classification: Complexity and performance. IEEE Transactions on Fuzzy Systems 8(5), 509–522 (2000)