Scholar Commons
USF University of South Florida

University of South Florida
Scholar Commons

Graduate Theses and Dissertations | Graduate School

10-27-2016

Scalable Clustering Using the Dempster-Shafer Theory of Evidence

Alireza Chakeri
University of South Florida, chakeri@mail.usf.edu

Follow this and additional works at: http://scholarcommons.usf.edu/etd

 Part of the Computer Sciences Commons

Scholar Commons Citation
Chakeri, Alireza, "Scalable Clustering Using the Dempster-Shafer Theory of Evidence" (2016). *Graduate Theses and Dissertations.*
http://scholarcommons.usf.edu/etd/6478

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact scholarcommons@usf.edu.

Scalable Clustering Using the Dempster-Shafer Theory of Evidence

by

Alireza Chakeri

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Computer Science
Department of Computer Science and Engineering
College of Engineering
University of South Florida

Major Professor: Lawrence O. Hall, Ph.D.
Dmitry Goldgof, Ph.D.
Rangachar Kasturi, Ph.D.

Date of Approval:
October 21, 2016

Keywords: Evidential Clustering, Single Pass Fuzzy C-Means

Copyright © 2016, Alireza Chakeri

ACKNOWLEDGMENTS

I would like to thank Dr. Lawrence Hall, Dr. Dmitry Goldgof and Dr. Rangachar Kasturi for providing me the opportunity to work on this project. The idea of this project has been enhanced in Dr. Hall's data mining class. I particularly thank Dr. Hall for spending time with all the discussion about the project and providing critical comments. Without his help this work would have not been completed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Clustering large data sets has become very important as the amount of available unlabeled data increases. Single Pass Fuzzy C-Means (SPFCM) is useful when memory is too limited to load the whole data set. The main idea is to divide dataset into several chunks and to apply fuzzy c-means (FCM) to each chunk. SPFCM uses the weighted cluster centers of the previous chunk in the next data chunks. However, when the number of chunks is increased, the algorithm shows sensitivity to the order in which the data is processed. Hence, we improved SPFCM by recognizing boundary and noisy data in each chunk and using it to influence clustering in the next chunks. The proposed approach transfers the boundary and noisy data as well as the weighted cluster centers to the next chunks. We show that our proposed approach is significantly less sensitive to the order in which the data is loaded in each chunk.

CHAPTER 1 : INTRODUCTION[1]

Data clustering aims to group a number of data points into different clusters such that data points in a cluster have more similarity to one another than those in other clusters [1, 2, 3, 4, 5, 6]. A data point consists of p features which describe an object. Using hard partitioning methods, an object only belongs to one cluster and it cannot be included in another cluster. On the other hand, in fuzzy clustering each object may belong to more than one cluster having different degrees of membership, such that the memberships sum to one. The most popular fuzzy partitioning method is Fuzzy C-means (FCM) clustering [2]. A brief survey of various variations of FCM can be found in [7, 8]. However, FCM has poor robustness against noise and outliers. Some approaches have been proposed to relax the normalization constraint in FCM. The resulting clusters are referred to as possibilistic clusters. Krishnapuram et al in [9] introduced the possibilistic clustering algorithm (PCM) by modifying the objective function of FCM. Also, rough C-means and fuzzy rough C-means were proposed in [10] by introducing the concept of lower and upper approximation of clusters. In [11], an approach to cluster noise objects poorly fit by the existing clusters was proposed. In addition, to gain a deeper insight into the data and improve robustness with respect to outliers, the credal partition was introduced in [12, 13] based on belief function theory, which can be stated from a Dempster-Shafer theory of evidence perspective. A credal partition is a generalization of the hard, fuzzy and possibilistic partitions. In this regard, based on evidential c-means (ECM) [14], a

[1]Portions of this chapter were reprinted from A. Chakeri, I. Nekooimehr, L. O. Hall, Dempster-Shafer theory of evidence in single pass fuzzy c means, IEEE International Conference on Fuzzy Systems, pp. 1-5, 2013, with permission from IEEE
Permission is included in Appendix A.

degree of membership is assigned for each object not only to single clusters, but also to any subsets of the set of clusters.

Applying a clustering algorithm to large data sets becomes a very important issue as the number of available large datasets increase. Some efforts have been made to apply clustering algorithms to large data sets [15, 16, 17]. They are mainly based on sampling the datasets or loading data incrementally. In [18], an approach to make FCM faster for large data sets was introduced based on sampling from data. It initialized the centroids for FCM using cluster centers from a limited random sampling from data. Also, in [19] a data reduction approach combined similar examples into weighted examples. On the other hand, recently a modified FCM algorithm was introduced for extremely large datasets when the limitation of memory has to be considered. Hore et al. in [20, 21] introduced a single pass FCM (SPFCM) and online FCM (OFCM) which do not need any complicated data compression technique. The SPFCM approach produces final partitions very similar to FCM by a single pass through the data. The method is based on partitioning the data set into several sub partitions and then clustering each of them. In this regard, after clustering the first chunk using fuzzy C-means, data of the second chunk and previous cluster centroids represented as weighted data are used when clustering the second chunk, and so on. They showed that SPFCM has a significant speed-up compared to fuzzy c means. The convergence proofs for SPFCM and OFCM were also given in [22]. In [17], kernel FCM has been also extended to very large data sets, including weighted kernel FCM, single pass kernel FCM and online kernel FCM.

However, when the number of chunks increases in SPFCM, i.e. the number of data in each chunk decreases, the extrema distribution obtained from the reformulated optimization function [16] may not be uniform. This can be clearly observed over experiments with different orderings

of the data. In fact, it is sensitive to the makeup of the data in each chunk. Hence, in this thesis, we introduce an improvement to SPFCM called SPECM by using boundary and noisy data. Based on this improvement, the boundary and noisy data in each chunk are identified by the ECM algorithm and their centroids are transferred to the next chunk as well as cluster centers. The idea of modification is that the boundary and noisy data may belong to the clusters in the next chunks, but SPFCM assigns those data to the cluster centers of the current chunk with almost equal degrees of membership. Hence, using SPECM, the boundary and noisy data have a chance to be put in appropriate clusters in the next chunks by including representations of them in the next chunks as well as cluster centers. The proposed approach will provide approximately similar clustering quality by loading a very small percentage of the data in each chunk compared to clustering all data at once using FCM.

CHAPTER 2 : DEMPSTER-SHAFER THEORY OF EVIDENCE[2]

Dempster-Shafer theory (D-S theory) is derived from Dempster's original work [23, 24] that relaxed some Bayesian restrictions to deal with unknown parameters. Later, Shafer in [25] extended the original concepts, and introduced what is now referred to as D-S theory of evidence. In this section, we review the important concepts of D-S theory.

2.1 Representation of Uncertainties in D-S Theory

Let $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$ denotes the set of all possible outcomes of a random variable. In conventional probability theory, the uncertainties about the elements in Θ are expressed by assigning probability values p_i to each element θ_i, that satisfy $\sum_{i=1}^{n} p_i = 1$. The uncertainties representation in D-S theory is similar to this traditional representation, with one significant difference: D-S theory assigns the probability to the subset of Θ as well as its individual elements. In D-S theory, the collection of all subsets of Θ is called the D-S frame of Θ. Also, the complete probabilities on the frame are referred to as a *Basic Probability Assignment* (BPA) [25, 26], which is denoted by $m(A)$ for any $A \subseteq \Theta$. Alternatively, $m(A)$ is also called the mass value of a focal element A. Any *normal* BPA must satisfy the following properties

[2]Portions of this chapter were reprinted from A. Chakeri, I. Nekooimehr, L. O. Hall, Dempster-Shafer theory of evidence in single pass fuzzy c means, IEEE International Conference on Fuzzy Systems, pp. 1-5, 2013, with permission from IEEE
Permission is included in Appendix A.

$$
\begin{aligned}
0 \leq m(A) \leq 1 \qquad & \forall A \subseteq \Theta \\
m(\varnothing) = 0 \qquad & \\
\sum_{A \subseteq \Theta} m(A) = 1 \qquad &
\end{aligned}
\tag{2.1}
$$

where $\varnothing$ represents the empty subset. The condition $m(\varnothing) = 0$ was originally imposed in [25] for any normal BPA, but it might be relaxed for the *open-world assumption* [27]. In the open-world assumption, the quantity $m(\varnothing) = 0$ is interpreted as a mass value to the hypothesis that the outcome might not be in Θ. However, in this thesis, we assume that Θ is complete and contains all of the outcomes. When the BPA has non-zero values only over the individual elements, which is called Bayesian BPA, it is equivalent to the conventional probability distribution. As a result, conventional probability distributions are special cases of BPA. On the other hand, when there is not any knowledge about Θ other than knowing that the truth lies within Θ, the person assign the entire BPA, i.e. 1, to the whole set Θ. This type of BPA is generally referred to as an ignorance BPA. Taking into account the above definition, a BPA provides extra freedom to express the uncertainties about Θ. This freedom makes the D-S theory attractive and useful. In fact, $m(A)$ represents the amount of belief that can be assigned to A, but to nothing smaller than A. This indicates that, the value of $m(A)$ can not be further subdivided among the elements in A.

2.2 Important Quantities

Once a BPA is defined, a number of functions can be determined for different purposes. Here, we discuss the main three functions, that are *Belief Function*, *Plausibility Function* and *Compatible Probability Distribution*.

2.2.1 Belief and Plausibility Functions

For any subset B of A, the mass value assigned to B also naturally supports A. The total belief of A, which is the collection of all mass values assigned to the subsets of A, is defined as the belief of A and denoted by $Bel(A)$ [25]

$$Bel(A) = \sum_{B \subset A} m(B) \tag{2.2}$$

When B is not necessarily contained in A but the intersection of B and A is not empty, the mass value assigned to B also provides possible support for A. These kind of possible supports to A defines the plausibility of A and denoted by $Pl(A)$ as follows

$$Pl(A) = \sum_{B \cap A \neq \varnothing} m(B) \tag{2.3}$$

The plausibility of A represents the potential support given to A. When BPA is a Bayesian BPA, the plausibility reduces to a probability measure. Also, when the focal elements are nested, it boils down to a possibility measure. As a result, probability and possibility measures can be considered as two special cases of belief functions.

Interestingly, Shafer proved a one to one correspondence between (2.1), (2.2) and (2.3) using the concept of Mobius inversion, i.e. if we have one of the belief function, plausibility function and BPA, then we can build the other two.

2.2.2 Compatible Probability Distributions

Since the mass values are assigned to the subsets of Θ, the true probability distribution is unknown over the elements of Θ. However, belief and plausibility functions of A induce a set of compatible probability distribution [23, 28, 29], $P(A)$, that satisfy

$$Bel(A) \leq P(A) \leq Pl(A) \tag{2.4}$$

Now, assume that $\{A_1, A_2, ...A_L\}$ contains all of the focal elements with their corresponding mass values $\{m_1, m_2, ..., m_L\}$. Then, the set of the compatible probability distributions, S, can be written as

$$S = \left\{ (p_1, \ldots, p_n) \;\middle|\; \begin{array}{lll} p_i = \sum_j m_{ij} & i = \{1, \ldots, n\} \\ m_j = \sum_i m_{ij} & j = \{1, \ldots, L\} \\ m_{ij} = 0 & \text{if} \quad \theta_i \notin A_j \end{array} \right\} \tag{2.5}$$

The motivation for defining m_{ij} is that all or part of m_j can be assigned to θ_i, but the distribution of m_j over A_j is not specified. In this definition, p_i is the sum of probabilities that θ_i obtains from different A_j's. Since S expressed by linear constraints is a convex set , the probability of set $B \subset \Theta$ lies in an interval called interval-valued probability. This uncertainty is a result of assigning mass values to sets instead of elements. In fact, Shafer in [25] showed that *Belief* and *Plausibility* of set B are the lower and upper bounds of its interval probability. i.e.,

$$Pr(B) = [Bel(B), Pl(B)] \tag{2.6}$$

where he also showed that

$$Bel(B) = \min_{(p_1,\ldots,p_n)\in S} \sum_{\theta_i \in B} p_i \tag{2.7}$$

$$pls(B) = \max_{(p_1,\ldots,p_n)\in S} \sum_{\theta_i \in B} p_i \tag{2.8}$$

2.2.3 Pignistic Transformation

Although the D-S framework provides a more flexible way to represent uncertainties, it is often necessary to make a decision and select one single outcome in Θ. There are different ways of doing this. One way is to choose the single element in Θ with the highest plausibility [30]. On the other hand, Smets [26] has proposed the use of a probability function. In particular, the pignistic transformation is the only transformation of a BPA into a probability distribution that satisfies elementary rationality requirements. In the pignistic transformation, each mass value is equally divided among its corresponding focal elements [31]. Thus, the pignistic probability for a normal BPA is defined as

$$BetP(\theta_i) \triangleq \sum_{\theta_i \in A} \frac{m(A)}{|A|} \tag{2.9}$$

where $|A|$ denotes the cardinality of the set A.

CHAPTER 3 : SCALABLE FUZZY C-MEANS[3]

Clustering is an exploratory data analysis tool that aims at grouping a set of objects into different groups so that similar objects are grouped together and dissimilar objects are assigned to different groups. A wide variety of methods for clustering data points have been developed [1, 2, 3, 4, 5, 6]. They can be generally classified into three main categories: *hard partitioning*, *probabilistic* and *soft partitioning* clustering approaches. Probabilistic clustering approaches give a probability of an example belonging to one or more clusters. We ignore probabilistic approaches like EM [32] here. In hard partitioning methods [1, 5, 33], data points are grouped in an exclusive way, such that each object only belongs to one cluster. On the other hand, using soft partitioning methods [2, 6, 9, 14, 34, 35, 36, 37], each object can belong to more than one cluster with different membership degrees. Soft partitioning methods carry more information about the data. One of the most popular soft partitioning method is *fuzzy c-means* (FCM) [2]. In this section, we review the FCM method and its single pass version.

3.1 Fuzzy C-Means

Let $\{x_1, x_2, ...x_n\}$ be the set of data points in $\mathbb{R}^p$, and c be the number of clusters. Let matrix V of size $c \times p$ represents the cluster centers. By minimizing a suitable objective function,

[3]Portions of this chapter were reprinted from A. Chakeri, I. Nekooimehr, L. O. Hall, Dempster-Shafer theory of evidence in single pass fuzzy c means, IEEE International Conference on Fuzzy Systems, pp. 1-5, 2013, with permission from IEEE
Permission is included in Appendix A.

FCM finds a fuzzy membership matrix $U = (u_{ik})$ of size $n \times c$ such that $\forall i \in \{1, ..., n\}$

$$\sum_{k=1}^{c} u_{ik} = 1 \tag{3.1}$$

and $\forall k \in \{1, ..., c\}$

$$\sum_{i=1}^{n} u_{ik} > 0 \tag{3.2}$$

Then, FCM looks for a fuzzy partitioning matrix U and cluster centers V by minimizing the following objective functions

$$J_{FCM}(U, V) = \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik}^{\beta} d_{ik}^2 \tag{3.3}$$

where d_{ik} shows the Euclidean distance between the kth cluster center and ith data point. In the objective function, β is a weighting parameter that controls the fuzziness of the partition. It is typically set to 2. Bezdek in [2] minimized the above objective function using an iterative algorithm. In particular, the cluster centers and the membership degrees are optimized alternately. The cluster centers v_k and the membership degrees u_{ij} are given by

$$v_k = \frac{\sum_{i=1}^{n} u_{ik}^{\beta} \, x_i}{\sum_{i=1}^{n} u_{ik}^{\beta}} \tag{3.4}$$

$$u_{ij} = \frac{d_{ij}^{2/(1-\beta)}}{\sum_{k=1}^{c} d_{ik}^{2/(1-\beta)}} \tag{3.5}$$

The above iterative algorithm changes the initialization of the partitioning matrix and the cluster centers over time until it converges with stopping condition $||V_{t+1} - V_t|| < \epsilon$.

Several authors have proposed to relax the normalization constraint in (3.1) to overcome the poor robustness of FCM against noise and outliers. In this regard, Krishnapuram and Keller in [9] modified the objective function in FCM and introduced the possibilistic clustering algorithm (PCM). In PCM, the resulting partitions are referred to as possibilistic partitions rather than fuzzy partitions. In a different approach, Dave in [11] added a noise cluster to FCM in order to group objects that are poorly represented by the clusters. The approaches have been further extended to cluster relational data [38, 39].

3.2 Sinlge Pass FCM

The single pass fuzzy C-means (SPFCM) algorithm was developed to cluster very large data sets fast [20]. It divides large data in smaller partitions called partial data accesses (PDA) based on the available memory allocation, assuming that the data is randomly scrambled on the disk. After scanning the first PDA, the algorithm divides the instances into c clusters using FCM and then the centers of the clusters are weighted and sent to be clustered again with new points loaded in the next PDA. In other words, instances in the first PDA are condensed into c weighted points, as

$$w_j = \sum_{i=1}^{n_1} u_{ij} \quad \forall j = 1, ..., c \tag{3.6}$$

where n_1 is the number of examples in the first PDA. These weighted points are the centers of the clusters such that their weights are calculated by summing the membership values of instances belong to them. For the second PDA up to the last one, the weighted points and the new loaded points are clustered again to form c new clusters, and then they will be weighted and sent to be clustered with the next PDA until all the examples will be scanned once. Note that the weights

for the first PDA will sum up to the number of instances within the PDA, n_1. The weights for the second PDA will sum up to the number of instances scanned in the first and second PDA, i.e. $n_1 + n_2$. For the third PDA it will be $n_1 + n_2 + n_3$, and for the last PDA it will be the total number of examples in the original dataset. The convergence of SPFCM algorithm has been proven in [22]. Table 3.1 shows how the algorithm works.

Table 3.1: Single pass FCM algorithm.

1. Cluster the first chunk of data by FCM.
2. Compute weights for the cluster centers based on the degree of membership assigned to the objects with equation (3.6).
3. Bring in the data in the next chunk and the c weighted cluster centers from the last chunk and apply FCM.
4. Repeat step 2-3 until all chunks are processed.

CHAPTER 4 : D-S THEORY OF EVIDENCE IN CLUSTERING[4]

Dempster-Shafer (DS) theory is a framework for reasoning with partial information. It has been applied successfully in various disciplines such as information fusion [40], classification [41, 42] and clustering [14, 43, 44]. Evidential clustering provides a rich and informative description of the structure of data points. It can be shown that the hard, fuzzy and rough partitions are special cases of evidential clustering. Also, it has been utilized in various domains such as medical image processing [45, 46] and social networks analysis [47]. In this section, we study the essentials of evidential c-means clustering

4.1 Credal Partition

Recently, the notion of a credal partition has been introduced based on D-S theory [12, 13]. A credal partition is an extension of the hard, fuzzy and possibilistic partitions. This framework allows us to handle different situations from complete ignorance to full certainty. The idea is that, for each object, it allocates a complete BPA to single clusters as well as any subset of the clusters. To clarify the concept, consider the following example. Assume the data set contains 5 examples with 3 clusters $\Theta = \{\theta_1, \theta_2, \theta_3\}$. Table 4.1 gives the complete BPA for each object. It illustrates different situations; the cluster of example 1 is known with certainty (hard partition); The BPA of examples 2 and 3 are partially known (m_2 specifies a fuzzy partition); The cluster of example 4 is

[4]Portions of this chapter were reprinted from A. Chakeri, I. Nekooimehr, L. O. Hall, Dempster-Shafer theory of evidence in single pass fuzzy c means, IEEE International Conference on Fuzzy Systems, pp. 1-5, 2013, with permission from IEEE
Permission is included in Appendix A.

completely unknown (complete ignorance); and the BPA of example 5 indicates that it does not belong to any of the clusters (outlier).

Table 4.1: An example for a credal partition.

A	$m_1(A)$	$m_2(A)$	$m_3(A)$	$m_4(A)$	$m_5(A)$
$\varnothing$	0	0	0	0	1
θ_1	1	0.2	0	0	0
θ_2	0	0.3	0.5	0	0
θ_3	0	0.5	0	0	0
$\{\theta_1, \theta_2\}$	0	0	0.3	0	0
$\{\theta_1, \theta_3\}$	0	0	0	0	0
$\{\theta_2, \theta_3\}$	0	0	0.2	0	0
Θ	0	0	0	1	0

4.1.1 Interpretation of a Credal Partition

From a credal partition, the classical partitions (hard, fuzzy and possibilistic partitions) can be recovered [14]. For instance, one can convert a credal partition into a possibilistic partition using the plausibility values, i.e.

$$poss_i(\theta_k) = \sum_{A \cap \{\theta_k\} \neq \varnothing} m_i(A) \tag{4.1}$$

where $poss_i(\theta_k)$ shows the possibility that object i belongs to cluster k. Similarily, one can use the pignistic distribution and convert a credal partition into a fuzzy partition. In addition, one can obtain a hard partition by assigning the objects to the cluster with the highest pignistic probability. Credal partitioning framework also provides rich information about the structure of data. It finds the objects that belong to one cluster, as well as the objects that lie at the boundaries of two or more clusters (see Figure 4.1). For instance, $\{\theta_1, \theta_2\}$ represents the boundary between clusters 1 and 2. As a result, from a credal partition, we can find the membership degree of each object belonging to each cluster or to the boundaries of the clusters. Also, the objects that belong to the empty subset

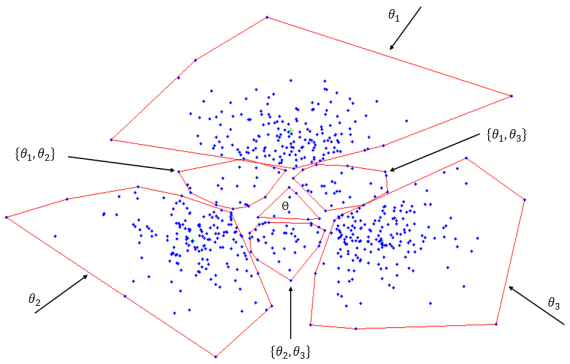

Figure 4.1: A hard partition derived from a credal partition. The clusters and their boundaries are very well separated.

are considered as outliers.

4.2 Evidential C-Means

Evidential C-Means (ECM) which is the application of D-S theory in clustering, was developed in [14] to obtain a credal partition. In ECM, an object not only can belong to each cluster, but it can also belong to a subset of clusters. The main idea is that in deriving a credal partition one needs to determine the distance between each object and any subset of the clusters. In particular, let's assume $\Theta = \{\theta_1, \theta_2, ..., \theta_c\}$ is the set of all possible clusters. For each subset A_j of Θ, ECM associates the barycenter $\bar{v}_j$ of the centers associated with the clusters present in A_j. More precisely, the barycenter $\bar{v}_j$ is defined as

$$\bar{v}_j = \frac{1}{c_j} \sum_{k=1}^{c} s_{kj} v_k \tag{4.2}$$

where $c_j = |A_j|$ denotes the cardinality of A_j and s_{kj} is

$$s_{kj} = \begin{cases} 1 & \text{if } \theta_k \in A_j \\ 0 & \text{otherwise} \end{cases} \tag{4.3}$$

Using this definition, the distance between object i and subset A_j can be written as

$$d_{ij}^2 = ||x_i - \bar{v_j}||^2 \tag{4.4}$$

ECM has a similar iterative optimization algorithm to FCM, but it improves FCM in some ways. First, the partitioning matrix U in FCM is replaced with the credal partition $M = (m_1, ..., m_n) \in \mathbb{R}^{n \times 2^c}$. Then, the objective function of ECM is modified as follows

$$J_{ECM}(M, V) = \sum_{i=1}^{n} \sum_{\{j | A_j \neq \varnothing, A_j \subseteq \Theta\}} c_j^{\alpha} m_{ij}^{\beta} d_{ij}^2 + \sum_{i=1}^{n} \delta^2 m_{i\varnothing}^{\beta} \tag{4.5}$$

subject to

$$\sum_{\{j | A_j \subseteq \Theta, A_j \neq \varnothing\}} m_{ij} + m_{i\varnothing} = 1, \quad \forall i = 1, ..., n \tag{4.6}$$

where $m_{ij} = m_i(A_j)$, and the term c_k^{α} penalizes subsets with high cardinalities, such that the exponent α controls the amount of penalization (the higher α is, the less data points in the boundaries). α is typically set to 2. Also, parameter β controls the number of data points considered as outliers. In this objective function, similar to PCM [9], the empty set is treated separately allowing the algorithm to find noisy data. The above objective function is similar to the objective function in [11]. What essentially differs the two objective functions is that here one considers all of the subsets of clusters rather than the individual clusters. The above objective function can be minimized using the Lagrangian multiplier and calculating the mass values as follows

$$m_{ij} = \frac{c_j^{\frac{\alpha}{1-\beta}} d_{ij}^{\frac{2}{1-\beta}}}{\sum_{A_k \neq \varnothing} c_k^{\frac{\alpha}{1-\beta}} d_{ik}^{\frac{2}{1-\beta}} + \delta^{\frac{2}{1-\beta}}}, \quad \forall i = 1, ..., n, \ \forall j : A_j \subseteq \Theta, A_j \neq \varnothing \tag{4.7}$$

and

$$m_{i\varnothing} = 1 - \sum_{A_j \neq \varnothing} m_{ij}, \qquad \forall i = 1, ..., n \tag{4.8}$$

Using the similar technique as in FCM, one can fix M and find the optimal matrix V. The authors in [14] proposed to first find B ($c \times p$ matrix) and H ($c \times c$ matrix) as

$$B_{lq} = \sum_{i=1}^{n} x_{iq} \sum_{j:\theta_l \in A_j} c_j^{\alpha-1} m_{ij}^{\beta}, \tag{4.9}$$

and

$$H_{lk} = \sum_{i=1}^{n} \sum_{j:\{\theta_l, \theta_k\} \subseteq A_j} c_j^{\alpha-2} m_{ij}^{\beta}, \tag{4.10}$$

Using these notations, V can be computed as follows

$$V = H^{-1} B \tag{4.11}$$

4.2.1 Time Complexity Limitation

ECM algorithm finds the degrees of membership for each object to all subsets of the set of clusters, i.e. 2^c subsets. As a result, the number of parameters in ECM is exponential in the number of clusters. This makes the algorithm very expensive for a high number of clusters. However, in [14], the authors suggested reducing the complexity of the algorithm by considering only the subsets with a limited number of focal elements. For instance, one may be only interested in the boundaries between every 2 clusters, not more than that. Hence, the complexity will be reduced to c^2. In this thesis, we do the same.

CHAPTER 5 : SINGLE PASS EVIDENTIAL C-MEANS[5]

Our proposed algorithm works pretty much the same as SPFCM except that it uses ECM instead of FCM to cluster the instances/data points and finds the boundary and noisy instances as well as cluster centers in each PDA. Our algorithm not only sends weighted cluster centers to the next PDA, it also sends weighted centers of the boundary points to be in the next chunk of data. Using ECM to cluster instances in each PDA also helps us to detect outliers, which are the instances with a tendency to belong to the empty set. The outliers are also transferred to the next PDA.

Assume we have a very large data set which needs space that is larger than memory size. To solve this problem, our algorithm randomly chooses n_i instances out of the whole data set each time without replacement. For the first n_1 instances, the proposed algorithm uses ECM to find the centers of clusters, the centers of boundary and noisy points and the membership value with which each instance belongs to each cluster or the borders of the clusters. Then the centers of clusters, the noisy data points, and the centers of boundary points are weighted and added to the next n_2 points to be clustered again using ECM. The centers of clusters, the noisy data points and centers of boundary points will be detected again and will be sent to the next random subset of overall data. This process will be repeated until all points are scanned. However, because it is not necessary to identify the boundary and noisy data in the last PDA, we use FCM in the last PDA to find

[5]Portions of this chapter were reprinted from A. Chakeri, I. Nekooimehr, L. O. Hall, Dempster-Shafer theory of evidence in single pass fuzzy c means, IEEE International Conference on Fuzzy Systems, pp. 1-5, 2013, with permission from IEEE
Permission is included in Appendix A.

the final cluster centers. The weights for the first n_1 instances are calculated by the summation of membership values of examples in a cluster or boundary set. For subsets after the first one, the weights are calculated by taking into account the previous PDA weights. The SPECM algorithm is summarized in Table 5.1.

Table 5.1: Single pass ECM algorithm.

1. Cluster the first chunk by ECM.
2. Create weights for the centers of clusters and the boundary points based on the degree of membership assigned to the them. Also find the noisy instances.
3. Bring in the next chunk and the weighted centers of clusters, the centers of boundary points and noisy data from the previous chunk and apply ECM.
4. Repeat steps 2-3 until the last chunk is processed.
5. Apply FCM to the last chunk.

5.1 Calculation of Weighted Points

For the first PDA, we only have n_1 instances $\{x_1, ..., x_{n_1}\}$ and there will be no previous weighted points (the weights of all examples are 1). Thus the algorithm uses regular ECM to find the centroids v_j, $1 \leq j \leq 2^c$ and the mass values m_{ij}, $1 \leq i \leq n_1$ and $1 \leq j \leq 2^c$. Hence, The weights of the centroids in the first PDA are calculated such that

$$w_j^{'} = \sum_{i=1}^{n_1} m_{ij} w_i, \quad \forall j = 1, ..., 2^c - 1 \tag{5.1}$$

where $w_i = 1$. Here, j is considered less than or equal to $2^c - 1$, since the last subset corresponds to the empty subset, which is considered to consist of outliers. Now, assume that the set $O_k = \{x_{k_1}, ..., x_{k_{|O_k|}}\}$ is the set of noisy data in the kth chunk, and $U_{\varnothing}^k = \{u_{\varnothing k_1}, ..., u_{\varnothing k_{|O_k|}}\}$ is the set of

degrees of belonging of the noisy data to the set $\varnothing$ in the kth chunk. Hence, for the second PDA, the weights for the centers and noisy data are

$$w_i = w_i^{'}, \qquad \forall i = 1, ..., 2^c - 1 \tag{5.2}$$

$$w_i = u_{\varnothing 1_i} \text{ s.t. } u_{\varnothing 1_i} \in U_{\varnothing}^1, \qquad \forall i = 2^c, ..., 2^c - 1 + |O_1| \tag{5.3}$$

Also the weights for the n_2 new instances just uploaded are

$$w_i = 1, \qquad \forall i = 2^c + |O_1|, ..., 2^c - 1 + |O_1| + n_2 \tag{5.4}$$

For the rest of PDA, we compute the weights for the centroids in a similar fashion. For instance, for the kth chunk, we use n_k, O_k and $U_{\varnothing}^k$.

5.2 Weighted ECM

As we mentioned before, our algorithm uses ECM in each PDA to find the centers of the clusters, the noisy data points and centers of boundary points, and then it weights them and sends them to be clustered again along with the new instances scanned in the next PDA. Hence, ECM needs to be modified for weighted points. We modified the ECM objective function such that it takes into account the weights of each example, as follows

$$J_{WECM}(M, V) = \sum_{i=1}^{n} w_i \sum_{\{j | A_j \neq \varnothing, A_j \subseteq \Theta\}} c_j^{\alpha} m_{ij}^{\beta} d_{ij}^2 + \sum_{i=1}^{n} \delta^2 w_i m_{i\varnothing}^{\beta} \tag{5.5}$$

subject to

$$\sum_{\{j|A_j\subseteq\Theta,A_j\neq\varnothing\}} m_{ij} + m_{i\varnothing} = 1, \quad \forall i = 1, ..., n \tag{5.6}$$

The above objective function can be minimized using the Lagrangian multiplier and one can calculate the mass values as follows

$$m_{ij} = \frac{c_j^{\frac{\alpha}{1-\beta}} d_{ij}^{\frac{2}{1-\beta}}}{\sum_{A_k\neq\varnothing} c_k^{\frac{\alpha}{1-\beta}} d_{ik}^{\frac{2}{1-\beta}} + \delta^{\frac{2}{1-\beta}}}, \qquad \forall i = 1, ..., n,\ \forall j : A_j \subseteq \Theta, A_j \neq \varnothing \tag{5.7}$$

and

$$m_{i\varnothing} = 1 - \sum_{A_j\neq\varnothing} m_{ij}, \quad \forall i = 1, ..., n \tag{5.8}$$

Also, H and B are adjusted as

$$B_{lq} = \sum_{i=1}^{n} w_i x_{iq} \sum_{j:\theta_l\in A_j} c_j^{\alpha-1} m_{ij}^{\beta}, \tag{5.9}$$

and

$$H_{lk} = \sum_{i=1}^{n} w_i \sum_{j:\{\theta_l,\theta_k\}\subseteq A_j} c_j^{\alpha-2} m_{ij}^{\beta}, \tag{5.10}$$

Using these notations, V can be computed as the solution of the following linear system

$$HV = B \tag{5.11}$$

CHAPTER 6 : MATERIALS AND EXPERIMENTAL DESIGN[6]

Two small data sets, Iris and Pima, were used in this study to test the developed SPECM algorithm. The Iris plant data set consists of 50 samples from each of three species of Iris, i.e., Iris setosa, Iris virginica and Iris versicolor. Four numeric attributes (the length and the width of the sepals and petals) were measured from each sample. The visualization of the Iris data set indicates that one class is completely separable and the other two classes intersect each other. In this study, the number of clusters for Iris data set was set as 3. The Pima Indians Diabetes database consists of 8 medical measurements of 768 subjects. A binary class variable is also provided in the data set to indicate whether the subject has diabetes or not.

First, the data was randomized and the weights were initialized to one. The data samples were then divided into multiple chunks and sequentially provided to the SPECM model. For the Iris data, each chunk consists of 5 and 10 data samples and there are 30 and 15 chunks in total respectively. For the Pima data set, on the other hand, 16 chunks were created and there were 48 data in each chunk. Furthermore, for each data set, the experiment was repeated 50 times, each time a new randomization of the data was done. It is noted that in the experiment, we increased the exponent for cardinality (see equation (5.5)) as more PDAs were scanned. In this way, there would be fewer data samples on the boundaries of clusters as more instances are scanned. To illustrate

[6]Portions of this chapter were reprinted from A. Chakeri, I. Nekooimehr, L. O. Hall, Dempster-Shafer theory of evidence in single pass fuzzy c means, IEEE International Conference on Fuzzy Systems, pp. 1-5, 2013, with permission from IEEE
Permission is included in Appendix A.

the model performance, the reformulated optimization criteria R_m was computed as

$$R_m(V) = \sum_{i=1}^{n} \Big(\sum_{k=1}^{c} d_{ik}^{\frac{1}{1-m}} \Big)^{(1-m)} \tag{6.1}$$

where d_{ik} is the distance between the ith data point to the kth cluster center.

6.1 Experimental Results

Two features, i.e. petal length and petal width, are used for the visualizing a partition of the Iris data. The partition is visualized in Fig. 6.1. As discussed above, just 10 data samples form a chunk and are clustered by the SPECM model. It is noted that the separable class (shown with blue color) is well separated on the lower-left corner and the two overlapped classes (green and red symbols) are also separated effectively on the upper-right corner. In the SPFCM approach, on the Iris data set, the two overlapped classes came out as a single cluster and the separable one got split into 2 clusters [21] sometimes. This has not happened in our experiments. As a result, the average error of assigning classes to data points over 50 experiments in SPFCM is 15.6, but using SPECM the average error over 50 experiments is 14.7. The improvement is statistically significant (p-value is almost 0.001, less than 0.05). This is because the developed method models the data more completely than SPFCM, and hence is less sensitive to the order in which the data is loaded in each chunk. However, the average running time for SPECM is 0.3304 second compared to SPFCM which is 0.0724 seconds. Hence, SPECM delivers more accurate results almost 5 times slower.

The extrema distributions of FCM and SPECM on the Iris data over 50 experiments are shown in Figure 6.2 such that 5 data samples are in one chunk and in total 30 chunks are clustered sequentially by SPECM. The x-axis is the number of the experiment and y-axis is the R_m value.

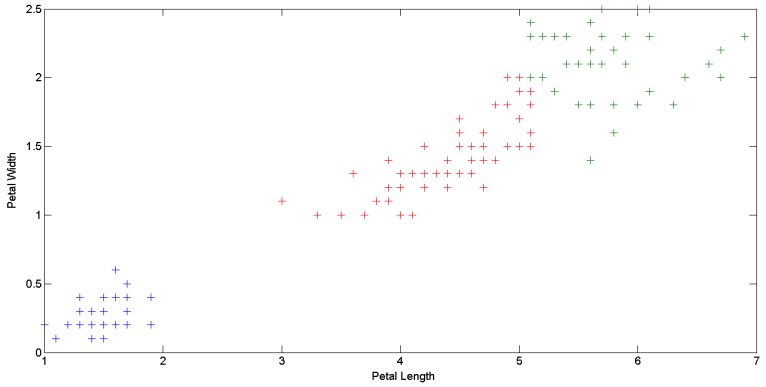

Figure 6.1: The three clusters are represented with different colors.

It can be observed that for the extrema of SPECM, they are distributed closely around 63.4 and no poor partition is found. However, as in [21], the extrema distribution of SPFCM shows that in some of the experiments the results are poor with high R_m values. As a result, the SPECM is more robust even when the number of chunks is high (or equivalently, when there are a few examples in each chunk). We also show the results of extrema distribution in Fig. 6.3 when there 10 examples in each chunk (15 chunks). Here, the R_m value is very close to the R_m value of FCM. This in fact true since there are more examples in each chunk, and is more similar to loading all data at once.

The extrema distributions of FCM and SPECM on the Pima data over 50 experiments are also shown in Fig. 6.4 such that 48 data samples are in one chunk and in total 16 chunks are clustered sequentially by SPECM. It can be observed that the extrema of SPECM, are distributed closely around 3.46 and no poor partition is found. The figure also shows that SPECM can cluster the Pima data set better than FCM in terms of R_m -2.88 percent less, while SPFCM performs 2.22 percent worse than FCM. However, the average running time for SPECM is 1.652 second compared to SPFCM which is 0.305 seconds.

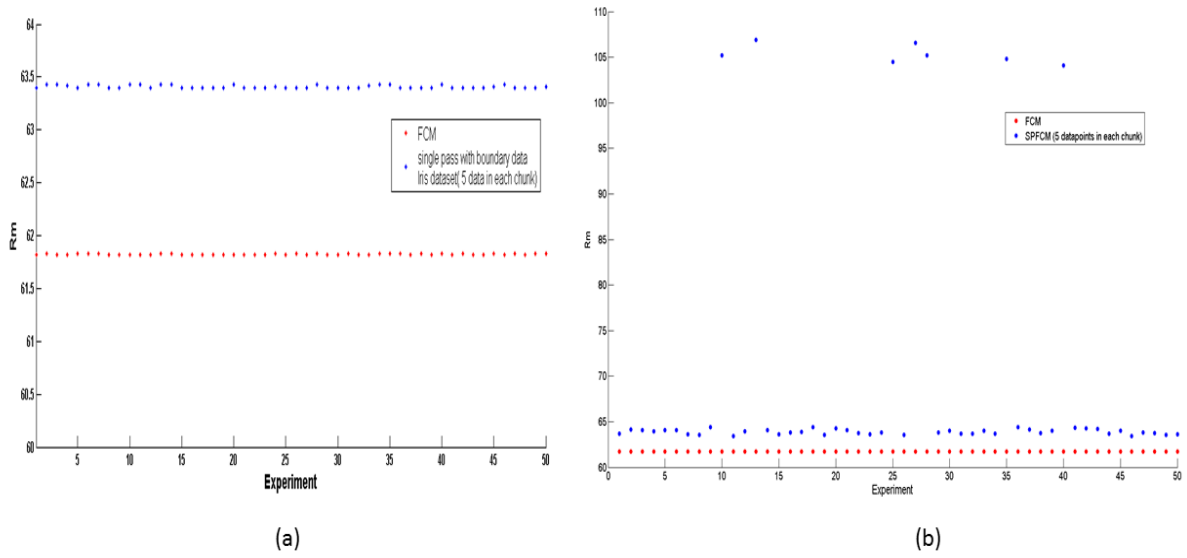

Figure 6.2: a) Extrema distribution of FCM and SPECM on the Iris data set when there are 30 chunks for SPECM. b) Extrema distribution of FCM and SPFCM on the Iris data set when there are 30 chunks for SPFCM. As the figure shows, the SPECM is more stable than SPFCM with regard to data randomization.

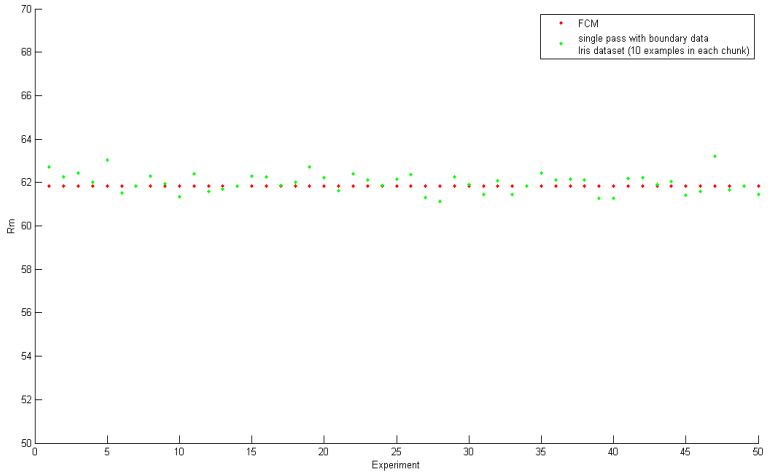

Figure 6.3: Extrema distribution of FCM and SPECM on the Iris data set when there are 15 chunks for SPECM.

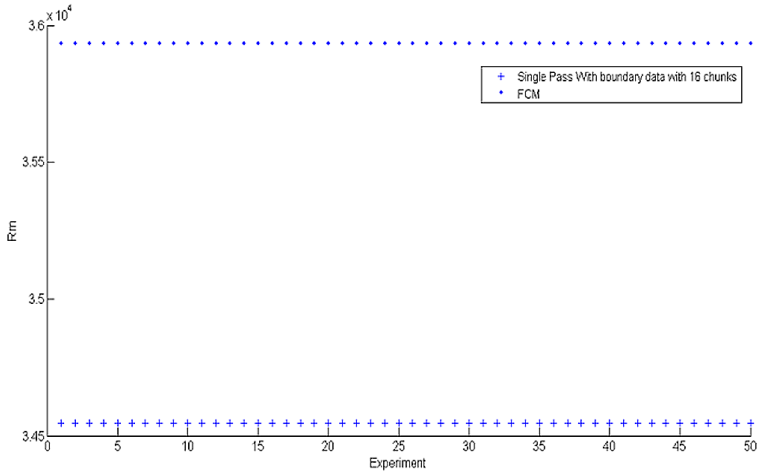

Figure 6.4: Extrema distribution of FCM and SPECM on the Pima data set when there are 16 chunks for SPECM.

CHAPTER 7 : CONCLUSIONS[7]

In this study, the SPECM algorithm was developed. The developed algorithm improves on SPFCM by incorporating cluster boundary and noisy data in each chunk. In each new chunk, the weighted cluster centroids as well as the data samples on the cluster boundaries are incorporated. Thus, the developed SPECM algorithm not only maintains the basic function of SPFCM, which enables partitioning a large amount of data accurately and quickly, but also eliminates the sensitivity to the loading of data samples. We extensively tested our algorithm using two small public data sets and both results show that the SPECM is more robust in terms of extrema obtained. The proposed approach can handle data with different numbers of clusters, not limited to the 3 clusters Iris data set and 2 clusters Pima data set. Although the proposed algorithm has higher time complexity compared to SPFCM, it has great potential in the fuzzy clustering of large data sets, e.g. streaming data.

[7]Portions of this chapter were reprinted from A. Chakeri, I. Nekooimehr, L. O. Hall, Dempster-Shafer theory of evidence in single pass fuzzy c means, IEEE International Conference on Fuzzy Systems, pp. 1-5, 2013, with permission from IEEE
Permission is included in Appendix A.

REFERENCES

[1] Anil K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8): 651 – 666, 2010. Award winning papers from the 19th International Conference on Pattern Recognition (ICPR)19th International Conference in Pattern Recognition (ICPR).

[2] James C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Kluwer Academic Publishers, Norwell, MA, USA, 1981. ISBN 0306406713.

[3] Lawrence O. Hall. Objective function-based clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(4):326–339, 2012.

[4] David M. Blei, Andrew Y. Ng, Michael I. Jordan, and John Lafferty. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:2003, 2003.

[5] J. Shi and J. Malik. Normalized cuts and image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(8):888–905, Aug 2000.

[6] M. Pavan and M. Pelillo. Dominant sets and pairwise clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(1):167–172, 2007.

[7] H. Zhou, G. Schaefer, A. H. Sadka, and M. E. Celebi. Anisotropic mean shift based fuzzy c-means segmentation of dermoscopy images. *IEEE Journal of Selected Topics in Signal Processing*, 3(1):26–34, Feb 2009.

[8] Huiyu Zhou, Gerald Schaefer, and Chunmei Shi. *Fuzzy C-Means Techniques for Medical Image Segmentation*, pages 257–271. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[9] R. Krishnapuram and J. M. Keller. A possibilistic approach to clustering. *Trans. Fuz Sys.*, 1 (2):98–110, May 1993. ISSN 1063-6706.

[10] G. Schaefer, Q. Hu, Huiyu Zhou, J.F. Peters, and A.E. Hassanien. Rough c-means and fuzzy rough c-means for colour quantisation. *Fundamenta Informaticae*, 119(1):113–120, 2012. ISSN 0169-2968. doi: 10.3233/FI-2012-729.

[11] Sumit Sen and R. N. Dave. Clustering of relational data containing noise and outliers. In *Fuzzy Systems Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on*, volume 2, pages 1411–1416 vol.2, May 1998.

[12] B Bouchon-Meunier, L Foulloy, and RR Yager. Clustering of proximity data using belief functions. *Intelligent Systems for Information Processing: From Representation to Applications*, page 291, 2003.

[13] T. Denoeux and M. H. Masson. Evclus: evidential clustering of proximity data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1):95–109, Feb 2004. ISSN 1083-4419.

[14] Ecm: An evidential version of the fuzzy c-means algorithm. *Pattern Recognition*, 41(4):1384 – 1397, 2008.

[15] P. S. Bradley, Usama Fayyad, and Cory Reina. Scaling clustering algorithms to large databases. In *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining*, KDD'98, pages 9–15. AAAI Press, 1998.

[16] N. R. Pal and J. C. Bezdek. Complexity reduction for "large image" processing. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 32(5):598–611, Oct 2002.

[17] T. C. Havens, J. C. Bezdek, C. Leckie, L. O. Hall, and M. Palaniswami. Fuzzy c-means algorithms for very large data. *IEEE Transactions on Fuzzy Systems*, 20(6):1130–1146, Dec 2012.

[18] Tai Wai Cheng, Dmitry B. Goldgof, and Lawrence O. Hall. Fast fuzzy clustering. *Fuzzy Sets and Systems*, 93(1):49 – 56, 1998.

[19] S. Eschrich, Jingwei Ke, L. O. Hall, and D. B. Goldgof. Fast accurate fuzzy clustering through data reduction. *IEEE Transactions on Fuzzy Systems*, 11(2):262–270, Apr 2003.

[20] Prodip Hore, Lawrence O Hall, and Dmitry B Goldgof. Single pass fuzzy c means. In *2007 IEEE International Fuzzy Systems Conference*, pages 1–7. IEEE, 2007.

[21] P. Hore, L. O. Hall, and D. B. Goldgof. Creating streaming iterative soft clustering algorithms. In *NAFIPS 2007 - 2007 Annual Meeting of the North American Fuzzy Information Processing Society*, pages 484–488, June 2007.

[22] L. O. Hall and D. B. Goldgof. Convergence of the single-pass and online fuzzy c-means algorithms. *IEEE Transactions on Fuzzy Systems*, 19(4):792–794, Aug 2011.

[23] A. P. Dempster. Upper and lower probabilities induced by a multivalued mapping. *Ann. Math. Statist.*, 38(2):325–339, 04 1967. doi: 10.1214/aoms/1177698950.

[24] A. P. Dempster. A generalization of bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 30(2):205–247, 1968. ISSN 00359246.

[25] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, Princeton, 1976.

[26] Philippe Smets and Robert Kennes. The transferable belief model. *Artificial Intelligence*, 66 (2):191 – 234, 1994. ISSN 0004-3702.

[27] P. Smets. The combination of evidence in the transferable belief model. *IEEE Trans. Pattern Anal. Mach. Intell.*, 12(5):447–458, May 1990. ISSN 0162-8828.

[28] Larry A. Wasserman. Belief functions and statistical inference. *Canadian Journal of Statistics*, 18(3):183–196, 1990.

[29] Larry Alan Wasserman. Prior envelopes based on belief functions. *The Annals of Statistics*, 18 (1):454–464, 1990.

[30] Barry R. Cobb and Prakash P. Shenoy. On the plausibility transformation method for translating belief function models to probability models. *International Journal of Approximate Reasoning*, 41(3):314 – 330, 2006.

[31] Philippe Smets. Decision making in the tbm: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38(2):133 – 147, 2005.

[32] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38, 1977.

[33] J. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pages 281–297, Berkeley, Calif., 1967. University of California Press.

[34] L.O. Hall and A. Chakeri. Relational data partitioning using evolutionary game theory. In *Computational Intelligence and Data Mining (CIDM), 2014 IEEE Symposium on*, pages 113–120, Dec 2014.

[35] A. Chakeri and L. O. Hall. Large data clustering using quadratic programming: A comprehensive quantitative analysis. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 806–813, Nov 2015.

[36] A. Chakeri and L.O. Hall. Dominant sets as a framework for cluster ensembles: An evolutionary game theory approach. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 3457–3462, Aug 2014.

[37] A. Chakeri, I. Nekooimehr, and L. O. Hall. Dempster-shafer theory of evidence in single pass fuzzy c means. In *Fuzzy Systems (FUZZ), 2013 IEEE International Conference on*, pages 1–5, July 2013. doi: 10.1109/FUZZ-IEEE.2013.6622476.

[38] Richard J. Hathaway, James C. Bezdek, and John W. Davenport. On relational data versions of c-means algorithms. *Pattern Recognition Letters*, 17(6):607 – 612, 1996.

[39] Miquel De Cáceres, Francesc Oliva, and Xavier Font. On relational possibilistic clustering. *Pattern Recogn.*, 39(11):2010–2024, November 2006. ISSN 0031-3203.

[40] Frederic Pichon, Didier Dubois, and Thierry Denoeux. Relevance and truthfulness in information correction and fusion. *International Journal of Approximate Reasoning*, 53(2):159 – 175, 2012.

[41] Chunfeng Lian, Su Ruan, and Thierry Denoeux. An evidential classifier based on feature selection and two-step classification strategy. *Pattern Recognition*, 48(7):2318 – 2327, 2015.

[42] T. Denoeux. A neural network classifier based on dempster-shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 30(2):131–150, Mar 2000.

[43] Thierry Denoeux, Songsak Sriboonchitta, and Orakanya Kanjanatarakul. Evidential clustering of large dissimilarity data. *Knowledge-Based Systems*, 106:179 – 195, 2016.

[44] Cecm: Constrained evidential c-means algorithm. *Computational Statistics and Data Analysis*, 56(4):894 – 914, 2012.

[45] Benoit Lelandais, Su Ruan, Thierry Denoeux, Pierre Vera, and Isabelle Gardin. Fusion of multi-tracer {PET} images for dose painting. *Medical Image Analysis*, 18(7):1247 – 1259, 2014.

[46] Nasr Makni, Nacim Betrouni, and Olivier Colot. Introducing spatial neighbourhood in evidential c-means for segmentation of multi-source images: Application to prostate multi-parametric {MRI}. *Information Fusion*, 19:61 – 72, 2014. Special Issue on Information Fusion in Medical Image Computing and Systems.

[47] Kuang Zhou, Arnaud Martin, Quan Pan, and Zhun ga Liu. Median evidential c-means algorithm and its application to community detection. *Knowledge-Based Systems*, 74:69 – 88, 2015.

APPENDIX A: COPYRIGHT CLEARANCE FORMS

Below is the permission for Abstract and Chapters 1, 2, 3, 4, 5, 6 and 7.

Rightslink® by Copyright Clearance Center https://s100.copyright.com/AppDispatchServlet#formTop

Copyright Clearance Center RightsLink® Home Create Account Help

IEEE Requesting permission to reuse content from an IEEE publication

Title: Dempster-Shafer theory of evidence in Single Pass Fuzzy C Means
Conference Proceedings: Fuzzy Systems (FUZZ), 2013 IEEE International Conference on
Author: Alireza Chakeri
Publisher: IEEE
Date: July 2013
Copyright © 2013, IEEE

LOGIN
If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already a RightsLink user or want to learn more?

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK CLOSE WINDOW

Copyright © 2016 Copyright Clearance Center, Inc. All Rights Reserved. Privacy statement. Terms and Conditions.
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

1 of 1 10/24/2016 10:37 AM