About Some Clustering Algorithms in Evidence Theory

Alexander Lepskiy

Abstract The Dempster–Shafer theory of evidence considers data that have a frequency-set nature (the so-called body of evidence). In recent years, there has been interest in clustering such objects to approximate them with simpler bodies of evidence, to analyze the inconsistency of information, reducing the computational complexity of processing algorithms, revealing the structure of the set of focal elements, etc. The article discusses some existing algorithms for clustering evidence bodies and suggests some new algorithms and approaches in such clustering.

1 Introduction

The Dempster–Shafer evidence theory [2, 15] considers data that is represented by a pair of objects $F = (\mathcal{A}, m)$, where $\mathcal{A}$ is the set of non-empty subsets (focal elements) of some base set X, m is a non-negative numerical function of sets (mass function) defined on the set of all subsets of the base set. The focal element $A \in \mathcal{A}$ describes the membership set of the true alternative $x \in A$ (for example, the air temperature forecast), and the mass $m(A)$ of this focal element A specifies the degree of belief that $x \in A$. Some set functions are put into one-to-one correspondence with the body of evidence. For example, there are such functions as belief and plausibility functions, which can be considered as the lower and upper bounds of the probability measure.

The tools for aggregating information presented by bodies of evidence, considering the reliability of information sources, their inconsistency and inaccuracy, are widely developed in the theory of evidence. However, many of the evidence body processing operations are computationally complex. In addition, it is required to reveal the enlarged structure of the set of focal elements in several problems, to

Alexander Lepskiy
National Research University Higher School of Economics, 20 Myasnitskaya Ulitsa, Moscow, 101000, Russia, e-mail: alepskiy@hse.ru

analyze the degree of homogeneity of the body of evidence, its internal inconsistency, etc. Therefore, there is a need to approximate complex evidence bodies with many focal elements by simpler evidence bodies with a smaller number of focal elements. Both an approximation of a set function (for example, a belief function) corresponding to the body of evidence by another set function from a given class, and an approximation based on clustering of a set of focal elements are considered. Pignistic probability is an example of the first type of approximation [16]. Below we consider only an approximation based on the clustering of the set of focal elements.

Evidential data have their own frequency-set specifics. Therefore, direct analogues of the well-known clustering algorithms for 'point' data either need deep modernization or additional interpretability.

This article will analyze some modern methods for clustering bodies of evidence. The article is of an overview and methodological nature, but it will consider a new method, which is an analogue of the k-means method for evidence bodies.

2 Necessary Information from Evidence Theory

Let X be some finite (for simplicity) basic set, 2^X be the set of all subsets from X. Let us consider some subset of non-empty sets (focal elements) $\mathcal{A}$ from 2^X and a non-negative set function (mass function) $m : 2^X \to [0,1]$ that satisfies the conditions: $m(A) > 0 \Leftrightarrow A \in \mathcal{A}$, $\sum_{A\in\mathcal{A}} m(A) = 1$. A pair $F = (\mathcal{A}, m)$ is called a body of evidence. Let $\mathcal{F}(X)$ be the set of all evidence bodies on X.

There is a one-to-one correspondence between the body of evidence $F = (\mathcal{A}, m)$ and the belief function $Bel(A) = \sum_{B\subseteq A} m(B)$ or the plausibility function $Pl(A) = \sum_{B\cap A\neq\emptyset} m(B)$, which can be considered as lower and upper bounds for the probability $P(A)$, respectively. The following special cases of evidence bodies are distinguished:

1) a categorical body of evidence of the form $F_A = (\{A\}, 1)$, i.e., a non-empty set A is the only focal element with unit mass;

2) a vacuous body of evidence $F_X = (\{X\}, 1)$.

An arbitrary body of evidence $F = (\mathcal{A}, m)$ can be represented as $F = \sum_{A\in\mathcal{A}} m(A) F_A$.

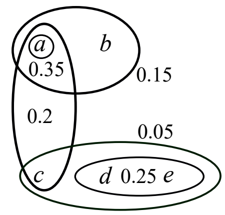
The body of evidence of the type $F_A^\alpha = \alpha F_A + (1-\alpha) F_X$ is called simple.

The body of evidence $F = (\mathcal{A}, m)$ on X can be represented as a weighted hypergraph with a set of vertices X, a set of hyperedges $\mathcal{A}$ and their weights $m(A)$, $A \in \mathcal{A}$.

Example 1. Let we have $X = \{a, b, c, d, e\}$ and the body of evidence $F = 0.35F_{\{a\}} + 0.15F_{\{a,b\}} + 0.2F_{\{a,c\}} + 0.25F_{\{d,e\}} + 0.05F_{\{c,d,e\}}$ is given on X, i.e. $\mathcal{A} = \{\{a\}, \{a, b\}, \{a, c\}, \{d, e\}, \{c, d, e\}\}$. The hypergraph of the evidence body F is shown in Fig. 1. □

If two sources of information are represented by the bodies of evidence $F_1 = (\mathcal{A}_1, m_1)$ and $F_2 = (\mathcal{A}_2, m_2)$ on X, then the degree of conflict (contradiction) between these sources can be assessed using some functional (measure of external

Fig. 1 Evidence body hypergraph



conflict) [10] $Con : \mathcal{F}(X)\times\mathcal{F}(X) \to [0,1]$, which takes on greater values the more pairs of non-overlapping (or 'weakly over-lapping') focal elements of two evidence bodies with large masses exist. The classical measure of external conflict is [2]

$$Con(F_1, F_2) = \sum_{A\cap B=\emptyset} m_1(A)m_2(B),$$

which we will use below. In addition to the measure of external conflict, the measure of internal conflict $Con_{in} : \mathcal{F}(X) \to [0,1]$ of one body of evidence is also considered [11]. The ability to evaluate internal conflict is one possible application of evidence body clustering (see Remark 2 below).

3 Basic Approaches for Clustering Body of Evidence

The clustering of the body of evidence $F = (\mathcal{A}, m)$ is primarily related to the clustering of the set of its focal elements $\mathcal{A}$. There are two formulations of the problem of clustering a set of focal elements.

1. It is required to find such a subset of $\mathcal{A}' \subseteq 2^X$ that would be 'close' to $\mathcal{A}$ in some sense, but $|\mathcal{A}'| \ll |\mathcal{A}|$. The new mass function $m'(A)$, is found either by a local redistribution of the masses $m(B)$ of the sets B involved in the formation of a new focal element $A \in \mathcal{A}'$, or by a global redistribution that minimizes the discrepancy functional between $F = (\mathcal{A}, m)$ and $F' = (\mathcal{A}', m')$.
2. It is required to find such a partition (or cover) of the set $\mathcal{A}$ of focal elements into subsets (clusters) $\{\mathcal{A}_1, \ldots, \mathcal{A}_l\}$ that would correspond in some sense to the structure of the set $\mathcal{A}$.

The first type of clustering is used to reduce the computational complexity of algorithms for processing evidence bodies or solving other approximation problems. The second type of clustering is used to identify the structure of a set of focal elements, to estimate the degree of heterogeneity, inconsistency, etc.

Next, we consider some implementations of clustering of these two types, namely:

1) hierarchical clustering;

2) clustering based on the density function of the distribution of conflict focal elements;

3) clustering based on conflict optimization (including an analogue of the k-means method for evidence bodies).

3.1 Hierarchical Inner and Outer Clustering

The simplest approximation procedure by clustering was proposed in [12], where 'close' focal elements or focal elements with small masses were combined. In this case, the masses of the combined focal elements were summed up. A more complex clustering scheme, which is analogous to divisional-agglomerative algorithms [13] in some sense, has been proposed in [7] and [14, 3]. Two clusterings are the result of this algorithm. One of them is internal in the form of evidence body $F^{-} = (\mathcal{A}^{-}, m^{-})$, the other is external in the form $F^{+} = (\mathcal{A}^{+}, m^{+})$. The set of focal elements $\mathcal{A}^{-}$ of internal clustering is the intersection of some sets from $\mathcal{A}$. While the set of focal elements $\mathcal{A}^{+}$ of external clustering is the union of some sets from $\mathcal{A}$. The masses of focal elements that are united in $\mathcal{A}^{+}$ or intersect in $\mathcal{A}^{-}$ are summarized: $m^{-}(B) = \sum_A m(A)$ if $B = \bigcap A \in \mathcal{A}^{-}$ and $m^{+}(C) = \sum_A m(A)$ if $C = \bigcup A \in \mathcal{A}^{+}$. In this case, such a pair (A, B) of focal elements is chosen for union/intersection, which delivers the minimum increment of the measure of imprecision [5] $f(F) = \sum_{A\in\mathcal{A}} m(A)\,|A|$.

The increments of this measure at the union/intersection of two sets and will be equal

$$\delta_{\cup}(C, D) = (m(C) + m(D))\,|C \cup D| - m(C)\,|C| - m(D)\,|D|$$

and

$$\delta_{\cap}(C, D) = m(C)\,|C| + m(D)\,|D| - (m(C) + m(D))\,|C \cap D|\,,$$

respectively. Therefore, the algorithm unites (intersects) those focal elements step by step, which deliver the minimum to the functional $\delta_{\cup}(C, D)$ ($\delta_{\cap}(C, D)$) at $C \neq D$. These procedures are repeated until a predetermined number $l < |\mathcal{A}|$ of focal elements remains, or some proximity condition between the original body of evidence $F = (\mathcal{A}, m)$ and its clustering is satisfied. As a result of such clustering, bodies of evidence F^{-} and F^{+} are obtained, which in the theory of belief functions are called specialization and generalization of the body of evidence F, respectively [4]. Thus, the algorithm for hierarchical inner and outer clustering will be as follows.

Algorithm 1.

Input data: body of evidence $F = (\mathcal{A}, m)$, the number of focal elements in clustering l.

Output data: bodies of evidence $F^{-} = (\mathcal{A}^{-}, m^{-})$ and $F^{+} = (\mathcal{A}^{+}, m^{+})$.

1. Let $F^{-} = F^{+} = F$.

2. Let's find the pairs $(A^{-}, B^{-}) = \arg\min_{C\neq D}\delta_{\cap}(C, D)$ and $(A^{+}, B^{+}) = \arg\min_{C\neq D}\delta_{\cup}(C, D)$ in $\mathcal{A}^{-}$ and $\mathcal{A}^{+}$, respectively. Let's replace a pair (A^{-}, B^{-}) with a set $A^{-} \cap B^{-}$ in $\mathcal{A}^{-}$, and a pair (A^{+}, B^{+}) with a set $A^{+} \cup B^{+}$ in $\mathcal{A}^{+}$. We get new sets $\mathcal{A}^{-}$ and $\mathcal{A}^{+}$. Let's recalculate the masses: $m^{-}(A^{-} \cap B^{-}) \leftarrow m^{-}(A^{-}) + m^{-}(B^{-})$, $m^{+}(A^{+} \cup B^{+}) \leftarrow m^{+}(A^{+}) + m^{+}(B^{+})$, the masses of the remaining focal elements from $\mathcal{A}^{-}$ and $\mathcal{A}^{+}$ do not change.

3. Step 2 is repeated until $l < |\mathcal{A}^-| = |\mathcal{A}^+|$. □

Example 2. We have the following transformations of sets of focal elements for the outer and inner approximations of the body of evidence from Example 1 and $l = 2$, respectively (pairs of merged/intersected focal elements at each step are marked in bold):

$$\mathcal{A} = \{\{a\}, \{a, b\}, \{a, c\}, \{\boldsymbol{d}, \boldsymbol{e}\}, \{\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{e}\}\} \to \{\{\boldsymbol{a}\}, \{\boldsymbol{a}, \boldsymbol{b}\}, \{a, c\}, \{c, d, e\}\} \to$$

$$\to \{\{\boldsymbol{a}, \boldsymbol{b}\}, \{\boldsymbol{a}, \boldsymbol{c}\}, \{c, d, e\}\} \to \{\{a, b, c\}, \{c, d, e\}\} = \mathcal{A}^+,$$

$$\mathcal{A} = \{\{a\}, \{a, b\}, \{a, c\}, \{\boldsymbol{d}, \boldsymbol{e}\}, \{\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{e}\}\} \to \{\{\boldsymbol{a}\}, \{a, b\}, \{\boldsymbol{a}, \boldsymbol{c}\}, \{d, e\}\} \to$$

$$\to \{\{\boldsymbol{a}\}, \{\boldsymbol{a}, \boldsymbol{b}\}, \{d, e\}\} \to \{\{a\}, \{d, e\}\} = \mathcal{A}^-.$$

We obtain the outer and inner approximations of the evidence body F, respectively $F^+ = 0.7F_{\{a,b,c\}} + 0.3F_{\{c,d,e\}}$ and $F^- = 0.7F_{\{a\}} + 0.3F_{\{d,e\}}$ (see Fig.2). □

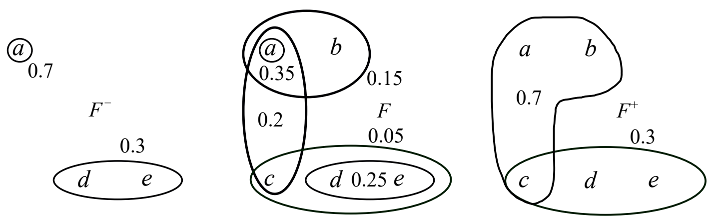


Fig. 2 Inner and outer clustering

3.2 Clustering Based on Conflict Density Distribution

Another approach to clustering is to find a small (by cardinality) subset $\mathcal{A}' \subseteq \mathcal{A}$ of 'significant' focal elements. What characteristics of focal elements can be considered significant? These can be the mass of the focal element, its cardinality (a measure in the case of a measurable X), the number of other focal elements that intersect with the given one, etc. In [1], these characteristics were combined in the concept of the density of distribution of conflict focal elements. Non-overlapping focal elements are called conflicting.

A function $\psi_F : 2^X \to [0, 1]$ is called the conflict density distribution of the evidence body $F = (\mathcal{A}, m)$ if it satisfies the conditions:

1. $\psi_F(A) = 0$ if$B \cap A \neq \emptyset \ \forall B \in \mathcal{A}$;
2. $\psi_F(A) = 1$ if $B \cap A = \emptyset \ \forall B \in \mathcal{A}$;
3. $\psi_{\alpha F_1 + \beta F_2} = \alpha \psi_{F_1} + \beta \psi_{F_2} \ \forall F_1, F_2 \in \mathcal{F}(X)$, where $\alpha + \beta = 1$, $\alpha \geq 0$, $\beta \geq 0$.

It can be shown that the conflict density function will be equal to $\psi_F(A) = \sum_{B: A \cap B = \emptyset} m(B) = 1 - Pl(A)$. 'Significant' focal elements in [1] were those that maximize the function $\varphi_F(A) = m(A)\psi_F(A)$, $A \in \mathcal{A}$. The distance between the selected focal elements was another characteristic that was considered in [1] when choosing elements for $\mathcal{A}' \subseteq \mathcal{A}$. This distance should not be too small. Thus, the set $\mathcal{A}' \subseteq \mathcal{A}$ will consist of sets that provide a large value of the function φ_F and are located at a sufficiently large distance from each other. Let $d(A, B)$ be a metric on the set of focal elements. Then the algorithm for choosing the set $\mathcal{A}' \subseteq \mathcal{A}$ will be as follows.

Algorithm 2.

Input data: body of evidence $F = (\mathcal{A}, m)$, the minimum possible value $h_1 > 0$ of $\varphi_F(A)$ for every $A \in \mathcal{A}'$; the minimum possible distance $h_2 > 0$ between focal elements from $\mathcal{A}'$.

Output data: the body of evidence $\mathcal{A}' \subseteq \mathcal{A}$.

1. Let the set of focal elements be ordered in descending order of the function φ_F: $\varphi_F(A_1) \geq \varphi_F(A_2) \geq \ldots \geq \varphi_F(A_k)$. Put $\mathcal{A}' = \{A_1\}$, $s := 2$.

2. If $\varphi_F(A_s) \leq h_1$, then the end. Otherwise, go to step 3.

3. If $\min_{A\in\mathcal{A}'} d(A, A_s) > h_2$, then $\mathcal{A}' := \mathcal{A}' \cup \{A_s\}$, $s := s + 1$, go to step 2. □

The function $d(A, B) = |A\Delta B|$ can be used as a metric between focal elements, where Δ is the symmetric difference of sets. To take into account not only the mutual position of focal elements, but also their masses, one can use metrics on the set of all evidence bodies $\mathcal{F}(X)$ [9]. For example, if a certain metric ρ is chosen on $\mathcal{F}(X)$, then the metric on 2^X can be defined as $d(A, B) = \rho(F_A^{m(A)}, F_B^{m(B)})$, where $F_A^{m(A)}$, $F_B^{m(B)}$ are simple evidence bodies. For example, $F_A^{m(A)} = m(A)F_A + (1 - m(A))F_X$. In particular, the following metric that is popular in evidence theory [8] between evidence bodies $F_1 = (\mathcal{A}_1, m_1)$ and $F_2 = (\mathcal{A}_2, m_2)$ can be used:

$$\rho_J(F_1, F_2) = \sqrt{\frac{1}{2} \sum_{A,B\in 2^X\setminus\{\emptyset\}} s_{A,B}(m_1(A) - m_2(A))(m_1(B) - m_2(B))},$$

where $s_{A,B} = |A \cap B|/|A \cup B|$ is the Jaccard index. It is easy to see that $\rho_J(F_1, F_2) \in [0, 1]$ $\forall F_1, F_2 \in \mathcal{F}(X)$. It can be shown that then the metric $d_J(A, B) = \rho_J(F_A^{m(A)}, F_B^{m(B)})$ takes the form.

Lemma 1. $d_J^2(A, B) = (m(A) - m(B))^2 + m(A)m(B)\frac{|A\Delta B|}{|A\cup B|} -$ $(m(B) - m(A))\ \frac{|B|m(B)-|A|m(A)}{|X|}$.
In particular, if $m(A) = m(B) = m$, *then* $d_J(A, B) = m\sqrt{|A\Delta B|/|A \cup B|}$.

Note that Algorithm 2 can be considered as an evidential analogue of the popular 'point' the DBSCAN algorithm (Density Based Spatial Clustering of Applications with Noise, [6]).

Example 3. Algorithm 2 will give the following result for the evidence body from Example 1 using the metric d_J, $h_1 = 0.1$, $h_2 = 0.2$.

Step 1. $\varphi_F(\{a\}) = 0.105$, $\varphi_F(\{a,b\}) = 0.045$, $\varphi_F(\{a,c\}) = 0.05$, $\varphi_F(\{d,e\}) = 0.175$, $\varphi_F(\{c,d,e\}) = 0.025$. Therefore, the set of focal elements will be ordered as follows: $\mathcal{A} = \{\{d,e\},\{a\},\{a,c\},\{a,b\},\{c,d,e\}\}$, $\mathcal{A}' = \{\{d,e\}\}$, $s := 2$.

Step 2. $\varphi_F(\{a\}) = 0.105 > h_1 \Rightarrow$ go to step 3.

Step 3. $d_J(\{d,e\},\{a\}) \approx 0.317 > h_2 \Rightarrow \mathcal{A}' := \mathcal{A}' \cup \{\{a\}\} = \{\{d,e\},\{a\}\}$, $s := 3$.

Step 2.1. $\varphi_F(\{a,c\}) = 0.05 < h_1 \Rightarrow$ the end.

As a result, we get a new set of focal elements $\mathcal{A}' = \{\{d,e\},\{a\}\}$.

The general view of the body of evidence with the set of focal elements $\mathcal{A}'$ will be as follows $F'(x) = xF_{\{a\}} + (1-x)F_{\{d,e\}}$, $x \in [0,1]$. The masses of the focal elements of the body of evidence F' can be found from the condition of minimizing the distance between F and F'. For example, if we use the metric ρ_J, then the solution to the problem $\rho_J(F, F'(x)) \to \min$ will be as follows $x_0 = \frac{149}{240} \approx 0.62$. Then $F' = 0.62F_{\{a\}} + 0.38F_{\{d,e\}}$ and $\rho_J(F, F'(x_0)) = 0.196$. □

3.3 Clustering Based on Conflict Optimization

These methods are based on the assumption of the heterogeneity of those bodies of evidence that need clustering. This heterogeneity, in particular, may be a consequence of the aggregation in a given body of evidence $F = (\mathcal{A}, m)$ of information from different, sometimes contradictory, sources. In this case, it is required to find such a partition (or cover) of the set of focal elements $\mathcal{A}$ into subsets (clusters) $\{\mathcal{A}_1,\dots,\mathcal{A}_l\}$ in order to optimize intracluster or intercluster conflict.

If a certain subset $\mathcal{A}' \subseteq \mathcal{A}$ of focal elements is selected, then we will further consider the following local redistribution of masses from $\mathcal{A}$ to $\mathcal{A}'$ (and such a body of evidence will be denoted by $F(\mathcal{A}') = (\mathcal{A}', m')$): $m'(A) = m(A)\ \forall A \in \mathcal{A}'$, $m'(X) = 1 - \sum_{A\in\mathcal{A}'} m(A)$. In particular, if , then $F(\{A\}) = F_A^{m(A)} = m(A)F_A + (1-m(A))F_X$ (simple evidence).

Then the following clustering optimization problem can be formulated. It is required to find such a partition (or cover) of the set of focal elements $\mathcal{A}$ into subsets (clusters) $C = \{\mathcal{A}_1,\dots,\mathcal{A}_l\}$ in order to maximize the external conflict between evidence clusters: $Con(F(\mathcal{A}_1),\dots,F(\mathcal{A}_l)) \to \max$.

In the following algorithm, Algorithm 2 can be used to extract the set from l centers of new clusters $C = \{\mathcal{A}_1,\dots,\mathcal{A}_l\}$. The remaining focal elements from the set are redistributed among l clusters so that $Con(F(\mathcal{A}_1),\dots,F(\mathcal{A}_l)) \to \max$.

Algorithm 3.

Input data: body of evidence $F = (\mathcal{A}, m)$, a selected small set $\mathcal{A}' = \{A_1,\dots,A_l\}$ of l focal elements that will be the centers of new clusters.

Output data: partition (cover) $C = \{\mathcal{A}_1,\dots,\mathcal{A}_l\}$ of the set of all focal elements $\mathcal{A}$.

1. Let $\mathcal{A}_i^{(0)} = \{A_i\}$, $i = 1,\dots,l$.

2. Focal elements from A are redistributed among clusters $\mathcal{A}_1^{(0)},\dots,\mathcal{A}_l^{(0)}$ according to the principle of conflict maximization between evidence clusters. The focal element $B\in\mathcal{A}\backslash\left\{\mathcal{A}_1^{(0)},\dots,\mathcal{A}_l^{(0)}\right\}$ will be assigned to that cluster $\mathcal{A}_i^{(0)}$ for which the maximum conflict measure is reached:

$$\mathcal{A}_i^{(0)}=\underset{j:B\in\mathcal{A}_j^{(0)}}{\arg\max}\, Con\left(F\left(\mathcal{A}_1^{(0)}\right),\dots,F\left(\mathcal{A}_j^{(0)}\cup\{B\}\right),\dots,F\left(\mathcal{A}_l^{(0)}\right)\right).$$

If equal maximum conflict values are obtained when assigning the element B to several clusters $\mathcal{A}_j^{(0)}$, $j\in J$, then this element B is included in all these clusters, and the mass value $m(B)$ is evenly distributed over the updated clusters, i.e. element B will be included in each cluster $\mathcal{A}_j^{(0)}$, $j\in J$ with weight $m(B)/|J|$. □

Example 4. Let's redistribute the remaining focal elements $\mathcal{A}\backslash\mathcal{A}'=\{\{a,b\},\{a,c\},\{c,d,e\}\}$ in accordance with Algorithm 4 for the body of evidence from Example 1 and the set of focal elements $\mathcal{A}'=\{\{d,e\},\{a\}\}$ selected in Example 3.

Step 1. $\mathcal{A}_1^{(0)}=\{\{d,e\}\}$, $\mathcal{A}_2^{(0)}=\{\{a\}\}$.

Step 2. Let $B=\{a,b\}$. If $B\in\mathcal{A}_1$, then we obtain:

$F\left(\{B\}\cup\mathcal{A}_1^{(0)}\right)=0.15F_{\{a,b\}}+0.25F_{\{d,e\}}+0.6F_X$,

$F\left(\mathcal{A}_2^{(0)}\right)=0.35F_{\{a\}}+0.65F_X$.

Then $Con\left(F\left(\{B\}\cup\mathcal{A}_1^{(0)}\right),F\left(\mathcal{A}_2^{(0)}\right)\right)=0.25\cdot 0.35=0.0875$.

If $B\in\mathcal{A}_2$, then we obtain:

$F\left(\mathcal{A}_1^{(0)}\right)=0.25F_{\{d,e\}}+0.75F_X$,

$F\left(\{B\}\cup\mathcal{A}_2^{(0)}\right)=0.35F_{\{a\}}+0.15F_{\{a,b\}}+0.5F_X$

and $Con\left(F\left(\mathcal{A}_1^{(0)}\right),F\left(\{B\}\cup\mathcal{A}_2^{(0)}\right)\right)=0.125$.

Thus, the element $B=\{a,b\}$ will be assigned to the cluster $\mathcal{A}_2$.

Let $B=\{a,c\}$. If $B\in\mathcal{A}_1$, then we obtain:

$F\left(\{B\}\cup\mathcal{A}_1^{(0)}\right)=0.2F_{\{a,c\}}+0.25F_{\{d,e\}}+0.55F_X$,

$F\left(\mathcal{A}_2^{(0)}\right)=0.35F_{\{a\}}+0.65F_X$.

Then $Con\left(F\left(\{B\}\cup\mathcal{A}_1^{(0)}\right),F\left(\mathcal{A}_2^{(0)}\right)\right)=0.25\cdot 0.35=0.0875$.

If $B\in\mathcal{A}_2$, then we obtain:

$F\left(\mathcal{A}_1^{(0)}\right)=0.25F_{\{d,e\}}+0.75F_X$,

$F\left(\{B\}\cup\mathcal{A}_2^{(0)}\right)=0.35F_{\{a\}}+0.2F_{\{a,c\}}+0.45F_X$

and $Con\left(F\left(\mathcal{A}_1^{(0)}\right),F\left(\{B\}\cup\mathcal{A}_2^{(0)}\right)\right)=0.1375$.

Thus, the element $B=\{a,c\}$ will be assigned to the cluster $\mathcal{A}_2$.

Let $B=\{c,d,e\}$. If $B\in\mathcal{A}_1$, then we obtain:

$F\left(\{B\}\cup\mathcal{A}_1^{(0)}\right)=0.25F_{\{d,e\}}+0.05F_{\{c,d,e\}}+0.7F_X$,

$F\left(\mathcal{A}_2^{(0)}\right)=0.35F_{\{a\}}+0.65F_X$.

Then $Con\left(F\left(\{B\}\cup\mathcal{A}_1^{(0)}\right), F\left(\mathcal{A}_2^{(0)}\right)\right) = 0.105$.
If $B \in \mathcal{A}_2$, then we obtain:
$F\left(\mathcal{A}_1^{(0)}\right) = 0.25F_{\{d,e\}} + 0.75F_X$,
$F\left(\{B\}\cup\mathcal{A}_2^{(0)}\right) = 0.35F_{\{a\}} + 0.05F_{\{c,d,e\}} + 0.6F_X$
and $Con\left(F\left(\mathcal{A}_1^{(0)}\right), F\left(\{B\}\cup\mathcal{A}_2^{(0)}\right)\right) = 0.0875$.
Thus, the element $B = \{a, c\}$ will be assigned to the cluster $\mathcal{A}_1$.
Thus, we get a partition $C = \{\mathcal{A}_1, \mathcal{A}_2\}$, where $\mathcal{A}_1 = \{\{d, e\}, \{c, d, e\}\}$, $\mathcal{A}_2 = \{\{a\}, \{a, b\}, \{a, c\}\}$. □

Another variant of the optimization problem of evidence body clustering will be considered below. It is required to find such a partition (or cover) of the set of focal elements $\mathcal{A}$ into subsets (clusters) $C = \{\mathcal{A}_1, \dots, \mathcal{A}_l\}$ in order to minimize the total internal conflict within evidence clusters: $\Phi = \sum_{i=1}^{l} Con_{in}(F(\mathcal{A}_i)) \to \min$, where Con_{in} is a measure of internal conflict. The total external conflict $Con_{in}(F(\mathcal{A}_i)) = \sum_{B\in\mathcal{A}_i} Con(F(\{B\}, C_i))$ between each body of evidence $F(\{B\})$, $B \in \mathcal{A}_i$ and some reference evidence (center) C_i of the i-th cluster can be considered as an internal conflict by analogy with the classical k-means algorithm

We will assume that center C_i has the form

$$C_i = \sum_{A\in\mathcal{A}_i} \alpha_i(A) F_A, \quad (1)$$

where $\alpha_i = (\alpha_i(A))_{A\in\mathcal{A}_i} \in S_{|\mathcal{A}_i|}$, $S_k = \left\{(t_1, ..., t_k) : t_i \geq 0, i = 1, ..., k, \sum_{i=1}^{k} t_i = 1\right\}$ is an k-dimensional simplex. The following theorem is true.

Theorem 1 *Let* $Pl_{\mathcal{A}_i}(A) = \sum_{\substack{B\in\mathcal{A}_i:\\ A\cap B\neq\emptyset}} m(B)$ *be the restriction of the plausibility function to the set* $\mathcal{A}_i$. *Then the minimum of the functional* Φ *for a fixed cover* $C = \{\mathcal{A}_1, \dots, \mathcal{A}_l\}$ *will be achieved at*

$$\alpha_i = (\alpha_i(A))_{A\in\overline{\mathcal{A}_i}} \in S_{\left|\overline{\mathcal{A}_i}\right|}, \quad i = 1, \dots, l, \quad (2)$$

where $\overline{\mathcal{A}_i} = \left\{A \in \mathcal{A}_i : A = \arg\max_{A\in\mathcal{A}_i} Pl_{\mathcal{A}_i}(A)\right\}$.

Then the clustering algorithm (analogous to k-means) will be as follows.

Algorithm 4.

Input data: body of evidence $F = (\mathcal{A}, m)$; number of clusters l; initial centers of clusters – bodies of evidence $C_i^{(0)}$, $i = 1, ..., l$; maximum conflict threshold within clusters $Con_{\max} \in [0, 1]$; $s = 0$.

Output data: partition (covering) $C = \{\mathcal{A}_1, \dots, \mathcal{A}_l\}$ of the set of all focal elements $\mathcal{A}$.

1. Focal elements are redistributed among clusters according to the principle of minimizing the conflict between evidence clusters and cluster centers. The focal element $B \in \mathcal{A}$ refers to the cluster $\mathcal{A}_i^{(s)}$ for which is achieved $\min_i Con\left(F(\{B\}), C_i^{(s)}\right)$ $\le Con_{\max}$. If it is true that $\min_i Con\left(F(\{B\}), C_i^{(s)}\right) > Con_{\max}$, then the focal element B is assigned as the center of the new cluster. As a result, clusters $\mathcal{A}_i^{(s)}$, $i = 1, ..., l$ are obtained.

2. New cluster centers are calculated using formulas (1), (2), $s \leftarrow s + 1$.

3. Steps 1 and 2 are repeated until the clusters (or their centers) stabilize. □

Corollary 1. *Algorithm 4 converges in a finite number of steps.*

Example 5. Algorithm 4 will give the following result for the evidence body from Example 1. Let $l = 2$ be set, and the initial centers of the clusters coincide with the focal elements identified by Algorithm 2: $C_1^{(0)} = F(\{d, e\}) = 0.25F_{\{d,e\}} + 0.75F_X$, $C_2^{(0)} = F(\{a\}) = 0.35F_{\{a\}} + 0.65F_X$; $Con_{\max} = 1$; $s = 0$.

Step 1.1. We have

$Con\left(F(\{a\}), C_1^{(0)}\right) = 0.0875$, $Con\left(F(\{a, b\}), C_1^{(0)}\right) = 0.0375$,

$Con\left(F(\{a, c\}), C_1^{(0)}\right) = 0.05$,

$Con\left(F(\{d, e\}), C_1^{(0)}\right) = Con\left(F(\{c, d, e\}), C_1^{(0)}\right) = 0$,

$Con\left(F(\{a\}), C_2^{(0)}\right) = Con\left(F(\{a, b\}), C_2^{(0)}\right) = Con\left(F(\{a, c\}), C_2^{(0)}\right) = 0$,

$Con\left(F(\{d, e\}), C_2^{(0)}\right) = 0.0875$, $Con\left(F(\{c, d, e\}), C_2^{(0)}\right) = 0.0175$

Then the initial clustering will have the form according to the principle of minimizing the conflict between evidence clusters and cluster centers: $\mathcal{A}_1^{(0)} = \{\{d, e\}, \{c, d, e\}\}$, $\mathcal{A}_2^{(0)} = \{\{a\}, \{a, b\}, \{a, c\}\}$.

Step 1.2. New cluster centers are calculated using formulas (1), (2):

$Pl_{\mathcal{A}_1^{(0)}}(\{d, e\}) = Pl_{\mathcal{A}_1^{(0)}}(\{c, d, e\}) = 0.3$,

$Pl_{\mathcal{A}_2^{(0)}}(\{a\}) = Pl_{\mathcal{A}_2^{(0)}}(\{a, b\}) = Pl_{\mathcal{A}_2^{(0)}}(\{a, c\}) = 0.7$.

Therefore

$C_1^{(1)} = \alpha F_{\{d,e\}} + (1 - \alpha)F_{\{c,d,e\}}$, $C_2^{(1)} = \beta F_{\{a\}} + \gamma F_{\{a,b\}} + (1 - \beta - \gamma)F_{\{a,c\}}$, where $\alpha, \beta, \gamma \in [0, 1]$, $\beta + \gamma \le 1$.

Step 2.1. Focal elements are redistributed:

$Con\left(F(\{a\}), C_1^{(1)}\right) = 0.35$, $Con\left(F(\{a, b\}), C_1^{(1)}\right) = 0.15$,

$Con\left(F(\{a, c\}), C_1^{(1)}\right) = 0.2\alpha$,

$Con\left(F(\{d, e\}), C_1^{(1)}\right) = Con\left(F(\{c, d, e\}), C_1^{(1)}\right) = 0$,

$Con\left(F(\{a\}), C_2^{(1)}\right) = Con\left(F(\{a, b\}), C_2^{(1)}\right) = Con\left(F(\{a, c\}), C_2^{(1)}\right) = 0$,

$Con\left(F(\{c, d, e\}), C_2^{(1)}\right) = 0.05(\beta + \gamma)$, $Con\left(F(\{d, e\}), C_2^{(1)}\right) = 0.25$.

Then $\mathcal{A}_1^{(1)} = \{\{d, e\}, \{c, d, e\}\}$, $\mathcal{A}_2^{(1)} = \{\{a\}, \{a, b\}, \{a, c\}\}$. The clusters have stabilized. □

Remark 1. Since cluster centers may depend on parameters $\alpha = (\alpha(A))_{A\in\overline{\mathcal{A}_i}} \in S_{\left|\overline{\mathcal{A}_i}\right|}$ (see formula (2)), additional procedures for choosing these parameters can be used in the algorithm, such as:

1) cover minimization $C = \{\mathcal{A}_1, \ldots, \mathcal{A}_l\}$. For example, $\sum_{i=1}^{l} |\mathcal{A}_i| \to \min$.

2) minimizing the uncertainty of evidence bodies C_i, $i = 1, ..., l$. For example, (measure of imprecision [5]) $H(C_i) = \sum_{A\in\overline{\mathcal{A}_i}} \alpha_i(A)\,|A| \to \min$.

3) minimizing the distance between the centers of clusters and the original body of evidence with respect to some metric ρ: $\rho(C_i, F) \to \min$, $i = 1, ..., l$; etc.

Remark 2. Clustering a body of evidence $F = (\mathcal{A}, m)$ can be used to evaluate its internal conflict. If $C = \{\mathcal{A}_1, \ldots, \mathcal{A}_l\}$ is a cover (or partition) of the set of focal elements $\mathcal{A}$, then the internal conflict can be estimated by the formula $Con_{in}(F) = Con(F(\mathcal{A}_1), \ldots, F(\mathcal{A}_l))$. So, the measure of internal conflict of the body of evidence from Example 1 using the clustering of Example 5 (or Example 4) will be equal to $Con_{in}(F) = Con(F(\{d, e\}, \{c, d, e\}_1), F(\{a\}, \{a, b\}, \{a, c\})) = 0.2$.

4 Conclusion

The article discusses the main known and currently being developed areas of evidence body clustering. In particular, the following classes of algorithms are considered: a) hierarchical clustering algorithms; b) clustering algorithms based on the density function; c) clustering algorithms based on conflict optimization.

On the one hand, many of the considered algorithms are analogues of the corresponding algorithms for ”point” data. On the other hand, the dual frequency-multiple nature of the bodies of evidence imposes peculiar restrictions, the need to use “one’s own” measures of proximity (for example, based on measures of conflict), etc. Some algorithms (for example, hierarchical ones) are explained by the peculiar goals of such clustering (for example, generating generalizations and specializations of the body of evidence).

All these features leave a lot of room for creativity in the development of algorithms for clustering bodies of evidence.

Acknowledgements The financial support from the Government of the Russian Federation within the framework of the implementation of the 5-100 Programme Roadmap of the National Research University Higher School of Economics is acknowledged.

References

1. Bronevich, A., Lepskiy, A.: Measures of conflict, basic axioms and their application to the clusterization of a body of evidence. Fuzzy Sets and Systems **446**, 277–300 (2022)
2. Dempster, A.P.: Upper and lower probabilities induced by multivalued mapping. Ann. Math. Statist. **38**, 325–339 (1967)

3. Denœux, T.: Inner and outer approximation of belief structures using a hierarchical clustering approach. Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems **9(4)**, 437–460 (2001)
4. Dubois, D., Prade, H.: A Set-Theoretic View on Belief Functions: Logical Operations and Approximations by Fuzzy Sets. Int. J. of General Systems **12**, 193–226 (1986)
5. Dubois, D., Prade, H.: Consonant approximations of belief measures. Int. J. of Approximate Reasoning **4** 419–449 (1990)
6. Ester, M., Kriegel, H-P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proc. of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96), pp. 226–231. AAAI Press (1996)
7. Harmanec, D.: Faithful approximations of belief functions. In: Laskey, K. B. and Prade, H., (eds.) Uncertainty in Artificial Intelligence 15 (UAI99), Stockholm, Sweden, (1999)
8. Jousselme, A.-L., Grenier, D., Bossé, É.: A new distance between two bodies of evidence. Information Fusion **2**, 91–101 (2001)
9. Jousselme, A.-L., Maupin, P.: Distances in Evidence Theory: Comprehensive Survey and Generalizations. Int. J. of Approximate Reasoning **53**, 118–145 (2012)
10. Lepskiy, A.: Analysis of Information Inconsistency in Belief Function Theory. Part I: External Conflict. Control Sciences **5**, 2–16 (2021)
11. Lepskiy, A.: Analysis of Information Inconsistency in Belief Function Theory. Part II: Internal Conflict. Control Sciences 6, 2–12 (2021)
12. Lowrance, J.D, Garvey, T.D., Strat, T.M.: A framework for evidential reasoning systems. In: Kehler, T. et al. (eds.), proceedings of AAAI'86, vol 2, pp.896–903, Philadelphia, August (1986)
13. Mirkin, B.: Core Data Analysis: Summarization, Correlation, and Visualization. Springer, Cham, Switzerland, (2019)
14. Petit-Renaud, S., Denœux, T.: Handling different forms of uncertainty in regression analysis: a fuzzy belief structure approach. In: Hunter, A. and Pearsons, S., (eds.) Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU'99), pp. 340–351. Springer Verlag (1999)
15. Shafer, G.: A mathematical theory of evidence. Princeton Univ. Press (1976)
16. Smets, P.: Decision Making in TBM: The Necessity of the Pignistic Transformation. Int. J. of Approximate Reasoning. **38** 133–147 (2005)