Cluster Decomposition of the Body of Evidence ⋆

Alexander Lepskiy[1] [0000−0002−1051−2857]

Higher School of Economics, 20 Myasnitskaya Ulitsa, Moscow, 101000, Russia,
alepskiy@hse.ru
https://www.hse.ru/en/org/persons/10586209

Abstract. Two algorithms for the body of evidence clustering are developed and studied in this paper. The first algorithm is based on the use of the distribution density function of conflicting focal elements of the body of evidence. The second algorithm is similar to the k-means algorithm, but it uses the external conflict measure instead of the metric. It is shown that cluster decomposition can be used to evaluate the internal conflict of the body of evidence.

Keywords: Body of evidence · Measure of conflict · Decomposition of evidence.

1 Introduction

The body of evidence may have a complex structure in applied problems of the theory of belief functions. For example, it may consist of many focal elements with a complex intersection structure. Such evidence is difficult to interpret. In addition, since many of the operations of the theory of belief functions (for example, combined rules) are computationally difficult, applying these operations to evidence bodies with many focal elements also becomes computationally difficult.

Therefore, the following tasks are relevant: 1) analysis of the structure of the set of focal elements $\mathcal{A}$ of the body of evidence $F = (\mathcal{A}, m)$ (m is the mass function); 2) finding an enlarged (simplified) structure of the set of focal elements $\widetilde{\mathcal{A}}$; 3) redistribution of masses of focal elements of the set $\mathcal{A}$ to focal elements from $\widetilde{\mathcal{A}}$. As a result, we obtain a new mass function $\widetilde{m}$, etc.

The paper proposes to solve these problems based on the clustering of a set of focal elements. We suggest that the complex structure of the body of evidence may be the result of aggregation of heterogeneous information. This information, which is obtained from various sources, may be contradictory (conflict). Therefore, the general approach to clustering the body of evidence can be as follows. The inconsistency should be minimal within clusters and maximal between clusters in the resulting partition into clusters of the original set of focal elements.

⋆ 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.

This approach is similar to the compactness principle in cluster analysis. Distances should be minimal between elements of the same cluster and maximal between clusters. But in this article, we will use the measure of external conflict (contradiction) between evidence bodies [8] as a proximity functional between clusters instead of a metric, and/or the measure of internal conflict of evidence bodies [9] as the proximity functional of focal elements within a cluster.

The idea of approximating the body of evidence F by a simpler body of evidence $\widetilde{F}$ using hierarchical clustering of focal elements was proposed in [6, 10] and developed in [5]. Clustering was carried out by union and intersection of 'close' focal elements and summing their masses. A clustering algorithm based on the concept of conflict density was proposed in [2].

In the general case, we get a partition (or coverage) $\{\mathcal{A}_1, \dots, \mathcal{A}_l\}$ of the set of focal elements $\mathcal{A}$ as a result of its clustering, which can be associated with the set of bodies of evidence $\{F_1, ..., F_l\}$, where $F_i = (\mathcal{A}_i, m_i)$, $i = 1, ..., l$.

In addition to revealing the structure and simplifying the body of evidence, clustering can be used to evaluate the internal conflict $Con_{in}(F)$ of the original body of evidence $F = (\mathcal{A}, m)$ according to the formula $Con_{in}(F) = Con(F_1, ..., F_l)$, where Con is a measure of external conflict.

Two algorithms for clustering the body of evidence are proposed in the article. The first algorithm is based on the use of the conflict distribution density function. The second algorithm is analogous to the k-means algorithm, but instead of a metric, a measure of external conflict is used.

2 Basic Concepts of the Evidence Theory

Let us briefly recall the basic concepts of evidence theory [4, 12]. Let X be a finite set, 2^X be the set of all subsets on X. A body of evidence on the set X is a pair $F = (\mathcal{A}, m)$, where $\mathcal{A}$ is a set of non-empty subsets (focal elements) from the X, $m : 2^X \to [0, 1]$ is a mass function that satisfies the conditions: $m(A) > 0 \Leftrightarrow A \in \mathcal{A}$, $\sum_A m(A) = 1$. Let $\mathcal{F}(X)$ be the set of all bodies of evidence on the X.

Special cases of bodies of evidence are categorical evidence $F_A = (\{A\}, 1)$, vacuous evidence $F_X = (\{X\}, 1)$, simple evidence $F_A^\alpha = \alpha F_A + (1 - \alpha) F_X$, $\alpha \in [0, 1]$. An arbitrary body of evidence $F = (\mathcal{A}, m)$ can be represented in the form $F = \sum_{A \in \mathcal{A}} m(A) F_A$.

Some set functions are associated with the body of evidence $F = (\mathcal{A}, m)$: belief function $Bel(A) = \sum_{B \subseteq A} m(B)$, plausibility function $Pl(A) = 1 - Bel(\neg A) = \sum_{A \cap B \neq \emptyset} m(B)$, etc. These functions uniquely define the entire body of evidence.

Inconsistency (conflict) is an important joint characteristic of two or more bodies of evidence. The external conflict of two evidence bodies $F_1 = (\mathcal{A}_1, m_1)$ and $F_2 = (\mathcal{A}_2, m_2)$ is some measure $Con(F_1, F_2) : \mathcal{F}(X) \times \mathcal{F}(X) \to [0, 1]$, which takes on a greater value in the case of the existence of a large number of pairs (A, B) with large masses of non-overlapping focal elements of two bodies of evidence: $A \in \mathcal{A}_1$, $B \in \mathcal{A}_2$, $A \cap B = \emptyset$. An overview of articles on external conflict measures can be found in [8]. Below we will use the canonical measure of external conflict associated with the Dempster rule: $Con(F_1, F_2) = \sum_{A\cap B=\emptyset} m_1(A)m_2(B)$.

Along with the measure of external conflict between several bodies of evidence, the inconsistency of information provided by one body of evidence is also considered. An evidence body that results from combining multiple bodies of conflicting evidence bodies can have a large internal conflict. The inconsistency of the body of evidence $F = (\mathcal{A}, m)$ is evaluated using some internal conflict measure $Con_{in}(F) : \mathcal{F}(X) \to [0, 1]$ [9].

3 Evidence Clustering

3.1 Restriction and Extension of the Mass Function

Let $F = (\mathcal{A}, m)$ be the body of evidence, where $\mathcal{A}$ is the set of focal elements of this evidence. Let's consider some subset $\mathcal{A}' \subseteq \mathcal{A}$. The set function $m' : 2^X \to [0, 1]$, $m'(A) = m(A)\ \forall\ A \in \mathcal{A}'$ and $m'(A) = 0\ \forall\ A \notin \mathcal{A}'$ is called the restriction of the mass function m to the set $\mathcal{A}' \subseteq \mathcal{A}$.

In the general case, the mass function m' does not satisfy the normalization condition $\sum_A m'(A) = 1$. Therefore, the pair $(\mathcal{A}', m')$ does not define any body of evidence. It is necessary to extend the set function m' to the mass function $\tilde{m}'$ so that this extension reflects the distribution of the m'.

This can be done in many ways. Examples of some extensions:

1) proportional extension: $\tilde{m}'(A) = m'(A) \big/ \sum_{B\in\mathcal{A}'} m'(B)\ \forall\ A \in \mathcal{A}'$.

2) vacuous extension: $\tilde{m}'(A) = m'(A)$, $\tilde{m}'(X) = m'(X) + 1 - \sum_{B\in\mathcal{A}'} m'(B)$.

Note that various extensions of the set function to the mass function of some body of evidence are used in the combination rules. For example, proportional extension is used in Dempster's rule [4], and inconsistent continuation is used in Yager's rule [13].

If a certain rule for the extension of the mass function is fixed, then the new body of evidence $F' = (\mathcal{A}', \tilde{m}')$ will be uniquely determined by the original body of evidence $F = (\mathcal{A}, m)$ by the set $\mathcal{A}' \subseteq \mathcal{A}$. Therefore, such a body of evidence will be denoted as $F(\mathcal{A}') = (\mathcal{A}', \tilde{m}')$.

In particular, if the vacuous extension is used, then the body of evidence $F(\{A\}) = F_A^{m(A)} = m(A)F_A + (1 - m(A))F_X$ will be simple for any set $A \in \mathcal{A}$.

3.2 Statement of the Problem of Clustering the Body of Evidence Based on Conflict Optimization

Various formulations of the clustering problem are possible. Let's note some of them. Suppose we have a body of evidence $F = (\mathcal{A}, m)$. It is required to find a partition of the set of all focal elements $\mathcal{A}$ into subsets $\{\mathcal{A}_1, \ldots, \mathcal{A}_l\}$ such that:

1) maximize external conflict between bodies of evidence (clusters):
$Con(F(\mathcal{A}_1), \ldots, F(\mathcal{A}_l)) \to \max$, where Con is a measure of external conflict;

2) minimize total internal conflict within clusters
$\sum_{i=1}^{l} Con_{in}(F(\mathcal{A}_i)) \to \min$, where Con_{in} is a measure of internal conflict;

3) minimize the overall conflict between the centers of clusters and the bodies of evidence formed by the focal elements of these clusters $\sum_{i=1}^{l}\sum_{B\in\mathcal{A}_i} Con\left(F(\{B\}), C_i\right) \to \min$, where C_i is the reference body of evidence corresponding to the i-th cluster, $i = 1, ..., l$.

In a more general setting, it is required to find a covering of the set of focal elements $\mathcal{A}$ instead of a partition.

3.3 Cluster Decomposition of Evidence Based on the Conflict Density Function

Density Function. The concept of conflict density was introduced in [2]. Let $F = (\mathcal{A}, m)$ be the body of evidence. A mapping $\psi_F : 2^X \to [0,1]$ is called a conflict density function of the body of evidence F if it satisfies the following 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 is easy to show [2] that a set function satisfying conditions 1)-3) is equal to $\psi_F(A) = \sum_{B:A\cap B=\emptyset} m(B) = 1 - Pl(A)$. Note that the function was considered in [11] and was called the inconsistency function.

The main idea of the clustering algorithm for a set of focal elements $\mathcal{A}$ based on the conflict density function is that the 'centers' of the clusters should have a large value of the conflict density function.

We will use the function $\varphi_F(A) = m(A)\psi_F(A)$, $A \in \mathcal{A}$ instead of the density function ψ_F itself. The function φ_F will take on large values for those focal elements that have not only a high density, but also a large mass.

Algorithm for Cluster Decomposition of Evidence Based on Conflict Density Functions. This algorithm will consist of the following steps.

Algorithm 1.

1. Let's calculate the values of the set function $\varphi_F(A)$, $A \in \mathcal{A}$. If we have $\varphi_F(A) = 0$ for all $A \in \mathcal{A}$, then we stop the algorithm. In this case, we have non-conflict body of evidence $\mathcal{A}$: $B \cap A \neq \emptyset\ \forall A, B \in \mathcal{A}$. Therefore, there will be no clustering.

2. If there are $A \in \mathcal{A}$ for which $\varphi_F(A) > 0$, then we arrange such focal elements in descending order of function values φ_F: $\varphi_F(A_1) \geq \varphi_F(A_2) \geq ....$ We choose the number of clusters l by analyzing the rate of decrease of the sequence $\{\varphi_F(A_i)\}$. Selected focal elements will be initial clusters: $\mathcal{A}_i^{(0)} = \{A_i\}$, $i = 1, ..., l$.

3. The remaining focal elements are redistributed among clusters $\mathcal{A}_1^{(0)}, ..., \mathcal{A}_l^{(0)}$ according to the principle of maximizing the conflict between evidence clusters. We will assign a focal element $B \in \mathcal{A}\backslash\left\{\mathcal{A}_1^{(0)}, ..., \mathcal{A}_l^{(0)}\right\}$ to the cluster $\mathcal{A}_i^{(0)}$ for which the maximum conflict measure is reached:

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

If equal maximum values of the conflict are obtained by assigning B to several clusters $\mathcal{A}_j^{(0)}$, $j \in J$, then we include B in all these clusters, and the mass value $m(B)$ is evenly distributed among the updated clusters. In this case, B will be included in each cluster with weight $m(B)/|J|$. As a result, we obtain a coverage $\{\mathcal{A}_1, \dots, \mathcal{A}_l\}$ of the set of all focal elements $\mathcal{A}$. The values of the mass function m_i on the $\mathcal{A}_i$, $i = 1, ..., l$ are calculated using the given restriction and extension procedures.

Example 1. Let we have $X = \{1, 2, 3\}$ and the body of evidence $F = 0.3F_{\{1\}} + 0.2F_{\{2\}} + 0.3F_{\{1,3\}} + 0.2F_{\{2,3\}}$ is given on X, i.e. $\mathcal{A} = \{\{1\}, \{2\}, \{1, 3\}, \{2, 3\}\}$.

Step 1. Find the values of the function φ_F: $\varphi_F(\{1\}) = \varphi_F(\{2\}) = 0.12$, $\varphi_F(\{1, 3\}, \{2, 3\}) = 0.06$.

Step 2. Let us assign the number of clusters $l = 2$ and $\mathcal{A}_1^{(0)} = \{\{1\}\}$, $\mathcal{A}_2^{(0)} = \{\{2\}\}$.

Step 3. Let's distribute the remaining two focal elements among clusters.

We have for $B = \{1, 3\}$. If $B \in \mathcal{A}_1$, then $F\left(\{B\} \cup \mathcal{A}_1^{(0)}\right) = 0.3F_{\{1\}} + 0.3F_{\{1,3\}} + 0.4F_X$, $F\left(\mathcal{A}_2^{(0)}\right) = 0.2F_{\{2\}} + 0.8F_X$. Consequently $Con\left(F\left(\{B\} \cup \mathcal{A}_1^{(0)}\right), F\left(\mathcal{A}_2^{(0)}\right)\right) = 0.12$. But if $B \in \mathcal{A}_2$, then $F\left(\mathcal{A}_1^{(0)}\right) = 0.3F_{\{1\}} + 0.7F_X$, $F\left(\{B\} \cup \mathcal{A}_2^{(0)}\right) = 0.2F_{\{2\}} + 0.3F_{\{1,3\}} + 0.5F_X$ and $Con\left(F\left(\mathcal{A}_1^{(0)}\right), F\left(\{B\} \cup \mathcal{A}_2^{(0)}\right)\right) = 0.06$. Thus, we assign $B = \{1, 3\}$ to the cluster $\mathcal{A}_1$.

We have for focal element $B = \{2, 3\}$. If $B \in \mathcal{A}_1$, then $F\left(\{B\} \cup \mathcal{A}_1^{(0)}\right) = 0.3F_{\{1\}} + 0.2F_{\{2,3\}} + 0.5F_X$, $F\left(\mathcal{A}_2^{(0)}\right) = 0.2F_{\{2\}} + 0.8F_X$ and $Con\left(F\left(\{B\} \cup \mathcal{A}_1^{(0)}\right), F\left(\mathcal{A}_2^{(0)}\right)\right) = 0.06$. But if $B \in \mathcal{A}_2$, then $F\left(\mathcal{A}_1^{(0)}\right) = 0.3F_{\{1\}} + 0.7F_X$, $F\left(\{B\} \cup \mathcal{A}_2^{(0)}\right) = 0.2F_{\{2\}} + 0.2F_{\{2,3\}} + 0.6F_X$ and $Con\left(F\left(\mathcal{A}_1^{(0)}\right), F\left(\{B\} \cup \mathcal{A}_2^{(0)}\right)\right) = 0.12$. Thus, we assign $B = \{2, 3\}$ to the cluster $\mathcal{A}_2$ and we get the final focal element clustering $\mathcal{A}_1 = \{\{1\}, \{1, 3\}\}$, $\mathcal{A}_2 = \{\{2\}, \{2, 3\}\}$.

Remark 1. The distance between the selected focal elements can also be considered at step 2 of the algorithm in addition to calculating the conflict density (function φ_F), as was done in [2]. In this case, the focal elements are selected in descending order of the function φ_F. If $\mathcal{A}'$ is a set of already selected focal elements, then the next element A_k is added to this set, provided that $\min_{A \in \mathcal{A}'} d(F(\{A\}), F(\{A_k\})) > h$, where d is some metric on the set of evidence bodies [7], h is the threshold value.

3.4 The k-means algorithm for the body of evidence

Let $F = (\mathcal{A}, m)$ be the body of evidence. It is required to find such a coverage of the set of all focal elements $\mathcal{A}$ by subsets (clusters) $\mathcal{C} = \{\mathcal{A}_1, \dots, \mathcal{A}_l\}$ that would minimize intracluster conflict. We will use the concept of center of a set (cluster) of focal elements by analogy with the classical k-means algorithm. By the center of the i-th cluster $\mathcal{A}_i$, we mean some body of evidence C_i constructed from the pair $(\mathcal{A}_i, m_i)$, where m_i is the restriction of the mass function to $\mathcal{A}_i \subseteq \mathcal{A}$,

$i = 1, \ldots, l$. We will consider the total conflict between the centers of clusters and the bodies of evidence generated by the focal elements of these clusters as a minimized functional by analogy with the k-means algorithm:

$$\Phi = \sum_{i=1}^{l} \sum_{B \in \mathcal{A}_i} Con\,(F(\{B\}), C_i), \tag{1}$$

where $F(\{B\})$ is the evidence generated from the set $\{B\}$ using the restriction and extension procedures (see subsection 3.1). In particular, if the vacuous extension is chosen, then $F(\{B\}) = m(B)F_B + (1 - m(B))F_X$.

In this algorithm, the number of evidence bodies l into which the evidence body $F = (\mathcal{A}, m)$ is decomposed will be considered predetermined (it is determined from some other heuristic considerations). Also, the method of extension the mass function will be considered predetermined.

Let us assume that the covering $\mathcal{C} = \{\mathcal{A}_1, \ldots, \mathcal{A}_l\}$ is fixed and the center of the i-th cluster has the form

$$C_i = \sum_{A \in \mathcal{A}_i} \alpha_i(A) F_A, \tag{2}$$

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 the k-dimensional simplex. Then we have for the vacuous extension

$$\Phi = \sum_{i=1}^{l} \sum_{B \in \mathcal{A}_i} Con\,(F(\{B\}), C_i) = \sum_{i=1}^{l} \sum_{B \in \mathcal{A}_i} m(B) \sum_{\substack{A \in \mathcal{A}_i: \\ A \cap B = \emptyset}} \alpha_i(A) =$$

$$= \sum_{i=1}^{l} \sum_{B \in \mathcal{A}_i} m(B) \left(1 - \sum_{\substack{A \in \mathcal{A}_i: \\ A \cap B \neq \emptyset}} \alpha_i(A) \right) = k_{\mathcal{C}} - \sum_{i=1}^{l} Q_i(\alpha_i),$$

where $k_{\mathcal{C}} = \sum_{i=1}^{l} \sum_{B \in \mathcal{A}_i} m(B) \geq 1$ ($k_{\mathcal{C}} = 1 \Leftrightarrow \mathcal{C} = \{\mathcal{A}_1, \ldots, \mathcal{A}_l\}$ is the partition of the set of focal elements), $Q_i(\alpha_i) = \sum_{A \in \mathcal{A}_i} \alpha_i(A) Pl_{\mathcal{A}_i}(A)$ and $Pl_{\mathcal{A}_i}(A) = \sum_{\substack{B \in \mathcal{A}_i: \\ A \cap B \neq \emptyset}} m(B)$ is the restriction of the plausibility function to the set $\mathcal{A}_i$. The minimum of the functional Φ for a fixed coverage $\mathcal{C} = \{\mathcal{A}_1, \ldots, \mathcal{A}_l\}$ will be achieved when the linear functions $Q_i(\alpha_i)$ reach maxima on the simplices $S_{|\mathcal{A}_i|}$, $i = 1, \ldots, l$. But

$$\max_{\alpha \in S_{|\mathcal{A}_i|}} Q_i(\alpha) = \max_{A \in \mathcal{A}_i} Pl_{\mathcal{A}_i}(A), \quad i = 1, \ldots, l.$$

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

$$C_i = \sum_{A \in \overline{\mathcal{A}_i}} \alpha_i(A) F_A, \quad \alpha_i = (\alpha_i(A))_{A \in \overline{\mathcal{A}_i}} \in S_{|\overline{\mathcal{A}_i}|}, \quad i = 1, \ldots, l, \tag{3}$$

then the functional Φ will reach a minimum in the case of a fixed coverage $\mathcal{C} = \{\mathcal{A}_1, \dots, \mathcal{A}_l\}$ with such a choice of cluster centers. This minimum will be

$$\min \Phi = k_{\mathcal{C}} - \sum_{i=1}^{l} \max_{A \in \mathcal{A}_i} Pl_{\mathcal{A}_i}(A) \tag{4}$$

and does not depend on the choice of parameters $\alpha_i = (\alpha_i(A))_{A \in \overline{\mathcal{A}}_i} \in S_{|\overline{\mathcal{A}_i}|}$, $i = 1, \dots, l$.

Then the evidence body clustering algorithm, by analogy with the classical k-means algorithm, will be as follows.

Algorithm 2.

1. Let's choose and fix the number of clusters l. Let's assign some evidence bodies as initial cluster centers $C_i^{(0)}$, $i = 1, ..., l$. We fix the threshold of maximum conflict within clusters $Con_{\max} \in [0, 1]$. We install $s = 0$.

2. We redistribute focal elements among clusters according to the principle of minimizing the conflict between evidence clusters and cluster centers. The focal element $B \in \mathcal{A}$ is assigned to the cluster $\mathcal{A}_i^{(s)} = \arg\min_j Con\left(F(\{B\}), C_j^{(s)}\right)$ and $\min_i Con\left(F(\{B\}), C_i^{(s)}\right) \le Con_{\max}$. If $\min_i Con\left(F(\{B\}), C_i^{(s)}\right) > Con_{\max}$, then the focal element B is assigned as the center of the new cluster. As a result, we get clusters $\mathcal{A}_i^{(s)}$, $i = 1, ..., l$.

3. Let us calculate new cluster centers using the formulac. We increase the counter $s \leftarrow s + 1$.

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

Proposition 1. *Algorithm 2 converges in a finite number of steps.*

The proof follows from the fact that the functional Φ does not increase at the 2^{nd} and 3^{rd} steps of the algorithm and we have a finite number of possible configurations.

Remark 2. Evidence bodies $C_i^{(0)} = F_{A_i}$, $i = 1, ..., l$ can be chosen as the initial centers of clusters at the 1^{st} step of the algorithm, where focal elements A_i, $i = 1, ..., l$ are chosen arbitrarily or, for example, using the density maximization algorithm.

Remark 3. Cluster centers may depend on parameters $\alpha = (\alpha(A))_{A \in \overline{\mathcal{A}}_i} \in S_{|\overline{\mathcal{A}_i}|}$ (see formula (3)). In this case, it is necessary to use additional procedures for choosing parameters at the 2^{nd} or 3^{rd} steps of the algorithm. The selection criteria can be considered, for example:

1) coverage minimization, i.e., we choose the parameters at the 2^{nd} step of the algorithm so that the coverage $\mathcal{C} = \{\mathcal{A}_1, \dots, \mathcal{A}_l\}$ is 'closer' to the partition. For example, $\sum_{i=1}^{l} |\mathcal{A}_i| \to \min$.

2) minimizing the uncertainty of evidence-centers of clusters C_i, $i = 1, ..., l$. This procedure is applied at the 3^{rd} step of the algorithm. The uncertainty of evidence can be assessed using one of the imprecision indices [1]. For example, it can be the generalized Hartley measure $H(C_i) = \sum_{A \in \overline{\mathcal{A}_i}} \alpha_i(A) \ln |A|$.

3) minimizing the distance between cluster centers and the original evidence body with respect to some metric between evidence bodies [7]: $d(C_i, F) \to \min$, $i = 1, ..., l$;

4) maximizing distance between cluster centers $d(C_i, C_j) \to \max$ or maximizing conflict $Con(C_i, C_j) \to \max$, $i, j = 1, ..., l$ $(i \neq j)$ etc.

Remark 4. One way to evaluate the internal conflict [9] of a body of evidence F on the X is based on finding the maximum of the contour function [3]: $Con(F) = 1 - \max_{x \in X} Pl(x)$. Then formula (4) can be interpreted as a total intra-cluster internal conflict.

Remark 5. It is possible to search for cluster centers C_i, $i = 1, ..., l$ in the form (2), minimizing the functional (1) with a fixed coverage $\mathcal{C} = \{\mathcal{A}_1, \ldots, \mathcal{A}_l\}$ and under the condition that uncertainties of cluster centers C_i are bounded:

$$H(C_i) = \sum_{A \in \mathcal{A}_i} \alpha_i(A) \ln |A| \leq u_i, \quad i = 1, ..., l,$$

where u_i, $i = 1, ..., l$ are some threshold values. Then the problem of minimizing the functional (1) for a fixed coverage $\mathcal{C} = \{\mathcal{A}_1, \ldots, \mathcal{A}_l\}$ is reduced to solving l linear programming problems:

$$\sum_{A \in \mathcal{A}_i} \alpha_i(A) Pl_{\mathcal{A}_i}(A) \to \max$$

$$\text{subject to } \alpha_i = (\alpha_i(A))_{A \in \mathcal{A}_i} \in S_{|\mathcal{A}_i|}, \quad \sum_{A \in \mathcal{A}_i} \alpha_i(A) \ln |A| \leq u_i, \quad i = 1, ..., l.$$

Example 2. Algorithm 2 will give the following result of clustering the body of evidence from Example 1 ($X = \{1, 2, 3\}$, $F = 0.3F_{\{1\}} + 0.2F_{\{2\}} + 0.3F_{\{1,3\}} + 0.2F_{\{2,3\}}$ into two clusters and the vacuous extension.

Step 1. We have $l = 2$. Let the initial centers of the clusters be equal to $C_1^{(0)} = F_{\{1\}}$, $C_2^{(0)} = F_{\{2\}}$; $s = 0$.

Step 2. We have $Con\left(F(\{1\}), C_1^{(0)}\right) = Con\left(F(\{1,3\}), C_1^{(0)}\right) = 0$,

$Con\left(F(\{2\}), C_1^{(0)}\right) = Con\left(F(\{2,3\}), C_1^{(0)}\right) = 0.2$,

$Con\left(F(\{1\}), C_2^{(0)}\right) = Con\left(F(\{1,3\}), C_2^{(0)}\right) = 0.3$, $Con\left(F(\{2\}), C_2^{(0)}\right) =$ $Con\left(F(\{2,3\}), C_2^{(0)}\right) = 0$ (for example, $Con\left(F(\{1,3\}), C_2^{(0)}\right) =$ $Con\left(0.3F_{\{1,3\}} + 0.7F_X, F_{\{2\}}\right) = 0.3$).

Therefore, according to the principle of minimizing the conflict between evidence clusters and cluster centers, the initial clustering will have the form $\mathcal{A}_1^{(0)} = \{\{1\}, \{1,3\}\}$, $\mathcal{A}_2^{(0)} = \{\{2\}, \{2,3\}\}$.

Step 3. Let us calculate the new cluster centers using the formula (3)): $Pl_{\mathcal{A}_1^{(0)}}(\{1\}) = 0.3 + 0.3 = 0.6$, $Pl_{\mathcal{A}_1^{(0)}}(\{1,3\}) = 0.3 + 0.3 = 0.6$, $Pl_{\mathcal{A}_2^{(0)}}(\{2\}) = 0.2 + 0.2 = 0.4$, $Pl_{\mathcal{A}_2^{(0)}}(\{2,3\}) = 0.2 + 0.2 = 0.4$.

Therefore $C_1^{(1)} = \alpha F_{\{1\}} + (1 - \alpha) F_{\{1,3\}}$ and $C_2^{(1)} = \beta F_{\{2\}} + (1 - \beta) F_{\{2,3\}}$, $\alpha, \beta \in [0, 1]$.

If we require the minimization of the generalized Hartley measure (see Remark 2), we get $C_1^{(1)} = \underset{0\le\alpha\le1}{\arg\min} H\left(\alpha F_{\{1\}} + (1-\alpha)F_{\{1,3\}}\right) = F_{\{1\}}$, $C_2^{(1)} = \underset{0\le\beta\le1}{\arg\min} H\left(\beta F_{\{2\}} + (1-\beta)F_{\{2,3\}}\right) = F_{\{2\}}$ and the algorithm will stop its work, since the centers of the clusters have not changed. If, however, we apply the coverage minimization rule (see Remark 3), we move on to the next step.

Step 4. We redistribute focal elements according to the criterion of least conflict with new centers:

$Con\left(F(\{1\}), C_1^{(1)}\right) = Con\left(F(\{1,3\}), C_1^{(1)}\right) = 0$, $Con\left(F(\{2\}), C_1^{(1)}\right) = 0.2$,

$Con\left(F(\{2,3\}), C_1^{(1)}\right) = 0.2\alpha$, $Con\left(F(\{1\}), C_2^{(1)}\right) = 0.3$,

$Con\left(F(\{1,3\}), C_2^{(1)}\right) = 0.3\beta$, $Con\left(F(\{2\}), C_2^{(1)}\right) = Con\left(F(\{2,3\}), C_2^{(1)}\right) = 0$.

We will get clusters $\mathcal{A}_1^{(1)} = \{\{1\},\{1,3\}\}$, $\mathcal{A}_2^{(1)} = \{\{2\},\{2,3\}\}$ after applying the coverage minimization rule (see Remark 3). Clusters have stabilized. Stop of the algorithm.

As a result, we get, in fact, a new body of evidence defined on the base set $\mathcal{A}$, the found coverage sets (clusters) $\mathcal{C} = \{\mathcal{A}_1, \dots, \mathcal{A}_l\}$ will be focal elements, the mass function will be equal to $m(\mathcal{A}_i) = \sum_{B\in\mathcal{A}_i} m(B)/n(B)$, where $n(B) = |\{\mathcal{A}_i : B \in \mathcal{A}_i\}|$ (the number of clusters containing the set B). Such a body of evidence can be considered second-order evidence, which reflects the enlarged structure of the original evidence.

4 Evaluation of the internal conflict of the body of evidence based on its clustering

Let us assume that in one way or another, a cluster coverage $\mathcal{C} = \{\mathcal{A}_1, \dots, \mathcal{A}_l\}$ (in a particular case, partitioning) of the body of evidence $F = (\mathcal{A}, m)$ is obtained. Then we can offer the following ways to evaluate the internal conflict of this body of evidence using some measure of external conflict Con:

1) $Con_1(F) = Con(F(\mathcal{A}_1), \dots, F(\mathcal{A}_l))$;

2) $Con_2(F) = Con(C_1, \dots, C_l)$, where $C_1, \dots, C_l$ are centers of clusters $\mathcal{A}_1, \dots, \mathcal{A}_l$ respectively.

For example, we will obtain the following estimates of the internal conflict for the body of evidence considered in Example 2, the vacuous extension, and the canonical measure of the external conflict. We have $\mathcal{A}_1 = \{\{1\},\{1,3\}\}$, $\mathcal{A}_2 = \{\{2\},\{2,3\}\}$ and $Con_1(F) = Con(F(\mathcal{A}_1), F(\mathcal{A}_2)) = 0.18$, $Con_2(F) = Con(C_1, C_2) = \alpha + (1-\alpha)\beta$, $\alpha, \beta \in [0,1]$.

Proposition 2. *The following equality is true*

$$Con_1(F) = Con(F(\mathcal{A}_1), \dots, F(\mathcal{A}_l)) = \sum_{A_1\in\mathcal{A}_1,\dots,A_l\in\mathcal{A}_l} Con(F(\{A_1\}), \dots, F(\{A_l\})).$$

5 Conclusion

Two methods of evidence body clustering are discussed in this paper. Each of these methods assumes that weakly conflicting focal elements should belong to one cluster, and strongly conflicting focal elements should belong to different clusters. This requirement is similar to the basic principle of compactness in cluster analysis: the distances between elements of one cluster should be minimal, and between clusters should be maximum.

The first algorithm is based on the use of the distribution density function of conflicting focal elements. The second algorithm implements the idea of the k-means method. In this case, the cluster centers are formed in some optimal way. Further, focal elements are redistributed according to the principle of minimizing conflict with cluster centers.

It shows how clustering can be used to evaluate the internal conflict of a body of evidence.

References

1. Bronevich, A., Lepskiy, A.: Imprecision indices: axiomatic, properties and applications. International Journal of General Systems 44(7-8), 812–832 (2015)
2. Bronevich, A., Lepskiy, A.: Measures of conflict, basic axioms and their application to the clusterization of a body of evidence. Fuzzy Sets and Systems (online) (2021) https://doi.org/10.1016/j.fss.2021.04.016
3. Daniel, M.: Properties of Plausibility Conflict of Belief Functions. In: Rutkowski, L., Korytkowski, M., Scherer, R., et al. (eds.) ICAISC 2013, part I, LNAI, vol. 7894, pp. 235–246. Springer-Verlag (2013)
4. Dempster, A.P.: Upper and lower probabilities induced by multivalued mapping. Ann. Math. Statist. 38, 325–339 (1967)
5. 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)
6. Harmanec, D.: Faithful approximations of belief functions. In: Laskey, K. B. and Prade, H., (eds) Uncertainty in Artificial Intelligence 15 (UAI99), Stockholm, Sweden, (1999)
7. Jousselme, A.-L., Maupin, P.: Distances in Evidence Theory: Comprehensive Survey and Generalizations. Int. J. of Approximate Reasoning 53, 118–145 (2012)
8. Lepskiy, A.: Analysis of Information Inconsistency in Belief Function Theory. Part I: External Conflict. Control Sciences 5, 2–16 (2021)
9. Lepskiy, A.: Analysis of Information Inconsistency in Belief Function Theory. Part II: Internal Conflict. Control Sciences 6, 2–12 (2021)
10. 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)
11. Pichon, F., Jousselme, A. L., Abdallah, N. Ben : Several shades of conflict. Fuzzy Sets Syst. 366, 63–84 (2019)
12. Shafer, G.: A mathematical theory of evidence. Princeton Univ. Press (1976)
13. Yager, R.: On the Dempster-Shafer Framework and New Combination Rules. Information Sciences 41, 93–137 (1987)