



**HAL**  
open science

# Representation of uncertainty and imprecision between clusters with belief functions

Zuowei Zhang

► **To cite this version:**

Zuowei Zhang. Representation of uncertainty and imprecision between clusters with belief functions. Other [cs.OH]. Université de Rennes; Northwestern Polytechnical University (Chine), 2022. English. NNT : 2022REN1S003 . tel-04416424

**HAL Id: tel-04416424**

**<https://theses.hal.science/tel-04416424>**

Submitted on 25 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1

ECOLE DOCTORALE N° 601  
*Mathématiques et Sciences et Technologies  
de l'Information et de la Communication*  
Spécialité : *Informatique*

Par

« **Zuwei ZHANG** »

« **Representation of uncertainty and imprecision between clusters with belief functions** »

Thèse présentée et soutenue à Rennes, le 18 Janvier 2022  
Institut de recherche en informatique et systèmes aléatoires (IRISA)

## Rapporteurs avant soutenance :

Pr. Thierry DENOEUX Professeur, Université Technologique de Compiègne  
Pr. Zied ELOUEDI Professeur, Institut Supérieur de Gestion de Tunis

## Composition du Jury :

Président :	Pr. Christelle VRAIN	Professeur, Université de Orléans
Examineurs :	Pr. Christelle VRAIN	Professeur, Université de Orléans
	Pr. Thierry DENOEUX	Professeur, Université Technologique de Compiègne
	Pr. Zied ELOUEDI	Professeur, Institut Supérieur de Gestion de Tunis
	Pr. Deqiang HAN	Professeur, Xi'an Jiaotong University
	Pr. Kuang ZHOU	Associate professor, Northwestern Polytechnical University
Dir. de thèse :	Pr. Arnaud MARTIN	Professeur, Université de Rennes 1
Co-dir. de thèse :	Pr. Zhunga LIU	Professeur, Northwestern Polytechnical University



## Abstract

Clustering aims to divide massive data without any prior information into groups with similar data structures or physical structures. To derive such a proposal, several different clustering strategies have been proposed, such as partition based methods, hierarchical based methods, distribution based methods, spectral based methods, density based methods, grid based methods, machine learning based methods, and neural networks based methods. But most of the methods cannot characterize the uncertainty and imprecision between clusters at the same time. In recent years, evidential clustering (EC), based on the concept of *credal partition*, has received a lot of attention for its ability to deal with this problem. It inherits the advantages of the theory of belief functions (TBF) in reasoning about uncertain and imprecise knowledge. However, since EC is still in the development stage, there are still issues such as basic concepts, high computational complexity, and inability to detect imbalanced or arbitrary clusters, limiting the applications of EC.

In this thesis, we work on proposing some alternative solutions to solve these issues. Specifically, four works are brought forward to handle them one by one.

We start from a systematic study of EC. In this work, we present the concepts and definitions of data (inputs), methods (models), and clusters (results) of EC, as well as that of the other types (*i.e.* hard/fuzzy/possibilistic ones), based on different theories (*i.e.* probability theory, fuzzy set theory, possibility theory, and the theory of belief functions). This is because EC is regarded as the evidential version of hard, fuzzy, and possibilistic clustering under the TBF. These concepts and definitions explain why EC can better characterize the uncertainty and imprecision between clusters. Moreover, we also study EC from the seminal to state-of-the-art methods in the context of data-based inputs, resulting in a coherent and comprehensive survey to analyze these methods. Based on the analysis of some representative techniques from different perspectives (*e.g.* center, measure type, complexity), we provide a guiding scheme to help users to choose appropriate evidential methods in their cases.

Afterward, we introduce a dynamic evidential clustering (DEC) algorithm for the issue of the high computational complexity of traditional EC when characterizing the uncertainty and imprecision between clusters. In DEC, most query objects are considered to have precise cluster information, so an FCM-like objective function is first employed and minimized to obtain the support levels of the real singletons (specific) clusters to which the query objects belong. Then the query object is initially adaptively assigned to the outlier, precise or imprecise one via a new rule based on the conflicts between the different support levels. Each imprecise object is finally reassigned to the singleton clusters or related meta-cluster by partial credal redistribution with the corresponding dynamic edited framework to reduce the computational burden. The proposed DEC method can reduce the complexity to a level similar to that

of fuzzy/possibilistic clustering, extending EC’s application in big data.

Then, we extend EC to detect imbalanced clusters (clusters with different sizes) by combining mean shift with traditional EC under the TBF, mainly containing two characteristics. First, the query object is preliminarily assigned as the noise, precise, or imprecise object based on the notion of “belief shift”. Second, partial credal redistribution with dynamic cluster centers, to avoid the “uniform effect” (imbalanced clusters), is established to reassign imprecise objects to the singleton cluster or related meta-cluster. Once an object is assigned to a meta-cluster, it indicates that the imbalanced singleton clusters involved in the meta-cluster cannot be distinguished because this object may be located in the overlapping or intermediate areas of these imbalanced singleton clusters. By doing this, the BSC can reasonably characterize the uncertainty and imprecision between imbalanced singleton clusters.

To avoid losing generality, we also investigate the representation of uncertainty and imprecision between clusters regardless of their shape, size, and dimensionality based on density peaks clustering and the TBF. First, we consider that different neighbors can provide complementary evidence supporting the object as a cluster center and redefine a distance-based density function to obtain more robust cluster centers in the decision graph. Then, we present a new evidential convergence rule to assign the remaining objects to different clusters. Finally, similar to BSC, the objects located in the overlapping or intermediate areas of different arbitrary singleton clusters are assigned to corresponding meta-clusters to characterize the uncertainty and imprecision between these arbitrary clusters.

The effectiveness of the proposed algorithms is estimated on different artificial and natural datasets. Experiments show that our proposed algorithms effectively improve the execution efficiency of traditional EC and detect imbalanced or arbitrary clusters, and characterizes the uncertainty and imprecision between these clusters.

## Résumé

Le clustering vise à diviser des données massives, sans aucune information préalable, en groupes présentant des structures de données ou des structures physiques similaires. Pour y parvenir, plusieurs stratégies de clustering ont été proposées : des méthodes de partitionnement, des méthodes hiérarchiques, des méthodes fondées sur des distributions de probabilité, des méthodes spectrales, des méthodes fondées sur la densité, des méthodes à base de grilles, et des méthodes fondées sur l'apprentissage automatique tels que les réseaux neuronaux. Cependant, la plupart de ces méthodes ne peuvent pas caractériser en même temps l'incertitude et l'imprécision entre les clusters. Ces dernières années, le clustering crédibiliste (Evidential clustering EC), fondé sur le concept de *credal partition*, a reçu beaucoup d'attention pour sa capacité à traiter ce problème. Il hérite des avantages de la théorie des fonctions de croyance (TBF) qui permet la modélisation des connaissances incertaines et imprécises. Cependant, comme le clustering crédibiliste est encore au stade de développement, il reste des problèmes tels que des concepts sous-jacents peu clairs, une complexité de calcul élevée et l'incapacité de détecter des clusters déséquilibrés ou arbitraires, ce qui limite les applications de la méthode EC.

Dans cette thèse, nous travaillons à proposer des schémas alternatifs pour résoudre ces problèmes. Plus précisément, quatre travaux sont mis en avant pour les traiter un par un.

Nous commençons par une étude systématique des approches crédibiliste de clustering. Dans ce travail, nous présentons les concepts et les définitions des données (entrées), des méthodes (modèles) et des groupes (résultats) de l'EC, ainsi que ceux des autres types (durs/flous/possibilistes), fondés sur différentes théories de l'incertain (théorie des probabilités, théorie des ensembles flous, théorie des possibilités et théorie des fonctions de croyance). En effet, l'EC est considéré comme la version crédibiliste du clustering dur, flou et possibiliste dans le cadre de la TBF. Ces concepts et définitions expliquent pourquoi l'EC peut mieux caractériser l'incertitude et l'imprécision entre les clusters. En outre, nous étudions l'EC depuis les méthodes séminales jusqu'aux méthodes de pointe dans le contexte des entrées fondées sur les données, ce qui donne lieu à une étude cohérente et complète pour analyser ces méthodes. Sur la base de l'analyse de certaines techniques représentatives sous différents angles (centre, type de mesure, complexité), nous fournissons un schéma directeur pour aider les utilisateurs à choisir les méthodes crédibilistes appropriées à leur données et cas d'étude.

Ensuite, nous introduisons un algorithme de clustering dynamique (DEC) pour résoudre le problème de la complexité de calcul élevée de l'EC traditionnel lors de la caractérisation de l'incertitude et de l'imprécision entre les clusters. Dans l'algorithme DEC, la plupart des objets de la requête sont considérés comme ayant des informations précises sur les clusters. Ainsi, une fonction objective de type FCM est d'abord

employée et minimisée pour obtenir les niveaux de soutien des clusters singletons (spécifiques) auxquels les objets de la requête appartiennent. Ensuite, l’objet de la requête est initialement assigné de manière adaptative à l’objet aberrant, précis ou imprécis, via une nouvelle règle fondée sur les conflits entre les différents niveaux de support. Chaque objet imprécis est finalement réaffecté aux clusters singleton ou aux méta-cluster appartenés par redistribution crédale partielle avec une approche dynamique correspondant pour réduire la charge de calcul. La méthode DEC proposée peut réduire la complexité à un niveau similaire à celui du clustering flou/possibiliste, ce qui étend l’application de l’EC aux données volumineuses.

Ensuite, nous étendons l’EC pour détecter les clusters déséquilibrés en combinant le déplacement de la moyenne avec l’EC traditionnel sous le TBF, contenant principalement deux caractéristiques. Premièrement, l’objet de la requête est préalablement assigné comme étant le bruit, précis ou imprécis, en se fondant sur la notion de “changement de croyance”. Deuxièmement, une redistribution crédale partielle avec des centres de grappes dynamiques, pour éviter “l’effet uniforme” (pour les clusters déséquilibrés), est établie pour réaffecter les objets imprécis à un singleton cluster ou à un méta-cluster connexe. Une fois qu’un objet est assigné à un méta-cluster, cela indique que les clusters singletons déséquilibrés impliqués dans le méta-cluster ne peuvent pas être distingués car cet objet peut être situé dans les zones de chevauchement ou intermédiaires de ces clusters singletons déséquilibrés. En procédant ainsi, l’approche BSC proposée peut raisonnablement caractériser l’incertitude et l’imprécision entre les clusters singletons déséquilibrés.

Pour éviter de perdre en généralité, nous étudions également la représentation de l’incertitude et de l’imprécision entre les clusters, indépendamment de leur forme, de leur taille et de leur dimensionnalité, sur la base des pics de densité et du TBF. Tout d’abord, nous considérons que différents voisins peuvent fournir des preuves complémentaires soutenant l’objet comme centre de cluster et redéfinissons une fonction de densité fondée sur la distance pour obtenir des centres de cluster plus robustes dans le graphe de décision. Ensuite, nous présentons une nouvelle règle de convergence crédibiliste pour affecter les objets restants à différents clusters. Enfin, comme dans le cas de l’approche BSC, les objets situés dans les zones de chevauchement ou intermédiaires de différents clusters singletons arbitraires sont assignés aux méta-clusters correspondants pour caractériser l’incertitude et l’imprécision entre ces clusters arbitraires.

L’efficacité des algorithmes proposés est estimée sur différents jeux de données artificiels et réels. Les expériences montrent que les algorithmes proposés améliorent effectivement l’efficacité d’exécution de l’EC traditionnel et détectent les clusters déséquilibrés ou arbitraires, et caractérisent l’incertitude et l’imprécision entre ces clusters.

# Acknowledgements

I am indebted to many people for their help and support during my Ph.D. study and research at team DRUID, IRISA, University of Rennes 1 in France, and the Key Laboratory of Information Fusion Technology (LIFT) of Northwestern Polytechnical University (NPU) in China. In addition, I would like to thank China Scholarship Council for supporting my study in France.

I would like to express my sincere thanks to my supervisor at the University of Rennes 1, Prof. Arnaud Martin. His help and efforts allow me to quickly overcome the language barrier and adapt to life at Lannion. The weekly meetings motivate me to delve deeper into my research, in which I learn many valuable methods and suggestions. Prof. Arnaud Martin also teaches me to face problems in my studies and in my life. It is better to confront and solve problems head-on than back to back. He teaches me a lot not only about knowledge but also how to be a researcher.

I would also like to gratefully thank my supervisors at NPU, Prof. Zhun-ga Liu and Assoc. prof. Kuang Zhou. They always give me some far-sighted suggestions for my study and life and provide me with great support for my study in France. Their academic rigor also has a positive impact on me and guides my future research. They have also provided sound and practical advice and guidance for my long-term research, which has been an invaluable asset in my life.

I would also like to thank all members in my jury committee, especially Profs. Thierry Denceux and Zied Elouedi, who have also been in the comité de suivi individuel.

I would also like to thank my colleagues and friends at Lannion, Yi-ru Zhang, Na Li, Constance Thierry, and many others. They helped me solve many problems in my life and studies, which helped me focus my research better.

I am grateful for the continuous support and love of my parents and my wife. Their understanding and support help me make the brave decision to pursue this Ph.D. and have kept me going ever since.





# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivations . . . . .	2
1.2	Objectives . . . . .	3
1.3	Contributions . . . . .	4
1.4	Structure of the thesis . . . . .	6
<b>2</b>	<b>Background</b>	<b>7</b>
2.1	Introduction . . . . .	8
2.2	The theory of belief functions . . . . .	8
2.3	Some links with other related theories . . . . .	10
2.3.1	Probability theory . . . . .	10
2.3.2	Fuzzy set theory . . . . .	11
2.3.3	Possibility theory . . . . .	11
2.3.4	Rough set theory . . . . .	12
2.4	Conclusion . . . . .	12
<b>3</b>	<b>Concepts, definitions and techniques of evidential clustering</b>	<b>13</b>
3.1	Introduction . . . . .	15
3.2	Concepts and definitions of data . . . . .	15
3.2.1	Hard data . . . . .	16
3.2.2	Fuzzy data . . . . .	16
3.2.3	Possibilistic data . . . . .	17
3.2.4	Evidential data . . . . .	18
3.2.5	The difference and relations of different data . . . . .	19
3.3	Concepts and definitions of clusters . . . . .	21
3.3.1	Hard clusters . . . . .	22
3.3.2	Fuzzy clusters . . . . .	22
3.3.3	Possibilistic clusters . . . . .	23
3.3.4	Evidential clusters . . . . .	23
3.3.5	The difference and relations of different clusters . . . . .	25
3.4	Concepts and definitions of methods . . . . .	26
3.4.1	Hard methods . . . . .	27
3.4.2	Fuzzy methods . . . . .	27

3.4.3	Possibilistic methods . . . . .	28
3.4.4	Evidential methods . . . . .	29
3.4.5	The overview of different data, methods, and clusters . . . . .	30
3.5	Clustering techniques based on TBF . . . . .	30
3.5.1	Object data-based methods . . . . .	31
3.5.2	Relational data-based methods . . . . .	33
3.5.3	Applications of evidential methods . . . . .	34
3.5.4	Analysis of different evidential methods . . . . .	35
3.5.5	Challenges of evidential clustering . . . . .	36
3.6	Conclusion . . . . .	37
<b>4</b>	<b>Dynamic evidential clustering algorithm</b>	<b>39</b>
4.1	Introduction . . . . .	40
4.2	Review of evidential $c$ -means and credal $c$ -means . . . . .	40
4.2.1	Evidential $c$ -means (ECM) . . . . .	40
4.2.2	credal $c$ -means (CCM) . . . . .	41
4.3	Dynamic evidential clustering (DEC) . . . . .	43
4.3.1	Preliminary credal partition . . . . .	43
4.3.2	Partial credal redistribution . . . . .	46
4.3.3	The involved parameters . . . . .	48
4.3.4	The complexity of DEC . . . . .	49
4.4	Experiments . . . . .	50
4.4.1	Performance evaluation . . . . .	50
4.4.2	Gaussian and round shape data . . . . .	51
4.4.3	Large data . . . . .	53
4.4.4	Real world data . . . . .	55
4.4.5	Discussion . . . . .	58
4.5	Conclusion . . . . .	60
<b>5</b>	<b>Belief shift clustering algorithm</b>	<b>63</b>
5.1	Introduction . . . . .	64
5.2	Review of mean-shift . . . . .	64
5.3	Belief shift clustering (BSC) . . . . .	65
5.3.1	Belief shift for preliminary credal partition . . . . .	65
5.3.2	Evidential clustering rule for credal redistribution . . . . .	69
5.3.3	The involved parameters . . . . .	73
5.4	Experiments . . . . .	73
5.4.1	Synthetic data . . . . .	74
5.4.2	Imbalanced data . . . . .	75
5.4.3	Medical and natural data . . . . .	77

---

5.4.4	Some real world data . . . . .	78
5.4.5	Olivetti Face data . . . . .	81
5.4.6	Discussion . . . . .	81
5.5	Conclusion . . . . .	84
<b>6</b>	<b>Density peaks-based evidential convergence algorithm</b>	<b>85</b>
6.1	Introduction . . . . .	86
6.2	Review of density peaks clustering . . . . .	86
6.3	Density peaks-based evidential convergence (DPEC) . . . . .	87
6.3.1	Evidence-based density peaks . . . . .	87
6.3.2	Evidential convergence . . . . .	89
6.3.3	The involved parameters . . . . .	91
6.4	Experiments . . . . .	92
6.4.1	Performance evaluation . . . . .	92
6.4.2	Synthetic data . . . . .	92
6.4.3	Olivetti Face data . . . . .	96
6.4.4	Natural data . . . . .	98
6.5	Discussions . . . . .	100
6.5.1	Complexity analysis . . . . .	100
6.5.2	Parametric sensitivity . . . . .	100
6.5.3	Comparisons of proposed algorithms . . . . .	101
6.6	Conclusion . . . . .	107
<b>7</b>	<b>Conclusion and future work</b>	<b>109</b>
7.1	Conclusion . . . . .	110
7.2	Future work . . . . .	111
	<b>Appendix</b>	<b>115</b>
	<b>Author's Publication List</b>	<b>129</b>
	<b>Bibliography</b>	<b>131</b>



# List of Tables

3.1	Evidence bodies of $\mathbf{x}_i$ from different sources . . . . .	18
3.2	Basic information of 3 people . . . . .	19
3.3	Fuzzy data for 3 people . . . . .	20
3.4	Possibilistic data for 3 people . . . . .	20
3.5	Evidential data for 3 people . . . . .	21
3.6	Comparisons of different evidential methods . . . . .	36
4.1	The numerical example . . . . .	45
4.2	Basics of the used datasets . . . . .	55
4.3	Clustering results of different datasets (in %) . . . . .	61
4.4	Clustering results of different datasets . . . . .	62
5.1	The basic information of the <b>SD4</b> dataset . . . . .	76
5.2	The results of different methods with <b>DI2</b> and <b>NI2</b> . . . . .	79
5.3	Basic information of the UCI datasets . . . . .	79
5.4	Selection of the parameters in different methods . . . . .	79
5.5	Clustering results of different methods with the UCI datasets (In %) . .	80
5.6	Clustering results of different methods with <b>FI5</b> . . . . .	83
6.1	Clustering results of different methods on the eight synthetic datasets .	94
6.2	Clustering results of different methods on the Olivetti face dataset . . .	97
6.3	Clustering results of different methods on the BS dataset . . . . .	98
6.4	Execution time on the Berkeley Segmentation and Olivetti face datasets	101
6.5	Clustering results of DPEC with different $\mathcal{K}_1$ , $\mathcal{K}_2$ , and $\zeta$ on the SD1 dataset . . . . .	102
7.1	Comparisons of different evidential methods . . . . .	112
7.2	Some given parameters in comparison methods . . . . .	115



# List of Figures

1.1	Illustration of uncertainty between different clusters. . . . .	2
3.1	Venn diagram of hard, fuzzy, possibilistic, and evidential data. . . . .	19
3.2	Relationship between hard, fuzzy, possibilistic, and evidential clusters. . . . .	25
3.3	Illustration of data, methods, and clusters of existing methods. . . . .	30
3.4	The basic principles for choosing different evidential methods. . . . .	37
4.1	Flowchart of the adaptive credal partition. . . . .	47
4.2	Clustering results of 3-class of gaussian dataset by different methods. . . . .	52
4.3	Clustering results of 3-class of round shape dataset by different methods. (a) Original artificial data. (b) CCM ( $Re = 7.07, Ri = 17.34, T = 0.5897$ ). (c) DEC ( $Re = 4.23, Ri = 20.81, T = 0.0072$ ). . . . .	52
4.4	Clustering results of 6-class of dataset by different methods. (a) Original artificial data. (b) ECM ( $Re = 1.13, Ri = 1.23, T = 4354.21$ ). (c) CCM ( $Re = 0, Ri = 0.30, T = 3833.51$ ). (d) BPEC ( $Re = 0.82, Ri = 1.14, T = 10534.04$ ). (e) DEC with $\varphi = 0.2$ ( $Re = 0, Ri = 0.21, T = 85.31$ ). (f) DEC with $\varphi = 0.3$ ( $Re = 0, Ri = 0.44, T = 110.90$ ). . . . .	54
4.5	Clustering results of Iris dataset. . . . .	56
4.6	ROC curves for imprecision and error rates of ECM and DEC. . . . .	57
4.7	Clustering results of different datasets for various $\varphi$ . . . . .	59
5.1	Illustration of the belief shift process. . . . .	67
5.2	Illustration of preliminary adaptive credal partition. . . . .	69
5.3	Illustration of the simulated cluster centers based on imbalanced data. . . . .	70
5.4	The basic information of the <b>SD15</b> dataset. . . . .	74
5.5	Clustering results of the SD15 dataset by different methods. . . . .	75
5.6	Clustering results of the <b>SD4</b> dataset by different methods. . . . .	76
5.7	Clustering results of the medical image by different methods. . . . .	77
5.8	Clustering results of different methods with a goose floating on the lake. . . . .	78
5.9	Detection performance of BSC with different $\gamma$ . . . . .	80
5.10	Clustering results of <b>FI5</b> by different methods. . . . .	82
5.11	The results of BSC with different $\alpha$ . . . . .	83
6.1	Clustering results by DPEC on the eight synthetic datasets. . . . .	93



---

6.2	Clustering results by DPEC on the Olivetti faces dataset. . . . .	96
6.3	Decision graphs of DPC, CDPC, and DPEC on the Olivetti face dataset. . . . .	97
6.4	Clustering results by DPC and DPEC on the BS dataset. . . . .	99
6.5	Clustering results of 3-class of gaussian dataset by different methods. . . . .	103
6.6	Clustering results of the <b>SD4</b> dataset by different methods. . . . .	104
6.7	Clustering results of the SD1 dataset by different methods. . . . .	105
6.8	Clustering results of the BS dataset by different methods. . . . .	106
7.1	The basic principles for choosing different evidential methods. . . . .	113
7.2	Clustering results by comparison methods on the SD1 dataset. . . . .	116
7.3	Clustering results by comparison methods on the SD2 dataset. . . . .	117
7.4	Clustering results by comparison methods on the SD3 dataset. . . . .	118
7.5	Clustering results by comparison methods on the SD4 dataset. . . . .	119
7.6	Clustering results by comparison methods on the SD5 dataset. . . . .	120
7.7	Clustering results by comparison methods on the SD6 dataset. . . . .	121
7.8	Clustering results by comparison methods on the SD7 dataset. . . . .	122
7.9	Clustering results by comparison methods on the SD8 dataset. . . . .	123
7.10	Clustering results by different methods on the Olivetti face dataset. . . . .	125
7.11	Clustering results by different methods on the BS dataset. . . . .	126
7.12	Clustering results by DPEC with different $\mathcal{K}_1$ on the SD1 dataset. . . . .	127
7.13	Clustering results by DPEC with different $\mathcal{K}_2$ on the SD1 dataset. . . . .	127
7.14	Clustering results by DPEC with different $\zeta$ on the SD1 dataset. . . . .	128

# Abbreviations and notations

In the following, a list as exhaustive as possible of basic abbreviations and notations used in this thesis:

## Basic clustering

- $\Omega$ : is the frame of discernment (FoD);
- $\{\omega_1\}, \{\omega_2\}, \dots, \{\omega_c\}$ : hypothesis in  $\Omega$ ; they are singletons;
- $\mathcal{X}$ : the set of objects in the analyzed dataset;
- $\mathbf{x}_i$ : denotes the  $i$ -th object in  $\mathcal{X}$ ;
- $x_{ij}$ : denotes the  $j$ -th attribute in  $\mathbf{x}_i$ ;
- $c$ : the number of clusters in  $\mathcal{X}$ ;
- $n$ : the number of objects in  $\mathcal{X}$ ;
- $d_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|$ : denotes the Euclidean distance between objects  $\mathbf{x}_i$  and  $\mathbf{x}_j$ .
- $A$  or  $A_j$ : is a focal set, and it can be a singleton or a compound focal set;
- $|A_j|$ : is the cardinality of  $A_j$ ;
- $M$ : is the mass function matrix;
- $V$ : is the center matrix;
- $m_{ij}, m(\{\omega_j\})$ : is the mass of belief of object  $\mathbf{x}_i$  on focal set  $\{\omega_j\}$  or  $A_j$ ;
- $m_i$ : is the mass function of object  $x_i$ ;
- $\mathbf{v}_j$ : is the center of singleton cluster  $\{\omega_j\}$ ;
- $\bar{\mathbf{v}}_j$ : is the center of the cluster that is associated with focal set  $A_j$ ;

## Belief functions

- $2^\Omega$ : the power-set on the FoD  $\Omega$ ;
- $\omega_{ij} \triangleq \{\omega_i, \omega_j\}$ : the compound focal set including  $\{\omega_i\}$  and  $\{\omega_j\}$ ;
- $m, m_j$ : is a mass function,  $m$  is a mass function defined on any frame of discernment  $\Omega$ ;  $m_j$  is the mass function provided by a source  $j$ ;
- $\mathcal{K}_{ij} = 1 - m_{1 \cap 2}(\emptyset)$ : is the degree of conflict between  $m_i$  and  $m_j$ ;
- $Bel(A)$ : is the belief function  $Bel$  on the focal set  $A_j$ ;
- $Pl(A)$ : is the plausibility function  $Pl$  on the focal set  $A_j$ ;
- $\oplus$ : is the resulting operation of Dempster's rule of combination;
- $BetP_m(\{\omega_j\})$ : is the pignistic probability function  $BetP_m$  on the focal set  $\{\omega_j\}$ .

## Related to concepts and definitions

- $\pi$ : is a possibility distribution (function) in possibility theory;
- $\Pi$ : is a possibility measure in possibility theory;
- $N$ : is a necessity measure in possibility theory;
- $f(\mathbf{x}_i)$ : is the probability function in probability theory;
- $p_{ij}$ : is the probability of the event  $\mathbf{x}_i \in A_j$ ;
- $\Psi = (\psi_{ij})$ : is a  $n \times n$  dissimilarity matrix;
- $\psi_{ij}$ : is the degree of dissimilarity between objects  $\mathbf{x}_i$  and  $\mathbf{x}_j$ ;
- $\Theta = \{\theta_1, \dots, \theta_d\}$ : is a FoD including  $d$  events or classes;
- $u_{ij}^{(f)}$ : is the fuzzy membership degree of object  $\mathbf{x}_i$  in class  $\{\theta_j\}$ ;
- $u_{ij}^{(p)}$ : is the possibilistic membership degree of object  $\mathbf{x}_i$  in class  $\{\theta_j\}$ .

## Related to DEC

- $\mathcal{M}_i$ : is the set of potential singleton clusters for  $\mathbf{x}_i$ ;
- $\mathcal{M}_i^\Omega$ : is the new frame for  $\mathbf{x}_i$  under the TBF;
- $n_1$ : is the number of imprecise objects;
- $\beta$ : is the weighting exponent;

- $\gamma$ : is the weighting factor of the distance;
- $\varphi$ : is a chosen meta-cluster threshold.

### Related to BSC

- $\tilde{\Omega}$ : is a FoD describing  $\mathbf{x}_i$  as the cluster center ( $\mathcal{C}$ ) or unknown ( $\bar{\mathcal{C}}$ );
- $\mathcal{X}_{im}$ : is the set of imprecise objects;
- $\mathcal{X}_{pr}$ : is the set of precise objects;
- $\mathcal{T}_i$ : is the total number of searched times for  $\mathbf{x}_i$ ;
- $\bar{\mathcal{T}}$ : is the average of  $\mathcal{T}_i$  for all objects in  $\mathcal{X}$ ;
- $\mathcal{F}_i$ : is a new edited dynamic framework for  $\mathbf{x}_i$  under the TBF;
- $[\cdot]$ : is the rounding symbol;
- $K_1$ : is the number of neighbors of  $\mathbf{x}_i$ ;
- $K_2$ : is number of neighbors of  $\mathbf{x}_i$  in each related cluster  $\omega_l$ ;
- $n_1$ : is the number of imprecise objects;
- $n_2$ : is the number of precise objects;
- $\mathbf{y}_k$ : is the  $k$ -th neighbor of  $\mathbf{x}_i$ ;
- $\alpha$ : is the outlier adjustment factor;
- $\beta$ : is the weighting exponent;
- $\gamma$ : is the weighting factor of the distance;

### Related to DPEC

- $\Omega_{\mathcal{C}}$ : is a FoD describing  $\mathbf{x}_i$  as the cluster center ( $\mathcal{C}$ ) or unknown ( $\mathcal{U}$ );
- $\mathcal{S}_{\mathcal{K}_1}(\mathbf{x}_i)$ : is the set of neighbors of  $\mathbf{x}_i$ ;
- $\mathcal{S}_{\mathcal{K}_2}(\mathbf{x}_i)$ : is the set of potential singleton clusters for  $\mathbf{x}_i$ ;
- $\mathcal{K}_1$ : is the number of neighbors in  $\mathcal{S}_{\mathcal{K}_1}(\mathbf{x}_i)$ ;
- $\mathcal{K}_2$ : is the number of neighbors in each related cluster  $\omega_\phi$ ;
- $\Phi$ : is the number of clusters in  $\mathcal{S}_{\mathcal{K}_2}(\mathbf{x}_i)$ ;

- $\rho_i$ : is the local density in DPC;
- $d_c$ : is a cutoff distance in DPC;
- $\gamma_i$ : is a quantity, for example,  $\gamma_i = \rho_i \delta_i$  in DPC;
- $\delta_i$ : is the minimum distance;
- $\vartheta_i$ : is an adjustable distance factor;
- $\rho_f$ : is the distance  $\delta_i$ -based function;
- $\zeta$ : is a threshold controlling the number of objects assigned to meta-clusters.

# 1

## Introduction

**Abstract:** In this chapter, we first state the motivation for the research of this thesis, i.e., representation of uncertainty and imprecision between clusters with the theory of belief functions. After that, the main objectives of this thesis are discussed. Then, we highlight four significant contributions of this thesis. Finally, the structure of this thesis is presented in detail.

### Contents

---

<b>1.1</b>	<b>Motivations . . . . .</b>	<b>2</b>
<b>1.2</b>	<b>Objectives . . . . .</b>	<b>3</b>
<b>1.3</b>	<b>Contributions . . . . .</b>	<b>4</b>
<b>1.4</b>	<b>Structure of the thesis . . . . .</b>	<b>6</b>

---

## 1.1 Motivations

Clustering analysis, also called unsupervised learning, has been widely used as an exploratory data analysis tool in many fields (Jain et al., 1999; Saxena et al., 2017). The goal of clustering is to group a set of objects  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  into  $c$  (small value) clusters  $\Omega = \{\omega_1, \dots, \omega_c\}$  in such a way that objects in the same cluster are as similar as possible while objects in different clusters are as dissimilar as possible. To measure the similarities (or dissimilarities), the datasets are described by either object data or relational data. Object data are described explicitly by a  $p$ -dimensional vector. For relational data, the available information arises from the pairwise similarities or dissimilarities, which is usually stored in an  $n \times n$  matrix named the similarity (or dissimilarity) matrix. A clustering algorithm is usually formulated by taking the requirements of the particular task and the nature of the dataset to be handled into consideration. A number of clustering technologies with various philosophies have been proposed, such as objective functions-based (Jain, 2010; Bezdek, 2013), distribution-based (McLachlan and Krishnan, 2007), density-based methods (Fayyad et al., 1996; Rodriguez and Laio, 2014). These above methods achieve good results for clusters, however, they ignore a very common problem in some applications. There may be some uncertain and imprecise information between these clusters. Imprecision refers to that the cluster information of objects distributed in the overlapping or intermediate areas of different specific clusters may be imprecise or ambiguous. Uncertainty refers to that the process of assigning objects to these clusters may be uncertain.

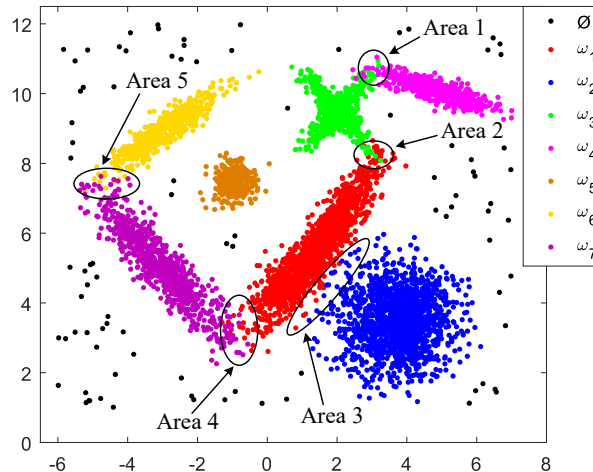


Fig. 1.1. Illustration of uncertainty between different clusters.

As a simple example, Fig. 1.1 shows a 7-class problem with two dimensions. It can easily find that the objects, distributed in the overlapping areas (*e.g.* Area 1) or intermediate areas (*e.g.* Area 3) of different clusters, are difficult to be recognized by specific clusters. It is unreasonable to accurately assign these objects to specific

clusters only depending on the current knowledge and it may increase the risk of errors if they are forced to be assigned. Thus, there may be some uncertainty and imprecision between these clusters. In such a case, these objects are imprecise and can be assigned to the union of some specific clusters, which is further explained as follows.

In clustering problems, one object simultaneously close to several specific clusters can be difficult to classify correctly since these close specific clusters appear not very distinguishable for this object. In such a case, we consider that the object can belong to the union of these specific clusters, called meta-cluster. In contrast, if an object is significantly close to a specific cluster, we assign it to that cluster. By doing this, we can characterize the uncertainty and imprecision in the clustering process and results.

Recently, a new way, called evidential clustering (EC), is appealing for dealing with such a case based on the theory of belief functions (TBF) (Shafer, 1976; Dempster, 2008; Smets and Kennes, 1994; Smets, 1990). TBF, also known as Dempster-Shafer theory or evidence theory, is famous for dealing with uncertain and imprecise information, and EC inherits this ability. Inspired by hard, fuzzy, and possibilistic clustering (Jain et al., 1999; Bezdek, 2013; Pal et al., 2005), Dencœux and Masson first define this way (EC) on a new concept, named *credal partition* (Dencœux and Masson, 2004; Masson and Dencœux, 2008). It can generate three clusters, named singleton (specific) cluster, meta-cluster, and noise cluster. Meta-cluster, considered as a new cluster, is defined as the union of several singleton (specific) clusters. In *credal partition*, it allocates, for each object, a “mass of belief”, not only to singleton clusters but also to any meta-clusters of  $\Omega = \{\omega_1, \dots, \omega_c\}$ . Experiments have shown that, compared with other ones, this additional flexibility allows to gain a deeper insight into the data and to improve robustness with respect to outliers. However, there are still some issues with EC. First, as a variant of other classical methods, EC is a relatively new branch. The differences and connections between EC and other classical methods need to be further clarified. Second, EC extends  $\Omega$  to the power-set  $2^\Omega$  to offer flexible and rich tools to model uncertainty and imprecision. Still, the introduction of meta-clusters also brings high complexity in classical EC compared with hard, fuzzy, and possibilistic clustering. Third, EC cannot handle uncertainty and imprecision between imbalanced clusters (clusters with different sizes) or clusters with arbitrary shapes and sizes. These issues hinder the applications of EC in different scenarios. Therefore, in this thesis, we aim to give some alternatives to address these issues.

## 1.2 Objectives

The goal of this thesis is to normalize some basics of evidential clustering (EC) and improve the ability of EC in characterizing uncertainty and imprecision between clusters, especially for imbalanced or clusters with arbitrary shapes and sizes. To achieve this goal, we first review the origin of EC, related concepts and definitions, and give the



differences and connections with other methods. Then, for the limitations of classical EC methods, including high complexity and unable to handle imbalanced clusters and arbitrary clusters, we propose corresponding solutions for each to improve the ability of EC in characterizing uncertainty and imprecision between clusters. More precisely, we propose a dynamic evidential clustering algorithm to solve the high computational complexity of classical methods. Afterward, we present two new algorithms to deal with imbalanced clusters and clusters with arbitrary shapes and sizes, respectively.

### 1.3 Contributions

In the process of achieving our objectives, we have proposed several contributions in the field of clustering, which can be summarized in the following four parts.

- **Normalization of concepts, definitions, and methodologies of evidential clustering** A systematic study of evidential clustering (EC) is presented. We first give the concepts and definitions of data (inputs), methods (models), and clusters (results) of EC, as well as that of the other types (*i.e.* hard, fuzzy, possibilistic ones), based on different theories (*i.e.* probability theory, fuzzy set theory, possibility theory, and the theory of belief functions). According to these concepts and definitions, we analyze the differences and connections between EC and other types and thus declare the advantages of EC in characterizing uncertainty and imprecision between clusters. Afterward, we also study EC from the seminal to state-of-the-art methods in the context of data-based inputs, resulting in a coherent and comprehensive survey to analyze these methods. Based on the analysis of some representative techniques from different perspectives (*e.g.* center, measure type, complexity), we provide a guiding scheme to help users choose appropriate methods in their cases.

– This work is in preparation and is detailed below.

Zuo-wei Zhang, Arnaud Martin, Yi-ru Zhang, Zhun-ga Liu, Kuang Zhou. A survey of evidential clustering: definitions, methods, and applications. Under Review.

- **DEC: Dynamic evidential clustering algorithm** A dynamic evidential clustering algorithm (DEC) is introduced to address the computational burden of classical methods. In DEC, an FCM-like objective function is first employed and minimized to obtain the support levels of the real singletons (specific) clusters to which the query objects belong. Then, the query object is initially adaptively assigned to the outlier, precise or imprecise one *via* a new rule based on the conflicts between the different support levels. Each imprecise object is finally reassigned to the singleton clusters or related meta-cluster by partial credal redistribution with the corresponding dynamic edited framework to reduce the computational

burden. The proposed DEC method can reduce the complexity to a level similar to that of fuzzy and possibilistic clustering, which can effectively extend the applications of EC, especially in big data. The effectiveness of the DEC method is tested by four experiments with artificial and real datasets.

– This work has been published and the details are as follows.

Zuo-wei Zhang, Zhe Liu, Arnaud Martin, Zhun-ga Liu, Kuang Zhou. Dynamic evidential clustering algorithm. *Knowledge-based systems*, 213:106643, 2021. DOI: 10.1016/j.knosys.2020.106643.

- **BSC: Belief shift clustering** A new belief shift clustering (BSC) algorithm is introduced to deal with imbalanced data *via* extending mean shift under the TBF, which mainly contains two characteristics. First, the query object is preliminarily assigned as the noise, precise, or imprecise one based on the notion of “belief shift”. Second, partial credal redistribution with dynamic cluster centers, to avoid the “uniform effect”, is established to reassign imprecise objects to the singleton cluster or related meta-cluster. Once an object is assigned to a meta-cluster, it indicates that the singleton clusters involved in the meta-cluster cannot be distinguished because this object may be located in the overlapping or intermediate areas of these singleton clusters. By doing this, the BSC can reasonably characterize the uncertainty and imprecision between imbalanced clusters. The effectiveness of BSC has been validated on several synthetic and real datasets by critically comparing with some related methods.

– This work is in the second round of review and is detailed below.

Zuo-wei Zhang, Zhun-ga Liu, Arnaud Martin. BSC: Belief shift clustering. Under Review.

- **DPEC: Density peaks-based evidential convergence** We investigate the representation of uncertainty and imprecision between clusters regardless of their shape, size, and dimensionality based on density peaks clustering and the TBF. First, we consider that different neighbors can provide complementary evidence supporting the object as a cluster center and redefine a distance-based density function to obtain more robust cluster centers in the decision graph. Then, we present a new evidential convergence rule to assign the remaining objects to different clusters. Similar to BSC, the objects located in the overlapping or intermediate areas of different singleton clusters are assigned to corresponding meta-clusters to characterize the uncertainty and imprecision between these arbitrary clusters. The effectiveness of DPEC is validated by comparing with related methods.

– This work is under review and the details are as follows.

Zuo-wei Zhang, Zhun-ga Liu, Liang-bo Ning, Arnaud Martin, Kuang Zhou. Representation of uncertainty and imprecision between clusters with arbitrary shapes and sizes. Under Review.

## 1.4 Structure of the thesis

The thesis is organized in the following six chapters:

In Chapter 2, some related preliminary knowledge is introduced, including the theory of belief functions (TBF) and other uncertainty theories, such as probability theory, fuzzy set theory and possibility theory.

In Chapter 3, the concepts and definitions of data, clusters, and methods of EC and other type methods, *i.e.* hard, fuzzy, and possibilistic clustering, are presented. These concepts and definitions are used as benchmarks for the studies of EC. Besides, the differences and connections between EC and these types are discussed, thereby declaring EC's advantages in characterizing uncertainty and imprecision between clusters. In addition, clustering techniques based on TBF are also reviewed and discussed from the seminal to state-of-the-art methods.

In Chapter 4, a dynamic evidential clustering algorithm is presented to make *credal partitions* converge quickly and characterize the uncertainty and imprecision between spherical clusters. It is regarded as an improvement of classical EC in terms of reducing the computational complexity.

Chapter 5 and Chapter 6 present two different algorithms to detect non-spherical clusters, *i.e.* imbalanced clusters and arbitrary clusters. In Chapter 5, a new belief shift clustering algorithm is introduced to deal with imbalanced clusters *via* extending mean shift under the TBF. In Chapter 6, we investigate the representation of uncertainty and imprecision between clusters regardless of their shape, size, and dimensionality based on density peaks clustering and the TBF.

Chapter 7 concludes this thesis and presents some perspectives.

# 2

## Background

**Abstract:** In this chapter, we mainly introduce some basics and related concepts used in this thesis, including the theory of belief functions and other theories related to dealing with uncertainty.

### Contents

---

<b>2.1</b>	<b>Introduction</b>	<b>8</b>
<b>2.2</b>	<b>The theory of belief functions</b>	<b>8</b>
<b>2.3</b>	<b>Some links with other related theories</b>	<b>10</b>
2.3.1	Probability theory	10
2.3.2	Fuzzy set theory	11
2.3.3	Possibility theory	11
2.3.4	Rough set theory	12
<b>2.4</b>	<b>Conclusion</b>	<b>12</b>

---

## 2.1 Introduction

Due to the importance of clustering in various tasks of exploratory data mining, considerable efforts have been made to develop clustering algorithms by scholars or engineers from multiple disciplines. Common open issues in clustering analysis include how to determine the number of clusters, robustness, scalability, and additional challenges in clustering some specific types of datasets. Instead of discussing general issues and providing an exhaustive survey of various clustering approaches, we concentrate mainly on issues and clustering approaches related to credal partition. To provide a basis for subsequent studies, the review of some related background knowledge in this chapter contains the following two components: the theory of belief functions (TBF) in Section 2.2, and some links with related theories in Section 2.3, such as Probability theory, Fuzzy set theory, and Possibility theory.

## 2.2 The theory of belief functions

In this section, we briefly recall some basic notions of the theory of belief functions (TBF), also well-known as Dempster-Shafer theory or evidence theory, (Shafer, 1976; Dempster, 2008; Smets and Kennes, 1994; Denœux, 2008a, 2016; Shafer, 2016; Smets, 2005). In this theory, the set of values that an uncertain quantity can take is defined as a frame of discernment (FoD) consisting of  $c$  events. In FoD, the  $c$  events are exclusive and exhaustive.

**Definition 2.1** *The frame of discernment (FoD) is a finite set of disjoint elements, defining the domain of reference, formally:*

$$\Omega = \{\omega_1, \dots, \omega_c\}, \quad (2.1)$$

where  $\{\omega_j\}$  are exclusive and exhaustive.

The TBF extends this FoD to the power-set  $2^\Omega$  to offer flexible and rich tools to model uncertainty and imprecision.

**Definition 2.2** *Given a FoD  $\Omega$ , the power-set  $2^\Omega$  is defined by:*

$$2^\Omega = \{\emptyset, \{\omega_1\}, \dots, \{\omega_1, \omega_2\}, \dots, \Omega\}. \quad (2.2)$$

If we have  $c = 3$ , for example,  $2^\Omega = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_3\}, \Omega\}$ .

For generalizing well-known models such as probabilities or possibility distributions, an *evidential body* or *basic belief assignment (BBA)* is given by a mass function.

**Definition 2.3** Given a FoD  $\Omega$ , a function  $m$  is defined as a mapping from  $2^\Omega \rightarrow [0,1]$  and called mass function on  $2^\Omega$ , such that:

$$\sum_{A \subseteq \Omega} m(A) = 1. \quad (2.3)$$

The subsets  $A$  of  $\Omega$  such that  $m(A) > 0$  are called the *focal sets* of  $m$  with  $A$  named *focal element*. A mass function  $m$  is called normalized if we have  $m(\emptyset) = 0$ .

For any *evidential body* or *BBA*  $m$ , there are other equivalent representations of a mass function such as the *belief*, and *plausibility* functions to represent different cases (Shafer, 1976; Dempster, 2008).

**Definition 2.4** Given a function  $m$  and focal elements  $A$  and  $B$ , the belief function  $Bel$  and plausibility function  $Pl$  are defined by:

$$Bel(A) = \sum_{\emptyset \neq B \subseteq A} m(B), \quad (2.4)$$

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B). \quad (2.5)$$

$Bel$  and  $Pl$  respectively represent the lower bound and upper bound of the support degree to each event  $A$  in  $m$ . The combination of mass functions plays a critical role in uncertain information fusion, called *Dempster's rule of combination*.

**Definition 2.5** Let  $m_1$  and  $m_2$  be two mass functions from two different sources that are reliable and cognitively independent. The conjunctive combination of  $m_1$  and  $m_2$  yields the unnormalized mass function:

$$m_{1 \cap 2}(A) = \sum_{B \cap C = A} m_1(B)m_2(C), \forall A \subseteq \Omega. \quad (2.6)$$

If necessary, the normality condition  $m(\emptyset) = 0$  may be recovered by dividing each mass  $m_{1 \cap 2}(A)$  by  $1 - m_{1 \cap 2}(\emptyset)$ . The resulting operation is noted  $\oplus$  and is called *Dempster's rule of combination*:

$$m_{1 \oplus 2}(A) = \frac{m_{1 \cap 2}(A)}{1 - m_{1 \cap 2}(\emptyset)}, \emptyset \neq A \subseteq \Omega. \quad (2.7)$$

Both rules are commutative, associative and admit the total ignorance  $\Omega$  as a unique neutral element.

In the TBF, a well-known pignistic probability transformation method (Smets and Kennes, 1994; Smets, 2005), which approximates a mass function to probabilities, is usually used for decision-making. The pignistic probability of the singleton class  $\{\omega_j\}$  is defined as follows.

**Definition 2.6** Let  $m$  be a mass function on  $\Omega$ , a pignistic probability function  $BetP_m: \Omega \rightarrow [0, 1]$  corresponding to  $m$  is defined by:

$$BetP_m(\{\omega_j\}) = \sum_{A \subseteq 2^\Omega, \{\omega_j\} \in A} \frac{1}{|A|} \frac{m(A)}{1 - m(\emptyset)}, \quad m(\emptyset) \neq 1 \quad (2.8)$$

where  $|A|$  is the cardinality of the subset  $A$ , i.e. the number of elements included in  $A$ .

The DS rule is inapplicable when there is a high conflict between different evidence. Thus, there have been many evidence combination methods for DS rule improvements, such as the well-known Smets' conjunctive rule (used in his TBM (Smets, 1990, 2007)), Dubois-Prade (DP) rule (Dubois and Prade, 1988), and the very sophisticated proportional conflict redistributions (PCR) rules (Smarandache and Dezert, 2005, 2013; Martin, 2019). However, these methods may still have some problems in applications, and we will propose some new fusion rules according to the actual scenarios in the subsequent works.

## 2.3 Some links with other related theories

In this section, we introduce three common theories characterizing uncertainty that are both distinct from and related to TBF. Many clustering methods based on these theories have been developed, which will be discussed in detail in the next chapter.

### 2.3.1 Probability theory

Probability theory is the branch of mathematics concerned with probability (Jeffreys, 1998; Durrett, 2019). Although there are several different probability interpretations, probability theory treats the concept in a rigorous mathematical manner by expressing it through a set of axioms. Typically these axioms formalize probability in terms of a probability space, which assigns a measure taking values between 0 and 1, termed the probability measure, to a set of outcomes called the sample space.

**Definition 2.7** Given a FoD  $\Omega$ , it is assumed that for each element  $\mathbf{x}_i \in \Omega$ , an intrinsic "probability" value  $f(\mathbf{x}_i)$  is attached, which satisfies the following properties:

$$\begin{cases} f(\mathbf{x}_i) \in [0, 1], \mathbf{x}_i \in \Omega; \\ \sum_{\mathbf{x}_i \in \Omega} f(\mathbf{x}_i) = 1. \end{cases} \quad (2.9)$$

We can find that the probability function  $f(\mathbf{x}_i)$  lies between 0 and 1 for  $\mathbf{x}_i$ , and the sum of  $f(\mathbf{x}_i)$  over all  $\mathbf{x}_i$  in  $\Omega$  is equal to 1. An event is defined as any subset  $A_j$  of  $\Omega$ .

**Definition 2.8** *The probability of the event  $\mathbf{x}_i \in A_j$ , is defined as:*

$$p_{ij} = \sum_{\mathbf{x}_i \in A_j} f(\mathbf{x}_i). \quad (2.10)$$

In this case, the probability of the FoD is 1, and the probability of the null event is 0. The function  $f(\mathbf{x}_i)$ , mapping a point in the FoD to the “probability” value is called a probability mass function. That is, mass functions generalize probabilities.

### 2.3.2 Fuzzy set theory

In mathematics, fuzzy sets are sets whose elements have degrees of membership. In classical set theory (Zadeh, 1996), the membership of an element in a set is either 1 or 0 (either belongs or does not belong to the set). By contrast, fuzzy set theory permits the gradual assessment of the membership of elements in a set with the help of a membership function valued in the interval  $[0, 1]$ . The classical bivalent sets are usually called crisp (hard) sets. Fuzzy sets generalize classical sets, since the indicator functions of classical sets are special cases of the membership functions of fuzzy sets, if the latter only take values 0 or 1.

### 2.3.3 Possibility theory

Possibility theory is another popular choice for representing uncertain information (Dubois and Prade, 2012, 1988). At the semantic level, a basic function in possibility theory is a possibility distribution denoted as  $\pi$  which assigns each possible world in the FOD  $\Omega$  with a value in  $[0,1]$  (or a set of graded values). From a possibility distribution, two measures are derived, a possibility measure (denoted as  $\Pi$ ) and a necessity measure (denoted as  $N$ ). The former estimates to what extent the true event is believed to be in the subset and the latter evaluates the degree of necessity that the subset is true.

**Definition 2.9** *Given  $\pi$ ,  $\Pi$ , and  $N$ , we have:*

$$\Pi(A) = \max(\{\pi(\omega) | \omega \in \Omega\}), \quad N(A) = 1 - \Pi(A). \quad (2.11)$$

$$\Pi(\Omega) = 1, \quad \Pi(\emptyset) = 0. \quad (2.12)$$

$$\Pi(A \cup B) = \max(\Pi(A), \Pi(B)), \quad N(A \cap B) = \min(N(A), N(B)). \quad (2.13)$$

A belief function is called a consonant function if its focal elements are nested (Shafer, 1976). That is, if  $A_1, A_2, \dots, A_n$  are focal elements then  $A_1 \subset A_2 \subset \dots \subset A_n$ .



**Definition 2.10** *Suppose the credibility function and plausibility function of the consonant bba are Bel and Pl, we have:*

$$Bel(A \cap B) = \min(Bel(A), Bel(B)), \quad Pl(A \cup B) = \max(Pl(A), Pl(B)). \quad (2.14)$$

These two properties are exactly the requirements of necessity and possibility measures in possibility theory. Necessity and possibility measures are special cases of credibility and plausibility functions when the focal sets of BBA are nested. The TBF thus, in a sense, is more general than the theory of possibilities.

### 2.3.4 Rough set theory

Rough set theory is another mathematical tool to deal with uncertainty (Pawlak, 1982, 1991; Pawlak and Skowron, 2007). The rough set concept can be defined quite generally by means of topological operations, interior and closure, called approximations. Given a space represented by the pair  $(U, R)$ ,  $U$  is a universe of objects, and  $R \subseteq U \times U$  is an indiscernibility relation defined by an attribute set (*i.e.*  $R = I(A)$  for some attribute set  $A$ ). In this case,  $R$  is the equivalence relation. Let  $[\mathbf{x}]_R$  denote an equivalence class of an element  $\mathbf{x} \in U$  under the indiscernibility relation  $R$ , where  $[\mathbf{x}]_R = \{\mathbf{y} \in U : \mathbf{x}R\mathbf{y}\}$ .

In this context,  $R$ -approximations of any set  $X \subseteq U$  are based on the exact (crisp) containment of sets. Then set approximations are defined as follows:

- $\mathbf{x} \in U$  belongs with certainty to  $X \subseteq U$  (*i.e.*  $\mathbf{x}$  belongs to the  $R$ -lower approximation of  $X$ ), if  $[\mathbf{x}]_R \subseteq X$ .
- $\mathbf{x} \in U$  possibly belongs  $X \subseteq U$  (*i.e.*  $\mathbf{x}$  belongs to the  $R$ -upper approximation of  $X$ ), if  $[\mathbf{x}]_R \cap X \neq \emptyset$ .
- $\mathbf{x} \in U$  belongs with certainty neither to the  $X$  nor to  $U - X$  (*i.e.*  $\mathbf{x}$  belongs to the  $R$ -boundary region of  $X$ ), if  $[\mathbf{x}]_R \cap (U \cap X) \neq \emptyset$  and  $[\mathbf{x}]_R \cap X \neq \emptyset$ .

## 2.4 Conclusion

In this chapter, we provided some background knowledge on uncertainty theories. Specifically, we start this chapter by introducing the TBF. Then we describe the links between TBF and other related theories. In the sequel of this thesis, some of the challenges mentioned will be solved by some alternative solutions.

# 3

## Concepts, definitions and techniques of evidential clustering

**Abstract:** In this chapter, we redefine the concepts related to evidential clustering based on different theories from three perspectives: input (data), model (method), and output (cluster). We also explain that uncertainty and imprecision can be characterized simultaneously by evidential clustering and discuss the differences and connections between evidential clustering and other related clustering methods. After that, we review the existing works related to evidential clustering, from the seminal to state-of-the-art methods. Finally, the chapter is concluded.

### Contents

---

<b>3.1</b>	<b>Introduction</b>	<b>15</b>
<b>3.2</b>	<b>Concepts and definitions of data</b>	<b>15</b>
3.2.1	Hard data	16
3.2.2	Fuzzy data	16
3.2.3	Possibilistic data	17
3.2.4	Evidential data	18
3.2.5	The difference and relations of different data	19
<b>3.3</b>	<b>Concepts and definitions of clusters</b>	<b>21</b>
3.3.1	Hard clusters	22
3.3.2	Fuzzy clusters	22
3.3.3	Possibilistic clusters	23
3.3.4	Evidential clusters	23
3.3.5	The difference and relations of different clusters	25
<b>3.4</b>	<b>Concepts and definitions of methods</b>	<b>26</b>
3.4.1	Hard methods	27
3.4.2	Fuzzy methods	27
3.4.3	Possibilistic methods	28
3.4.4	Evidential methods	29
3.4.5	The overview of different data, methods, and clusters	30
<b>3.5</b>	<b>Clustering techniques based on TBF</b>	<b>30</b>

3.5.1	Object data-based methods . . . . .	31
3.5.2	Relational data-based methods . . . . .	33
3.5.3	Applications of evidential methods . . . . .	34
3.5.4	Analysis of different evidential methods . . . . .	35
3.5.5	Challenges of evidential clustering . . . . .	36
<b>3.6</b>	<b>Conclusion . . . . .</b>	<b>37</b>

---

### 3.1 Introduction

In the previous chapter, we introduced the basics of different theories, *i.e.* probability theory, fuzzy set theory, possibility theory, rough set theory. These theories are widely used in data mining and machine learning because of their ability to handle uncertainty well. Many different clustering methods that deal with uncertainty between clusters have also been proposed based on these theories, such as rough clustering (Lingras and Peters, 2011; Peters et al., 2012). Besides, some methods are not based on these theories but also consider the uncertainty between clusters, such as overlapping clustering (Banerjee et al., 2005; Baadel et al., 2016). All these methods have achieved good clustering results. Since evidential clustering (EC) (Masson and Dencœux, 2008) is considered as an extension of hard/fuzzy/possibilistic  $c$ -means under the theory of belief functions (TBF), we only consider the underlying theories and the associated clustering algorithms in this thesis, *i.e.* hard clustering (Jain, 2010), fuzzy clustering (Bezdek et al., 1984), and possibilistic clustering (Krishnapuram and Keller, 1993). Although EC inherits the advantages of TBF in characterizing uncertainty and imprecision, as one of the new branches of cluster analysis, still EC has some concepts and definitions that need to be standardized. Therefore, we aim to give some related concepts and definitions to regulate the subsequent in-depth study here. Specifically, this chapter contains three sections. They are about the concepts and definitions of data in Section 3.2, clusters in Section 3.3, and methods in Section 3.4 under different theoretical frameworks respectively. These concepts and definitions correspond to the inputs (data), models (methods), and outputs (clusters).

### 3.2 Concepts and definitions of data

In clustering tasks, empirical information can generally be grouped into two families, namely, certain and precise information, uncertain and imprecise information, which are normalized and characterized by “data” due to different requests. In the subsequent research, we focus on modeling and characterizing the uncertainty and imprecision between clusters, so the corresponding definitions should be given first. Here uncertainty is defined as a state of limited (insufficient) information (knowledge) where it is impossible to describe the existing state exactly. By contrast, imprecision is defined as a state of fuzzy (imprecise) information (knowledge) where it is impossible to describe the existing state precisely. Hard data is generally used to represent certain and precise empirical information. By contrast, fuzzy, possibilistic, and evidential data are used to represent uncertain and imprecise empirical information. In this section, basic concepts and definitions of hard, fuzzy, possibilistic, and evidential data are given as follows.

### 3.2.1 Hard data

In clustering, the certain and precise information records in the form of “data” for further analysis, which can be observations, items, or feature vectors, etc. Here we refer to these data as *hard data*, such as object data, and relational data<sup>1</sup>. That is, given a dataset  $\mathcal{H}$ , if regarded as a hard one, there is no uncertainty and no imprecision on  $\mathcal{H}$ . For example, object data is often presented in the form of feature vectors, and can be defined as follows.

**Definition 3.1** *Given a dataset  $\mathcal{X}$  with  $n$  objects, and each object is explicit described using  $s$  numeric attributes. In this case, the dataset  $\mathcal{X}$  mathematically defined by:*

$$\mathcal{X} = \{\mathbf{x}_i = (x_{i1}, \dots, x_{ip}) : i = 1, \dots, n; p \geq 1\} \quad (3.1)$$

where  $\mathcal{X}$  is an object dataset, and  $\mathbf{x}_i$  is the  $i$ -th object in  $\mathcal{X}$ .

### 3.2.2 Fuzzy data

Following fuzzy set theory (Zadeh, 1996; Zimmermann, 2011), *fuzzy data* is used to represent imprecise/fuzzy information (Coppi et al., 2012; Han et al., 2011). There are also a few definitions of fuzzy data. For example, in many real-life situations, measurements may be imprecise and the observations vaguely defined. These can be appropriately represented by fuzzy numbers (Dubois and Prade, 1978), which in turn evolve into *LR fuzzy data* (Coppi et al., 2012; Denoeux and Masson, 2004). In contrast, since a fuzzy membership function quantifies the degree of membership of the object belonging to different classes, an intuitionistic fuzzy set is proposed based on fuzzy membership-like functions (Atannasov, 1999; Han et al., 2011, 2013). This set can also be considered as fuzzy data if it is used as an input. Since we focus on fuzzy set theory based clustering methods, such as fuzzy  $c$ -means (Bezdek et al., 1984), we consider *fuzzy data* as the set of fuzzy membership functions defined as follows.

**Definition 3.2** *Given a dataset  $\mathcal{X}$  and FoD  $\Theta = \{\theta_1, \dots, \theta_d\}$ , for an object  $\mathbf{x}_i$ , it is used to represent imprecise information for the object  $\mathbf{x}_i$  belonging to different classes by a function  $\mathbf{u}^{(f)}$ . The function  $u_{ij}^{(f)} : \Theta \rightarrow [0, 1]$  is called fuzzy membership function on  $\Theta$ , such that:*

$$\sum_{j=1}^d u_{ij}^{(f)} = 1, u_{ij}^{(f)} \in [0, 1] \quad (3.2)$$

where  $u_{ij}^{(f)}$  represents the fuzzy membership degree of the object  $\mathbf{x}_i$  in the class  $\{\theta_j\}$ .

<sup>1</sup>In the following, object data will be employed as the benchmark (example) when a generalized description of other concepts or definitions is needed.

In fuzzy set theory, the object  $\mathbf{x}_i$  is allowed to belong to different classes at the same time, for example, the events  $\mathbf{x}_i \in \{\theta_j\}$  and  $\mathbf{x}_i \in \{\theta_{j'}\}$ ,  $j \neq j'$ , can occur simultaneously. In fuzzy data,  $u_{ij}^{(f)}$  just represents the fuzzy degree of  $\mathbf{x}_i \in \{\theta_j\}$ . In such a case, if  $u_{ij}^{(f)} \in (0, 1)$ , it is used to characterize the imprecision of the event  $\mathbf{x}_i \in \{\theta_j\}$ . In other words, fuzzy data can represent imprecise information (imprecision of knowledge). Of course, fuzzy data can also represent precise information in a specific case defined as follows.

**Definition 3.3** Given a set  $U^{(f)}$  including  $n$  fuzzy membership functions  $\mathbf{u}^{(f)}$ , the set  $U^{(f)}$  will be a hard one if we have:

$$u_{ij}^{(f)} = 1, \forall i, \exists j \quad (3.3)$$

where  $u_{ij}^{(f)} = 1$  indicates that  $\mathbf{x}_i \in \{\theta_j\}$  is a event with precise information.

### 3.2.3 Possibilistic data

It is not enough to describe the imprecision of information, and in some cases, it is considered necessary to provide a reasonable description of the uncertainty (Aggarwal and Philip, 2008; McClean et al., 2001). Following possibility theory (Dubois and Prade, 2012, 1988), *possibilistic data* can represent uncertain information. In clustering analysis, there are a few shortcomings of a fuzzy membership function in semantic interpretation and applications (Krishnapuram and Keller, 1993). Therefore, in clustering, the fuzzy membership function is further improved and called a possibilistic membership function based on possibility theory. Here we consider *possibilistic data* as the set of possibilistic membership functions defined as follows.

**Definition 3.4** Given a dataset  $\mathcal{X}$  and FoD  $\Theta$ , for an object  $\mathbf{x}_i$ , it is used to represent uncertain information for the object  $\mathbf{x}_i$  belonging to different classes by a function  $\mathbf{u}^{(p)}$ . The function  $u_{ij}^{(p)} : \Theta \rightarrow [0, 1]$  is called *possibilistic membership function* on  $\Theta$ , such that:

$$\sum_{i=1}^d u_{ij}^{(p)} \leq 1, u_{ij}^{(p)} \in [0, 1] \quad (3.4)$$

where  $u_{ij}^{(p)}$  represents the *possibilistic membership degree* of  $i$ -th object in  $j$ -th class.

In possibilistic membership function  $\mathbf{u}^{(p)}$ ,  $\sum_{i=1}^c u_{ij}^{(p)} < 1$  is allowed, which is used to against noise and outliers (Krishnapuram and Keller, 1993). From possibility theory perspective, the object  $\mathbf{x}_i$  belonging to different classes, for example,  $\mathbf{x}_i \in \{\theta_j\}$  and  $\mathbf{x}_i \in \{\theta_{j'}\}$ ,  $j \neq j'$ , are mutually exclusive events. That is, the object  $\mathbf{x}_i$  is only allowed to belong to one class  $\{\theta_j\}$  on the FoD  $\Theta$ , and  $u_{ij}^{(p)}$  is used to represent the possibilistic

degree of  $\mathbf{x}_i \in \{\theta_j\}$ . In such a case, if  $u_{ij}^{(p)} \in (0, 1)$ , it can characterize the uncertainty of the event  $\mathbf{x}_i \in \{\theta_j\}$ . That is, possibilistic data can not only represent uncertain information (uncertainty of knowledge) but can also represent certain information (certainty of knowledge).

**Definition 3.5** Given a set  $U^{(p)}$  including  $n$  possibilistic membership functions  $\mathbf{u}^{(p)}$ , the set  $U^{(p)}$  will be a hard one if we have:

$$u_{ij}^{(p)} = 1, \forall i, \exists j \quad (3.5)$$

where  $u_{ij}^{(p)} = 1$  indicates that  $\mathbf{x}_i \in \{\theta_j\}$  is a event with certain information.

### 3.2.4 Evidential data

To make it easier to analyze a wide variety of information expression, following TBF, a new form, called *evidential data* or *evidence body*, is proposed to unify all the *imprecise* or *uncertain data*, which can characterize in greater depth the uncertainty and imprecision caused by data or model. Therefore, *evidential data* is defined as the set of mass functions. A mass function have been introduced in Eq. (2.3). In specific cases, given a mass function, it can transform into fuzzy, possibilistic, or probability ones, for example, by pignistic probability transformation (Smets and Kennes, 1994; Smets, 2005). Thus, evidential data can unify various kinds of uncertain and imprecise information.

**Example 3.1:** Let consider an object  $\mathbf{x}_i$  identified by different sensors with a FoD  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  and get the corresponding evidence bodies, as shown in Table 3.1.

Table 3.1: Evidence bodies of  $\mathbf{x}_i$  from different sources

	$\emptyset$	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_3\}$	$\{\theta_1, \theta_2\}$	$\{\theta_1, \theta_3\}$	$\{\theta_2, \theta_3\}$	$\Theta$
$m_1$	1	0	0	0	0	0	0	0
$m_2$	0	0.3	0.2	0.5	0	0	0	0
$m_3$	0	0.1	0.05	0.3	0	0.5	0.05	0
$m_4$	0	0.1	0.15	0	0.7	0.02	0.03	0
$m_5$	0	0	0	0	0	0	0	1

We can see from the evidence body  $m_1$  that the object  $\mathbf{x}_i$  does not belong to any of the three classes and is an *outlier*. In this case,  $m_1$  characterizes certain and precise information. By contrast,  $m_5$  makes us believe that  $\mathbf{x}_i$  belongs to any class, but we are completely ignorant. It can be regarded as an imprecise empirical information. For  $m_2$ , we find that  $m(A) = 0$  if  $A = \emptyset$  or  $|A| \neq 1$ . In this case, we have a probability mass function/possibilistic membership function, which can be regarded as possibilistic data. From  $m_3$  and  $m_4$  we can find that the class information is partially ignorant. For  $m_3$ ,  $\mathbf{x}_i$  may belong to  $\{\theta_1\}$  or  $\{\theta_3\}$ , and according to  $m_4$ ,  $\mathbf{x}_i$  may belong to  $\{\theta_1\}$  or  $\{\theta_2\}$ .

Thus, we consider that  $m_3$  and  $m_4$  contain both uncertain and imprecise information. From this numerical example, we can see that it is reasonable to use evidential data to unify different empirical information into one form.

### 3.2.5 The difference and relations of different data

Based on the above analysis, we have Venn diagram of different data in Fig. 3.1.

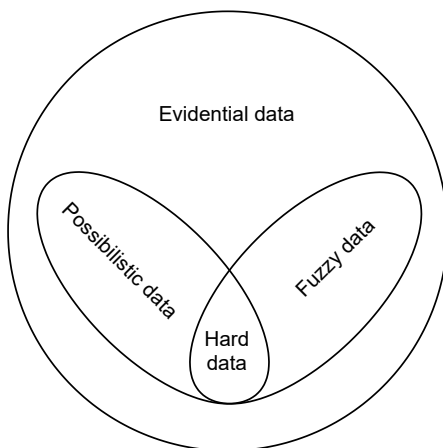


Fig. 3.1. Venn diagram of hard, fuzzy, possibilistic, and evidential data.

To illustrate more intuitively the characterization of different forms of data, we give the following example.

**Example 3.2:** Given a set of classes  $\Theta = \{\theta_1, \theta_2, \theta_3\}$  with “ $\{\theta_1\} = \text{young}$ ”, “ $\{\theta_2\} = \text{adult}$ ”, and “ $\{\theta_3\} = \text{old}$ ”, we discuss the following cases based on 3 people.

**Case 3.1:** Let us consider that one has the basic information of 3 people ( $\mathbf{x}_i$ ,  $i = 1, 2, 3$ ), as shown in Table 3.2.

Table 3.2: Basic information of 3 people

	Age (years)	Salary (euro)	Children (number)
$\mathbf{x}_1$	34	1900	2
$\mathbf{x}_2$	25	1800	0
$\mathbf{x}_3$	42	2200	1

- Assume that the given set  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$  with, for example,  $\mathbf{x}_2 = (25, 1800, 0)$ , then it is regarded as an **object data** or **object dataset**;
- Assume that the given matrix  $\Psi = (\psi_{ii'})_{3 \times 3}$  with, for example,  $\psi_{1 \times 2}$  characterizing the dissimilarity between the objects  $\mathbf{x}_1$  and  $\mathbf{x}_2$ , then this dissimilarity matrix  $\Psi$  is regarded as a **relational data** or **dissimilarity data**. If the squared



*Euclidean*  $\Delta$  is used to measure the dissimilarity between objects, then we have  $\Delta = (\psi_{ii'})$ . In this case,  $\psi_{ii'} = \|\mathbf{x}_i - \mathbf{x}_{i'}\|^2$ , for example,  $\psi_{1 \times 2} = \|\mathbf{x}_1 - \mathbf{x}_2\|^2 = (34 - 25)^2 + (1900 - 1800)^2 + (2 - 0)^2 = 10085$ . Thus, the relational data defined on the matrix  $\Psi$  is given by:

$$\Psi = \begin{bmatrix} 0 & 10085 & 90065 \\ 10085 & 0 & 160290 \\ 90065 & 160290 & 0 \end{bmatrix}.$$

Based on the above example, we can find that there is no uncertainty and no imprecision on hard data.

**Case 3.2:** Let us consider a set of fuzzy membership function  $\mathbf{u}^{(f)}$  on the set of classes  $\Theta$ , as shown Table 3.3.

Table 3.3: Fuzzy data for 3 people

	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_3\}$
$\mathbf{u}_1^{(f)}$	0.3	0.5	0.2
$\mathbf{u}_2^{(f)}$	0.7	0.3	0
$\mathbf{u}_3^{(f)}$	0.2	0.7	0.1

- Following fuzzy set theory, the data defined by a function  $\mathbf{u}^{(f)}$  on  $\Theta$  is considered as a **fuzzy membership function**, and is also regarded as a **fuzzy data**. For example, the fuzzy data  $\mathbf{u}_2^{(f)} = [0.7, 0.3, 0]$  is given for the second person (the object  $\mathbf{x}_2$ ), where  $u_{21}^{(f)} = 0.7$ ,  $u_{22}^{(f)} = 0.3$  represent the fuzzy membership degrees of the second people (the object  $\mathbf{x}_2$ ) belonging to “young” or “adult” (the classes  $\{\theta_1\}$  or  $\{\theta_2\}$ ), respectively. The concepts of “young”, “adult”, and “old” may be fuzzy or imprecise in some cases, for example, we can say that a person is a young or adult. In this case, fuzzy data represent the imprecision of knowledge.

**Case 3.3:** Let us consider a set of possibilistic membership function  $\mathbf{u}^{(p)}$  on the set of classes  $\Theta$ , as shown Table 3.4.

Table 3.4: Possibilistic data for 3 people

	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_3\}$
$\mathbf{u}_1^{(p)}$	0.2	0.7	0.1
$\mathbf{u}_2^{(p)}$	0.9	0.1	0
$\mathbf{u}_3^{(p)}$	0.1	0.8	0.1

- Following possibility theory, the data defined by a function  $\mathbf{u}^{(p)}$  on  $\Theta$  is considered as a **possibilistic membership function**, and is also regarded as a **possibilistic data**. For example, the possibilistic data  $\mathbf{u}_1^{(p)} = [0.2, 0.7, 0.1]$  is given for the

first person (the object  $\mathbf{x}_1$ ), where  $u_{1i}^{(p)}$  represents the possibilistic membership degree of the object  $\mathbf{x}_1$  belonging to the class  $\{\theta_i\}$ ,  $i = 1, 2, 3$ . Although the definitions of “young”, “adult”, and “old” are clear in some cases, limited knowledge may lead to uncertainty. For example,  $u_{11}^{(p)} = 0.2$ ,  $u_{12}^{(p)} = 0.7$ , and  $u_{13}^{(p)} = 0.1$  represent the possibilistic degrees of the object  $\mathbf{x}_1$  belonging to the classes  $\{\theta_1\}$ ,  $\{\theta_2\}$  and  $\{\theta_3\}$ , respectively. Thus, possibilistic data can represent the uncertainty of knowledge.

**Case 3.4:** Let us consider a set of mass functions  $m$  on the power-set  $2^\Theta$ , as shown Table 3.5.

Table 3.5: Evidential data for 3 people

	$\emptyset$	$\{\theta_1\}$	$\{\theta_2\}$	$\{\theta_3\}$	...	$\Theta$
$m_1$	0	0.2	0.3	0	...	0.5
$m_2$	0	0.3	0.0	0	...	0.7
$m_3$	0	0	0.4	0.1	...	0.5

- Following the TBF, the data defined by a function  $m$  on  $2^\Theta$  is considered as a **mass function**, and is also regarded as a **evidential data**. For example, for the object  $\mathbf{x}_3$ , we have the mass function  $m_3$ . In this case,  $m_3(\{\theta_2\})$  represents the degree of the object  $\mathbf{x}_3$  associated with the focal element (class)  $\{\theta_2\}$ , and  $m_3(\{\theta_2\}) = 0.4$  is used to characterize the degree of uncertainty due to limited information.  $\Theta$  represents total ignorance, and  $m_3(\Theta) = 0.4$  is used to characterize the degree of this ignorance. In terms of spatial distribution, once the object  $\mathbf{x}_3$  belongs to the total ignorance  $\Theta$ , it means that  $\mathbf{x}_3$  may be in the overlapping or intermediate region of all the classes (*i.e.*  $\{\theta_1\}, \{\theta_2\}, \{\theta_3\}$ ). In this case,  $\mathbf{x}_3$  may belong to any of these classes contained in  $\Theta_3$ . The class of  $\mathbf{x}_3$  is imprecise or fuzzy at this point. Thus, we can say that evidential data can characterize both uncertainty and imprecision of knowledge.

### 3.3 Concepts and definitions of clusters

In clustering tasks, different data, *i.e.* empirical information, will be grouped into several *clusters* in different ways. In fact, there is no universally agreed upon definition of “*cluster*”, which are also called *groups*, *subsets*, or *categories* in some cases (Xu and Wunsch, 2005). Here we use the word “*cluster*” and consider that a partition of the given dataset  $\mathcal{X}$  is a set of *clusters*. That is, different kinds of clusters give different kinds of partitions, for example, *fuzzy cluster* gives *fuzzy partition*. Typically, given a FoD  $\Omega = \{\omega_1, \dots, \omega_c\}$ , a *hard/fuzzy/possibilistic c-partition* of  $\mathcal{X}$  is usually defined by a set of values  $p_{ij}/u_{ij}$  that can be conveniently arrayed as a  $(c \times n)$  matrix (Pal

et al., 2005), *i.e.* *hard partition* with  $P = (p_{ij})$ , *fuzzy partition* with  $U = (u_{ij}^{(f)})$ , and *possibilistic partition* with  $U = (u_{ij}^{(p)})$ . By contrast, a collection of  $n$  mass functions on the power-set  $2^\Omega$  is called *credal partition* with  $M = (m_{ij})$  (Dencœux and Masson, 2004, 2003). In this section, we will give the concepts and definitions of *clusters* from hard, fuzzy, possibilistic, and evidential perspectives as follows.

### 3.3.1 Hard clusters

In clustering, if a cluster collects a number of objects in a certain/precise way, we call it a *hard cluster*. That is, *hard cluster* is defined as a set consisting of one or more certain/precise objects with some specific properties (both concrete and abstract). The objects in hard cluster have a clear and certain affiliation. Following probability theory (Jeffreys, 1998; Durrett, 2019), we give the definition as follows.

**Definition 3.6** *Given a FoD  $\Omega = \{\omega_1, \dots, \omega_c\}$  and a cluster  $\{\omega_j\}$ ,  $\{\omega_j\} \in \Omega$ , for any object  $\mathbf{x}_i \in \{\omega_j\}$ , the cluster  $\{\omega_j\}$  with  $\mathbf{p}_{\cdot j} : \Omega \rightarrow 0/1$  is called a hard cluster, such that:*

$$p_{ij} = 1, p_{ij'} = 0, \forall \mathbf{x}_i \in \{\omega_j\} \neq \{\omega_{j'}\} \in \Omega \quad (3.6)$$

where  $p_{ij}$  represents the probabilistic degree of  $i$ -th object belonging to  $j$ -th cluster.

Although  $p_{ij} = 1$ ,  $\mathbf{p}_{\cdot j}$  is still constrained by  $\sum_{i=1}^c p_{ij} = 1$ . Essentially,  $\mathbf{p}_{\cdot j} : \Omega \rightarrow 0/1$  is a mapping function based on probability theory. Following this definition, once an object is assigned to a hard cluster, the probability that the object belongs to this cluster is 1. Hence, there is no uncertain/imprecise information in hard clusters.

### 3.3.2 Fuzzy clusters

In clustering, unlike hard clusters, there are also many fuzzy concepts that describe properties of objects that cannot be answered simply by “yes” or “no”. A *fuzzy cluster* is considered to be a set of objects that have the properties described by a fuzzy concept. Following fuzzy set theory (Zadeh, 1996), *fuzzy cluster* is also called *fuzzy set*, and we give the definition of fuzzy cluster as follows.

**Definition 3.7** *Given a FoD  $\Omega$  and a cluster  $\{\omega_j\} \in \Omega$ , then the cluster  $\{\omega_j\}$  with  $\mathbf{u}_{\cdot j}^{(f)} : \Omega \rightarrow [0, 1]$  is called a fuzzy cluster/set on  $\Omega$ . The mapping  $\mathbf{u}_{\cdot j}^{(f)}$  is called fuzzy membership function of the fuzzy cluster/set  $\{\omega_j\}$ . Given a dataset  $\mathcal{X}$ , for any  $\mathbf{x}_i \in \mathcal{X}$ , we have:*

$$u_{ij}^{(f)}, u_{ij'}^{(f)} \in [0, 1] \quad (3.7)$$

where  $u_{ij}^{(f)}$  is used to represent the fuzzy membership degree of the object  $\mathbf{x}_i$  belonging to the  $j$ -th cluster.

Unlike hard one, fuzzy cluster has imprecise information. If  $u_{ij}^{(f)} < 1$ , it indicates that although the object  $\mathbf{x}_i$  is assigned to the cluster  $\{\omega_j\}$ , it may belong to other clusters at the same time. Thus, fuzzy clusters can characterize the imprecision of results in clustering task. Of course, for any  $\mathbf{x}_i \in \{\omega_j\}$ , if  $u_{ij}^{(f)} = 1$ , then  $\{\omega_j\}$  degenerates to a hard cluster. In this case, the cluster  $\{\omega_j\}$  does not contain any imprecise information.

### 3.3.3 Possibilistic clusters

In clustering, it is known that an event must happen (the object belongs to a specific cluster), but due to insufficient information, we cannot completely certain this event. In this case, it is a good choice to use *possibilistic clusters* to characterize the uncertainty of this event. Following possibility theory (Dubois and Prade, 2012, 1988), *possibilistic clusters* can represent the uncertainty of clustering results.

**Definition 3.8** *Given a FoD  $\Omega$  and a cluster  $\{\omega_j\}$ , then the cluster  $\{\omega_j\}$  with  $\mathbf{u}_{\cdot j}^{(p)} : \Omega \rightarrow [0, 1]$ , is a possibilistic cluster on  $\Omega$ . The mapping  $\mathbf{u}_{\cdot j}^{(p)}$  is possibilistic membership function of the cluster  $\{\omega_j\}$ . Given a dataset  $\mathcal{X}$ , for any  $\mathbf{x}_i \in \mathcal{X}$ , we have:*

$$u_{ij}^{(p)}, u_{ij}^{(p)} \in [0, 1] \quad (3.8)$$

where  $u_{ij}^{(p)}$  is used to represent the possibilistic membership degree of the object  $\mathbf{x}_i$  belonging to the  $j$ -th cluster, i.e. it characterizes the possibility of event  $\mathbf{x}_i \in \{\omega_j\}$ .

Note that once an object  $\mathbf{x}_i$  is assigned to cluster  $\{\omega_j\}$ , it indicates that for  $\mathbf{x}_i$ , the event  $\mathbf{x}_i \in \{\omega_j\}$  is most likely to occur. Here different events are mutually exclusive, i.e. the events  $\mathbf{x}_i \in \{\omega_j\}$  and  $\mathbf{x}_i \in \{\omega_{j'}\}$ ,  $j \neq j'$ , cannot occur at the same time, which is different from fuzzy clusters. Specifically,  $u_{ij}^{(p)}$  characterizes the uncertainty of event  $\mathbf{x}_i \in \{\omega_j\}$  and  $u_{ij}^{(f)}$  characterizes the imprecision of event  $\mathbf{x}_i \in \{\omega_j\}$ . Thus, possibilistic clusters can characterize the uncertainty of results in clustering task. Of course, for any  $\mathbf{x}_i \in \{\omega_j\}$ , if  $u_{ij}^{(p)} = 1$ , then we have a hard cluster  $\{\omega_j\}$ . In this case, the cluster  $\{\omega_j\}$  does not contain any uncertain information.

### 3.3.4 Evidential clusters

Given a FoD  $\Omega$ , unlike other ones, the TBF extends it to the power-set framework  $2^\Omega$ . For example, if  $c = 3$ , we have  $2^\Omega = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_1, \omega_2\}, \{\omega_1, \omega_3\}, \{\omega_2, \omega_3\}, \Omega\}$ . In other words, the object can be assigned to three clusters: singleton (specific) cluster (e.g.  $\{\omega_1\}$ ), meta-cluster (e.g.  $\{\omega_1, \omega_2\}$ ) and the noise cluster represented by  $\emptyset$ . These clusters are collectively referred to as evidential cluster here. In this theory, the object  $\mathbf{x}_i$  belonging to different clusters are considered independent events, and these events cannot occur simultaneously. That is, in the decision, each object must be assigned to one of these clusters. The definition of these clusters are given as follows.

**Definition 3.9** *Singleton cluster, also called specific one, similar to possibilistic cluster, aims to characterize the uncertainty of the results. Given an object  $\mathbf{x}_i$ , once it is assigned to the singleton cluster  $A_j$ , i.e.  $\mathbf{x}_i \in A_j$ , we have:*

$$m_i(A_j) > m_i(A_{j'}), \forall j', A_j \neq A_{j'} \subseteq \Omega, |A_j| = 1. \quad (3.9)$$

where the mass of belief  $m_i(A_j)$  denotes the degree of the object  $\mathbf{x}_i$  associated with the focal element (singleton cluster)  $A_j$ .

As analyzed, the mass of belief  $m_i(A_j)$  can be fuzzy, possibility, or probability values in special cases (Masson and Dencœux, 2008).

**Definition 3.10** *Meta-cluster is used to characterize imprecision in results. Given a meta-cluster  $A_j$ , it is defined by the disjunction (union) of several singleton clusters and has the following form:*

$$A_j = \{\omega_k, \dots, \omega_l\}, k \neq l \in [1, c]. \quad (3.10)$$

Meta-cluster is regarded as a real and new cluster with the same properties as a singleton cluster and can be considered as a transition cluster among these different close singleton clusters. If an object is assigned to a meta-cluster, this indicates that the object may belong to any of the singleton clusters in the meta-cluster (Liu et al., 2015; Zhang et al., 2021d). For example, if  $\mathbf{x}_i \in A_j = \{\omega_1, \omega_2\}$ , we consider that  $\mathbf{x}_i$  may belong to both cluster  $\{\omega_1\}$  and cluster  $\{\omega_2\}$ . In this case, event  $\mathbf{x}_i \in \{\omega_1\}$  and event  $\mathbf{x}_i \in \{\omega_2\}$  can occur simultaneously. In this case, after a transformation, *e.g.* Eq. (2.8), the mass values obtained, *i.e.*  $m_{i1}$ ,  $m_{i2}$ , can be viewed as the fuzzy membership degrees of  $\mathbf{x}_i$  belonging to  $\{\omega_1\}$  and  $\{\omega_2\}$ , respectively. Hence, meta-cluster is used to characterize imprecise information in results.

In spatial distribution, if  $\mathbf{x}_i$  is assigned to meta-cluster, the object  $\mathbf{x}_i$  may simultaneously close to several singleton clusters included in the meta-cluster. In this case, we can say that the object  $\mathbf{x}_i$  is *imprecise* or we have an *imprecise object*  $\mathbf{x}_i$ . By contrast, for the object  $\mathbf{x}_i$ , once it is assigned to a singleton cluster, it indicates that the object  $\mathbf{x}_i$  is obviously close to one and only one singleton cluster. In this case, we thereby say that this object  $\mathbf{x}_i$  is *precise*. In other words, we can say that  $\mathbf{x}_i$  is a *precise object* if we are sure that  $\mathbf{x}_i$  belongs to one singleton cluster only. Thus, we have the following definition of *precise/imprecise* object.

**Definition 3.11** *Given an object  $\mathbf{x}_i$ , we have a precise object  $\mathbf{x}_i$  if it is assigned to the singleton cluster  $A_j$ , such that:*

$$\mathbf{x}_i \in A_j, |A_j| = 1, A_j \neq \emptyset. \quad (3.11)$$

**Definition 3.12** Given an object  $\mathbf{x}_i$ , we have an imprecise object  $\mathbf{x}_i$  if it is assigned to the meta-cluster  $A_j$ , such that:

$$\mathbf{x}_i \in A_j, |A_j| \geq 2. \quad (3.12)$$

By contrast, some objects may be far from all singleton clusters at the same time. In this case, they are considered as a set of *noise*. Thus, we have the following definition.

**Definition 3.13** Noise cluster, represented by  $\emptyset$ , is considered as a separate cluster and is defined as the set of those objects that are far from all singleton clusters. Given an object  $\mathbf{x}_i$ , if it is assigned to noise cluster, we have:

$$m_i(\emptyset) > m_i(A_j), A_j \neq \emptyset, A_j \subseteq \Omega. \quad (3.13)$$

Once object  $\mathbf{x}_i$  is assigned to noise cluster, it is regarded as a *noise* or *outlier*.

Each of these clusters has a unique and important role in the clustering task, especially the meta-cluster and noise cluster allows us to gain a deeper insight into the data and to improve robustness with respect to outliers (Masson and Dencœux, 2008).

### 3.3.5 The difference and relations of different clusters

Inspired from (Dencœux and Kanjanatarakul, 2016), we have the relationship between different clusters, as shown in Fig. 3.2.

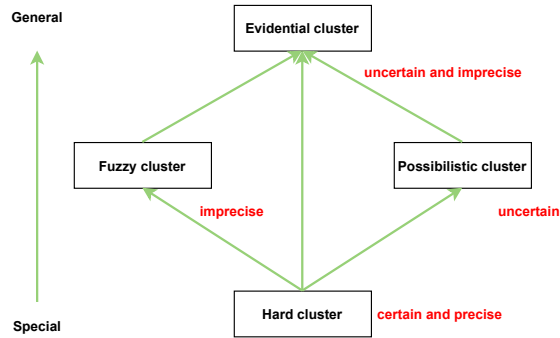


Fig. 3.2. Relationship between hard, fuzzy, possibilistic, and evidential clusters.

**Example 3.3:** Let us consider that a dataset  $\mathcal{X}$  is a collection of  $n$  people, defined as  $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n\}$ , where the basic information of  $\mathbf{x}_i, i = 1, 2, 3$  has been shown in Table 3.2. A partition of the given dataset  $\mathcal{X}$  is a set of  $c$  clusters defined on the FoD  $\Omega$ . Specifically, given a set of clusters  $\Omega = \{\omega_1, \omega_2, \omega_3\}$  with  $c = 3$ , let us consider that all objects in  $\mathcal{X}$  are grouped into these 3 clusters and we have the following cases.

- **Case 1:** In the results, if we are sure that all the objects are assigned to these clusters with no uncertainty and no imprecision, we consider these clusters, such

as  $\{\omega_1\}$ , to be **hard clusters**. Following probability theory, for example, if  $\mathbf{x}_1$  is assigned to cluster  $\{\omega_1\}$ , we have  $p_2(\{\omega_1\}) = 1$ . Thus, hard clusters can characterize the certainty and precision of the results.

- **Case 2:** In the results, if there is some fuzzy or imprecise knowledge in these clusters, following fuzzy set theory, they can be considered as **fuzzy sets** or **fuzzy clusters**. For example, for the object  $\mathbf{x}_2$  in cluster  $\{\omega_1\}$ , we thereby have  $u_2^{(f)}(\{\omega_1\}) > u_2^{(f)}(\{\omega_i\})$  ( $i = 2, 3$ ) with  $u_2^{(f)}(\{\omega_1\}) \in (0, 1)$ . In this case, we can say that the object  $\mathbf{x}_2$  in cluster  $\{\omega_1\}$  with the biggest fuzzy degree, and if  $u_2^{(f)}(\{\omega_2\}) > 0$ , we can also say that it may be in cluster  $\{\omega_2\}$ , only with a smaller fuzzy degree. Thus, fuzzy clusters can characterize the imprecision of the results.
- **Case 3:** In the results, if there is some uncertain knowledge in these clusters, following possibility theory, they can be considered as **possibilistic clusters**. For example, for the object  $\mathbf{x}_3$  in cluster  $\{\omega_1\}$ , we thus have  $u_3^{(p)}(\{\omega_1\}) > u_3^{(p)}(\{\omega_i\})$  with  $u_3^{(p)}(\{\omega_1\}) \in (0, 1)$ ,  $i = 2, 3$ . In this case, we can only say that the object  $\mathbf{x}_3$  belongs to cluster  $\{\omega_1\}$ , but there is uncertainty in this proposition due to limited knowledge. Thus, possibilistic clusters can characterize the uncertainty of the results.
- **Case 4:** In the results, following the TBF, all the objects are assigned to different clusters, including the noise cluster  $\emptyset$ , the singleton clusters  $\{\omega_i\}$ ,  $i = 1, 2, 3$ , and the meta-clusters  $\{\omega_1, \omega_2\}$ ,  $\{\omega_1, \omega_3\}$ ,  $\{\omega_2, \omega_3\}$ ,  $\Omega$ . These clusters are called **evidential clusters**. For example, for the object  $\mathbf{x}_1$  in the singleton cluster  $\{\omega_1\}$ , we have  $m_1(\{\omega_1\}) > m_1(A_{j'})$  with  $m_1(\{\omega_1\}) \in (0, 1)$ , where  $A_{j'}$  represents all the evidential clusters except  $\{\omega_1\}$  under the power-set  $2^\Omega$ . In this case, we can say that the object  $\mathbf{x}_1$  belongs to cluster  $\{\omega_1\}$ , but there is uncertainty in this proposition due to limited knowledge. Again, for the object  $\mathbf{x}_2$  in the meta-cluster  $\{\omega_2, \omega_3\}$ , we can say  $\mathbf{x}_2$  belongs to cluster  $\{\omega_2\}$  or cluster  $\{\omega_3\}$ . In this case, the meta-cluster  $\{\omega_2, \omega_3\}$  is used to describe partial ignorance of information (knowledge), *i.e.* there is imprecise or fuzzy information (knowledge) in this proposition. Thus, evidential clusters can characterize the uncertainty and imprecision in the results.

### 3.4 Concepts and definitions of methods

In clustering, empirical information, *i.e.* data or dataset, is usually grouped into several clusters by the ways that we call *methods*. Given the dataset  $\mathcal{X}$ , the methods will group  $n$  objects  $\mathbf{x}_i$  ( $\mathbf{x}_i \in \mathcal{X}$ ,  $i = 1, \dots, n$ ) into different clusters in a hard/soft way, where the members (objects) are similar in each cluster. In this section, we refer to those methods that group objects into different clusters in a certain way as *hard methods*,

*e.g.*  $K$ -means (Jain, 2010). By contrast, those methods that group objects into clusters in a soft (uncertain/imprecise) way are called *fuzzy, possibilistic, or evidential methods*, *e.g.* Fuzzy  $c$ -means (Bezdek, 2013), Possibilistic  $c$ -means (Krishnapuram and Keller, 1996), Evidential  $c$ -means (Masson and Dencœux, 2008). We will give clear definitions and review some classical fuzzy/possibilistic methods in this section.

### 3.4.1 Hard methods

In clustering, *hard method* usually submit the object to a singleton (specific) cluster in a completely certain way under the established rules, models, etc. That is, hard method give hard clusters. In this case, each object only belongs to one singleton cluster, which is regarded as a certain event. If allowed, from the view of probability theory, we give the following definition.

**Definition 3.14** *Given a dataset  $\mathcal{X}$  and a FoD  $\Omega$ , if we have a hard method, each object  $\mathbf{x}_i$  should be submitted to a singleton cluster  $\{\omega_j\}$ , such that:*

$$p_{ij} = 1 \tag{3.14}$$

where  $p_{ij}$  represents the probabilistic degree of  $i$ -th object belonging to the  $j$ -th cluster.

In applications, a number of hard methods have been proposed. In hierarchy-based clustering (Murtagh, 1983; Johnson, 1967), it attempts to construct a tree-like nested structure partition where neighbors are adapt and gradually merge into a cluster as needed. In partition-based  $K$ -means/ $K$ -medoids (Jain, 2010), the objects are always assigned to the nearest center. An objective function, typically the sum of the distance to a set of putative cluster centers, is optimized (Kaufman and Rousseeuw, 2009; Frey and Dueck, 2007; Höppner et al., 1999) until the best cluster centers candidates are found. In distribution-based EM methods (McLachlan and Krishnan, 2007), one attempts to reproduce the observed realization of objects as a mix of predefined probability distribution functions. In density-based DPC methods (Rodriguez and Laio, 2014), inspired by DBSCAN (Ester et al., 1996) and Mean-shift (mode seeking) (Cheng, 1995), one characterizes the cluster centers by a higher density than their neighbors and by a relatively large distance from objects with higher densities. Then the non-central objects gradually converge to different clusters based on the known information of neighbors. There also exist some methods based on machine learning (Ben Hur et al., 2001) and neural networks (Vesanto and Alhoniemi, 2000). These methods have been applied with good results.

### 3.4.2 Fuzzy methods

Unlike hard methods, following the fuzzy set theory, *fuzzy methods* believe that an object  $\mathbf{x}_i$  can belong to all singleton clusters with different fuzzy membership degrees.



The definition is shown as follows.

**Definition 3.15** *Given a dataset  $\mathcal{X}$  and FoD  $\Omega$ , for each object  $\mathbf{x}_i$ , we have a fuzzy method, such that:*

$$\sum_{j=1}^c u_{ij}^{(f)} = 1, \forall i, \sum_{i=1}^n u_{ij}^{(f)} < n, \forall j, u_{ij}^{(f)} \in [0, 1] \quad (3.15)$$

where  $u_{ij}^{(f)}$  represents the fuzzy membership degree of the  $i$ -th object in the  $j$ -th cluster.

*Fuzzy methods* have a complete theoretical and mathematical framework, and a number of fuzzy methods have been proposed and applied in the past decades.

In these methods, the most popular is Bezdek’s Fuzzy  $c$ -means (FCM) algorithm (Bezdek, 2013; Bezdek et al., 1984) for object data. It can be considered as the fuzzy version of  $K$ -means because they are similar regarding the definition of the objective function and the optimization mechanism. The difference is that FCM also needs to optimize the fuzzy matrix  $U = (u_{ij}^{(f)})$  in Eq. (3.15). Fuzzy methods have been widely used in many fields. For example, for clustering images, spatial information, considered as the summation of the membership function in the neighborhood of each pixel, is incorporated into the membership function (Chuang et al., 2006).

### 3.4.3 Possibilistic methods

*Possibilistic methods* are considered as the possibilistic version of FCM under possibility theory. It interprets and normalizes fuzzy methods from possibility distribution.

**Definition 3.16** *Given a dataset  $\mathcal{X}$  and FoD  $\Omega$ , for each object  $\mathbf{x}_i$ , we have a possibilistic method, such that:*

$$\sum_{j=1}^c u_{ij}^{(p)} \leq 1, \forall i, \sum_{i=1}^n u_{ij}^{(p)} < n, \forall j, u_{ij}^{(p)} \in [0, 1] \quad (3.16)$$

where  $u_{ij}^{(p)}$  represents the possibilistic membership degree of  $i$ -th object in  $j$ -th cluster.

In Eq. (3.16), the normalization constraint  $\sum_{i=1}^c u_{ij}^{(p)} = 1$  is relaxed to against noise and outliers, and  $1 - \sum_{i=1}^c u_{ij}^{(p)}$  can be regarded as the possibilistic degree of  $i$ -th object belonging to the noise and outliers.

In the early version, the possibilistic clustering (PCM) (Krishnapuram and Keller, 1993, 1996) works by modifying the objective function to be minimized. In PCM, the membership  $u_{ij}^{(p)}$  is interpreted as a *typicality* degree or a *possibilistic* degree (under possibility theory) that the object  $\mathbf{x}_i$  belongs to the cluster  $\{\omega_j\}$ . Experiments have

shown that PCM can improve robustness with respect to noise or outliers. In (Zhang and Leung, 2004), an improved possibilistic  $c$ -means (IPCM) is proposed to address the issue that PCM tend to find identical clusters. The key is to integrate the fuzzy partition into PCM, so that the IPCM can determine proper clusters. The idea of combining the *fuzzy partition* and the *possibilistic partition* has been verified in improving the robustness of the methods.

### 3.4.4 Evidential methods

The seminal evidential methods, following the TBF, has been introduced in (Dencœux and Masson, 2004; Masson and Dencœux, 2008). It extends the existing concepts of hard, fuzzy, and possibilistic partitions by allocating, for each object, a “mass of belief”, not only to singleton clusters, but also to any subsets of the given FoD  $\Omega = \{\omega_1, \dots, \omega_c\}$ . Roughly speaking, a mass function can be seen as a collection of sets with corresponding masses. A collection of such mass functions for  $n$  objects is called *credal partition* (Dencœux and Masson, 2004, 2003; Su and Dencœux, 2018). Experiments have shown that this additional flexibility allows to gain a deeper insight in the data and to improve robustness with respect to outliers (Masson and Dencœux, 2008). Evidential methods have been reviewed in Section 3.5, and here we give some basic concepts and definitions.

**Definition 3.17** *Let consider a dataset  $\mathcal{X}$  with  $n$  objects and a power-set  $2^\Omega$  on FoD  $\Omega$ , given a set of  $n$  mass functions, we have an evidential method, for each  $m$ , such that:*

$$\sum_{A_j \subseteq \Omega} m_i(A_j) = 1, \quad \forall i, m_i(A_j) \in [0, 1] \quad (3.17)$$

where  $m_i(A_j)$ , a mass of belief, denotes the degree of the object  $\mathbf{x}_i$  associated with the focal element (cluster)  $A_j$ .

Since credal partition is considered as an extended version of hard, fuzzy, and possibilistic partitions of the given dataset  $\mathcal{X}$  under the TBF, it can be degraded to either hard, fuzzy, or possibilistic partitions in specific cases.

**Definition 3.18** *Given a set of  $n$  mass functions, for the dataset  $\mathcal{X}$ , we have a set of hard mass functions, such that:*

$$m_i(A_j) = 1, \quad \forall m, A_j \neq \emptyset, |A_j| = 1. \quad (3.18)$$

In this case, each object  $\mathbf{x}_i$  is completely associated with the focal element (cluster)  $A_j$ ,  $|A_j| = 1$ , *i.e.*  $A_j = \{\omega_j\}$ .

**Definition 3.19** *Given a set of  $n$  mass functions, for the dataset  $\mathcal{X}$ , we have a set of fuzzy membership functions, such that:*

$$m_i(A_j) = 0, \quad \forall m, A_j \neq \emptyset, |A_j| \geq 2. \quad (3.19)$$

In this case, for each object  $\mathbf{x}_i$ , it is allowed to belong to different singleton clusters simultaneously and the sum of mass of belief that the object  $\mathbf{x}_i$  belongs to these singleton clusters is equal to 1.

**Definition 3.20** *Given a set of  $n$  mass functions, for the dataset  $\mathcal{X}$ , we have a set of possibilistic membership functions, such that:*

$$\left\{ \begin{array}{l} \sum_{j=1}^c m_i(A_j) \leq 1, |A_j| = 1, \\ m_i(A_j) = 0, |A_j| \neq 1, \\ m_i(A_j) \in [0, 1]. \end{array} \right. \quad (3.20)$$

In this case, the object  $\mathbf{x}_i$  belongs to one of these specific clusters, and if  $m_i(A_j) < 1$ , it represents the possibilistic degree of the  $i$ -th object in the  $j$ -th cluster. The relationship between credal partition and other partition structures has been discussed (Denceux and Kanjanatarakul, 2016).

### 3.4.5 The overview of different data, methods, and clusters

In this subsection, we will give a diagram of hard, fuzzy, possibilistic, and evidential clustering from data, methods and clusters perspectives, as shown in Fig. 3.3. The solid path in Fig. 3.3 indicates that it is feasible and relevant literature is available, while the others are not found.

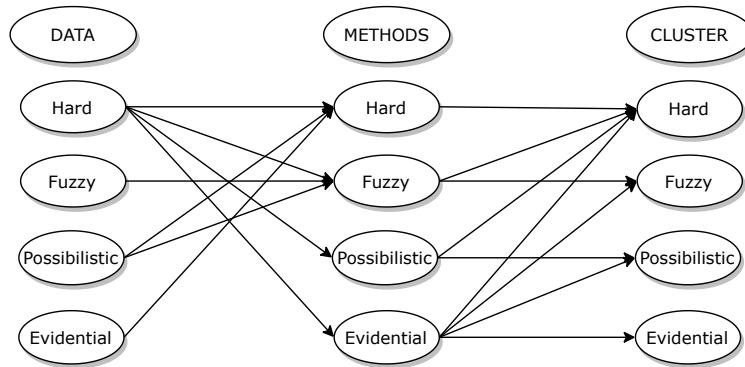


Fig. 3.3. Illustration of data, methods, and clusters of existing methods.

## 3.5 Clustering techniques based on TBF

Evidential clustering is still under development and existing methods are mainly for object data and relational data. In this section, therefore, the existing methods are

reviewed from these two aspects in the Subsections 3.5.1 and 3.5.2, respectively. Afterward, some practical applications of evidential clustering are discussed in Subsection 3.5.3. Then, the advantages and disadvantages of different methods are analyzed and the applicable scenarios are given in Subsection 3.5.4. Finally, the challenges faced by evidential clustering are discussed.

### 3.5.1 Object data-based methods

Evidential clustering based on the concept of *credal partition* (Dencœux and Masson, 2004, 2003; Dencœux and Kanjanatarakul, 2016; Masson and Dencœux, 2004) is proposed by Dencœux and Masson. Afterward, they present an early version for *object data* in 2008, named evidential *c*-means (ECM) (Masson and Dencœux, 2008), which can be considered as the evidential version of fuzzy *c*-means (FCM) (Bezdek, 2013) and noise clustering (NC) (Dave, 1991; Sen and Davé, 1998) under the TBF. In ECM, it allows the object to be in any singleton clusters and meta-clusters with different masses of beliefs, and the noise cluster is represented by  $\emptyset$ . Inspired by FCM and NC, the objective function of ECM is defined by:

$$J_{ECM} = \sum_{i=1}^n \sum_{A_j \neq \emptyset, A_j \subseteq \Omega} |A_j|^a m_{ij} d_{ij} + \sum_{i=1}^n \delta^2 m_{i\emptyset}^\beta, \quad (3.21)$$

subject to

$$\sum_{A_j \neq \emptyset, A_j \subseteq \Omega} m_{ij} + m_{i\emptyset} = 1 \quad (3.22)$$

where  $d_{ij}$  represents the Euclidean distance between the  $i$ -th object and the center of  $j$ -th cluster. The center ( $\bar{\mathbf{v}}_j$ ) of meta-cluster ( $A_j$ ) is the average of the involved singleton cluster centers, defined by:

$$\bar{\mathbf{v}}_j = \frac{1}{|A_j|} \sum_{k=1}^c s_{kj} \mathbf{v}_k, \quad s_{kj} = \begin{cases} 1, & \text{if } \{\omega_k\} \in A_j \\ 0, & \text{otherwise} \end{cases} \quad (3.23)$$

where  $|A_j|$  is the cardinality of  $A_j$  and  $\mathbf{v}_k$  is the center of the singleton cluster  $\{\omega_k\}$ .

Similar to FCM and NC, ECM also assigns objects to the nearest centers by minimizing an objective function. Since the mass of belief only depends on the distance between the object and the center of this cluster, it may produce unreasonable results when meta-cluster centers are close to the centers of singleton clusters. To overcome this limitation, new methods, belief *c*-means (BCM) (Liu et al., 2012) and credal *c*-means (CCM) (Liu et al., 2015) are introduced by redefining the distance between the object and the centers. For example, in CCM, the distance between the object and meta-cluster is not only related to the center of this meta-cluster but also related to the associated singleton clusters included in this meta-cluster. By doing this, it can prevent some objects from being unreasonably assigned to meta-clusters.

In applications, prior knowledge about cluster membership is sometimes available. To integrate such auxiliary information, an interesting constrained evidential  $c$ -means algorithm (CECM) (Antoine et al., 2010, 2012) is proposed. In CECM, the pairwise constraints are translated into the framework of TBF and integrated into the cost function. Besides, based on a small set of labeled data (or seeds), an evidential seed-based semi-supervised clustering (SECM) is proposed in (Antoine et al., 2014a), which is considered as the first semi-supervised extension of evidential clustering. Interestingly, Dencœux presents a calibrated model-based evidential clustering method recently (Dencœux, 2020a). It first computes the bootstrap percentile confidence intervals for all pairwise probabilities, *i.e.* the probabilities for any two objects to belong to the same class. Then, the pairwise belief and plausibility degrees are made to approximate the bounds of the confidence intervals by calibrating the evidential partition. The practical applicability of the method is verified in the real world. In (Zhang et al., 2021c), evidential clustering is first applied to missing data, suggesting an advantage in characterizing the uncertainty and imprecision due to missing values.

In addition to the above methods, there are some approaches that combine with other techniques to better extend the applications of evidential clustering on object data. For example, inspired by other methods (Fred and Lourenço, 2008; Hadjitodorov et al., 2006), an evidential version of ensemble clustering is proposed in (Masson and Dencœux, 2011). In this method, belief functions, defined on the lattice of intervals partitions of a set of objects, are investigated as a suitable framework for combining multiple clusterings. Based on classical  $K$ -modes (Huang, 1998; Huang and Ng, 1999) and DBSCAN (Ester et al., 1996), a series of evidential methods (Hariz et al., 2006; Hariz and Elouedi, 2010a; Hariz et al., 2007; Hariz and Elouedi, 2011, 2010b; Bessrouer et al., 2020) in uncertain contexts are developed, aiming to handle uncertainty in the attribute values of objects, such as belief  $k$ -modes method (BKM) (Hariz et al., 2006), and to handle the belongs of the objects to different clusters with various mass degrees, such as evidential DBSCAN (E-DBSCAN) (Bessrouer et al., 2020). In (Dencœux, 2011), a variant of the EM algorithm (McLachlan and Krishnan, 2007; Dempster et al., 1977), called evidential EM ( $E^2M$ ), is presented under the TBF and the maximization of a generalized likelihood criterion. This criterion can be interpreted as a degree of agreement between the statistical model and the uncertain observations and the proposed method can iteratively maximize this criterion. In particular, in a new method, named belief-peaks evidential clustering (BPEC) (Su and Dencœux, 2018; Gong et al., 2021), cluster centers can be adaptively obtained and credal partition is achieved by combining evidential clustering with density peak algorithm (Rodriguez and Laio, 2014). To overcome the shortcomings of BPEC in manual center selection and inability to detect arbitrary clusters, a new evidential clustering algorithm, named CBP-EKNN (Gong et al., 2020), is proposed by finding the “cumulative belief peaks” and evidential  $K$ -nearest neighbor rule. In (Zhou et al., 2019), an evolutionary version of evidential

clustering for multi-objective, called MOE2CM, is proposed with an underlying evolution strategy called NSGA-II (Deb et al., 2002). It aims to present a number of non-dominated solutions for the decision-maker based on the problem requirements. Interestingly, they further propose an evidential evolving  $c$ -means (E2CM) (Su et al., 2020) clustering method in the framework of evolutionary computation. It aims to actively obtaining the cluster centers by encoding the centers in a population of variable strings (or particles) while searching for the optimal number and location. Recently, an evidential version of the Gaussian mixture model (EGMM) for clustering object data has also been recently announced (Jiao et al., 2020), which also allows to automatic determination of the proper number of clusters. In (Zhou et al., 2021), a prototype-based evidential transfer clustering algorithm, named transfer evidential  $c$ -means (TECM), is introduced to handle the insufficiency and uncertainty problems in the clustering task simultaneously. The proposed algorithm employs the cluster prototypes of the source data as references to guide the clustering process of the target data. These methods can model uncertainty and imprecision in the clustering process and achieve good results from different perspectives.

### 3.5.2 Relational data-based methods

The TBF is also first extended to clustering analysis for *relational data* (*proximity data*) by Dencœux and Masson and introduced the concept of *credal partition* in detail (Dencœux and Masson, 2004, 2003). Inspired by some fuzzy methods (Sen and Davé, 1998; Dave, 1991), they call this relational data-oriented method as Evidential CLUSTERING (EVCLUS) (Dencœux and Masson, 2004). It does not require the data to be characterized as a vector, but in order not to lose generality, we take the example of object data here. The input is a  $n \times n$  dissimilarity matrix  $\Psi = (\psi_{ij})$ , where  $\psi_{ij} \geq 0$  measures the degree of dissimilarity between objects  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . The objective function of EVCLUS is defined by:

$$J_{EVCLUS} = \frac{1}{cons} \sum_{i < j} \frac{(a\mathcal{K}_{ij} + b - \psi_{ij})^2}{\psi_{ij}} \quad (3.24)$$

where  $a$  and  $b$  are two coefficients,  $cons$  is a normalizing constant.  $\mathcal{K}_{ij}$  is the degree of conflict between  $m_i$  and  $m_j$ , defined by:

$$\mathcal{K}_{ij} = \sum_{A \cap B = \emptyset} m_i(A)m_j(B). \quad (3.25)$$

The above objective function  $J_{EVCLUS}$  can be implemented by optimizing the credal partition matrix  $M = (m_{ij})$  and the parameters  $a, b$  (Dencœux and Masson, 2004, 2003). To control the complexity of the model, they suggest adding a penalty term to  $J_{EVCLUS}$  for supporting simple and more informative evidence. The informativeness of the mass

function  $m_i$  can be defined in terms of entropy by:

$$E(m_i) = \sum_{A \neq \emptyset} \log_2\left(\frac{|A|}{m_i(A)}\right) + m_i(\emptyset) \log_2\left(\frac{|\Omega|}{m_i(\emptyset)}\right). \quad (3.26)$$

The value of  $E(m_i)$  is smaller when more confidence is assigned to the focal element with smaller potential. The final objective function to be optimized is written by:

$$J_{EVCLUS1} = J_{EVCLUS} + \lambda \sum_{i=1}^n E(m_i) \quad (3.27)$$

where the weight factor  $\lambda$  is used to balance the two terms.

Afterward, EVCLUS is extended to a semi-supervised learning framework, called CEVCLUS (Antoine et al., 2011, 2014b). This method aims to improve the clustering accuracy by considering the must-link and cannot-link restrictions between some objects in the optimization objective function. Furthermore, two much faster and efficient versions of CEVCLUS, such as called  $k$ -CEVCLUS (Li et al., 2018) are introduced to make it applicable to large datasets (Li et al., 2018; Denceux and Kanjanatarakul, 2016). Interestingly, a new clustering method for relational data, named EK-NNclus (Denceux et al., 2015), is proposed recently. It initially assumes that all objects individually form a cluster, and then gradually merges the clusters through the EK-NN rule (Denceux, 2008b) and the idea of label propagation, eventually forming a stable cluster structure. Besides, a new neural-network based evidential clustering method is presented, called NN-EVCLUS (Denceux, 2020b), where the neural network can be paired with a one-class support vector machine to make it robust to outliers and allow for novelty detection. Unlike ECM (Masson and Denceux, 2008), most of these above methods only includes singleton clusters and the total ignorance  $\Omega$ . Fortunately, the relational version of ECM, called RECM (Masson and Denceux, 2009), has been derived. In particular, in some cases, it may not adequately model different types of group structure by using one center to represent a class, and thereby a new prototype-based clustering method with multiple prototypes, named evidential  $c$ -medoids (ECMdd), is proposed for relational data (Zhou et al., 2015b, 2016a). In ECMdd, a single or multiple weighted medoids are used to flexibly represent one cluster. These aforementioned methods have made important contributions in clustering relational data.

### 3.5.3 Applications of evidential methods

In fact, evidential clustering has been applied to many fields (Mahamadou et al., 2019; Saidi et al., 2018b; Lian et al., 2017b; Ayed et al., 2017; Serir et al., 2012; Zhou et al., 2015a, 2018b, 2014, 2016b, 2018a; Saidi et al., 2018a; Abdelkhalek et al., 2019a,b, 2018). For example, to model, fuse, and reason with uncertain and imprecise knowledge from noisy and blurry PET-CT images, a co-clustering algorithm is proposed to concur-

rently segment 3D tumors in PET-CT images (Lian et al., 2017b,a). In this method, it is considered that the two PET-CT and mono-modal imaging modalities can provide complementary information to improve segmentation performance by combining functional and anatomical information. The validity has been demonstrated on PET-CT images for non-small cell lung cancer patients. In (Serir et al., 2012), a new online clustering method called evidential evolving Gustafson-Kessel (E2GK) is introduced. It aims to achieve a credal partition of data streams online by applying an algorithm resulting from the adaptation of the evolving Gustafson-Kessel (EGK) algorithm (Filev et al., 2010). In particular, a new prototype-based clustering method, called median evidential  $c$ -means (MECM) (Zhou et al., 2015c), is presented, which can be considered as an extension of median  $c$ -means (Cottrell et al., 2006) and median fuzzy  $c$ -means (Geweniger et al., 2010) under the framework of TBF. Since it relaxes the restrictions on the object embedding metric space and restricts the prototypes to be in the original dataset, the proposed MECM method based on credal partition has a better understanding of the graph structures in detecting community for social networks. Afterward, a few domain knowledge is incorporated into the community detection and thereby a semi-supervised clustering based on an evidential label propagation strategy is proposed (Zhou et al., 2016b, 2018a). Also, evidential collaborative filtering is proposed in recommender systems for helping and guiding users towards items of interest (Abdelkhalek et al., 2019a,b, 2018). The successful practice of these methods enhances the prospects for the applications of evidential clustering in different scenarios.

### 3.5.4 Analysis of different evidential methods

To further advance the applications of evidential clustering in various fields, we summarize the advantages and disadvantages of some representative methods from different perspectives, such as data type, center, complexity, as shown in Table 3.6. It should be noted that some methods provide schemes to optimize the number ( $c$ ) of cluster centers, but we generally consider it necessary to provide  $c$  in our applications, *e.g.* ECM (Masson and Dencœux, 2008), MECM (Zhou et al., 2015c). In this case, we group these methods as those that require the number ( $c$ ) of cluster centers. By the way, in order to easily distinguish the advantages and disadvantages of the different methods, we evaluate the practicality of these methods in 5 main aspects, which are included in the notes to comment. It is easy to find from Table 3.6 that different methods have their own scope of applications, so users can choose a suitable method as a benchmark according to different scenarios. To make it as easy as possible for the user in applications, we give the basic principles for choosing some classical methods as potential solutions, as shown in Fig. 3.4.



Table 3.6: Comparisons of different evidential methods

Method	Data type	Center	Measure type	Cluster number	Complexity	Disadvantage
ECM	Object	Centroids	Euclidean	Known	$\mathcal{O}(n2^c)$	(1), (3), (4)
CCM	Object	Centroids	Euclidean	Known	$\mathcal{O}(n2^c)$	(1), (3), (4)
CECM	Object	Centroids	Adaptive metric	Known	$\mathcal{O}(n2^c)$	(1), (2), (3)
BPEC	Object	Medoids	Euclidean	Unknown	$\mathcal{O}(n^2 + n2^c)$	(3), (4)
CBP-EKNN	Object	Medoids	Euclidean	Unknown	$\mathcal{O}(n^3)$	(3), (5)
EGMM	Object	Centroids	Covariance matrix	Known	$\mathcal{O}(n2^c)$	(1), (3)
EVCLUS	Relational	/	Dissimilarity matrix	Known	$\mathcal{O}(n^2c^2)$	(1), (3)
CEVCLUS	Relational	/	Dissimilarity matrix	Known	$\mathcal{O}(n^22^c)$	(1), (2), (3), (5)
k-CEVCLUS	Relational	/	Dissimilarity matrix	Known	$\mathcal{O}(n2^c)$	(1), (2), (5)
RECM	Relational	Centroids	Euclidean	Known	$\mathcal{O}(nc^2 + cn^2)$	(1), (3), (4)
MECM	Relational	Medoids	Euclidean	Known	$\mathcal{O}(n2^c)$	(1), (3), (4)
EK-NNclus	Relational	/	Dissimilarity matrix	Unknown	$\mathcal{O}(n^22^c)$	(3)
NN-EVCLUS	Relational	/	Dissimilarity matrix	Known	$\mathcal{O}(n^22^c)$	(1), (2), (3)
ECMdd	Relational	Medoids	Dissimilarity matrix	Known	$\mathcal{O}(cn^2 + n2^c)$	(1), (3), (4)

(1) The users need to know the number of clusters; (2) The users know some prior knowledge about the objects/clusters; (3) The method has a high computational complexity; (4) This method is not applicable to non-spherical symmetric data; (5) This method cannot detect partial ignorance (*e.g.* the meta-cluster  $A_j$ ) among singleton clusters,  $|A_j| \geq 2$ ,  $A_j \subset \Omega$ .

### 3.5.5 Challenges of evidential clustering

The successful applications of evidential clustering, such as in image segmentation, community detection, have demonstrated the potential, but many challenges remain, at least but not limited to the following.

- The relevant concepts, definitions, etc., involved in evidential clustering need to be further standardized.
- How to reduce the computational burden due to the introduction of meta-clusters has not been effectively addressed;
- Existing evidential methods are mainly focused on the detection of spherical clus-

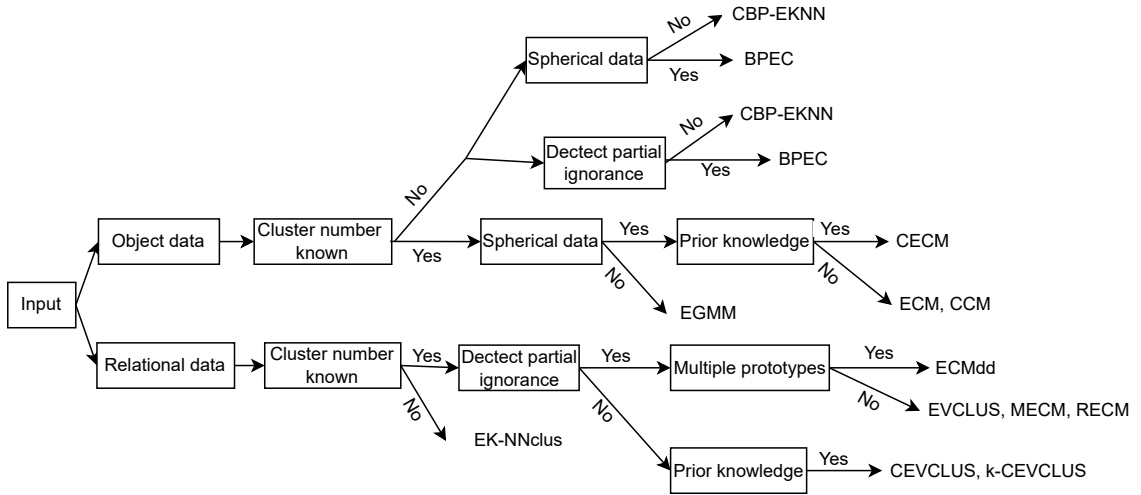


Fig. 3.4. The basic principles for choosing different evidential methods.

ters, with insufficient attention to clusters with arbitrary shape and size in space;

- Existing evidential methods are mainly developed for object and relational data. However, exploring soft data, including fuzzy/possibilistic/evidential ones, remains a challenging task;
- The applications of evidential clustering in different fields are still in the exploration stage, and many of the advantages are still to be explored.

### 3.6 Conclusion

In this chapter, we gave the concepts and definitions related to clustering. These normative fundamentals facilitate the in-depth study of evidential clustering as well as its applications. Besides, we also discussed the differences and connections between evidential clustering and other clustering methods. Afterward, we comprehensively reviewed evidential clustering (EC) from five aspects, *i.e.* object data-based methods, relational data-based methods, applications and analysis of different methods, and challenges of existing methods. Compared with other clustering methods, evidential clustering has the advantage of characterizing both uncertainty and imprecision between clusters. However, the existing evidential methods still have some defects, such as high computational complexity, and not applicable to non-spherical data. Based on this, we will propose three clustering algorithms in the subsequent Chapters 4, 5, and 6 to reduce the complexity and reasonably characterize the uncertainty and imprecision between imbalanced clusters or even clusters with arbitrary shapes and sizes.



# 4

## Dynamic evidential clustering algorithm

**Abstract:** In this chapter, we present a dynamic evidential clustering algorithm. To this end, we first review the related clustering algorithms such as evidential  $c$ -means and credal  $c$ -means. After that, we present in detail the implementation of the proposed algorithm. Then, we validate the effectiveness of the proposed algorithm on different datasets and conclude the chapter at the end.

### Contents

---

<b>4.1</b>	<b>Introduction</b>	<b>40</b>
<b>4.2</b>	<b>Review of evidential <math>c</math>-means and credal <math>c</math>-means</b>	<b>40</b>
4.2.1	Evidential $c$ -means (ECM)	40
4.2.2	credal $c$ -means (CCM)	41
<b>4.3</b>	<b>Dynamic evidential clustering (DEC)</b>	<b>43</b>
4.3.1	Preliminary credal partition	43
4.3.2	Partial credal redistribution	46
4.3.3	The involved parameters	48
4.3.4	The complexity of DEC	49
<b>4.4</b>	<b>Experiments</b>	<b>50</b>
4.4.1	Performance evaluation	50
4.4.2	Gaussian and round shape data	51
4.4.3	Large data	53
4.4.4	Real world data	55
4.4.5	Discussion	58
<b>4.5</b>	<b>Conclusion</b>	<b>60</b>

---

## 4.1 Introduction

In the previous chapter, the concepts and definitions related to evidential clustering (EC) are given, and the differences and connections between evidential clustering and other clustering methods are discussed. Based on this, EC is proved to characterize the uncertainty and imprecision between clusters well. However, due to the introduction of meta-cluster, traditional EC, such as ECM (Masson and Dencœux, 2008) and CCM (Liu et al., 2015), often requires a large amount of computation in extracting cluster structures, which limits the applications of EC in big data. In this chapter, we propose a dynamic evidential clustering (DEC) algorithm (Zhang et al., 2021d) to improve the efficiency of traditional EC. Since the DEC is an improvement of ECM and CCM, we first give the corresponding basics in Section 4.2, respectively. Afterward, the proposed DEC algorithm is presented in detail in Section 4.3. Then, experiments based on synthetic and real datasets illustrate the performances of DEC and show its difference from other methods in Section 4.4. Finally, we conclude this chapter in Section 4.5.

## 4.2 Review of evidential $c$ -means and credal $c$ -means

### 4.2.1 Evidential $c$ -means (ECM)

Evidential  $c$ -means (ECM) (Masson and Dencœux, 2008) is regarded as the evidential version of the fuzzy  $c$ -means (FCM) (Bezdek, 2013) and noise clustering (NC) (Sen and Davé, 1998; Dave, 1991) to characterize the uncertainty and imprecision between different clusters, and it will be briefly introduced as follows.

Let us consider a data set  $\mathcal{X}$  including  $n$  objects with  $s$  attributes over the FoD  $\Omega = \{\omega_1, \dots, \omega_c\}$ . For the object  $\mathbf{x}_i \in \mathcal{X}, i = 1, \dots, n$ , the mass of belief  $m_{ij} \triangleq m_i(A_j)$  for associating the object  $\mathbf{x}_i$  with an element  $A_j$  of the power-set  $2^\Omega$ . Particularly,  $A_j \subseteq \Omega, A_j \neq \emptyset$ , *i.e.*  $A_j$  can be any singleton cluster or meta-cluster included in  $2^\Omega$ . The cluster center  $\bar{\mathbf{v}}_j$  associated to  $A_j$  has been defined in Eq. (3.23).

In ECM, the value of  $m_{ij}$  depends on the distance  $d_{ij}$  between the object  $\mathbf{x}_i$  and the cluster center  $\bar{\mathbf{v}}_j$  of  $A_j$ , *i.e.* the higher distance  $d_{ij}$  leads to lower  $m_{ij}$ . ECM looks for the matrix  $M$  of credal partition and the matrix  $V$  of cluster centers by minimizing the objective function defined in Eqs. (3.21)-(3.22). In the function, the noise (outlier) threshold, denoted as  $\delta$ , represents the distance between any object  $\mathbf{x}_i$  ( $i = 1, \dots, n$ ) and the noise cluster.  $m_{i\emptyset}$  represents the mass of belief that the object assigned to the noise cluster and it can be adjusted by the threshold  $\delta$ . A bigger threshold  $\delta$  will lead to a lower mass of belief  $m_{i\emptyset}$ , and the object may be far away from the other objects if it is assigned to the noise cluster.

Afterward, the object function  $\mathcal{J}_{ECM}$  is minimized by the Lagrange multipliers to provide the matrix  $M$  of credal partition for the objects and the matrix  $V$  of cluster

centers, defined by:

$$\left\{ \begin{array}{l} m_{ij} = \frac{|A_j|^{-\alpha/(\beta-1)} d_{ij}^{-2/(\beta-1)}}{\sum_{A_k \neq \emptyset} |A_k|^{-\alpha/(\beta-1)} d_{ik}^{-2/(\beta-1)} + \delta^{-2/(\beta-1)}}, \text{ if } A_j \neq \emptyset \\ m_{ij} = 1 - \sum_{A_j \neq \emptyset} m_{ij}, \text{ if } A_j = \emptyset \end{array} \right. \quad (4.1)$$

where  $d_{ij}$  represents the distance between the  $\mathbf{x}_i$  and the center of cluster  $A_j$ . The exponent  $\alpha$  is used to control the degree of penalization and the  $\beta$  is a weighting exponent and it is generally set  $\beta = 2$  as default.

The centers of the cluster are given by the rows of the matrix  $V_{c \times s}$ , given by:

$$V_{c \times p} = H_{c \times c}^{-1} B_{c \times p} \quad (4.2)$$

subject to

$$B_{lq} = \sum_{i=1}^n \mathbf{x}_{iq} \sum_{\{\omega_l\} \in A_j} |A_j|^{\alpha-1} m_{ij}^\beta \quad (4.3)$$

$$H_{lk} = \sum_{i=1}^n \sum_{\{\omega_l, \omega_k\} \subseteq A_j} |A_j|^{\alpha-2} m_{ij}^\beta \quad (4.4)$$

where  $B_{lq}$  ( $l \in [1, c], q \in [1, p]$ ) and  $H_{lk}$  ( $l, k \in [1, c]$ ) represent the elements in the matrix  $B_{c \times p}$  and  $H_{c \times c}$ , respectively.

### 4.2.2 credal $c$ -means (CCM)

Credal  $c$ -means clustering (CCM) method (Liu et al., 2015) addresses the disadvantages of ECM when the singleton cluster centers and the meta-cluster centers are very close, which may produce unreasonable results.

Compared with ECM, the mass of belief that the object  $\mathbf{x}_i$  assigned to the meta-cluster in CCM depends not only on the distance from  $\mathbf{x}_i$  to the meta-cluster center but also the distance between  $\mathbf{x}_i$  and the singleton clusters included in meta-cluster, which can avoid the unreasonable result provided by ECM in clustering the dataset with special distribution. Additionally, CCM sets a threshold  $t_c \in [2, 2^c]$  to eliminate some meta-clusters with big cardinality so as to reduce the computational complexity especially in the dataset with abundant clusters. Thereby, the set of the selected available clusters  $S^\Omega$  is given by  $S^\Omega = \{A_j, |A_j| < t_c\}$ . According to the above basic

principle, the object function  $\mathcal{J}_{CCM}$  of CCM can be defined by:

$$\mathcal{J}_{CCM}(M, V) = \sum_{i=1}^n \sum_{A_j \in S^\Omega} m_{ij}^\beta D_{ij}^2, \quad (4.5)$$

subject to

$$\sum_{A_j \in S^\Omega} m_{ij} = 1 \quad (4.6)$$

and

$$D_{ij}^2 = \begin{cases} \delta^2, & \text{if } |A_j| = \emptyset \\ d_{ij}^2, & \text{if } |A_j| = 1 \\ \frac{\sum_{\omega_l \in A_j} d_{il}^2 + \gamma d_{ij}^2}{|A_j| + \gamma}, & \text{if } |A_j| > 1 \end{cases} \quad (4.7)$$

where  $d_{ij}$  is the Euclidean distance between  $\mathbf{x}_i$  and the meta-cluster ( $A_j$ ) center, and  $d_{il}$  represents the distances from  $\mathbf{x}_i$  to the singleton cluster ( $\{\omega_l\}$ ) center in meta-cluster  $A_j$  such that  $|A_j| > 1$ .  $\gamma$  is the weighting factor of the distance between the object and the meta-cluster center, and it is used to control the imprecision rate. The bigger the value of  $\gamma$  is, the more objects will be assigned to the meta-clusters, and it is generally taken  $\gamma \in [0.5, 3]$ . The weighting exponent  $\beta$  is set  $\beta = 2$  as default.

The function  $\mathcal{J}_{CCM}(M, V)$  is minimized by the Lagrange multipliers method to get the matrix  $M$  of credal partition and the matrix  $V$  of cluster centers, defined by:

$$m_{ij} = \frac{D_{ij}^{-2/(\beta-1)}}{\sum_{A_k \in S^\Omega} D_{ik}^{-2/(\beta-1)}}. \quad (4.8)$$

The centers of the cluster are given by the rows of the matrix  $V_{c \times p}$ , given by:

$$V_{c \times p} = H_{c \times c}^{-1} B_{c \times n} \mathcal{X}_{n \times p} \quad (4.9)$$

subject to

$$B_{li} = m_{il}^\beta + \sum_{A_l \in A_j} m_{ij}^\beta \frac{1 + \gamma}{|A_j| + \gamma} \quad (4.10)$$

$$H_{ll} = \sum_{i=1}^n m_{il}^\beta + \sum_{i=1}^n \sum_{A_l \in A_j} m_{ij}^\beta \frac{1 + \frac{\gamma}{|A_j|^2}}{|A_j| + \gamma} \quad (4.11)$$

$$H_{lq} = \sum_{i=1}^n \sum_{\{A_l, A_q\} \in A_k} m_{ij}^\beta \frac{\gamma}{|A_k|^2 (|A_k| + \gamma)}, \quad l \neq q \quad (4.12)$$

where  $B_{li}$  ( $l \in [1, c], i \in [1, n]$ ) and  $H_{lq}$  ( $l, q \in [1, c]$ ) represent the elements in the matrix

$B_{c \times n}$  and  $H_{c \times c}$ , respectively.

Since the proposed DEC algorithm in this chapter is an improvement of EC in improving the computational efficiency. After introducing the classical ECM and CCM algorithms, we next present the proposed DEC algorithm in detail.

### 4.3 Dynamic evidential clustering (DEC)

To reduce the computational complexity of existing EC methods, a new alternative version, named dynamic evidential clustering (DEC) (Zhang et al., 2021d), is proposed in this section. It is based on the following two assumptions.

1) For the same query set, the centers obtained in the fuzzy (possibilistic) partition and that of singleton clusters obtained by the credal partition are very similar. This means that the meta-clusters can be ignored in the initial iterations because the centers of meta-clusters are defined on the related singleton clusters.

2) Only a part of objects in the query set is difficult to be accurately assigned to singleton clusters. They are then assigned to the related meta-clusters composed of only several close singleton clusters. Thus, it is not necessary to assign all the objects under the power-set  $2^\Omega$ .

Based on the above assumptions, the DEC method can be summarized as two steps: 1) Preliminary credal partition; 2) Partial credal redistribution.

#### 4.3.1 Preliminary credal partition

The purpose of this subsection is to preliminary assign each object in the query set as the outlier, precise or imprecise one adaptively. To derive such a proposal, let's consider a query set  $\mathcal{X}$  including  $n$  objects in  $s$ -dimensions with  $\Omega = \{\omega_1, \dots, \omega_c\}$ . The support degrees of each object belonging to different singleton (specific) clusters and the noise cluster, called the mass of beliefs in credal partition, can be minimized by an FCM-like objective function at first. There exist many methods to obtain the mass of beliefs. For example, Davé's noise-clustering (Sen and Davé, 1998; Dave, 1991) can be applied for the query set, and we have modified it as the version of the credal partition to facilitate the presentation. The objective function can be expressed as follows:

$$J_{DEC-NC}(M_1, V_1) = \sum_{i=1}^n \sum_{j=1}^c m_{ij}^\beta \cdot d_{ij}^2 + \sum_{i=1}^n \delta^2 \cdot m_{i\emptyset}^\beta \quad (4.13)$$

subject to

$$\sum_{j=1}^c m_{ij} + m_{i\emptyset} = 1, \quad \forall i = 1, n \quad (4.14)$$



where  $M_1 = (m_1, \dots, m_n)$  is the mass of belief matrix for  $n$  objects in  $\mathcal{X}$ , and  $V_1$  is the matrix of the centers of singleton clusters.  $d_{ij}$  is the Euclidean distance between the object  $\mathbf{x}_i$  and the center of singleton cluster  $\{\omega_j\}$ . Parameters  $\beta$ ,  $\delta$  are adjustable with the same meanings as those in NC, ECM and CCM. Since it has the same structure as NC, we directly give the mass of beliefs of the query object  $\mathbf{x}_i$  belonging to different clusters (*i.e.* the noise and singleton clusters) as follows:

$$m_{ij} = \frac{d_{ij}^{-2/(\beta-1)}}{\sum_{k=1}^c d_{ik}^{-2/(\beta-1)} + \delta^{-2/(\beta-1)}} \quad (4.15)$$

and

$$m_{i\emptyset} = 1 - \sum_{j=1}^c m_{ij}, \quad \forall i = 1, n \quad (4.16)$$

where  $m_{ij}$  is the mass of belief of the object  $\mathbf{x}_i$  belonging to the cluster  $\{\omega_j\}$ , and  $m_{i\emptyset}$  represents the possibility of  $\mathbf{x}_i$  belonging to the noise cluster (*i.e.*  $\emptyset$ ).

By doing this, we can assign the query object to different clusters according to the mass of belief matrix. Here the maximum of belief function  $Bel(\cdot)$  given in Eq. (2.4) is used as the criteria for the decision-making of the cluster which is strongly supported by the various mass of beliefs. Specifically, the object  $\mathbf{x}_i$  is directly assigned to the noise cluster if  $Bel(\emptyset)$  is larger than the others, and defined by:

$$Bel_i(\emptyset) > Bel_i(\{\omega_j\}), \quad j \in [1, c]. \quad (4.17)$$

where  $Bel_i(\emptyset) = m_{i\emptyset}$  and  $Bel_i(\{\omega_j\}) = m_{ij}$  since  $Bel(\cdot)$  has a straight corresponding relationship in such particular BBAs structure (Liu et al., 2014). In such a case, we don't further assign the object since it is obviously considered as the outlier. Whereas if the cluster  $\{\omega_j\}$  is the most believed cluster for  $\mathbf{x}_i$ , defined by:

$$Bel_i(\{\omega_j\}) = \max[Bel_i(\{\omega_1\}), \dots, Bel_i(\{\omega_c\})]. \quad (4.18)$$

We need to further judge the possibility that the object belongs to  $\{\omega_j\}$ , because  $Bel_i(\{\omega_j\})$  maybe not significantly different from others.

**Example 4.1:** The 4-tuple  $M = (m_1, m_2, m_3, m_4)$  in Table 4.1 is an example of preliminary credal partition. One can see that the objects  $\mathbf{x}_1$  and  $\mathbf{x}_2$  likely belong to the outlier  $\emptyset$  and the singleton cluster  $\omega_1$ , respectively. In contrast, the objects  $\mathbf{x}_3$  and  $\mathbf{x}_4$  will be assigned to the clusters  $\{\omega_2\}$  and  $\{\omega_1\}$  respectively from the probabilistic perspective, but it has the risk of error since this slight difference may be caused by multiple reasons. The objects like  $\mathbf{x}_3$  and  $\mathbf{x}_4$  should be further assigned because they would be better to assigned to the meta-clusters  $\{\omega_2, \omega_3\}$  and  $\{\omega_1, \omega_2\}$ , respectively.

In practice, it also happens that the belief  $Bel_i(\{\omega_l\})$  of the other cluster  $\{\omega_l\}$

Table 4.1: The numerical example

	$\emptyset$	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_3\}$
$m_1$	1.0	0	0	0
$m_2$	0	0.9	0	0.1
$m_3$	0	0.01	0.5	0.49
$m_4$	0	0.43	0.42	0.15

( $l \in [1, c]$ ,  $l \neq j$ ) is very close (or equal) to the  $Bel_i(\{\omega_j\})$  of the strongest cluster  $\{\omega_j\}$ . In such a case, the object can also potentially belong to  $\{\omega_l\}$  with a high likelihood, and we should adopt a more cautious strategy in the preliminary credal partition. That is, we need to consider all the very likely specific clusters as the potential solution for  $\mathbf{x}_i$ . The set of these potential singleton clusters can be called  $\mathcal{M}_i^2$ , defined by:

$$\mathcal{M}_i = \{\omega_j, \dots, \omega_l\}, Bel_i(\{\omega_j\}) - Bel_i(\{\omega_l\}) \leq \varphi \quad (4.19)$$

where  $\varphi \in [0, 1]$  is a chosen meta-cluster threshold depending on the imprecision one can accept. In fact,  $\varphi$  is a very interesting threshold, which determines the several frameworks of clustering as follows.

- **Fuzzy/possibilistic partition:** We have a fuzzy/possibilistic partition with  $\varphi = 0$  since all objects will be assigned directly to specific (singleton) clusters in the Bayesian framework. We don't need to partial credal redistribution for imprecise objects since there is no meta-cluster at this time. In contrast, the query set will be clustered under credal partition with  $\varphi > 0$ .
- **Traditional credal partition:** We have a traditional credal partition with  $\varphi = 1$ , which means that all objects are imprecise ones in preliminary credal partition. All query objects need to be redistributed under the power-set  $2^\Omega$ .
- **Dynamic credal partition:** We have a dynamic (partial) credal partition with  $\varphi \in (0, 1)$  to consider  $\{\omega_l\}$  differently because only a part of objects in the query set is difficult to be accurately assigned to singleton clusters.

Based on the above analysis, all clusters in  $\mathcal{M}_i$  may very likely correspond to the real (unknown) clusters for  $\mathbf{x}_i$ , and they appear indistinguishable with respect to the threshold  $\varphi$ . That is, the imprecise object  $\mathbf{x}_i$  needs to be further assigned under the

<sup>2</sup>Here we provide a corresponding  $\mathcal{M}_i$  for each imprecise object  $\mathbf{x}_i$ , called the specific dynamic edited framework. In fact, since in DEC, each  $\mathcal{M}_i$  is specific, we can also call it an adaptive edited framework for each imprecise object.

new frame  $\mathcal{M}_i^\Omega$ , and defined by:

$$\mathcal{M}_i^\Omega = \{\{\omega_j\}, \dots, \{\omega_l\}, \{\omega_j \cup \omega_l\}, \dots, \mathcal{M}_i\} \quad (4.20)$$

By doing this, one can easily find that only a few objects need to be further reassigned. It can greatly reduce the computational complexity, and each imprecise object has a specific dynamic edited framework. An example is given to illustrate this.

**Example 4.2:** Let us consider a chosen meta-cluster threshold  $\varphi = 0.1$ , and the objects  $\mathbf{x}_3, \mathbf{x}_4$  in Table 4.1 are employed here to illustrate the specific dynamic edited framework. For the object  $\mathbf{x}_3$ , one can obtain that  $\omega_2$  is the most likely cluster with  $Bel_3(\{\omega_2\}) = \max[Bel_3(\{\omega_1\}), Bel_3(\{\omega_2\}), Bel_3(\{\omega_3\})]$ . However, it does not mean that  $\mathbf{x}_3$  is directly assigned to the singleton cluster  $\{\omega_2\}$  since there is another cluster (*i.e.*  $\{\omega_3\}$ ) that the object may belong to under the chosen meta-cluster threshold  $\varphi = 0.1$  with  $\mathcal{M}_3 = \{\omega_2 \cup \omega_3 \mid Bel_3(\{\omega_2\}) - Bel_3(\{\omega_3\}) \leq 0.1\}$ . The object  $\mathbf{x}_3$  thereby is considered as an imprecise object by preliminary credal partition, and it will be reclustered by partial credal redistribution with the specific dynamic edited framework  $\mathcal{M}_3^\Omega = \{\omega_2, \omega_3, \{\omega_2, \omega_3\}\}$ . One can also deduce that the imprecise object  $\mathbf{x}_4$  will be reassigned under the specific dynamic edited framework  $\mathcal{M}_4^\Omega = \{\omega_1, \omega_2, \{\omega_1, \omega_2\}\}$ . The flowchart in Fig. 4.1 explicitly shows how the preliminary credal partition works.

### 4.3.2 Partial credal redistribution

In DEC, the center  $\bar{\mathbf{v}}_j$  of the meta-cluster  $A_j$  is also defined as mean value of that of the singleton clusters included in  $A_j$ , which is similar to ECM and defined in Eq. (3.23).

Whereas the distance between the object  $\mathbf{x}_i$  and the meta-cluster  $A_j$  depends not only on the distance between the object  $\mathbf{x}_i$  and the center  $\bar{\mathbf{v}}_j$ , but also on the distance between the object  $\mathbf{x}_i$  and the center  $\mathbf{v}_k$  of all the singleton clusters included in  $A_j$ . Thus, the distance  $D_{ij}^2$  from the object to different clusters can be defined as follows:

$$D_{ij}^2 = \begin{cases} d_{ij}^2, & \text{if } |A_j| = 1 \\ \frac{\sum_{A_k \in A_j} d_{ik}^2 + d_{ij}^2}{|A_j| + 1}, & \text{if } |A_j| > 1 \end{cases} \quad (4.21)$$

where  $d_{ij}$  is the Euclidean distance between the object  $\mathbf{x}_i$  and the center  $\bar{\mathbf{v}}_j$  of the cluster  $A_j$ . If  $A_j$  is a singleton cluster,  $\bar{\mathbf{v}}_j$  is the center of  $A_j$ . In contrast,  $\bar{\mathbf{v}}_j$  is the mean value of the included singleton cluster centers if  $A_j$  is a meta-cluster.

We can find that the distance between the object and the meta-cluster is similar to that in CCM, but they are different. In DEC,  $|A_j| + 1$  in Eq. (4.21) denotes the number of distances from the object to different clusters. Whereas  $|A_j| + \gamma$  varies with the change of  $\gamma$  of CCM, it may lead to some specific objects being assigned to meta-clusters and then increase the number of objects in meta-clusters unreasonably. Once

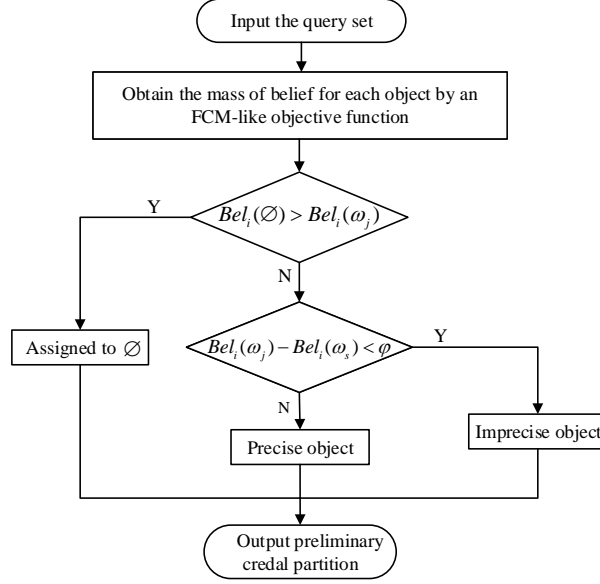


Fig. 4.1. Flowchart of the adaptive credal partition.

$\mathcal{M}_i^\Omega$  is obtained, we only need to calculate the distance between the object and the related clusters in one iteration. This is the reason why the DEC can greatly reduce the computing time and produce similar clustering results with ECM and CCM.

For the imprecise object  $\mathbf{x}_i$  with  $\mathcal{M}_i^\Omega$ , we need to update its mass of beliefs  $m(\cdot)$  belonging to different clusters under the edited framework  $\mathcal{M}_i^\Omega$ . Assuming that there are  $n_1$  imprecise objects after preliminary credal partition. Since the centers of singleton and meta-clusters have been obtained, we can directly update the mass of belief  $m(\cdot)$  for each imprecise object. Inspired by ECM and CCM, the update formula of mass values can be derived by minimizing the other sub-objective function denoted as follows:

$$J_{DEC-CR}(M_2, V_2) = \sum_{i=1}^{n_1} \sum_{j/A_j \in \mathcal{M}_i^\Omega} m_{ij}^\beta D_{ij}^2 \quad (4.22)$$

subject to

$$\sum_{j/A_j \in \mathcal{M}_i^\Omega} m_{ij} = 1 \quad (4.23)$$

where  $n_1$  is the number of imprecise objects, and  $V_2$  is the center matrix of singleton clusters and meta-clusters. Although Eq. (4.22) is inspired by ECM and CCM, the noise cluster is no longer considered here since noise has been well identified in preliminary credal partition.

Since we have known the centers of singleton clusters and meta-clusters, *i.e.*  $V_2$  is known, Lagrange multipliers  $\lambda_i$  are used to solve the constrained minimization problem

with respect to  $M_2$  as follows:

$$\mathcal{L}(M_2, \lambda_1, \dots, \lambda_n) = J_{DEC-CR}(M_2, V_2) - \sum_{i=1}^{n_1} \lambda_i \left( \sum_{j=1}^c m_{ij} - 1 \right). \quad (4.24)$$

By differentiating the Lagrangian with respect to the  $m_{ij}$  and  $\lambda_i$  and setting the derivatives to zero, we obtain:

$$\frac{\partial \mathcal{L}}{\partial m_{ij}} = \beta m_{ij}^{\beta-1} D_{ij}^2 - \lambda_i = 0, \quad (4.25)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = \sum_{j/A_j \in \mathcal{M}_i^\Omega} m_{ij} - 1 = 0. \quad (4.26)$$

From Eq. (4.25), we thus have:

$$m_{ij} = \left( \frac{\lambda_i}{\beta} \right)^{\frac{1}{\beta-1}} \left( \frac{1}{D_{ij}^2} \right)^{\frac{1}{\beta-1}}, \quad (4.27)$$

Using Eqs. (4.26) - (4.27):

$$\left( \frac{\lambda_i}{\beta} \right)^{\frac{1}{\beta-1}} = \frac{1}{\sum_{j/A_j \in \mathcal{M}_i^\Omega} D_{ij}^{\frac{-2}{\beta-1}}}. \quad (4.28)$$

Returning in Eq. (4.27), one obtains the necessary condition of optimality for  $M_2$ :

$$m_{ij} = \frac{D_{ij}^{\frac{-2}{\beta-1}}}{\sum_{k/A_k \in \mathcal{M}_i^\Omega} D_{ik}^{\frac{-2}{\beta-1}}}. \quad (4.29)$$

We can find that only one iteration is needed to obtain the mass of beliefs for each imprecise object  $\mathbf{x}_i$  because  $V_2$  is known. Once the imprecise object is reassigned to a specific cluster, it indicates that the object is precise under the current knowledge. In contrast, if the object is final assigned to a meta-cluster, the object may belong to any singleton clusters included in the meta-cluster depending on the known information.

### 4.3.3 The involved parameters

There are three parameters involved in the DEC method: the meta-cluster threshold  $\varphi$ , the weighting exponent  $\beta$  and the outlier threshold  $\delta$ . The meta-cluster threshold  $\varphi$  can control the number of objects assigned in meta-clusters, and the larger the  $\varphi$ , the more objects in meta-clusters. Thus,  $\varphi$  can be adjusted according to the number

of objects in meta-clusters that one can accept, and  $\varphi = 0.3$  is the default value. The use of the tuning parameter  $\beta$  is similar to ECM and CCM, and  $\beta = 2$  can be used as default value. Parameter  $\delta$  is strongly dependent on the dataset to be clustered and controls the amount of data considered as outliers which has been discussed in NC.

#### 4.3.4 The complexity of DEC

In order to reduce the computational complexity of existing EC, the DEC first uses an FCM-like objective function to iteratively obtain the real centers of singleton clusters and preliminary adaptively assign the query object as the outlier, precise or imprecise one by the preliminary credal partition, and then the results of the partial credal redistribution for imprecise objects are used as the final decision-making. Since the partial credal redistribution is only for a part of objects and only needs to be iterated once, the computational complexity of DEC mainly depends on the preliminary credal partition. Thus, the computational complexity is  $\mathcal{O}(n \cdot c)$  which is much lower than that of ECM ( $\mathcal{O}(n \cdot 2^c)$ ) where  $n$  is the number of the objects in the query set and  $c$  is the real number of singleton clusters.

By the way, the proposed DEC can obtain the real singleton cluster centers in the first step, which is very important to reduce the complexity. After that, we can directly update the mass value under the TBF framework. This is heuristic, and from this point of view, we can take a similar way to obtain the real singleton cluster centers and then globally optimize all the parameters, keeping the power-set  $2^\Omega$  unchanged to reduce the complexity of other classical EC methods (Zhang et al., 2021a). In addition, there are some improved EC methods to handle large-scale relational data. For example, the classical EVCLUS (Denceux and Masson, 2004) method is improved to  $k$ -EVCLUS (Denceux and Kanjanatarakul, 2016), making it applicable to very large relational data. First, the gradient-based optimization procedure in the original EVCLUS is replaced by a much faster iterative row-wise quadratic programming method (Ter Braak et al., 2009). Second, they also show that EVCLUS can be provided with only a random sample of the dissimilarities, reducing the time and space complexity from quadratic to roughly linear. The complexity of  $k$ -EVCLUS is  $\mathcal{O}(k \cdot n)$ , where  $k$  is a given parameter in applications. We can find that DEC and  $k$ -EVCLUS are similar in terms of computational complexity, which means that we can use the corresponding methods when dealing with different large-scale object and relational datasets.

The pseudo-code shown in **Algorithm 1** is used to clearly explain the working principle of the DEC method.

---

**Algorithm 1** Dynamic Evidential Clustering Algorithm

---

**Require:** Data to cluster:  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  in  $\mathbb{R}^p$ ; Parameters:  $c, \beta, \delta, \varphi$ .

**Ensure:** Cluster decision results

**Step 1**

Construct the objective function  $J_{DEC-NC}$  without meta-clusters by Eqs. (4.13)-(4.14);

Iterate the mass of beliefs for each object by Eqs. (4.15)-(4.16);

**for** the  $1$ -th to  $n$ -th query object Preliminary assign the object as the outlier, precise or imprecise one using Eqs. (4.17)-(4.20);

**end**

**Subreturn:** Preliminary credal partition.

**Step 2**

Calculate the centers of meta-cluster using Eq. (3.23);

Reconstruct the objective function  $J_{DEC-CR}$  for credal redistribution using Eqs. (4.22)-(4.23);

Reiterate the mass of beliefs for the  $n_1$  imprecise objects using Eqs. (4.21), (4.29);

**for** the  $1$ -th to  $n_1$ -th imprecise object

Reassign the object to singleton cluster or meta-cluster.

**end**

**Subreturn:** Partial credal redistribution.

**Return:** Dynamic credal partition.

---

## 4.4 Experiments

### 4.4.1 Performance evaluation

In this section, four experiments are conducted to evaluate the performance of the proposed DEC method compared with  $K$ -means (Jain, 2010), FCM (Bezdek, 2013), NC (Sen and Davé, 1998), ECM (Masson and Dencœux, 2008), CCM (Liu et al., 2015), BPEC (Su and Dencœux, 2018). Except for the parameters set in the experiments, the others are all defaults. Since the introduction of meta-cluster, the error rate and imprecision rate (Liu et al., 2015) are used as the indicators of different methods. The error rate denoted by  $Re$  (in %) is calculated by  $Re = n_e/n$ , where  $n_e$  is the number of clustering errors, and  $n$  is the number of objects under the test. The imprecision rate denoted by  $Ri$  (in %) is calculated by  $Ri = n_i/n$ , where  $n_i$  is number of objects assigned to meta-clusters. Moreover, we also evaluate the clustering results using Credal Rand Index (CRI) (Dencœux et al., 2017), which is employed to measure the closeness of credal partition and ground truth. It is defined by:

$$CRI = (M, M^*) = \frac{\sum_{i < j} pl_{ij}(s)^{r_{ij}^*} pl_{ij}(\neg s)^{1-r_{ij}^*}}{n(n-1)/2} \quad (4.30)$$

where  $M$  represents the evidential partition and  $M^*$  is the true hard partition,  $n$  is the number of objects in the dataset and  $r_{ij}^* = 1$  if the  $i$ -th and  $j$ -th object truly belong to the same cluster, and  $r_{ij}^* = 0$  otherwise. Particularly,  $CRI$  and  $ARI$  are equal when comparing the closeness of hard partitions to truth, so the performance index is uniformly denoted by “ $CRI$ ” in this thesis. The upper bound of  $CRI$  is 1, a larger value corresponds to a better clustering result. In order to compare the computational complexity of all algorithms more fairly, in ECM, CCM, and BPEC, the number of focal elements in meta-clusters is limited to 2 by a given threshold  $t_c = 2$ . For more detailed explanation, please see (Liu et al., 2015; Su and Dencœux, 2018). In addition to the above indexes, pieces of literature put forward some other optimality criteria. Here we employ the objective function value and the number of iterations to evaluate the performance of DEC. In the tests, we report the average of these indexes based on all methods that have been run 10 times. The elapsed time denoted by  $T$  (in second).

#### 4.4.2 Gaussian and round shape data

This experiment is mainly to explain the use of the DEC in clustering on two 3-class datasets clearly, which we called Gaussian Data and Round Shape Data, respectively.

**(1) Gaussian Data:** A 3-class gaussian dataset is employed, as shown in Fig. 4.2, to test ECM and DEC. The set has  $3 \times 100 = 300$  data points from three  $2D$  Gaussian distributions with the mean variance parameters as follows:

$$\mu_1 = (3, 3), \Sigma_1 = 1.5\mathbf{I};$$

$$\mu_2 = (0, 0), \Sigma_2 = \mathbf{I};$$

$$\mu_3 = (8, 8), \Sigma_3 = \mathbf{I}.$$

where  $\mu_i$  ( $i = 1, 2, 3$ ) is the mean vector,  $\Sigma_i$  ( $i = 1, 2, 3$ ) is the covariance matrix, and  $\mathbf{I}$  is the identity matrix.

In ECM and DEC, the noise threshold is set as  $\delta = 5$ , and other parameters are the defaults. The clustering results of different methods are shown in Fig. 4.2 (b)-(c).

We can see from Fig. 4.2 (a) that there are some objects in the overlapping areas of different clusters  $\{\omega_1\}$  and  $\{\omega_2\}$ , which are difficult to be correctly distinguished for the clusters. The clusters  $\{\omega_2\}$  and  $\{\omega_3\}$  are far apart, however, many of the objects belonging to  $\{\omega_1\}$  are incorrectly assigned to the meta-cluster  $\{\omega_2, \omega_3\}$  labeled by pink plus in ECM, as shown in Fig. 4.2 (b). In addition, a lot of objects belonging to  $\{\omega_1\}$  are assigned to the total ignorant cluster  $\Omega$  labeled by the black multiplier. Meanwhile, some objects belonging to  $\{\omega_1\}$  are assigned to the meta-cluster  $\{\omega_1, \omega_3\}$  labeled by the green plus, although there is no overlapping area between  $\{\omega_1\}$  and  $\{\omega_3\}$ . These unreasonable clustering results are mainly caused by the close proximity of different cluster centers (*i.e.*  $\mathbf{v}_1, \bar{\mathbf{v}}_{2,3}, \bar{\mathbf{v}}_{1,2,3}$ ). In contrast, there is no object in DEC assigned



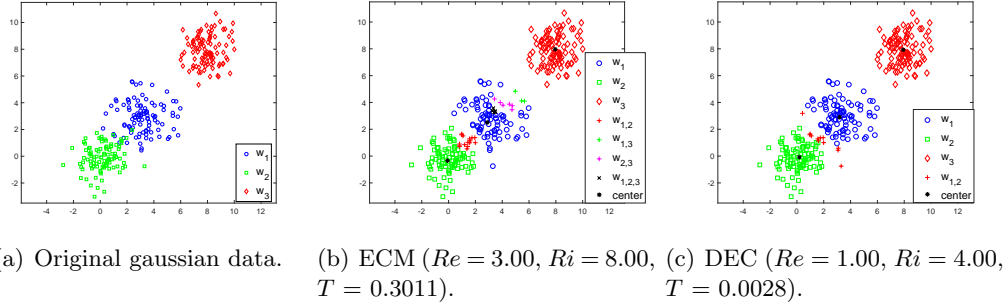


Fig. 4.2. Clustering results of 3-class of gaussian dataset by different methods.

to the meta-clusters  $\{\omega_2, \omega_3\}$  and  $\Omega$ , as shown in Fig. 4.2 (c), and the objects in the overlapping area are reasonably assigned to the corresponding meta-cluster  $\{\omega_1, \omega_2\}$ . The DEC has obtained reasonable results, which not only reduces the clustering error rate but also reduces the clustering imprecision compared with ECM.

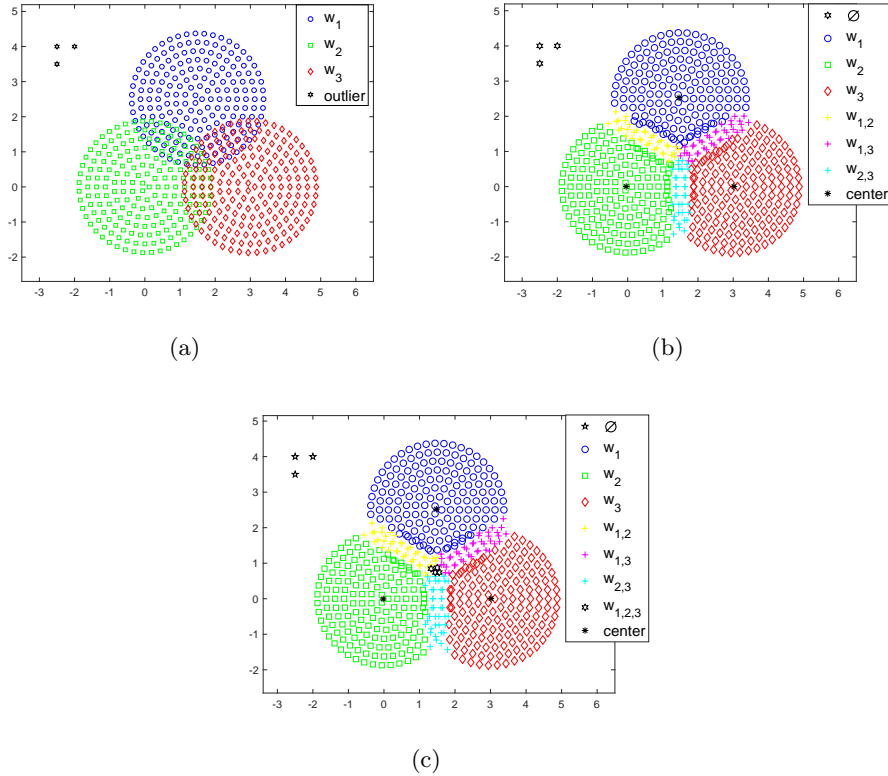


Fig. 4.3. Clustering results of 3-class of round shape dataset by different methods. (a) Original artificial data. (b) CCM ( $Re = 7.07$ ,  $Ri = 17.34$ ,  $T = 0.5897$ ). (c) DEC ( $Re = 4.23$ ,  $Ri = 20.81$ ,  $T = 0.0072$ ).

**(2) Round Shape Data:** We consider a particular 3-class dataset in the round shape as shown in Fig. 4.3 (a), and some objects are in the overlapping areas. The dataset consists of 594 data points, including 3 outlier data points. The radius of the round is  $r = 2$  and the centers of the three rounds are given by the points  $c_1 = (1.5, 2.5)$ ,  $c_2 = (0, 0)$ ,  $c_3 = (3, 0)$ .

Here CCM and DEC are applied to clustering the dataset with  $\delta = 3$ . Fig. 4.3 (b)-(c) show the clustering results of CCM and DEC respectively. The error rate (in %), the imprecision rate (in %), and computation time are given in the caption. We can see from Fig. 4.3 (a) that the objects in the clusters  $\{\omega_1\}$ ,  $\{\omega_2\}$ , and  $\{\omega_3\}$  overlap partially at their edges, and these objects in overlapping areas are difficult to cluster accurately and clearly. In Fig. 4.3 (b), we can see that CCM assigns data points in the overlapping areas of different clusters to the corresponding meta-clusters:  $\{\omega_1, \omega_2\}$ ,  $\{\omega_1, \omega_3\}$ ,  $\{\omega_2, \omega_3\}$ . However, some objects in the middle area of these three clusters are not accurately assigned. This unreasonable assignment is mainly because the meta-cluster threshold is selected as  $t_c = 2$  in CCM. This also shows that it may be unreasonable for ECM and CCM to reduce the computational burden by limiting the number of elements in the meta-cluster in a few cases.

In Fig. 4.3 (c), the DEC assigns the objects in the overlapping areas of  $\{\omega_1\}$  and  $\{\omega_2\}$ ,  $\{\omega_1\}$  and  $\{\omega_3\}$ ,  $\{\omega_2\}$  and  $\{\omega_3\}$  to meta-clusters  $\{\omega_1, \omega_2\}$ ,  $\{\omega_1, \omega_3\}$  and  $\{\omega_2, \omega_3\}$ . Meanwhile, the DEC can reasonably assign the objects in the overlapping area of these three clusters to the meta-cluster  $\Omega$  because these objects are totally unrecognized. Three objects labeled with the black hexagram are far away from the others, the CCM and DEC have better detection, and all of them are regarded as the outliers. Meanwhile, the execution time of DEC is much shorter than that of CCM, which indicates that the computational complexity of DEC is significantly less than CCM.

#### 4.4.3 Large data

In this experiment, we test the methods on the dataset consisting of 180000 points generated by different Gaussian distributions where each class contains 30000 data points. The dataset is shown in Fig. 4.4 (a). Here we take  $\delta = 10$ , and the meta-cluster threshold  $\varphi = 0.2$  and  $\varphi = 0.3$  in DEC. The results obtained by different methods are specified in the caption of each subfigure.

In Fig. 4.4 (a), we can clearly see that the objects in the overlapping areas are difficult to be accurately assigned to singleton clusters. The ECM and BPEC obtain similar clustering results in Fig. 4.4 (b) and (d), and they assign the imprecise objects to the corresponding meta-clusters with credal partition, for instance, the objects lying in the overlapping areas of  $\{\omega_1\}$  and  $\{\omega_2\}$  are assigned to the meta-cluster  $\{\omega_1, \omega_2\}$ . However, ECM and BPEC also unreasonably assign the objects that originally belong to singleton clusters to the unrelated meta-clusters. In Fig. 4.4 (a),  $\{\omega_2\}$  and  $\{\omega_6\}$  are

completely separated, for example, ECM and BPEC still assign some objects belonging to  $\{\omega_1\}$  to the meta-cluster  $\{\omega_2, \omega_6\}$  labeled by purple plus.

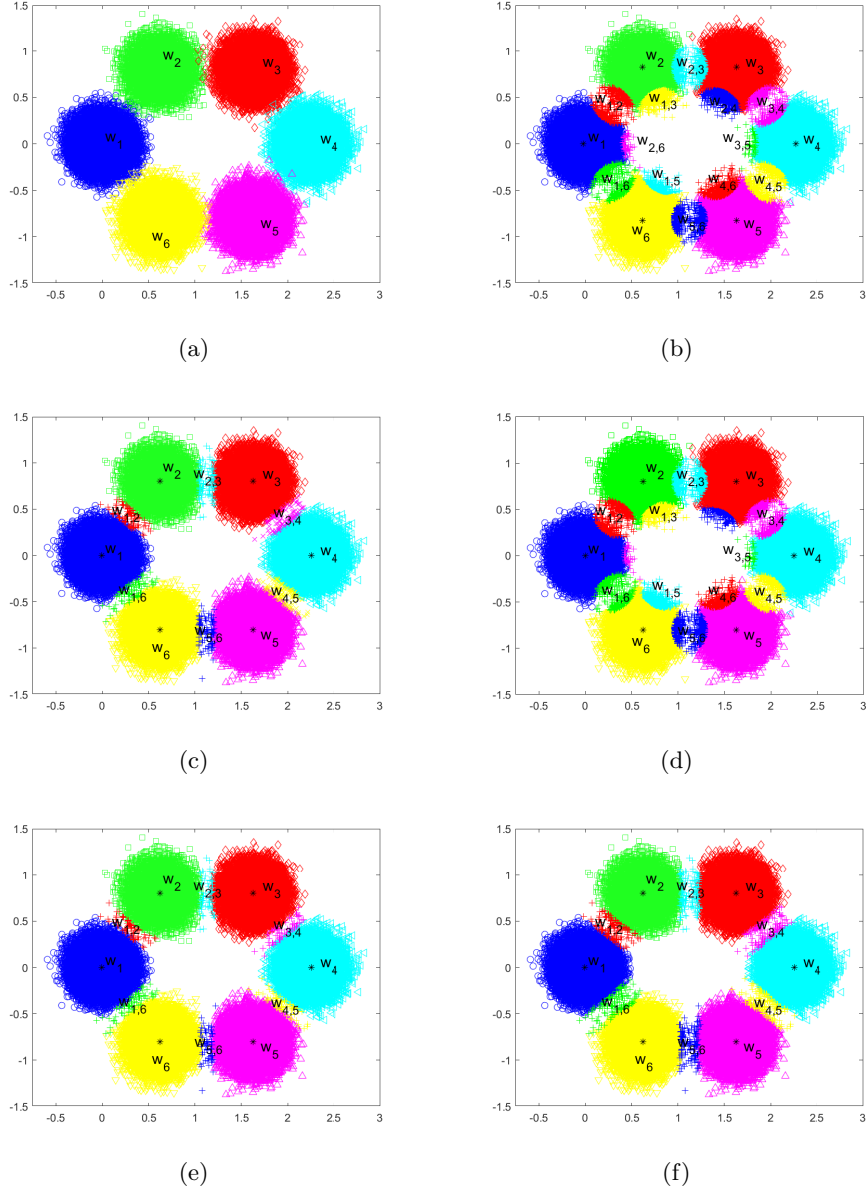


Fig. 4.4. Clustering results of 6-class of dataset by different methods. (a) Original artificial data. (b) ECM ( $Re = 1.13, Ri = 1.23, T = 4354.21$ ). (c) CCM ( $Re = 0, Ri = 0.30, T = 3833.51$ ). (d) BPEC ( $Re = 0.82, Ri = 1.14, T = 10534.04$ ). (e) DEC with  $\varphi = 0.2$  ( $Re = 0, Ri = 0.21, T = 85.31$ ). (f) DEC with  $\varphi = 0.3$  ( $Re = 0, Ri = 0.44, T = 110.90$ ).

CCM and DEC overcome the limitation since the mass of belief of the object be-

longing to the meta-cluster is related not only to the distance from the object to the meta-cluster center but also to the distance between the object and the centers of the singleton clusters included in the meta-cluster. Therefore, in CCM and DEC, only imprecise objects in overlapping areas of different clusters are assigned to appropriate meta-clusters as shown in Fig. 4.4 (c) and (e)-(f). Whereas CCM and DEC are different in some cases, especially the computational complexity of DEC is much lower than that of CCM, which greatly expands the use of DEC especially in big data. From Fig. 4.4, we can see that the DEC take only  $T = 85.31(s)$  and  $T = 110.90(s)$  under different meta-clusters threshold, while ECM, CCM and BPEC take  $T = 4354.21(s)$ ,  $T = 3833.51(s)$  and  $T = 10534.04(s)$  respectively. In DEC, it will lead to a slight increase in the number of objects assigned to meta-clusters with the increase of  $\varphi$  (e.g.  $\varphi = 0.2$  and  $\varphi = 0.3$ ). In applications, we can adjust the parameter  $\varphi$  according to the acceptable imprecision rate.

#### 4.4.4 Real world data

Table 4.2: Basics of the used datasets

Name	#Clus.	#Attr.	#Inst.
Iris	3	4	150
Seeds	3	7	210
Haberman	2	3	306
Wine	3	13	178
Magic	2	10	19020
Contraceptive	3	9	1573
Vehicle	4	18	846
Satimage	7	36	6435
Glass	7	9	214

In this experiment, nine real-world datasets are selected from the UCI repository to test the performance of DEC with respect to ECM, CCM, and BPEC. The basic information of nine datasets including the number of clusters (#Clus.), attributes (#Attr.), and instances (#Inst.) are shown in Table 4.2, and all the detailed information can be found with the link<sup>3</sup>.

For the Iris and Seeds datasets, the outlier thresholds of ECM, CCM, BPEC, and

<sup>3</sup><http://archive.ics.uci.edu/ml/>.

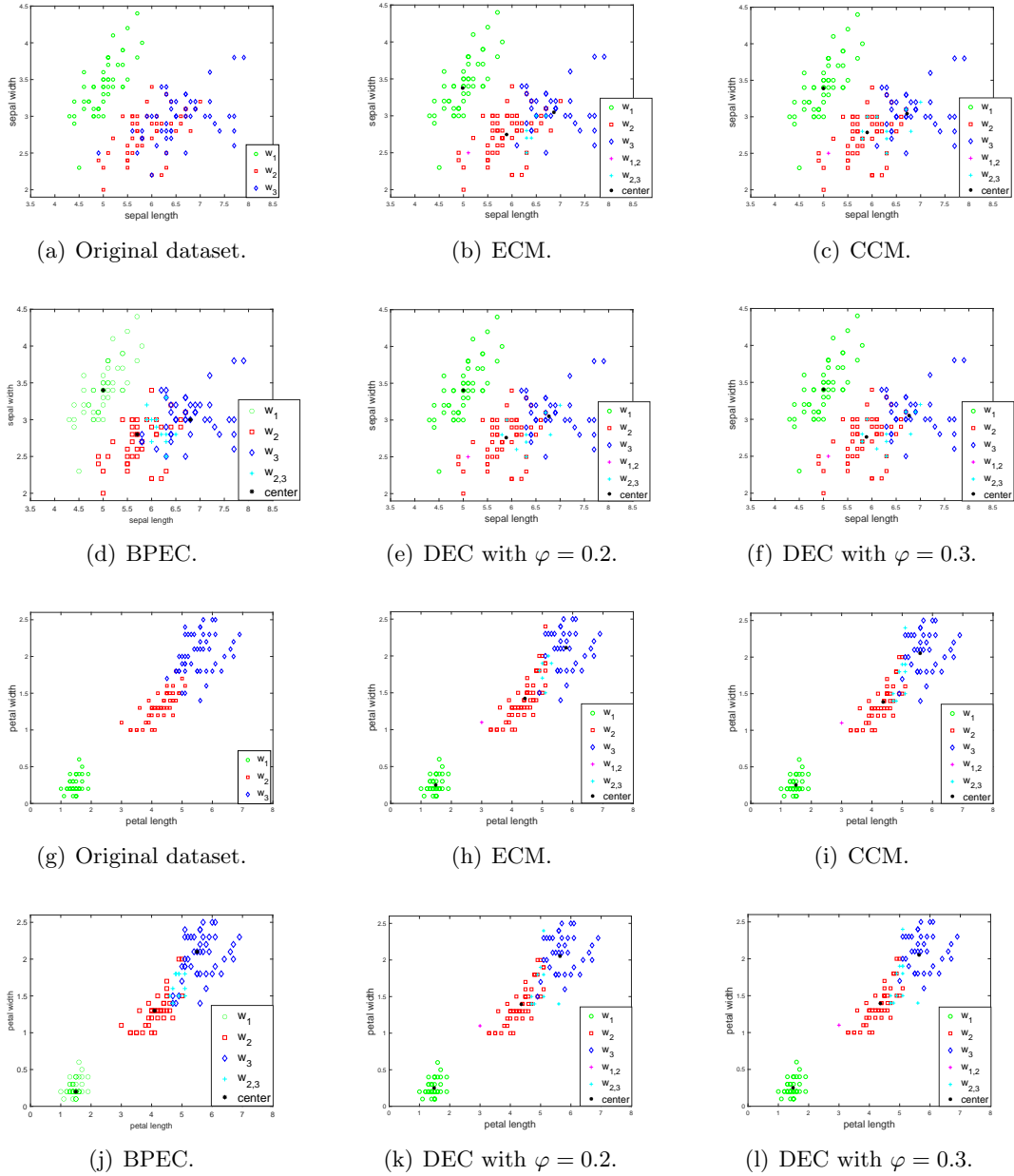


Fig. 4.5. Clustering results of Iris dataset.

DEC are set as  $\delta = 10$ , and the Haberman and Appendicitis datasets with  $\delta = 100$ , while for other datasets, the outlier thresholds are  $\delta = 1000$ . Furthermore, the number of nearest neighbors and quantile of these nearest neighbors are set for BPEC as follows, for the nine datasets from Iris to Appendicitis: (30,0.5), (20,0.9), (20,0.9), (15,0.5), (300,0.9), (115,0.9), (100, 0.9), (500, 0.9) and (5, 0.5). The other parameters in this experiment are default, and the clustering results are shown in Table 4.3.

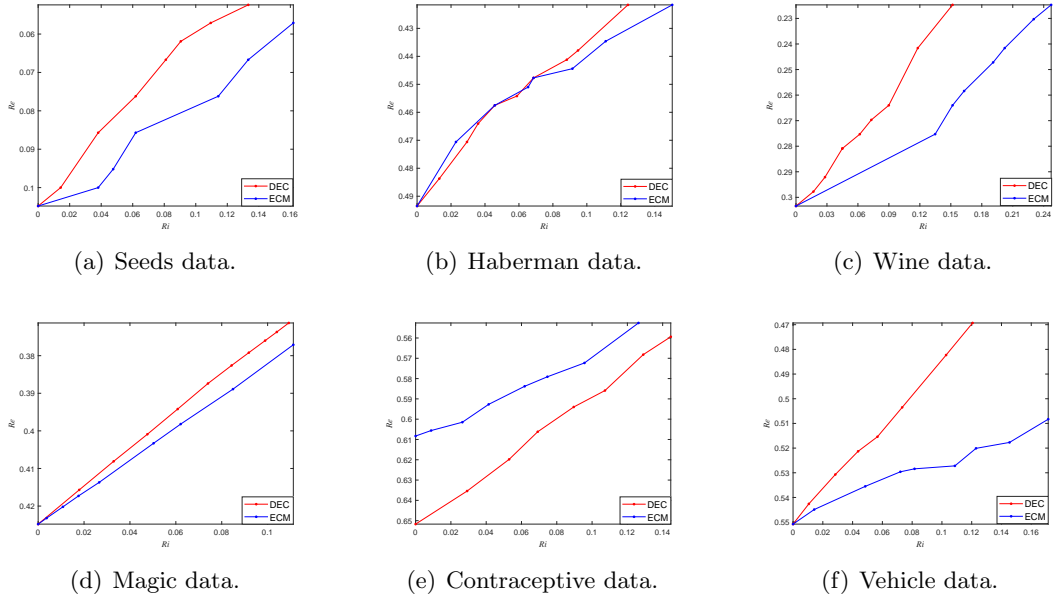


Fig. 4.6. ROC curves for imprecision and error rates of ECM and DEC.

For most of the datasets, CCM and DEC usually obtain similar results, they provide fewer errors than ECM and BPEC. For the Wine dataset, the error rate of ECM and BPEC is lower than that of DEC, whereas its imprecision is much higher, and even some objects are assigned to total ignorant cluster  $\Omega$ . In parallel, DEC has the highest CRI values in most situations, which truly shows the results obtained by the DEC are much more reasonable. The execution time of DEC is much lower than that of ECM, CCM, and BPEC, especially in the case of a large number of instances or clusters, which indicates that the DEC can still ensure good performance.

Furthermore, we visualize the clustering result of Iris real dataset in Fig. 4.5. Four features are measured from each object. They are the length and width of sepal and petal. From the original dataset shown in Fig. 4.5 (a) and (g), we can see that  $\{\omega_1\}$  is significantly distinct from  $\{\omega_2\}$  and  $\{\omega_3\}$ , but  $\{\omega_2\}$  and  $\{\omega_3\}$  are partially overlapped. These four methods can easily separate  $\{\omega_1\}$  from  $\{\omega_2\}$  and  $\{\omega_3\}$ . However, it is difficult to separate  $\{\omega_2\}$  and  $\{\omega_3\}$  in the overlapping area. In ECM, CCM, and BPEC, although several objects in the overlapping area are assigned to the meta-clusters, there are still some objects that are misclassified. In DEC, the error rate of clustering is significantly reduced when the meta-cluster threshold is from  $\varphi = 0.2$  to  $\varphi = 0.3$ . More precisely, the number of objects with error clustering is reduced from 11 to 8, while the number of objects in the meta-clusters is increased from 10 to 13. The experiment shows that some objects in the overlapping area, which are misclassified by other methods, are regarded as imprecise objects and assigned to corresponding meta-clusters in DEC.

Fig. 4.6 shows the ROC curves for imprecision rates and error rates of ECM and DEC based on different datasets, where the  $x$ -axis denotes the mean of the imprecision rate, and the  $y$ -axis denotes the mean of the error rate with scale reversed. From Fig. 4.6, we can find that with the reduction of error rate, the imprecision rate will be incremented correspondingly. Besides, the error rate of DEC, in most cases, is lower than that of ECM when the methods have the same imprecision rate. In practical applications, both ECM and DEC can control the number of objects assigned to the meta-clusters by adjusting the parameters thereby controlling the imprecision rate. Moreover, the value of  $\varphi$  in DEC determines different partition methods. The DEC becomes a fuzzy/possibilistic partition when  $\varphi = 0$ , *i.e.* all objects are assigned to different singleton clusters, and DEC has a traditional credal partition when  $\varphi = 1$ , *i.e.* all objects are imprecise ones, which are assigned to singleton clusters or meta-clusters under the power-set  $2^\Omega$ . The result of DEC also present that the known information does not allow us to accurately classify the objects in the meta-clusters. If we want to get more accurate results, some other (possibly expensive) techniques seem necessary to distinguish and classify these imprecise objects.

In order to verify the effectiveness of the DEC method from different perspectives, Table 4.4 shows the objective function values and number of iterations of all datasets. One can see from Table 4.4 that the number of iterations required for convergence of the DEC method is significantly less than that of ECM and CCM. For some datasets, however, the final objective function value of DEC is greater than that of ECM and CCM. The reason is that the objective function value of an object depends mainly on the mass of belief of its most believed cluster and its distance from the cluster center. We can find that although DEC, ECM, and CCM can assign objects to the most believed singleton cluster, the mass of beliefs for the objects belonging to the most believed cluster in DEC is greater than that of ECM and CCM. This is because most precise objects are directly assigned to the  $c$  singleton clusters in DEC without considering meta-clusters, which will result in a greater mass of beliefs compared to that of ECM and CCM. Thus, the objective function value of DEC is sometimes greater than that of ECM and CCM, especially when the number of clusters is large.

#### 4.4.5 Discussion

The management of parameters and sensitivity analysis are important components for the application of the proposed method. In this subsection, three parameters, *i.e.* the meta-cluster threshold  $\varphi$ , the weighting exponent  $\beta$  and the outlier threshold  $\delta$ , are involved in the proposed DEC method. One can find that the proposed DEC method can deal with fuzzy partition/classical credal partition/dynamic credal partition mode switching by controlling the threshold  $\varphi$ . This can help us to achieve fuzzy partition when imprecise information is not allowed in the results. However, sometimes users are

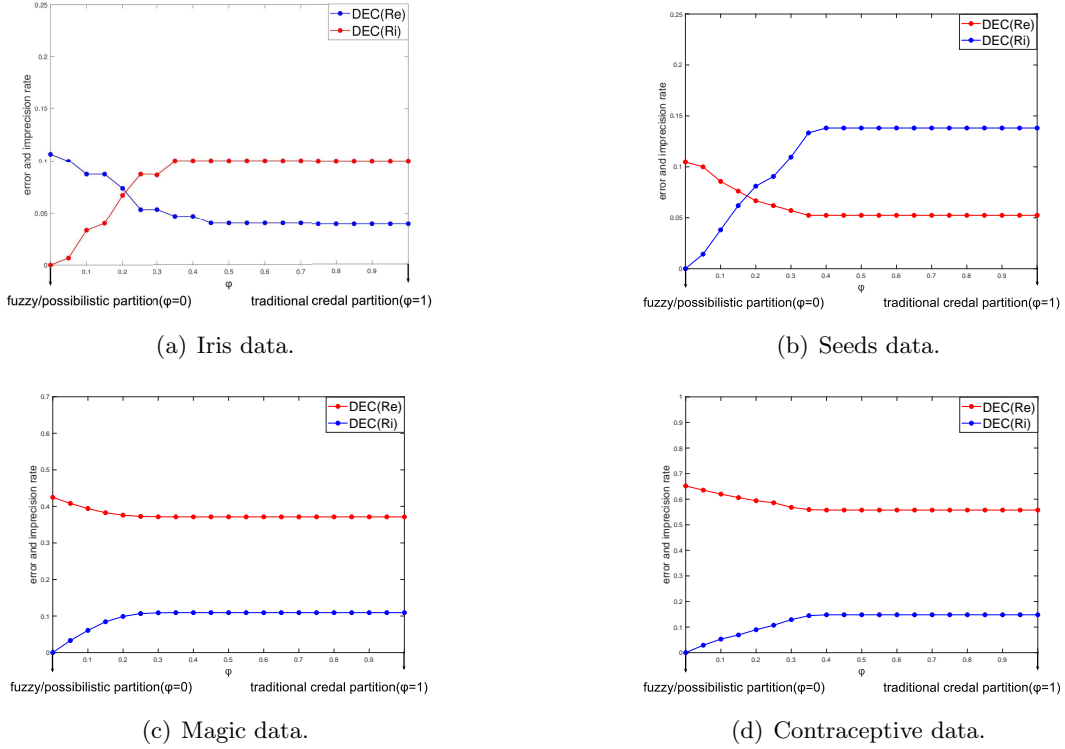


Fig. 4.7. Clustering results of different datasets for various  $\varphi$ .

more willing to obtain reliable imprecise results because the risk of incorrect clustering may be unbearable. In such a case, DEC can switch to dynamic credal partition, which greatly reduces the computational complexity compared to traditional credal partition.

Fig. 4.7 shows the statistical results of the error and imprecision rates based on the Iris, Seeds, Magic, and Contraceptive datasets as  $\varphi$  changes from 0 to 1. It is easy to see that as  $\varphi$  increases the error rate decreases and the inaccuracy rate increases, but this does not last forever. The method has a high sensitivity when  $\varphi = 0.4$ , as measured from the point of view of prudent decision-making, which is consistent with our intuition. This also demonstrates the validity of the method, *i.e.* it is time-consuming and pointless to subject all query objects to the power-set  $2^\Omega$ . However, this does not imply that  $\varphi = 0.4$  is the only value to be taken, as there are also applications where it is desirable to obtain precise clustering results. Therefore, we suggest that  $\varphi \in [0, 0.4]$  and  $\varphi = 0.3$  as the default. Users could set the threshold  $\varphi$  based on an acceptable imprecision rate.

Moreover, the parameters  $\beta$  and  $\delta$  also should be selected in applications. We generally consider the selection of  $\beta$  to be relatively easy because there has been a lot of work on  $\beta$ . In general,  $\beta = 2$  can be taken as the default and is applicable in most cases. However, the selection of  $\delta$  is very difficult because it is highly dependent on the



dataset. We can find that different  $\delta$  values are chosen manually in different datasets, which is a reactive strategy. We generally select  $\delta$  based on the acceptable noise level in applications.

## 4.5 Conclusion

We introduced a new dynamic evidential clustering (DEC) algorithm in this chapter. It is an improved version of traditional evidential clustering, such as ECM and CCM, to reduce the computational complexity of existing methods. It provides a dynamic edited framework  $\mathcal{M}_i^\Omega$  ( $\mathcal{M}_i^\Omega \subset 2^\Omega$ ) including several close singleton clusters and related meta-clusters for each imprecise object to be assigned more reasonably. The DEC overcomes the shortcomings of ECM and CCM and greatly reduces the complexity. In addition, it eliminates the possible negative impact of other meta-clusters on imprecise objects and avoids many invalid computations compared with the framework of the power-set  $2^\Omega$ . The DEC can effectively extend the applications of credal partition, especially in big data, because it reduces the complexity of credal partition to a level similar to that of the fuzzy and possibilistic partition (such as FCM, PCM, and NC). Four experiments with artificial and real datasets have been done to verify the performance of DEC with respect to other methods. It should be noted that in credal partition, once the object is assigned to the meta-cluster, the object may belong to any singleton cluster included in the meta-cluster under the known information, which is also one of the advantages of credal partition – it can reasonably characterize and represent the imprecision and uncertainty caused by a variety of reasons. However, the proposed DEC algorithm in this chapter only improves the execution efficiency of traditional EC methods and does not effectively detect non-spherical clusters. For this reason, we will present two algorithms for detecting imbalanced or even clusters with arbitrary shapes and sizes in subsequent Chapters 5 and 6, respectively.

Table 4.3: Clustering results of different datasets (in %)

Datasets	Indexes	ECM	CCM	BPEC	DEC
Iris	<i>Re</i>	8.00	5.33	6.00	5.33
	<i>Ri</i>	4.67	8.00	8.00	8.67
	<i>CRI</i>	0.8509	0.8631	0.8549	<b>0.9184</b>
	<i>T</i>	0.1600	0.0695	0.2949	<b>0.0050</b>
Seeds	<i>Re</i>	7.62	5.71	7.62	5.71
	<i>Ri</i>	11.43	10.00	10.00	10.95
	<i>CRI</i>	0.7977	0.7912	0.7760	<b>0.8367</b>
	<i>T</i>	0.0754	0.0730	0.3810	<b>0.0075</b>
Haberman	<i>Re</i>	43.46	42.16	39.87	42.16
	<i>Ri</i>	11.11	12.42	15.03	12.42
	<i>CRI</i>	0.6631	<b>0.7064</b>	0.6821	0.5494
	<i>T</i>	0.0912	0.0883	0.2943	<b>0.0104</b>
Wine	<i>Re</i>	17.98	27.53	18.54	26.97
	<i>Ri</i>	32.02	9.55	29.21	7.30
	<i>CRI</i>	0.8207	0.7343	0.8076	<b>0.8814</b>
	<i>T</i>	0.5020	0.3813	0.9520	<b>0.0086</b>
Magic	<i>Re</i>	43.32	39.33	40.85	37.15
	<i>Ri</i>	0.37	8.25	18.05	10.88
	<i>CRI</i>	0.6245	<b>0.7421</b>	0.6987	0.5754
	<i>T</i>	9.15	9.05	34.66	<b>0.6273</b>
Contraceptive	<i>Re</i>	57.91	56.48	54.38	56.82
	<i>Ri</i>	7.49	8.35	12.90	12.90
	<i>CRI</i>	0.6245	0.6419	0.6519	<b>0.7712</b>
	<i>T</i>	2.63	1.53	4.29	<b>0.0600</b>
Vehicle	<i>Re</i>	52.84	51.73	54.14	48.23
	<i>Ri</i>	8.16	6.03	6.38	10.28
	<i>CRI</i>	0.7490	<b>0.7973</b>	0.7558	0.7251
	<i>T</i>	2.77	1.69	3.92	<b>0.0626</b>
Satimage	<i>Re</i>	28.08	30.54	31.45	24.07
	<i>Ri</i>	5.44	8.66	4.83	12.56
	<i>CRI</i>	0.7279	0.7523	0.7101	<b>0.8482</b>
	<i>T</i>	23.72	18.17	44.58	<b>1.70</b>
Glass	<i>Re</i>	48.13	47.66	49.07	46.26
	<i>Ri</i>	61 2.34	6.07	2.34	13.08
	<i>CRI</i>	0.6470	0.7007	0.6411	<b>0.7805</b>
	<i>T</i>	0.6993	0.3673	2.15	<b>0.0678</b>

Table 4.4: Clustering results of different datasets

Datasets	Methods	Objective function values	Number of iterations
Iris	ECM	47.57	33
	CCM	42.44	18
	DEC	56.91	15
Seeds	ECM	294.99	27
	CCM	263.81	24
	DEC	371.24	16
Haberman	ECM	18071.77	20
	CCM	15519.06	23
	DEC	18860.38	17
Wine	ECM	796031.24	75
	CCM	1177975.85	86
	DEC	1647084.50	40
Magic	ECM	115788607.23	40
	CCM	91457280.53	66
	DEC	109617350.83	38
Contraceptive	ECM	12886.81	67
	CCM	11627.91	33
	DEC	16199.99	26
Vehicle	ECM	1460608.50	97
	CCM	1366180.72	87
	DEC	2182377.66	56
Satimage	ECM	4859854.19	140
	CCM	3120338.23	136
	DEC	5325419.18	63
Glass	ECM	96.97	31
	CCM	56.30	31
	DEC	76.45	29

# 5

## Belief shift clustering algorithm

**Abstract:** In this chapter, we present a belief shift clustering algorithm. The purpose is to characterize the uncertainty and imprecision between imbalanced clusters (clusters with different sizes). We first review the benchmark algorithm, mean shift. After that, we present in detail the implementation steps of the proposed algorithm. Then, the effectiveness of the proposed algorithm is demonstrated experimentally on several datasets. Finally, we conclude the chapter.

### Contents

---

<b>5.1</b>	<b>Introduction</b>	<b>64</b>
<b>5.2</b>	<b>Review of mean-shift</b>	<b>64</b>
<b>5.3</b>	<b>Belief shift clustering (BSC)</b>	<b>65</b>
5.3.1	Belief shift for preliminary credal partition	65
5.3.2	Evidential clustering rule for credal redistribution	69
5.3.3	The involved parameters	73
<b>5.4</b>	<b>Experiments</b>	<b>73</b>
5.4.1	Synthetic data	74
5.4.2	Imbalanced data	75
5.4.3	Medical and natural data	77
5.4.4	Some real world data	78
5.4.5	Olivetti Face data	81
5.4.6	Discussion	81
<b>5.5</b>	<b>Conclusion</b>	<b>84</b>

---

## 5.1 Introduction

In the previous chapter, we proposed a dynamic evidential clustering (DEC) algorithm (Zhang et al., 2021d) to reduce the computational complexity of traditional ones, but they are only applicable to spherical data. In this chapter, we work on detecting cluster structures of imbalanced data and characterizing the uncertainty and imprecision between clusters. To drive this goal, we combine traditional evidential clustering (EC) and mean-shift techniques. The main content of this chapter thus consists of three parts. We briefly review the classical mean-shift algorithm in Section 5.2. Afterward, a belief shift clustering algorithm based on the TBF is proposed in Section 5.3 and then verified in Section 5.4. Section 5.5 concludes this chapter.

## 5.2 Review of mean-shift

Mean shift, a simple nonparametric iterative procedure that shifts each object to the average of the objects included in its neighborhood, is introduced by Fukunaga and Hostetler (Fukunaga and Hostetler, 1975; Cheng, 1995). The generalized version of mean shift procedure is briefly introduced as follow.

Let us consider a dataset  $\mathcal{X}$  including  $n$  objects with  $s$  attributes over the FoD  $\Omega = \{\omega_1, \dots, \omega_c\}$ . For the object  $\mathbf{o}_i$  in the  $s$ -dimensional space, if it is the initial cluster center (mode), then the new cluster center  $\mathbf{o}_{i+1}$  which  $\mathbf{o}_i$  shifts to is given by:

$$\mathbf{o}_{i+1} = \frac{\sum_{\mathbf{x}_j \in \mathcal{X}} K(\mathbf{x}_j - \mathbf{o}_i) \mathbf{x}_j}{\sum_{\mathbf{x}_j \in \mathcal{X}} K(\mathbf{x}_j - \mathbf{o}_i)} \quad (5.1)$$

with

$$K(\mathbf{x}_j - \mathbf{o}_i) = \begin{cases} 1, & \text{if } d_{jo} \leq h \\ 0, & \text{if } d_{jo} > h \end{cases} \quad (5.2)$$

where  $d_{jo}$  is the Euclidean distance between the object  $\mathbf{x}_j$  and the object  $\mathbf{o}_i$ , and  $h$  is called the bandwidth.

The function  $K(\mathbf{x}_j - \mathbf{o}_i)$  is the *unit flat kernel* here and it can also be other kernels (*e.g.* Gaussian kernel (Cheng, 1995)). The cluster center  $\mathbf{o}_i$  shifts to the new cluster center  $\mathbf{o}_{i+1}$ , denoted as  $\mathbf{o}_i \leftarrow \mathbf{o}_{i+1}$ , according to the mean shift vector, named  $\mathbf{m}(\mathbf{o}_i)$ , can be concluded by  $\mathbf{m}(\mathbf{o}_i) = \mathbf{o}_{i+1} - \mathbf{o}_i$ . The mean shift algorithm repeatedly updates the cluster center  $\mathbf{o}_i$  using Eqs. (5.1)–(5.2) until the deviation between the estimations of two consecutive cluster centers is less than the value of threshold  $\varepsilon$ , *i.e.*  $\|\mathbf{o}_{i+1} - \mathbf{o}_i\| < \varepsilon$ . In the process of iterations, each object will be searched by other ones that coverage to one or more clusters. That is, the object is searched by one or multiple clusters, and it will be assigned to the cluster that searches it the most times.

### 5.3 Belief shift clustering (BSC)

In this section, we propose a new belief shift clustering (BSC) algorithm based on the notion of “belief shift”, which can be considered as an evidential version of mean shift (mode seeking) under the TBF. The BSC mainly consists of two parts.

1) Belief shift for preliminary credal partition. It mainly assigns each object as a noise (outlier) that far away from other objects, or a precise object belonging to a specific cluster, or an imprecise object that may belong to several possible clusters;

2) Evidential clustering rule for partial credal redistribution. It can further assign the imprecise objects to different clusters including singleton clusters or meta-clusters composed of several singleton clusters that the object likely belongs to.

These two parts will be discussed in Subsections 5.3.1 and 5.3.2, respectively.

#### 5.3.1 Belief shift for preliminary credal partition

Let us consider a dataset  $\mathcal{X}$  including  $n$  objects with  $s$  attributes over the FoD  $\Omega = \{\omega_1, \dots, \omega_c\}$ . Belief shift for preliminary credal partition can be detailed as follows. For a specific object  $\mathbf{x}_i$  ( $i = 1, \dots, n$ ), the neighbors can be obtained by using the Euclidean distances between  $\mathbf{x}_i$  and the others, defined by:

$$d_{ij} = \sqrt{\sum_{q=1}^s (x_{iq} - x_{jq})^2} \quad (5.3)$$

where  $\mathbf{x}_j \in \mathcal{X}$  and  $\mathbf{x}_j \neq \mathbf{x}_i$ .

The distances are ranked from small to large, and the  $K_1$  neighbors,  $\mathbf{y}_1, \dots, \mathbf{y}_{K_1}$ , with the corresponding minimum distances can be obtained. These neighbors are selected from the entire dataset and each neighbor provides a piece of evidence represented by a mass function  $m_{ik}(\cdot)$  for the object  $\mathbf{x}_i$  being a cluster center in a new frame of discernment  $\tilde{\Omega} = \{\mathcal{C}, \bar{\mathcal{C}}\}$ , which is defined to describe the belief degree of the object as the cluster center ( $\mathcal{C}$ ) or unknown ( $\bar{\mathcal{C}}$ ). The basic principle is that cluster centers are most likely to be distributed in areas with the highest global (local) density. Thus, if an object is very close to all its neighbors, it has the potential to become a cluster center. Inspired by the pieces of literature (Denœux, 2008b; Liu et al., 2013), we define the mass function  $m_{ik}(\cdot)$  on  $\tilde{\Omega}$  as follows:

$$m_{ik}(B) = \begin{cases} \frac{1}{K_1} \cdot e^{-d_{ik}^2}, & \text{if } B = \mathcal{C} \\ 1 - \frac{1}{K_1} \cdot e^{-d_{ik}^2}, & \text{if } B = \bar{\mathcal{C}} \end{cases} \quad (5.4)$$

where  $d_{ik}$  is the Euclidean distance between the object  $\mathbf{x}_i$  and the object  $\mathbf{y}_k$ . The mass function  $m_{ik}(\mathcal{C})$ , for example, can be regarded as the support degree that the object

$\mathbf{y}_k$  believes the object  $\mathbf{x}_i$  as a cluster center. We can see from Eq. (5.4) that the nearer the neighbor to  $\mathbf{x}_i$ , the larger the  $m_{ik}(\mathcal{C})$  obtained, *i.e.* this neighbor strongly supports it as a cluster center.

Afterward, the  $K_1$  pieces of evidence can be fused by the DS rule to obtain the degree of belief  $m_i(\cdot)$  that the object  $\mathbf{x}_i$  is a cluster center, calculated by:

$$m_i(A) = \bigoplus_{k \in [1, K_1]} m_{ik}(A) \quad (5.5)$$

where  $\bigoplus$  represents the DS rule operation symbol<sup>4</sup>. We can get  $m_i(A) = Bel_i(A)$  by deriving the Eq. (5.5), and the proof process is similar to that of the literature (Su and Dencoux, 2018). Thus, the degree of belief  $Bel_i(A)$  can be further defined by:

$$Bel_i(A) = \begin{cases} 1 - \prod_{k=1}^{K_1} (1 - m_{ik}(\mathcal{C})) , & \text{if } A = \mathcal{C}; \\ \prod_{k=1}^{K_1} m_{ik}(\bar{\mathcal{C}}) , & \text{if } A = \bar{\mathcal{C}}. \end{cases} \quad (5.6)$$

By doing this, we can calculate the belief degrees of all the  $n$  objects, and each of them then is regarded as the initial cluster center for belief shift<sup>5</sup>, similar to mean shift (Cheng, 1995). In the belief shift process, for the specific object  $\mathbf{x}_i$  ( $i = 1, \dots, n$ ), as the initial cluster center, it will shift to the new center (object), called  $\mathbf{x}_\mu$ , corresponding to the neighbor  $\mathbf{y}_k$  with the highest belief degree, and defined by:

$$\mathbf{x}_\mu = \arg \max \{ Bel_{i1}(\mathcal{C}), \dots, Bel_{iK_1}(\mathcal{C}) \} \quad (5.7)$$

where  $Bel_{ik}(\mathcal{C})$  ( $k = 1, \dots, K_1$ ) represents the belief degree (the possibility) of the  $k$ -th neighbor  $\mathbf{y}_k$  to become a cluster center.

Then, the new cluster center  $\mathbf{x}_\mu$  keeps shifting until the belief degree  $Bel_\mu(\mathcal{C})$  is higher than that of its new neighbors, and defined by:

$$Bel_\mu(\mathcal{C}) \geq \max \{ Bel_{\mu 1}(\mathcal{C}), \dots, Bel_{\mu K_1}(\mathcal{C}) \} \quad (5.8)$$

By doing this, we can obtain the final cluster center  $\mathbf{x}_\mu$  that  $\mathbf{x}_i$  converges to. After all the objects as the initial cluster center to shift, they will eventually converge to  $c$  different cluster centers, *i.e.*  $c$  different real objects, with the highest belief degree

<sup>4</sup>Since these  $K_1$  pieces of evidence come from neighbors, they are generally no conflict or low conflict. In this case, compared with other combination rules (Dubois and Prade, 1988; Martin, 2019), the DS rule is more suitable, which is commutative and associative.

<sup>5</sup>In the process of belief shift, each object will be selected as initial cluster center to eliminate the negative impact of random selection, and the final cluster centers are all from the real objects since the advantage of such prototype based clustering method is their intuitive understanding by prototypes as representatives for the clusters (Zhou et al., 2015c; Hariz et al., 2007; Haykin and Network, 2004).

compared with the corresponding neighbors. We can easily find that although the object  $\mathbf{x}_i$  will converge to a specific (singleton) cluster, it may also be searched by other singleton clusters as the neighbor of different query objects. In this case, for an object  $\mathbf{x}_i$ , we will temporarily divide it as an imprecise object. To more intuitively show the principle of belief shift and explain its difference from the classical mean shift, we give the following example.

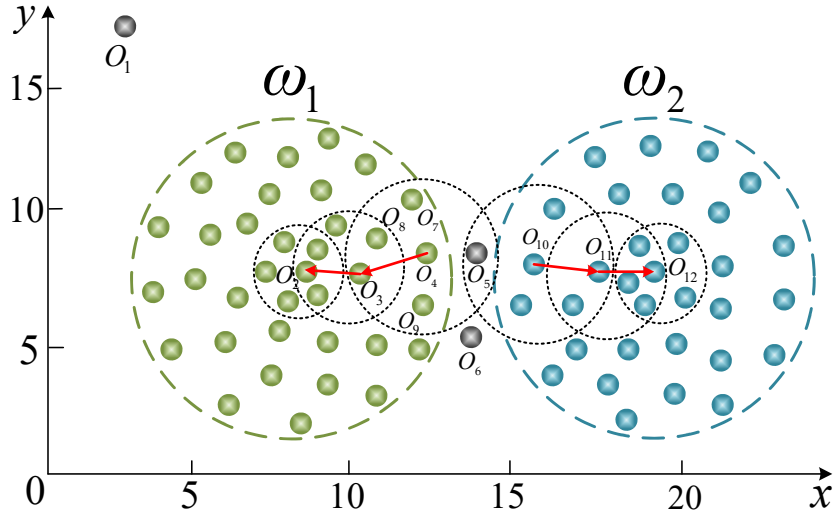


Fig. 5.1. Illustration of the belief shift process.

As a simple example, Fig. 5.1 shows a 2-class problem with two attributes corresponding to  $x$ -coordinate and  $y$ -coordinate. It is assumed that the objects  $O_2$  and  $O_{12}$  have the highest degree of beliefs, *i.e.*  $Bel_2(\mathcal{C})$  and  $Bel_{12}(\mathcal{C})$ , to be the centers of the clusters  $\{\omega_1\}$  and  $\{\omega_2\}$ , respectively, and the number of neighbors  $K_1 = 5$  for each query object is the default in the belief shift process. More intuitively, the  $K_1$  neighbors of the object are in the range of black dotted line and the trajectory of belief shift is marked as a red line with an arrow in Fig. 5.1. The object  $O_4$ , for example, as the initial cluster center will find the  $K_1$  neighbors (*i.e.*  $O_3, O_5, O_7, O_8, O_9$ ), and then shifts to the neighbor  $O_3$  with the highest belief degree, and finally converges to the cluster center  $O_2$ , *i.e.*  $O_4 \in \{\omega_1\}$ . Of course, the object  $O_{10}$  will converge to the cluster center  $O_{12}$ , *i.e.*  $O_{10} \in \{\omega_2\}$ . For objects like  $O_4$  and  $O_{10}$ , both classical mean shift or belief shift will directly assign them to specific clusters like  $\{\omega_1\}$  and  $\{\omega_2\}$ , which is the similarity between the methods.

For the object  $O_5$ , however, the classical mean shift will also directly assign it to the cluster  $\{\omega_1\}$  or  $\{\omega_2\}$  depending only on the times of the object searched by these two clusters, *i.e.* the cluster with the most search times will own the object. This may exist a high risk of error, since the search times are related to the randomly initialized cluster centers in mean shift and depending only on the search times is unreasonable.



Different from mean shift, the proposed belief shift considers that once the object like  $O_5$  is searched as a neighbor by different clusters, it indicates that the object may be distributed in the overlapping or middle areas of these clusters<sup>6</sup>. We can also get a similar inference from the object  $O_6$  if interesting. In this case, it is unwise to assign the object  $O_5$  (or  $O_6$ ) to the singleton cluster  $\{\omega_1\}$  or  $\{\omega_2\}$ , since it may increase error risks. Thus, they are temporarily assigned to the specific edited framework  $\mathcal{M}_i = \{\omega_1, \omega_2\}$ ,  $\mathcal{M}_i \subseteq \Omega$  and  $i = 5, 6$ , as imprecise objects to wait for credal redistribution.

By contrast, we can also infer that although the object  $O_1$  as the initial center can converge to the cluster  $\{\omega_1\}$ , it will not be searched, as a neighbor of other query objects, by different clusters because it is too far away from all objects except itself. In such a case, the object  $O_1$  is more suitable to be assigned to the noise cluster, *i.e.* the object  $O_1$  is regarded as a noise (outlier). The classical mean shift, however, does not have the ability to deal with such a case.

Based on the above analysis, the query object  $\mathbf{x}_i$  will have two general indexes after belief shift, *i.e.* the number of different clusters that search the object, denoted as  $|\mathcal{M}_i|$ , and the number of times the object is searched by other ones as a neighbor, named  $\mathcal{T}_i$ . In fact,  $\mathcal{M}_i$  is the corresponding specific dynamic sub-framework that the query object  $\mathbf{x}_i$  will be assigned to, and  $\mathcal{M}_i$  (if  $|\mathcal{M}_i| > 1$ ) is a set of clusters that have searched the object  $\mathbf{x}_i$  in the process. The total number  $\mathcal{T}_i$  of searched times for  $\mathbf{x}_i$  is defined by:

$$\mathcal{T}_i = \sum_{\{\omega_j\} \in \mathcal{M}_i} \mathcal{T}_i^j \quad (5.9)$$

where  $\mathcal{T}_i^j$  represents the number of times that the object  $\mathbf{x}_i$ , as the neighbor(s), is searched by the other objects that converge to cluster  $\{\omega_j\}$ . As we analyzed earlier, if  $\mathbf{x}_i$  is searched significantly less than that of others, it will be assigned directly to the noise cluster, *i.e.*  $\emptyset$ , defined by:

$$\emptyset = \{\mathbf{x}_i \mid \mathcal{T}_i \leq \lceil \bar{\mathcal{T}}^\alpha \rceil\} \quad (5.10)$$

with

$$\bar{\mathcal{T}} = \frac{1}{n} \sum_{i=1}^n \mathcal{T}_i \quad (5.11)$$

where  $\lceil \cdot \rceil$  is the rounding symbol and  $\bar{\mathcal{T}}$  is the average of  $\mathcal{T}_i$  for all objects in the query set. The parameter  $\alpha$  is the outlier adjustment factor, which controls the number of objects assigned to the noise cluster, *i.e.*  $\emptyset$ .

The flowchart of preliminary adaptive credal partition by belief shift is shown in Fig. 5.2. We can see from the flowchart that the query object  $\mathbf{x}_i$  will be regarded as the

<sup>6</sup>At this time, the belief shift will not use the simple search times as the indicator to assign the object  $O_5$  to a specific cluster.

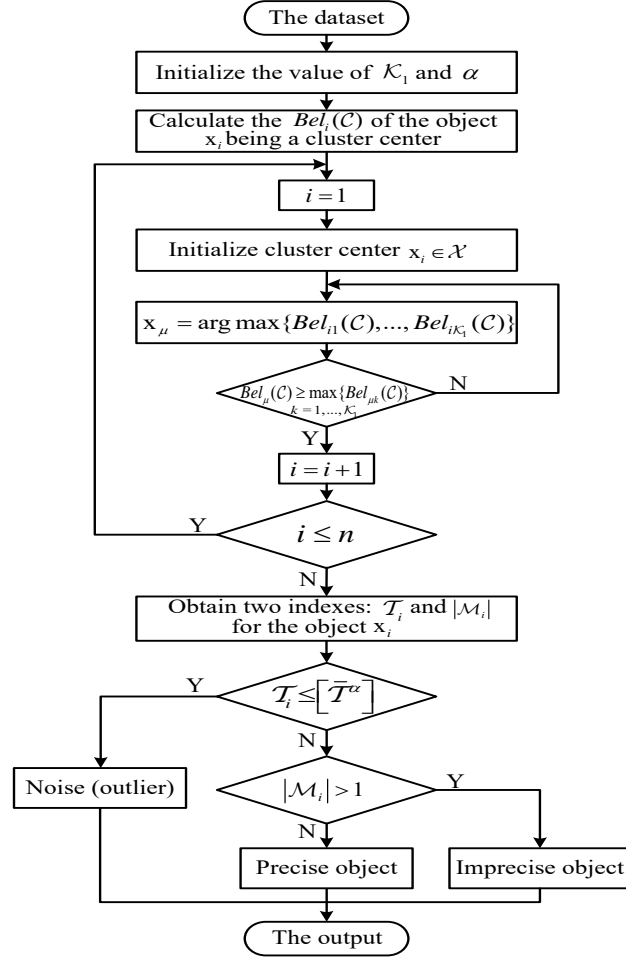


Fig. 5.2. Illustration of preliminary adaptive credal partition.

noise (outlier) if it is searched as a neighbor by very few or even no other objects after belief shift, which also means that it is far away from others. If  $|\mathcal{M}_i| = 1$  with  $\mathcal{T}_i > [\bar{\mathcal{T}}^\alpha]$ , it indicates that the object  $\mathbf{x}_i$  is an precise object with definite cluster information and should be assigned to the singleton cluster  $\mathcal{M}_i$ , otherwise, it will be temporarily regarded as an imprecise object and processed in the next part.

### 5.3.2 Evidential clustering rule for credal redistribution

For the imprecise object  $\mathbf{x}_i$ , the masses of belief will be partially redistributed by EC rule to assign  $\mathbf{x}_i$  to different clusters under the corresponding specific dynamic sub-framework  $\mathcal{M}_i$  ( $|\mathcal{M}_i| > 1$ ). We consider that there are  $n_1$  ( $0 < n_1 < n$ ) imprecise object in the set  $\mathcal{X}_{im} = (\mathbf{x}_1, \dots, \mathbf{x}_{n_1}) \in \mathbb{R}^{n_1 \times s}$  and  $n_2$  ( $0 < n_2 < n$ ) precise object in the set  $\mathcal{X}_{pr} = (\mathbf{x}_1, \dots, \mathbf{x}_{n_2}) \in \mathbb{R}^{n_2 \times s}$ ,  $n_1 + n_2 = n$ , after preliminary credal partition based on belief shift. For the query object  $\mathbf{x}_i \in \mathcal{X}_{im}$ , it will be identified under the new edited dynamic framework  $\mathcal{F}_i$  with  $2^{|\mathcal{M}_i|} - 1$  elements and  $\mathcal{F}_i \subset 2^\Omega$  under the TBF. The credal

partition  $m_{ij} \triangleq m_i(A_j) \in \mathbb{R}^{2^{|\mathcal{M}_i|-1}}$  with the  $j$ -th focal element  $A_j$  in  $\mathcal{F}_i$ , *i.e.*  $A_j \in \mathcal{F}_i$ , is provided for each imprecise object  $\mathbf{x}_i$  ( $i = 1, \dots, n_1$ ). For instance, if  $\Omega = \{\omega_1, \omega_2, \dots, \omega_5\}$  with  $|\Omega| = 5$  for the query set and  $\mathcal{M}_i = \{\omega_1, \omega_3, \omega_5\}$  with  $|\mathcal{M}_i| = 3$  for the imprecise object  $\mathbf{x}_i$ , then  $\mathcal{F}_i = \{\{\omega_1\}, \{\omega_3\}, \{\omega_5\}, \{\omega_1, \omega_3\}, \{\omega_1, \omega_5\}, \{\omega_3, \omega_5\}, \mathcal{M}_i\}$ .

In credal partition, the dataset converges in  $c$  clusters and related meta-clusters by alternating iterations of the center matrix and the mass of belief matrix (Liu et al., 2015; Zhou et al., 2015c). Although the real (final) cluster centers have been obtained by belief shift, they cannot be used directly to iterate the masses of belief for imprecise objects, since it is unreasonable to use one fixed center to represent the cluster with arbitrary shape and size. For imbalanced clusters (Liang et al., 2012), for example, these centers based on belief shift tend to assign the objects of the majority clusters into the minority clusters. Thus, the BSC will provide specific simulation centers for related clusters supervised by  $\mathcal{F}_i$  when clustering the imprecise object  $\mathbf{x}_i$ , which can be explained by a concrete example, as shown in Fig. 5.3, with a 2-class dataset.

In Fig. 5.3, the black pentagrams represent the cluster centers of  $\{\omega_1\}$  and  $\{\omega_2\}$  after belief shift, and the objects specifically assigned to different clusters are precise objects while the objects  $O_1, O_2$  and  $O_3$  are imprecise objects. One can see from Fig. 5.3 that the center of meta-cluster  $\{\omega_1, \omega_2\}$  marked as purple pentagram is obtained by calculating the mean values of  $\{\omega_1\}$  and  $\{\omega_2\}$  in credal partition (Liu et al., 2015; Zhou et al., 2015c). However, the center is located in the cluster  $\{\omega_1\}$ , which is obviously unreasonable and runs counter to our intuitive perception since the cluster center of meta-cluster should be located in the overlapped area of the different singleton clusters or midpoint of the edges of the different clusters. Those objects like  $O_4$  that originally belongs to the cluster  $\{\omega_1\}$  will be assigned to the meta-cluster  $\{\omega_1, \omega_2\}$  and the objects like  $O_1$  will be assigned to the cluster  $\{\omega_2\}$  if we use the meta-cluster center to assign them directly in this case.

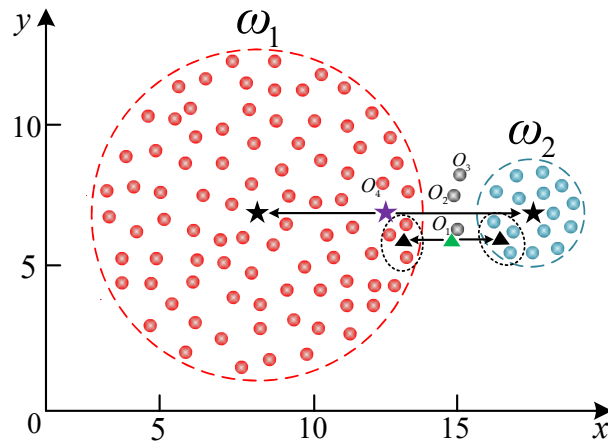


Fig. 5.3. Illustration of the simulated cluster centers based on imbalanced data.

To address this problem, the simulated cluster centers based on KNNs technology

are used here to partial assign the imprecise objects. Specifically, the object  $\mathbf{x}_i$  will find  $K_2$  neighbors from the different clusters contained in  $\mathcal{M}_i$  to simulate the corresponding centers of the singleton clusters by calculating the mean value of the neighbors, respectively. Then, the simulated singleton cluster centers are employed to calculate the related meta-cluster centers, *i.e.* we adaptively provide particular centers of clusters included in  $\mathcal{F}_i$  for each imprecise object  $\mathbf{x}_i$  according to its distribution. For example, the imprecise object  $O_1$  finds different neighbors included in black dashed circle form precise objects in clusters  $\{\omega_1\}$  and  $\{\omega_2\}$  as shown in Fig. 5.3, respectively. The black triangle represents the simulated singleton cluster center obtained from the average value of neighbors, and the midpoint marked as the green triangle between the two cluster centers is the meta-cluster center for  $O_1$ . We can intuitively see that the meta-cluster center obtained by simulated centers is located at the halfway of  $\{\omega_1\}$  and  $\{\omega_2\}$ , which is more reasonable than the center marked as the purple pentagram.

On the basis of the above analysis, the singleton cluster center  $\mathbf{v}_{il}$  is given by:

$$\mathbf{v}_{il} = \frac{1}{K_2} \sum_{k=1}^{K_2} \mathbf{y}_k^l, \quad l = 1, \dots, c \quad (5.12)$$

where  $\mathbf{y}_k^l$  ( $k = 1, \dots, K_2$ ) represents the  $k$ -th neighbor of  $\mathbf{x}_i$  from the cluster  $\{\omega_l\} \subset \mathcal{X}_{pr}$ . The method calculating meta-cluster centers is the same way as ECM (Masson and Dencoux, 2008). That is, the center of meta-cluster  $A_j$  with  $|A_j| > 1$  is the mean of the singleton cluster centers included in  $A_j$ , which has been defined in Eq. (3.23).

Inspired by CCM (Liu et al., 2015), we can thereby use the given EC rule with the simulated center matrix  $V$  to update the masses of belief for the object  $\mathbf{x}_i$ , defined by:

$$m_{ij} = \frac{D_{ij}^{-2/(\beta-1)}}{\sum_{j|A_j \in \mathcal{F}_i} D_{ij}^{-2/(\beta-1)}} \quad (5.13)$$

subject to

$$\mathcal{J}_{BSC}(M, V) = \sum_{i=1}^{n_1} \sum_{j|A_j \in \mathcal{F}_i} m_{ij}^\beta D_{ij}^2 \quad (5.14)$$

and

$$D_{ij}^2 = \begin{cases} d_{ij}^2, & \text{if } |A_j| = 1 \\ \frac{\sum_{\{\omega_l\} \in A_j} \gamma^{-1} d_{il}^2 + d_{ij}^2}{|A_j| + 1}, & \text{if } |A_j| > 1 \end{cases} \quad (5.15)$$

where  $d_{ij}$  is the Euclidean distance between  $\mathbf{x}_i$  and the meta-cluster ( $A_j$ ) center, and  $d_{il}$  represents the distances from  $\mathbf{x}_i$  to the singleton cluster ( $\{\omega_l\}$ ) center in meta-cluster  $A_j$  such that  $|A_j| > 1$ , and  $n_1$  is the number of imprecise objects. The tuning parameter

$\beta$ , such that  $\beta > 1$ , is a weighting exponent (Masson and Denœux, 2008) and  $\gamma$  is the threshold to control the number of objects in meta-clusters (Liu et al., 2015). One can find that the distance between the object  $\mathbf{x}_i$  and the meta-cluster  $A_j$  with  $|A_j| > 1$  depends not only on the distance from  $\mathbf{x}_i$  to the center of  $A_j$ , but also on the distance between  $\mathbf{x}_i$  and the centers of all the singleton clusters included in  $A_j$ , which is similar to CCM and has been detailed explained in CCM.

However, different from CCM, we do not consider noise clustering here because we have solved this problem well in the preliminary credal partition. Since the simulated cluster centers are reliable, only one update is needed to produce the masses of beliefs that each imprecise object  $\mathbf{x}_i$  belongs to different clusters, which can decrease the computation brought by the iterative process while ensuring the rationality. By doing this, each imprecise object in the set  $\mathcal{X}_{im}$  is partial credal redistributed again based on EC. In the process, some objects will be assigned to singleton clusters, which means that these objects are precise ones with definite cluster information. By contrast, the others will be assigned to related meta-clusters to model the imprecision of their cluster information. This prudent decision-making method can well characterize the uncertainty and imprecision between different clusters with arbitrary shapes and sizes in the space, and this may be very important in some applications.

The pseudo-code is presented in **Algorithm 2** to clearly show how BSC works and illustrate its basic principle.

---

**Algorithm 2** Belief shift clustering.

---

**Require:** Dataset:  $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ ; Given the parameters:  $K_1, K_2, \alpha, \beta, \gamma$ .

**Ensure:** Cluster decision results.

**Step 1**

Search the neighbors for all objects using Eq. (5.3);

Calculate  $(Bel_i(C))$  for all objects using Eqs. (5.4)–(5.6);

**for**  $i = 1$  to  $n$

**repeat**

Each object is employed to belief shift using Eq. (5.7);

**until** Satisfy the judgment condition of Eq. (5.8).

**end**

Assign the outlier using Eqs. (5.9)–(5.11);

Assign precise and imprecise objects using  $|\mathcal{M}_i|$  and  $\mathcal{T}_i$ ;

**Step 2**

**for**  $i = 1$  to  $n_1$

Calculate simulated cluster centers using Eq. (5.12);

Obtain meta-cluster centers based on that of simulated singleton clusters;

Reassign the imprecise object again using Eq. (5.13).

**end**

**Return:** Output the results.

---

### 5.3.3 The involved parameters

In BSC, some parameters including  $K_1, K_2, \alpha, \beta, \gamma$  play a very important role, and they should be selected in advance to implement the proposed BSC method.  $K_1$  is the number of not only the neighbors that are used to provide the pieces of evidence for the object being a cluster, but also the neighbors that the objects are looking for in the process of belief shift. The value of  $K_1$  should not be too small since it may cause the object to fall into the local maximum belief degree during belief shift. Whereas some clusters with very close data distribution may not be able to correctly distinguish if too large  $K_1$  value is set. Here the value of  $K_1$  is determined by the number  $n$  of objects in the dataset, and we find that  $K_1 \in [0.05n, 0.15n]$  can be used as the default in most cases according to the experience.  $K_2$  is the number of neighbors that are used to get simulated cluster centers of different clusters, and it does not need to take too large and it is an open value. Thus, we recommend a common default value, *i.e.*  $K_2 = 7$ . The parameter  $\alpha$  is the outlier adjustment factor, which controls the number of objects regarded as the noise. In general, the bigger  $\alpha$  causes the more objects assigned to the noise cluster, and we recommend that  $\alpha \in [0, 0.6]$  and take  $\alpha = 0.3$  in applications. The use of the tuning parameter  $\beta$  is similar to FCM (Bezdek, 2013), ECM (Masson and Dencœux, 2008), and  $\beta = 2$  can be used as default value. The weighting factor  $\gamma$  can be used to control the number of objects in meta-clusters. The smaller  $\gamma$  is, the fewer objects are assigned to the meta-clusters, which will increase the number of misclassified objects. Whereas,  $\gamma$  is not the bigger the better since a big value of  $\gamma$  will lead to high imprecision and is not conducive to make decisions. Therefore, the selection of  $\gamma$  should be based on the imprecision rate one can accept, and we recommend  $\gamma \in [0.5, 2.5]$ , similar to CCM (Liu et al., 2015), and take  $\gamma = 1$ .

## 5.4 Experiments

Five experiments have been done to evaluate the performance of the proposed BSC method with respect to MC (Liang et al., 2012), DPC (Rodriguez and Laio, 2014), MS-type (Yamasaki and Tanaka, 2019), and BPEC (Su and Dencœux, 2018). Subsections 5.4.1 and 5.4.2, based on particular synthetic data, are used to illustrate the use of BSC and the limitations of other methods. Subsections 5.4.3 and 5.4.4 with real images is presented to evaluate the effectiveness of BSC compared with other methods. Subsection 5.4.5 reveals the potential of BSC in image classification with face datasets.

All parameters are default except the ones we adjust in the sequel experiments so as to make the comparison experiment intuitive. The error rate  $R_e$ , the impression rate  $R_i$ , and the Credal Rand Index  $CRI$  are used as indexes of the performances of different methods (Liu et al., 2015; Masson and Dencœux, 2008; Dencœux et al., 2017).

### 5.4.1 Synthetic data

In this experiment, a synthetic dataset named **SD15** is employed to validate the effectiveness of BSC and reveal the limitations of hard partition methods including MC, DPC, and MS-type in clustering data with high overlap and noise (outlier) in the space. **SD15** can be obtained from (Fränti and Virmajoki, 2006) that contains 5000 data points, and it consists of fifteen classes. In addition, we take three noisy points marked by black dots in the top left corner of **SD15**. The original data points of **SD15** is shown in Fig. 5.4(a) while Fig. 5.4(b) reveals their probability distribution. The data points included in different clusters marked by various colors. All the attributes of **SD15** are normalized into  $[0,1]$  by the min-max rule introduced in (Su and Dencœur, 2018) to eliminate the influence of differences in various dimensions.

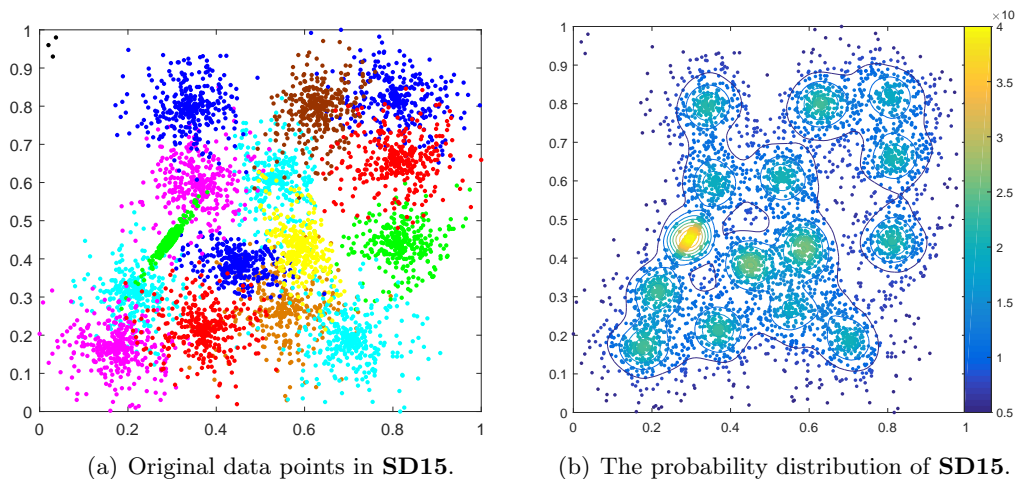


Fig. 5.4. The basic information of the **SD15** dataset.

Here we choose  $f = 0.02$  in DPC and the bandwidth  $r = 0.08$  in MS-type, respectively.  $K_1 = 80$  are taken in BSC. We can see from Fig. 5.4(a) that the different clusters of the dataset **SD15** are partly overlapped on their borders, and the points in these areas are really difficult to be clearly classified, which can be verified by the clustering results of MC, DPC, and MS-type as shown in Fig. 5.5(a)–(c). We can see that the points in the overlapping areas are all assigned to singleton clusters by these methods and most of them are misclassification. It is worth noting that the noisy points marked with black dots are far from the other points, and they cannot be detected by MC and DPC but clustered into specific clusters. Although MS-type yields a singleton cluster for the noisy data, it also does not correctly assign these special points due to the limitations of the probability framework that does not introduce the noise cluster. Fig. 5.5(d) shows the trajectory of the belief shift in BSC, where the trajectories of the objects in the process of belief shift are marked as red lines and they finally coverage to

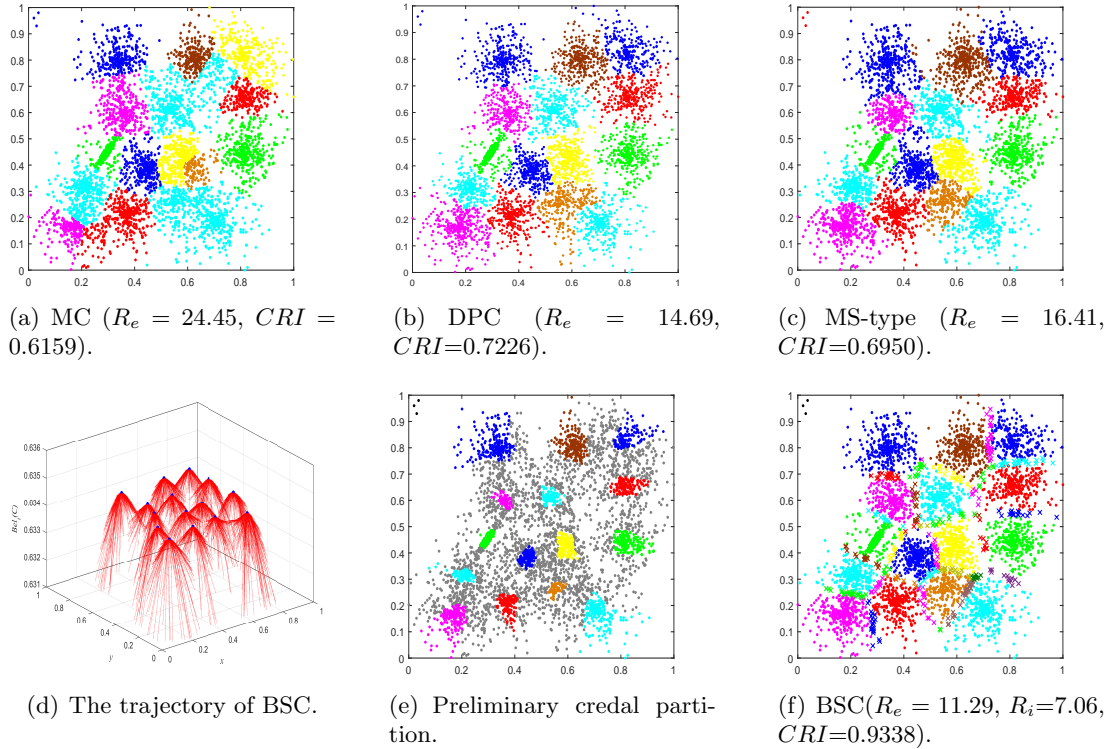


Fig. 5.5. Clustering results of the SD15 dataset by different methods.

the fifteen points with the highest belief degrees marked by blue dots. One can see from Fig. 5.5(e) that BSC preliminary yields credal partitions for the points according to the results of belief shift. The three points marked by black dots are the noisy points that are only searched by few in the belief shift process since they are far away from others. By contrast, the points that are searched multiple times as a neighbor by other ones and searched by only one cluster can be directly assigned to specific clusters. Whereas the points, marked as gray and searched by multiple clusters in the process, are imprecise points and most of them are lying in overlapping areas of different clusters. The final clustering results of BSC as shown in Fig. 5.5(f), where these imprecise points are prudently assigned to the exact singleton clusters and proper meta-clusters marked by points and crosses with different colors, respectively. By doing this, BSC can effectively reduce the errors and reasonably characterize uncertainty between the different clusters since it can provide the credal partition for the imprecise objects under the TBF.

#### 5.4.2 Imbalanced data

This experiment is used to reveal the limitations of traditional EC in dealing with imbalanced clusters, *i.e.* the clusters with imbalanced sizes. We investigate the performance of BSC with respect to BPEC in clustering a 4-class 2-D dataset named **SD4** that con-



tains not only overlapping areas in different clusters but also has nonspherical shapes and imbalanced sizes in the space. The dataset is made up of 3300 data points with two dimensions. The points arise from a mixture of four bivariate Gaussian densities are given in Table 5.1, where  $\mu_i$  is the means vector and  $\Sigma_i$  is the covariance matrices, and  $N_i$ ,  $i = 1, 2, 3, 4$ , represents the number of data points in different clusters.

Table 5.1: The basic information of the **SD4** dataset

Data	Indexes	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_3\}$	$\{\omega_4\}$
	$\mu_i$	$[2.5, 5]^T$	$[7.5, 5]^T$	$[5, 7]^T$	$[5, 1.5]^T$
<b>SD4</b>	$\Sigma_i$	$\begin{pmatrix} 1 & 0 \\ 0 & 0.05 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 0.05 \end{pmatrix}$	$\begin{pmatrix} 0.05 & 0 \\ 0 & 0.5 \end{pmatrix}$	$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$
	$N_i$	500	500	300	2000

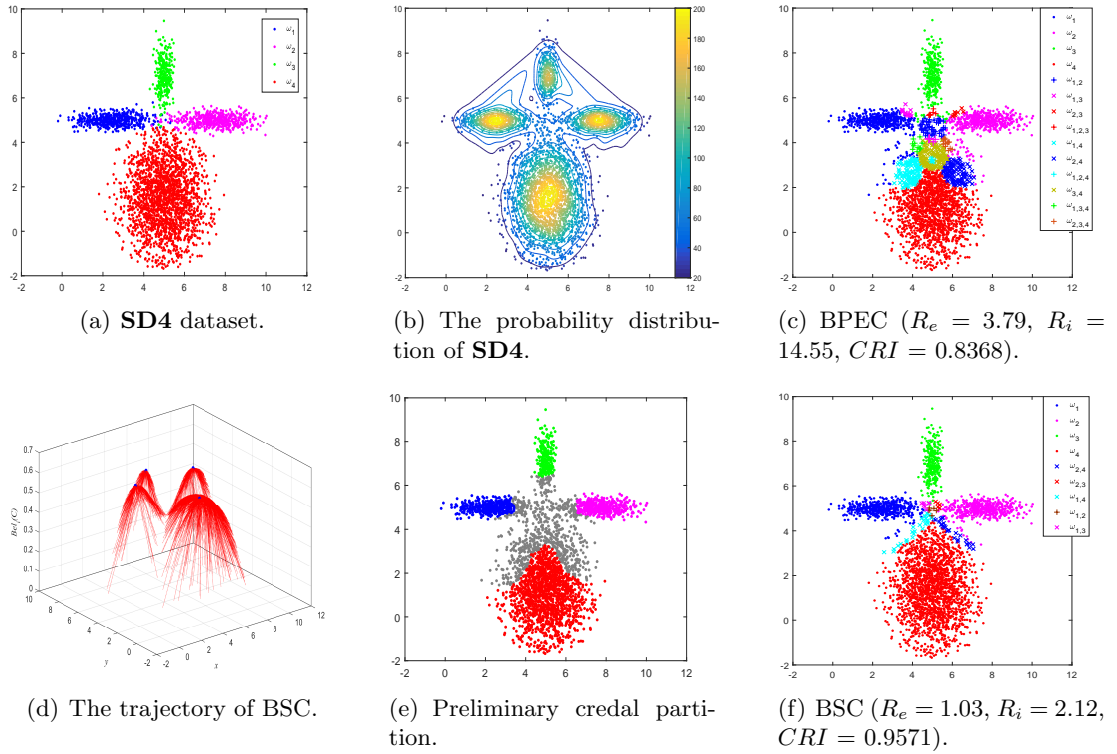


Fig. 5.6. Clustering results of the **SD4** dataset by different methods.

Fig. 5.6(a) intuitively shows the distribution of the points in this dataset and their probability distribution are given by Fig. 5.6(b). We take  $K = 70$ ,  $\alpha = 3$ ,  $\Delta = 5$  in BPEC and  $K_1 = 200$  in BSC, respectively. The clustering results of dataset **SD4** by different methods are given in Fig. 5.6(c)-(f). The BPEC and BSC based on the TBF can assign the points in overlapping areas of different classes to appropriate meta-

clusters so as to characterize the uncertainty and imprecision of these points and reduce the risk of errors. We can see that BSC yields a lower error rate and imprecision rate than BPEC. Although BPEC can also provide credal partition, it assigns a part of points belonging to the majority clusters to the minority clusters in dealing with the imbalanced data, for instance, some points in the cluster  $\{\omega_4\}$  are assigned to  $\{\omega_1\}$  and  $\{\omega_2\}$ , since the method based on symmetric distance only considers the symmetry of points in the feature space and does not take into account the distribution or the number of points in different clusters. In such a case, the center of meta-cluster tends to shift to the singleton cluster with majority points, so we may obtain the unreasonable clustering results. By contrast, BSC simulates cluster centers using neighbors at the boundary of different clusters, which can effectively decrease the negative impact of clusters with arbitrary sizes, *i.e.* imbalanced data, on the methods based on symmetric distance.

### 5.4.3 Medical and natural data

In this experiment, a medical image and a natural image are employed to demonstrate the effectiveness of BSC applying to image segmentation.

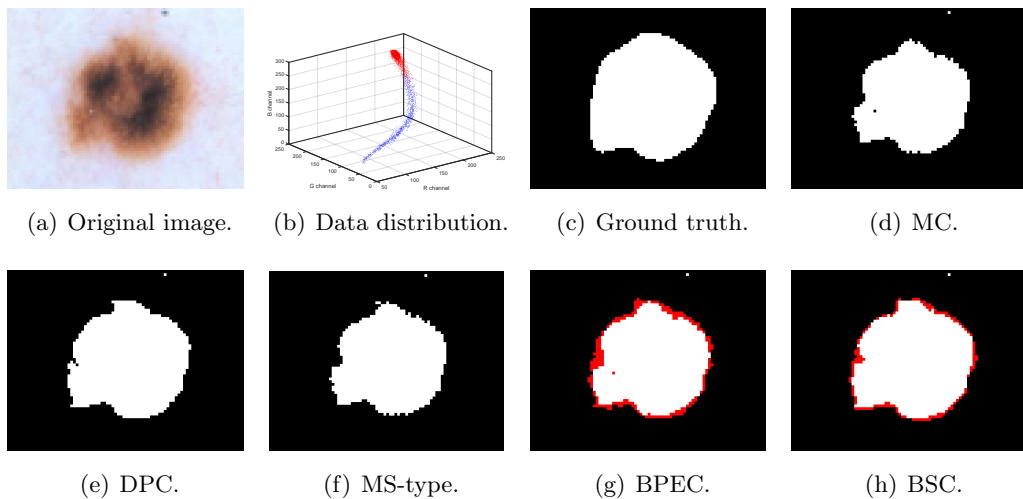


Fig. 5.7. Clustering results of the medical image by different methods.

The true color dermoscopic image (invasive malignant melanoma) of  $66 \times 86$  pixels, named **DI2**, as shown in Fig. 5.7(a), can be obtained from the EDRA Interactive Atlas of Dermoscopy (Argenziano et al., 2002). It consists of two classes including lesion and non-lesion and its ground truth is given by Fig. 5.7(c). The distribution of the pixels is shown in Fig. 5.7(b). The blue and red points with three dimensions including  $R$ ,  $G$ , and  $B$  value represent the pixels of the lesion and non-lesion according to the

ground truth, respectively. Here we choose  $f = 0.02$  in DPC and  $r = 150$  in MS-type.  $K = 2000$ ,  $\alpha = 3$ ,  $\Delta = 500$  and  $K_1 = 1000$  are set in BPEC and BSC, respectively.

We can see from Fig. 5.7(a) that the lesion edge is ambiguous and the distribution of the pixels given in Fig. 5.7(b) intuitively reveals that there are some pixels distributed in the overlapping area of different clusters. These pixels correspond to the lesion edge, and they are really difficult to be accurately distinguished. The segmentation results of MC, DPC, and MS-type are shown in Fig. 5.7(e)-(g), where some pixels in the lesion edge are assigned to the specific cluster (*i.e.* lesion or non-lesion), and most of them are misclassification. Interestingly, we can observe from Fig. 5.7(h),(d) that BPEC and BSC cautiously assign the pixels in lesion edge to the meta-cluster composed of the lesion and non-lesion under the TBF. That is, the pixels in this area are indistinguishable for clusters of the lesion or non-lesion. By doing this, it cannot only reduce the risk of error but also characterize the uncertain information between different clusters. The analysis of a natural gray image of  $60 \times 90$  pixels from the Berkeley Segmentation Dataset (Martin et al., 2001), named **NI2**, a goose floating on the lake, is given in Fig. 5.8. The clustering results of different methods based on these two images are given in Table 5.2.

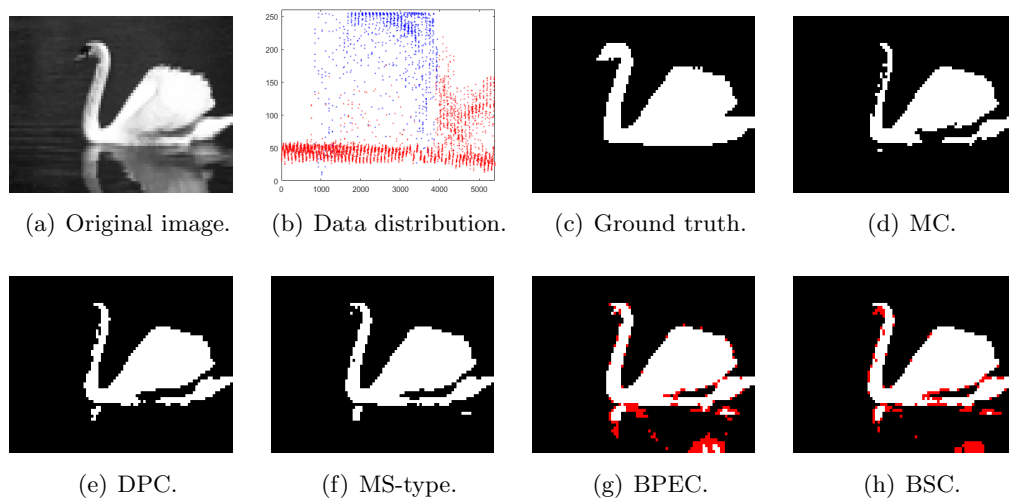


Fig. 5.8. Clustering results of different methods with a goose floating on the lake.

#### 5.4.4 Some real world data

In this experiment, we evaluate the performances of BSC compared with other methods based on the UCI databases<sup>7</sup>. The basic information of these datasets including the number of clusters (#Clus.), attributes (#Attr.), instances (#Inst.), and the number

<sup>7</sup>available at <http://archive.ics.uci.edu/ml/>.

Table 5.2: The results of different methods with **DI2** and **NI2**

Data	Indexes	MC	DPC	MS-type	BPEC	BSC
<b>DI2</b>	<i>Re</i>	4.61	3.91	6.93	3.11	<b>2.24</b>
	<i>Ri</i>	/	/	/	3.46	<b>2.40</b>
	<i>CRI</i>	0.8201	0.8554	0.7255	0.9031	<b>0.9476</b>
<b>NI2</b>	<i>Re</i>	5.76	4.83	3.85	2.85	<b>2.63</b>
	<i>Ri</i>	/	/	/	3.50	<b>3.17</b>
	<i>CRI</i>	0.7464	0.7878	0.8313	0.9106	<b>0.9235</b>

of objects of different clusters are reported in Table 5.3.

Table 5.3: Basic information of the UCI datasets

Data	#Clus	#Attr.	#Inst.	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_3\}$
Appendicitis(Ap)	2	7	106	21	85	/
Biodeg(Bi)	2	41	1055	356	699	/
Spambase(Sp)	2	56	4597	1812	2785	/
Abalone(Ab)	3	7	600	233	239	128
Seeds(Se)	3	7	210	70	70	70
Contraceptive(Co)	3	8	1473	629	333	511

Table 5.4: Selection of the parameters in different methods

Method	DPC	MS-type	BPEC			BSC
Indicators	$f$	$r$	$K$	$\alpha$	$\Delta$	$\mathcal{K}_1$
Ap	0.02	0.55	50	2	30	10
Bi	0.02	70	100	2	200	400
Sp	0.02	12000	200	2	5000	2000
Ab	0.02	3	30	6	30	60
Se	0.02	2	20	3	30	20
Co	0.02	7	30	2	30	200

The main parameters of DPC, MS-type, and BPEC in clustering different datasets are given in Table 5.4 and the other parameters in this experiment are default. Here the cluster label of these real datasets does not contain noise and we take proper  $\Delta$  in BPEC and  $\alpha = 0$  in BSC, respectively. That is, the clustering results of different methods do not include outlier. The clustering results of different methods are shown

Table 5.5: Clustering results of different methods with the UCI datasets (In %)

Data	Indicators	MC	DPC	MS-type	BPEC	BSC
Ap	<i>Re</i>	24.53	13.21	20.75	19.81	<b>12.26</b>
	<i>Ri</i>	/	/	/	2.83	<b>2.83</b>
	<i>CRI</i>	0.2410	0.4194	-0.0139	0.7127	<b>0.7795</b>
Bi	<i>Re</i>	37.63	33.84	34.60	41.99	<b>32.13</b>
	<i>Ri</i>	/	/	/	2.09	<b>13.65</b>
	<i>CRI</i>	-0.0268	-0.0009	-0.0081	0.6121	<b>0.7138</b>
Sp	<i>Re</i>	37.11	35.81	39.37	33.41	<b>28.52</b>
	<i>Ri</i>	/	/	/	4.74	<b>7.50</b>
	<i>CRI</i>	0.0077	0.0450	0.0005	0.5927	<b>0.6426</b>
Ab	<i>Re</i>	55.00	54.50	60.17	57.17	<b>52.17</b>
	<i>Ri</i>	/	/	/	5.00	<b>0</b>
	<i>CRI</i>	0.0455	0.0985	-0.0008	<b>0.5896</b>	0.5570
Se	<i>Re</i>	18.10	11.43	13.33	8.57	<b>7.62</b>
	<i>Ri</i>	/	/	/	6.19	<b>7.62</b>
	<i>CRI</i>	0.5702	0.7027	0.6592	0.8242	<b>0.9027</b>
Co	<i>Re</i>	61.10	59.06	60.56	57.98	<b>56.96</b>
	<i>Ri</i>	/	/	/	7.26	<b>6.65</b>
	<i>CRI</i>	0.0236	0.0099	0.0186	<b>0.6861</b>	0.6244

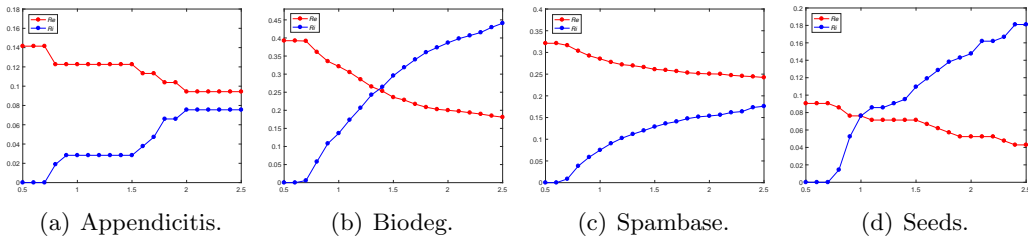


Fig. 5.9. Detection performance of BSC with different  $\gamma$ .

in Table 5.5. From these results, we can see that the error of BSC is obviously lower than that of MC, DPC, MS-type, and BPEC, and its imprecision rates are within an acceptable range, which can truly reflect that BSC is superior to other methods. The data sets Appendicitis, Biodeg, Spambase, and Seeds are employed to test the effect of different parameter  $\gamma$  on the clustering results of BSC as shown in Fig. 5.9, where the  $x$ -coordinate denotes the value of  $\gamma$ , ranging from 0.5 to 2.5, and the  $y$ -coordinate represents its error rate and imprecision rate, which is expressed in  $[0, 1]$ . We can see

that the error rate of BSC gradually decreases as the parameter  $\gamma$  changes from 0.5 to 2.5, while the imprecision rate increases, which indicates that  $\gamma$  can well adjust the degree of the imprecision and help to reduce the error rate. In real applications, the parameter  $\gamma$  should correspond to a compromise between the error rate and imprecision rate and it can be chosen depending on the imprecision rate that we can accept.

#### 5.4.5 Olivetti Face data

In this experiment, the Olivetti Face Database (Rodriguez and Laio, 2014), a widespread benchmark for machine learning, is applied to further evaluate the potential of the BSC in unsupervised image classification. The image data, called **FI5**, contains five people, each of whom has 10 face pictures with different shooting angles and expressions. The original data of these faces are shown in Fig. 5.10(a), where the faces of the same color belong to the same cluster, *i.e.* the same person. Here we take  $f = 0.2$  in DPC and  $r = 9$  is set in MS-type.  $K = 8$ ,  $\alpha = 3$ ,  $\Delta = 10$  in BPEC and  $K_1 = 6$ ,  $\alpha = 0.1$  are taken in BSC, respectively. The other parameters are default<sup>8</sup>.

The images of the clustering results of comparison methods are shown in Fig. 5.10. It can be intuitively seen that women’s faces are accurately identified by these methods. However, for the faces of four other people, MC, DPC, and MS-type mistakenly cluster them, since some faces are not clearly distinguished. For instance, the faces of the fourth and fifth people are very similar, and they are assigned exactly by MC, DPC, and MS-type, which may increase the risk of errors. We can see from Fig. 5.10(b),(f) that BPEC and BSC both produce the credal partitions to reduce errors. However, it is worth noting that BPEC assigns two faces marked by black of the fourth person to meta-cluster composed of the second and the fifth person, *i.e.* it believes that the two faces are difficult to distinguish between the second and the fifth person, which is obviously unreasonable. BSC can accurately recognize the first three people, and for two faces marked by cyan of the fifth people, it considers that they are hard to distinguish between the fourth and the fifth people only using the existing information, which is a prudent decision that can effectively reduce the error rates and fits with what we reasonably expect. These imprecise images in meta-cluster can be eventually distinguished using some other techniques or with extra information sources. The clustering results of different methods are shown in Table 5.6 and the results verify that BSC has potential in image classification.

#### 5.4.6 Discussion

Since the parameters  $\beta$ ,  $\gamma$  involved in BSC are very similar to these in ECM (Masson and Dencoux, 2008), CCM (Liu et al., 2015), MECM (Zhou et al., 2015c), we will

<sup>8</sup>We take  $K_2 = 5$  since only a few objects in each class. In this case, BSC needs to obtain different neighbors from various clusters of the precise objects to yield simulated cluster centers.

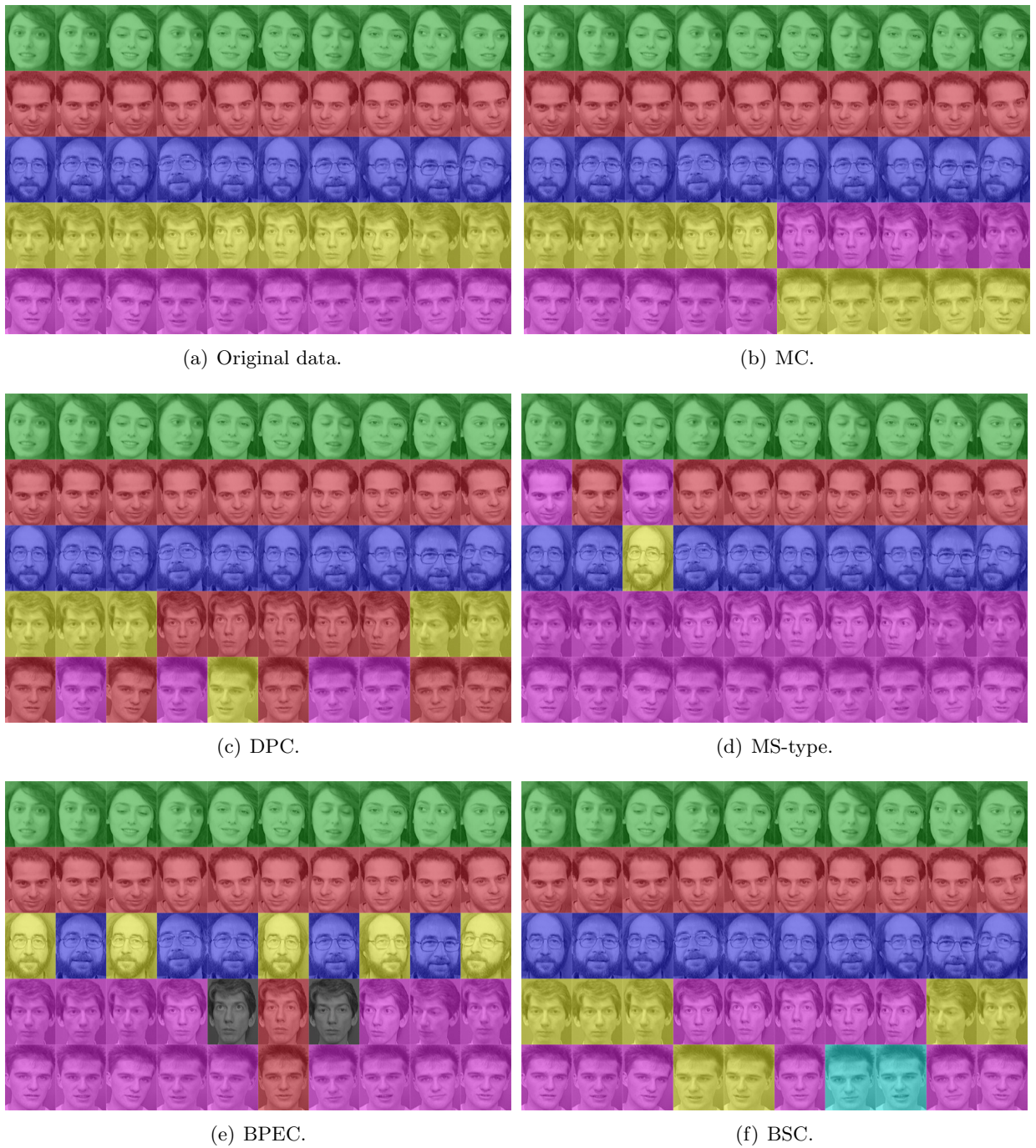


Fig. 5.10. Clustering results of **FI5** by different methods.

not discuss them separately here. In this section, the outlier adjustment factor  $\alpha$ , controlling the number of objects assigned to the noise cluster, will be discussed in detail. In order to verify the effectiveness of parameter  $\alpha$  in controlling the number of

Table 5.6: Clustering results of different methods with **FI5**

Data	Indexes	MC	DPC	MS-type	BPEC	BSC
	<i>Re</i>	20.00	22.00	26.00	32.00	<b>14.00</b>
<b>FI5</b>	<i>Ri</i>	/	/	/	0	<b>4.00</b>
	<i>CRI</i>	0.7278	0.5571	0.6251	0.7701	<b>0.9343</b>

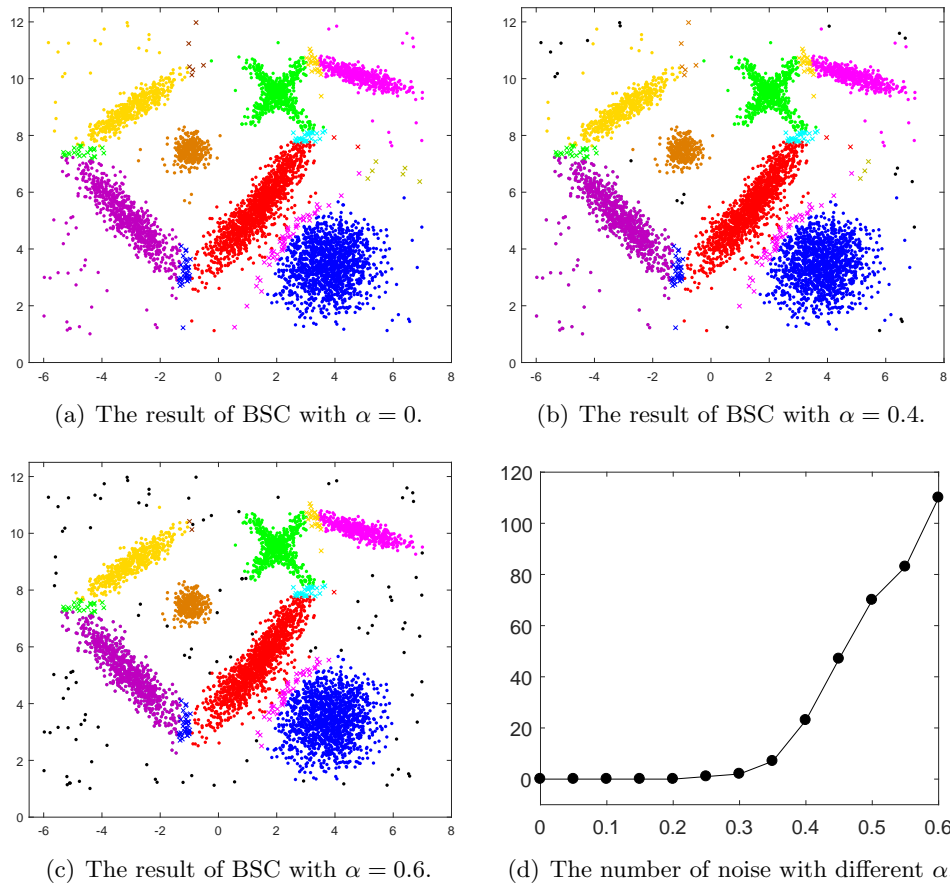


Fig. 5.11. The results of BSC with different  $\alpha$ .

noise, the dataset, as shown in Fig. 1.1, is employed and clustered by BSC with various  $\alpha$  values. The clustering results with three values of  $\alpha$ , *i.e.*  $\alpha \in [0, 0.4, 0.6]$ , are shown in Fig. 5.11(a)-(c), where the exact singleton clusters and proper meta-clusters marked by points and crosses with different colors, respectively. The black points represent the noise (outlier), and the number of noise with  $\alpha$  from 0 to 0.6 is given by Fig. 5.11(d). The value of  $\alpha$  corresponds to the  $x$ -coordinate and the  $y$ -coordinate represents the number of noise. Interestingly, we can intuitively observe that BSC can assign all



points to proper clusters without noise if we take a very small  $\alpha$ , *e.g.*  $\alpha = 0$ , as shown in Fig. 5.11(a). As  $\alpha$  increases, of course, those objects far from the clusters will be gradually assigned as noise, which is consistent with our intuitive perception. Some objects that are far from the clusters, for example, are regarded as noise if one takes a big  $\alpha$ , *e.g.*  $\alpha = 0.6$ , as shown in Fig. 5.11(c). We can also continue to infer that with the continuous increase of  $\alpha$ , increasingly objects will be assigned as noise. Fig. 5.11(d) reveals that the parameter  $\alpha$  can effectively adjust the number of noise in BSC and it can be taken depending on the number of noise that we can accept.

## 5.5 Conclusion

In this chapter, we proposed a new belief shift clustering (BSC) method to capture the uncertainty and imprecision between imbalanced clusters, which can be regarded as the evidential version of mean shift or mode seeking under the TBF. For each query object, it is preliminarily assigned as the noise, precise, or imprecise one based on the notion of “belief shift”. For the imprecise object, the BSC tries to reassign it to the singleton cluster or related meta-cluster again by partial credal redistribution with EC rule. The objects clustered into the meta-clusters, of course, are usually lying in the overlapping or middle areas of different singleton clusters. We need to combine with other technologies if further distinguish them. By doing this, it can effectively reduce the error rates and reasonably capture the uncertainty and imprecision between the imbalanced clusters, which has been verified by four experiments. However, we may deal with clusters of arbitrary shapes and sizes in some specific cases. The proposed BSC method in this chapter is still based on masses of belief obtained from the minimization of the objective function in the second step (credal redistribution), which may not apply to clusters of arbitrary shapes and sizes in space. For this reason, we will propose a new method in the next Chapter 6 that can detect clusters of arbitrary shapes and sizes, and characterize the uncertainty and imprecision between these clusters.

# 6

## Density peaks-based evidential convergence algorithm

**Abstract:** In this chapter, we present a density peaks-based evidential convergence algorithm to detect arbitrary clusters and characterize the uncertainty and imprecision in results. To this end, we first review the relevant benchmark algorithm, density peaks clustering. After that, we present the proposed algorithm and discuss the involved parameters in detail. Then, we demonstrate the effectiveness of the proposed algorithm based on several datasets. Before concluding the chapter, we also discuss the complexity of the algorithm and the sensitivity of the involved parameters.

### Contents

---

<b>6.1</b>	<b>Introduction</b>	<b>86</b>
<b>6.2</b>	<b>Review of density peaks clustering</b>	<b>86</b>
<b>6.3</b>	<b>Density peaks-based evidential convergence (DPEC)</b>	<b>87</b>
6.3.1	Evidence-based density peaks	87
6.3.2	Evidential convergence	89
6.3.3	The involved parameters	91
<b>6.4</b>	<b>Experiments</b>	<b>92</b>
6.4.1	Performance evaluation	92
6.4.2	Synthetic data	92
6.4.3	Olivetti Face data	96
6.4.4	Natural data	98
<b>6.5</b>	<b>Discussions</b>	<b>100</b>
6.5.1	Complexity analysis	100
6.5.2	Parametric sensitivity	100
6.5.3	Comparisons of proposed algorithms	101
<b>6.6</b>	<b>Conclusion</b>	<b>107</b>

---

## 6.1 Introduction

In the previous chapter, we proposed a belief shift clustering (BSC) algorithm, which can effectively extract the structure of non-spherical and imbalanced clusters. However, its core is still based on traditional evidential clustering (EC) methods, *i.e.* it is designed based on minimization of the objective function. In this chapter, we propose a method that can detect clusters of arbitrary shapes and sizes. Recently, a novel method called density peaks clustering (DPC) (Rodriguez and Laio, 2014) has attracted a lot of attention and is known for its ability to rapidly converge to clusters regardless of the shape and of the dimensionality of the space in which they are embedded. It should be noted that DPC is a medoid-based clustering algorithm where the cluster center matrix is composed of real objects (Krishnapuram et al., 2001; Zhou et al., 2016a). Our proposed clustering algorithm for evidential convergence is precisely based on DPC, and named density peaks-based evidential convergence (DPEC) algorithm. The rest of this chapter is organized as follows. After a brief introduction of DPC in Section 6.2, the DPEC is introduced in Section 6.3. The performance of DPEC is then tested in Section 6.4 and compared with several classical methods, followed by conclusions.

## 6.2 Review of density peaks clustering

Density peaks clustering (DPC) (Rodriguez and Laio, 2014) has attracted a lot of attention and is known for its ability to rapidly converge clusters regardless of the shape and of the dimensionality of the space in which they are embedded. In DPC, it is assumed that cluster centers are surrounded by neighbors of lower local density and that their distances from points of higher local density are relatively large. Thus, for each object  $\mathbf{x}_i$ , two quantities can be computed, namely the local density  $\rho_i$  and the distance  $d_{ij}$  from objects of higher density. The local density  $\rho_i$  is first defined by:

$$\rho_i = \sum_{j \neq i} \chi(d_{ij} - d_c) \quad (6.1)$$

where  $d_c$  is a cutoff distance and  $\rho_i$  is the number of objects that are closer than  $d_c$ . For another quantity  $\delta_i$ , it is measured by calculating the minimum distance between  $\mathbf{x}_i$  and any other object of higher density:

$$\delta_i = \min_{j: \rho_j > \rho_i} (d_{ij}) \quad (6.2)$$

For the object  $\mathbf{x}_i$  of the highest density,  $\delta_i = \max_j (d_{ij})$  is the default. Next, a decision graph with  $\rho_i$  and  $\delta_i$  is drawn to choose the cluster centers, specifically, defined as the objects that have both high density  $\rho_i$  and large distance  $\delta_i$ . For convenience, DPC also gives a quantity  $\gamma_i$ , denoted as  $\gamma_i = \rho_i \delta_i$ , and the objects with the largest  $\gamma_i$  values

are selected as the centers. Afterward, each remaining point is converged to the same cluster as its nearest neighbor of higher density. Due to the significant advantages in recognizing clusters with arbitrary shapes and dimensionality of the space, a number of interesting works on DPC have emerged, for example, see (Zhang et al., 2016; Bian et al., 2020; Pizzagalli et al., 2019; Yao Hui et al., 2017; Su and Dencœux, 2018).

### 6.3 Density peaks-based evidential convergence (DPEC)

In this section, we present a detailed description of the DPEC method. Similar to DPC, the main procedure also consists of two steps: 1) Definition of cluster centers by evidence-based density peaks, and 2) Assignment of the remaining objects by evidential convergence. These are discussed in Subsections 6.3.1 and 6.3.2, respectively.

#### 6.3.1 Evidence-based density peaks

The purpose of this subsection is to define a small number of objects as cluster centers. For the object  $\mathbf{x}_i$ , whether it is selected as a center depends on the quantities:  $\rho_f$  and  $\delta_i$ . Thus, we will describe the process of calculating each of these two quantities.

First of all, the neighbors are the basis for providing useful information of  $\mathbf{x}_i$  becoming a center. Let's consider a dataset  $\mathcal{X}$  including  $n$  objects with  $\Omega = \{\omega_1, \dots, \omega_c\}$ . For  $\mathbf{x}_i$ , the set  $\mathcal{S}_{\mathcal{K}_1}(\mathbf{x}_i)$  of its neighbors is defined by:

$$\mathcal{S}_{\mathcal{K}_1}(\mathbf{x}_i) = \{\mathbf{x}_k \in \mathcal{X} \mid d_{ik} \leq d_{i\mathcal{K}_1}\} \quad (6.3)$$

where  $\mathcal{K}_1$  is the number of neighbors, and  $d_{i\mathcal{K}_1}$  denotes the Euclidean distance from  $\mathbf{x}_i$  to the  $\mathcal{K}_1$ -th neighbor. In other words, every object whose distance to  $\mathbf{x}_i$  is less than or equal to  $d_{i\mathcal{K}_1}$  provides one evidence to support  $\mathbf{x}_i$  as a center. The evidence is denoted by a mass function  $m_{ik}(\cdot)$  in a new frame of discernment  $\Omega_{\mathcal{C}} = \{\mathcal{C}enter(\mathcal{C}), \mathcal{U}nknown(\mathcal{U})\}$ , which describes whether the degree of belief of  $\mathbf{x}_i$  as a center ( $\mathcal{C}$ ) or unknown ( $\mathcal{U}$ ). The mass function  $m_{ik}(\cdot)$  is defined by:

$$m_{ik}(A) = \begin{cases} \frac{1}{\mathcal{K}_1} \cdot e^{-\vartheta_i d_{ik}}, & A = \mathcal{C} \\ 1 - \frac{1}{\mathcal{K}_1} \cdot e^{-\vartheta_i d_{ik}}, & A = \mathcal{U} \end{cases} \quad (6.4)$$

where  $d_{ik}$  is the Euclidean distance from  $\mathbf{x}_i$  to the  $k$ -th neighbor.  $\vartheta_i$  is an adjustable distance factor, and it has been discussed in detail in (Su and Dencœux, 2018; Dencœux, 2008b). Here, it is defined as the inverse of a quantile of the distances between  $\mathbf{x}_i$  and its  $k$ -th neighbor  $\mathbf{x}_k$ , *i.e.*  $\vartheta_i = 1/\text{quantile}(d_{ik}, 0.9)$ , and 0.9 is the quantile number. If interesting, see (Su and Dencœux, 2018).

We can find from Eq. (6.4) that the smaller the distance  $d_{ik}$ , in general, the more the

neighbor  $\mathbf{x}_k$  strongly supports  $\mathbf{x}_i$  as a center. In essence, the principle of the computed  $m_{ik}(\mathcal{C})$  is the same as that of the density  $\rho_i$  in DPC. That is, if the chosen  $\mathcal{K}_1$  neighbors are all very close to  $\mathbf{x}_i$ , it also precisely means that for a given cutoff distance  $d_c$ , a larger density  $\rho_i$  can be obtained in DPC. By combining these mass functions using Eq. (2.7), *i.e.* the DS rule, a normalized mass function  $m_i$  can be obtained as well as equal to the belief function  $Bel_i$  (Su and Denceux, 2018), defined by:

$$Bel_i(A) = \begin{cases} 1 - \prod_{j=1}^{\mathcal{K}_1} (1 - m_{ij}(\mathcal{C})), & A = \mathcal{C} \\ 1 - \prod_{j=1}^{\mathcal{K}_1} m_{ij}(\mathcal{C}), & A = \mathcal{U} \end{cases} \quad (6.5)$$

where  $Bel_i(\mathcal{C})$  provide us with the final mass of belief on the possibility of the object  $\mathbf{x}_i$  becoming a center. Thus, similar to DPC, we can first redefine the quantity  $\delta_i$  depending on  $Bel_i(\mathcal{C})$ , and denoted as follows:

$$\delta_i = \begin{cases} \max_{\mathbf{x}_j} \{d_{ij}\}, & \text{if } \forall \mathbf{x}_j, Bel_i \geq Bel_j; \\ \min_{\mathbf{x}_j: Bel_i < Bel_j} \{d_{ij}\}, & \text{otherwise.} \end{cases} \quad (6.6)$$

In the dataset  $\mathcal{X}$ , there may be large variability in the distribution of different clusters, we consider  $\delta_i$  to be more reliable than  $Bel_i(\mathcal{C})$  (or  $\rho_i$  in DPC) in supporting  $\mathbf{x}_i$  as a cluster center. Therefore, another quantity  $\rho_f$  is introduced to replace  $\rho_i$  as the density of  $\mathbf{x}_i$ , and  $\rho_f$  is defined as the distance  $\delta_i$ -based function. If the object  $\mathbf{x}_i$ , as a cluster center, already has a relatively large  $\delta_i$ , at this point, the quantity  $\rho_f$  is also desired to be relatively large. In this case, we have a larger  $\gamma_i = \rho_f \delta_i$ , which is helpful to correctly choose  $\mathbf{x}_i$  as a center in the decision graph. Based on the above analysis, it is easy to find that  $\rho_f$  and  $\delta_i$  are positively correlated, so a common and easy to understand method is employed here to define the function  $\rho_f$  as follows:

$$\rho_f = e^{\delta_i}. \quad (6.7)$$

By doing so, the two quantities  $\rho_f$  and  $\delta_i$  of each object in the set  $\mathcal{X}$  can be computed by Eqs. (6.6)-(6.7). Thus, we can construct a  $\rho_f$ - $\delta$  decision graph by plotting  $\rho_f$  versus  $\delta$ , where the objects with higher  $\rho_f$  and larger  $\delta_i$  are chosen as the cluster centers. The center objects obtained by this evidence-based density peaks method are more distinct in the decision graph, which provides a better comparison for the definition of centers.

### 6.3.2 Evidential convergence

The purpose of this subsection is to assign the remaining objects to different clusters, *i.e.* singleton clusters, meta-clusters, and the noise cluster, by a new evidential convergence rule. After the cluster centers have been found, similar to DPC, the remaining objects will be converged to different clusters one by one. The difference is that, in DPC, each object is assigned only on the basis of cluster information from the nearest neighbor. In contrast, we think that cluster information based on multiple neighbors would be more robust. This allows for both the potential representation of imprecision and avoidance of noise interference. More precisely, the remaining objects are sorted by  $Bel_i$  from largest to smallest, and the object with larger  $Bel_i$  is assigned first. In this case, it is conceivable that the neighbors for the first assigned object are exactly the cluster centers. Next, we will generalize the convergence process.

For the remaining object  $\mathbf{x}_t$ , refers to Eq. (6.3), the set  $\mathcal{S}_{\mathcal{K}_2}(\mathbf{x}_t)$  of the neighbors can be obtained, and  $\mathcal{K}_2$  is the number of objects in the set. Assume that the set  $\mathcal{S}_{\mathcal{K}_2}(\mathbf{x}_t)$  contains  $\Phi$  cluster information, *i.e.* these neighbors are from  $\Phi$  clusters,  $1 \leq \Phi \leq \mathcal{K}_2$ . Inspired by (Dencœux, 2008b; Su et al., 2018; Liu et al., 2013), each neighbor provides a piece of evidence by a mass function  $m_{\mathbf{x}_t}^k(\cdot)$  for the assignment, defined by:

$$\begin{cases} m_{\mathbf{x}_t}^k(\{\omega_\phi\}) = e^{-\vartheta_t d_{tk}} \\ m_{\mathbf{x}_t}^k(\{\bar{\omega}_\phi\}) = 1 - e^{-\vartheta_t d_{tk}} \end{cases} \quad (6.8)$$

where  $m_{\mathbf{x}_t}^k(\{\omega_\phi\})$  denotes the degree to which the  $k$ -th neighbor  $\mathbf{x}_k$  supports the object  $\mathbf{x}_i$  is assigned to the same cluster  $\{\omega_\phi\}$ ,  $\phi = 1, \dots, \Phi$ . In contrast,  $m_{\mathbf{x}_t}^k(\{\bar{\omega}_\phi\})$  is the probability that  $\mathbf{x}_t$  does not belong to  $\{\omega_\phi\}$ , *i.e.* the sum of the probabilities<sup>9</sup> that  $\mathbf{x}_t$  belongs to all other singleton clusters, including the noise cluster  $\emptyset$ , except for  $\{\omega_\phi\}$ . The factor  $\vartheta_t$  is the same as  $\vartheta_i$  in Eq. (6.4). From Eq. (6.8), we can find that if  $\mathbf{x}_k$  is closer to  $\mathbf{x}_t$ , the more it supports  $\mathbf{x}_t$  belonging to  $\{\omega_\phi\}$ .

For  $\Phi$ , if we have  $\Phi = 1$ , it implies that all  $\mathcal{K}_2$  neighbors are from the cluster  $\{\omega_\phi\}$  and that  $\mathbf{x}_t$  is naturally assigned to  $\{\omega_\phi\}$ ; If we have  $1 < \Phi < \mathcal{K}_2$ , it implies that some neighbors come from the same clusters. In this case, we can use the DS rule to preferentially fuse these pieces of evidence from the same clusters directly because they are non-conflicting or low-conflicting. Thus, for  $1 < \Phi \leq \mathcal{K}_2$ , we end up with  $\Phi$  pieces of evidence, and they are the key to assigning  $\mathbf{x}_t$  to different clusters. Since these  $\Phi$  pieces of evidence support  $\mathbf{x}_t$  assigned to  $\Phi$  different singleton clusters, some evidence may be highly conflicting. For example, we have two evidence,

<sup>9</sup>Note that each neighbor only provides some uncertain information so here we get is probabilistic knowledge.  $m_{\mathbf{x}_t}^k(\{\bar{\omega}_\phi\})$  includes the probability that  $\mathbf{x}_t$  belongs to any other clusters, e.g.,  $\{\omega_{\phi'}\}$ ,  $\{\omega_{\phi'}\} = \emptyset$  or  $\phi' \in [1, \Phi]$ ,  $\phi \neq \phi'$ , but the quantification of the probability is unknown. In evidence theory,  $m_{\mathbf{x}_t}^k(\{\bar{\omega}_\phi\})$  can be regarded as the (partial) ignorance, and it plays a neutral role in the fusion process. This is critical and will also be mentioned in the new fusion rules.

$m^1(\cdot) = [m_{\mathbf{x}_t}^1(\{\omega_1\}) = 0.93, m_{\mathbf{x}_t}^1(\{\bar{\omega}_1\}) = 0.07]$  and  $m^2(\cdot) = [m_{\mathbf{x}_t}^2(\{\omega_2\}) = 0.94, m_{\mathbf{x}_t}^2(\{\bar{\omega}_2\}) = 0.06]$ . Clearly, these two evidence strongly support the assignment of  $\mathbf{x}_t$  to the singleton clusters  $\{\omega_1\}$  and  $\{\omega_2\}$ , respectively, suggesting that both neighbors are very close to  $\mathbf{x}_t$ . In this case, DPC may assign  $\mathbf{x}_t$  to  $\{\omega_2\}$  because the neighbor  $\{\mathbf{x}_2\}$  seems to be a little closer, but this could be a huge risk of error. Here, we consider that  $\mathbf{x}_t$  may be distributed in the middle or overlapping region of these two singleton clusters. Therefore, it is more reasonable to assign  $\mathbf{x}_t$  to the meta-cluster  $\{\omega_1, \omega_2\}$  to represent imprecision. To achieve this goal, inspired by (Liu et al., 2013, 2014), we generalize the process and propose a new set of fusion rules, described as follows.

For simplicity and notation convenience, assume that there is a set of quantities, listed as  $m_{\mathbf{x}_t}^1(\{\omega_1\}), \dots, m_{\mathbf{x}_t}^\Phi(\{\omega_\Phi\})$ , and that  $\{\omega_{\max}\}$  is the most believed singleton cluster of  $\mathbf{x}_t$  with  $m_{\mathbf{x}_t}^{\{\max\}}(\{\omega_{\max}\}) = \max\{m_{\mathbf{x}_t}^\phi(\{\omega_\phi\}) \mid \phi \in [1, \Phi]\}$ . There are also other masses of belief, *e.g.*,  $m_{\mathbf{x}_t}^{\phi'}(\{\omega_{\phi'}\})$ , very close or equal to  $m_{\mathbf{x}_t}^{\{\max\}}(\{\omega_{\max}\})$  but  $\{\omega_{\max}\} \neq \{\omega_{\phi'}\}$ . In this case,  $\mathbf{x}_t$  can also potentially belong to  $\{\omega_{\phi'}\}$  with a high likelihood. Thus, we define a new dynamic framework  $\mathcal{M}_t$  by a threshold  $\zeta$  that contains all possible solutions to the singleton cluster of  $\mathbf{x}_t$ , defined by:

$$\mathcal{M}_t = \{\omega_{\max}, \dots, \omega_\phi \mid m_{\mathbf{x}_t}^{\{\max\}}(\{\omega_{\max}\}) - m_{\mathbf{x}_t}^\phi(\{\omega_\phi\}) \leq \zeta\} \quad (6.9)$$

where  $\mathcal{M}_t \subseteq \Omega$ , and  $\zeta \in [0, 1]$  is a chosen threshold. With the viewpoint of evidence theory,  $\mathbf{x}_t$  will be assigned in a new power-set  $2^{|\mathcal{M}_t|} \subseteq 2^\Omega$ . From Eq. (6.9), we can see that the clusters that are excluded under  $\zeta$  are not in the possible solutions, for example, the cluster  $\{\omega_l\}$  with  $m_{\mathbf{x}_t}^{\{\max\}}(\{\omega_{\max}\}) - m_{\mathbf{x}_t}^l(\{\omega_l\}) > \zeta$ . That is,  $\zeta$  can further compress the assignment range, which can eliminate noise due to oversized  $\mathcal{K}_2$ . In this case, the  $\mathcal{K}_2$  neighbors only provide a small possibility of  $\mathbf{x}_t$  belonging to  $\{\omega_l\}$ , because the neighbor  $\mathbf{x}_l$  is relatively far away from  $\mathbf{x}_t$ . After having the new power-set  $2^{|\mathcal{M}_t|}$ , these  $|\mathcal{M}_t|$  evidence can be fused by a new set of fusion rules, defined by:

$$m^{\mathcal{M}_t}(A) = \begin{cases} \text{If } A \in \mathcal{M}_t, A = \{\omega_\phi\}, |A| = 1, \\ m_{\mathbf{x}_t}^\phi(\{\omega_\phi\}) \cdot \prod_{\phi' \neq \phi, \phi'=1}^{|\mathcal{M}_t|} m_{\mathbf{x}_t}^{\phi'}(\{\bar{\omega}_{\phi'}\}) \\ \text{If } A \in \mathcal{M}_t, |A| \geq 2, \{\omega_\phi\} \in A, \\ \prod_{\phi=1}^{|A|} m_{\mathbf{x}_t}^\phi(\{\omega_\phi\}) \cdot \prod_{\phi'=|A|+1}^{|\mathcal{M}_t|} m_{\mathbf{x}_t}^{\phi'}(\{\bar{\omega}_{\phi'}\}) \\ \text{If } A = \emptyset, \\ \prod_{\phi'=1}^{|\mathcal{M}_t|} m_{\mathbf{x}_t}^{\phi'}(\{\bar{\omega}_{\phi'}\}) \end{cases} \quad (6.10)$$

where  $|\cdot|$  represents the cardinality of the set. For example, if  $\mathcal{M}_t = \{\omega_1, \omega_2, \omega_3\}$ , then  $|\mathcal{M}_t| = 3$ , and  $|\mathcal{M}_t| \leq \Phi$ .

In Eq. (6.10), the mass of belief (*i.e.* BBA) that  $\mathbf{x}_t$  belongs to a cluster is the continuous product of  $|\mathcal{M}_t|$  values. More precisely, in the first part of Eq. (6.10), when calculating the BBA of  $\mathbf{x}_t$  belonging to the singleton cluster  $\{\omega_\phi\}$ , it considers not only the BBA, *i.e.*  $m_{\mathbf{x}_t}^\phi(\{\omega_\phi\})$ , provided by neighbors belonging to  $\{\omega_\phi\}$ , but also the other  $|\mathcal{M}_t| - 1$  BBAs, *e.g.*  $m_{\mathbf{x}_t}^{\phi'}(\{\bar{\omega}_{\phi'}\})$ , provided by other neighbors since they also provide evidence that  $\mathbf{x}_t$  belongs to  $\{\omega_\phi\}$ . Moreover,  $m_{\mathbf{x}_t}^{\phi'}(\{\bar{\omega}_{\phi'}\})$  can be regarded as the ignorance, and it plays a neutral role in the fusion process. In the second part of

Eq. (6.10),  $\prod_{\phi=1}^{|\mathcal{A}|} m_{\mathbf{x}_t}^\phi(\{\omega_\phi\})$  represents the partial conflict and this product characterizes

in fact the joint belief that the object simultaneously belongs to these  $|\mathcal{A}|$  singleton clusters. In contrast, for any BBA, *i.e.*  $m_{\mathbf{x}_t}^\phi(\{\omega_\phi\})$ ,  $\phi \in [1, \Phi]$ , if it is small, it indicates that  $\mathbf{x}_t$  is far away from all the neighbors. In this case, we can obtain a larger  $m_{\mathbf{x}_t}^{\phi'}(\{\bar{\omega}_{\phi'}\})$

and define  $\prod_{\phi'=1}^{|\mathcal{M}_t|} m_{\mathbf{x}_t}^{\phi'}(\{\omega_{\phi'}\})$  as the BBA of  $\mathbf{x}_t$  belonging to the noise  $\emptyset$ , as shown in the

third part of Eq. (6.10). In fact, these conjunctive combinations are very similar to the unnormalized DS rule, thus, we can normalize these combined BBAs before making a final decision.

### 6.3.3 The involved parameters

In this new DPEC method, the parameters  $\mathcal{K}_1$  in Eq. (6.3),  $\mathcal{K}_2$  in Eq. (6.8), and the threshold  $\zeta$  in Eq. (6.9) should be tuned in applications. **(i)** The parameter  $\mathcal{K}_1$  plays a critical role on determining the distribution of objects in the decision graph, and we give a simple suggestion to increase  $\mathcal{K}_1$  until it is possible to visually separate some objects (*i.e.* clustering centers) from others. In most cases, the distribution of cluster centers in the decision graph is robust to  $\mathcal{K}_1$ . It has been discussed in detail in (Su and Denceux, 2018); **(ii)** Similar to KNN-based classifiers,  $\mathcal{K}_2$  is an open value. The principle of selecting  $\mathcal{K}_2$  is not to be too large or too small, because the choice of extremes may introduce noise. Interestingly, the DPEC is not too sensitive to the change of  $\mathcal{K}_2$ . **(iii)** The threshold  $\zeta$  controls the number of objects assigned to meta-clusters, adjusted by users according to the acceptable imprecision rate. In fact,  $\zeta$  is an interesting threshold since it allows DPEC to degrade to the DPC algorithm.

**(1) The DPC algorithm:** We have a classical DPC algorithm with  $\zeta = 0$  because it only needs to consider the cluster information of the nearest neighbor. That is,  $\mathbf{x}_t$  will be assigned to the cluster  $\omega_{\max}$ , which is essentially the same as the convergence mechanism of DPC.

**(2) The DPEC algorithm:** We have the DPEC algorithm with  $\zeta \in (0, 1]$ . In this case,  $\mathbf{x}_t$  may be assigned to meta-clusters to represent imprecision if it is distributed



in the middle or overlapping region of different singleton clusters. And if  $\zeta = 1$ , all clusters contained in the  $\mathcal{K}_2$  neighbors are potential solutions.

## 6.4 Experiments

### 6.4.1 Performance evaluation

Three experiments have been conducted to test and evaluate the performances of the proposed DPEC method with respect to DPC (Rodriguez and Laio, 2014), MC (Liang et al., 2012), CDPC (Zhang et al., 2016), ECM (Dencœux, 2008b), MS-type (Yamasaki and Tanaka, 2019), BPEC (Su and Dencœux, 2018), DPC-DBFN (Lotfi et al., 2020) and TCASP (Pizzagalli et al., 2019). Among them, DPC and its improved methods, *i.e.* CDPC, BPEC, DPC-DBFN, TCASP, are the benchmark references; MC, ECM and BPEC are representatives of partition-based methods; DPC and MS-type are classical density-based methods; ECM and BPEC are pioneering techniques under evidence theory. Besides, Subsection 6.4.2 is employed to illustrate the uses of DPEC and the limitations of comparison methods based on eight synthetic data with arbitrary shapes and sizes. Subsections 6.4.3 and 6.4.4 are used to demonstrate the potential of DPEC for image classification and segmentation based on face and natural data, respectively.

Moreover, since the DPEC method is proposed under the TBF and introduces the notion of meta-cluster to represent imprecision, some common evaluation criteria and their evidential versions are employed to quantify the performances of different methods. They are error rate ( $R_e$ ), imprecision rate ( $R_i$ ), Precision ( $P$ ), Recall ( $R$ ), Rand Index ( $RI$ ), Evidential Precision (EP), Evidential Recall (ER) and Evidential Rand Index (ERI). If interesting, see (Liu et al., 2015; Zhou et al., 2015c; Rand, 1971). For evidential clustering, we can employ pignistic probabilities, defined in Eq. (2.8), to make the clustering results crisp. Note that when comparing the closeness of crisp partition to the ground truth, the  $EP$ ,  $ER$  and  $ERI$  equal to traditional  $P$ ,  $R$  and  $RI$ , respectively. The upper bound of these criteria is 1. Except for  $R_e$  and  $R_i$ , a larger value corresponds to a better clustering performance.

### 6.4.2 Synthetic data

In this experiment, we conduct 8 clustering tasks to validate the effectiveness of DPEC with  $\mathcal{K}_1 = 0.06N$ ,  $\mathcal{K}_2 = 9$ ,  $\zeta = 0.1$ . The original datasets and the corresponding results by DPEC are shown in Fig. 6.1. The clustering results of comparison methods and the corresponding hyper-parameters are presented in **Appendix (A)**. Table. 6.1 reports the clustering results of different methods on the given evaluation indexes. We can see that the distributions of these eight datasets are diverse, where some classes are arbitrary shapes and sizes. Based on the clustering results of different methods, some limitations of comparison methods can be discussed as follows.

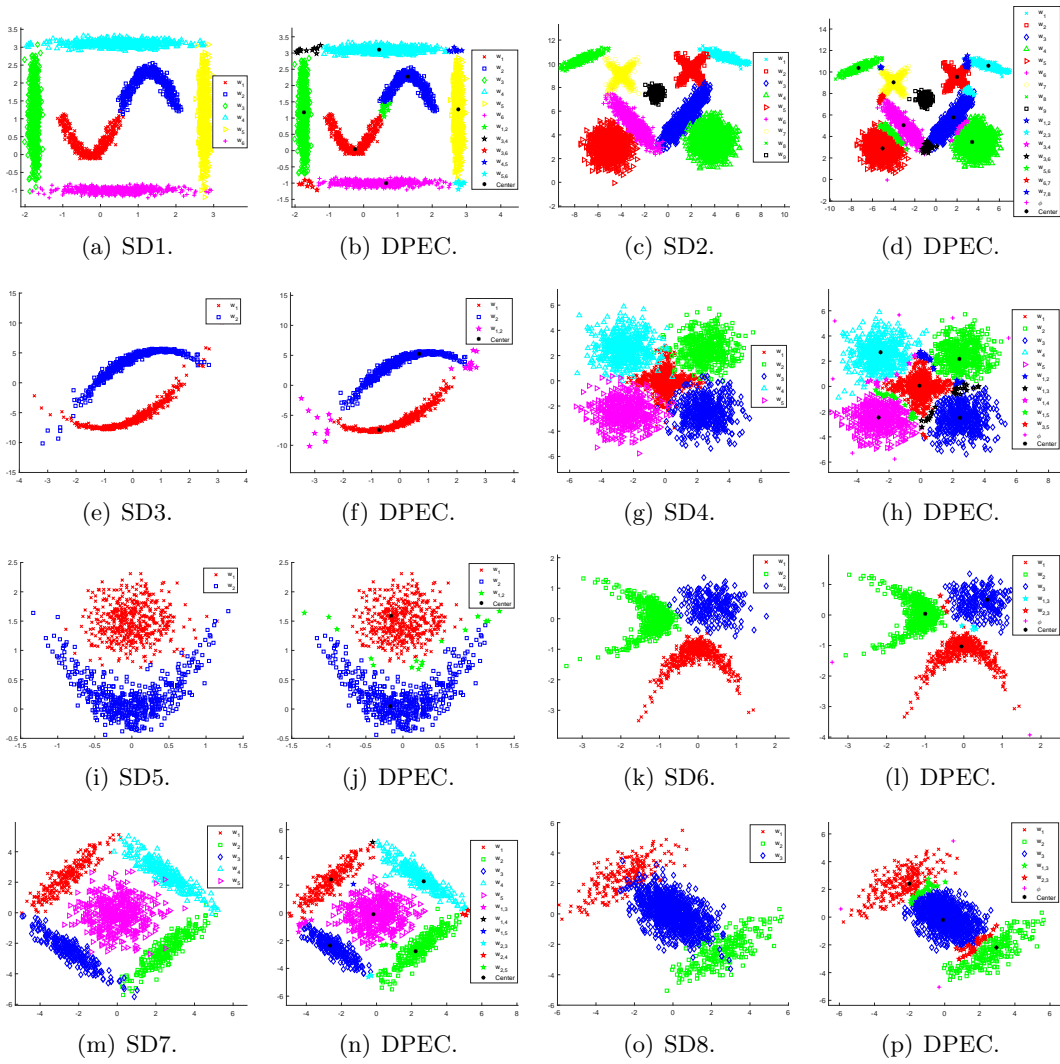


Fig. 6.1. Clustering results by DPEC on the eight synthetic datasets.

Table 6.1: Clustering results of different methods on the eight synthetic datasets

Datasets	Methods	$R_e$	$R_i$	$P$	$R$	$RI$	$EP$	$ER$	$ERI$
SD1	DPC	0.0085	\	0.9819	0.9822	0.9938	0.9819	0.9822	0.9938
	MC	0.1012	\	0.8310	0.8227	0.9408	0.8310	0.8227	0.9408
	CDPC	0.2556	\	0.6967	0.6944	0.8956	0.6967	0.6944	0.8956
	ECM	0.1075	0.0031	0.8263	0.8091	0.9380	0.8263	0.8091	0.9380
	MS-type	0.1138	\	0.7783	0.8654	0.9346	0.7783	0.8654	0.9346
	BPEC	0.0836	0.0695	0.8332	0.8458	0.9444	0.8755	0.8504	0.9509
	DPC-DBFN	0.0343	0.0000	0.9315	0.9469	0.9789	0.9315	0.9469	0.9789
	TCASP	0.0346	\	0.9397	0.9464	0.9804	0.9397	0.9464	0.9804
	DPEC	<b>0.0000</b>	0.0160	<b>0.9849</b>	<b>0.9850</b>	<b>0.9948</b>	<b>1.0000</b>	<b>1.0000</b>	<b>1.0000</b>
SD2	DPC	0.0107	\	<b>0.9834</b>	0.9780	0.9948	0.9834	0.9780	0.9948
	MC	0.1578	\	0.8474	0.8537	0.9594	0.8474	0.8537	0.9594
	CDPC	0.3387	\	0.6937	0.6534	0.9140	0.6937	0.6534	0.9140
	ECM	0.1703	0.0086	0.8572	0.7553	0.9498	0.8671	0.7655	0.9524
	MS-type	0.0206	\	0.9673	0.9605	0.9903	0.9673	0.9605	0.9903
	BPEC	0.0335	0.1669	0.7583	0.7963	0.9381	0.9435	0.9216	0.9813
	DPC-DBFN	0.0246	0.0001	0.9564	0.9490	0.9872	0.9564	0.9490	0.9872
	TCASP	0.0115	\	0.9800	0.9798	0.9946	0.9800	0.9798	0.9946
	DPEC	<b>0.0059</b>	0.0186	0.9818	<b>0.9801</b>	<b>0.9948</b>	<b>0.9899</b>	<b>0.9879</b>	<b>0.9970</b>
SD3	DPC	0.0160	\	0.9683	0.9684	0.9684	0.9683	0.9684	0.9684
	MC	0.0532	\	0.8991	0.8992	0.8992	0.8991	0.8992	0.8992
	CDPC	0.0481	\	0.9072	0.9094	0.9083	0.9072	0.9094	0.9083
	ECM	0.0532	0.0000	0.8991	0.8992	0.8992	0.8991	0.8992	0.8992
	MS-type	0.1474	\	0.7328	0.7809	0.7483	0.7328	0.7809	0.7483
	BPEC	0.1545	0.0000	0.7193	0.7815	0.7385	0.7193	0.7815	0.7385
	DPC-DBFN	0.0532	0.0000	0.8991	0.8992	0.8992	0.8991	0.8992	0.8992
	TCASP	0.0160	\	0.9683	0.9684	0.9684	0.9683	0.9684	0.9684
	DPEC	<b>0.0010</b>	0.0231	<b>0.9703</b>	<b>0.9703</b>	<b>0.9703</b>	<b>0.9980</b>	<b>0.9979</b>	<b>0.9979</b>
SD4	DPC	0.0320	\	0.9168	0.9492	0.9696	0.9168	0.9492	0.9696
	MC	0.1887	\	0.6994	0.6970	0.8663	0.6994	0.6970	0.8663
	CDPC	0.3487	\	0.5513	0.5379	0.8003	0.5513	0.5379	0.8003
	ECM	0.0330	0.0020	0.9188	0.9332	0.9669	0.9219	0.9399	0.9690
	MS-type	0.0450	\	0.8915	0.9221	0.9578	0.8915	0.9221	0.9578
	BPEC	0.0527	0.1053	0.9066	0.8759	0.9524	0.8454	0.9128	0.9442
	DPC-DBFN	0.0553	0.0120	0.8434	0.9210	0.9445	0.8424	0.9220	0.9441
	TCASP	0.1640	\	0.5889	0.8118	0.8324	0.5889	0.8118	0.8324
	DPEC	<b>0.0177</b>	0.0303	<b>0.9237</b>	<b>0.9505</b>	<b>0.9716</b>	<b>0.9546</b>	<b>0.9697</b>	<b>0.9830</b>
SD5	DPC	0.0343	\	0.9333	0.9340	0.9337	0.9333	0.9340	0.9337
	MC	0.0414	\	0.9182	0.9234	0.9206	0.9182	0.9234	0.9206
	CDPC	0.2886	\	0.5671	0.7491	0.5890	0.5671	0.7491	0.5890
	ECM	0.0404	0.0010	0.9182	0.9234	0.9206	0.9199	0.9252	0.9224
	MS-type	0.0363	\	0.9273	0.9328	0.9299	0.9273	0.9328	0.9299
	BPEC	0.0494	0.0182	0.8976	0.9080	0.9023	0.8998	0.9096	0.9043
	DPC-DBFN	0.0525	0.0000	0.8956	0.9064	0.9005	0.8956	0.9064	0.9005
	TCASP	0.0172	\	0.9661	0.9663	0.9662	0.9661	0.9663	0.9662
	DPEC	<b>0.0071</b>	0.0161	<b>0.9721</b>	<b>0.9721</b>	<b>0.9721</b>	<b>0.9857</b>	<b>0.9857</b>	<b>0.9857</b>
SD6	DPC	0.0056	\	0.9885	0.9929	0.9933	0.9885	0.9929	0.9933
	MC	0.0056	\	0.9876	0.9939	0.9933	0.9876	0.9939	0.9933
	CDPC	0.0313	\	0.9330	0.9692	0.9639	0.9330	0.9692	0.9639
	ECM	0.0064	0.0000	0.9858	0.9930	0.9924	0.9858	0.9930	0.9924
	MS-type	0.0754	\	0.8423	0.9106	0.9066	0.8423	0.9106	0.9066
	BPEC	0.0048	0.0225	0.9806	0.9905	0.9895	0.9891	0.9921	0.9931
	DPC-DBFN	0.0192	0.0016	0.9581	0.9801	0.9775	0.9580	0.9800	0.9774
	TCASP	0.0056	\	0.9885	0.9929	0.9933	0.9885	0.9929	0.9933
	DPEC	<b>0.0016</b>	0.0080	<b>0.9894</b>	<b>0.9947</b>	<b>0.9943</b>	<b>0.9964</b>	<b>0.9982</b>	<b>0.9981</b>
SD7	DPC	0.0284	\	0.9564	0.9417	0.9775	0.9564	0.9417	0.9775
	MC	0.1955	\	0.7004	0.7035	0.8674	0.7004	0.7035	0.8674
	CDPC	0.3762	\	0.5363	0.5258	0.7938	0.5363	0.5258	0.7938
	ECM	0.0595	0.0020	0.8990	0.8400	0.9435	0.9043	0.8482	0.9465
	MS-type	0.1313	\	0.7848	0.7111	0.8926	0.7848	0.7111	0.8926
	BPEC	0.0731	0.1103	0.8444	0.7697	0.9174	0.8365	0.8554	0.9368
	DPC-DBFN	0.0325	0.0000	0.9496	0.9226	0.9720	0.9496	0.9226	0.9720
	TCASP	0.0237	\	0.9643	0.9544	0.9820	0.9643	0.9544	0.9820
	DPEC	<b>0.0169</b>	0.0142	<b>0.9732</b>	<b>0.9642</b>	<b>0.9862</b>	<b>0.9737</b>	<b>0.9650</b>	<b>0.9864</b>
SD8	DPC	0.0396	\	0.9320	<b>0.9408</b>	0.9357	0.9320	0.9408	0.9357
	MC	0.1423	\	0.8786	0.6819	0.7927	0.8786	0.6819	0.7927
	CDPC	0.5483	\	0.5626	0.4407	0.5465	0.5626	0.4407	0.5465
	ECM	0.1409	0.0020	0.8792	0.6831	0.7935	0.8792	0.6838	0.7941
	MS-type	0.0624	\	0.9490	0.8503	0.9018	0.9490	0.8503	0.9018
	BPEC	0.1013	0.1389	0.8806	0.6836	0.7943	0.8825	0.7691	0.8487
	DPC-DBFN	0.0879	0.0094	0.8182	0.9193	0.8567	0.8241	0.9208	0.8600
	TCASP	0.0799	\	0.8426	0.9250	0.8754	0.8426	0.9250	0.8754
	DPEC	<b>0.0255</b>	0.0416	<b>0.9487</b>	0.9228	<b>0.9360</b>	<b>0.9710</b>	<b>0.9422</b>	<b>0.9564</b>

As partition-based methods, MC, ECM, and BPEC are not able to detect non-spherical clusters because they always assign the object to the nearest cluster center. Thus, they do not work well when dealing with these clusters with arbitrary shapes and sizes. However, since MC proposes representing a real class with multiple centers, it can detect spherical data with imbalanced clusters. ECM and BPEC based on evidence theory can characterize the imprecision between clusters but are still only applicable to spherical data, although BPEC is an improved DPC method under the TBF.

DPC and its improved methods, *i.e.* CDPC, DPC-DBFN, TCASP, are often effective in detecting these clusters of arbitrary shapes and sizes. Moreover, they can detect the structures of these clusters in most cases but do not characterize the imprecision of objects in overlapping or middle regions. Furthermore, these objects often have characteristics of different clusters, which can easily result in misclassification once they are forced to be assigned. CDPC designs a relative distance and proposes a new density measure based on the tree structure but sometimes does not find reasonable cluster centers. See the results on these datasets. DPC-DBFN uses a density-based KNN graph to label backbones to prevent the chain reaction. TCASP proposes to substitute the local association rule of DPC with the solution of a global optimization problem on a graph. Although DPC-DBFN, TCASP can find reasonable cluster centers in most cases, the assignment of the remaining objects is unreasonable sometimes, especially those prone to misclassification. For example, DPC-DBFN tries to assign true labels to those objects located on the border regions based on the labeled backbones, but the reliability of the backbones sometimes needs to be further evaluated. See the results on the SD1, SD2, SD5 datasets. In addition, DPC-DBFN assigns some objects away from the backbones as noise. Here we count them as imprecise objects for a fair comparison. BPEC argues that all objects in the neighbors of each object provide pieces of evidence that induce belief on the possibility of such object becoming a cluster center. This suggestion is inherited by DPEC. However, although BPEC can find reasonable cluster centers in most cases, it still does not address the small relative density of one real center object due to the diversity of data distribution. Besides, BPEC detects cluster structures based on obtained centers and is still based on the objective function strategy, a step backward of the DPC method. Thus, although it improves DPC under evidence theory to characterize imprecision, the ability is limited in such a case.

Like the DPC method, MS-type, another branch (mean shift or mode seeking) of density-based methods, can easily detect cluster structures of arbitrary shape and size. However, it may produce irrational results when different clusters are very close and have similar densities because assigning objects based on the number of visits by different clusters may be unreasonable in this case. See the results on the SD1, SD3, SD5 datasets. Moreover, MS-type may be more suitable for low-dimensional data because objects are density-sensitive in the shifting process. Thus, the results are highly dependent on the ability to measure density (distance). In addition, MS-type cannot

reasonably characterize the imprecision between detected clusters of arbitrary shapes and sizes. Therefore, it can only forcibly assign objects in overlapping or middle regions to specific clusters based on the number of times the objects are visited.

The proposed DPEC method integrates the advantages of both DPC and evidence theory by detecting clusters of arbitrary shapes and sizes and reasonably characterizing the imprecision between these clusters. In DPEC, we redefine a distance-based density function that strives to allow a real center object to have a more significant density regardless of data distribution in different clusters. Thus DPEC can obtain more robust center objects. In converging the remaining objects, the neighbors (including centers) with cluster labels are employed as prior knowledge. Therefore, we present a new evidential convergence rule to assign these objects one by one to different clusters. In the process, some imprecise objects are converged to related meta-clusters to represent imprecision. The results on these datasets can demonstrate the power of DPEC.

### 6.4.3 Olivetti Face data

In this experiment, we conduct a clustering task to evaluate the performance of DPEC on high-dimensional datasets. The Olivetti faces dataset (Samaria and Harter, 1994) is a database widely used in the field of machine learning, which includes the facial images of 40 people, each of whom has 10 images with different angles. Five faces, *i.e.* 50 images, are employed to test. Furthermore, to reduce the complexity of storage and computation, we reduce the dimensions to 6 by principal component analysis (PCA) (Wold et al., 1987; Ke and Sukthankar, 2004). The faces with the same color belong to the same class, *i.e.* the same person. Here we take  $r = 10$  in MS-type,  $\alpha = 1$  and  $\delta = 8$  in ECM,  $K = 5$  in BPEC,  $K = 15$  in DPC-DBFN, and  $\mathcal{K}_1 = 2$ ,  $\mathcal{K}_2 = 4$ ,  $\zeta = 0.1$  in DEPC, respectively. The other parameters of different methods are defaults.

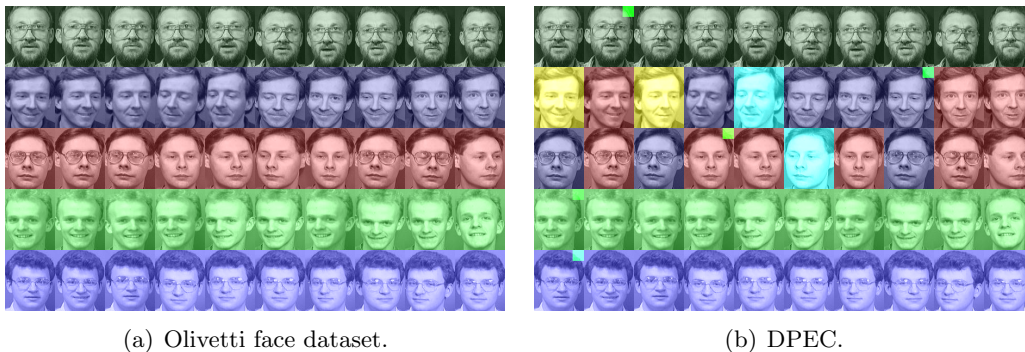


Fig. 6.2. Clustering results by DPEC on the Olivetti faces dataset.

Fig. 6.2 shows the Olivetti faces dataset and the clustering result by DPEC. The clustering results of comparison methods are included in **Appendix (B)**. Table 6.2

reports the clustering results of different methods on this real dataset. Based on these results, we can find that the proposed DPEC method has potential in unsupervised image classification. In Fig. 6.2, the real cluster centers are labeled with green rectangles, and DPEC can find them correctly. Furthermore, we see that DPEC can cluster the images of the first, the fourth, and the fifth person accurately. For some images of the second and the third person, DPEC can cautiously assign these imprecision images, labeled with yellow and azure, to proper meta-clusters, respectively. Once these imprecise images are assigned to meta-clusters, the current knowledge cannot classify them exactly, and forcing the assignment only increases the risk of errors.

Table 6.2: Clustering results of different methods on the Olivetti face dataset

Dataset	Methods	$R_e$	$R_i$	$P$	$R$	$RI$	$EP$	$ER$	$ERI$
Olivetti faces	DP	0.3000		0.6667	<b>0.8889</b>	0.8980	0.6667	<b>0.8889</b>	0.8980
	MC	0.3600		0.5140	0.6533	0.8229	0.5140	0.6533	0.8229
	CDPC	0.2600		0.6715	0.8178	0.8931	0.6715	0.8178	0.8931
	ECM	0.2200	0.0400	0.7544	0.7644	0.9110	0.6930	0.7198	0.8901
	MS-type	0.2400		0.6771	0.8667	0.8996	0.6771	0.8667	0.8996
	BPEC	0.5600	0.0000	0.2047	0.6533	0.4702	0.2047	0.6533	0.4702
	DPC-DBFN	0.3400	0.0000	0.5923	0.6844	0.8555	0.5923	0.6844	0.8555
	TCASP	0.4200		0.5486	0.7022	0.8392	0.5486	0.7022	0.8392
	DPEC	<b>0.1200</b>	0.0800	<b>0.7588</b>	0.7689	<b>0.9127</b>	<b>0.8438</b>	0.8438	<b>0.9420</b>

In Table 6.2, the two indexes  $R$  and  $ER$  of DPC and MS-type are higher than those of DPEC, but this does not indicate that they outperform DPEC. In the definitions of  $R$  and  $ER$ , their values are higher when the number of image pairs simultaneously assigned to the same clusters is higher. In fact, we can find that DPC and MS-type do not separate the second and third person, which means a cluster is not detected at all. Furthermore, we can find that the error rate ( $R_e$ ) of DPC and MS-type is much larger than that of DPEC, so the classification performance of DPEC on the Olivetti faces dataset is better than that of the comparison methods including DPC and MS-type.

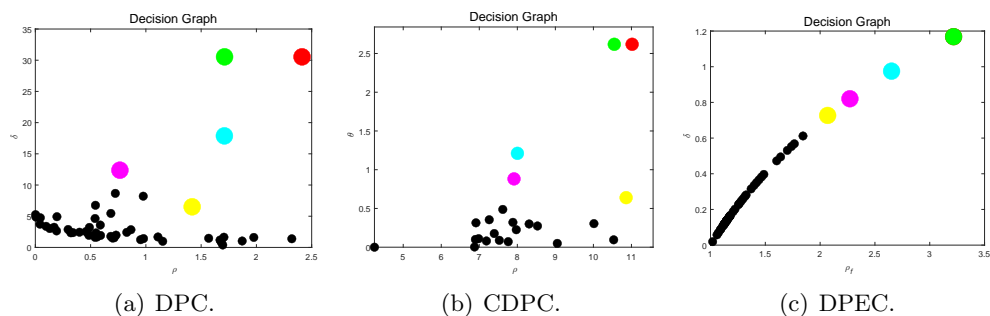


Fig. 6.3. Decision graphs of DPC, CDPC, and DPEC on the Olivetti face dataset.

Moreover, to illustrate the advantages of DPEC in selecting cluster centers, we draw the decision graphs of DPC, CDPC, and DPEC based on the Olivetti faces dataset, as

shown in Fig. 6.3. It is well known that whether an object  $\mathbf{x}_i$  is selected as a cluster center or not is determined by its corresponding  $\gamma_i = \rho_i \delta_i$  in DPC ( $\gamma_i = \rho_i \theta_i$  in CDPC, and  $\gamma_i = \rho_f \delta_i$  in DPEC). So, if  $\mathbf{x}_i$  is a real cluster center, it is expected to have larger  $\rho_i$  and  $\delta_i$  values at the same time. We can see from Figs. 6.3(a),(b) that DPC and CDPC don't satisfy this case because they do not consider the distribution diversity of different clusters when calculating the density  $\rho_i$ . Thus, for DPC and CDPC, the real center object may also has a small  $\rho_i$ . Besides, the relative distance  $\theta$  defined by CDPC may not apply to this case either, as shown in Fig. 6.3(b), where some real center objects do not have a relatively large  $\theta_i$  value. As a result, the cluster centers selected by DPC and CDPC are not reasonable, as shown in 7.10(b),(d). DPEC overcomes these limitations. In DPEC, we define  $\rho_f = e^{\delta_i}$  because a real center object always has a relatively large  $\delta_i$  value while defining  $\rho_f$  as a function of  $\delta_i$  has significantly better robustness. In this case, we can find that the center objects in DPEC always have larger  $\rho_f$  and  $\delta_i$  values, *i.e.* larger  $\gamma_i = \rho_f \delta_i$  values, as shown in Fig. 6.3(c). This helps us to find the real center objects. Overall, the proposed DPEC can obtain more robust cluster centers, as shown in Fig. 6.2(b) and Fig. 6.3(c).

#### 6.4.4 Natural data

In the experiment, we conduct a clustering task to evaluate the performance of DPEC in image segmentation on a natural image, *i.e.* Berkeley Segmentation (BS) dataset (Martin et al., 2001). The BS dataset is composed of bird and background, the resolution ratio of which is  $182 \times 134$ , as shown in Fig. 6.4(a). Besides, Fig. 6.4 shows the ground truth. Here we take  $r = 0.5$  in MS-type,  $\alpha = 5$  and  $\delta = 5$  in ECM,  $K = 250$  in BPEC,  $K = 250$  in DPC-DBFN and  $\mathcal{K}_1 = 800$ ,  $\mathcal{K}_2 = 12$ ,  $\zeta = 0.1$  in DPEC. The other parameters of different methods are defaults.

Table 6.3: Clustering results of different methods on the BS dataset

Dataset	Methods	$R_e$	$R_i$	$P$	$R$	$RI$	$EP$	$ER$	$ERI$
BS	DPC	0.0217	\	0.9633	0.9834	0.9576	0.9633	0.9834	0.9576
	MC	0.0208	\	0.9673	0.9814	0.9593	0.9673	0.9814	0.9593
	CDPC	0.0433	\	0.9174	0.9829	0.9171	0.9174	0.9829	0.9171
	ECM	0.0184	0.0044	0.9670	0.9816	0.9593	0.9695	0.9848	0.9637
	MS-type	0.0211	\	0.9627	0.9855	0.9587	0.9627	0.9855	0.9587
	BPEC	0.0147	0.0128	0.9667	0.9823	0.9595	0.9755	0.9878	0.9707
	DPC-DBFN	0.0163	0.0072	0.9667	0.9823	0.9595	0.9747	0.9847	0.9678
	TCASP	0.0227	\	0.9564	<b>0.9885</b>	0.9557	0.9564	0.9885	0.9557
	DPEC	<b>0.0049</b>	0.0446	<b>0.9786</b>	0.9716	<b>0.9611</b>	<b>0.9941</b>	<b>0.9932</b>	<b>0.9897</b>

Figs. 6.4(c),(d) shows the clustering results by DPC and DPEC. The clustering results of comparison methods are included in **Appendix (C)**. It is intuitive from Fig. 6.4(a) that the tail of the first bird is not very different from the background pixels, making it difficult to segment these parts accurately. However, the comparison methods, *i.e.* DPC, MC, MS-type, DPC-DBFN, and TCASP, all incorrectly assign

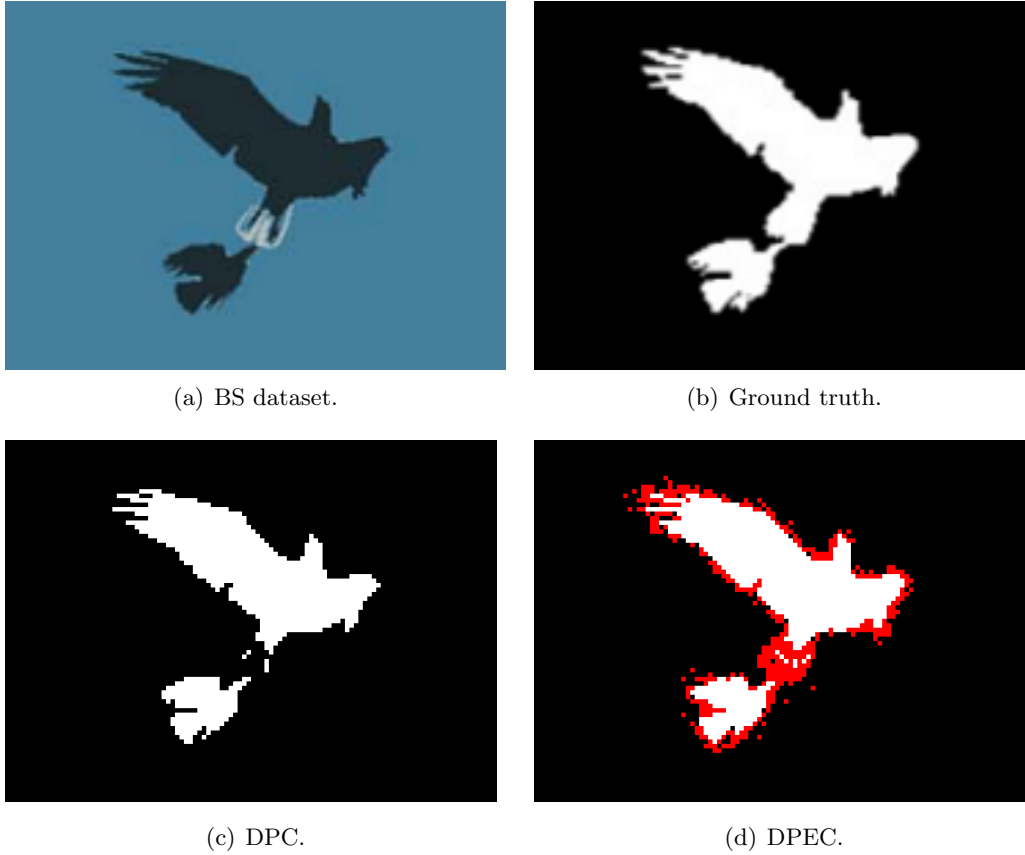


Fig. 6.4. Clustering results by DPC and DPEC on the BS dataset.

the tail part of the bird to the background, as shown in Figs. 7.11(c),(d),(g),(i),(j). In particular, CDPC can no longer detect the birds' silhouettes, as shown in Figs. 7.11(e). In this case, the methods based on the TBF, *i.e.* ECM, BPEC, DPEC, can characterize this imprecision caused by the only small difference between the birds and background (the local overlap of image pixel values) to reduce the risk of incorrect segmentation, as shown in Fig. 6.4(d) and Figs. 7.11(f),(h). However, ECM and BPEC are objective function-based methods, so they can only partially characterize the imprecision and incorrectly assign the bird's tail part to the background. In contrast, DPEC can reasonably assign indistinguishable pixels to the meta-cluster to reduce clustering errors. Moreover,  $\zeta$  can control the number of pixels assigned to the meta-cluster in the DPEC method. Table 6.3 reports the clustering results of different methods on the BS dataset. We can see that the error rate ( $R_e$ ) of the DPEC method is much lower than that of comparison methods, and the evidential indexes, *i.e.*  $EP$ ,  $ER$ , and  $ERI$ , are significantly higher than that of comparison methods. This also indicates that it is reasonable to assign those imprecise pixels to the meta-cluster.



## 6.5 Discussions

### 6.5.1 Complexity analysis

Let us consider that there is a dataset  $\mathcal{X}$  containing  $n$  objects under the FoD  $\Omega = \{\omega_1, \dots, \omega_c\}$ . The proposed DPEC method consists of two steps. In the first step, each object should find  $\mathcal{K}_1$  neighbors from  $\mathcal{X}$ . The complexity of this process is  $\mathcal{O}(n^2)$ . In the second step, there are  $n-c$  remaining objects. Each object should find  $\mathcal{K}_2$  neighbors from the already assigned objects. The complexity of this process is  $\mathcal{O}\left(\sum_{i=c}^{n-1} i\right)$ . This can be simplified to  $\mathcal{O}(n^2)$  and the simplification process is presented in (6.11). Therefore, the total computational complexity of DPEC is  $\mathcal{O}(n^2 + n^2) = \mathcal{O}(n^2)$ .

$$\begin{aligned}
 \mathcal{O}\left(\sum_{i=c}^{n-1} i\right) &= \mathcal{O}(c + (c+1) + \dots + n-1) \\
 &= \mathcal{O}\left(\frac{(n+c-1) \times (n-c)}{2}\right) \\
 &= \mathcal{O}\left(\frac{n^2 - c^2 - n + c}{2}\right) \\
 &= \mathcal{O}(n^2)
 \end{aligned} \tag{6.11}$$

Moreover, to compare the execution times of the different methods more visually, Table 6.4 shows the execution time in seconds of DPEC and comparison methods on the Olivetti face and Berkeley Segmentation datasets. We can see from Table 6.4 that the execution time of DPEC is higher than MC, ECM, MS-type, BPEC, DPC-DBFN, but lower than DPC, CDPC, TCASP, etc. This indicates that the proposed DPEC method has a relative execution advantage over DPC and its improved algorithms. However, DPEC is still time-consuming compared to traditional methods. The reason is that finding neighbors based on the entire dataset is inefficient. Therefore, the proposed DPEC method is more suitable for applications requiring high classification accuracy and the ability to characterize imprecision between different clusters, while efficient computation is not a vital requirement. In the future, we will try to combine with other KNN techniques to reduce the reduction range and thus the computational complexity.

### 6.5.2 Parametric sensitivity

There are three hyper-parameters  $\mathcal{K}_1$ ,  $\mathcal{K}_2$  and  $\zeta$  involved in the proposed DPEC method. In most cases, the number of nearest neighbors is considered an open problem and needs to be determined based on the application. To verify the robustness of DPEC to the number of nearest neighbors, we set  $\mathcal{K}_1$  and  $\mathcal{K}_2$  in  $\{200, 250, 300, 350, 400\}$  and  $\{3, 6, 9, 12, 15\}$  to study the sensitivity of  $\mathcal{K}_1$  and  $\mathcal{K}_2$  on the SD1 dataset, respectively.

Table 6.4: Execution time on the Berkeley Segmentation and Olivetti face datasets

Dataset	Methods	Time(s)	Dataset	Methods	Time(s)
BS	DPC	235.8706	Olivetti faces	DPC	5.627320
	MC	2.716033		MC	0.594996
	CDPC	133.5187		CDPC	0.975217
	ECM	4.325599		ECM	0.276391
	MS-type	4.881048		MS-type	0.202248
	BPEC	9.621040		BPEC	0.527620
	DPC-DBFN	8.978732		DPC-DBFN	0.455501
	TCASP	122.2539		TCASP	4.541309
	DPEC	26.76184		DPEC	0.633895

In the experiments,  $\mathcal{K}_1 = 0.06n$ ,  $\mathcal{K}_2 = 12$ , and  $\zeta = 0.1$  are the defaults. Afterward, we also set  $\zeta$  in  $\{0, 0.03, 0.06, 0.09, 0.12\}$  to study the imprecision of DPEC on the SD1 dataset. The clustering results of DPEC on the SD1 dataset is presented in **Appendix (D)**. Table 6.5 reports the evaluation criteria of DPEC with different  $\mathcal{K}_1$ ,  $\mathcal{K}_2$  and  $\zeta$  on the SD1 dataset. From these results, we can see that DPEC is robust to  $\mathcal{K}_1$ , which is consistent with our previous analysis. For  $\mathcal{K}_2$ , it is used to assign the remaining objects in the evidential convergence rule, so as  $\mathcal{K}_2$  increases, we can find more and more different clusters are included in  $\mathcal{M}_t$ . This means that local imprecision increases as  $\mathcal{K}_2$  increases, so the error rate  $R_e$  decreases, and the imprecision rate  $R_i$  gradually increases. However, too large  $\mathcal{K}_2$  may introduce noise. Interestingly, DPEC is not too sensitive to the change of  $\mathcal{K}_2$  because DPEC controls  $R_i$  by adjusting  $\zeta$ . We can find that as  $\zeta$  increases, more and more objects are assigned to meta-clusters, and thus the imprecision rate increases while the error rate decreases. It indicates that  $\zeta$  can effectively control the number of objects assigned to meta-clusters. In particular, when  $\zeta = 0$ , no objects are assigned to meta-clusters, and then DPEC degenerates to the DPC algorithm. At this point, each remaining object is assigned to a specific cluster with a maximum mass of belief. Although this increases the risk of misclassification, it is suitable for cases that do not allow imprecise results. Thus,  $\zeta$  is adjusted by users according to the acceptable imprecision rate.

### 6.5.3 Comparisons of proposed algorithms

To present the differences and connections of the three proposed algorithms, *i.e.* DEC, BSC, DPEC, in this thesis more intuitively, we have done some experiments based on four given datasets. The parameters of the proposed algorithms are set to the defaults. Figs. 6.5-6.8 show the clustering results of DEC, BSC, DPEC on the given four datasets, respectively. Since these algorithms are proposed to characterize the uncertainty and imprecision between different clusters, we can find that the performance of these algorithms on each dataset is different. Specifically, DEC is used to reduce the

complexity of traditional evidential clustering. BSC aims to detect imbalanced clusters, while DPEC can detect clusters with arbitrary shapes, sizes, and dimensionality.

More precisely, Fig. 6.5 shows the clustering results of DEC, BSC, and DPEC on spherical clusters. We can find that the proposed algorithms effectively detect the cluster structure and characterize the uncertainty and imprecision between spherical clusters. Although BSC and DPEC deal with imbalanced and arbitrary clusters, respectively, they can also deal with spherical clusters. It is important to note that DEC performs more efficiently when dealing with spherical clusters.

Fig. 6.6 shows the clustering results of DEC, BSC, and DPEC on imbalanced clusters. We can find that DEC cannot effectively detect these imbalanced clusters and, subsequently, cannot characterize the uncertainty and imprecision between these clusters. In contrast, BSC and DPEC can effectively detect these imbalanced clusters, which is our motivation for proposing the BSC algorithm. DPEC, as the final generalization scheme, can naturally effectively detect the imbalanced clusters and characterize the uncertainty and imprecision between these imbalanced clusters.

Fig 6.7 shows the clustering results of DEC, BSC, and DPEC on arbitrary clusters, and we can find that neither DEC nor BSC can effectively detect these arbitrary clusters. In contrast, DPEC can effectively detect arbitrary clusters and characterize the uncertainty and imprecision between these arbitrary clusters, which is the desired result. In summary, our proposed DEC, BSC, and DPEC can achieve the given results and gradually generalize the constraints on clusters of shapes, sizes, and dimensionality. In other words, this thesis provides alternatives for characterizing uncertainty and imprecision between clusters, which is also verified on the real dataset, as shown in Fig 6.8. Furthermore, we need to state that the proposed DEC, BSC, and DPEC can control the number of imprecise objects in meta-clusters and the imprecision rate in results. Therefore, users can set them in applications.

Table 6.5: Clustering results of DPEC with different  $\mathcal{K}_1$ ,  $\mathcal{K}_2$ , and  $\zeta$  on the SD1 dataset

Dataset	Parameters	$R_e$	$R_i$	$P$	$R$	$RI$	$EP$	$ER$	$ERI$
SD1	$\mathcal{K}_1 = 200$	0.0000	0.0179	0.9854	0.9855	0.9950	1.0000	1.0000	1.0000
	$\mathcal{K}_1 = 250$	0.0000	0.0148	0.9858	0.9859	0.9951	1.0000	1.0000	1.0000
	$\mathcal{K}_1 = 300$	0.0000	0.0148	0.9858	0.9859	0.9951	1.0000	1.0000	1.0000
	$\mathcal{K}_1 = 350$	0.0000	0.0141	0.9854	0.9855	0.9950	1.0000	1.0000	1.0000
	$\mathcal{K}_1 = 400$	0.0009	0.0340	0.9847	0.9850	0.9948	0.9981	0.9987	0.9994
	$\mathcal{K}_2 = 3$	0.0041	0.0022	0.9904	0.9905	0.9967	0.9917	0.9918	0.9972
	$\mathcal{K}_2 = 6$	0.0016	0.0104	0.9883	0.9884	0.9960	0.9968	0.9968	0.9989
	$\mathcal{K}_2 = 9$	0.0000	0.0160	0.9849	0.9850	0.9948	1.0000	1.0000	1.0000
	$\mathcal{K}_2 = 12$	0.0000	0.0179	0.9854	0.9855	0.9950	1.0000	1.0000	1.0000
	$\mathcal{K}_2 = 15$	0.0000	0.0242	0.9854	0.9855	0.9950	1.0000	1.0000	1.0000
	$\zeta = 0.00$	0.0069	0.0000	0.9854	0.9855	0.9950	0.9854	0.9855	0.9950
	$\zeta = 0.03$	0.0003	0.0145	0.9854	0.9855	0.9950	0.9993	0.9993	0.9998
	$\zeta = 0.06$	0.0003	0.0170	0.9854	0.9855	0.9950	0.9993	0.9993	0.9998
	$\zeta = 0.09$	0.0000	0.0179	0.9854	0.9855	0.9950	1.0000	1.0000	1.0000
	$\zeta = 0.12$	0.0000	0.0220	0.9854	0.9855	0.9950	1.0000	1.0000	1.0000

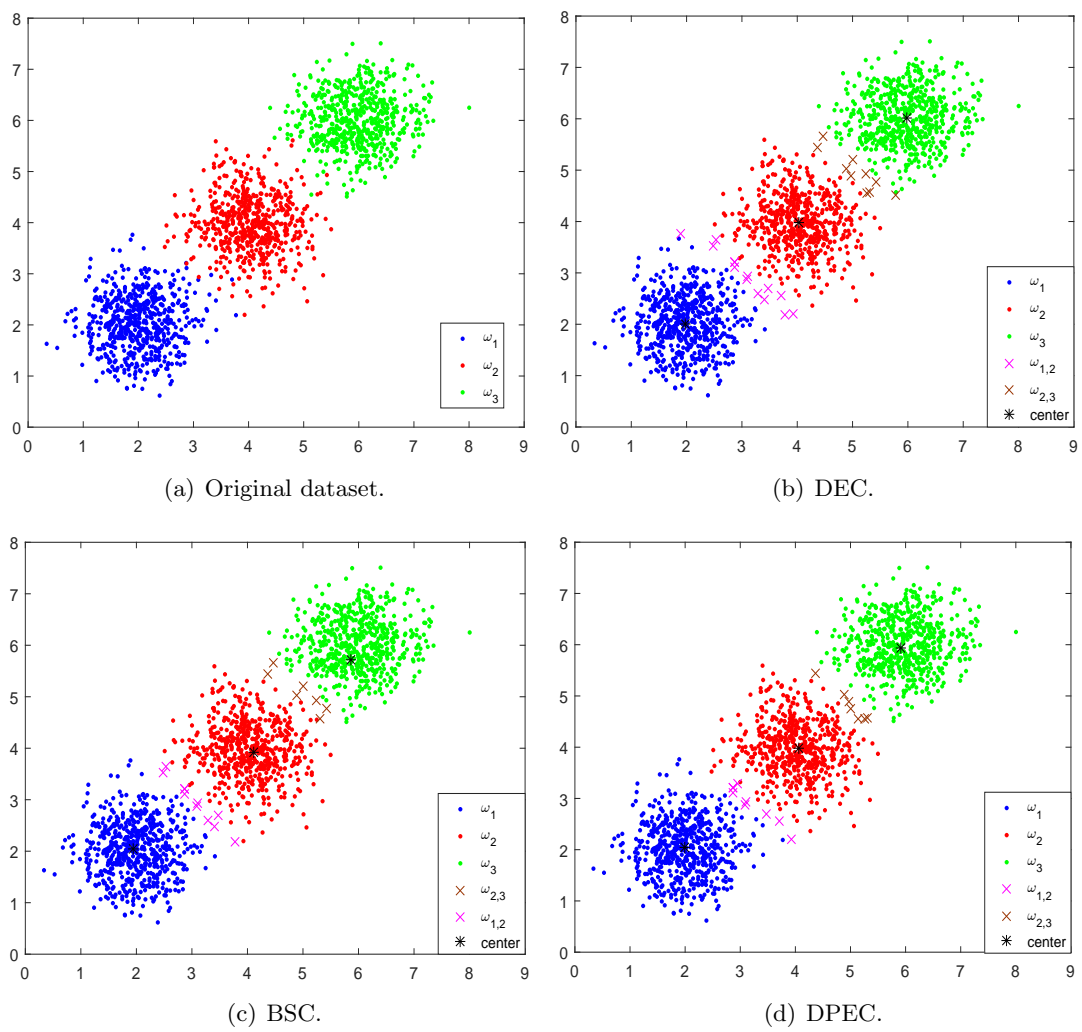


Fig. 6.5. Clustering results of 3-class of gaussian dataset by different methods.

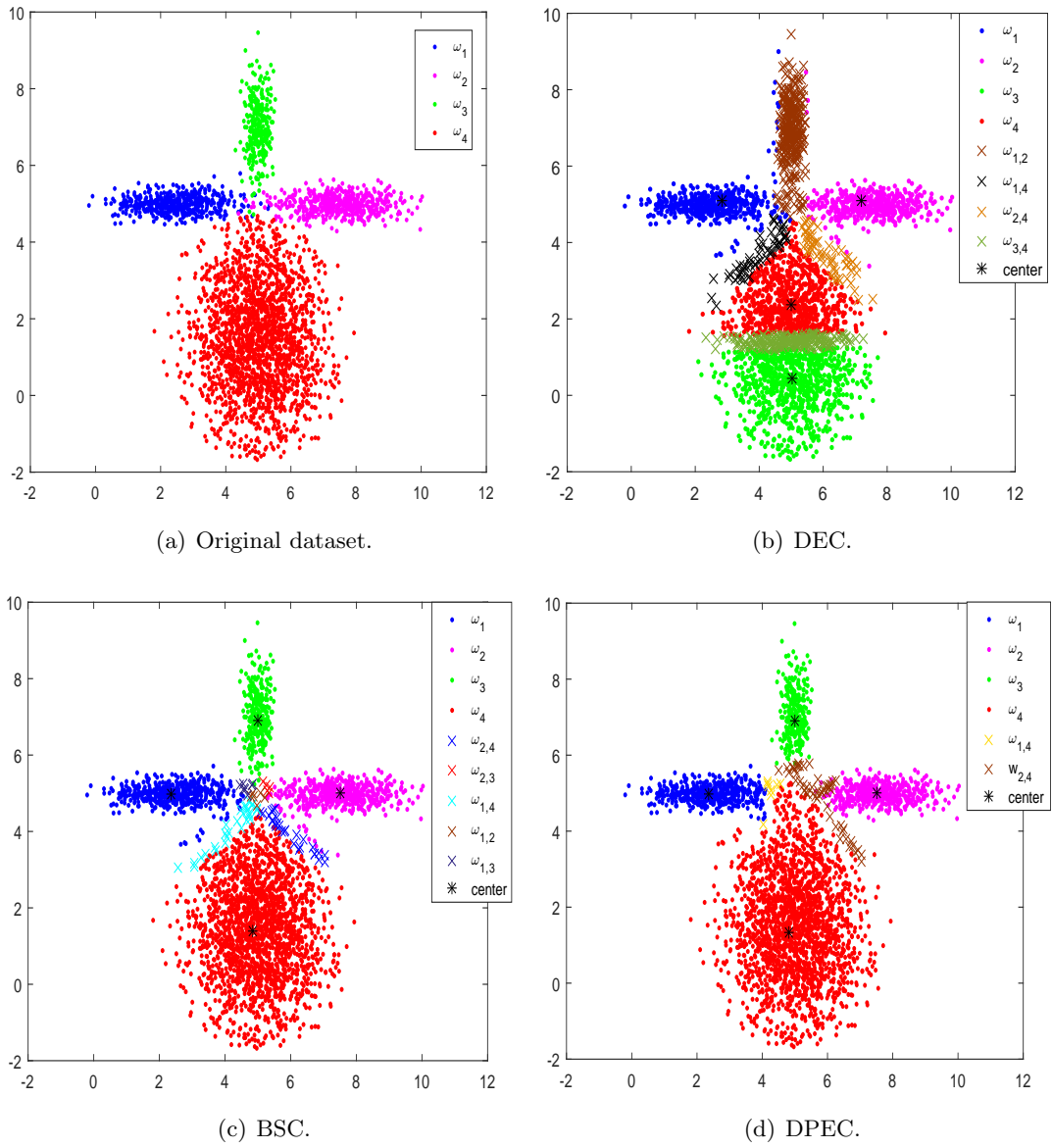


Fig. 6.6. Clustering results of the **SD4** dataset by different methods.

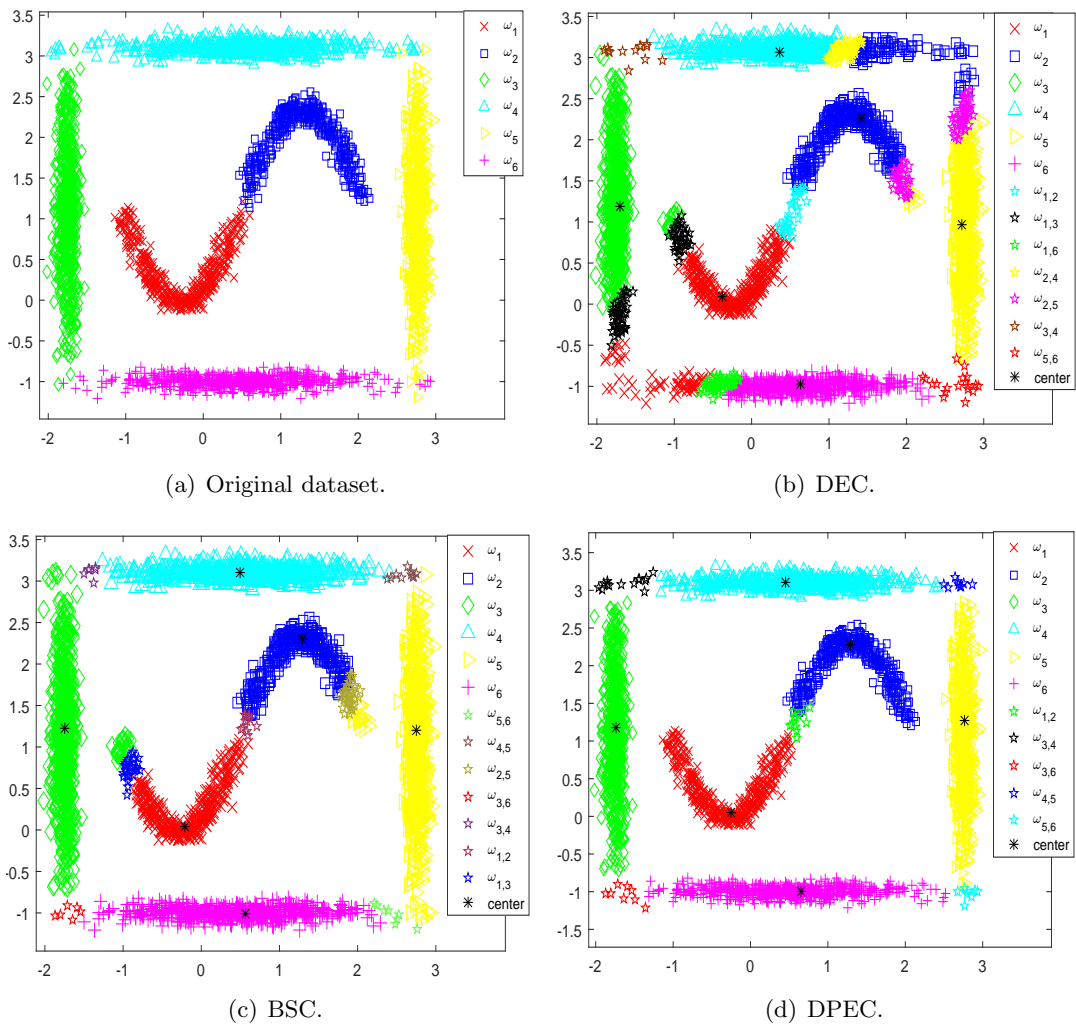


Fig. 6.7. Clustering results of the SD1 dataset by different methods.



(a) Original dataset.



(b) DEC.



(c) BSC.



(d) DPEC.

Fig. 6.8. Clustering results of the BS dataset by different methods.

## 6.6 Conclusion

In this chapter, we investigated a density peak-based evidence convergence (DPEC) algorithm that works to detect clusters of arbitrary shapes and sizes and characterize the uncertainty and imprecision between these clusters. We hope that the proposed DPEC can generalize EC in different scenarios, and the proposed DPEC method consists of two steps. First, we design a reasonable density function to obtain more robust cluster centers. Second, we develop an evidential convergence rule to provide evidence using the assigned objects to aid in assigning the remaining ones. DPEC can reasonably assign those imprecise objects between specific clusters to proper meta-clusters to characterize the imprecision in the results. Doing so can greatly reduce the risk of misclassification. The experimental results show that DPEC is powerful and has a strong generalization ability. Moreover, we also analyzed the computational complexity of DPEC and its sensitivity to involved parameters. In particular, the differences and connections between DPEC and the previously proposed DEC and BSC. The results also show that DPEC is very robust to these parameters. In addition, we argue that DPEC can be degraded to the DPC algorithm if users do not retain imprecision in results.





# 7

## Conclusion and future work

**Abstract:** In this chapter, we conclude all the contents of this thesis and present some ideas for future work.

### Contents

---

7.1	Conclusion . . . . .	110
7.2	Future work . . . . .	111

---

## 7.1 Conclusion

Clustering is a key process in different data mining applications and it offers many potential benefits to a wide-range of fields including data mining, pattern recognition, information retrieval, bioinformatics, and business intelligence. Different requirements and challenges need to be taken into consideration in designing clustering algorithms applied to different tasks. In this thesis, we focused our study on normalizing some basics of evidential clustering (EC) and improving the ability of EC to characterize uncertainty and imprecision between different clusters, especially for imbalanced or clusters with arbitrary shapes and sizes.

With this objective structure in mind, we first normalized the concepts and definitions related to EC and compared the differences and connections of the corresponding clustering algorithms under different theories. Specifically, we first gave the concepts and definitions of data (inputs), methods (models), and clusters (results) of EC, as well as that of the other types (*i.e.* hard, fuzzy, possibilistic ones), based on different theories (*i.e.* probability theory, fuzzy set theory, possibility theory, and the theory of belief functions). According to these concepts and definitions, we analyzed the differences and connections between EC and other types and thus declared the advantages of EC in characterizing uncertainty and imprecision between clusters. Afterward, we also studied EC from the seminal to state-of-the-art methods in the context of data-based inputs, resulting in a coherent and comprehensive survey to analyze these methods. Based on the analysis of some representative techniques from different perspectives (*e.g.* center, measure type, complexity), we provided a guiding scheme to help users choose appropriate methods in their cases.

Moreover, three clustering methods, DEC, BSC, and DPEC were proposed. DEC worked to reduce the computational complexity of traditional EC to extend applications, especially for big data. In DEC, an FCM-like objective function was first employed and minimized to obtain the support levels of the real singletons (specific) clusters to which the query objects belong. Then, the query object was initially adaptively assigned to the outlier, precise or imprecise one *via* a new rule based on the conflicts between the different support levels. Each imprecise object was finally reassigned to the singleton clusters or related meta-cluster by partial credal redistribution with the corresponding dynamic edited framework to reduce the computational burden. The proposed DEC method can reduce the complexity to a level similar to that of fuzzy and possibilistic clustering, which can effectively extend the applications of EC, especially in big data. The effectiveness of the DEC method was tested by different experiments with artificial and real datasets. However, traditional EC can only detect symmetric spherical clusters, which is not the full form of data distribution.

Furthermore, we further proposed the BSC algorithm to detect non-spherical and imbalanced clusters. In BSC, the query object was preliminarily assigned as the noise,

precise, or imprecise one based on the notion of “belief shift”. Then, partial credal redistribution with dynamic cluster centers, to avoid the “uniform effect”, was established to reassign imprecise objects to the singleton cluster or related meta-cluster. Once an object was assigned to a meta-cluster, it indicated that the singleton clusters involved in the meta-cluster cannot be distinguished because this object may be located in the overlapping or intermediate areas of these singleton clusters. In this case, the BSC can reasonably characterize the uncertainty and imprecision between imbalanced clusters. The effectiveness of BSC had been validated on several synthetic and real datasets by critically comparing with some related methods.

Finally, we presented a DPEC algorithm that can characterize the uncertainty and imprecision between clusters of arbitrary sizes and shapes. In DPEC, We first considered that different neighbors can provide complementary evidence supporting the object as a cluster center and redefined a distance-based density function to obtain more robust cluster centers in the decision graph. Then, we presented a new evidential convergence rule to assign the remaining objects to different clusters. Similar to BSC, the objects located in the overlapping or intermediate areas of different singleton clusters were assigned to corresponding meta-clusters to characterize the uncertainty and imprecision between these arbitrary clusters. The effectiveness of DPEC was also validated by comparing with some typical techniques.

It can be found that our research is a process of continuous enhancement and improvement based on previous works, and our goal is to propose a flexible EC under the TBF to detect the structure of arbitrary clusters. A series of experiments have shown that our proposed methods can achieve good results.

To conclude, our work on EC aims to improve clustering results by combining EC with other classical methods, providing explanatory information to describe each detected cluster, and characterizing uncertainty and imprecision between these clusters. Based on our proposed algorithms in this thesis, the comparison of different methods and the principles of selection can be updated, as shown in Table 7.1 and Fig. 7.1.

## 7.2 Future work

In this section, we try to develop some possible future works based on the current work that we have already presented in this thesis. We are mostly interested in several directions as below:

**Characterization of overlapping cluster edges.** Compared to traditional clustering methods, the algorithms we proposed in this thesis are able to characterize the uncertainty and imprecision between (arbitrary) overlapping clusters based on mass functions as well as meta-clusters. However, by assigning imprecise objects to meta-clusters, we also destroy the original distributions of overlapping clusters, i.e., the edges of the clusters in the overlapping regions.

Table 7.1: Comparisons of different evidential methods

Method	Data type	Center	Measure type	Cluster number	Complexity	Disadvantage
ECM	Object	Centroids	Euclidean	Known	$\mathcal{O}(n2^c)$	(1), (3), (4)
CCM	Object	Centroids	Euclidean	Known	$\mathcal{O}(n2^c)$	(1), (3), (4)
CECM	Object	Centroids	Adaptive metric	Known	$\mathcal{O}(n2^c)$	(1), (2), (3)
<b>DEC</b>	<b>Object</b>	<b>Centroids</b>	<b>Euclidean</b>	<b>Known</b>	$\mathcal{O}(nc)$	<b>(1), (4)</b>
BPEC	Object	Medoids	Euclidean	Unknown	$\mathcal{O}(n^2 + n2^c)$	(3), (4)
CBP-EKNN	Object	Medoids	Euclidean	Unknown	$\mathcal{O}(n^3)$	(3), (5)
EGMM	Object	Centroids	Covariance matrix	Known	$\mathcal{O}(n2^c)$	(1), (3)
<b>BSC</b>	<b>Object</b>	<b>Centroids</b>	<b>Euclidean</b>	<b>Unknown</b>	$\mathcal{O}(n^2)$	<b>(3)</b>
<b>DPEC</b>	<b>Object</b>	<b>Medoids</b>	<b>Euclidean</b>	<b>Unknown</b>	$\mathcal{O}(n^2)$	<b>(3)</b>
EVCLUS	Relational	/	Dissimilarity matrix	Known	$\mathcal{O}(n^2c^2)$	(1), (3)
CEVCLUS	Relational	/	Dissimilarity matrix	Known	$\mathcal{O}(n^22^c)$	(1), (2), (3), (5)
k-CEVCLUS	Relational	/	Dissimilarity matrix	Known	$\mathcal{O}(n2^c)$	(1), (2), (5)
RECM	Relational	Centroids	Euclidean	Known	$\mathcal{O}(nc^2 + cn^2)$	(1), (3), (4)
MECM	Relational	Medoids	Euclidean	Known	$\mathcal{O}(n2^c)$	(1), (3), (4)
EK-NNclus	Relational	/	Dissimilarity matrix	Unknown	$\mathcal{O}(n^22^c)$	(3)
NN-EVCLUS	Relational	/	Dissimilarity matrix	Known	$\mathcal{O}(n^22^c)$	(1), (2), (3)
ECMdd	Relational	Medoids	Dissimilarity matrix	Known	$\mathcal{O}(cn^2 + n2^c)$	(1), (3), (4)

(1) The users need to know the number of clusters; (2) The users know some prior knowledge about the objects/clusters; (3) The method has a high computational complexity; (4) This method is not applicable to non-spherical symmetric data; (5) This method cannot detect partial ignorance (*e.g.* the meta-cluster  $A_j$ ) among singleton clusters,  $|A_j| \geq 2$ ,  $A_j \subset \Omega$ .

**Characterization of uncertainty and imprecision in missing data.** The causes of uncertainty and imprecision in (unsupervised) classification are diverse, such as missing data. Some original well-distributed clusters/classes may have local overlapping due to missing values. We have conducted researches on this uncertainty and imprecision between clusters/classes due to missing values (Zhang et al., 2021c,b,e). However, it is not established a long-term effective mechanism to generally characterize the uncertainty and imprecision due to missing data in the modeling process and results. Therefore, it is still a challenging task.

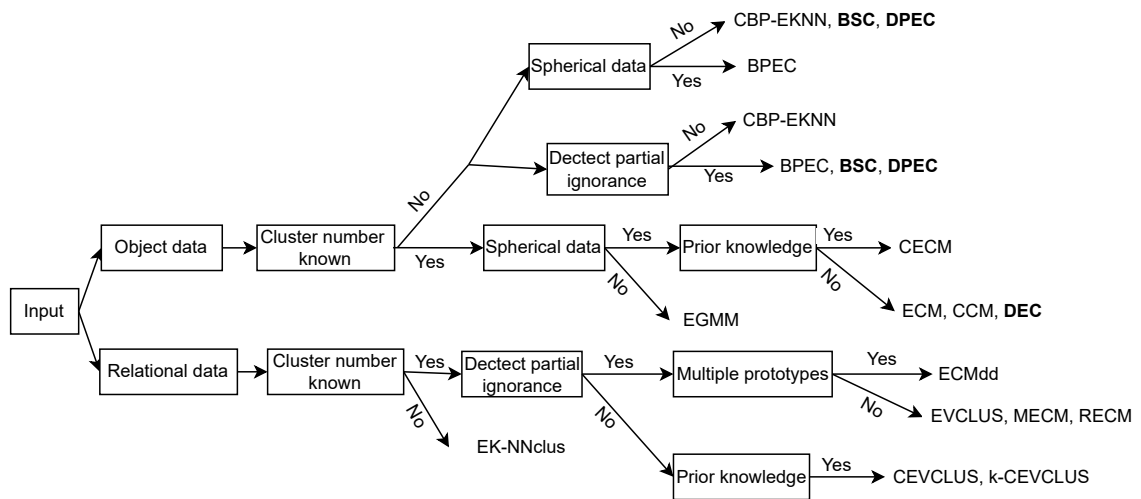


Fig. 7.1. The basic principles for choosing different evidential methods.



# Appendix

## A: The results of comparison methods on the eight datasets

This part is complementary to Subsection 6.4.2. Some hyperparameters need to be set in the comparison methods. In general, we use the defaults provided by comparison methods. However, some parameters should be reset based on the datasets. Here we set those parameters as feasible solutions when obtaining better results. Table 7.2 reports some parameters that need to be reset in comparison methods. Figs. 7.2-7.9 shows the results of comparison methods on the eight datasets. The analysis of these results has been presented in the relevant experimental parts.

Table 7.2: Some given parameters in comparison methods

Datasets	ECM	MS-type	BPEC	DPC-DBFN
SD1	$\alpha = 8, \delta = 8$	$r = 0.9$	$K = 250$	$K = 10$
SD2	$\alpha = 8, \delta = 8$	$r = 1.0$	$K = 250$	$K = 10$
SD3	$\alpha = 8, \delta = 8$	$r = 10$	$K = 250$	$K = 10$
SD4	$\alpha = 8, \delta = 8$	$r = 2.0$	$K = 250$	$K = 10$
SD5	$\alpha = 8, \delta = 8$	$r = 0.9$	$K = 250$	$K = 10$
SD6	$\alpha = 8, \delta = 8$	$r = 1.18$	$K = 250$	$K = 30$
SD7	$\alpha = 8, \delta = 8$	$r = 2.0$	$K = 250$	$K = 30$
SD8	$\alpha = 8, \delta = 8$	$r = 1.3$	$K = 250$	$K = 50$

## B: The results of comparison methods on the face dataset

This part is complementary to Subsection 6.4.3. Fig. 7.10 shows the results of different methods on the Olivetti face dataset. From Fig. 7.10 we can find that DPC and its improved methods, *i.e.*, CDPC, BPEC, DPC-DBFN, TCASP, fail to select the five real (best) cluster centers because they do not consider the diversity of distribution of different clusters. DPC and DPC-DBFN find only four valid centers, and TCASP finds only three. Moreover, the results of these methods are not reasonable. For example, DPC, CDPC, and DPC-DBFN could not classify the second and third person, while BPEC and TCASP could hardly classify these five person. For MC, ECM and MS-type, they are unable to classify some images in two or more classes, and they even



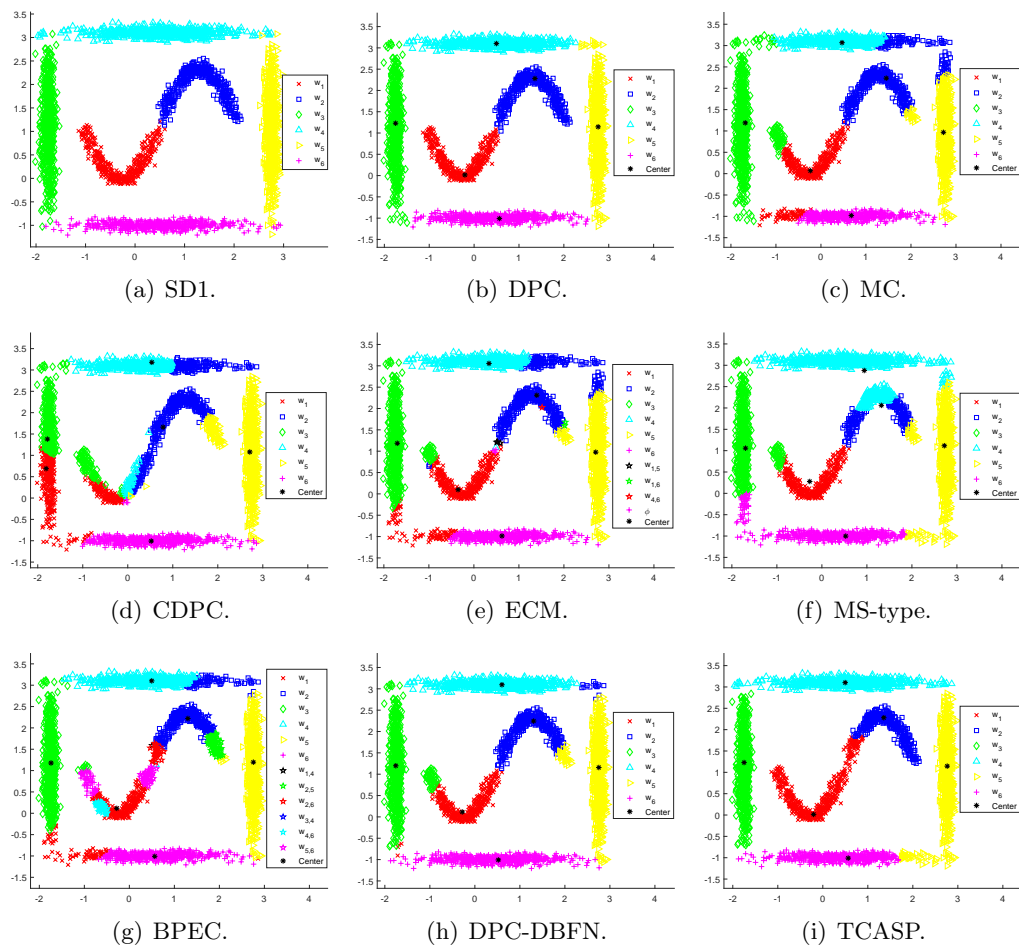


Fig. 7.2. Clustering results by comparison methods on the SD1 dataset.

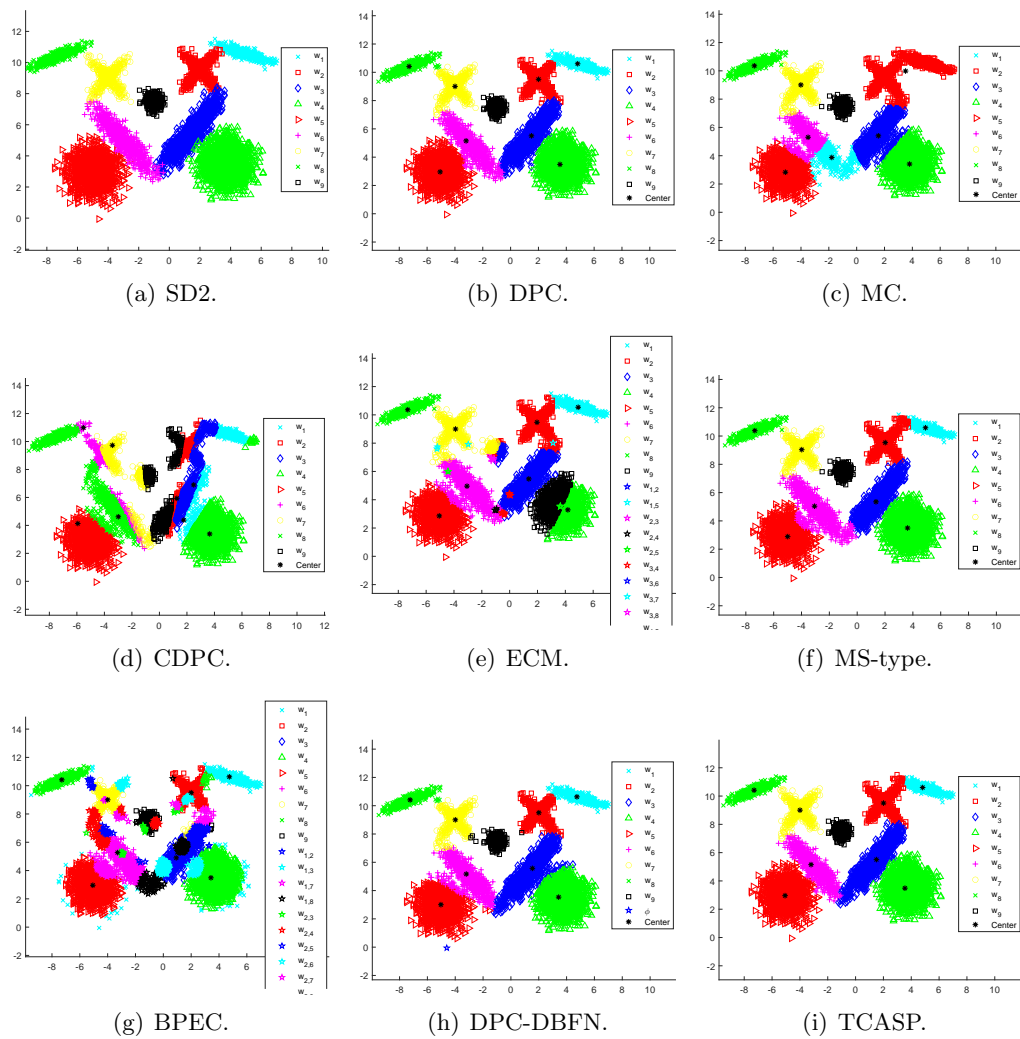


Fig. 7.3. Clustering results by comparison methods on the SD2 dataset.

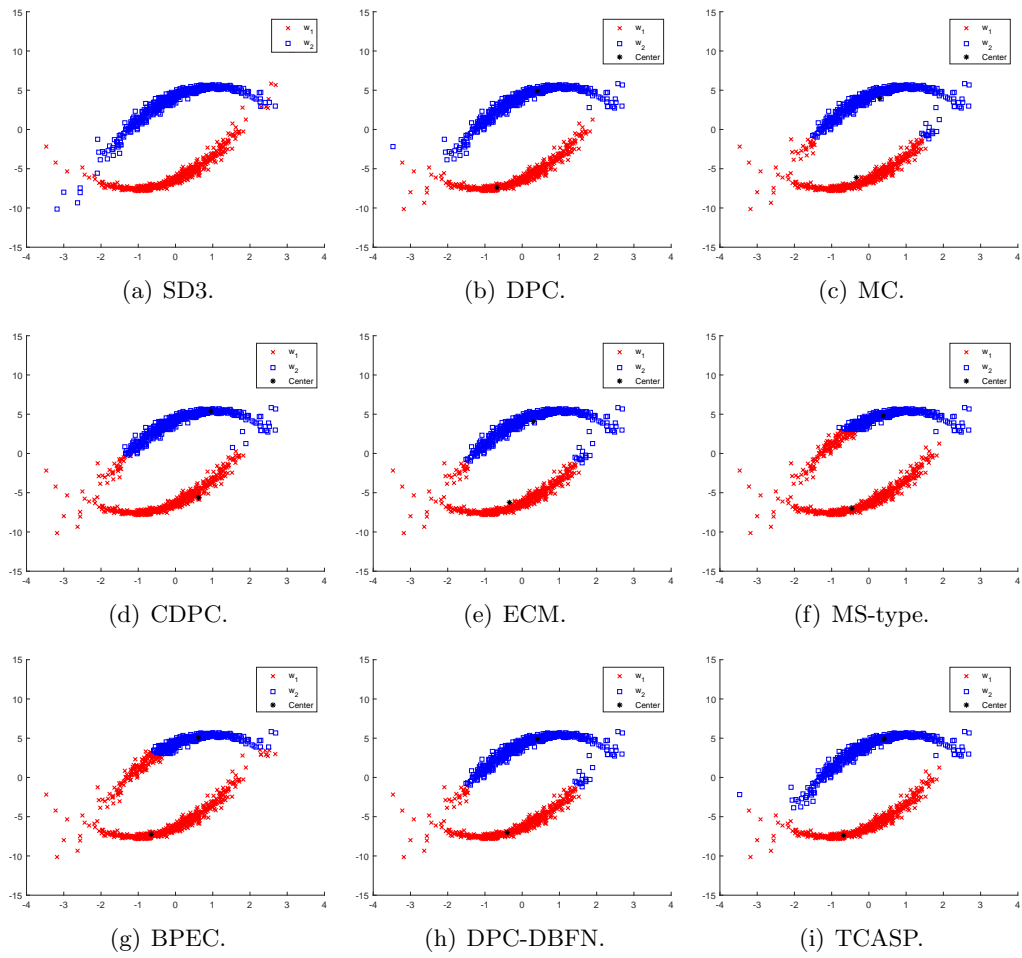


Fig. 7.4. Clustering results by comparison methods on the SD3 dataset.

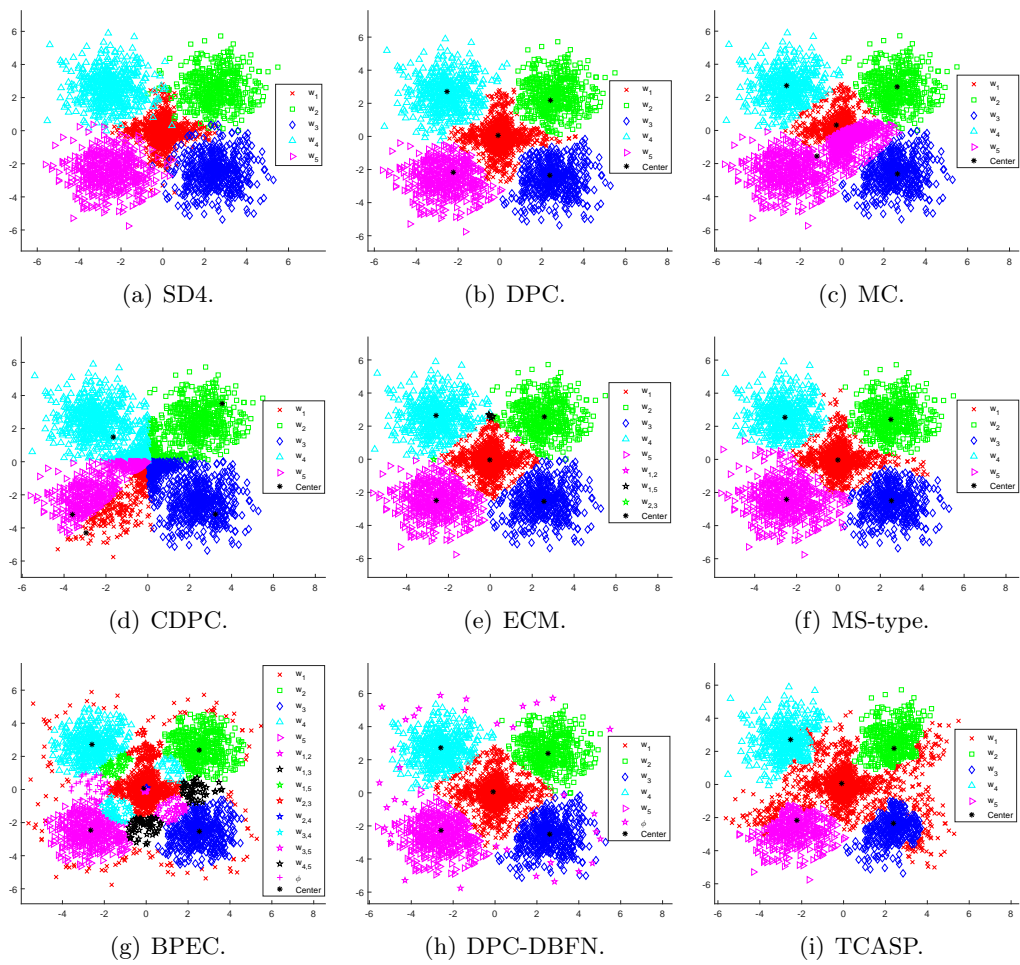


Fig. 7.5. Clustering results by comparison methods on the SD4 dataset.

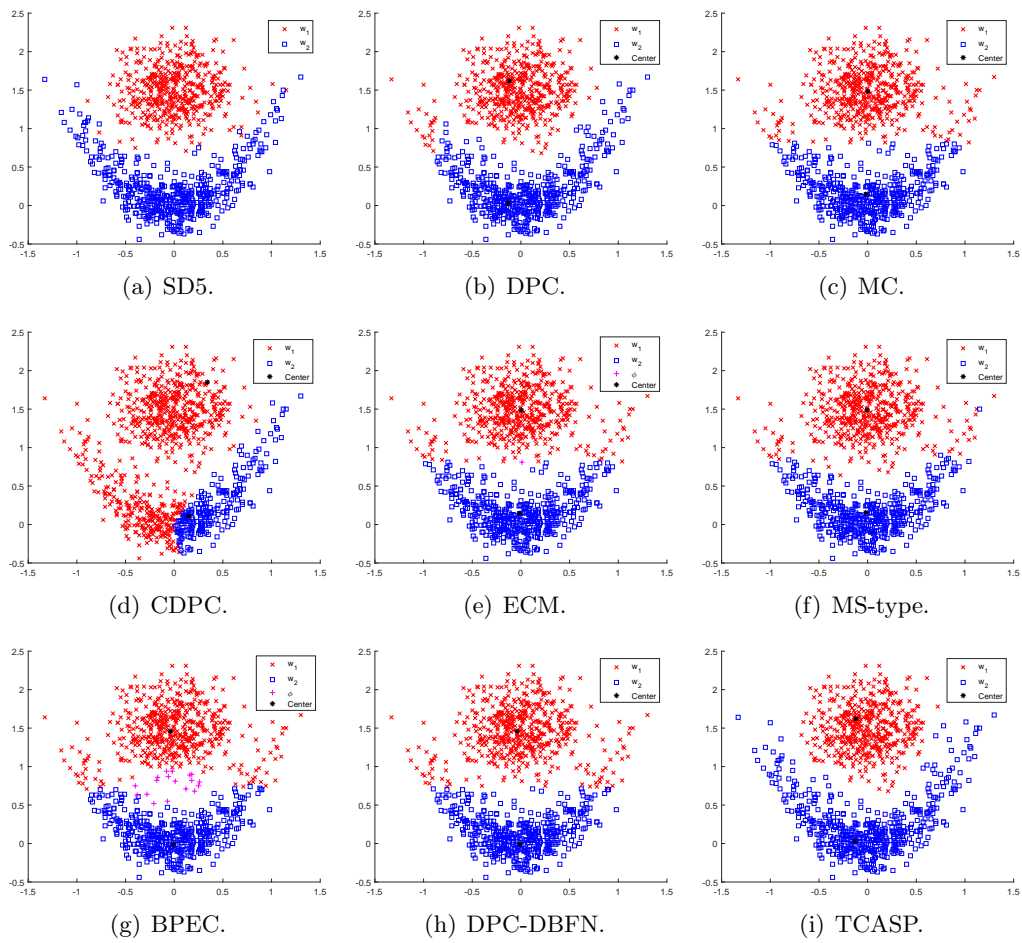


Fig. 7.6. Clustering results by comparison methods on the SD5 dataset.

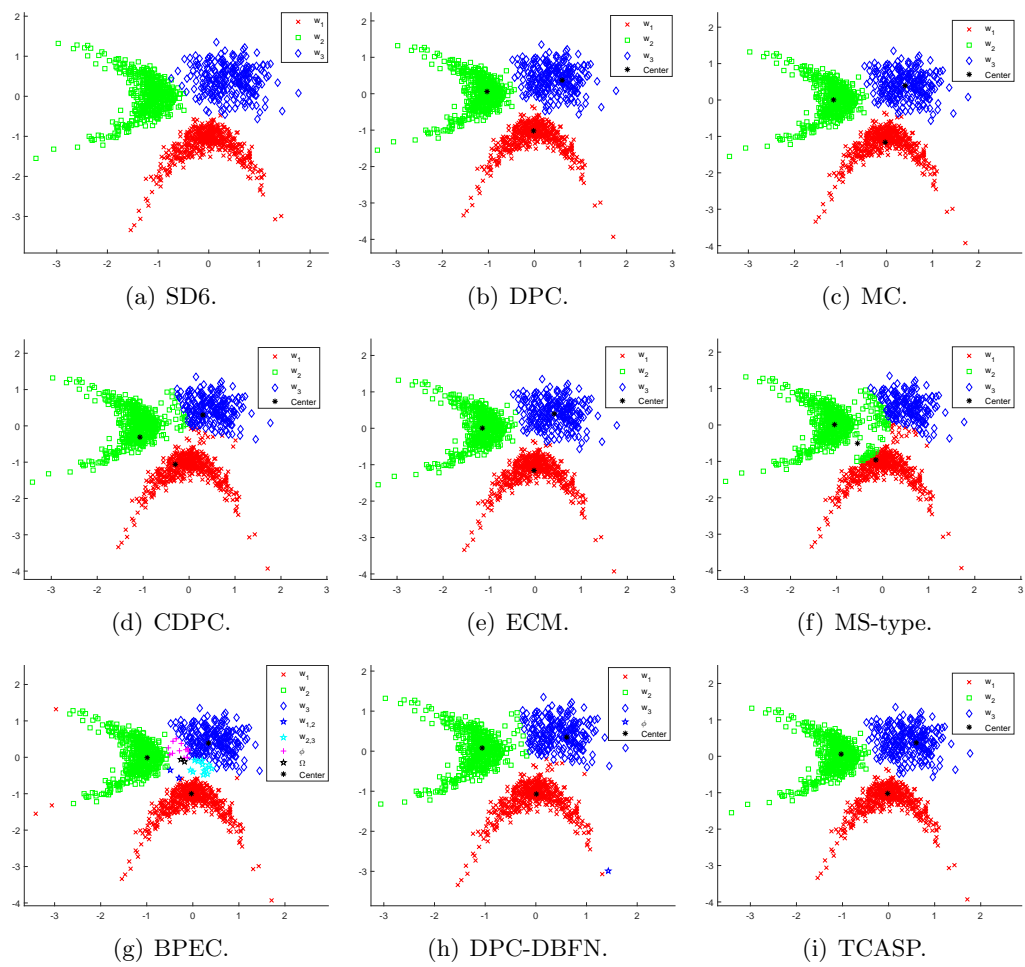


Fig. 7.7. Clustering results by comparison methods on the SD6 dataset.

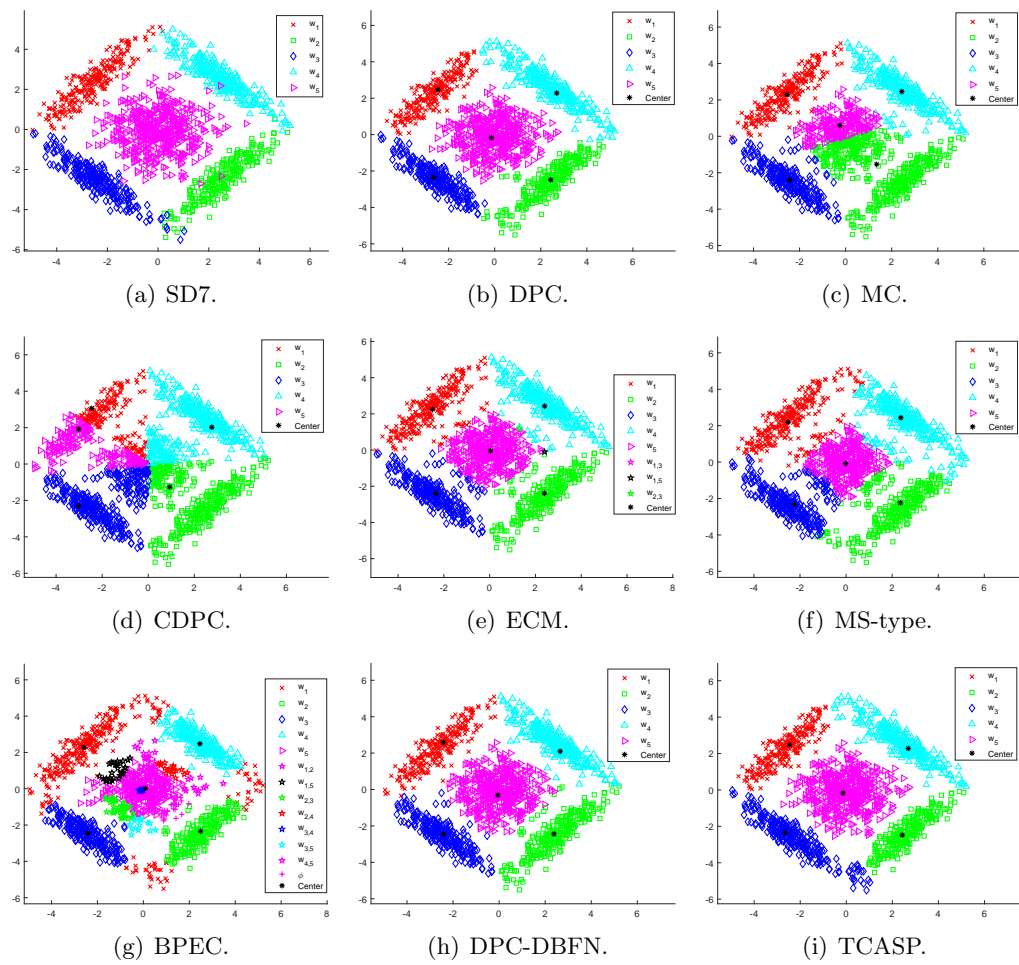


Fig. 7.8. Clustering results by comparison methods on the SD7 dataset.

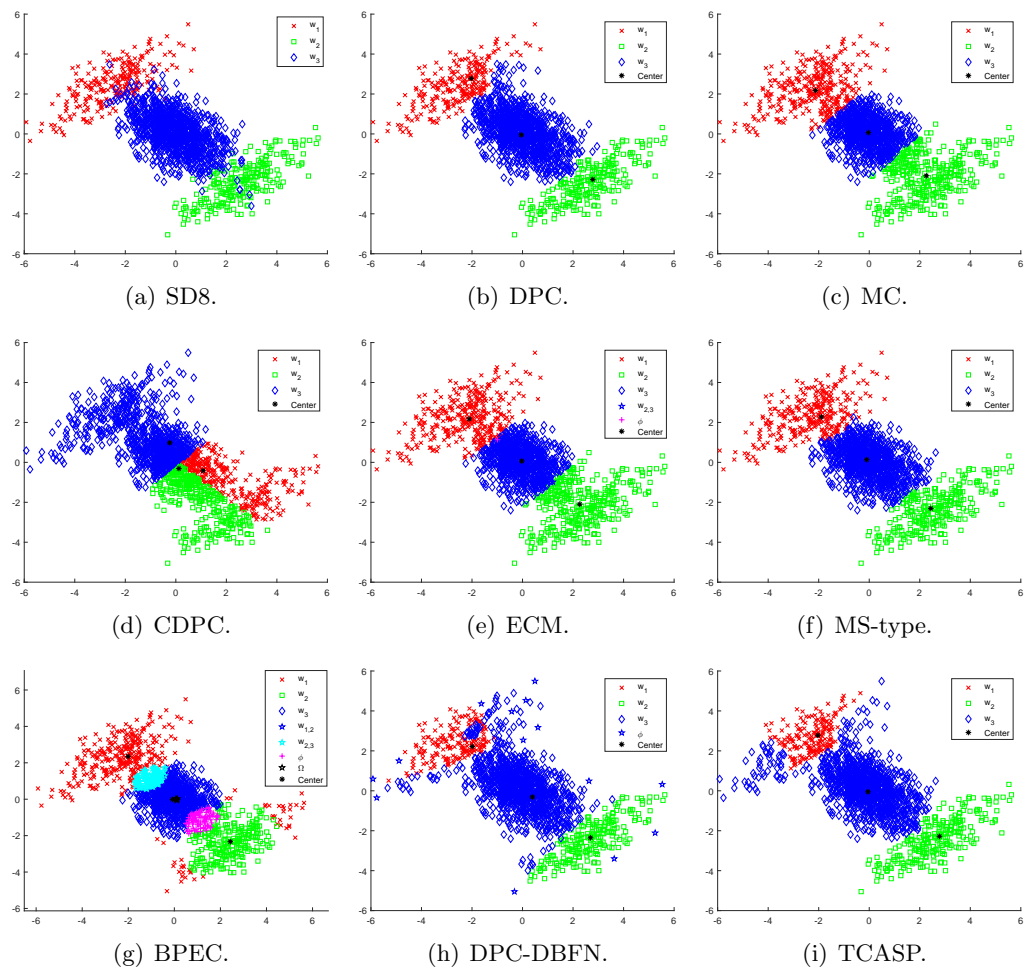


Fig. 7.9. Clustering results by comparison methods on the SD8 dataset.



assign images from different classes to one cluster. In contrast, the proposed DPEC method is able to select the real/reasonable cluster centers and can assign most of the images correctly. For those imprecise images, they are assigned to proper meta-clusters by the evidential convergence rule. Thus, the proposed DPEC method can effectively characterize the imprecision between clusters.

### **C: The results of comparison methods on the natural dataset**

This part is complementary to Subsection 6.4.4. Fig. 7.11 shows the Berkeley Segmentation dataset and the clustering results by different methods on the given dataset.

### **D: The discussions of parametric sensitivity**

This part is complementary to Subsection 6.5.2. Fig. 7.12 shows the clustering results of DPEC on the SD1 dataset when we set  $\mathcal{K}_1 \in \{200, 250, 300, 350, 400\}$ ,  $\mathcal{K}_2 = 12$ , and  $\zeta = 0.1$ . In contrast, Fig. 7.13 shows the clustering results when  $\mathcal{K}_2 \in \{3, 6, 9, 12, 15\}$  with  $\mathcal{K}_1 = 12$  and  $\zeta = 0.1$ . Fig. 7.14 is used to show how DPEC can control the imprecision rate by assigning objects to meta-clusters when setting  $\zeta \in \{0, 0.03, 0.06, 0.09, 0.12\}$ . We can find that DPEC has good robustness to these three hyper-parameters, *i.e.*  $\mathcal{K}_1$ ,  $\mathcal{K}_2$ , and  $\zeta$ .



(a) Olivetti faces dataset.



(b) DPC.



(c) MC.



(d) CDPC.



(e) ECM.



(f) MS-type.



(g) BPEC.



(h) DPC-DBFN.



(i) TCASP.



(j) DPEC.

Fig. 7.10. Clustering results by different methods on the Olivetti face dataset.

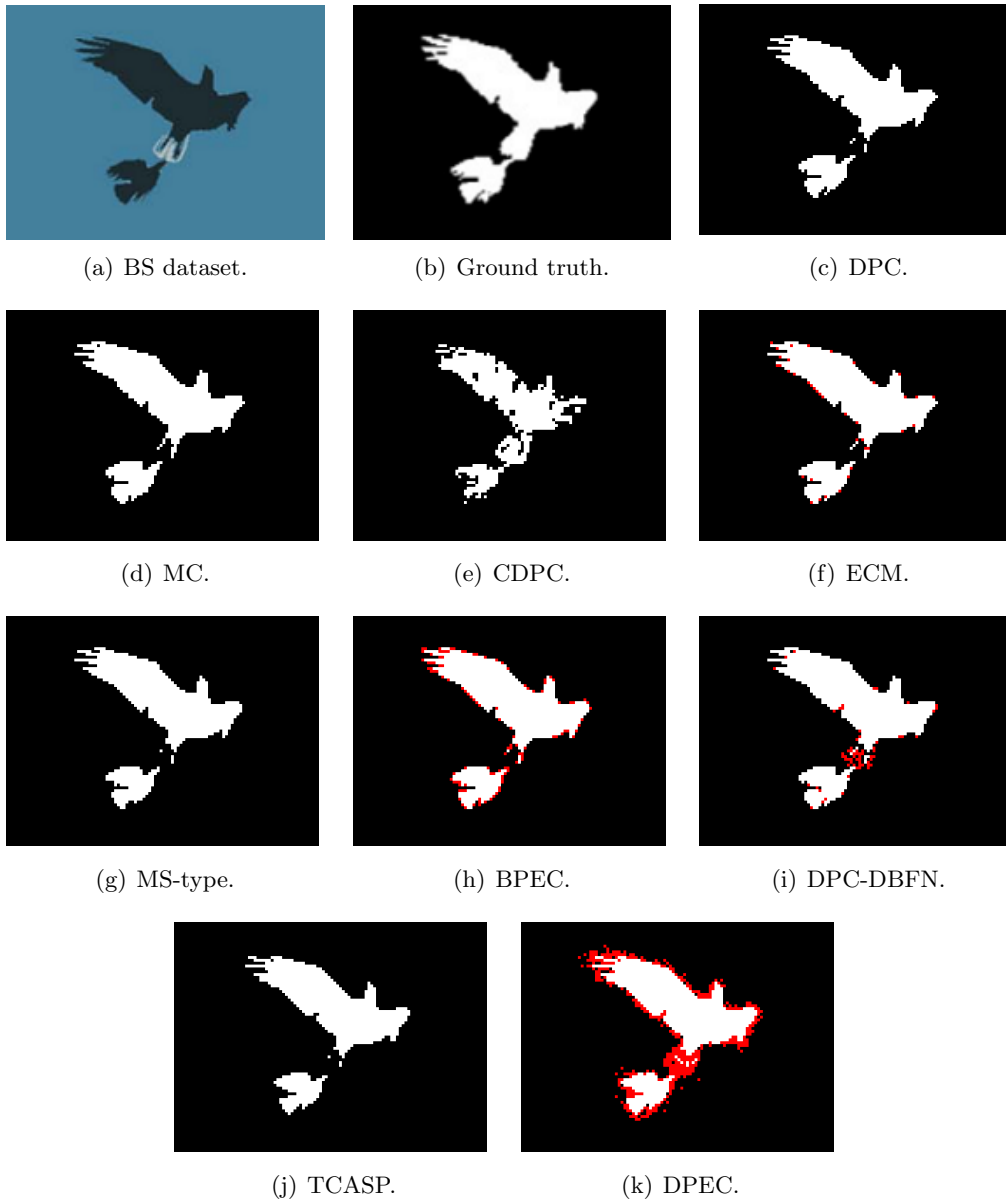


Fig. 7.11. Clustering results by different methods on the BS dataset.

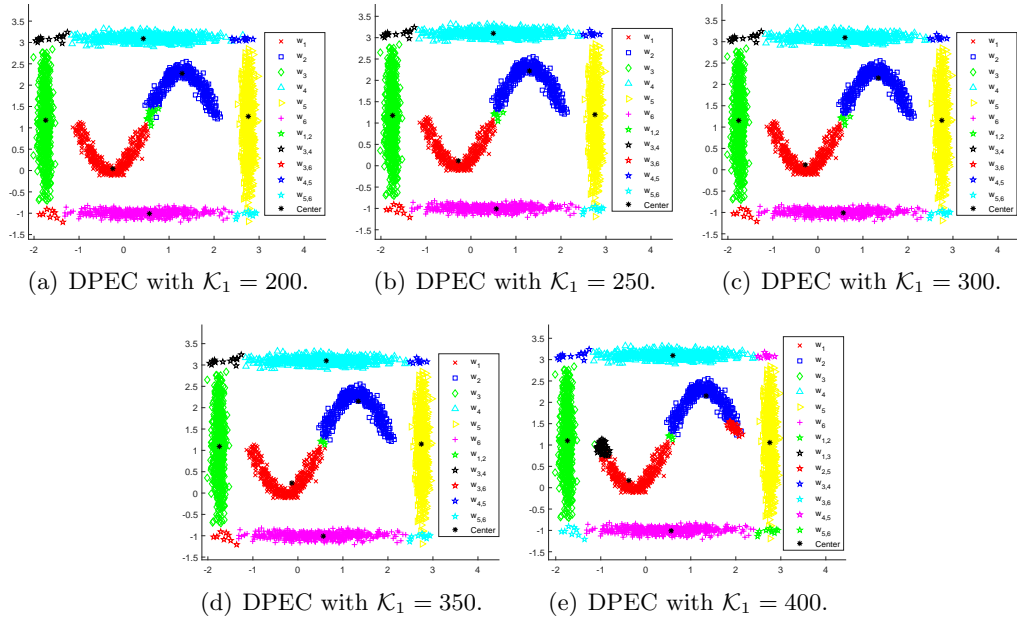


Fig. 7.12. Clustering results by DPEC with different  $\mathcal{K}_1$  on the SD1 dataset.

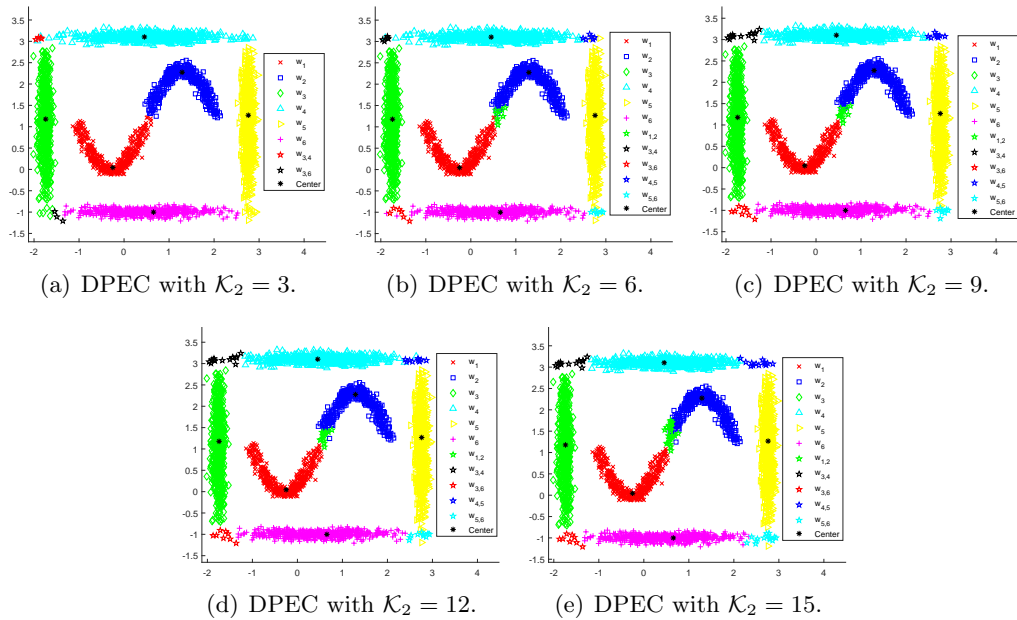


Fig. 7.13. Clustering results by DPEC with different  $\mathcal{K}_2$  on the SD1 dataset.

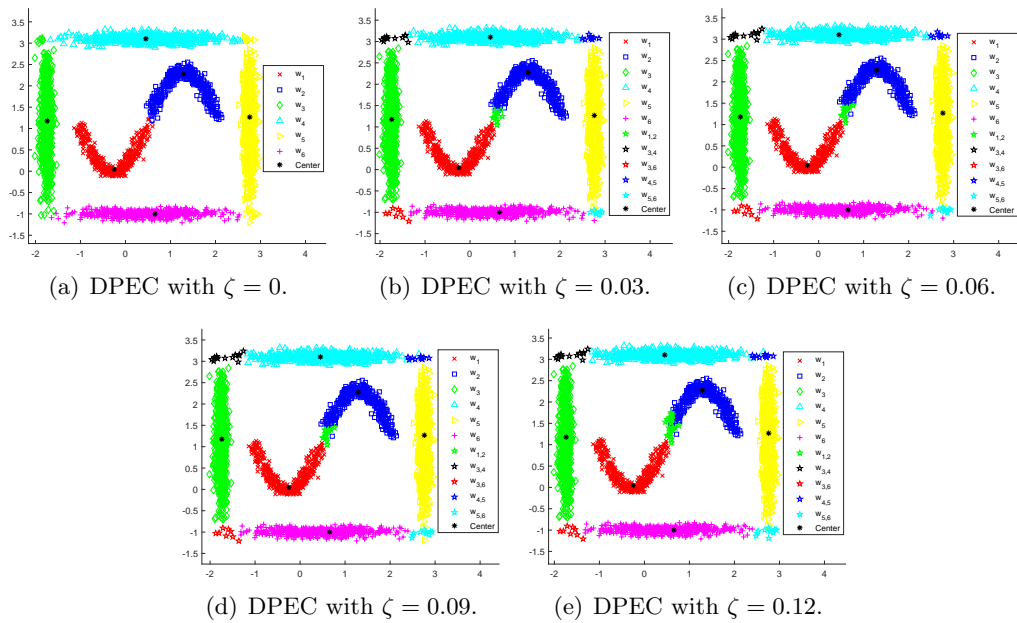


Fig. 7.14. Clustering results by DPEC with different  $\zeta$  on the SD1 dataset.

# Author's Publication List

## International Journals

(1) Zhun-ga Liu, **Zuo-wei Zhang**, Yu Liu, Jean Dezert, Quan Pan. A new pattern classification improvement method with local quality matrix based on K-NN. *Knowledge-based systems*, 164:336-347, 2019. DOI: 10.1016/j.knosys.2018.11.001.

(2) Zong-fang Ma, Hong-peng Tian, Ze-chao Liu, **Zuo-wei Zhang**. A new incomplete pattern belief classification method with multiple estimations based on KNN. *Applied Soft Computing*, 90:106175, 2020. DOI: 10.1016/j.asoc.2020.106175.

(3) **Zuo-wei Zhang**, Zhe Liu, Arnaud Martin, Zhun-ga Liu, Kuang Zhou. Dynamic evidential clustering algorithm. *Knowledge-based systems*, 213:106643, 2021. DOI: 10.1016/j.knosys.2020.106643.

(4) **Zuo-wei Zhang**, Zhe Liu, Zong-fang Ma, Yi-ru Zhang, Hao Wang. A new belief-based incomplete pattern unsupervised classification method. *IEEE Transactions on Knowledge and Data Engineering*, 2021. DOI: 10.1109/TKDE.2021.3049511.

(5) **Zuo-wei Zhang**, Zhe Liu, Zong-fa Ma, Ji-huan He, Xing-yu Zhu. Evidence integration credal classification algorithm versus missing data distributions. *Information Sciences*, 569:39-54, 2021. DOI: 10.1016/j.ins.2021.04.008.

(6) **Zuo-wei Zhang**, Hong-peng Tian, Ling-zhi Yan, Arnaud Martin, Kuang Zhou. Learning a credal classifier with optimized and adaptive multi-estimation for missing data imputation. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, 2021. DOI: 10.1109/TSMC.2021.3090210.

(7) Zhun-ga Liu, **Zuo-wei Zhang**, Quan Pan Liang-bo Ning. Unsupervised Change detection from heterogeneous data based on image translation. *IEEE Transactions on Geoscience and Remote Sensing*, 2021. DOI: 10.1109/TGRS.2021.3097717.

(8) **Zuo-wei Zhang**, Song-tao Ye, Yi-ru Zhang, Wei-ping Ding, Hao Wang. Belief combination of classifiers for missing data. *IEEE-CAA Journal of Automatica Sinica*, 9(4): 652-667, 2022. DOI: 10.1109/JAS.2022.105458.

## Under Review

(8) **Zuo-wei Zhang**, Zhun-ga Liu, Arnaud Martin. BSC: Belief shift clustering. *IEEE Transactions on Systems, Man and Cybernetics: Systems*, Major Revision.

- (9) **Zuo-wei Zhang**, Zhun-ga Liu, Hong-peng Tian, Arnaud Martin, Kuang Zhou. Mixed-type Imputation for Missing Data Credal Classification via Quality Matrices. *IEEE Transactions on Knowledge and Data Engineering*, Under Review.
- (10) **Zuo-wei Zhang**, Zhun-ga Liu, Liang-bo Ning, Arnaud Martin, Kuang Zhou. Representation of imprecision between clusters with arbitrary shape and size. *IEEE Transactions on Cybernetics*, Under Review.
- (11) Zhe Liu, **Zuo-wei Zhang**, Laurence T. Yang, Arnaud Martin, Sahil Garg. Belief ensemble clustering for missing data in IoMT via fuzzy reliability matrix. *IEEE Transactions on Fuzzy Systems*, Under Review.
- (12) Zhun-ga Liu, Liang-bo Ning, **Zuo-wei Zhang**. A New Progressive Multi-Source Domain Adaptation Network with Weighted Decision Fusion. *IEEE Transactions on Neural Networks and Learning Systems*, Under Review.
- (13) Yi-ru Zhang Sébastien Destercke, **Zuo-wei Zhang**, Tassadit Bouadi, Arnaud Martin. On computing evidential centroid through conjunctive combination: an impossibility theorem. *IEEE Transactions on Artificial Intelligence*, Under Review.
- (14) **Zuo-wei Zhang**, Arnaud Martin, Yi-ru Zhang, Zhun-ga Liu, Kuang Zhou. A survey of evidential clustering: definitions, methods, and applications. *ACM Computing Surveys*, Under Review.
- (15) **Zuo-wei Zhang**, Zhun-ga Liu, Liangbo Ning, Arnaud Martin, Kuang Zhou, Representation of Imprecision in Deep Techniques: Active Learning via Belief Functions. *IEEE Transactions on Image Processing*, Under Review.

#### Conference Papers

- (1) **Zuo-wei Zhang**, Xu-xia Zhang, Zhun-ga Liu. A new belief-based classification fusion for incomplete data. In 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP) (pp. 1-5). IEEE.
- (2) Xu-xia Zhang, Jing-fei Duan, **Zuo-wei Zhang**, Zhun-ga Liu. A new weighted classifier combination method with two-step evidential discounting operations. In 2019 International Applied Computational Electromagnetics Society Symposium-China (ACES) (Vol. 1, pp. 1-2). IEEE.
- (3) **Zuo-wei Zhang**, Arnaud Martin, Zhun-ga Liu, Kuang Zhou, Yi-ru Zhang. Fast unfolding of credal partitions in evidential clustering. In 6th International Conference on Belief Functions (BELIEF 2021), pp. 3-12. Springer.
- (4) **Zuo-wei Zhang**, Zhun-ga Liu, Kuang Zhou, Arnaud Martin, Yi-ru Zhang. Credal clustering for imbalanced data. In 6th International Conference on Belief Functions (BELIEF 2021), pp. 13-21. Springer.

# Bibliography

- Abdelkhalek, R., Boukhris, I., and Elouedi, Z. (2018). An evidential collaborative filtering approach based on items contents clustering. In *International Conference on Belief Functions*, pages 1–9. Springer.
- Abdelkhalek, R., Boukhris, I., and Elouedi, Z. (2019a). An evidential clustering for collaborative filtering based on users preferences. In *International Conference on Modeling Decisions for Artificial Intelligence*, pages 224–235. Springer.
- Abdelkhalek, R., Boukhris, I., and Elouedi, Z. (2019b). Joining items clustering and users clustering for evidential collaborative filtering. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 310–318. Springer.
- Aggarwal, C. C. and Philip, S. Y. (2008). A survey of uncertain data algorithms and applications. *IEEE Transactions on Knowledge and Data Engineering*, 21(5):609–623.
- Antoine, V., Labroche, N., and Vu, V. V. (2014a). Evidential seed-based semi-supervised clustering. In *2014 Joint 7th International Conference on Soft Computing and Intelligent Systems (SCIS) and 15th International Symposium on Advanced Intelligent Systems (ISIS)*, pages 706–711. IEEE.
- Antoine, V., Quost, B., Masson, M. H., and Dencœux, T. (2010). CECM: Adding pairwise constraints to evidential clustering. In *International Conference on Fuzzy Systems*, pages 1–8. IEEE.
- Antoine, V., Quost, B., Masson, M.-H., and Dencœux, T. (2011). CEVCLUS: Constrained evidential clustering of proximity data. In *7th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT 2011)*, pages 876–882. Atlantis Press.
- Antoine, V., Quost, B., Masson, M. H., and Dencœux, T. (2012). CECM: Constrained evidential c-means algorithm. *Computational Statistics & Data Analysis*, 56(4):894–914.
- Antoine, V., Quost, B., Masson, M. H., and Dencœux, T. (2014b). CEVCLUS: evidential clustering with instance-level constraints for relational data. *Soft Computing*, 18(7):1321–1335.



- Argenziano, G., Soyer, H., De Giorgi, V., Piccolo, D., Carli, P., and Delfino, M. (2002). Dermoscopy: a tutorial. *EDRA, Medical Publishing & New Media*, 16.
- Atanasov, K. (1999). Intuitionistic fuzzy sets: Theory and applications. *Physica-Verlag, New York*.
- Ayed, S. B., Elouedi, Z., and Lefevre, E. (2017). ECTD: evidential clustering and case types detection for case base maintenance. In *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, pages 1462–1469. IEEE.
- Baadel, S., Thabtah, F., and Lu, J. (2016). Overlapping clustering: A review. In *2016 SAI Computing Conference (SAI)*, pages 233–237. IEEE.
- Banerjee, A., Krumpelman, C., Ghosh, J., Basu, S., and Mooney, R. J. (2005). Model-based overlapping clustering. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 532–537.
- Ben Hur, A., Horn, D., Siegelmann, H. T., and Vapnik, V. (2001). Support vector clustering. *Journal of Machine Learning Research*, 2(Dec):125–137.
- Bessrouer, M., Elouedi, Z., and Lefèvre, E. (2020). E-DBSCAN: An evidential version of the DBSCAN method. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 3073–3080. IEEE.
- Bezdek, J. C. (2013). *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media.
- Bezdek, J. C., Ehrlich, R., and Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & geosciences*, 10(2-3):191–203.
- Bian, Z. K., Chung, F. L., and Wang, S. T. (2020). Fuzzy density peaks clustering. *IEEE Transactions on Fuzzy Systems*.
- Cheng, Y. Z. (1995). Mean shift, mode seeking, and clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):790–799.
- Chuang, K. S., Tzeng, H. L., Chen, S., Wu, J., and Chen, T. J. (2006). Fuzzy c-means clustering with spatial information for image segmentation. *Computerized Medical Imaging and Graphics*, 30(1):9–15.
- Coppi, R., DUrso, P., and Giordani, P. (2012). Fuzzy and possibilistic clustering for fuzzy data. *Computational Statistics & Data Analysis*, 56(4):915–927.
- Cottrell, M., Hammer, B., Hasenfuß, A., and Villmann, T. (2006). Batch and median neural gas. *Neural Networks*, 19(6-7):762–771.

- Dave, R. N. (1991). Characterization and detection of noise in clustering. *Pattern Recognition Letters*, 12(11):657–664.
- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197.
- Dempster, A. P. (2008). Upper and lower probabilities induced by a multivalued mapping. In *Classic works of the Dempster-Shafer theory of belief functions*, pages 57–72. Springer.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.
- Denceux, T. (2008a). Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172(2-3):234–264.
- Denceux, T. (2008b). A k-nearest neighbor classification rule based on Dempster-Shafer theory. In *Classic works of the Dempster-Shafer theory of belief functions*, pages 737–760. Springer.
- Denceux, T. (2011). Maximum likelihood estimation from uncertain data in the belief function framework. *IEEE Transactions on Knowledge and Data Engineering*, 25(1):119–130.
- Denceux, T. (2016). 40 years of Dempster-Shafer theory. *International Journal of Approximate Reasoning*, 79(C):1–6.
- Denceux, T. (2020a). Calibrated model-based evidential clustering using bootstrapping. *Information Sciences*, 528:17–45.
- Denceux, T. (2020b). NN-EVCLUS: Neural network-based evidential clustering. *arXiv preprint arXiv:2009.12795*.
- Denceux, T. and Kanjanatarakul, O. (2016). Evidential clustering: a review. In *International symposium on integrated uncertainty in knowledge modelling and decision making*, pages 24–35. Springer.
- Denceux, T., Kanjanatarakul, O., and Sriboonchitta, S. (2015). EK-NNclus: a clustering procedure based on the evidential k-nearest neighbor rule. *Knowledge-Based Systems*, 88:57–69.
- Denceux, T., Li, S. M., and Sriboonchitta, S. (2017). Evaluating and comparing soft partitions: An approach based on Dempster–Shafer theory. *IEEE Transactions on Fuzzy Systems*, 26(3):1231–1244.

- Denceux, T. and Masson, M. (2003). Clustering of proximity data using belief functions. In *Intelligent Systems for Information Processing*, pages 291–302. Elsevier.
- Denceux, T. and Masson, M.-H. (2004). EVCLUS: evidential clustering of proximity data. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1):95–109.
- Denoeux, T. and Masson, M.-H. (2004). Principal component analysis of fuzzy data using autoassociative neural networks. *IEEE Transactions on Fuzzy Systems*, 12(3):336–349.
- Dubois, D. and Prade, H. (1978). Operations on fuzzy numbers. *International Journal of systems science*, 9(6):613–626.
- Dubois, D. and Prade, H. (1988). Representation and combination of uncertainty with belief functions and possibility measures. *Computational intelligence*, 4(3):244–264.
- Dubois, D. and Prade, H. (2012). *Possibility theory: an approach to computerized processing of uncertainty*. Springer Science & Business Media.
- Durrett, R. (2019). *Probability: theory and examples*, volume 49. Cambridge University Press.
- Ester, M., Kriegel, H. P., Sander, J., and Xu, X. W. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, volume 96, pages 226–231.
- Fayyad, U. M., Piatetsky Shapiro, G., Smyth, P., and Uthurusamy, R. (1996). Advances in knowledge discovery and data mining. American Association for Artificial Intelligence.
- Filev, D., Georgieva, O., Angelov, P., and Kasabov, A. (2010). An extended version of the Gustafson-Kessel algorithm for evolving data stream clustering. *Evolving intelligent systems: Methodology and applications*, pages 273–300.
- Franti, P. and Virmajoki, O. (2006). Iterative shrinking method for clustering problems. *Pattern Recognition*, 39(5):761–775.
- Fred, A. and Loureno, A. (2008). Cluster ensemble methods: from single clusterings to combined solutions. In *Supervised and Unsupervised Ensemble Methods and Their Applications*, pages 3–30. Springer.
- Frey, B. J. and Dueck, D. (2007). Clustering by passing messages between data points. *science*, 315(5814):972–976.

- Fukunaga, K. and Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on Information Theory*, 21(1):32–40.
- Geweniger, T., Zülke, D., Hammer, B., and Villmann, T. (2010). Median fuzzy c-means for clustering dissimilarity data. *Neurocomputing*, 73(7-9):1109–1116.
- Gong, C. Y., Su, Z. G., Wang, P. H., and Wang, Q. (2020). Cumulative belief peaks evidential k-nearest neighbor clustering. *Knowledge-Based Systems*, 200:105982.
- Gong, C. Y., Su, Z. G., Wang, P. H., and Wang, Q. (2021). An evidential clustering algorithm by finding belief-peaks and disjoint neighborhoods. *Pattern Recognition*, 113:107751.
- Hadjitodorov, S. T., Kuncheva, L. I., and Todorova, L. P. (2006). Moderate diversity for better cluster ensembles. *Information Fusion*, 7(3):264–275.
- Han, D., Han, C., and Deng, Y. (2013). Novel approaches for the transformation of fuzzy membership function into basic probability assignment based on uncertainty optimization. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 21(02):289–322.
- Han, D., Onera, J. D., Han, C., and Yang, Y. (2011). New dissimilarity measures in evidence theory. In *14th International Conference on Information Fusion*, pages 1–7. IEEE.
- Hariz, S. B. and Elouedi, Z. (2010a). DK-BKM: decremental k belief k-modes method. In *International Conference on Scalable Uncertainty Management*, pages 84–97. Springer.
- Hariz, S. B. and Elouedi, Z. (2010b). IK-BKM: An incremental clustering approach based on intra-cluster distance. In *ACS/IEEE International Conference on Computer Systems and Applications-AICCSA 2010*, pages 1–8. IEEE.
- Hariz, S. B. and Elouedi, Z. (2011). Ranking-based feature selection method for dynamic belief clustering. In *International Conference on Adaptive and Intelligent Systems*, pages 308–319. Springer.
- Hariz, S. B., Elouedi, Z., and Mellouli, K. (2006). Clustering approach using belief function theory. In *International Conference on Artificial Intelligence: Methodology, Systems, and Applications*, pages 162–171. Springer.
- Hariz, S. B., Elouedi, Z., and Mellouli, K. (2007). Selection initial modes for belief k-modes method. *International Journal of Applied Science, Engineering and Technology*, 4(4).

- Haykin, S. and Network, N. (2004). A comprehensive foundation. *Neural networks*, 2(2004):41.
- Höppner, F., Klawonn, F., Kruse, R., and Runkler, T. (1999). *Fuzzy cluster analysis: methods for classification, data analysis and image recognition*. John Wiley & Sons.
- Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, 2(3):283–304.
- Huang, Z. X. and Ng, M. K. (1999). A fuzzy k-modes algorithm for clustering categorical data. *IEEE Transactions on Fuzzy Systems*, 7(4):446–452.
- Jain, A. K. (2010). Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651–666.
- Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys (CSUR)*, 31(3):264–323.
- Jeffreys, H. (1998). *The theory of probability*. OUP Oxford.
- Jiao, L. M., Dencœur, T., Liu, Z. G., and Pan, Q. (2020). EGMM: an evidential version of the gaussian mixture model for clustering. *arXiv preprint arXiv:2010.01333*.
- Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254.
- Kaufman, L. and Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis*, volume 344. John Wiley & Sons.
- Ke, Y. and Sukthankar, R. (2004). Pca-sift: A more distinctive representation for local image descriptors. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, pages II–II. IEEE.
- Krishnapuram, R., Joshi, A., Nasraoui, O., and Yi, L. (2001). Low-complexity fuzzy relational clustering algorithms for web mining. *IEEE transactions on Fuzzy Systems*, 9(4):595–607.
- Krishnapuram, R. and Keller, J. M. (1993). A possibilistic approach to clustering. *IEEE Transactions on Fuzzy Systems*, 1(2):98–110.
- Krishnapuram, R. and Keller, J. M. (1996). The possibilistic c-means algorithm: insights and recommendations. *IEEE Transactions on Fuzzy Systems*, 4(3):385–393.
- Li, F., Li, S., and Dencœur, T. (2018). k-CEVCLUS: constrained evidential clustering of large dissimilarity data. *Knowledge-Based Systems*, 142:29–44.

- Lian, C. F., Ruan, S., Dencœux, T., Li, H., and Vera, P. (2017a). Spatial evidential clustering with adaptive distance metric for tumor segmentation in fdg-pet images. *IEEE Transactions on Biomedical Engineering*, 65(1):21–30.
- Lian, C. F., Ruan, S., Dencœux, T., Li, H., and Vera, P. (2017b). Tumor delineation in FDG-PET images using a new evidential clustering algorithm with spatial regularization and adaptive distance metric. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 1177–1180. IEEE.
- Liang, J. Y., Bai, L., Dang, C. Y., and Cao, F. Y. (2012). The  $k$ -means-type algorithms versus imbalanced data distributions. *IEEE Transactions on Fuzzy Systems*, 20(4):728–745.
- Lingras, P. and Peters, G. (2011). Rough clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(1):64–72.
- Liu, Z. G., Dezert, J., Mercier, G., and Pan, Q. (2012). Belief  $c$ -means: An extension of fuzzy  $c$ -means algorithm in belief functions framework. *Pattern Recognition Letters*, 33(3):291–300.
- Liu, Z. G., Pan, Q., and Dezert, J. (2013). A new belief-based  $k$ -nearest neighbor classification method. *Pattern Recognition*, 46(3):834–844.
- Liu, Z. G., Pan, Q., Dezert, J., and Mercier, G. (2015). Credal  $c$ -means clustering method based on belief functions. *Knowledge-Based Systems*, 74:119–132.
- Liu, Z. G., Pan, Q., Mercier, G., and Dezert, J. (2014). A new incomplete pattern classification method based on evidential reasoning. *IEEE Transactions on Cybernetics*, 45(4):635–646.
- Lotfi, A., Moradi, P., and Beigy, H. (2020). Density peaks clustering based on density backbone and fuzzy neighborhood. *Pattern Recognition*, 107:107449.
- Mahamadou, A. J. D., Antoine, V., Christie, G. J., and Moreno, S. (2019). Evidential clustering for categorical data. In *2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, pages 1–6. IEEE.
- Martin, A. (2019). Conflict management in information fusion with belief functions.
- Martin, D., Fowlkes, C., Tal, D., and Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423. IEEE.
- Masson, M. H. and Dencœux, T. (2004). Clustering interval-valued proximity data using belief functions. *Pattern Recognition Letters*, 25(2):163–171.

- Masson, M. H. and Denceux, T. (2008). ECM: An evidential version of the fuzzy c-means algorithm. *Pattern Recognition*, 41(4):1384–1397.
- Masson, M. H. and Denceux, T. (2009). RECM: relational evidential c-means algorithm. *Pattern Recognition Letters*, 30(11):1015–1026.
- Masson, M. H. and Denceux, T. (2011). Ensemble clustering in the belief functions framework. *International Journal of Approximate Reasoning*, 52(1):92–109.
- McClellan, S., Scotney, B., and Shapcott, M. (2001). Aggregation of imprecise and uncertain information in databases. *IEEE Transactions on Knowledge and Data Engineering*, 13(6):902–912.
- McLachlan, G. J. and Krishnan, T. (2007). *The EM algorithm and extensions*, volume 382. John Wiley & Sons.
- Murtagh, F. (1983). A survey of recent advances in hierarchical clustering algorithms. *The Computer Journal*, 26(4):354–359.
- Pal, N. R., Pal, K., Keller, J. M., and Bezdek, J. C. (2005). A possibilistic fuzzy c-means clustering algorithm. *IEEE Transactions on Fuzzy Systems*, 13(4):517–530.
- Pawlak, Z. (1982). Rough sets. *International journal of computer & information sciences*, 11(5):341–356.
- Pawlak, Z. (1991). *Rough sets: Theoretical aspects of reasoning about data*, volume 9. Springer Science & Business Media.
- Pawlak, Z. and Skowron, A. (2007). Rough sets: some extensions. *Information sciences*, 177(1):28–40.
- Peters, G., Weber, R., and Nowatzke, R. (2012). Dynamic rough clustering and its applications. *Applied Soft Computing*, 12(10):3193–3207.
- Pizzagalli, D. U., Gonzalez, S. F., and Krause, R. (2019). A trainable clustering algorithm based on shortest paths from density peaks. *Science Advances*, 5(10):3770.
- Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850.
- Rodriguez, A. and Laio, A. (2014). Clustering by fast search and find of density peaks. *science*, 344(6191):1492–1496.
- Saidi, F., Trabelsi, Z., and Ghazela, H. B. (2018a). An evidential clustering based framework for cyber terrorist cells topology identification. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*, pages 436–443. IEEE.

- Saidi, F., Trabelsi, Z., and Ghazela, H. B. (2018b). A novel approach for terrorist sub-communities detection based on constrained evidential clustering. In *2018 12th International Conference on Research Challenges in Information Science (RCIS)*, pages 1–8. IEEE.
- Samaria, F. S. and Harter, A. C. (1994). Parameterisation of a stochastic model for human face identification. In *Proceedings of 1994 IEEE workshop on applications of computer vision*, pages 138–142. IEEE.
- Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W., and Lin, C.-T. (2017). A review of clustering techniques and developments. *Neurocomputing*, 267:664–681.
- Sen, S. and Davé, R. N. (1998). Clustering of relational data containing noise and outliers. In *1998 IEEE International Conference on Fuzzy Systems Proceedings. IEEE World Congress on Computational Intelligence (Cat. No. 98CH36228)*, volume 2, pages 1411–1416. IEEE.
- Serir, L., Ramasso, E., and Zerhouni, N. (2012). Evidential evolving Gustafson–Kessel algorithm for online data streams partitioning using belief function theory. *International Journal of Approximate Reasoning*, 53(5):747–768.
- Shafer, G. (1976). *A mathematical theory of evidence*, volume 42. Princeton university press.
- Shafer, G. (2016). A mathematical theory of evidence turns 40. *International Journal of Approximate Reasoning*, 79:7–25.
- Smarandache, F. and Dezert, J. (2005). Information fusion based on new proportional conflict redistribution rules. In *2005 7th International Conference on Information Fusion*, volume 2, pages 8–pp. IEEE.
- Smarandache, F. and Dezert, J. (2013). On the consistency of pcr6 with the averaging rule and its application to probability estimation. In *Proceedings of the 16th International Conference on Information Fusion*, pages 1119–1126. IEEE.
- Smets, P. (1990). The combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(5):447–458.
- Smets, P. (2005). Decision making in the TBM: the necessity of the pignistic transformation. *International Journal of Approximate Reasoning*, 38(2):133–147.
- Smets, P. (2007). Analyzing the combination of conflicting belief functions. *Information fusion*, 8(4):387–412.



- Smets, P. and Kennes, R. (1994). The transferable belief model. *Artificial intelligence*, 66(2):191–234.
- Su, Z. G. and Dencœux, T. (2018). BPEC: Belief-peaks evidential clustering. *IEEE Transactions on Fuzzy Systems*, 27(1):111–123.
- Su, Z. G., Dencœux, T., Hao, Y. S., and Zhao, M. (2018). Evidential K-NN classification with enhanced performance via optimizing a class of parametric conjunctive t-rules. *Knowledge-Based Systems*, 142:7–16.
- Su, Z. G., Zhou, H. Y., and Hao, Y. S. (2020). Evidential evolving c-means clustering method based on artificial bee colony algorithm with variable strings and interactive evaluation mode. *Fuzzy Optimization and Decision Making*, pages 1–21.
- Ter Braak, C. J., Kourmpetis, Y., Kiers, H. A., and Bink, M. C. (2009). Approximating a similarity matrix by a latent class model: a reappraisal of additive fuzzy clustering. *Computational statistics & data analysis*, 53(8):3183–3193.
- Vesanto, J. and Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11(3):586–600.
- Wold, S., Esbensen, K., and Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52.
- Xu, R. and Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3):645–678.
- Yamasaki, R. and Tanaka, T. (2019). Properties of mean shift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(9):2273–2286.
- Yao Hui, L., Zheng Ming, M., and Fang, Y. (2017). Adaptive density peak clustering based on k-nearest neighbors with aggregating strategy. *Knowledge-Based Systems*, 133:208–220.
- Zadeh, L. A. (1996). Fuzzy sets. In *Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh*, pages 394–432. World Scientific.
- Zhang, J. S. and Leung, Y. W. (2004). Improved possibilistic c-means clustering algorithms. *IEEE Transactions on Fuzzy Systems*, 12(2):209–217.
- Zhang, Y. F., Chen, S. M., and Yu, G. (2016). Efficient distributed density peaks for clustering large data sets in mapreduce. *IEEE Transactions on Knowledge and Data Engineering*, 28(12):3218–3230.
- Zhang, Z., Martin, A., Liu, Z., Zhou, K., and Zhang, Y. (2021a). Fast unfolding of credal partitions in evidential clustering. In *International Conference on Belief Functions*, pages 3–12. Springer.

- Zhang, Z.-w., Liu, Z., Ma, Z.-f., He, J.-h., and Zhu, X.-y. (2021b). Evidence integration credal classification algorithm versus missing data distributions. *Information Sciences*, 569:39–54.
- Zhang, Z. W., Liu, Z., Ma, Z. F., Zhang, Y. R., and Wang, H. (2021c). A new belief-based incomplete pattern unsupervised classification method. *IEEE Transactions on Knowledge and Data Engineering*.
- Zhang, Z. W., Liu, Z., Martin, A., Liu, Z. G., and Zhou, K. (2021d). Dynamic evidential clustering algorithm. *Knowledge-Based Systems*, 213:106643.
- Zhang, Z.-W., Tian, H.-P., Yan, L.-Z., Martin, A., and Zhou, K. (2021e). Learning a credal classifier with optimized and adaptive multiestimation for missing data imputation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*.
- Zhou, H. Y., Su, Z. G., and Hao, Y. S. (2019). Multi-objective evolutionary evidential c-means clustering. In *2019 International Conference on Electronic Engineering and Informatics (EEI)*, pages 426–429. IEEE.
- Zhou, K., Guo, M., and Martin, A. (2021). Evidential clustering based on transfer learning. In *International Conference on Belief Functions*, pages 56–65. Springer.
- Zhou, K., Martin, A., and Pan, Q. (2014). Evidential communities for complex networks. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 557–566. Springer.
- Zhou, K., Martin, A., and Pan, Q. (2015a). A similarity-based community detection method with multiple prototype representation. *Physica A: Statistical Mechanics and its Applications*, 438:519–531.
- Zhou, K., Martin, A., Pan, Q., and Liu, Z. G. (2015b). Evidential relational clustering using medoids. In *2015 18th International Conference on Information Fusion (Fusion)*, pages 413–420. IEEE.
- Zhou, K., Martin, A., Pan, Q., and Liu, Z. G. (2015c). Median evidential c-means algorithm and its application to community detection. *Knowledge-Based Systems*, 74:69–88.
- Zhou, K., Martin, A., Pan, Q., and Liu, Z. G. (2016a). ECMdd: Evidential c-medoids clustering with multiple prototypes. *Pattern Recognition*, 60:239–257.
- Zhou, K., Martin, A., Pan, Q., and Liu, Z. G. (2016b). Evidential label propagation algorithm for graphs. In *2016 19th International Conference on Information Fusion (FUSION)*, pages 1316–1323. IEEE.

- Zhou, K., Martin, A., Pan, Q., and Liu, Z. G. (2018a). SELP: Semi-supervised evidential label propagation algorithm for graph data clustering. *International Journal of Approximate Reasoning*, 92:139–154.
- Zhou, K., Pan, Q., and Martin, A. (2018b). Evidential community detection based on density peaks. In *International Conference on Belief Functions*, pages 269–277. Springer.
- Zimmermann, H. J. (2011). *Fuzzy set theory and its applications*. Springer Science & Business Media.

---

**Titre :** Représentation de l'incertitude et de l'imprécision entre les clusters avec des fonctions de croyance

**Mots clés :** Fonctions de croyance, clustering crédibiliste, incertitude, imprécision

**Résumé :** Le clustering crédibiliste fondé sur la théorie des fonctions de croyance est devenu un sujet important de la fouille de données en raison de sa capacité à caractériser l'incertitude et l'imprécision entre les clusters. Cependant, il existe encore des problèmes tels que l'ambiguïté des concepts de base, la complexité de calcul élevée et l'incapacité à détecter efficacement les clusters déséquilibrés et arbitraires. Cette thèse est consacrée à la résolution des problèmes ci-dessus et propose des solutions correspondantes, qui comprennent quatre parties. Tout d'abord, nous passons en revue de manière exhaustive les algorithmes de clustering crédibiliste existants et nous donnons les concepts et définitions associés, en détaillant en quoi le clustering crédibiliste peut caractériser l'incertitude et l'imprécision entre les clusters.

Deuxièmement, nous proposons un algorithme de clustering dynamique qui peut réduire efficacement la complexité de calcul des algorithmes existants pour étendre les applications. Troisièmement, nous proposons un algorithme de regroupement par décalage de croyance pour caractériser l'incertitude et l'imprécision entre les grappes déséquilibrées. Quatrièmement, nous proposons un algorithme de convergence crédibiliste fondé sur les pics de densité qui généralise la détection de clusters de forme, de taille et de structure spatiale arbitraires. Des données artificielles et réelles ont permis de vérifier l'efficacité des algorithmes proposés. Enfin, nous discutons également de quelques recherches potentielles à l'issue de cette thèse.

---

**Title :** Representation of uncertainty and imprecision between clusters with belief functions

**Keywords :** Belief functions, evidential clustering, uncertainty, imprecision

**Abstract :** Evidential clustering based on the theory of belief functions has become one of the topics of machine learning due to its ability to characterize the uncertainty and imprecision between clusters. However, there are still problems such as ambiguity of basic concepts, high computational complexity, and inability to detect imbalanced and arbitrary clusters effectively. This thesis is devoted to addressing the above problems and proposes corresponding solutions, which include four parts. First, we comprehensively survey existing evidential clustering algorithms and give related concepts and definitions, detailing why evidential clustering can characterize the uncertainty and imprecision between clusters.

Second, we propose a dynamic evidential clustering algorithm that can effectively reduce the computational complexity of existing algorithms to extend the applications. Third, we propose a belief shift clustering algorithm to characterize the uncertainty and imprecision between imbalanced clusters. Fourth, we propose a density peaks-based evidential convergence algorithm that generalizes the detection of clusters of arbitrary shape, size, and spatial structure. Artificial and real data have verified the effectiveness of the proposed algorithms. Finally, we also discuss some potential research based on this thesis.

## Résumé

L'analyse par regroupement, également appelée apprentissage non supervisé, a été largement utilisée comme outil d'analyse exploratoire des données dans de nombreux domaines. Elle vise à diviser des données massives sans aucune information préalable en groupes présentant des structures de données ou des structures physiques similaires. Pour y parvenir, plusieurs stratégies de clustering ont été proposées: des méthodes de partitionnement, des méthodes hiérarchiques, des méthodes fondées sur des distributions de probabilité, des méthodes spectrales, des méthodes fondées sur la densité, des méthodes à base de grilles, et des méthodes fondées sur l'apprentissage automatique telles que les réseaux de neurones. Cependant, la plupart de ces méthodes ne peuvent pas caractériser en même temps l'incertitude et l'imprécision entre les clusters. L'incertitude signifie que le processus d'affectation des objets à ces clusters peut être incertain. L'imprécision signifie que les informations sur les objets distribués dans les zones de chevauchement ou intermédiaires de différents groupes spécifiques peuvent être imprécises ou ambiguës. Ces dernières années, le clustering crédibiliste (Evidential clustering EC), fondé sur le concept de *credal partition*, a reçu beaucoup d'attention pour sa capacité à traiter ce problème. Il hérite des avantages de la théorie des fonctions de croyance (TBF) qui permet la modélisation des connaissances incertaines et imprécises. Cependant, comme le clustering crédibiliste est encore au stade de développement, il reste des problèmes tels que des concepts sous-jacents peu clairs, une complexité de calcul élevée et l'incapacité de détecter des clusters déséquilibrés ou arbitraires, ce qui limite les applications de la méthode EC.

Dans cette thèse, nous travaillons à proposer des schémas alternatifs pour résoudre ces problèmes. L'objectif est de normaliser certaines bases de l'Evidential Clustering (EC) et d'améliorer la capacité de l'EC à caractériser l'incertitude et l'imprécision entre les clusters, en particulier pour les clusters déséquilibrés ou de forme et de taille arbitraires. Plus précisément, quatre travaux sont mis en avant pour les traiter un par un.

Nous commençons par une étude systématique des approches crédibilistes de clustering. Dans ce travail, nous présentons les concepts et les définitions des données (entrées), des méthodes (modèles) et des groupes (résultats) de l'EC, ainsi que ceux des autres types (durs/flous/possibilistes), fondés sur différentes théories de l'incertain (théorie des probabilités, théorie des ensembles flous, théorie des possibilités et théorie des fonctions de croyance). En effet, l'EC est considéré comme la version crédibiliste du clustering dur, flou et possibiliste dans le cadre de la TBF. Ces concepts et définitions expliquent pourquoi l'EC peut mieux caractériser l'incertitude et l'imprécision entre les clusters. En outre, selon ces concepts et définitions, nous analysons les différences et les connexions entre EC et les autres types et déclarons ainsi les avantages d'EC dans la caractérisation de l'incertitude et de l'imprécision entre les clusters. De plus, nous étudions

l'EC depuis les méthodes séminales jusqu'aux méthodes de pointe dans le contexte des entrées fondées sur les données, ce qui donne lieu à une étude cohérente et complète pour analyser ces méthodes. Sur la base de l'analyse de certaines techniques représentatives sous différents angles (centre, type de mesure, complexité), nous fournissons un schéma directeur pour aider les utilisateurs à choisir les méthodes crédibilistes appropriées à leur données et cas d'étude.

Ensuite, nous introduisons un algorithme de clustering dynamique (DEC) pour résoudre le problème de la complexité de calcul élevée de l'EC traditionnel lors de la caractérisation de l'incertitude et de l'imprécision entre les clusters. Dans l'algorithme DEC, la plupart des objets de la requête sont considérés comme ayant des informations précises sur les clusters. Ainsi, une fonction objective de type FCM est d'abord employée et minimisée pour obtenir les niveaux de soutien des clusters singletons (spécifiques) auxquels les objets de la requête appartiennent. Ensuite, l'objet de la requête est initialement assigné de manière adaptative à l'objet aberrant, précis ou imprécis, via une nouvelle règle fondée sur les conflits entre les différents niveaux de support. Chaque objet imprécis est finalement réaffecté aux clusters singleton ou aux méta-cluster apparentés par redistribution crédale partielle avec une approche dynamique correspondant pour réduire la charge de calcul. La méthode DEC proposée peut réduire la complexité à un niveau similaire à celui du clustering flou/possibiliste, ce qui étend l'application de l'EC aux données volumineuses. Nous analysons également la complexité informatique de l'algorithme DEC proposé de manière théorique et effectuons une comparaison et une analyse détaillées avec la complexité informatique des méthodes EC connexes. Nous démontrons par des expériences approfondies mises en œuvre sur des ensembles de données artificielles et réelles que DEC peut obtenir des résultats de regroupement similaires à ceux des méthodes EC classiques, mais avec un temps d'exécution considérablement réduit. Cela démontre pleinement que la méthode DEC proposée est efficace et adaptée à l'application aux regroupements de données probantes. En outre, nous discutons également des paramètres impliqués dans DEC et nous constatons que la conversion de DEC en EC classique et en partition floue/possibiliste peut être réalisée en fixant les paramètres. En d'autres termes, la méthode DEC est une méthode plus généralisée.

Ensuite, nous étendons l'EC pour détecter les clusters déséquilibrés (clusters de tailles différentes) en combinant le déplacement de la moyenne avec l'EC traditionnel sous le TBF, appelé belief shift clustering (BSC). Le BSC proposé contient principalement deux caractéristiques. Premièrement, l'objet de la requête est préalablement assigné comme étant le bruit, précis ou imprécis, en se fondant sur la notion de "changement de croyance". Deuxièmement, une redistribution crédale partielle avec des centres de grappes dynamiques, pour éviter "l'effet uniforme" (pour les clusters déséquilibrées), est établie pour réaffecter les objets imprécis à un singleton cluster ou à un méta-cluster connexe. Une fois qu'un objet est assigné à un méta-cluster, cela indique que les clusters

singletons déséquilibrés impliqués dans le méta-cluster ne peuvent pas être distingués car cet objet peut être situé dans les zones de chevauchement ou intermédiaires de ces clusters singletons déséquilibrés. En procédant ainsi, l’approche BSC proposée peut raisonnablement caractériser l’incertitude et l’imprécision entre les clusters singletons déséquilibrés. Ensuite, nous démontrons l’efficacité de l’algorithme BSC proposé sur la base de données synthétiques équilibrées/déséquilibrées ainsi que de données réelles, notamment des données médicales et des données de visage. De nombreux résultats expérimentaux montrent que le BSC proposé peut bien traiter non seulement les données sphériques mais aussi les données déséquilibrées et caractériser l’incertitude et l’imprécision entre les clusters. En outre, nous discutons également des paramètres impliqués dans le BSC. Les expériences montrent que le BSC est robuste aux paramètres et qu’il est capable de contrôler le taux d’imprécision d’une bonne manière, ce qui peut être réglé en fonction de la situation réelle.

Pour éviter de perdre en généralité, nous étudions également la représentation de l’incertitude et de l’imprécision entre les clusters, indépendamment de leur forme, de leur taille et de leur dimensionnalité, sur la base des pics de densité et du TBF, appelé algorithme de *density peaks-based evidential convergence* (DPEC). Tout d’abord, nous considérons que différents voisins peuvent fournir des preuves complémentaires soutenant l’objet comme centre de cluster et redéfinissons une fonction de densité fondée sur la distance pour obtenir des centres de cluster plus robustes dans le graphe de décision. Ensuite, nous présentons une nouvelle règle de convergence crédibiliste pour affecter les objets restants à différents clusters. Enfin, comme dans le cas de l’approche BSC, les objets situés dans les zones de chevauchement ou intermédiaires de différents clusters singletons arbitraires sont assignés aux méta-clusters correspondants pour caractériser l’incertitude et l’imprécision entre ces clusters arbitraires. De nombreuses expériences ont montré que le DPEC proposé peut bien détecter des clusters arbitraires et caractériser l’incertitude et l’imprécision entre ces clusters. En outre, nous analysons la complexité de calcul du DPEC proposé. Les expériences montrent que le DPEC proposé a un temps d’exécution légèrement plus élevé que les autres algorithmes, mais qu’il en vaut la peine, en particulier pour certaines décisions prudentes.

Par ailleurs, nous comparons également les algorithmes DEC, BSC et DPEC proposés et analysons les différences et les liens entre eux. Comme ces algorithmes sont proposés pour caractériser l’incertitude et l’imprécision entre les différents clusters, leurs performances sont différentes. Plus précisément, DEC est utilisé pour réduire la complexité du clustering crédibiliste traditionnel. BSC vise à détecter les clusters déséquilibrés, tandis que DPEC peut détecter des clusters de formes, de tailles et de dimensionnalités arbitraires. L’efficacité des algorithmes proposés est estimée sur différents jeux de données artificielles et réelles. Les expériences montrent que les algorithmes proposés améliorent effectivement l’efficacité d’exécution de l’EC traditionnel et détectent les clusters déséquilibrés ou arbitraires, et caractérisent l’incertitude

et l'imprécision entre ces clusters. Plus précisément, les algorithmes DEC, BSC et DPEC proposés sont tous efficaces pour détecter les clusters sphériques. En revanche, BSC et DPEC peuvent détecter efficacement les clusters déséquilibrés, tandis que DPEC peut détecter les clusters arbitraires. En d'autres termes, cette thèse est consacrée à la généralisation des méthodes EC existantes afin de parvenir progressivement à la détection de clusters généralisés et de caractériser efficacement l'incertitude et l'imprécision entre les clusters. Cependant, cela ne signifie pas que la méthode DPEC est utilisée dans tous les cas, qui doit être choisie par l'utilisateur en fonction des exigences.

Ensuite, nous concluons cette thèse, notamment sur la base des algorithmes DEC, BSC et DPEC proposés, nous fournissons des directives de base actualisées pour la sélection de différentes méthodes crédibilistes, ce qui facilite grandement la sélection et l'utilisation par l'utilisateur en fonction de cas spécifiques.

Enfin, nous discutons des travaux en perspective de cette thèse, qui consistent en deux aspects principaux. Premièrement, par rapport aux méthodes traditionnelles de clustering, les algorithmes que nous avons proposés dans cette thèse sont capables de caractériser l'incertitude et l'imprécision entre des clusters (arbitraires) se chevauchant fondés sur des fonctions de masse ainsi que des méta-clusters. Cependant, en assignant des objets imprécis aux méta-clusters, nous détruisons également les distributions originales des clusters qui se chevauchent, c'est-à-dire les bords des clusters dans les régions de chevauchement. Deuxièmement, les causes d'incertitude et d'imprécision dans la classification (non supervisée) sont diverses, comme les données manquantes. Certains clusters/classes bien distribués à l'origine peuvent présenter un chevauchement local en raison de valeurs manquantes. Nous avons mené des recherches sur cette incertitude et cette imprécision entre les clusters/classes dues aux valeurs manquantes. Cependant, il n'existe pas de mécanisme efficace à long terme pour caractériser généralement l'incertitude et l'imprécision dues aux données manquantes dans le processus de modélisation et les résultats. Par conséquent, il s'agit encore de tâches difficiles. Nous proposerons des solutions alternatives dans nos travaux futurs.