A cross-domain perspective to Clustering with Uncertainty

Salvatore F. Pileggi[0000-0001-9722-2205]

School of Computer Science, Faculty of Engineering and IT
University of Technology Sydney (Australia)
SalvatoreFlavio.Pileggi@uts.edu.au

Abstract. Clustering in presence of uncertainty may be considered, at the same time, to be a pressing need and a challenge to effectively address many real-world problems. This concise literature review aims to identify and discuss the associated body of knowledge according to a cross-domain perspective. A semi-systematic methodology has allowed the selection of 68 papers, with a priority on the most recent contributions. The analysis has re-marked the relevance of the topic and has made explicit a trend to domain-specific solutions over generic-purpose approaches. On one side, this trend enables a more specific set of solutions within specific communities; on the other side, the resulting distributed approach is not always well-integrated in the mainstream and may generate a further fragmentation of the body of knowledge, mostly because of some lack of abstraction in the definition of specific problems. While these gaps are largely understandable within the research community, a lack of implementations to provide ready-to-use resources is overall critical, looking at a more and more computational and data intensive world.

Keywords: Clustering · Uncertainty Modelling · Uncertainty Management · Unsupervised Learning · Data Analysis · Data Mining

1 Introduction

Empirical observations show an increasing quantity of data with a degree of uncertainty [14]. Indeed, real-world data naturally tends to present uncertainty due to different factors including, among others, human or instrumental errors [81], randomness, imprecision, vagueness and partial ignorance [17].

In general terms, the theoretical impact of data uncertainty, as well as the risk associated with ignoring it (e.g. [48][19]), are well-known issues within the scientific community. Always in general, it strongly suggests, wherever possible, a proper and explicit uncertainty model to effectively support representation, measuring and consequent analysis. From a more practical perspective, more and more studies present a specific focus in a variety of application domains, such as, among the very many, budget impact analysis [59], organizational environments [45] and hydrological data [56]. Such a critical modelling is intrinsically challenging and may require a domain specific-approach, such as for Big Data [29][79], Visualization [41] and Deep Learning [1].

On the other side, clustering techniques [86] group data points into different clusters based on their similarity. These techniques have been extensively used in a general scientific context and traditional approaches keep evolving as a response to an environment characterised by evolving needs [82]. For instance, clustering is a common class of unsupervised learning [73], often adapted to achieve concrete goals in the different application domains (e.g. [7]), as well as formal classification [8], ontological modelling [49, 62] and rule mining [76] commonly rely on clustering techniques.

Intuitively, clustering in a context of uncertainty, or even just potential uncertainty, proposes additional significant challenges on both (i) modelling similarity between uncertain objects and (ii) developing effective and efficient computational methods accordingly [39].

Alternative approaches to deal with uncertainty can be used for different reasons in different contexts. A classification of these techniques is not trivial. For instance, in [36] the authors have identified two main broad categories that aim, respectively, to complement and to generalise probabilistic representations. The former family addresses non-probabilistic uncertainty (typically imprecision, vagueness or gradedness), while the latter targets effective modelling of partial ignorance. More holistically, looking at extensions of traditional methods, three main categories have been summarised in [39]: partitioning clustering, density-based clustering, and possible world approaches. The resulting extended solutions integrate the original semantics with uncertainty modelling.

This papers aims to holistically review the most recent advances in the field of clustering in presence of uncertainty. The focus is on a cross-domain perspective resulting from a contextual analysis that considers the most relevant computational trends and related applications.

Related Work. This work can be framed in the very broad context of uncertain data algoritms and applications [5]. A valuable review specifically on clustering has been provided in 2017 [17]. The focus of such a work is on uncertainty management and associated theoretical formalisms. Other concise contributions aimed at summarising the body of knowledge are relatively old (e.g. [36] and [14]), given the strong and constant advances in the computational world. This paper proposes an additional contribution to the body of knowledge in the field by addressing recent advances minimising the overlapping with existing reviews. Additionally, the provided analysis is performed according to a simple analysis framework, which also include the application domain. It allows to distinguish between generic-purpose and application-specific solutions.

Structure of the paper. The introductory part of the paper is concluded by a description of the adopted methodology (Section 2). The core part of the paper includes two different sections that aims, respectively, to overview the most relevant contributions in literature according to a cross-domain perspective (Section 3) and to discuss results looking at major gaps and challenges (Section 4). Finally, a conclusions section provides an overview of the work.

2 Methodology and Approach

In order to generate a tangible contribution to the body of knowledge and avoid, in the limit of the possible, overlapping and lack of deepening, this literature review has been conducted by combining typical systematic processes with non-systematic practices.

The mainstream process assumes, as usual, paper retrieval from relevant databases. In this specific case, queries have been performed by simply combining two main keywords, namely *Clustering* and *Uncertainty*.

Selection Criteria & Saturation. The selection of the papers to include in the review has been performed by applying a critical analysis aimed at the identification of the most relevant contributions in the field. Although no pre-defined objective filter (e.g. on time-range) has been applied, the most recent papers have been systematically included in order to highlight the most recent advances in context and to maximise the value provided. The relatively soft selection criteria enabled the retrieval of an important number of papers. However, the selection process has been much more focused in fact. Indeed, after a number of iterations, a feeling of saturation naturally emerged as contributions started to present consolidations of existing concepts rather than novel solutions. This additional non-systematic element has been a determinant to facilitate de facto conciseness at a relevant scale.

Analysis Framework & Limitations. The analysis has been conducted according to two major dimensions: *domain* and *approach*. The former dimension aims primarily to distinguish between generic-purpose and domain specific solutions, while the latter wants to facilitate an overview of major techniques. The presentation of the review (Section 3) has been structured looking at the domain. Indeed, the classification of the different approaches is intrinsically more fragmented and not always explicit. In general terms, the classification followed the claims by authors and the original analysis. Non-systematic practices may have introduced biases. It applies mostly to selection criteria. Additionally, because of the high number of existing works distributed in a variety of domains, it is hard to assess the exhaustiveness of the review.

3 A cross-domain Analysis

This section has a descriptive focus as it provides an overview of the contributions included in this study.

We deal separately with solutions that present a completely generic focus (referred to as "generic-purpose" and reported in Table 1) and that have been designed within a specific application domain ("domain-specific", Table 2). This generic classification naturally introduces a cross-domain analysis. However, there are not always well defined boundaries as certain applications as identified in the context of this work may present a certain degree of genericness.

Table 1: Generic-purpose selected contributions.

Title/Ref.	**Year**	**Approach**
Cloud-Cluster: An uncertainty clustering algorithm based on cloud model [54]	2023	Method
Outlier-robust multi-view clustering for uncertain data [69]	2021	Multi-view Clustering
Multi-view spectral clustering for uncertain objects [68]	2021	Multi-view Clustering
Modeling uncertain data using Monte Carlo integration method for clustering [67]	2019	Monte-Carlo
Optimal clustering under uncertainty [15]	2018	Method
Locally weighted ensemble clustering [33]	2017	Ensemble Clustering
Self-adapted mixture distance measure for clustering uncertain data [51]	2017	Method
Novel density-based and hierarchical density-based clustering algorithms for uncertain data [88]	2017	Hierarchical Clustering
An information-theoretic approach to hierarchical clustering of uncertain data [24]	2017	Hierarchical Clustering
Active Clustering with Model-Based Uncertainty Reduction [84]	2016	Active Clustering
Robust ensemble clustering using probability trajectories [32]	2015	Ensemble Clustering
Large margin clustering on uncertain data by considering probability distribution similarity [85]	2015	PD Similarity
Representative clustering of uncertain data [90]	2014	Framework
Active spectral clustering via iterative uncertainty reduction [80]	2012	Active Clustering
Minimizing the variance of cluster mixture models for clustering uncertain objects [22]	2012	Method
Clustering uncertain data based on probability distribution similarity [39]	2011	PD Similarity
Clustering uncertain data using voronoi diagrams and r-tree index [44]	2010	Method
Subspace clustering for uncertain data [25]	2010	Sub-space clustering
Exceeding expectations and clustering uncertain data [20]	2009	Optimization
Clustering Uncertain Data with Possible Worlds [77]	2009	Method
Clustering Uncertain Data Via K-Medoids [21]	2008	Method
A hierarchical algorithm for clustering uncertain data via an information-theoretic approach [23]	2008	Hierarchical Clustering
Clustering Uncertain Data Using Voronoi Diagrams [43]	2008	Voronoi diagrams
Uncertain fuzzy clustering: Insights and recommendations [65]	2007	Fuzzy Logic
Uncertain fuzzy clustering: Interval type-2 fuzzy approach to c-means [38]	2007	Fuzzy Logic
Density-based clustering of uncertain data [47]	2005	Fuzzy Logic
Hierarchical density-based clustering of uncertain data [46]	2005	Fuzzy Logic

Table 2: Domain-specific selected contributions.

Title/Ref.	**Year**	**Approach**	**Domain**
Stochastic economic dispatch of wind power under uncertainty using clustering-based extreme scenarios [6]	2024	Stochastic Model	Energy
Uncertainty clustering internal validity assessment using Fréchet distance for unsupervised learning [64]	2022	Fuzzy Logic	Machine Learning
A three-stage automated modal identification framework for bridge parameters based on frequency uncertainty and density clustering [30]	2022	Method	Engineering
Clustering uncertain graphs using ant colony optimization (ACO) [37]	2022	Optimization	Graphs
Uncertainty-Aware Clustering for Unsupervised Domain Adaptive Object Re-Identification [78]	2022	Hierarchical Clustering	Machine Learning

Continued on next page

Table 2 – continued from previous page

Title/Ref.	**Year**	**Approach**	**Domain**
Decision-based scenario clustering for decision-making under uncertainty [31]	2022	Method	Decision Making
Active domain adaptation via clustering uncertainty-weighted embeddings [63]	2021	Active Learning	Machine Learning
UAC: An Uncertainty-Aware Face Clustering Algorithm [16]	2021	Method	Face Recognition
Uncertainty assessment in reservoir performance prediction using a two-stage clustering approach: Proof of concept and field application [26]	2021	Method	Petroleum Science
Handling uncertainty in financial decision making: a clustering estimation of distribution algorithm with simplified simulation [70]	2020	Method	Decision Making
Ride-sharing under travel time uncertainty: Robust optimization and clustering approaches [50]	2020	Method	Transportation
Deep semantic clustering by partition confidence maximisation [35]	2020	Method	Machine Learning
Uncertainty mode selection in categorical clustering using the rough set theory [58]	2020	Rough Set	Categorical Data
Clustering of electrical load patterns and time periods using uncertainty-based multi-level amplitude thresholding [11]	2020	Fuzzy Logic	Energy
Efficient Assessment of Reservoir Uncertainty Using Distance-Based Clustering: A Review [42]	2019	Review	Petroleum Science
Big-data clustering with interval type-2 fuzzy uncertainty modeling in gene expression datasets [71]	2019	Fuzzy Logic	Genetics
Clustering mining blocks in presence of geological uncertainty [75]	2019	Method	Geology
Efficient and effective algorithms for clustering uncertain graphs [28]	2019	Optimization	Graphs
A three-way clustering approach for handling missing data using GTRS [2]	2018	Three-way Clustering	Missing Data
Clustering uncertain graphs [9]	2017	Method	Graphs
Novel adaptive multi-clustering algorithm-based optimal ESS sizing in ship power system considering uncertainty [87]	2017	Optimization	Energy
Multiple clustering views from multiple uncertain experts [10]	2017	Bayesian Model	Collaborative Environments
Uncertain data clustering in distributed peer-to-peer networks [89]	2017	Distributed Clustering	P2P Network
Efficient clustering of large uncertain graphs using neighborhood information [27]	2017	Framework	Graphs
Clustering based unit commitment with wind power uncertainty [72]	2016	Method	Energy
A framework for clustering uncertain data [66]	2015	Framework	Visualization
Efficient clustering of uncertain data streams [40]	2014	Method	Data Stream
Uncertain data clustering-based distance estimation in wireless sensor networks [55]	2014	Method	Wireless Sensor Network
Weighted graph clustering with non-uniform uncertainties [13]	2014	Optimization	Graphs

Continued on next page

Table 2 – continued from previous page

Title/Ref.	Year	Approach	Domain
Clustering large data with uncertainty [18]	2013	Fuzzy Logic	Large Data
Reliable clustering on uncertain graphs [52]	2012	Possible Worlds	Graphs
Clustering uncertain trajectories [61]	2011	Method	Location Data
Hue-stream: Evolution-based clustering technique for heterogeneous data streams with uncertainty [57]	2011	Method	Data Stream
An algorithm for clustering heterogeneous data streams with uncertainty [34]	2010	Method	Data Stream
On high dimensional projected clustering of uncertain data streams [3]	2009	Method	Data Stream
Clustering trajectories of moving objects in an uncertain world [60]	2009	Method	Location Data
A Framework for Clustering Uncertain Data Streams [4]	2008	Method	Data Stream
Conceptual clustering categorical data with uncertainty [83]	2007	Quality Assessment	Categorical Data
Including probe-level uncertainty in model-based gene expression clustering [53]	2007	Method	Genetics
Pvclust: an r package for assessing the uncertainty in hierarchical clustering [74]	2006	Hierarchical Clustering	Genetics
Uncertain Data Mining: An Example in Clustering Location Data [12]	2006	UK-Means	Location Data

Generic-purpose solutions. Among the generic-purpose works, there are two clearly identifiable sub-sets of solutions adopting, respectively, mixed (or not uniquely classifiable) methods [21, 77, 44, 22, 51, 15, 54] and Fuzzy Logic [46, 47, 38, 65]. Smaller classes of solutions adopt Hierarchical Clustering [88][23][24], Probability Distribution Similarity [39][85], Ensemble Clustering [33][32], Multi-view Clustering [68][69] and Active Clustering [84][80]. Other methods focus on framework-based solutions [90], Voronoi diagrams [43], Monte-Carlo [67], optimization strategies [20] and Sub-space Clustering [25].

Domain-specific solutions. Mixed methods [53, 4, 60, 3, 34, 57, 61, 55, 40, 72, 9, 75, 35, 50, 70, 26, 16, 31, 30], as well as Fuzzy Logic [18, 71, 11, 64], Hierarchical Clustering [74, 78], Optimization [13, 87, 37, 28], and framework-based approaches [66, 27] play a significant role also in a context of domain-specific applications. Other contributions include the adaptation of traditional techniques [12], quality assessment [83], Possible Worlds [52], Distributed Clustering [89], Bayesian Modeling [10], Three-way Clustering [2], review for assessment purpose within a specific domain [42], Rough Set Theory [58], Active Learning [63] and Stochastic Models [6].

4 Discussion

In order to exhaustively discuss the review, the section is structured in two different subsections to address an overview of the results (Section 4.1) and a critical analysis of the major gaps emerged (Section 4.2).

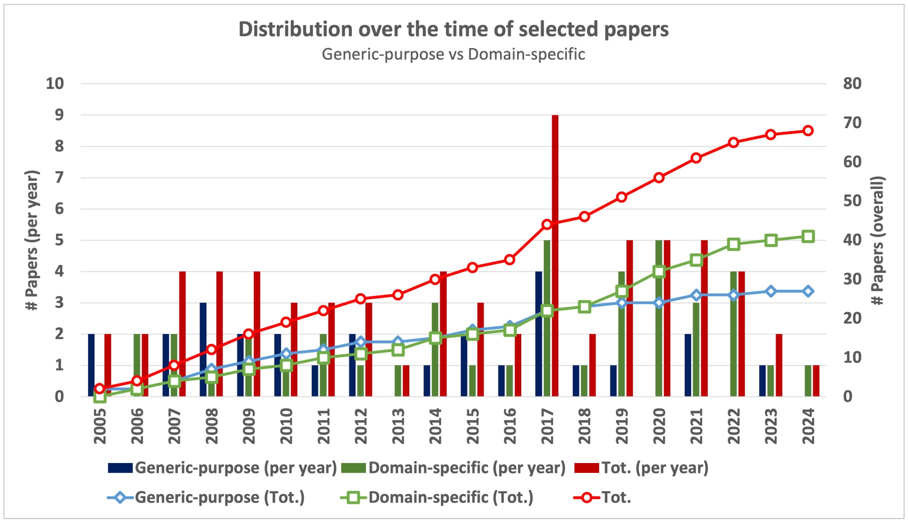


Fig. 1: Distribution over the time of the selected contributions.

4.1 Overview

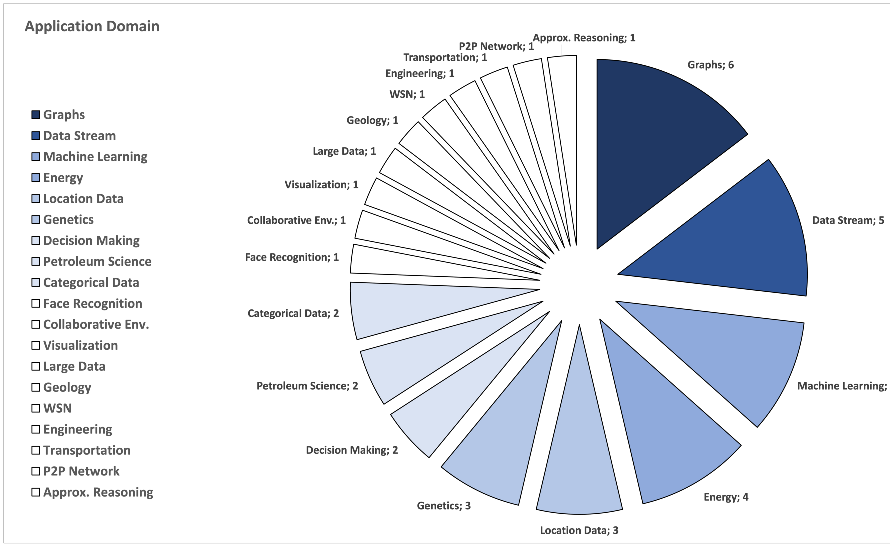
In quantitative terms, the majority (60%) of the 68 papers selected within the time range 2005-2024 presents a domain-specific focus. As shown in Figure 1, such a trend becomes more consistent and somehow predominant from 2018 onward. More holistically, the study confirms a substantial research interest in the topic throughout the observation period.

The analysis conducted in this study based on soft-classification allows to overview the application domain (Figure 2a). Looking at the 41 domain-specific papers, as expected, generic application fields, such as *Graphs*, *Data Stream* and *Machine learning*, are quantitatively more relevant, both with large domains (e.g. *Energy*, *Genetics* and *Location Data*). At a more fine-grained level, the review has identified a diverse spectrum reflecting a generic need for clustering in presence of uncertainty.

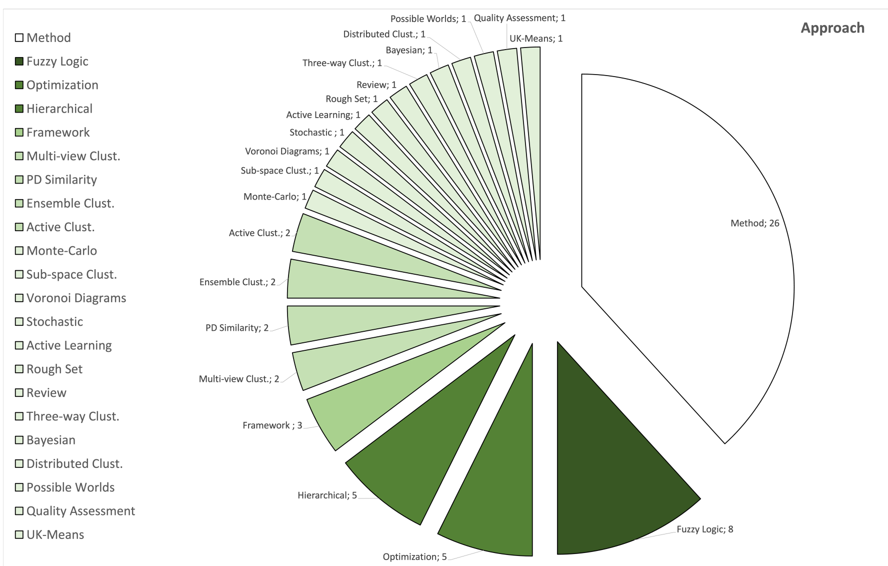
A more technical perspective is summarised in Figure 2b. A consistent part (38%) of the considered papers proposes a mixed-method approach, which is generically referred to as *method* in the adopted analysis framework. *Fuzzy Logic*, *Optimization* and *Hierarchical Clustering* are the most popular approaches. In addition, to note a focus on defining analysis *frameworks*, on *Multi-view Clustering*, *Probability Distribution Similarity*, *Ensemble* and *Active Clustering*.

4.2 Major Gaps and Challenges

From a critical perspective, the analysis conducted has allowed the identification of a number of gaps other than the originally reported in the different contributions that are summarised in Table 3.



(a) Application Domain.



(b) Approaches characterizing the selected solutions.

Fig. 2: Analysis overview.

Table 3: Main gaps.

	Gap
G1	Lack of implementations to provide ready-to-use computational resources
G2	Relationship between generic-purpose and domain-specific solutions not always clear
G3	Fine-grained application-specific approach that doesn't facilitate re-use in a different context
G4	Lack of abstraction in domain-specific approaches
G5	Despite a well-identified research field, solutions are not always discussed in context looking at the exiting body of knowledge

The review has reiterated the practical relevance of clustering in presence of uncertainty. In such a context, ready-to-use resources in the computational world are crucial and a determinant to consolidate and properly transfer innovation into practice (*G1*).

The cross-domain focus has highlighted and put emphasis on applications to solve real-world problems. The relationship between generic-purpose and domain-specific solutions not always clear (*G2*). The fine-grained application-specific approach makes re-using complex and costly (*G3*). That is because of a lack of abstraction in the formulation of domain-specific problems (*G4*) with a consequent difficulty to generalize solutions or re-using existing ones in a different context.

More in general, despite a well-identified research field, solutions are not always discussed in context looking at the exiting body of knowledge (*G5*).

5 Conclusions

Given the popularity of clustering techniques within the modern computational world and the intrinsic need to deal with uncertainty in the different application domains, this concise literature review aimed at a cross-domain analysis of the most recent solutions in the field.

Such analysis has firstly re-marked the relevance of the topic and the consequent related research activity. A trend to domain-specific solutions over generic-purpose approaches progressively emerged and became more consistent in the last few years. On one side, this trend enables a more specific set of solutions within specific communities; on the other side, the resulting distributed approach is not always well-integrated in the mainstream and may generate a further fragmentation of the body of knowledge, mostly because of some lack of abstraction in the definition of specific problems.

While these gaps are largely understandable within the research community, a lack of implementations to provide ready-to-use resources is overall critical, looking at a more and more computational and data intensive world.

Acknowledgements. I would like to thank the conference organisers, the workshop chair as well as the anonymous reviewers, who have provided valuable constructive feedback.

References

1. Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., et al.: A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Information fusion **76**, 243–297 (2021)
2. Afridi, M.K., Azam, N., Yao, J., Alanazi, E.: A three-way clustering approach for handling missing data using gtrs. International Journal of Approximate Reasoning **98**, 11–24 (2018)
3. Aggarwal, C.C.: On high dimensional projected clustering of uncertain data streams. In: 2009 IEEE 25th International Conference on Data Engineering. pp. 1152–1154. IEEE (2009)
4. Aggarwal, C.C., Philip, S.Y.: A framework for clustering uncertain data streams. In: 2008 IEEE 24th International Conference on Data Engineering. pp. 150–159. IEEE (2008)
5. Aggarwal, C.C., Philip, S.Y.: A survey of uncertain data algorithms and applications. IEEE Transactions on knowledge and data engineering **21**(5), 609–623 (2008)
6. Bhavsar, S., Pitchumani, R., Maack, J., Satkauskas, I., Reynolds, M., Jones, W.: Stochastic economic dispatch of wind power under uncertainty using clustering-based extreme scenarios. Electric Power Systems Research **229**, 110158 (2024)
7. Caron, M., Bojanowski, P., Joulin, A., Douze, M.: Deep clustering for unsupervised learning of visual features. In: Proceedings of the European conference on computer vision (ECCV). pp. 132–149 (2018)
8. Castellanos, A., Cigarrán, J., García-Serrano, A.: Formal concept analysis for topic detection: a clustering quality experimental analysis. Information Systems **66**, 24–42 (2017)
9. Ceccarello, M., Fantozzi, C., Pietracaprina, A., Pucci, G., Vandin, F.: Clustering uncertain graphs. Proceedings of the VLDB Endowment **11**(4), 472–484 (2017)
10. Chang, Y., Chen, J., Cho, M.H., Castaldi, P.J., Silverman, E.K., Dy, J.G.: Multiple clustering views from multiple uncertain experts. In: International Conference on Machine Learning. pp. 674–683. PMLR (2017)
11. Charwand, M., Gitizadeh, M., Siano, P., Chicco, G., Moshavash, Z.: Clustering of electrical load patterns and time periods using uncertainty-based multi-level amplitude thresholding. International Journal of Electrical Power & Energy Systems **117**, 105624 (2020)
12. Chau, M., Cheng, R., Kao, B., Ng, J.: Uncertain data mining: An example in clustering location data. In: Advances in Knowledge Discovery and Data Mining: 10th Pacific-Asia Conference, PAKDD 2006, Singapore, April 9-12, 2006. Proceedings 10. pp. 199–204. Springer (2006)
13. Chen, Y., Lim, S.H., Xu, H.: Weighted graph clustering with non-uniform uncertainties. In: International Conference on Machine Learning. pp. 1566–1574. PMLR (2014)
14. Cormode, G., McGregor, A.: Approximation algorithms for clustering uncertain data. In: Proceedings of the twenty-seventh ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. pp. 191–200 (2008)
15. Dalton, L.A., Benalcázar, M.E., Dougherty, E.R.: Optimal clustering under uncertainty. PLoS One **13**(10), e0204627 (2018)
16. Debnath, B., Coviello, G., Yang, Y., Chakradhar, S.: Uac: an uncertainty-aware face clustering algorithm. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3487–3495 (2021)

17. D'Urso, P.: Informational paradigm, management of uncertainty and theoretical formalisms in the clustering framework: A review. Information Sciences **400**, 30–62 (2017)
18. Ghosh, S., Mitra, S.: Clustering large data with uncertainty. Applied Soft Computing **13**(4), 1639–1645 (2013)
19. Griffin, S.C., Claxton, K.P., Palmer, S.J., Sculpher, M.J.: Dangerous omissions: the consequences of ignoring decision uncertainty. Health economics **20**(2), 212–224 (2011)
20. Guha, S., Munagala, K.: Exceeding expectations and clustering uncertain data. In: Proceedings of the twenty-eighth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. pp. 269–278 (2009)
21. Gullo, F., Ponti, G., Tagarelli, A.: Clustering uncertain data via k-medoids. In: International Conference on Scalable Uncertainty Management. pp. 229–242. Springer (2008)
22. Gullo, F., Ponti, G., Tagarelli, A.: Minimizing the variance of cluster mixture models for clustering uncertain objects. Statistical Analysis and Data Mining: The ASA Data Science Journal **6**(2), 116–135 (2013)
23. Gullo, F., Ponti, G., Tagarelli, A., Greco, S.: A hierarchical algorithm for clustering uncertain data via an information-theoretic approach. In: 2008 Eighth IEEE International Conference on Data Mining. pp. 821–826. IEEE (2008)
24. Gullo, F., Ponti, G., Tagarelli, A., Greco, S.: An information-theoretic approach to hierarchical clustering of uncertain data. Information sciences **402**, 199–215 (2017)
25. Günnemann, S., Kremer, H., Seidl, T.: Subspace clustering for uncertain data. In: Proceedings of the 2010 SIAM International Conference on Data Mining. pp. 385–396. SIAM (2010)
26. Haddadpour, H., Niri, M.E.: Uncertainty assessment in reservoir performance prediction using a two-stage clustering approach: Proof of concept and field application. Journal of Petroleum Science and Engineering **204**, 108765 (2021)
27. Halim, Z., Waqas, M., Baig, A.R., Rashid, A.: Efficient clustering of large uncertain graphs using neighborhood information. International Journal of Approximate Reasoning **90**, 274–291 (2017)
28. Han, K., Gui, F., Xiao, X., Tang, J., He, Y., Cao, Z., Huang, H.: Efficient and effective algorithms for clustering uncertain graphs. Proceedings of the VLDB Endowment **12**(6), 667–680 (2019)
29. Hariri, R.H., Fredericks, E.M., Bowers, K.M.: Uncertainty in big data analytics: survey, opportunities, and challenges. Journal of Big Data **6**(1), 1–16 (2019)
30. He, Y., Yang, J.P., Li, Y.F.: A three-stage automated modal identification framework for bridge parameters based on frequency uncertainty and density clustering. Engineering Structures **255**, 113891 (2022)
31. Hewitt, M., Ortmann, J., Rei, W.: Decision-based scenario clustering for decision-making under uncertainty. Annals of Operations Research **315**(2), 747–771 (2022)
32. Huang, D., Lai, J.H., Wang, C.D.: Robust ensemble clustering using probability trajectories. IEEE transactions on knowledge and data engineering **28**(5), 1312–1326 (2015)
33. Huang, D., Wang, C.D., Lai, J.H.: Locally weighted ensemble clustering. IEEE transactions on cybernetics **48**(5), 1460–1473 (2017)
34. Huang, G.Y., Liang, D.P., Hu, C.Z., Ren, J.D.: An algorithm for clustering heterogeneous data streams with uncertainty. In: 2010 International Conference on Machine Learning and Cybernetics. vol. 4, pp. 2059–2064. IEEE (2010)

35. Huang, J., Gong, S., Zhu, X.: Deep semantic clustering by partition confidence maximisation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8849–8858 (2020)
36. Hüllermeier, E.: Uncertainty in clustering and classification. In: Scalable Uncertainty Management: 4th International Conference, SUM 2010, Toulouse, France, September 27-29, 2010. Proceedings 4. pp. 16–19. Springer (2010)
37. Hussain, S.F., Butt, I.A., Hanif, M., Anwar, S.: Clustering uncertain graphs using ant colony optimization (aco). Neural Computing and Applications **34**(14), 11721–11738 (2022)
38. Hwang, C., Rhee, F.C.H.: Uncertain fuzzy clustering: Interval type-2 fuzzy approach to *c*-means. IEEE Transactions on fuzzy systems **15**(1), 107–120 (2007)
39. Jiang, B., Pei, J., Tao, Y., Lin, X.: Clustering uncertain data based on probability distribution similarity. IEEE Transactions on Knowledge and Data Engineering **25**(4), 751–763 (2011)
40. Jin, C., Yu, J.X., Zhou, A., Cao, F.: Efficient clustering of uncertain data streams. Knowledge and Information Systems **40**, 509–539 (2014)
41. Kamal, A., Dhakal, P., Javaid, A.Y., Devabhaktuni, V.K., Kaur, D., Zaientz, J., Marinier, R.: Recent advances and challenges in uncertainty visualization: a survey. Journal of Visualization **24**(5), 861–890 (2021)
42. Kang, B., Kim, S., Jung, H., Choe, J., Lee, K.: Efficient assessment of reservoir uncertainty using distance-based clustering: a review. Energies **12**(10), 1859 (2019)
43. Kao, B., Lee, S.D., Cheung, D.W., Ho, W.S., Chan, K.: Clustering uncertain data using voronoi diagrams. In: 2008 Eighth IEEE International Conference on Data Mining. pp. 333–342. IEEE (2008)
44. Kao, B., Lee, S.D., Lee, F.K., Cheung, D.W., Ho, W.S.: Clustering uncertain data using voronoi diagrams and r-tree index. IEEE Transactions on Knowledge and data engineering **22**(9), 1219–1233 (2010)
45. Karimi, J., Somers, T.M., Gupta, Y.P.: Impact of environmental uncertainty and task characteristics on user satisfaction with data. Information systems research **15**(2), 175–193 (2004)
46. Kriegel, H.P., Pfeifle, M.: Hierarchical density-based clustering of uncertain data. In: Fifth IEEE International Conference on Data Mining (ICDM'05). pp. 4–pp. IEEE (2005)
47. Kriegel, H.P., Pfeifle, M.: Density-based clustering of uncertain data. In: Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining. pp. 672–677 (2005)
48. Kuczenski, B.: False confidence: are we ignoring significant sources of uncertainty? The International Journal of Life Cycle Assessment **24**, 1760–1764 (2019)
49. Lee, C.S., Kao, Y.F., Kuo, Y.H., Wang, M.H.: Automated ontology construction for unstructured text documents. Data & Knowledge Engineering **60**(3), 547–566 (2007)
50. Li, Y., Chung, S.H.: Ride-sharing under travel time uncertainty: Robust optimization and clustering approaches. Computers & Industrial Engineering **149**, 106601 (2020)
51. Liu, H., Zhang, X., Zhang, X., Cui, Y.: Self-adapted mixture distance measure for clustering uncertain data. Knowledge-Based Systems **126**, 33–47 (2017)
52. Liu, L., Jin, R., Aggarwal, C., Shen, Y.: Reliable clustering on uncertain graphs. In: 2012 IEEE 12th international conference on data mining. pp. 459–468. IEEE (2012)
53. Liu, X., Lin, K.K., Andersen, B., Rattray, M.: Including probe-level uncertainty in model-based gene expression clustering. BMC bioinformatics **8**(1), 1–19 (2007)

54. Liu, Y., Liu, Z., Li, S., Guo, Y., Liu, Q., Wang, G.: Cloud-cluster: An uncertainty clustering algorithm based on cloud model. Knowledge-Based Systems **263**, 110261 (2023)
55. Luo, Q., Peng, Y., Peng, X., Saddik, A.E.: Uncertain data clustering-based distance estimation in wireless sensor networks. Sensors **14**(4), 6584–6605 (2014)
56. McMillan, H.K., Westerberg, I.K., Krueger, T.: Hydrological data uncertainty and its implications. Wiley Interdisciplinary Reviews: Water **5**(6), e1319 (2018)
57. Meesuksabai, W., Kangkachit, T., Waiyamai, K.: Hue-stream: Evolution-based clustering technique for heterogeneous data streams with uncertainty. In: Advanced Data Mining and Applications: 7th International Conference, ADMA 2011, Beijing, China, December 17-19, 2011, Proceedings, Part II 7. pp. 27–40. Springer (2011)
58. Naouali, S., Salem, S.B., Chtourou, Z.: Uncertainty mode selection in categorical clustering using the rough set theory. Expert Systems with Applications **158**, 113555 (2020)
59. Nuijten, M., Mittendorf, T., Persson, U.: Practical issues in handling data input and uncertainty in a budget impact analysis. The European journal of health economics **12**, 231–241 (2011)
60. Pelekis, N., Kopanakis, I., Kotsifakos, E., Frentzos, E., Theodoridis, Y.: Clustering trajectories of moving objects in an uncertain world. In: 2009 Ninth IEEE international conference on data mining. pp. 417–427. IEEE (2009)
61. Pelekis, N., Kopanakis, I., Kotsifakos, E.E., Frentzos, E., Theodoridis, Y.: Clustering uncertain trajectories. Knowledge and information systems **28**, 117–147 (2011)
62. Pileggi, S.F.: Ontological modelling and social networks: From expert validation to consolidated domains. In: International Conference on Computational Science. pp. 672–687. Springer (2023)
63. Prabhu, V., Chandrasekaran, A., Saenko, K., Hoffman, J.: Active domain adaptation via clustering uncertainty-weighted embeddings. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8505–8514 (2021)
64. Rendon, N., Giraldo, J.H., Bouwmans, T., Rodríguez-Buritica, S., Ramirez, E., Isaza, C.: Uncertainty clustering internal validity assessment using fréchet distance for unsupervised learning. Engineering Applications of Artificial Intelligence **124**, 106635 (2023)
65. Rhee, F.C.H.: Uncertain fuzzy clustering: Insights and recommendations. IEEE computational intelligence magazine **1**(2), 44–56 (2007)
66. Schubert, E., Koos, A., Emrich, T., Züfle, A., Schmid, K.A., Zimek, A.: A framework for clustering uncertain data. Proceedings of the VLDB Endowment **8**(12), 1976–1979 (2015)
67. Sharma, K.K., Seal, A.: Modeling uncertain data using monte carlo integration method for clustering. Expert systems with applications **137**, 100–116 (2019)
68. Sharma, K.K., Seal, A.: Multi-view spectral clustering for uncertain objects. Information Sciences **547**, 723–745 (2021)
69. Sharma, K.K., Seal, A.: Outlier-robust multi-view clustering for uncertain data. Knowledge-Based Systems **211**, 106567 (2021)
70. Shi, W., Chen, W.N., Gu, T., Jin, H., Zhang, J.: Handling uncertainty in financial decision making: a clustering estimation of distribution algorithm with simplified simulation. IEEE Transactions on Emerging Topics in Computational Intelligence **5**(1), 42–56 (2020)
71. Shukla, A.K., Muhuri, P.K.: Big-data clustering with interval type-2 fuzzy uncertainty modeling in gene expression datasets. Engineering Applications of Artificial Intelligence **77**, 268–282 (2019)

72. Shukla, A., Singh, S.: Clustering based unit commitment with wind power uncertainty. Energy Conversion and Management **111**, 89–102 (2016)
73. Sinaga, K.P., Yang, M.S.: Unsupervised k-means clustering algorithm. IEEE access **8**, 80716–80727 (2020)
74. Suzuki, R., Shimodaira, H.: Pvclust: an r package for assessing the uncertainty in hierarchical clustering. Bioinformatics **22**(12), 1540–1542 (2006)
75. Tabesh, M., Askari-Nasab, H.: Clustering mining blocks in presence of geological uncertainty. Mining Technology (2019)
76. Tew, C., Giraud-Carrier, C., Tanner, K., Burton, S.: Behavior-based clustering and analysis of interestingness measures for association rule mining. Data Mining and Knowledge Discovery **28**, 1004–1045 (2014)
77. Volk, P.B., Rosenthal, F., Hahmann, M., Habich, D., Lehner, W.: Clustering uncertain data with possible worlds. In: 2009 IEEE 25th International Conference on Data Engineering. pp. 1625–1632. IEEE (2009)
78. Wang, P., Ding, C., Tan, W., Gong, M., Jia, K., Tao, D.: Uncertainty-aware clustering for unsupervised domain adaptive object re-identification. IEEE Transactions on Multimedia (2022)
79. Wang, X., He, Y.: Learning from uncertainty for big data: future analytical challenges and strategies. IEEE Systems, Man, and Cybernetics Magazine **2**(2), 26–31 (2016)
80. Wauthier, F.L., Jojic, N., Jordan, M.I.: Active spectral clustering via iterative uncertainty reduction. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1339–1347 (2012)
81. Weng, C.H., Chen, Y.L.: Mining fuzzy association rules from uncertain data. Knowledge and Information Systems **23**, 129–152 (2010)
82. Wierzchoń, S.T., Kłopotek, M.A.: Modern algorithms of cluster analysis, vol. 34. Springer (2018)
83. Xia, Y., Xi, B.: Conceptual clustering categorical data with uncertainty. In: 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007). vol. 1, pp. 329–336. IEEE (2007)
84. Xiong, C., Johnson, D.M., Corso, J.J.: Active clustering with model-based uncertainty reduction. IEEE transactions on pattern analysis and machine intelligence **39**(1), 5–17 (2016)
85. Xu, L., Hu, Q., Hung, E., Chen, B., Tan, X., Liao, C.: Large margin clustering on uncertain data by considering probability distribution similarity. Neurocomputing **158**, 81–89 (2015)
86. Xu, R., Wunsch, D.: Survey of clustering algorithms. IEEE Transactions on neural networks **16**(3), 645–678 (2005)
87. Yao, C., Chen, M., Hong, Y.Y.: Novel adaptive multi-clustering algorithm-based optimal ess sizing in ship power system considering uncertainty. IEEE transactions on power systems **33**(1), 307–316 (2017)
88. Zhang, X., Liu, H., Zhang, X.: Novel density-based and hierarchical density-based clustering algorithms for uncertain data. Neural networks **93**, 240–255 (2017)
89. Zhou, J., Chen, L., Chen, C.P., Wang, Y., Li, H.X.: Uncertain data clustering in distributed peer-to-peer networks. IEEE transactions on neural networks and learning systems **29**(6), 2392–2406 (2017)
90. Züfle, A., Emrich, T., Schmid, K.A., Mamoulis, N., Zimek, A., Renz, M.: Representative clustering of uncertain data. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 243–252 (2014)