



Three-Way Decision for Handling Uncertainty in Machine Learning: A Narrative Review

Andrea Campagner, Federico Cabitza, and Davide Ciucci^(✉) 

Dipartimento di Informatica, Sistemistica e Comunicazione,
University of Milano-Bicocca, viale Sarca 336, 20126 Milan, Italy
davide.ciucci@unimib.it

Abstract. In this work we introduce a framework, based on three-way decision (TWD) and the trisecting-acting-outcome model, to handle uncertainty in Machine Learning (ML). We distinguish between handling uncertainty affecting the input of ML models, when TWD is used to identify and properly take into account the uncertain instances; and handling the uncertainty lying in the output, where TWD is used to allow the ML model to abstain. We then present a narrative review of the state of the art of applications of TWD in regard to the different areas of concern identified by the framework, and in so doing, we will highlight both the points of strength of the three-way methodology, and the opportunities for further research.

1 Introduction

Three-way decision (TWD) is a recent paradigm emerged from rough set theory (RST) that is acquiring its own status and visibility [46]. This paradigm is based on the simple idea of *thinking in three “dimensions”* (rather than in binary terms) when considering how to represent computational objects. This idea leads to the so-called *trisecting-acting-outcome* (TAO) model [82]: *Trisecting* addresses the question of how to divide the universe under investigation in three partitions; *Acting* explains how to deal with the three parts identified; and *Outcome* gives methodological indications on how to evaluate the adopted strategy.

Based on the TAO model, we propose a framework to handle uncertainty in Machine Learning: this model can be applied both to the input and the output of the Learning algorithm. Obviously, these two latter aspects are strictly related and they mutually affect each other in real applications. Schematically, the framework looks as illustrated in Table 1.

With reference to the table, we distinguish between applications that handle uncertainty in the input and those that handle uncertainty with respect to the output. By *uncertainty in the input* we mean different forms of uncertainty that are already explicitly present in the training datasets used by ML algorithms. By *uncertainty in the output* we mean mechanisms adopted by the ML algorithm in order to create more robust models or making the (inherent and partly insuppressible) predictive uncertainty more explicit.

Table 1. TAO model applied to Machine Learning

	Trisecting	Acting	Outcome
Input	The dataset contains different forms of uncertainty and it can be split in certain/uncertain instances	The ML-algorithm should take into account the dataset uncertainty and handle it	Ad-hoc measures should be introduced to quantify the dataset uncertainty, which should also be considered in the algorithm evaluation
Output	The output can contain instances with no decision (classification, clustering, etc.)	The ML algorithm abstains from giving the result on uncertain instances	New measures to evaluate ML algorithms with abstention should be introduced

In the following Sections, we will explain in more detail the different parts of the framework outlined in Table 1, and discuss the recent advances and current research in the framework areas by means of a narrative review of the literature indexed by the Google Scholar database. In particular, in Sect. 2, we describe the different steps of the proposed model with respect to the handling of uncertainty in the input, while in Sect. 3 we do the same for the handling of the uncertainty in the output. In Sect. 4, we will then discuss the advantages of incorporating TWD and the TAO model for uncertainty handling into Machine Learning, and some relevant future directions.

2 Handling Uncertainty in the Input

Real-world datasets are far from being perfect: typically they are affected by different forms of uncertainty (often missingness) that can be mainly related to either the data acquisition process or the complexity (e.g. in terms of volatility) of the phenomena under consideration or for both these factors.

These forms of uncertainty are usually distinguished in three common variants:

1. *Missing data*: this is usually the most common type of uncertainty in the input [6]. The dataset could contain missing values in its predictive features either because the original value was not recorded (e.g. the data was collected in two separate times, and the instrumentation to measure the feature was available only at one time), was subsequently lost or considered irrelevant (e.g. a doctor decided not to measure the BMI of a seemingly healthy person). This type of uncertainty has been the most studied, typically under the *data imputation* perspective, that is the task in which missing values are *filled in* before any subsequent ML process. This can be done in various ways, with techniques based on clustering [34,65], statistical or regression approaches [7], rough set or fuzzy rough set methods [4,51,67];

2. *Weak supervision*: in the case of supervised problems, the supervision (i.e. the target or decision variable) is only given in an imprecise form or only partially specified. This type of uncertainty has seen some increase in interest in the recent years [105], with a growing literature focusing specifically on *superset learning* [17, 29]; this is a specific type of weak supervision in which instances are associated with sets of *possible but mutually exclusive labels* that are guaranteed to contain the true value of the decision label;
3. *Multi-rater annotation*: this form of uncertainty is getting more and more impact due to the increasing use of *crowdsourcing* [5, 23, 69] for data annotation purposes, but it is also inherent in many domains where it is common (and in fact recommended) practice to involve multiple experts to increase the reliability of the Ground Truth, which is a crucial requirement in many situations where ML models are applied for sensitive or critical tasks (like in medicine for diagnostic tasks). Involving multiple raters who annotate the dataset independently of each others often results in multiple and conflicting decision labels for a given instance [9], for a common phenomenon that has been denoted with many expressions, like *observer variability* or *inter-rater reliability*.

While superficially similar (e.g. weak supervision could be seen as a form of missing data), the problems inherent to and the methods to handle these types of uncertainty are such that they should be distinguished. In the case of missing data, the main problem is to build reliable models of knowledge despite the incomplete information, and the *completion* of the dataset is but a means to an end, often under assumptions that are difficult to attain (or verify). In the case of weak supervision, on the other hand, the task of completion (which is usually called *disambiguation*) is of fundamental importance and the goal is, usually, to simultaneously build ML models and disambiguate the uncertain instances. Finally, in the case of multi-rater annotations, while the task of disambiguation is obviously present, there is also the problem of inferring the extent each single rater can be trusted (i.e., how accurate they are) and how to meaningfully aggregate the information they provide in order to build a consensus which is to be used to build the ground truth by which to train the ML model.

2.1 Trisecting and Acting Steps

In all three uncertainty forms, the *trisecting* act is at the basis of the process of uncertainty handling, as the uncertain instances (e.g., the instances missing some feature values, or those for which the provided annotations are only weak) must be necessarily recognised for any action to be considered: this also means that the trisecting act usually amounts to simply dividing the certain instances from the uncertain ones, and the bulk of the work is usually performed in the *acting* step in order to decide how differently handle the two kinds of instances. According to the three kinds of problems described at the beginning of the section, we present the following solutions.

Missing Data. Missing data is the type of uncertainty for which a TWD methodology to handle this kind of uncertainty is more mature, possibly because the problem has been well studied in RST and other theories for the management of uncertainty that are associated with TWD [21,22]. Most approaches in this direction have been based on the notion of *incomplete information table*, which is typically found in RST: Liu et al. [42] introduced a TWD model based on an incomplete information table augmented with interval-valued loss functions; Luo et al. [45] proposed a multi-step approach by which to distinguish different types of missing data (e.g. “don’t know”, “don’t care”) and similarity relations; Luo et al. [44] focused on how to update TWD in incomplete and multi-scale information systems using decision-theoretic rough sets; Sakai et al. [57–59] described an approach based on TWD to construct certain and possible rules using an algorithm which combines the classical A-priori algorithm [3] and possible world semantics [30]. Other approaches (not directly based on the incomplete information table notion) have also been considered: Nowicki et al. [52] proposed a TWD algorithm for classification with missing or interval-valued data based on rough sets and SVM; Yang et al. [75] proposed a method for TWD based on intuitionistic fuzzy sets that are construed based on a similarity relation of instances with missing values.

While all the above approaches propose techniques based on TWD with missing data for *classification* problems, there have also been proposals to deal with this type of uncertainty in *clustering*, starting from the original approach proposed by Yu [85,87], to deal with missing data in clustering using TWD: Afridi et al. [2] described an approach which is based, as for the classification case, on a simple trisecting step in which complete instances are used to produce an initial clustering and then use an approach based on game-theoretic rough sets to cluster the instances with missing values; Yang et al. [74] proposed a method for three-way clustering with missing data based on clustering density.

Weak Supervision. With respect to the case of weak supervision, the application of three-way based strategies is more recent and different techniques have been proposed in the recent years. Most of the work in this sense has focused on the specific case of *semi-supervised learning*, in which the uncertain instances have no supervision, and *active learning*, in which the missing labels can be requested to an external oracle (usually a human user) at some cost: Miao et al. [48] proposed a method for semi-supervised learning based on TWD; Yu et al. [88] proposed a three-way clustering approach for semi-supervised learning that uses an active learning approach to obtain labels for instances that are considered as uncertain after the initial clustering; Triff et al. [66] proposed an evolutionary semi-supervised algorithm based on rough sets and TWD and compare it with other algorithms obtaining interesting results when only the certainly classified objects are considered; Dai et al. [18] introduced a co-training technique for cost-sensitive semi-supervised learning based on sequential TWD and apply it to different standard ML algorithms (k-NN, PCA, LDA) in order to obtain a multi-view dataset; Campagner et al. [10,13] introduced a three-way Decision Tree model for semi-supervised learning and show that this model achieves

good performance with respect to standard ML algorithms for semi-supervised learning; Wang et al. [70,71] proposed a cost-sensitive three-way active learning algorithm based on the computation of label error statistics; Min et al. [49] proposed a cost-sensitive active learning strategy based on k-nearest neighbours and a tripartition of the instances in certain and uncertain ones.

In the case of more general weakly supervised learning, Campagner et al. [12] proposed a collection of approaches based on TWD and standard ML algorithms in order to take into account this type of uncertainty in the setting of *classification*. In particular, the authors considered an algorithm for Decision Tree (and ensemble-based extensions, such as Random Forest) learning, in which the trisecting and acting steps are dynamically and iteratively performed during the Decision Tree induction process on the basis of TWD and generalized information theory [33], and a generalized stochastic gradient descent algorithm based on interval analysis and TWD, in order to take into account the fact that the uncertain instances naturally determine interval-valued information with respect to the loss function to be optimized. In both cases, promising results were reported, showing that they outperform standard superset learning and semi-supervised techniques. A different approach, which is based on treating weakly supervision as a type of missing data, proposed by Sakai et al. [58], employs a three-way rule extraction algorithm that could also be applied in the case of weakly supervised data: this approach is of particular interest in that it suggests an integrated end-to-end approach to simultaneously handle missing data and weakly supervised data.

Multi-rater Annotation. With respect to the third type of uncertainty, that is multi-rater annotation, in [12] we noted that the issue has largely been ignored in the ML community. With respect to the application of TWD methodologies to handle this type of uncertainty, there has been some recent works with respect to aggregation methods and information fusion using TWD, mainly under the perspective of group decision making [25,39,53,96] and the modelling of multi-agent systems [76]. However, there has been so far a lack of studies concerning the application of these TWD based techniques to ML problems. Some related approaches have been explored under the perspective of *multi-source information tables* in RST, in which the multi-rater, and possibly conflicting, information is available not only for the decision variable but also for the predictor ones: Huang et al. [28] proposed a three-way concept learning method for multi-source data; Sang et al. [60] studied the application of decision-theoretic rough sets for TWD in multi-source information systems; Sang et al. [61] proposed an alternative approach which is not directly based on merging different information systems but instead it employs multi-granulation double-quantitative decision-theoretic rough set, which the authors show to be more fault tolerant with respect to traditional approaches. Campagner et al. [8,15] proposed a novel aggregation strategy, based on TWD, which can be applied to implement the trisecting step to handle the multi-rater annotation uncertainty type. In this case, the instances are categorized as certain or uncertain depending on the distribution of labels given by the raters and a set of parameters that have a cost-theoretic

interpretation. After the aggregation step, the problem is converted into a weakly supervised one and a learning algorithm is proposed that is shown to be significantly more effective than the traditional approach of simply assigning the most frequent labels (among the multi-rater annotations) to the uncertain instances.

2.2 Outcome Step: Evaluating the Results

All of the articles considered for this review mainly deal with the trisecting and acting step in the TAO model that we propose. The outcome step has rarely been considered and is usually addressed as it would be for traditional ML models: that is by simply considering the accuracy of the trained models, sometimes even in naive ways [14]. According to the framework that we propose, the main goal of employing TWD for ML is the handling of uncertainty. In this light, attention should also be placed on how much the TWD approach allows to reduce the initial uncertainty in the input data or at least to which degree the TWD-based algorithm is able to obtain good performances despite of the uncertainty. For example, with respect to the missing data problem, the outcome step should also consider the amount of missing values that have been correctly imputed (for imputation-based approaches), or the *robustness* of the induced ML algorithm with respect to different values that could be present in the missing features, for instance using interval-valued accuracy or information-theoretic metrics [11, 14], or by distinguishing which predictions made by the algorithm are certain (i.e., robust with respect to the missing values or the weakly supervised instance) or only possible. Similarly, with respect to the multi-rater annotation uncertainty type, besides the accuracy of the proposed approaches with respect to a known ground truth (when available), the outcome step should also consider the robustness of the proposed approach when varying the degree of conflicting information, and the level of noise of the raters who annotate the datasets, as we considered in [15]. In this sense, we believe that more attention should be put on the outcome step of the proposed framework, and further research in this sense should be performed.

3 Handling Uncertainty in the Output

The application of TWD to handle uncertainty in the output of the ML is a mature research area, and has possibly been considered since the original proposal of TWD, both for classification [80, 103] and for clustering [40]. In both cases, the uncertainty in the output of the ML model refers to the inability of the ML model to properly discriminate the instances and assign them a certain, precisely known, label. This could be due to a variety of issues: the chosen data representation (i.e., the selected features and/or their level of granularity) is not informative enough; the inability to distinguish different instances that are either identical or “too near” in the sample space, but are associated with different decision labels; the selected model class is not powerful enough to properly represent the concept to be learned. All these issues have been widely studied,

both under the perspective of RST with the notion of *indiscernibility* [54,55], and of more traditional ML approaches, with the notion of *decision boundary*. The approach suggested by TWD in this setting consists in allowing the classifier to abstain [81], even partially, that is excluding some of the possible alternative classifications. In so doing, the focus is on the trisecting step, which involves deciding on which instances the ML model (both for classification or clustering) should be considered uncertain, and hence the model should abstain on.

3.1 Trisecting and Acting Steps for Classification

With respect to classification, the traditional model of TWD applies only to binary classification cases, for which a third “uncertain” category is added, for which extensions of the most traditional ML methods are available. In all of the cases, the trisecting step is performed in a similar manner, on the basis of the original decision-theoretic rules proposed by Yao [81]; these rules are often embedded in different models, and the main variation relates to how the *acting step* is implemented. This step has usually been based on Bayesian decision analysis under the decision-theoretic rough set paradigm [31,36,79,103,104]. However, also other approaches to implement the acting step have been proposed, such as *structured approximations* in RST [27], or the combination of TWD with more traditional ML techniques, for instance, Deep Learning [37,100,101], optimization-based learning [41,43,95] or frequent pattern mining [38,50]: all of these implementations of the TWD model for the handling of uncertainty have been successfully applied to different fields, such as face recognition, spam filtering or recommender systems.

A particularly interesting use, with respect to the acting outcome, consists of integrating TWD in active learning methodologies: Chen et al. [16] proposed a three-way rule-based decision algorithm that employs active learning to re-classify the uncertain instances; Zhang et al. [94] proposed a random forest-based recommender systems with the capability to ask for user supervision on uncertain objects; Yao et al. [78] proposed a TWD model based on a game-theoretic rough set for medical decision systems that distinguish *certain rules* (for acceptance and rejection) from *deferment rules* which require intervention from the user.

In recent years, different proposals have also been considered for the extension to the multi-class case, mainly under two major approaches. The first one is based on sequential TWD [83], which essentially implements a hierarchical one-vs-all learning scheme; Yang et al. [77] considered a Bayesian extension of multi-class decision theoretic rough sets [102]; Savchenko [62,63] proposed sequential TWD and granular computing for speed-up of image classification when the number of classes is large; Zhang et al. [98] proposed a sequential TWD model based on the use of autoencoders for granular feature extraction. The second approach, which can be defined as natively multi-class, has been proposed by some authors (e.g., in [11,12]): it employs a decision-theoretic procedure to convert every standard probabilistic classifier into a multi-class TWD classifier. A similar approach, but based on decision-theoretic rough sets, have also been developed by Jia et al. [32].

While all the approaches mentioned above consider the combination of TWD and ML models in a *a posteriori* strategy in which the trisecting step is performed after, or as a consequence of, the standard ML training procedure, in [11, 12] we also considered how to directly embed TWD in the training algorithm of a wide class of standard ML models, either by a direct modification of the learning algorithm (for decision trees and related methods), or by adopting ad-hoc regularized loss functions (for optimization-based procedures such as SVM or logistic regression).

3.2 Trisecting and Acting Steps for Clustering

In regards clustering, various approaches have been proposed to implement the TWD-based handling of the uncertainty in the output, hence to construct clusterings in which the assignment of some instances to clusters is uncertain, mainly under the frameworks of *rough clustering* [40], *interval-set clustering* [84] and *three-way clustering* [90]. In all of the above approaches, the trisecting step is implemented as a modification of standard clustering assignment criteria, and it allows instances to be considered as *uncertain* with respect to their assignment to one or more clusters: Yu [90] proposed a three-way clustering algorithm that also works with incomplete data; Wang et al. [73] proposed a three-way clustering method based on mathematical morphology; Yu et al. [91] considered a flexible tree-based incremental three-way clustering algorithm; Yu et al. [86] proposed an optimized ensemble-based three-way clustering algorithm for large-scale datasets; Afridi et al. [1] proposed a variance-based three-way clustering algorithm; Zhang et al. [99] proposed a novel improvement on the original rough k-means based on a weighted Gaussian distance function; Li et al. [35] extended standard rough k-means with an approach based on decision-theoretic rough sets, Yu et al. [89] proposed a hybrid clustering/active learning based on TWD for multi-view data; Zhang [97] proposed a three-way c-means algorithm; Wang et al. [72] proposed a refinement three-way clustering algorithm based on the re-clustering of ensemble of traditional hard clustering algorithms; Yu et al. [93] proposed a density three-way clustering algorithm based on DBscan; Yu et al. [92] proposed a three-way clustering algorithm optimized for high-dimensionality datasets based on a modification of the k-medoids algorithm and the random projection method; Hu et al. [26] proposed a sequential TWD model for consensus clustering based on the notion of co-association matrix.

3.3 Outcome Step: Evaluating the Results

With respect to the outcome step, both clustering and classification techniques based on TWD have been shown to significantly improve the performance in comparison to traditional ML algorithms (see the referenced literature). Despite this promising assessment, one should also consider that the evaluation of ML algorithms using TWD to handle the uncertainty in output, at least in principle, cannot be made on the same grounds of traditional ML models (i.e., only on the basis of accuracy metrics). Indeed, since these models are allowed to *abstain* on

uncertain instances, metrics for their evaluation should take into account the trade-off between the accuracy on the classified/clustered instances but also the *coverage* of the algorithm, that is on how many instances the model defers its decision. As an example of this issue, suffice it to consider that a three-way classifier that abstains on all the instances but one, which is correctly classified/clustered, has perfect accuracy but it is hardly a useful predictive model. However, attention towards this trade-off has emerged only recently, where the majority of the surveyed papers only focus on the accuracy of the models on the classified/clustered instances: Peters [56] proposed a modified Davis-Bouldin index for evaluation of three-way clustering; Depaolini et al. [19] proposed generalizations of Rand, Jaccard and Fowlkes-Mallows indices; similarly, we proposed a generalization of information-theoretic measures of clustering quality [14] and generalization of accuracy metrics for classification [11]. Promisingly, the superior performance of TWD techniques for the handling of output uncertainty can be observed also under these more robust, and conservative, metrics.

4 Discussion

In this article, we proposed a TAO model for the management of uncertainty in Machine Learning that is based on TWD. After describing the proposed framework, we have reviewed the current state of the art for the different areas of concern identified by our framework, and discussed about the strengths, limitations and areas requiring further investigation of the main works considered.

In what follows, we emphasise both what we believe are the main advantages of adopting this methodology in ML and also delineate some topics that in our opinion are particularly in need of further study.

4.1 Advantages of Three-Way ML

It is undeniable that in the recent years, the application of TWD and the TAO model to ML applications has been growing and showing promising results. In this Section, we will emphasise the advantages of TWD under the perspective of uncertainty handling for ML. In this perspective, TWD and the TAO model look promising as a means to provide a principled way to handle uncertainty in the ML process in an end-to-end fashion, by directly using the information obtained in the *trisecting* act (i.e., the splitting of instances into certain/uncertain ones), in the subsequent *acting* and *outcome* steps, without the need to address and “correct” the uncertainty in a separate pre-processing step. This is particularly clear in our discussion about the *handling of the uncertainty in the input*: in this case, the TAO model enables one to directly deal with different forms of data uncertainty in a theoretically-sound, robust and *non-invasive* manner [20], while also obtaining higher predictive accuracy than with traditional ML methodologies. The same holds true also with respect to the handling of uncertainty in the output. In this case, the TAO model allows to obtain classifiers that are both more accurate and robust, thanks to the possibility of abstention that allows

more conservative decision boundaries. Abstention is a more *informative* strategy also from a decision-support perspective, in that the model that is enhanced with TWD can expose its predictive uncertainty by abstaining as a sign that the situation needs more information, or the careful consideration of the human decision maker.

4.2 Future Directions

Despite the increasing popularity of TWD to handle the uncertainty in ML pipelines, and the relative maturity of the application of this methodology with respect to the trisecting and acting steps of our framework (see Table 1), we believe that some specific aspects merit further investigations. Then, as already discussed in Sects. 2 and 3, the outcome step has not been sufficiently explored, especially with respect to the handling of uncertainty in the input. As discussed in Sect. 2.2, we believe that conceiving appropriate metrics to assess the robustness of TWD methods represents a particularly promising strand of research, which would also enable *counterfactual*-like reasoning [47] for ML models, a topic that has recently been considered important in the light of *eXplainable AI* [68]. For instance, this can be done by analyzing the robustness and performance of the ML models with respect to specific counterfactual instantiations of the instances affected by uncertainty that would most likely alter the learnt decision boundary. Similarly, while there have been more proposals for the outcome step for the output part of our framework, we believe that further work should be done towards the general adoption of these measures in the application of TWD-based ML. Similarly, a second promising direction of research regards the acting step for the management of the uncertainty in the output: as we previously discussed, active learning and *human-in-the-loop* [24] techniques to handle the instances recognized as uncertain by the ML algorithms are of particular interest. Similarly, it would be interesting to study the connection between the TWD model to handle uncertainty in the output and the *conformal prediction* paradigm [64], as both are based on the idea of providing set-valued predictions on uncertain instances. A third research direction regards the fact that the different steps have currently been studied mostly in isolation: so far, most studies applying TWD in ML focused either on the input *or* the output part of our framework. While some initial works with respect to a unified treatment of both types of uncertainty have recently been considered [12], we believe that further work toward such a uniform methodology would be particularly promising. Finally, missing data is usually understood as a problem of completeness: this is missing data at feature level, for instances at least partly observed. But there is also a “missingness” at row level, that is a source of uncertainty (which makes the data we have uncertain and less reliable) that regards instances that we have not observed or whose characteristics are not well represented in the data collected: more research is due to how TWD can tackle this important source of bias, which is usually called sampling bias.

References

1. Afridi, M.K., Azam, N., Yao, J.: Variance based three-way clustering approaches for handling overlapping clustering. *IJAR* **118**, 47–63 (2020)
2. Afridi, M.K., Azam, N., Yao, J., et al.: A three-way clustering approach for handling missing data using GTRS. *IJAR* **98**, 11–24 (2018)
3. Agrawal, R., Srikant, R., et al.: Fast algorithms for mining association rules. In: *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB*, vol. 1215, pp. 487–499 (1994)
4. Amiri, M., Jensen, R.: Missing data imputation using fuzzy-rough methods. *Neurocomputing* **205**, 152–164 (2016)
5. Awasthi, P., Blum, A., Haghtalab, N., et al.: Efficient PAC learning from the crowd. arXiv preprint [arXiv:1703.07432](https://arxiv.org/abs/1703.07432) (2017)
6. Brown, M.L., Kros, J.F.: Data mining and the impact of missing data. *Ind. Manag. Data Syst.* **103**(8), 611–621 (2003)
7. Buuren, S.V., Groothuis-Oudshoorn, K.: Mice: multivariate imputation by chained equations in R. *J. Stat. Softw.* **45**(3), 1–67 (2010)
8. Cabitza, F., Campagner, A., Ciucci, D.: New frontiers in explainable AI: understanding the GI to interpret the GO. In: Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) *CD-MAKE 2019*. LNCS, vol. 11713, pp. 27–47. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-29726-8_3
9. Cabitza, F., Locoro, A., Alderighi, C., et al.: The elephant in the record: on the multiplicity of data recording work. *Health Inform. J.* **25**(3), 475–490 (2019)
10. Campagner, A., Cabitza, F., Ciucci, D.: Exploring medical data classification with three-way decision tree. In: *Proceedings of BIOSTEC 2019 - Volume 5: HEALTHINF*, pp. 147–158. SCITEPRESS (2019)
11. Campagner, A., Cabitza, F., Ciucci, D.: Three-way classification: ambiguity and abstention in machine learning. In: Mihálydeák, T., et al. (eds.) *IJCERS 2019*. LNCS (LNAI), vol. 11499, pp. 280–294. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22815-6_22
12. Campagner, A., Cabitza, F., Ciucci, D.: The three-way-in and three-way-out framework to treat and exploit ambiguity in data. *IJAR* **119**, 292–312 (2020)
13. Campagner, A., Ciucci, D.: Three-way and semi-supervised decision tree learning based on orthopartitions. In: Medina, J., et al. (eds.) *IPMU 2018*. CCIS, vol. 854, pp. 748–759. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-91476-3_61
14. Campagner, A., Ciucci, D.: Orthopartitions and soft clustering: soft mutual information measures for clustering validation. *Knowl.-Based Syst.* **180**, 51–61 (2019)
15. Campagner, A., Ciucci, D., Svensson, C.M., et al.: Ground truthing from multi-rater labelling with three-way decisions and possibility theory. *IEEE Trans. Fuzzy Syst.* (2020, submitted)
16. Chen, Y., Yue, X., Fujita, H., et al.: Three-way decision support for diagnosis on focal liver lesions. *Knowl.-Based Syst.* **127**, 85–99 (2017)
17. Cour, T., Sapp, B., Taskar, B.: Learning from partial labels. *J. Mach. Learn. Res.* **12**, 1501–1536 (2011)
18. Dai, D., Zhou, X., Li, H., et al.: Co-training based sequential three-way decisions for cost-sensitive classification. In: *2019 IEEE 16th ICNSC*, pp. 157–162 (2019)
19. Depaolini, M.R., Ciucci, D., Calegari, S., Dominoni, M.: External indices for rough clustering. In: Nguyen, H.S., Ha, Q.-T., Li, T., Przybyła-Kasperek, M. (eds.) *IJCERS 2018*. LNCS (LNAI), vol. 11103, pp. 378–391. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99368-3_29

20. Düntsch, I., Gediga, G.: Rough set data analysis—a road to non-invasive knowledge discovery. *Methodos* (2000)
21. Greco, S., Matarazzo, B., Slowinski, R.: Dealing with missing data in rough set analysis of multi-attribute and multi-criteria decision problems. In: Zanakis, S.H., Doukidis, G., Zopounidis, C. (eds.) *Decision Making: Recent Developments and Worldwide Applications*, pp. 295–316. Springer, Boston (2000). https://doi.org/10.1007/978-1-4757-4919-9_20
22. Grzymala-Busse, J.W., Hu, M.: A comparison of several approaches to missing attribute values in data mining. In: Ziarko, W., Yao, Y. (eds.) *RSCTC 2000. LNCS (LNAI)*, vol. 2005, pp. 378–385. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-45554-X_46
23. Heinecke, S., Reyzin, L.: Crowdsourced PAC learning under classification noise. In: *Proceedings of AAAI HCOMP 2019*, vol. 7, pp. 41–49 (2019)
24. Holzinger, A.: Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Inf.* **3**(2), 119–131 (2016)
25. Hu, B.Q., Wong, H., Yiu, K.F.C.: The aggregation of multiple three-way decision spaces. *Knowl.-Based Syst.* **98**, 241–249 (2016)
26. Hu, M., Deng, X., Yao, Y.: A sequential three-way approach to constructing a co-association matrix in consensus clustering. In: Nguyen, H.S., Ha, Q.-T., Li, T., Przybyła-Kasperek, M. (eds.) *IJCRS 2018. LNCS (LNAI)*, vol. 11103, pp. 599–613. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-99368-3_47
27. Hu, M., Yao, Y.: Structured approximations as a basis for three-way decisions in rough set theory. *Knowl.-Based Syst.* **165**, 92–109 (2019)
28. Huang, C., Li, J., Mei, C., et al.: Three-way concept learning based on cognitive operators: an information fusion viewpoint. *IJAR* **83**, 218–242 (2017)
29. Hüllermeier, E., Cheng, W.: Superset learning based on generalized loss minimization. In: Appice, A., Rodrigues, P.P., Santos Costa, V., Gama, J., Jorge, A., Soares, C. (eds.) *ECML PKDD 2015. LNCS (LNAI)*, vol. 9285, pp. 260–275. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23525-7_16
30. Imieliński, T., Lipski Jr., W.: Incomplete information in relational databases. *J. ACM* **31**(4), 761–791 (1984)
31. Jia, X., Deng, Z., Min, F., Liu, D.: Three-way decisions based feature fusion for chinese irony detection. *IJAR* **113**, 324–335 (2019)
32. Jia, X., Li, W., Shang, L.: A multiphase cost-sensitive learning method based on the multiclass three-way decision-theoretic rough set model. *Inf. Sci.* **485**, 248–262 (2019)
33. Klir, G.J., Wierman, M.J.: *Uncertainty-based information: elements of generalized information theory*, vol. 15. Physica (2013)
34. Li, D., Deogun, J., Spaulding, W., Shuart, B.: Towards missing data imputation: a study of fuzzy k-means clustering method. In: Tsumoto, S., Słowiński, R., Komorowski, J., Grzymala-Busse, J.W. (eds.) *RSCTC 2004. LNCS (LNAI)*, vol. 3066, pp. 573–579. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-25929-9_70
35. Li, F., Ye, M., Chen, X.: An extension to rough c-means clustering based on decision-theoretic rough sets model. *IJAR* **55**(1), 116–129 (2014)
36. Li, H., Zhang, L., Huang, B., et al.: Sequential three-way decision and granulation for cost-sensitive face recognition. *Knowl.-Based Syst.* **91**, 241–251 (2016)
37. Li, H., Zhang, L., Zhou, X., et al.: Cost-sensitive sequential three-way decision modeling using a deep neural network. *IJAR* **85**, 68–78 (2017)

38. Li, Y., Zhang, Z.H., Chen, W.B., et al.: TDUP: an approach to incremental mining of frequent itemsets with three-way-decision pattern updating. *IJMLC* **8**(2), 441–453 (2017). <https://doi.org/10.1007/s13042-015-0337-6>
39. Liang, D., Pedrycz, W., Liu, D., Hu, P.: Three-way decisions based on decision-theoretic rough sets under linguistic assessment with the aid of group decision making. *Appl. Soft Comput.* **29**, 256–269 (2015)
40. Lingras, P., West, C.: Interval set clustering of web users with rough k-means. Technical report 2002-002, Department of Mathematics and Computing Science, St. Mary's University, Halifax, NS, Canada (2002)
41. Liu, D., Li, T., Liang, D.: Incorporating logistic regression to decision-theoretic rough sets for classifications. *IJAR* **55**(1), 197–210 (2014)
42. Liu, D., Liang, D., Wang, C.: A novel three-way decision model based on incomplete information system. *Knowl.-Based Syst.* **91**, 32–45 (2016). Three-way Decisions and Granular Computing
43. Liu, J., Li, H., Zhou, X., et al.: An optimization-based formulation for three-way decisions. *Inf. Sci.* **495**, 185–214 (2019)
44. Luo, C., Li, T., Huang, Y., et al.: Updating three-way decisions in incomplete multi-scale information systems. *Inf. Sci.* **476**, 274–289 (2019)
45. Luo, J., Fujita, H., Yao, Y., Qin, K.: On modeling similarity and three-way decision under incomplete information in rough set theory. *Knowl.-Based Syst.* **191**, 105251 (2020)
46. Ma, M.: Advances in three-way decisions and granular computing. *Knowl.-Based Syst.* **91**, 1–3 (2016)
47. Mandel, D.R.: Counterfactual and causal explanation: from early theoretical views to new frontiers. In: *The Psychology of Counterfactual Thinking*, pp. 23–39. Routledge (2007)
48. Miao, D., Gao, C., Zhang, N.: Three-way decisions-based semi-supervised learning. In: *Theory and Applications of Three-Way Decisions*, pp. 17–33 (2012)
49. Min, F., Liu, F.L., Wen, L.Y., et al.: Tri-partition cost-sensitive active learning through kNN. *Soft. Comput.* **23**(5), 1557–1572 (2019)
50. Min, F., Zhang, Z.H., Zhai, W.J., et al.: Frequent pattern discovery with tri-partition alphabets. *Inf. Sci.* **507**, 715–732 (2020)
51. Nelwamondo, F.V., Marwala, T.: Rough set theory for the treatment of incomplete data. In: *2007 IEEE International Fuzzy Systems Conference*, pp. 1–6. IEEE (2007)
52. Nowicki, R.K., Grzanek, K., Hayashi, Y.: Rough support vector machine for classification with interval and incomplete data. *J. Artif. Intell. Soft Comput. Res.* **10**(1), 47–56 (2020)
53. Pang, J., Guan, X., Liang, J., Wang, B., Song, P.: Multi-attribute group decision-making method based on multi-granulation weights and three-way decisions. *IJAR* **117**, 122–147 (2020)
54. Pawlak, Z.: *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer (1991)
55. Pawlak, Z., Skowron, A.: Rough sets: some extensions. *Inf. Sci.* **177**(1), 28–40 (2007)
56. Peters, G.: Rough clustering utilizing the principle of indifference. *Inf. Sci.* **277**, 358–374 (2014)
57. Sakai, H., Nakata, M.: Rough set-based rule generation and apriori-based rule generation from table data sets: a survey and a combination. *CAAI Trans. Intell. Technol.* **4**(4), 203–213 (2019)

58. Sakai, H., Nakata, M., Watada, J.: NIS-apriori-based rule generation with three-way decisions and its application system in SQL. *Inf. Sci.* **507**, 755–771 (2020)
59. Sakai, H., Nakata, M., Yao, Y.: Pawlak’s many valued information system, non-deterministic information system, and a proposal of new topics on information incompleteness toward the actual application. In: Wang, G., Skowron, A., Yao, Y., Ślęzak, D., Polkowski, L. (eds.) *Thriving Rough Sets. SCI*, vol. 708, pp. 187–204. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-54966-8_9
60. Sang, B., Guo, Y., Shi, D., et al.: Decision-theoretic rough set model of multi-source decision systems. *IJMLC* **9**(11), 1941–1954 (2018)
61. Sang, B., Yang, L., Chen, H., et al.: Generalized multi-granulation double-quantitative decision-theoretic rough set of multi-source information system. *IJAR* **115**, 157–179 (2019)
62. Savchenko, A.V.: Fast multi-class recognition of piecewise regular objects based on sequential three-way decisions and granular computing. *Knowl.-Based Syst.* **91**, 252–262 (2016)
63. Savchenko, A.V.: Sequential three-way decisions in multi-category image recognition with deep features based on distance factor. *Inf. Sci.* **489**, 18–36 (2019)
64. Shafer, G., Vovk, V.: A tutorial on conformal prediction. *J. Mach. Learn. Res.* **9**(Mar), 371–421 (2008)
65. Tian, J., Yu, B., Yu, D., Ma, S.: Missing data analyses: a hybrid multiple imputation algorithm using gray system theory and entropy based on clustering. *Appl. Intell.* **40**(2), 376–388 (2014)
66. Triff, M., Wiechert, G., Lingras, P.: Nonlinear classification, linear clustering, evolutionary semi-supervised three-way decisions: a comparison. *FUZZ-IEEE* **2017**, 1–6 (2017)
67. W. Grzymala-Busse, J.: Rough set strategies to data with missing attribute values. In: *Proceedings of ISMIS 2005*, vol. 542, pp. 197–212 (2005)
68. Wachter, S., Mittelstadt, B., Russell, C.: Counterfactual explanations without opening the black box: automated decisions and the GDPR. *Harv. JL Tech.* **31**, 841 (2017)
69. Wang, L., Zhou, Z.H.: Cost-saving effect of crowdsourcing learning. In: *IJCAI*, pp. 2111–2117 (2016)
70. Wang, M., Fu, K., Min, F., Jia, X.: Active learning through label error statistical methods. *Knowl.-Based Syst.* **189**, 105140 (2020)
71. Wang, M., Lin, Y., Min, F., Liu, D.: Cost-sensitive active learning through statistical methods. *Inf. Sci.* **501**, 460–482 (2019)
72. Wang, P., Liu, Q., Yang, X., Xu, F.: Ensemble re-clustering: refinement of hard clustering by three-way strategy. In: Sun, Y., Lu, H., Zhang, L., Yang, J., Huang, H. (eds.) *IScIDE 2017. LNCS*, vol. 10559, pp. 423–430. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67777-4_37
73. Wang, P., Yao, Y.: Ce3: a three-way clustering method based on mathematical morphology. *Knowl.-Based Syst.* **155**, 54–65 (2018)
74. Yang, L., Hou, K.: A method of incomplete data three-way clustering based on density peaks. In: *AIP Conference Proceedings*, vol. 1967, p. 020008. AIP Publishing LLC (2018)
75. Yang, X., Tan, A.: Three-way decisions based on intuitionistic fuzzy sets. In: Polkowski, L., Yao, Y., Artiemjew, P., Ciucci, D., Liu, D., Ślęzak, D., Zielosko, B. (eds.) *IJCRS 2017. LNCS (LNAI)*, vol. 10314, pp. 290–299. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60840-2_21
76. Yang, X., Yao, J.: Modelling multi-agent three-way decisions with decision-theoretic rough sets. *Fundam. Inform.* **115**(2–3), 157–171 (2012)

77. Yang, X., Li, T., Fujita, H., Liu, D.: A sequential three-way approach to multi-class decision. *IJAR* **104**, 108–125 (2019)
78. Yao, J., Azam, N.: Web-based medical decision support systems for three-way medical decision making with game-theoretic rough sets. *IEEE Trans. Fuzzy Syst.* **23**(1), 3–15 (2014)
79. Yao, Y.: Three-way decision: an interpretation of rules in rough set theory. In: Wen, P., Li, Y., Polkowski, L., Yao, Y., Tsumoto, S., Wang, G. (eds.) *RSKT 2009*. LNCS (LNAI), vol. 5589, pp. 642–649. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-02962-2_81
80. Yao, Y.: Three-way decisions with probabilistic rough sets. *Inf. Sci.* **180**(3), 341–353 (2010)
81. Yao, Y.: An outline of a theory of three-way decisions. In: Yao, J.T., et al. (eds.) *RSCCTC 2012*. LNCS (LNAI), vol. 7413, pp. 1–17. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-32115-3_1
82. Yao, Y.: Three-way decision and granular computing. *Int. J. Approx. Reason.* **103**, 107–123 (2018)
83. Yao, Y., Deng, X.: Sequential three-way decisions with probabilistic rough sets. In: *Proceedings of IEEE ICCI-CC 2011*, pp. 120–125. IEEE (2011)
84. Yao, Y., Lingras, P., Wang, R., Miao, D.: Interval set cluster analysis: a reformulation. In: Sakai, H., Chakraborty, M.K., Hassaniien, A.E., Ślęzak, D., Zhu, W. (eds.) *RSFDGrC 2009*. LNCS (LNAI), vol. 5908, pp. 398–405. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-10646-0_48
85. Yu, H.: A framework of three-way cluster analysis. In: Polkowski, L., et al. (eds.) *IJCRS 2017*. LNCS (LNAI), vol. 10314, pp. 300–312. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60840-2_22
86. Yu, H., Chen, Y., Lingras, P., et al.: A three-way cluster ensemble approach for large-scale data. *IJAR* **115**, 32–49 (2019)
87. Yu, H., Su, T., Zeng, X.: A three-way decisions clustering algorithm for incomplete data. In: Miao, D., Pedrycz, W., Ślęzak, D., Peters, G., Hu, Q., Wang, R. (eds.) *RSKT 2014*. LNCS (LNAI), vol. 8818, pp. 765–776. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11740-9_70
88. Yu, H., Wang, X., Wang, G.: A semi-supervised three-way clustering framework for multi-view data. In: Polkowski, L., et al. (eds.) *IJCRS 2017*. LNCS (LNAI), vol. 10314, pp. 313–325. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-60840-2_23
89. Yu, H., Wang, X., Wang, G., et al.: An active three-way clustering method via low-rank matrices for multi-view data. *Inf. Sci.* **507**, 823–839 (2020)
90. Yu, H., Wang, Y.: Three-way decisions method for overlapping clustering. In: Yao, J.T., et al. (eds.) *RSCCTC 2012*. LNCS (LNAI), vol. 7413, pp. 277–286. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-32115-3_33
91. Yu, H., Zhang, C., Wang, G.: A tree-based incremental overlapping clustering method using the three-way decision theory. *Knowl.-Based Syst.* **91**, 189–203 (2016)
92. Yu, H., Zhang, H.: A three-way decision clustering approach for high dimensional data. In: Flores, V., et al. (eds.) *IJCRS 2016*. LNCS (LNAI), vol. 9920, pp. 229–239. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-47160-0_21
93. Yu, H., Chen, L., Yao, J., et al.: A three-way clustering method based on an improved dbscan algorithm. *Phys. A* **535**, 122289 (2019)
94. Zhang, H.R., Min, F.: Three-way recommender systems based on random forests. *Knowl.-Based Syst.* **91**, 275–286 (2016)

95. Zhang, H.R., Min, F., Shi, B.: Regression-based three-way recommendation. *Inf. Sci.* **378**, 444–461 (2017)
96. Zhang, H.Y., Yang, S.Y.: Three-way group decisions with interval-valued decision-theoretic rough sets based on aggregating inclusion measures. *IJAR* **110**, 31–45 (2019)
97. Zhang, K.: A three-way c-means algorithm. *Appl. Soft Comput.* **82**, 105536 (2019)
98. Zhang, L., Li, H., Zhou, X., et al.: Sequential three-way decision based on multi-granular autoencoder features. *Inf. Sci.* **507**, 630–643 (2020)
99. Zhang, T., Ma, F.: Improved rough k-means clustering algorithm based on weighted distance measure with gaussian function. *Int. J. Comput. Math.* **94**(4), 663–675 (2017)
100. Zhang, Y., Miao, D., Wang, J., et al.: A cost-sensitive three-way combination technique for ensemble learning in sentiment classification. *IJAR* **105**, 85–97 (2019)
101. Zhang, Y., Zhang, Z., Miao, D., et al.: Three-way enhanced convolutional neural networks for sentence-level sentiment classification. *Inf. Sci.* **477**, 55–64 (2019)
102. Zhou, B.: Multi-class decision-theoretic rough sets. *IJAR* **55**(1), 211–224 (2014)
103. Zhou, B., Yao, Y., Luo, J.: A three-way decision approach to email spam filtering. In: Farzindar, A., Kešelj, V. (eds.) *AI 2010. LNCS (LNAI)*, vol. 6085, pp. 28–39. Springer, Heidelberg (2010). https://doi.org/10.1007/978-3-642-13059-5_6
104. Zhou, B., Yao, Y., Luo, J.: Cost-sensitive three-way email spam filtering. *JIS* **42**(1), 19–45 (2014)
105. Zhou, Z.H.: A brief introduction to weakly supervised learning. *Natl. Sci. Rev.* **5**(1), 44–53 (2018)