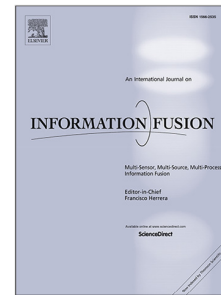


Journal Pre-proof

A survey on uncertainty reasoning and quantification in belief theory and its application to deep learning

Zhen Guo, Zelin Wan, Qisheng Zhang, Xujiang Zhao, Qi Zhang, Lance M. Kaplan, Audun Jøsang, Dong H. Jeong, Feng Chen, Jin-Hee Cho



PII: S1566-2535(23)00303-2
DOI: <https://doi.org/10.1016/j.inffus.2023.101987>
Reference: INFFUS 101987

To appear in: *Information Fusion*

Received date : 29 May 2023
Revised date : 22 August 2023
Accepted date : 23 August 2023

Please cite this article as: Z. Guo, Z. Wan, Q. Zhang et al., A survey on uncertainty reasoning and quantification in belief theory and its application to deep learning, *Information Fusion* (2023), doi: <https://doi.org/10.1016/j.inffus.2023.101987>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2023 Elsevier B.V. All rights reserved.

A Survey on Uncertainty Reasoning and Quantification in Belief Theory and Its Application to Deep Learning

Zhen Guo^{*a}, Zelin Wan^{*a}, Qisheng Zhang^{*a}, Xujiang Zhao^{*f}, Qi Zhang^a, Lance M. Kaplan^c, Audun Jøsang^e, Dong H. Jeong^d, Feng Chen^b, Jin-Hee Cho^a

^aDepartment of Computer Science, Virginia Tech, 7054 Haycock Road, Falls Church, 22043, VA, USA.

^bDepartment of Computer Science, University of Texas at Dallas, 800 W Campbell Rd, Richardson, 75080, TX, USA.

^cUS Army Research Laboratory, 2800 Powder Mill Rd, Adelphi, 20783, MD, USA.

^dDepartment of Computer Science of Information Technology, University of the District of Columbia, 4200 Connecticut Ave NW, 20008, Washington, DC, USA.

^eDepartment of Informatics, University of Oslo, Ole-Johan Dahls hus Gaustadalleen, Oslo, 23b 0373, Norway.

^fNEC Laboratories America Inc, 4 Independence Way, Suite 200, Princeton, 08540, NJ, USA.

Authors with * contributed equally to this work.

Abstract

An in-depth understanding of uncertainty is the first step to making effective decisions under uncertainty. Machine/deep learning (ML/DL) has been hugely leveraged to solve complex problems involved with processing high-dimensional data. However, reasoning and quantifying different uncertainties to achieve effective decision-making have been much less explored in ML/DL than in other Artificial Intelligence (AI) domains. In particular, belief/evidence theories have been studied in Knowledge representation and reasoning (KRR) since the 1960s to reason and measure uncertainties to enhance decision-making effectiveness. Based on our in-depth literature review, only a few studies have leveraged mature uncertainty research in belief/evidence theories in ML/DL to tackle complex problems under different types of uncertainty. Our present survey paper discusses major belief theories and their core ideas dealing with uncertainty causes and types and quantifying them, along with the discussions of their applicability in ML/DL. Particularly, we discuss three main approaches leveraging belief theories in Deep Neural Networks (DNNs), including Evidential DNNs, Fuzzy DNNs, and Rough DNNs, in terms of their uncertainty causes, types, and quantification methods along with their applicability in diverse problem domains. Through an in-depth understanding of the extensive survey on this topic, we discuss insights, lessons learned, limitations of the current state-of-the-art bridging belief theories and ML/DL, and future research directions. This paper conducts an extensive survey by bridging belief theories and deep learning in reasoning and quantifying uncertainty to help researchers initiate uncertainty and decision-making research.

Keywords: Belief theory, uncertainty reasoning, uncertainty quantification, decision making, machine/deep learning

PACS: 0000, 1111

2000 MSC: 0000, 1111

1. Introduction

1.1. Motivation

An in-depth understanding of uncertainty is a must to make effective and efficient decisions. Uncertainty reasoning and quantification in decision-making have been studied for several decades in various artificial intelligence (AI) domains (e.g., belief/evidence theory, game theory, and machine/deep learning). However, different uncertainty estimates considering other root causes of uncertainty have not been sufficiently examined. The Internet and Big Data era have brought a flood of information for decision-making. Yet, a large volume of information does not allow effective and efficient decision-making because of unavailable, unreliable, incomplete, or deceptive information, generating various uncertainties.

In AI, a series of belief or evidence theories have a long history of studying reasoning and decision-making under uncertainty. In particular, *Subjective Logic* (SL) discusses the estimates of different uncertainty types. SL offers the capability

of measuring uncertainty caused by a lack of information or knowledge (i.e., vacuity), failure to distinguish an observation (i.e., vagueness), and conflicting evidence (i.e., dissonance [1]). In addition, ML/DL algorithms have considered two common uncertainty types, *aleatoric uncertainty* and *epistemic uncertainty*, for decision-making [2]. However, there has been no consensus on multidimensional uncertainty because of different and limited understanding of uncertainty and its attributes.

Via conducting an in-depth survey on a series of belief models, our work aims to introduce a new, promising research direction that leverages uncertainty reasoning and quantification in belief/evidence theory to develop uncertainty-aware ML/DL solutions for effective decision-making. In particular, it is critical to quantify diverse types of uncertainty caused by different root causes, which may provide other courses of action for a decision-maker. For example, recent studies combined belief models (e.g., Subjective Belief Models [3, 4, 5], Fuzzy Logic [6]) with DL models to quantify different uncertainty types about the predictions of a DL model, such as vacuity,

vagueness, and dissonance [7]. This approach will help provide solutions for ML/DL that can meet explainable AI (XAI) by providing how uncertainty derives from, the reason behind it, and ultimately how it impacts a decision's effectiveness. The state-of-the-art decision-making research has fully recognized the significance of uncertainty in the decision-making process. However, few works have extensively examined the general trends of how to apply existing belief models to ML/DL techniques for solving complex decision-making problems in diverse application domains.

1.2. Other Existing Survey Papers

This section describes existing survey papers on uncertainty research and identifies the key differences between those papers and our survey paper.

Li et al. [8] discussed the causes of different uncertainties and how to process them in belief models for making effective decisions in various domains. According to the nature of aleatory and epistemic uncertainty, they classified uncertainty types processing in probability theory, fuzzy theory, information-gap theory, and derived uncertainty theory for their comparison. They focused on how different uncertainties can be processed in data management techniques. Kabir et al. [9] surveyed prediction interval techniques using deep neural networks (DNNs). The prediction interval techniques quantify the level of uncertainty or randomness and have been widely applied in the medical and electricity fields. They discussed aleatoric and epistemic uncertainty (see Section 2 for their definitions) to explain uncertainty in prediction using DNNs. They also discussed how a Bayesian method is used to optimize the weight of an NN during training and applied for NN-based prediction intervals in various fields. Hariri et al. [10] surveyed various AI techniques, including ML, Natural Language Processing (NLP), and computational intelligence, that can recognize and reduce uncertainty in Big Data. Abdar et al. [11] reviewed over 700 papers studying uncertainty quantification in ML/DL. They mainly focused on discussing Bayesian and ensemble techniques and their applications in image processing, computer vision, medical applications, NLP, and text mining.

We also discuss the contributions of the survey papers focusing on the uncertainty mainly in ML/DL in the following papers. Hüllermeier and Waegeman [2] distinguished aleatoric uncertainty from epistemic uncertainty. They explained how these two uncertainties are represented in various ML problems or models and can contribute to decision-making under the assessed uncertainty. Ulmer [12] surveyed the methods of quantifying uncertainty in the 'evidential deep learning model based on the conjugate prior and posterior distributions and unknown outlier samples. This model estimates uncertainty from the Dirichlet distribution by data (aleatoric) uncertainty, model (epistemic) uncertainty, and distributional uncertainty. Gawlikowski et al. [13] provided a comprehensive survey on the uncertainty in DNNs. They discussed two types of uncertainties: *reducible uncertainty* and *unreducible uncertainty*. The concept of reducible uncertainty is aligned with that of epistemic uncertainty, where the reducible uncertainty can be introduced by variability in the real world, errors in model structure, or in

training parameters (i.e., batch size, optimizer). Unreducible uncertainty means uncertainty by noises in measurement (i.e., sensor noise) and is in line with aleatoric uncertainty. The authors classified uncertainty estimation methods based on the cross-combination of the nature (i.e., deterministic or stochastic) and number (i.e., single or multiple) of DNNs. Since the discussions of uncertainty in [2, 12, 13] are very limited in scope, we did not include them in Table 1.

Some studies also investigated uncertainty measures based on entropy in a statistical model and deep learning. Kopetzki et al. [14] investigated whether Dirichlet-based uncertainty (DBU) models are robust under adversarial attacks. They found that uncertainty estimates of DBU models are not robust in detecting classified samples, adversarial examples, and detecting in-distribution and out-of-distribution data. However, this work only used a measure of uncertainty based on differential entropy, not other uncertainty measures studied in belief/evidence theories. Bengs et al. [15] discussed that evidential deep learning and prior networks do not precisely capture evidence adequately because epistemic uncertainty does not converge to zero as the training dataset size goes to infinity. This phenomenon is because a loss function is based upon minimizing first-order prediction error, associated with the probability's calibration but not the probability's variance.

Unlike the existing survey papers above [8, 9, 10, 11], our paper provides an in-depth survey of eight different belief models emphasizing how to reason and quantify uncertainty based on the root causes and types of the uncertainty. In addition, we discussed how a belief model is applicable in the DL domain. This will allow researchers to leverage both the solid methodologies of uncertainty reasoning/quantification in belief models and DL techniques for attaining effective decision-making. Finally, in Table 1, we summarize the key differences between our survey paper and the existing four survey paper on uncertainty research. We selected the key criteria based on the common discussion points covered by the existing survey papers considered in this paper as well as the key discussion points made in our survey paper.

1.3. Research Questions

This study will answer the following research questions:

- RQ1.** *What are the key causes and types of uncertainty studied in belief theory and deep learning?*
- RQ2.** *How can the taxonomy of uncertainty be defined based on the multidimensional aspects of uncertainty studied in belief models and deep learning?*
- RQ3.** *How has each belief model considered and measured uncertainty?*
- RQ4.** *How has each belief model been applied in deep learning and vice-versa for effective decision-making under uncertainty?*
- RQ5.** *What are the key differences between belief theory and deep learning about uncertainty reasoning and quantification?*
- RQ6.** *How can belief model(s) be applied in deep learning to solve complicated decision-making problems?*

The research questions above will be answered in Section 5.

Table 1: COMPARISON OF OUR SURVEY PAPER WITH THE EXISTING SURVEYS ON UNCERTAINTY RESEARCH

Key Criteria	Our Survey (2023)	Li et al. [8] (2012)	Kabir et al. [9] (2018)	Hariri et al. [10] (2019)	Abdar et al. [11] (2021)
Taxonomies of uncertainty	✓	▲	▲	▲	▲
Causes of uncertainty	✓	✓	✓	▲	✓
Uncertainty reasoning & quantification in DST	✓	✓	✗	✗	✗
Uncertainty reasoning & quantification in TBM	✓	✗	✗	✗	✗
Uncertainty reasoning & quantification in DSMT	✓	✗	✗	✗	✗
Uncertainty reasoning & quantification in IDM	✓	▲	✗	✗	✗
Uncertainty reasoning & quantification in TVL	✓	▲	✗	✗	✗
Uncertainty reasoning & quantification in Fuzzy Logic	✓	✓	✗	▲	✗
Uncertainty reasoning & quantification in Bayesian Inference	✓	✓	✗	▲	▲
Uncertainty reasoning & quantification in Subjective Logic	✓	✗	✗	▲	✗
Uncertainty reasoning & quantification in Bayesian Deep Learning	▲	✗	▲	✗	✓
Applicability of Belief Models in Deep Learning	✓	✗	▲	✗	✗
Insights, lessons, and limitations of the existing uncertainty-aware approaches	✓	▲	▲	▲	✓
Future research directions	✓	✓	✓	✓	✓

✓: Fully addressed; ▲: Partially addressed; ✗: Not addressed at all; DST: Dempster-Shafer Theory; TBM: Transferable Belief Model, DSMT: Dezert-Smarandache Theory; TVL: Three-Valued Logic.

1.4. Scope & Key Contributions

The *decision-making* (DM) term embraces a broad meaning with many implications. Our work mainly suggests a promising research direction for uncertainty reasoning and quantification methods explored in various belief/evidence theories to be applicable in machine/deep learning applications. Hence, we use DM in common ML/DL applications, such as inference, classification prediction, action selection, or object detection.

Different domains and disciplines have diverse understandings and interpretations of uncertainty and differently justify how to deal with it. For example, probabilistic theory considers uncertainty as an aleatory uncertainty where uncertainty comes from randomness [16, 17, 18]. Game/decision theory considers uncertainty an aleatory uncertainty where each player can take an action probabilistically [19, 20]. ML/DL researchers have investigated uncertainty in terms of two main uncertainty types, aleatory and epistemic uncertainty, and used entropy as a common measure of uncertainty [21, 2, 22, 23]. On the other hand, belief/evidence theorists have reasoned and quantified uncertainty in terms of vacuity due to a lack of evidence, fuzziness due to inconclusiveness, and ambiguity due to failing to distinguish a single observation or interpretation [24, 25, 26]. However, there is no prior work that can provide a comprehensive viewpoint about uncertainty, its diverse root causes, its reasoning and quantification, and its application to machine/deep learning. To fill this gap, our work conducts an extensive survey on belief theories dealing with uncertainty reasoning and quantification and discusses example studies applying belief models to deep learning, which is one of the emerging research areas in decision-making research. In addition, we outlook the promising directions of taking hybrid approaches that bridge belief models and machine/deep learning for effective decision-making under uncertainty.

This paper makes the following **key contributions**:

1. We conduct an extensive survey on identifying the causes and types of uncertainty studied in various belief models and deep learning using the ontology of uncertainty.
2. We investigate how various belief theories reasoned and quantified uncertainty for effective decision-making.

3. We discuss how belief theories can be effectively leveraged for DL-based solutions for decision-making.
4. We identify the key commonalities and differences about how each belief theory reasons and quantifies uncertainty and how it is applied in DL or along with it.
5. We provide the overall perspectives of insights and lessons learned and the limitations from our extensive survey and suggest promising future research directions.

Our paper mainly discusses quantitative measurements of uncertainty based on observations. Measuring qualitative uncertainty involved with human cognitive aspects is beyond the scope of this research.

The rest of the paper is organized as follows. Section 2 provides various classification types of uncertainty, the causes of different types of uncertainties, and **proposes an** uncertainty ontology based on the key concepts of uncertainty. Section 3 details eight belief models and discusses belief formation, causes and types of uncertainty, uncertainty quantification, and its application. Section 4 discusses how a belief theory can be applied in the context of DL as a decision-making application under uncertainty. Section 6 concludes our paper by discussing the limitations, insights, and lessons learned from our survey and future research directions.

Caveat of Mathematical Notations: In Sections 3 and 4, we discuss a set of belief theories and deep learning theories leveraging belief models, including Subjective Logic, Fuzzy theory, and rough set theory. The discussion of a theory needs to use mathematical notations, which are only used under the theory, not other theories. We keep the mathematical notations of original papers in order to deliver the common notations used to discuss each theory in the literature. The summary of all notations is given in Table A.1 of Appendix A.

2. Types, Causes, and Ontology of Uncertainty

In this work, we deal with uncertainty in data or information. We define an *uncertainty type* as a perceived state of data or information, such as fuzziness, discord, non-specificity, ambiguity, and so forth (see Section 2.1). We define the *causes of uncertainty* by the reason introducing uncertainty in a decision

maker's judgment. We also discuss the ontology of uncertainty, where ontology is studied as a branch of philosophy. Ontology means the "science of what it is" describing the structures of objects, properties, events, processes, and relations in every area of reality [27]. This section will describe the ontology of uncertainty in its types, causes, and outcomes of decision-making based on uncertainty reasoning and quantification.

2.1. Classification of Uncertainty Types

In the probabilistic uncertainty research community [24, 18], two types of *uncertainty natures* are widely discussed:

- *Aleatoric uncertainty*: This refers to statistical uncertainty about the long-term relative frequencies of possible outcomes [24]. For example, if we do not know whether a dice is loaded – and thereby unfair – then we are faced with aleatoric uncertainty. This uncertainty can be reduced to the true variance about the loaded dice by throwing the dice sufficiently many times. However, every time a dice is thrown, we cannot predict its outcome exactly but can have only a long-term probability [24]. In this sense, the long-term probability can reduce epistemic uncertainty through more and more observations. Therefore, aleatoric uncertainty is fundamentally related to the nature of randomness in which a variable is governed by a frequentist process [18, 24].
- *Epistemic uncertainty*: This uncertainty is related to a situation in which we cannot predict an event exactly because of a lack of knowledge. A typical example is the assassination of President Kennedy in 1963 [24], where the uncertainty is about whether he was killed by Lee Harvey Oswald and who organized it. The nature of epistemic uncertainty derives from a lack of knowledge or information (or data). This type of uncertainty is also called *systematic uncertainty* or *model uncertainty*. This means that the outcome of a specific future or past event can be known, but there is insufficient evidence to support it. This uncertainty can be reduced by more evidence, advanced technology, and/or scientific principles to interpret the evidence (e.g., forensic science) [18]. This follows a non-frequentist process representing the likelihood of an event [24].
- *Fuzziness*: This refers to a lack of sharp distinctions and has pertinent synonyms, such as vagueness, cloudiness, haziness, unclearness, indistinctness, or sharplessness.
- *Ambiguity*: In general, there exists *ambiguity* when an object cannot be specified as a single class due to conflicting evidence. Hence, these two cases can be further classified into the following two subclasses:
 - *Discord*: This is associated with disagreement among several alternatives and interchangeably used with synonyms, including dissonance, incongruity, or discrepancy.
 - *Nonspecificity*: This refers to a situation in which two or more alternatives are left unspecified. It is also called variety, generality, diversity, equivocation, and imprecision.

Based on our understanding, fuzziness introduces vagueness (i.e., failing to distinguish one from another), while ambiguity belongs to epistemic uncertainty. In addition to the above, the most common uncertainty is also derived from a lack of evidence, called *vacuity* (or ignorance) [24] as we do not know how to make a decision because of insufficient evidence, which belongs to epistemic uncertainty as more evidence can reduce vacuity. In the modeling and risk assessment research [29, 30], uncertainty associated with choices made by modelers has been studied, such as differences in problem formulation, model implementation, and parameter selection originating from subjective interpretation of the problem at hand. We call this 'modeler uncertainty,' categorized by the following three types:

Since the above two natures of uncertainty have been most widely discussed as the nature of uncertainty, we will discuss how a different type of uncertainty in different belief models and DL models is related to these two natures of uncertainty.

Uncertainty reasoning and quantification research have been heavily explored by several theories, such as probability theory, fuzzy sets theory, possibility theory, evidence theory, and rough sets theory. These theories can be seen as complementary as each of them is designed to deal with different types of uncertainty. In these theories, Dubois [28] identified three uncertainty types in terms of *fuzziness*, *discord*, and *nonspecificity*. The latter two terms, discord, and non-specificity, are combined as the term *ambiguity*. Each type is represented by a brief common-sense characterization and several pertinent synonyms as follows:

- *Parameter uncertainty*: This refers to uncertainty derived from the values of input parameters in a model, such as measurement errors, sampling errors, variability, and use of surrogate data. Hence, it is a type of epistemic uncertainty and can be reduced by collecting more reliable evidence to more accurately estimate the parameters used in the model.
- *Model uncertainty*: This indicates uncertainty about a model structure and the mathematical relationships of components defined in the model. For example, uncertainty can be introduced by making assumptions and simplifying mathematical equations in modeling real-world problems. This uncertainty is introduced by missing or incomplete information, which makes it hard to fully define the model. This belongs to epistemic uncertainty and can be reduced by gathering necessary, reliable information to accurately define the model.
- *Scenario uncertainty*: This represents uncertainty caused by normative choices made in constructing scenarios, including the choice of a functional unit, time horizon, geographical scale, and other methodological choices. This uncertainty arises from uncertain problem formulations and theoretic assumptions, which are not statistical in nature. Due to this nature of uncertainty, we understand this uncertainty as epistemic uncertainty. This uncertainty can be reduced by collecting more evidence, using more advanced technology, and/or considering scientific principles to interpret the evidence, such as finding a better choice of a time horizon or a geographical scale for crime hotspot detection.

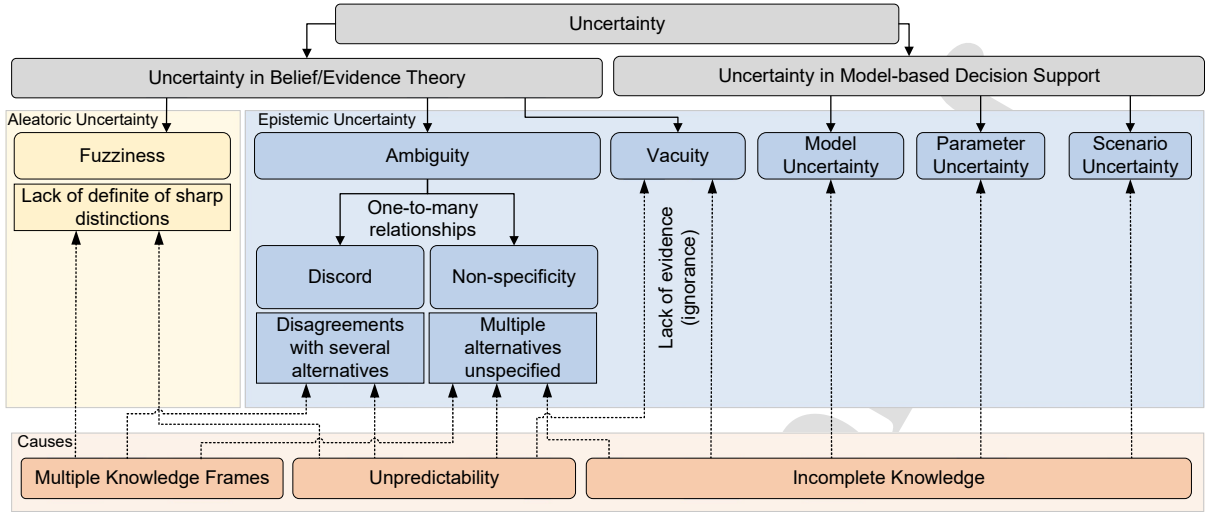


Figure 1: Classification of uncertainty types.

2.2. Causes of Uncertainty

From the perspective of framing research in decision-making, where frames are heuristic representations of the external world, there are three main causes of uncertainty: unpredictability, incomplete knowledge, and multiple knowledge frames [31]). Our work chooses the classification of the causes of uncertainty in [31] based on the nature and knowledge relationships producing uncertainty. From an engineering perspective, these three causes can be introduced by a lack of evidence, limited cognition to process a large amount of evidence, conflicting evidence, ambiguity, measurement errors, and subjective beliefs [32]. The three causes are described:

- **Unpredictability:** A system (or entity/data) may exhibit chaotic, variable behavior in space/time. In Statistics, confidence intervals have been used to measure uncertainty [32]. Statistical noise is a common factor triggering uncertainty, leading to unpredictability. The system exhibits highly variable behaviors even if it learns and adapts to dynamic, new conditions. The variability may be due to unreliability in information, data, or an entity caused by system/network dynamics, non-stationary environmental conditions, or adversarial attacks. If this is the case, this type of uncertainty can be reduced by detecting and excluding unreliable sources or data in the decision-making process [32]. Unpredictability is caused by multiple sources, such as unpredictable model error, unpredictable outcome due to lack of knowledge or evidence, or long-term stochastic variability [33]. Therefore, unpredictability is related to epistemic uncertainty, such as ambiguity (including discord and nonspecificity) and vacuity. More observations and evidence can reduce epistemic uncertainty. However, some extent of unpredictability still will remain due to the variability of an outcome, which is related to fuzziness in aleatory uncertainty [34].
- **Incomplete Knowledge:** This refers to situations where we do not know enough about a system to be managed or our

knowledge about the system is incomplete (i.e., epistemic uncertainty) [32]. This can be due to a lack of evidence (e.g., information/data) or knowledge, called vacuity, because we may not have sufficient theoretical understanding (e.g., ignorance) or reliable information or data. From the modeling perspective, the lack of evidence and knowledge will lead to ‘modeler uncertainty,’ such as parameter, model, and scenario uncertainties. This uncertainty can be reduced by considering more evidence or discarding unreliable evidence. In addition, when human decision-makers receive a large amount of information, which is often highly complex, they cannot process them properly because of their limited cognition and processing power. This may lead to non-specificity: the scenario where two or more alternatives can not be specified due to the lack of knowledge or the complexity of the alternatives. To deal with this, people usually transform available data into information with a rougher ‘granularity’ or focus on important features, neglecting other less important (or noisy) information or data. This uncertainty can be reduced by considering relevant information among the available information [32].

- **Multiple Knowledge Frames:** This refers to when the same information (e.g., evidence or opinions) is interpreted differently, resulting in different, conflicting views. Dewulf et al. [35] defined *ambiguity* as the presence of multiple, valid beliefs about a certain phenomenon. It is related to fuzziness because the information may lack definite or sharp distinctions in its states. The ways of understanding the system (or the external world) can differ in where to put the system’s boundaries or what and who to put as the focus of attention. The differences can also emerge from the way in which the information about the system is interpreted. This cause can also be interpreted as *conflicting evidence*, representing a situation where some of the information available may be incorrect, simply irrelevant, or the model to observe a system may not be correct at a given time. Further, multi-

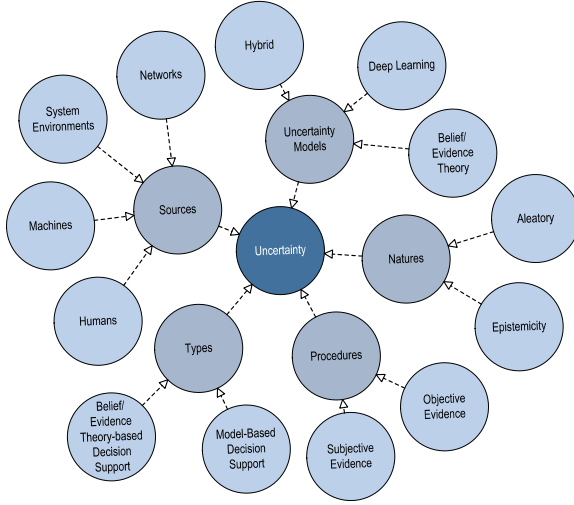


Figure 2: A taxonomy of uncertainty in data (or information).

ple observers may provide different opinions based on their subjective views. The related uncertainty types from this perspective are discord and non-specificity, as discussed in Section 2.1.

We demonstrated our view about the classifications of uncertainty types based on the existing classifications in Figure 1.

2.3. Taxonomies of Uncertainty

Rocha Souza et al. [36] discussed the taxonomy of spatiotemporal uncertainties by categorizing uncertainty in terms of intrinsic and extrinsic uncertainties. Smithson [37] also addressed the taxonomy of uncertainty and ignorance and interpreted ‘uncertainty’ as a specific type of ‘incompleteness.’ Lovell [38] also provided a taxonomy of uncertainty based on different sources of uncertainty, such as individual difference factors and information flow. Rowe [39] classified types of uncertainty in terms of uncertainty and variability of information. For the convenience of demonstrating their taxonomies, we summarized them in Table B.1.

To more extensively understand the concepts of uncertainty and its multiple causes, we develop the taxonomy of uncertainty in Figure 2 as an example that can be used as the basis to provide the taxonomy of uncertainty in data (or information). We placed the extended version of the taxonomies with subattributes of each factor in Figure 2 of Appendix B.

The sources of uncertainty can be as follows:

- **Humans:** Human beings make mistakes in daily life, consciously or unconsciously, leading to uncertainty. For example, humans make errors in perception, imperfect cognition, irrational propensity because of lacking the cognitive ability to judge fairly and accurately, bias, prejudice, delusional beliefs, or distorted memories. They will all impact decision-making. Even some humans have malicious intent, which adds uncertainty to an entity (e.g., other humans or systems) to disrupt its decision-making process.

- **System Environments:** System environments can be very complicated. A system is in a dynamic environment whose internal and external factors continuously generate uncertainty within the environment [40]. For instance, multiple internal factors (e.g., system dynamics, system structures - centralized, decentralized, distributed structures, high-tempo operations, etc.) often increase uncertainty because the dynamic environment changes its internal settings to support continuous communication with the environments. At the same time, uncertainty can also be generated by external factors, such as regulation changes, operating function transformations, and continued technological revolutions to run the system. All those aforementioned reasons can lead to generating uncertainty [40].
- **Machines:** Systems run on machines. Its certainty relies on machines as well. If machines should have non-functional hardware or software, or compromised software, an uncertainty level can increase.
- **Networks:** Networks, as the carrier of information, will create uncertainty. Congestion is normal in networks, causing transfer delays, and networks are vulnerable to cyber attacks or sometimes result in unavailability. All those conditions generate uncertainty in the systems that use networks.
- **Procedures:** Uncertainty is derived from two kinds of procedures. One is based on objective evidence, which should be able to be implemented by the formal and repeatable procedure. The other is based on subjective evidence, which includes beliefs, judgments, and opinions that are mostly generated by humans who use bias or subjective feeling in their decision-making process.

3. Uncertainty Reasoning and Quantification and Its Use in Decision-Making in Belief Theory

Different types of uncertainty affect the assessment and analysis of a specific situation. Underlying uncertainty comes from how to view and model a given part of the world which we call a *domain*. A domain is the abstract representation of states of the world, where analysts or decision-makers can have beliefs about the true states of a domain. Beliefs about domains can be easily biased by an analyst or a decision maker, which is often called the “framing effect” [41, 30], which can cause subjective beliefs about the world to deviate from the ground truth of the world (e.g., past or future events) [30]. The way a situation is formally modeled (i.e., elements in a domain) can also affect the types and levels of uncertainty perceived by a decision-maker.

Figure 1 is developed based on well-known classifications of uncertainty types in the belief/evidence theory [42] and ML/DL domains [5]. *Belief* has been based on the decision-making process. In the 1930s, Three-Valued Logic (TVL) [43] was defined as an algebra based on three values, including false, unknown, and truth. Many other theories expressing a belief based on probabilities have been proposed since the 1960s. They include Fuzzy Logic [44], Dempster-Shafer Theory (DST) [45], Transferable Belief Model (TBM) [46], Subjective Logic (SL) [47, 48], Dezert-Smarandache Theory (DSmT) [49], Bayesian Inference (BI) [50], and Imprecise Dirichlet Model (IDM) [51]. As

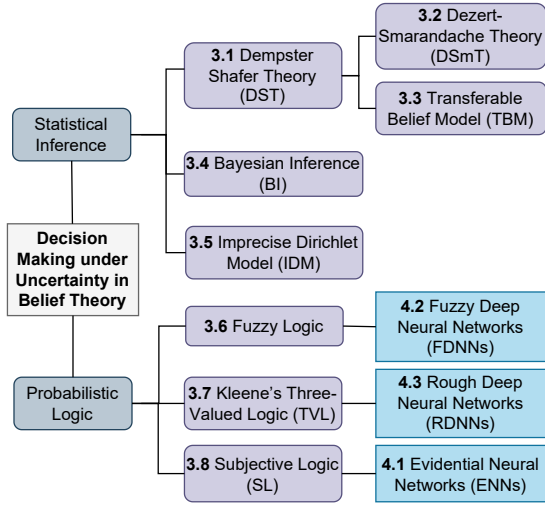


Figure 3: Taxonomies of the considered belief theories and their applications in deep learning.

shown in Figure 3, belief theories have been discussed in terms of *statistical inference* and *probabilistic logic*. This section will first discuss DST, which is most prevalent in many applications. DST [45] first defined a “frame of discernment,” the set of propositions considered. DST generalized Bayesian theory based on subjective probability [45]. However, it is well known that DST’s combination rule generates a counter-intuitive result when evidence-forming beliefs conflict with each. To resolve the issue in DST, TBM [46] has been proposed to deal with more knowledge and situations than DST. Further, DSmT [49] extended DST to deal with conflicting evidence in trust management systems [52, 53]. However, applying DST, DSmT, and TBM in ML/DL is limited. To our knowledge, TBM has not been considered in the ML/DL domain. BI was studied in the 18th century and used even in various modern statistical inference and machine learning applications [50].

We discuss Fuzzy Logic, TVL, and SL as a probabilistic logic theory. In the 1960s, Zadeh [44] introduced *fuzzy set theory* to represent an uncertain, subjective belief based on a membership function [54] and has been applied to various trust-based systems [55, 56, 57, 58, 59, 60, 61]. As TVL was proposed in the 1930s, its application was limited to the 1970s and 1980s classification research. SL is initially proposed in 1997 [62] and has been further explored until recent years [24]. SL has its root in DST and can estimate multidimensional uncertainty. SL has been recently leveraged to develop evidence neural networks (ENNs) [5, 63, 64, 65, 66, 67], discussed in Section 4. We discuss how each belief theory has been considered in the ML/DL research in Section 3.

3.1. Dempster Shafer Theory (DST)

In [68, 69], the authors presented the key concepts and notable results of DST used to develop a crucial tool for knowledge representation and uncertain reasoning. The authors reviewed how DST has evolved in terms of transitioning into a

computational tool for evidential reasoning in AI by discussing its various practical applications in business, engineering, and medicine. Kohlas and Monney [69] also discussed DST by emphasizing the process of combining multiple hints relative to a common question, and they drew a comprehensive relationship between these hints and their corresponding Dempster’s rule.

DST is a fusion technique for decision-making based on the *belief mass* (a.k.a. evidence) of various detection systems. Each system can be defined as a set of possible conclusions, called *proposition* [45]. The set of all propositions is denoted by Θ (a.k.a. the *frame of discernment (FOD)*). Given the set Θ , we can generate the power set $P(\Theta)$ (a.k.a. the *powerset of FOD*), where the $P(\Theta)$ represents all possible combinations of the set Θ , including an empty set $\emptyset$. So, $2^{|\Theta|}$ is the size of the $P(\Theta)$. As an example of $P(\Theta)$, if $\Theta = \{W, Z, L\}$, then $P(\Theta) = \{\emptyset, \{W\}, \{Z\}, \{L\}, \{W, Z\}, \{W, L\}, \{Z, L\}, \{W, Z, L\}\}$.

3.1.1. Belief Formation

The belief mass is an observed probability based on evidence. For example, assume we have a black ball, a black square, and a red ball. The mass for focal element *black* is $m(\text{black}) = \frac{2}{3}$, and the mass for focal element *red* is $m(\text{red}) = \frac{1}{3}$. We assign a belief mass to each element in power set $P(\Theta)$ and define the mass function as $m : P(\Theta) \rightarrow [0, 1]$. The belief mass, also called basic probability assignment (*bpa*) [70], is an observed probability based on evidence. It defines a mapping of the power set to the interval between 0 and 1, where the *bpa* of the null set is zero. The sum of the mass for each element in set $P(\Theta)$ is equal to one, that is [45]:

$$m : \Theta \rightarrow [0, 1], \sum_{A \in P(\Theta)} m(A) = 1, \text{ where } m(\emptyset) = 0. \quad (1)$$

Given the power set $P(\Theta)$ and the corresponding belief mass m for each focal element (i.e., a subset A in $P(\Theta)$), we can calculate the belief interval of each focal element A , and represent it as $[Bel(A), pl(A)]$. The belief $Bel(A)$ is the lower bound and plausibility $pl(A)$ indicates the upper bound [71]. The $Bel(A)$ and $pl(A)$ are obtained by:

$$Bel(A) = \sum_{B|B \subseteq A} m(B), \quad pl(A) = \sum_{B|B \cap A \neq \emptyset} m(B), \quad (2)$$

$$Dis(A) = 1 - pl(A).$$

For example, given $\Theta = \{W, Z, L\}$ and the belief mass $m(W)$, $m(Z)$, $m(W \text{ or } L)$, we can obtain beliefs of focal element W and $(W \text{ and } Z)$ by $Bel(W) = m(W)$ and $Bel(W \text{ and } Z) = m(W) \cdot m(Z)$, respectively, and plausibility of W by $pl(W) = m(W) + m(W \text{ or } L)$. The belief interval for W is denoted by $[m(W), m(W) + m(W \text{ or } L)]$.

The disbelief of focal element A is given as $Dis(A)$, which equals $Bel(\bar{A})$, where $\bar{A}$ means the complement of A (i.e., negation of A). The $Dis(A)$ is calculated by summing all masses of the focal elements that do not intersect with A . $Dis(A)$ can also be estimated by $Dis(A) = 1 - pl(A)$ where $pl(A)$, indicating uncertainty, can be considered as a potential credit to increase the given belief, which is different from the conventional interpretation of uncertainty as risk [72]. Figure 5 describes the key concept of DST.

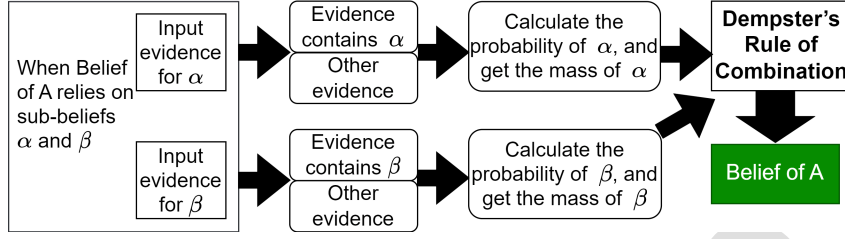


Figure 4: Dempster's Rule of Combination.

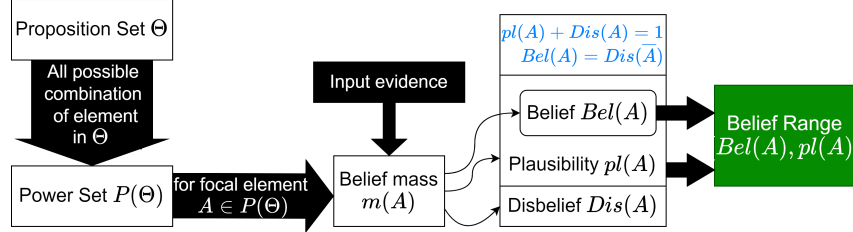


Figure 5: Belief, Plausibility, and Disbelief in DST.

Dempster's Rule of Combination is a belief mass combination function for two independent detection systems i and j over the same frame. The joint mass committed to the focal element A is given by:

$$m(A) = \kappa \sum_{A_i \cap B_j = A \neq \emptyset} m_1(A_i) \cdot m_2(B_j), \quad (3)$$

where A_i and B_j are values in set Θ of two different systems i and j that contain target value A . The κ is a renormalization constant, defined by $\kappa = (1 - \sum_{A_i \cap B_j = \emptyset} m_1(A_i)m_2(B_j))^{-1}$ [45]. For example, we have $m_1(W)$, $m_1(Z)$, $m_1(W, Z)$, $m_2(W)$, $m_2(Z)$, and $m_2(W \text{ or } Z)$. The joint mass for focal element W is calculated by $m(W) = m_1(W) \cdot m_2(W) + m_1(W) \cdot m_2(W \text{ or } Z) + m_1(W \text{ or } Z) \cdot m_2(W) + m_1(W \text{ or } Z) \cdot m_2(W \text{ or } Z)$. We summarize the key concept of Dempster's rule of combination in Figure 4 based on our discussion above.

Many DST variants have been proposed. Those variant techniques usually aim to solve the problem of the original DST. For example, Dempster's rule requires a normalization operation to move the zero belief into a null set. However, Zadeh discovered that such normalization operation could lead to a counterintuitive result/belief in some situations [73, 74, 75]. To eliminate this problem, Yager [74] proposed a DST-concept-based technique for generating beliefs from evidence combining in the Dempster Shafer framework. Murphy [76] also solved an issue in Dempster's combination rule, introducing greater uncertainty with greater conflict evidence by refining the normalization part with the average belief. This modification removes the mass assigned to null set $\emptyset$, and also combines weighted averages of the masses $n-1$ times in Eq. (3), where n is the number of values in set Θ . In addition, the transferable belief model (TBM) is one of the well-known variants of DST, which is described in detail in Section 3.3.

3.1.2. Causes and Types of Uncertainty

DST considers uncertainty in *plausibility* due to a lack of evidence. This implies that uncertainty in DST is closely related to epistemic uncertainty or vacuity. Hence, DST can quantify an uncertain opinion as a subjective belief in a given proposition [45].

3.1.3. Uncertainty Quantification

Smarandache et al. [71] measured uncertainty in DST based on its multiple dimensions, including *auto-conflict* (i.e., conflict in a belief function with conjunctive rule), *non-specificity* (i.e., a generic form of Hartley entropy with base 2), *confusion* (i.e., uncertainty by a lack of evidence), *dissonance* (i.e., all beliefs are mostly the same), *aggregate uncertainty measure* (AU) (i.e., generalized Shannon entropy), and *ambiguity measure* (AM) (i.e., non-specificity and discord).

Smarandache et al. [71] measured uncertainty in DST by its multiple dimensions as follows:

1. *Auto-conflict* quantifies how much a source of evidence disagrees with itself when it is combined with itself repeatedly using the conjunctive combination rule [77]. The auto-conflict of order n (a_n with $n \geq 1$) fuses n identical belief functions and is estimated by:

$$a_n(e) = \left(\bigoplus_{i=1}^n \right)(\emptyset), \quad (4)$$

where $\oplus$, conjunctive combination rule, is given by:

$$m_{Conj}(X) = \sum_{A \cap B = X} m_1(A_i)m_2(B_j) := (m_1) \oplus (m_2)(X). \quad (5)$$

$m_{Conj}(X)$ is a non-normalized conjunctive rule, which is a multinomial form of the DST combination rule shown in Eq. (3). The 'semantics of conflict' (k) quantifies the degree to which different sources of evidence disagree, while

the ‘auto-conflict’ quantifies the intrinsic conflict within a single source of evidence when combined with itself.

2. *Non-specificity* is a generalization of Hartley entropy, which is aligned with Shannon entropy when base two is used in a logarithmic term. Non-specificity increases when there are many subsets $A \subseteq \Theta$ because it implies more subsets of Θ overlap in their elements. This non-specificity can be treated as a weighted sum of the Hartley measure [78] for different focal elements A in $P(\Theta)$ and is measured by:

$$N(m) = \sum_{A \subseteq \Theta} m(A) \log_2 |A|. \quad (6)$$

3. *Confusion* [79] refers to uncertainty caused by a lack of evidence supporting each set $A \subseteq \Theta$ and estimated by:

$$\text{Confusion}(m) = - \sum_{A \subseteq \Theta} m(A) \log_2 (\text{Bel}(A)). \quad (7)$$

As shown above, confusion increases when a large number of subsets $A \subseteq \Theta$ exists, and each belief in set A is very small.

4. *Dissonance* [79] is similar to confusion but differs in using plausibility rather than belief. That is, dissonance increases when there are a large number of subsets $A \subseteq \Theta$, and its associated plausibility is very low. This implies that the difference between the lower bound (i.e., belief) and upper bound (i.e., plausibility) is small. Hence, the dissonance is more likely to be higher when there is a sufficient amount of evidence (i.e., uncertainty due to a lack of evidence is low) supporting all possible beliefs (i.e., subsets) generating each $pl(A)$ with very small probabilities. The dissonance is measured by:

$$\text{Dissonance}(m) = - \sum_{A \subseteq \Theta} m(A) \log_2 (pl(A)). \quad (8)$$

Note that in [68], natural logarithmic (i.e., $\ln$) is used rather than $\log_2$.

5. *Aggregate uncertainty measure* (AU) is a generalized *Shannon entropy* for total uncertainty and measured by:

$$\text{AU}(\text{Bel}) = \max \left[- \sum_{\theta \in \Theta} p_\theta \log_2 p_\theta \right]. \quad (9)$$

Recall that $\Theta = \{W, Z, L\}$ and θ refers an element in set Θ .

6. *Ambiguity measure* (AM) [80] represents non-specificity and discord where the *discord* means disagreement in choosing several alternatives and *non-specificity* refers to a situation with two or more alternatives left unspecified. The AM is obtained by:

$$\text{AM}(m) = \sum_{\theta \in \Theta} \text{BetP}_m(\theta) \log_2 (\text{BetP}_m(\theta)), \quad (10)$$

where BetP_m is defined by the Generalized pignistic transformation (GPT), a pignistic¹ generalized basic belief assignment in DSmt, as estimated in Eq. (32).

Blasch et al. [25] defined the *Interval of Uncertainty* (IOU) by:

$$\begin{aligned} \text{IOU}(A) &= pl(A) - \text{Bel}(A) \\ &= 1 - \text{Bel}(\bar{A}) - \text{Bel}(A) = 1 - \text{Dis}(A) - \text{Bel}(A). \end{aligned} \quad (11)$$

Klir and Ramer [82] measured the total uncertainty in DST, denoted by $U^T(A)$, by considering two types of uncertainty, non-specificity, and discord and capture them differently. AM captures them in a level of each proposition (i.e., element $\theta \in \Theta$) while $U^T(A)$ obtains them at the level of sets, $A \subset \Theta$. Hence, $U^T(A)$ is given by [82]:

$$U^T(A) = \sum_{A \subset \Theta} m(A) \log_2 \left(\frac{|A|}{\sum_{B \subset \Theta} m(B) \frac{|A \cap B|}{|B|}} \right). \quad (12)$$

The key merit of DST is to combine an amount of uncertain evidence from multiple sources and select elements based on the combined belief mass. However, the combination rule of DST fails to balance different sources, especially when sources provide conflicting evidence. Although many alternative combination rules have been proposed, Dubois and Prade [26] argued that no single combination rule could be a universal solution to all encountered situations.

3.1.4. Applications of DST on Machine/Deep Learning

Early use of DST in deep learning in [83] was focused on creating an adaptive pattern classifier. This model used basic belief assignments (BBAs) in DST within a multilayer neural network architecture for flexible decision rules and robust sensor fusion. Soua et al. [84] also proposed a framework using a Deep Belief Network (DBN) to independently predict traffic flow based on streams and events. DST was then utilized to fuse these predictions in handling heterogeneous data sources. Since the 2020s, DST has further applied to tackle various tasks, such as classification problems by combining Convolutional Neural Networks (CNNs) with DST [85], intrusion detection by applying DST within a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) [86], and fault diagnosis problems by applying DST in deep neural networks [87].

Besides the applications above, the evidential reasoning (ER) approach, rooted in Dempster-Shafer’s evidence theory, has been widely utilized in various machine/deep learning (ML/DL) domains. For instance, Xu et al. [88] introduced a classification method using the evidence reasoning (ER) rule, which enhances Dempster’s combination rule. This data-driven method has shown competitive classification accuracy on benchmark databases. Similarly, in the realm of recommendation systems, Le et al. [89] proposed a multi-criteria collaborative filtering method for hotel recommendations. This method integrates matrix factorization with a deep learning model and then employs the evidential reasoning approach to handle the uncertainty of multi-criteria ratings. Furthermore, Xu et al. [90] extended the Dempster-Shafer theory to ensemble learning, representing an ML paradigm to train and combine multiple weak models to solve the same problem and achieve

¹The pignistic means a probability assigned to an option by a person who is required to make a rational decision [81].

better solution quality. This method uses the ER rule as a combination strategy. Their approach focuses on determining the weight of evidence through a novel combination weighting method, demonstrating its effectiveness in image classification datasets. These studies emphasize the growing viability and importance of the ER approach in recent ML/DL applications. This evolution signifies the versatility of DST and its potential for further explorations in deep learning.

3.1.5. Algorithmic Complexity of DST

In DST, Dempster's Rule of Combination is directly proportional to the number of hypotheses or subsets in the frame of discernment. Given a frame of discernment with n elements, it has 2^n subsets, resulting in an exponential algorithmic complexity. Thus, the complexity of DST is represented by $O(2^n)$.

3.2. Dezert-Smarandache Theory (DSmT)

Dezert and Smarandache [49] introduced DSmT theory for data and information fusion problems as a general framework that provides new rules for handling highly imprecise, vague, and uncertain sources of evidence and making decisions under them. The main advantages of DSmT over DST are as follows. First, DSmT has a more general fusion space in a hyper-power set (discussed in Section 3.2.1) than a power set. Second, DSmT fits free and a hybrid model compared to a strict DST model (see Section 3.2.1). Third, DSmT also combines complex classes based on subsets or complements and introduces better fusion rules, such as proportional conflict redistribution rule 5 (PCR5), dynamic fusion by hybrid DSm rule (DSmH), a new probability transformation, qualitative operators for data with labels (e.g., linguistic labels in natural language), and new belief conditioning rules (BCRs), or new fusion rules for set-valued imprecise beliefs. However, DSmT has not been largely applied in applications compared to DST. In addition, the sparse distribution and high complexity lower efficiency leading to a lack of scalability. Further, inefficiency lies in requiring a specific combination rule for each problem with no universal rule.

3.2.1. Belief Formation

A *generalized basic belief assignment (gbba)* is formulated the same as a belief function of DST in Eqs. (1) and (3.1.1), but the domain of DSmT is the hyper-power set D^Θ , compared to a power set $P(\Theta)$ of DST. Note that $P(\Theta) \triangleq (\Theta, \cup)$, $D^\Theta \triangleq (\Theta, \cup, \cap)$, and $S^\Theta \triangleq (\Theta, \cup, \cap, c(\cdot))$. If $\Theta = \{a, b\}$, $P(\Theta) \triangleq (\emptyset, a, b, a \cup b)$, $D^\Theta \triangleq (\emptyset, a, b, a \cup b, a \cap b)$, and $S^\Theta = (\emptyset, a, b, a \cup b, a \cap b, c(\emptyset), c(a), c(b), c(a \cup b), c(a \cap b))$ where $c(X)$ refers to the complement of X .

Other various types of belief mass functions introduced in DSmT are formulated via the following concepts:

- A *hyper-power set*: The closed finite set (i.e., frame), denoted by Θ , has n hypotheses elements, represented by $\Theta = \theta_1, \dots, \theta_n$. The hyper-power set $D^\Theta \triangleq (\Theta, \cup, \cap)$ is defined as all composite subsets built from elements of Θ with $\cup$ and $\cap$ operators. In general, notation G^Θ covers the belief functions defined under any frames, such as a power set

$P^\Theta \triangleq (\Theta, \cup)$ for DST, a hyper-power set D^Θ or a super-power set for DSmT [91].

- *Free and hybrid DSm models*: DST is the most restricted hybrid DSm model, denoted by $M^0(\Theta)$, because all the elements are exhaustive and exclusive. However, in real-world fusion problems, the hypothesis can be vague and imprecise. The exclusive elements, θ_i 's, may not be identified and separated. When there is no constraint on the elements (i.e., hypothesis elements θ_i 's can overlap), this model is the free DSm model as $M^f(\Theta)$. Hybrid DSm models, $M(\Theta)$, take into account some exclusivity constraints and non-existential constraints (i.e., changes of frame Θ with time in a dynamic fusion problem). DSmT works with any kind of hybrid models, such as the free DSm model, hybrid model, or Shafer's model (i.e., DST) [91].
- *Generalized belief functions*: The generalized belief and plausibility functions, $Bel(A)$ and $pl(A)$, are defined in the same way DST defines a belief by Eq. (3.1.1).
- *Qualitative belief assignment*: The qualitative beliefs are a set of m linguistic labels $L = \{L_0, L_1, L_2, \dots, L_m, L_{m+1}\}$ with a total order relationship $<$ where $L_1 < L_2 < \dots < L_m$. The example of ordered labels is $L = \{\text{very poor, poor, good, very good}\}$. L_0 and L_{m+1} represent the boundary of numeric values in $[0, 1]$ for the labels. With numeric values, the labels can have qualitative operators of label addition, label multiplication, and label division in [49]. The qualitative belief assignment is a mapping function by [91]:

$$qm(\cdot) : G^\Theta \mapsto L. \quad (13)$$

- *Fusion of precise beliefs by the classic DSm rule of combination*: For the free DSm model, $M^f(\Theta)$, the belief functions combination, $m_{M^f(\Theta)} \equiv m(\cdot) \triangleq [m_1 \oplus m_2](\cdot)$, from the two independent sources is the conjunctive consensus of the sources [49]:

$$\forall C \in D^\Theta, \quad m_{M^f(\Theta)}(C) \equiv m(C) = \sum_{A, B \in D^\Theta, A \cap B = C} m_1(A)m_2(B), \quad (14)$$

where belief functions from two sources $Bel_1(\cdot)$ and $Bel_2(\cdot)$ are related to generalized basic belief assignments, $m_1(\cdot)$ and $m_2(\cdot)$. This function uses the same method as Dempster's rule of combination in Eq. (3), but the fusion space can be expanded to hyper-sets.

- *Fusion of precise beliefs by the hybrid DSm rule of combination*: The hybrid DSm rule (DSmH) works for the set with integrity constraints in a hybrid model, $M(\Theta) \neq M^f(\Theta)$. For all $A \in D^\Theta$ [49]:

$$m_{DSmH}(A) = m_{M(\Theta)}(A) \triangleq \phi(A)[S_1(A) + S_2(A) + S_3(A)], \quad (15)$$

$$S_2(A) = \sum_{\substack{X_1, X_2, \dots, X_k \in \emptyset \\ (U=A) \vee (U \neq \emptyset) \wedge (A=I_i)}} \prod_{i=1}^k m_i(X_i), \quad (16)$$

$$S_3(A) = \sum_{\substack{X_1, X_2, \dots, X_k \in D^\Theta \\ (X_1 \cup X_2 \cup \dots \cup X_k) = A, \\ X_1 \cap X_2 \cap \dots \cap X_k \in \emptyset}} \prod_{i=1}^k m_i(X_i), \quad (17)$$

where $\phi(A)$ indicates the non-emptiness of a set A , i.e., $\phi(A) = 1$ if $A \neq \emptyset$. $S_1(A) = m_{M(\emptyset)}(A)$ is the classic rule based on the free DSm model as shown in Eq. (14). The $S_2(A)$ is the relative or total ignorance with non-existential constraints in some dynamic problems, calculated from the mass of empty sets X_i , where i is one of the k sources of evidence. Total ignorance, I_t , is defined as $I_t \triangleq \theta_1 \cup \theta_2 \cup \dots \cup \theta_n$, meaning the union of n hypotheses elements. The mass of relatively empty sets in $S_2(A)$ is defined as the $X_1, X_2, \dots, X_k \in \emptyset$ where $\mathcal{U} = u(X_1) \cup u(X_2) \cup \dots \cup u(X_k)$ and $u(X)$ is the union of all the granules θ forming X . $S_3(A)$ is the canonical disjunctive form of non-empty sets X_i , calculated from relatively empty sets. The details of $S_2(A)$ and $S_3(A)$ are defined in [49] and their formulations are shown in Eq. (17).

- *Fusion of imprecise beliefs by the classic DSm rule of combination*: For k sources of evidence, the imprecise belief is defined by [49]:

$$\forall A \neq \emptyset \in D^\Theta, m^I(A) = \sum_{\substack{X_1, X_2, \dots, X_k \in D^\Theta \\ (X_1 \cap X_2 \cap \dots \cap X_k) = A}} \prod_{i=1, \dots, k} m_i^I(X_i), \quad (18)$$

where $\sum$ and $\prod$ are for summation and product operation of sets, respectively. The imprecise beliefs are defined by a set of points or continuous intervals in $[0, 1]$. The notations of intervals can either be open, closed, or half-open intervals (i.e., (a, b) , $[a, b]$, or $[a, b)$).

- *Fusion of imprecise beliefs by hybrid DSm rule of combination*: The hybrid model DSm rule from the sets of imprecise beliefs are defined by [49]:

$$m_{DSmH}^I(A) = m_{M(\emptyset)}^I(A) \triangleq \phi(A) \cdot [S_1^I(A) + S_2^I(A) + S_3^I(A)], \quad (19)$$

$$S_2^I(A) = \sum_{\substack{X_1, X_2, \dots, X_k \in \emptyset \\ (\mathcal{U}=A) \vee (\mathcal{U} \neq \emptyset) \wedge (A=I_t)}} \prod_{i=1}^k m_i^I(X_i), \quad (20)$$

$$S_3^I(A) = \sum_{\substack{X_1, X_2, \dots, X_k \in D^\Theta, (X_1 \cup X_2 \cup \dots \cup X_k) = A, \\ X_1 \cap X_2 \cap \dots \cap X_k \in \emptyset}} \prod_{i=1}^k m_i^I(X_i), \quad (21)$$

where the $+$ and $\cdot$ operators are both for two sets. $S_1^I(A) = m^I(A)$ in Eq. (18) and $S_2^I(A)$ and $S_3^I(A)$ in Eq. (21) are described similar to the precise belief DSm rule of combination in [49].

- *Fusion of qualitative precise beliefs*: The qualitative DSm Classic rule (q-DSmC) is similar to Eq. (14) by the qualitative conjunctive rule as below [49]:

$$qm_{qDSmC}(A) = \sum_{\substack{X_1, \dots, X_k \in D^\Theta \\ X_1 \cap \dots \cap X_k = A}} \prod_{i=1}^k qm_i(X_i), \quad \text{for } k \geq 2, \quad (22)$$

where $\forall A \in D^\Theta \setminus \{\emptyset\}$, $qm_{qDSmC}(\emptyset) = L_0$,

where k refers to the number of sources. The qualitative hybrid DSm rule (q-DSmH) is similarly de-

fined to Eq. (15) by [49]:

$$qm_{qDSmH}(A) \triangleq \phi(A)[qS_1(A) + qS_2(A) + qS_3(A)] \quad (23)$$

where $qm_{qDSmH}(\emptyset) = L_0$,

$$qS_2(A) = \sum_{\substack{X_1, X_2, \dots, X_k \in \emptyset \\ (\mathcal{U}=A) \vee (\mathcal{U} \neq \emptyset) \wedge (A=I_t)}} \prod_{i=1}^k qm_i(X_i), \quad (24)$$

$$qS_3(A) = \sum_{\substack{X_1, X_2, \dots, X_k \in D^\Theta, (X_1 \cup X_2 \cup \dots \cup X_k) = A, \\ X_1 \cap X_2 \cap \dots \cap X_k \in \emptyset}} \prod_{i=1}^k qm_i(X_i), \quad (25)$$

where $qS_1(A) = qm_{qDSmC}(A)$ and the details of $qS_2(A)$ and $qS_3(A)$ are defined in Eq. (17).

- *Proportional conflict redistribution (PCR) rules*: Compared to DSmH, which considers a direct transfer of partial conflicts to partial uncertainties, PCR transfers total or partial conflicting belief functions to non-empty sets in the conflicts in proportion to the belief functions by sources. The belief function redistribution generates several versions of PCR rules. We take the following two examples:

- *PCR5* is the most sophisticated PCR rule. The PCR5 formula of combining two sources, considering $m_{PCR5}(\emptyset) = 0$ and $\forall X \in G^\Theta \setminus \{\emptyset\}$, is defined by [49]:

$$m_{PCR5}(X) = m_{12}(X) + \sum_{\substack{Y \in G^\Theta \setminus \{X\} \\ X \cap Y = \emptyset}} \left(\frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)} + \frac{m_2(X)^2 m_1(Y)}{m_2(X) + m_1(Y)} \right), \quad (26)$$

where all sets are in canonical form and $m_{12}(X) \equiv m_{\cap}(X)$ for the conjunctive consensus on X and all denominators are not zero. If a denominator is zero, $m_{12}(X) = 0$. The PCR5 rule also provides good results for combining qualitative results into a refined label.

- *PCR6* is a more intuitive PCR formula than other PCR formulas [49]:

$$m_{PCR6}(X) = m_{12 \dots k}(X) + \sum_{\substack{X_1, X_2, \dots, X_k \in G^\Theta \setminus \{\emptyset\} \\ X_1 \cap X_2 \cap \dots \cap X_k = \emptyset}} \left(\sum_{r=1}^k \delta_{X_r}^X \cdot m_r(X_r) \right) \cdot \frac{m_1(X)^2 m_2(Y)}{m_1(X) + m_2(Y)}, \quad (27)$$

where $\delta_{X_r}^X = 1$ if $X = X_r$; $\delta_{X_r}^X = 0$ otherwise. The $m_{12 \dots s}(X) \equiv m_{\cap}(X)$ is the conjunctive consensus on X for $s > 2$ sources. For $s = 2$ sources, PCR5 and PCR6 have the same formulas.

3.2.2. Causes and Types of Uncertainty

DSmT handles various uncertainties as follows [91]:

- *Precise, uncertain beliefs from multiple sources*: The beliefs from multiple sources contribute uncertainty even if the belief of each proposition is a precise $m(X)$, where each $m(X)$

is only represented by one real number in $[0, 1]$ in D^Θ . Uncertainty exists when a single source provides beliefs about partial elements or multiple sources provide conflicting beliefs. For example, for $\Theta = \{\theta_1, \theta_2, \theta_3\}$, two independent sources provide beliefs $m_1(\theta_1) = 0.6, m_1(\theta_3) = 0.4$ and $m_2(\theta_2) = 0.8, m_2(\theta_3) = 0.2$, respectively.

- *Highly conflicting evidence from multiple sources*: If k multiple sources have conflicting evidence toward the same event, there is uncertainty about which source to trust. For example, for $\Theta = \{\theta_1, \theta_2, \theta_3\}$, two sources provide $m_1(\theta_1) = 0.2, m_1(\theta_2) = 0.1, m_1(\theta_3) = 0.7$ and $m_2(\theta_1) = 0.5, m_2(\theta_2) = 0.4, m_2(\theta_3) = 0.1$. The decision is based on conflicting evidence.
- *Imprecise beliefs*: Imprecise beliefs are represented by the admissible imprecise. Imprecise beliefs can be either quantitative or qualitative. Quantitative imprecise beliefs $m^l(\cdot)$ are real subunitary intervals of $[0, 1]$ or real subunitary sets over D^Θ . Qualitative $m^l(\cdot)$ is a set of labels $L = \{L_0, L_1, L_2, \dots, L_m, L_{m+1}\}$ in order. Imprecise beliefs are common in fusion problems because it is very hard to generate precise sources of evidence. For example, the set of ordered sentiment labels are $L = \{L_0, L_1, L_2\} = \{\text{negative}, \text{neutral}, \text{positive}\}$ and the set of elements is $\Theta = \{\theta_1, \theta_2\}$. The two sources can give qualitative beliefs by sentiment labels as $qm_1(\theta_1) = L_1, qm_1(\theta_2) = L_0$ and $qm_2(\theta_1) = L_2, qm_2(\theta_2) = L_1$, respectively.
- *Subjective probability (DSm probability or DSmP) transformation, fusion space, and fusion rules*: The criteria (i.e., frame Θ), the set of elements (i.e., G^Θ), the choice of combination rule, the probability function, and controllable parameter ϵ for DSmP, all contribute to the uncertainty that can significantly impact decision making.

3.2.3. Uncertainty Quantification

DSmT does not provide its own uncertainty measure. It borrows other methods and helps decision-making using the following uncertainty measures:

- In probability theory, uncertainty in proposition A can be defined as [92]:

$$U(A) = \sum_{B \in S^\Theta \setminus \{\emptyset, B \cap A \neq \emptyset, B \cap C(A) \neq \emptyset\}} m(B), \quad (28)$$

where A, B , and C are three different elements (i.e., propositions), Θ is a set of the elements, and S^Θ is a super power set. The $C(A)$ is the complement of A . Uncertainty and IOU can also be defined in the same way as DST in Eq. (3.1.3).

- Degree of uncertainty can be evaluated in the probability transformation. Normalized Shannon's entropy is a measure of uncertainty in probability theory and is given by:

$$E_H = - \frac{\sum_{i=1}^n m(\theta_i) \log_2(m(\theta_i))}{H_{max}}, \quad (29)$$

where H_{max} is the maximal entropy for the uniform distribution.

- Probabilistic information content (PIC) score refers to the degree of certainty which can be estimated by $PIC = 1 - E_H$. Less uncertainty (or higher certainty) can lead to a correct and reliable decision.

For decision making, DSmT extends the probability function, called *classical pignistic transformation* (CPT) in DST, into either the *Generalized pignistic transformation* (GPT), or a generalization of a subjective probability measure of $m(\cdot)$ by $\epsilon \geq 0$ in the new probability transformation, $DSmP_\epsilon$, with a subjective measure, ϵ .

1. *Generalized pignistic transformation (GPT)*: To make a rational decision, a subjective/pignistic probability function $BetP(\cdot)$ in Eq. (32) is constructed, based on any generalized basic belief assignment $m(\cdot)$ from the DSmT rules of combination by expanding the set to D^Θ and counting the cardinal of propositions in the hyper-power set.
2. *DSmP transformation*: In GPT, the decision to map belief to the probability is in a fixed way but DSmP provides variations in subjective probabilities. DSmP is a new probability transformation that considers both the values and the cardinality of the elements involved in each ignorance in the proportional redistribution process. This mapping is defined by $DSmP_\epsilon(\emptyset) = 0$ and $\forall X \in G^\Theta \setminus \{\emptyset\}$:

$$DSmP_\epsilon(X) = \sum_{Y \in G^\Theta} \frac{\sum_{\substack{Z \subseteq X \cap Y \\ |Z|_M=1}} m(Z) + \epsilon \cdot |X \cap Y|_M}{\sum_{\substack{Z \subseteq Y \\ |Z|_M=1}} m(Z) + \epsilon \cdot |Y|_M} m(Y), \quad (30)$$

where $|\cdot|_M$ is the DSm cardinal of proposition in D^Θ and $\epsilon \geq 0$ is parameter for a subjective probability measure of $m(\cdot)$. The smaller ϵ , the bigger the PIC value. When $\epsilon = 1$, $DSmP_{\epsilon=1} = BetP$ (i.e., Eq. (30) is reduced to Eq. (32)). DSmP is a more accurate distribution of the ignorance masses to the elements than $BetP$. DSmP and $BetP$ work in both DST (Shafer's model) and DSmT (free or hybrid models).

Figure 6 demonstrates the following steps/criteria to make a decision. In Figure 6, (b) and (c) combines the belief masses from different sources. Uncertainty is not combined into input evidence and is used to make final decisions in (c).

1. Belief functions and models are defined in the proper frame Θ of a given problem. The closed finite set (i.e., frame), denoted by Θ , has n elements of hypotheses. These steps decide the elements of the given problem.
2. The belief functions are defined in the proper set of G^Θ (e.g., power, hyper, or super set) where G^Θ means any set of items, including power set P^Θ , hyper-power set D^Θ , and super-power set S^Θ . DSmT works on any G^Θ , but normally D^Θ is used to distinguish from P^Θ in DST. This step means the choice of P^Θ , D^Θ , or S^Θ .
3. Choose an efficient rule to combine belief functions (see Section 3.2.1).
4. Before making a decision, one must use a probability function (e.g., GPT, see Eq. (30)) or DSmP with a subjective

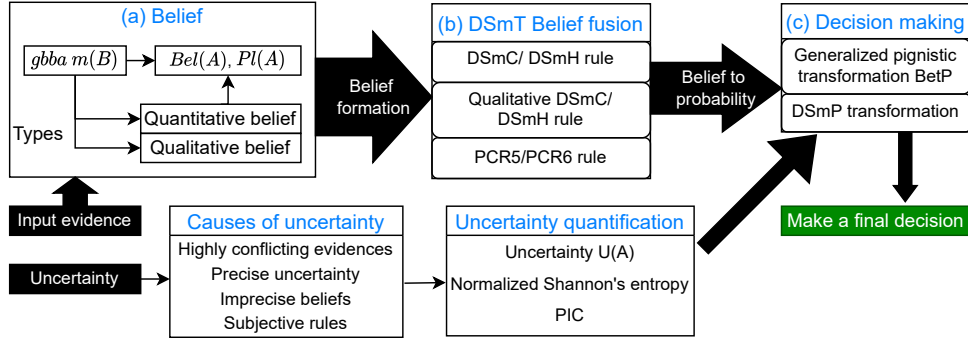


Figure 6: Decision-making process using DSMT where the generalized basic belief assignment (*gbba*) is the formal name of $m(\cdot)$ and BetP refers to a pignistic transformation in *gbba* domain.

measure) which is from the belief functions. The maximum of the GPT function can be used as a decision criterion between two choices.

Making decisions by DSMP can improve the previous probabilistic transformations and increase the strength of a critical decision from the total knowledge.

3.2.4. Applications of DSMT on Machine/Deep Learning

DSMT covers broad information fusion topics of data and sensors in robotics, biometrics, image fusion, trust management, situation analysis, or object tracking. DSMT was mainly applied in developing machine/deep learning-based information fusion tools in binary or multi-class classification problems [93, 94]. DSMT is an extension to the ML multi-class classification models, such as Support Vector Machine (SVM). For example, as a fusion step in SVM One-Against-All, DSMT quantifies partial ignorance by combining conflicting evidence from two complementary SVM results through the PCR6 rule in Section 3.2.1. This PCR6 combination can reduce focal elements in the two sources. The complex, layered ML/DL multi-classifier structures, including CNNs, LSTM, and Random Forests (RF), PCR6 fusion rules can also be integrated as a DSMT-based fusion step [94]. DSMT has been applied to the final decision-making process by combining multi-signal sources of fault characteristics.

3.2.5. Algorithmic Complexity of DSMT

Due to the considered hyper-power domain, given the number of elements n , DSMT's algorithmic complexity in Big-O is $O(2^{2^n})$.

3.3. Transferable Belief Model (TBM)

TBM is a variant of DST to resolve unreasonable results of the DST combination rule (see Section 3.1) when multiple sources provide conflicting evidence [46]. TBM is based on the open-world assumption with two levels of belief reasoning: *credal level* and *pignistic level*. The *credal level* quantifies and updates a belief through a belief function. The *pignistic level* transfers a belief into a probability using the so-called *pignistic probability function* for making a decision [46].

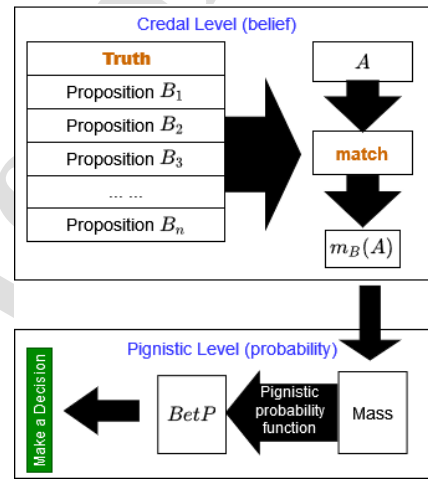


Figure 7: A belief at the credal and pignistic levels and their relationships in TBM.

3.3.1. Belief Formation

TBM defines basic belief masses the same as DST [45] (see Eq. (1)). The *credal level belief function* updating a belief upon the arrival of new evidence is formulated by [46]:

$$m_B(A) = \begin{cases} \frac{\sum_{C \subseteq \bar{B}} m(A \cup C)}{1 - \sum_{C \subseteq \bar{B}} m(C)} & \text{for } (A \subseteq B) \wedge (A \neq \emptyset); \\ 0 & \text{otherwise.} \end{cases} \quad (31)$$

Here $m_B(A)$ means the belief mass supporting propositions A when conditional evidence does not support proposition B as the truth [46]. The $\sum_{C \subseteq \bar{B}} m(C)$ refers to the sum of beliefs supporting a set not supporting B and $\sum_{C \subseteq \bar{B}} m(A \cup C)$ is the sum of beliefs not supporting B or supporting A . For $m_B(A) : \mathbb{R} \rightarrow [0, 1]$, Eq. (31) needs to hold $\sum_{C \subseteq \bar{B}} m(A \cup C) < 1 - \sum_{C \subseteq \bar{B}} m(C)$.

The probability transformed through the pignistic probability function for decision-making is denoted by *BetP*:

$$BetP(x) = \sum_{x \in A \subseteq X} \frac{m(A)}{|A|} = \sum_{x \in A \subseteq X} m(A) \frac{|x \cap A|}{|A|}, \quad (32)$$

where $\frac{m(A)}{|A|}$ means belief mass, $m(A)$, is evenly distributed into the atoms of A , a set of atoms, and $|A|$ means the number of

atoms x in set A (i.e., $x \in A$). The X is the *Boolean algebra* of the subset of Ω , where Ω is a set of worlds (truth). The probability distribution calculated from the pignistic probability function is used for decision-making.

3.3.2. Causes and Types of Uncertainty

TBM considers epistemic uncertainty caused by a lack of evidence.

3.3.3. Uncertainty Quantification

In TBM, uncertainty has not been explicitly discussed. Since a belief at the pignistic level is for decision-making in real-world settings, the pignistic probability function gives a belief mass that considers uncertainty in practice, while the credal level belief function estimates a belief based on observed evidence. Figure 7 shows how a belief is constructed at the credal level upon the arrival of evidence and how the credal level belief is transferred to the pignistic level belief for decision-making.

3.3.4. Applications of TBM on Machine/Deep Learning

Applications of TBM have evolved from the classical associative classification [95] leveraging TBM to produce a robust and compact classifier composed of (pignistic) probability functions summarizing vast mined rule sets [95]. Next, the concept of TBM was broadened by Quost et al. [96] to improve the efficiency of multi-class classification through classifier combination within the belief function framework, marking a crucial step towards handling more complex classification tasks. TBM began to integrate more prominently with deep learning techniques. Soua et al. [84] proposed a Deep Belief Network (DBN) based framework to accurately predict traffic flow using TBM to fuse predictions based on diverse data types. Henni et al. [97] developed a TBM-based localization algorithm to manage ambiguities during zone transitions. On the other hand, Honer and Hettmann [98] focused on improving Occupancy Grid Mapping (OGM) algorithm for autonomous vehicles by using TBM to deliver a more accurate map, integrating information from both LIDAR and OGM maps. Finally, in the fight against cyber threats, Ioannou et al. [99] proposed the Markov Multi-phase TBM (MM-TBM) to deal with multi-phase cyberattacks characterized by uncertainty and conflicting information. These steps reveal the expanding application of TBM in ML/DL and underpin its potential in dealing with a variety of complex real-world scenarios.

3.3.5. Algorithmic Complexity of TBM

The algorithmic complexity of the TBM is largely determined by the operations used to combine belief functions. Similar to the DST, when applying the combination rule in TBM, we consider all subsets of a frame of discernment with n elements, leading to 2^n subsets. Consequently, the algorithmic complexity in Big-O is $O(2^n)$.

3.4. Bayesian Inference (BI)

Bayesian theory has evolved for more than a hundred years [50]. **Bayesian inference (BI) is the process of inductive learning using Bayer's rule [100]. Inductive learning is the**

process of estimating the characteristics of a population from a subset of members of the entire population. Although some literature treats BI as an ML technique due to its statistical nature [101], we treat BI as a belief model because it deals with a subjective probability representing a belief.

3.4.1. Belief Formation

Bayes' rule offers a rational tool for updating beliefs of unknown information, which connects probabilities and information [100]. Beliefs are statements that can have overlapping domains, such as two beliefs A and B and $A \cap B \neq \emptyset$. A higher value returned from a belief function indicates a higher degree of a given belief. Bayesian inference estimates population characteristics θ from a single dataset sample y . A belief is formed via three steps:

1. Prior distribution $p(\theta)$ describes that the belief of θ being true population characteristics.
2. Sampling model $p(y|\theta)$ shows the belief where y means a sample of the huge sample space $\mathcal{Y}$ if θ is true and y needs to be estimated.
3. Posterior distribution $p(\theta|y)$ updates the belief about θ from Bayes' rule based on observed datasets y [100], for the set of all possible parameter values in the parameter space, Θ :

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int_{\Theta} p(y|\tilde{\theta})p(\tilde{\theta})d\tilde{\theta}}. \quad (33)$$

Bayesian inference includes conjugate (i.e., prior and posterior distributions are in the same class) prior distributions, posterior inference, predictive distributions, and confidence regions. There are variants of quantifying the uncertainty of variables depending on different sampling methods.

There are variants of quantifying the uncertainty of variables depending on different sampling methods as follows:

1. *Binomial model*: A random variable $Y \in \{0, 1, \dots, n\}$ has a binomial distribution [100] if

$$\Pr(Y = y|\theta) = \text{dbinom}(y, n, \theta) = \binom{n}{y} \theta^y (1 - \theta)^{n-y}, \quad \text{where } y \in \{0, 1, \dots, n\}, \quad (34)$$

where $E[Y|\theta] = n\theta$ and $\text{Var}[Y|\theta] = n\theta(1 - \theta)$. The prior is an uncertain variable θ from Beta(a, b) distribution [100] if

$$p(\theta) = B(\theta, a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} \quad \text{for } 0 \leq \theta \leq 1, \quad (35)$$

where $\Gamma(x+1) = x!$, $E[\theta] = \frac{a}{a+b}$, and $\text{Var}[\theta] = \frac{E[\theta]E[1-\theta]}{a+b+1}$. Then the posterior distribution [100] is given by:

$$p(\theta|y) = \text{dbeta}(\theta, a+y, b+n-y), \quad (36)$$

where $E[\theta|y] = \frac{a+y}{a+b+n}$ and $\text{Var}[\theta|y] = \frac{E[\theta|y]E[1-\theta|y]}{a+b+n+1}$. This reflects that beta priors are conjugate for the binomial sampling model.

2. *Poisson model*: A random variable Y has a Poisson distribution [100] if

$$\Pr(Y = y|\theta) = \text{dpois}(y, \theta) = \frac{\theta^y e^{-\theta}}{y!} \text{ for } y \in \{0, 1, 2, \dots\}, \quad (37)$$

with $E[Y|\theta] = \theta$ and $\text{Var}[Y|\theta] = \theta$. The conjugate prior can be a gamma(a, b) distribution [100] if

$$p(\theta) = \Gamma(\theta, a, b) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} \text{ for } \theta, a, b > 0, \quad (38)$$

with $E[\theta] = \frac{a}{b}$, $\text{Var}[\theta] = \frac{a}{b^2}$. Combining the gamma family priors and Poisson sampling model, the posterior distribution of θ is given by [100]:

$$p(\theta|y) = \Gamma(a + y, b + n), \quad (39)$$

where b is the number of prior observations, a is the sum of counts from b prior observations, and $E[\theta|y] = \frac{a+y}{b+n}$.

3. *Monte Carlo approximation*: For the arbitrary posterior distributions, Monte Carlo approximation generates random sampling with large S samples to model $p(\theta|y_1, y_2, \dots, y_n)$ [100]. The empirical distribution of the samples $\{\theta^{(1)}, \dots, \theta^{(S)}\}$ can represent $p(\theta|y_1, y_2, \dots, y_n)$. For $S \rightarrow \infty$, the mean θ (i.e., $\bar{\theta}$) and the variance of the samples are given by:

$$\begin{aligned} \bar{\theta} &= \sum_{s=1}^S \frac{\theta^{(s)}}{S} \rightarrow E[\theta|y_1, \dots, y_n], \\ \sum_{s=1}^S \frac{(\theta^{(s)} - \bar{\theta})^2}{(S-1)} &\rightarrow \text{Var}[\theta|y_1, \dots, y_n]. \end{aligned} \quad (40)$$

4. *The normal model*: Normal model is a two-parameter model with mean θ and variance σ^2 . If the sampling model is normal, $\{Y_1, \dots, Y_n|\theta, \sigma^2\} \sim \text{normal}(\theta, \sigma^2)$, the joint sampling density is [100]:

$$p(y_1, \dots, y_n|\theta, \sigma^2) = \prod_{i=1}^n p(y_i|\theta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta)^2}. \quad (41)$$

When σ^2 is known, a posterior distribution can be inferred with conditional prior $p(\theta|\sigma^2)$. The conjugate prior distribution is normal based on $\theta \sim \text{normal}(\mu_0, \tau_0^2)$. The (conditional) posterior distribution is [100]:

$$\begin{aligned} p(\theta|\sigma^2, y_1, \dots, y_n) &\propto \exp\left(-\frac{1}{2}\left(\frac{\theta - \mu_n}{1/\tau_n}\right)^2\right), \\ \tau_n^2 &= \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}, \quad \mu_n = \frac{\kappa_0}{\kappa_0 + n} \mu_0 + \frac{n}{\kappa_0 + n} \bar{y}, \quad \tau_0^2 = \frac{\sigma^2}{\kappa_0}. \end{aligned} \quad (42)$$

When σ^2 is also unknown, the joint inference for the posterior distribution can be calculated based on the prior distributions and sampling model normal (θ, σ^2) as: prior $\frac{1}{\sigma^2} \sim \text{gamma}(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2})$, prior $\{\theta|\sigma^2\} \sim \text{normal}(\mu_0, \frac{\sigma^2}{\kappa_0})$ [100]. Then the joint posterior distribution is given by:

$$\begin{aligned} p(\theta, \sigma^2|y_1, \dots, y_n) &= \\ p(\theta|\sigma^2, y_1, \dots, y_n) p(\sigma^2|y_1, \dots, y_n), \end{aligned} \quad (43)$$

where

$$\begin{aligned} \{\theta|\sigma^2, y_1, \dots, y_n\} &\sim \text{normal}\left(\mu_n, \frac{\sigma^2}{\kappa_n}\right), \\ \text{with } \kappa_n &= \kappa_0 + n, \quad \mu_n = \frac{(\kappa_0 \mu_0 + n \bar{y})}{\kappa_n}, \\ \left\{\frac{1}{\sigma^2}|y_1, \dots, y_n\right\} &\sim \text{gamma}\left(\frac{\nu_n}{2}, \frac{\nu_n \sigma_n^2}{2}\right), \\ \text{with } \nu_n &= \nu_0 + n, \quad \sigma_n^2 = \frac{\left[\frac{\nu_0 \sigma_0^2}{2} + (n-1)s^2 + \frac{(\bar{y} - \mu_0)^2 \kappa_0 n}{\kappa_n}\right]}{\nu_n}. \end{aligned} \quad (44)$$

The probability of an event can be obtained by [100]: (1) determining proper parameter θ and sample spaces; (2) selecting sampling model $p(y|\theta)$ and collecting samples; (3) observing prior distribution $p(\theta)$ by experience or select uninformative prior; (4) calculating posterior distribution $p(\theta|y)$ based on prior and sampling methods; (5) performing sensitivity analysis for a range of parameter values; and (6) finalizing general estimation of a population mean. The reliable estimate of θ contains a best guess and degree of its confidence. Figure 8 shows the diagram of uncertainty and belief process of Bayesian inference.

3.4.2. Causes and Types of Uncertainty

A belief is formed with the unknown values of random variables. In a population, the parameter of population characteristics θ may be unknown. This means the conjugate prior belief $p(\theta)$ is unknown. Before obtaining a dataset y , the subset of a population is also unknown. A sample of dataset y can help to reduce the uncertainty about the population characteristics. This type of uncertainty is caused by a lack of evidence.

3.4.3. Uncertainty Quantification

In single-parameter sampling models, such as Binomial, Poisson, and Monte Carlo approximation, the posterior inference variance of the estimated mean θ measures uncertainty from the current belief formation. The uncertainty is measured by a variance in the Binomial model, Poisson model, and Monte Carlo sampling by:

$$\text{Var}^{\text{Bin}}[\theta|y] = \frac{E[\theta|y]E[1-\theta|y]}{a+b+n+1}, \quad \text{Var}^{\text{Pois}}[\theta|y] = \frac{a+y}{b+n}, \quad (45)$$

$$\text{Var}^{\text{MC}}[\theta|y] = \sum_{s=1}^S \frac{(\theta^{(s)} - \bar{\theta})^2}{(S-1)}, \quad (46)$$

where n is the number of choices of y , a and b are the parameters in Beta(a, b) distribution for a Binomial model, a and b are the parameters of $\Gamma(a, b)$ distribution for a Poisson model, and θ is the estimation of parameters and $\bar{\theta}$ is the mean of θ for Monte Carlo sampling.

In the normal model with mean θ and variance σ^2 , a joint distribution can be transformed to a conditional probability by Eq. (4). The distribution $p(\theta|\sigma^2, y_1, \dots, y_n)$ is defined by Eq. (42) with variance $\tau_n^2 = 1/(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2})$. The posterior inverse variance $\frac{1}{\tau_n^2} = \frac{1}{\tau_0^2} + \frac{n}{\sigma^2}$ indicates that the posterior inverse variance (a.k.a. precision) $1/\tau_n^2$ combines sampling precision $1/\sigma^2$ and prior precision $1/\tau_0^2$.

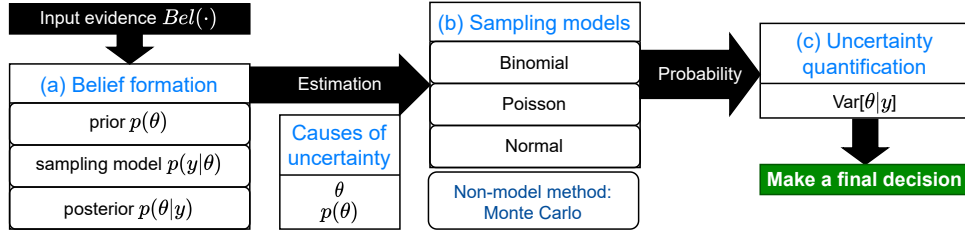


Figure 8: Uncertainty-aware decision making process using Bayesian inference where $p(\theta)$ refers to the probability estimation of θ that causes uncertainty. $Bel(\cdot)$ is the belief of evidence in Eq. (3.1.1). When no model is identified, the Monte Carlo approximation can be used.

3.4.4. Applications of Bayesian Inference on ML/DL

Tipping [101] introduced how Bayesian inference (BI) is used in ML. BI solves a non-deterministic relationship between dependent (Y) and independent (X) variables. Given N data examples and many parameters $\mathbf{w}$, the model of probability of Y given X is computed by $P(Y|X) = f(X; \mathbf{w})$ [101]. The distribution over parameters $\mathbf{w}$ can be inferred from Bayes' rule. Approximation techniques are the key points, such as least-square, maximum likelihood, and regularization. The common choice of a prior is a zero-mean Gaussian prior. The Bayesian way of estimating Maximum A Posteriori (MAP) is for posterior inference. Marginalization serves an important role in the Bayesian framework [101]. Sofman et al. [102] used improved robot navigation in a linear Gaussian model to estimate the posterior distribution of the general Bayesian features and the locale-specific (i.e., preferred by users) features. Tripathi and Govindaraju [103] used relevance vector machines to predict uncertainty in hydrology. Tian et al. [104] analyzed brain image segmentation by applying the Gaussian mixture model (GMM) with a genetic algorithm (GA) and the variational expectation-maximization algorithm.

Many parameters in the ML/DL were expanded in the Bayesian neural network (BNN) [105]. BNN is natural to capture uncertainty for prediction by putting a prior distribution over its weights, such as Gaussian prior distribution: $\theta \sim \mathcal{N}(0, I)$, where θ is the model weights (parameters). Specifically, given a dataset $D = \{X = \{x_1, \dots, x_N\}, Y = \{y_1, \dots, y_N\}\}$, instead of optimizing the deterministic model weights via maximum likelihood estimation (MLE), BNN refers to extending standard networks with posterior inference, which learns a posterior over model weights $p(\theta|D)$ such that model output $f(x, \theta)$ is stochastic.

3.4.5. Algorithmic Complexity of BI

The BI's algorithmic complexity is simply considered as $O(N)$, where N is the number of features.

3.5. Imprecise Dirichlet Model (IDM)

Walley [51] proposed IDM to derive beliefs based on objective statistical inference from multinomial data without prior information. The inference is expressed in terms of posterior upper and lower probabilities. A typical application is predicting the color of the next marble from a bag whose contents are initially unknown. Objective Bayesian does not satisfy this principle because the predicted outcome is unknown, and we

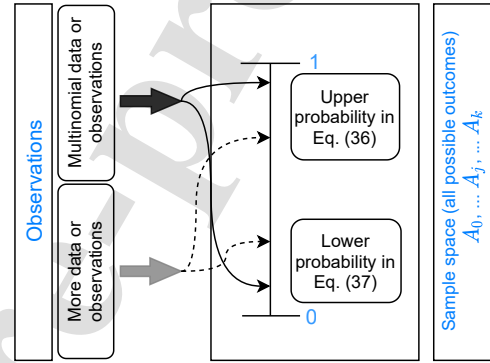


Figure 9: Derivation of the upper and lower bounds in IDM.

cannot formulate the sample space. In IDM, the inferences are expressed as the posterior upper and lower probabilities, $\bar{P}(A|n)$ and $\underline{P}(A|n)$, where A refers to an event and n is the number of observations towards event A . In a multinomial sampling (i.e., $k \geq 2$), the sample space, any event of interest can be identified as a subset of Ω . IDM generates the lower and upper bounds for each value in Beta/Dirichlet PDFs (Probability Density Functions).

3.5.1. Belief Formation

According to Walley [51], IDM can be defined as the set of all Dirichlet (s, t) distribution, such that $0 < t_j < 1$ for $j = 1, 2, \dots, k$ and $\sum_{j=1}^k t_j = 1$ and s is a specified positive constant that does not depend on Ω . Walley suggests $s \leq 2$ where s determines how quickly the upper and lower probabilities converge as the observation data accumulate. This is a prior set and is denoted as μ_0 to model the prior ignorance about chance θ . Given $\theta = \{\theta_1, \theta_2, \dots, \theta_k\}$, which refers to the identical probability distribution of observations. The corresponding set of a posterior distribution, denoted by μ_N , is composed of all Dirichlet ($N + s, \mathbf{t}^*$) distribution (i.e., $\mathbf{t}^* = \{t_1^*, t_2^*, \dots, t_j^*, \dots, t_k^*\}$), where $t_j^* = \frac{n_j + s \times t_j}{N + s}$ and n_j is the number of observations of category ω_j in N trials.

For example, let A_j be the event with outcome ω_j from the next trial. The predictive probability $P(A_j|n)$ under Dirichlet ($N + s, \mathbf{t}^*$) is equal to the posterior mean of θ . By maximizing and minimizing t_j^* with respect to t_j (i.e., $t_j \rightarrow 1$ and $t_j \rightarrow 0$), the posterior upper and lower probabilities of A_j are given by:

$$\bar{P}(A_j|n) = \frac{n_j + s}{N + s} \text{ for } t_j \rightarrow 1; \quad (47)$$

$$\underline{P}(A_j|n) = \frac{n_j}{N+s} \text{ for } t_j \rightarrow 0. \quad (48)$$

If s is the hidden observation and N is the number of revealed observations, those values can be interpreted as the upper and lower bound of the relative frequency of A_j . For example, before making any observation, $n_j = N = 0$, so that $\bar{P}(A_j|n) = \frac{s}{s} = 1$ and $\underline{P}(A_j|n) = \frac{0}{s} = 0$. However, the interval of an IDM bound may be out of range under insufficient evidence (s) conditions. For example, if a bag has nine red balls and one black ball, we randomly pick a ball and obtain a black ball. Now we have evidence $r(\text{black}) = 1$, which gives $\underline{P}(\text{black}) = \frac{1}{2+1} = \frac{1}{3}$. However, we know that the actual probability of having a black ball is $p(\text{black}) = \frac{1}{10}$. So $\underline{P}(\text{black}) > p(\text{black})$ when the number of trials is not sufficient. This case shows that actual probability may be outside the range of IDM under a lack of evidence [24].

3.5.2. Causes and Types of Uncertainty

In IDM, uncertainty decreases as more amount of evidence is received. Hence, it is aligned with the concept of epistemic uncertainty, which can be reduced by increasing the number of observations (or evidence).

3.5.3. Uncertainty Quantification

In IDM, the uncertainty is associated with the imprecision whose degree is captured by the difference between the posterior upper and lower probabilities as:

$$\bar{P}(A_j|n) - \underline{P}(A_j|n) = \frac{s}{N+s}. \quad (49)$$

From the above, we conclude that the imprecision does not depend on event A_j . That is, uncertainty due to the imprecision is based on the number of hidden observations.

3.5.4. Applications of IDM on Machine/Deep Learning

IDM has undergone a progressive development over the years, with its application on ML/DL evolving in complexity and utility. Utkin [106] first proposed the IDMBost algorithm and demonstrated its adaptability and effectiveness in improving classic ML algorithms, such as AdaBoost, by addressing the overfitting problem and reducing iterations. Serafin et al. [107] also devised an adaptation of Credal Decision Trees (CDTs), leveraging the Non-Parametric Predictive Inference Model (NPI-M) for imprecise probabilities. Corani and de Campos [108] presented the TANC classifier, a tree-augmented naïve model that employs imprecise probabilities and models prior near-ignorance using the Extreme Imprecise Dirichlet Model (EDM). This work illustrated how IDM could handle missing data conservatively, providing a reliable and efficient classifier with better performance than previous TANs. These advancements illustrate the potential of IDM for enhancing machine and deep learning applications.

3.5.5. Algorithmic Complexity of IDM

IDM calculates the posterior upper and lower probabilities of an event. The algorithmic complexity of IDM arises primarily from its generation of posterior distributions and the calculation of these probabilities. Considering that the IDM performs

a relatively small amount of operations (e.g., a few arithmetic operations and comparisons), the computational complexity is constant, often expressed as $O(1)$. This indicates that the computational cost remains constant regardless of the size of the input data or the number of outcomes.

3.6. Fuzzy Logic

Łukasiewicz and Alfred Tarski [109] first proposed Łukasiewicz logic, which is the most typical case of many-valued logic. Our discussion focuses on the real-valued semantics of Łukasiewicz logic as the backbone of fuzzy logic. Assume α and β are two propositional formulas with truth values $v(\alpha) = x$ and $v(\beta) = y$, we adopt these semantics in the following context:

$$\begin{aligned} v(\alpha \vee \beta) &= \max\{x, y\}, \quad v(\alpha \wedge \beta) = \min\{x, y\}, \\ v(\sim \alpha) &= 1 - x, \end{aligned} \quad (50)$$

where $\vee$ refers to logical disjunction, $\wedge$ is logical conjunction, and $\sim$ indicates negation.

3.6.1. Belief Formation

Fuzzy logic [110] is a kind of infinite-valued logic defined on type 1 fuzzy sets [44]. A fuzzy logic truth value set $\mathcal{T}$ is a set of linguistic truth values, which is a language generated from a context-free grammar G :

$$\mathcal{T} = L(G). \quad (51)$$

For each truth value $\tau \in \mathcal{T}$, τ is defined as a fuzzy subset of a truth-value set I_τ of Łukasiewicz logic, which is given by:

$$\tau = \int_0^1 \frac{\mu_\tau(v)}{v}, \quad (52)$$

where $\mu_{I_\tau} : [0, 1] \rightarrow c_\tau \in [0, 1]$ and $\mu_\tau : [0, 1] \rightarrow [0, c_\tau]$ are defined as the membership function of I_τ and τ , respectively. Suppose τ has a finite support set $\{v_1, v_2, \dots, v_n\} \subset [0, 1]$, then we can write:

$$\tau = \frac{\mu_1}{v_1} + \frac{\mu_2}{v_2} + \dots + \frac{\mu_n}{v_n}, \quad (53)$$

where $\mu_i = \mu_\tau(v_i)$ for $i \in [1, n]$ and '+' stands for an union operation. Since truth values are fuzzy subsets of truth-value sets of Łukasiewicz logic, logic operations between them can be similarly defined by:

$$\begin{aligned} \mu_{\neg\tau_0} &= 1 - \mu_{\tau_0}, \quad \mu_{\tau_0 \vee \tau_1} = \max\{\mu_{\tau_0}, \\ &\mu_{\tau_1}\}, \mu_{\tau_0 \wedge \tau_1} = \min\{\mu_{\tau_0}, \mu_{\tau_1}\}, \end{aligned} \quad (54)$$

where $\tau_0, \tau_1 \in \mathcal{T}$. From this, we can then derive $\mu_{\tau_0 \Rightarrow \tau_1} = \mu_{\neg\tau_0 \vee \tau_1} = \max\{1 - \mu_{\tau_0}, \mu_{\tau_1}\}$ based on *Kleene-Dienes implication*.

Generally, a type n fuzzy set has a membership function defined based on the set of fuzzy sets of type $n-1$, where $n \geq 2$. Fuzzy numbers [111] can also be formulated as instances of fuzzy sets. In other words, each fuzzy number is attached to a membership function that defines a fuzzy set.

A decision-making process under fuzzy logic often consists of three phases: fuzzification, inference, and defuzzification. A

fuzzifier transforms crispy data into fuzzy sets. An inference engine makes the logical deduction based on given fuzzy rules. A defuzzifier transforms the fuzzy relationships into crispy relationships and makes a final decision.

Zadeh [112] defined $P(A)$ as the probability of a fuzzy event A :

$$P(A) = \int_{\mathbb{R}^n} \mu_A(x) dP = E(\mu_A), \quad (55)$$

where $A \subseteq \mathbb{R}^n$, $\mu_A : \mathbb{R}^n \rightarrow [0, 1]$ is the membership function of A , and $P(A)$ represents the belief of a fuzzy event A .

Let S be a nonempty set. A fuzzy set $\mathcal{P}$ in S is characterized by a membership function: $\mu_p : S \rightarrow [0, 1]$. That is

$$\mu_p(x) = \begin{cases} 1, & \text{if } x \in S \\ 0, & \text{if } x \notin S \\ (0, 1) & \text{if } x \text{ is partly in } S \end{cases} \quad (56)$$

Alternatively, a fuzzy set $\mathcal{P}$ in S is an object having the form

$$\mathcal{P} = \{ \langle x, \mu_p(x) \rangle \mid x \in S \} \quad (57)$$

where the function: $\mu_p(x) : S \rightarrow [0, 1]$ defines the degree of membership of the element, $x \in S$.

The closer the membership value $\mu_p(x)$ to 1, the more x belongs to S , where grades 1 and 0 represent full membership and full non-membership. A fuzzy set is a collection of objects with graded membership, that is, having a degree of membership.

3.6.2. Causes and Types of Uncertainty

Uncertainty in fuzzy logic mostly comes from *linguistic imprecision* or *vagueness*, leading to generating unpredictability, multiple knowledge frames, and/or incomplete knowledge.

3.6.3. Uncertainty Quantification

Zadeh [112] defined two types of fuzzy sets: Type-1 fuzzy set and Type-2 fuzzy set. In Type-1 fuzzy sets, the uncertainty of fuzzy events introduces unpredictability and multiple knowledge frames. Zadeh formulated the uncertainty of a fuzzy event A based on the entropy of event A , $H^P(A)$, which is given by:

$$H^P(A) = - \sum_{i=1}^n \mu_A(x_i) P(x_i) \log P(x_i), \quad (58)$$

where $A = \{x_1, x_2, \dots, x_n\}$, μ_A is the membership function of A , and $P = \{P(x_1), P(x_2), \dots, P(x_n)\}$. Here, $P(x_i)$ refers to the probability of occurring event x_i .

Fuzzy logic research mainly focused on investigating uncertainty measures on Type-2 fuzzy sets, which can provide a way of accurately and effectively measuring fuzziness and uncertainty characteristics of fuzzy complex systems with two membership functions [111]. Wu and Mendel [113] proposed five novel uncertainty metrics, called *centroid*, *cardinality*, *fuzziness* (*entropy*), *variance*, and *skewness*, to measure uncertainty in interval Type-2 fuzzy sets. They further evaluated these metrics with inter-uncertainty and intra-uncertainty raised in words paradigms [114]. Zhai and Mendel [115] extended the five metrics to *general Type-2 fuzzy sets*.

3.6.4. Applications of Fuzzy Logic on ML/DL

In general, fuzzy set theory has two types of applications. First, it was used in the data preprocessing stage for data selection and preparation, such as modeling vague data [116] and summarizing crisp data [117] with fuzzy sets. As described in Figure 10, the deterministic data must be fuzzified before being processed by ML/DL module. Similarly, the fuzzy output must be defuzzified to have a deterministic meaning. Second, the fuzzy data analysis [118] can further process the fuzzy data. This can be implemented by either extending the traditional data analysis methods to fuzzy data or embedding the data to a fuzzy space [119].

Recently, *fuzzy deep neural networks* (FDNNs) are considered for a system using both fuzzy logic and deep neural networks (DNNs) to deal with uncertainty or ambiguity in data [120]. According to fuzzy data preprocessing and analysis discussed earlier, FDNNs fall into *ensemble models* and *integrated models*. The ensemble models refer to ensembles of DL and fuzzy logic systems with three models, including the models with fuzzy inputs, fuzzy outputs, and parallel models. Wang et al. [121] proposed a DL model that takes fuzzy feature points for input for damaged fingerprint classification. Zhang et al. [122] proposed a DL model with fuzzy granulation features to predict time-series data. Chopade and Narvekar [6] proposed an ensemble of fuzzy logic and DL to predict fuzzy memberships for document summarization. Deng et al. [123] proposed a DL architecture with DL layers and fuzzy membership functions running in parallel. FDNNs have been applied in various application domains, such as traffic control [124, 125], surveillance and security [126, 127], text processing [128, 129], image processing [130], and time-series prediction [131]. The integrated models incorporated fuzzy logic as a part of DL models. In particular, Pythagorean Fuzzy Deep Boltzmann Machine (PFDBM) [132] was developed based upon the DBM [133]. PFDBM used the Pythagorean Fuzzy Set (PFS) [134] to replace standard real-valued parameters. El Hatri and Boumhidi [135] developed a DL model in which a network architecture was designed based on stacked-auto-encoders (SAE) where multiple hyperparameters, such as the learning rate and the momentum, were determined using fuzzy logic systems.

3.6.5. Algorithmic Complexity of Fuzzy Logic

Fuzzy logic is based on the support set and corresponding membership function. Specifically, each fuzzy event is a subset of the support set, and its uncertainty is the combination of uncertainties on each element in the support set. According to Eq. 57 the support set has the same cardinality as the fuzzy set. Hence, the algorithmic complexity is $O(|\mathcal{P}|)$, where $|\mathcal{P}|$ is the cardinality of the fuzzy set.

3.7. Kleene's Three-Valued Logic (TVL)

3.7.1. Belief Formation.

Kleene [43] first proposed TVL in 1938. Its truth table is shown in Table 2, where p_1 and p_2 are two logical variables. TVL's belief distribution is determined by the logical values of logical variables, i.e., $\mathbf{b}(p)_q = 1$ when $p = q$ and 0 otherwise.

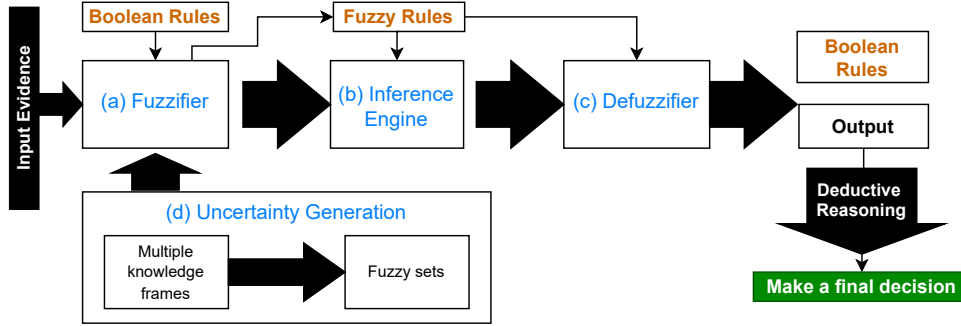


Figure 10: Uncertainty-aware decision making in Fuzzy Logic.

Table 2: TRUTH TABLE OF THE TVL

$p_1 \wedge p_2$				$p_1 \vee p_2$			
$p_1 \backslash p_2$	T	U	F	$p_1 \backslash p_2$	T	U	F
T	T	U	F	T	T	T	T
U	U	U	F	U	T	U	U
F	F	F	F	F	T	U	F

The p is a logical variable and q is a logical value from $\{T, U, F\}$ where T is true, U is unknown, and F is false.

Kleene Algebras and TVL. Kleene's TVL is a special case of Kleene algebras. The properties of Kleene algebra, $\mathcal{K} = (K, \vee, \wedge, \sim, F, T)$, are:

- $\mathcal{K}$ is a bounded distributive lattice; and
- $\forall a, b \in K, \sim(a \wedge b) = \sim a \vee \sim b, \sim \sim a = a$, and $a \wedge \sim a \leq b \vee \sim b$.

Here we apply the semantics where $\vee$ means logical disjunction, $\wedge$ means logical conjunction and $\sim$ means negation. It can be easily derived that Kleene's TVL, in which $K = \{T, U, F\}$, is a Kleene algebra where $T = \sim F, F = \sim T$, and $U = \sim U$.

Rough Sets and Kleene Algebras. Kleene algebras are related to rough sets. The relationships between these two concepts are as follows. Given an information system, $I = (S, \mathbb{A})$, where S is a set of objects and $\mathbb{A}$ is a set of attributes $a : x \mapsto a(x)$ for any $x \in S$, we can define the set of equivalence relationships, $IND(I) : IND(I) = \{IND(A) : A \subseteq \mathbb{A}\}$, where $IND(A) = \{(x, y) \in S^2 : \forall a \in A, a(x) = a(y)\}$. Given any equivalence relationship $R \in IND(I)$, a rough set $X \in (S \times S)/R$ is a pair $(RX, \bar{R}X)$, where RX and $\bar{R}X$ are called the R -lower and R -upper approximation of X , respectively. More specifically,

$$\begin{aligned} RX &= \bigcup \{Y \in S/R : Y \subseteq X\}, \\ \bar{R}X &= \bigcup \{Y \in S/R : Y \cap X \neq \emptyset\}, \end{aligned} \quad (59)$$

where S/R is the collection of equivalence classes corresponding to R .

Given any set S with $|S| \geq 2$, universal equivalence relationship $R := S \times S$ and information system $I = (S, \mathbb{A})$, we can induce a three-valued algebra on a collection of rough sets, $\mathcal{RS}$, with the Kleene semantics by:

$$\mathcal{RS} = \{(\underline{R}A, \bar{R}A) : A \subseteq S\} = \{(S, S), (\emptyset, S), (\emptyset, \emptyset)\}. \quad (60)$$

Here, if we define $\sim X := (RX^c, \bar{R}X^c)$, we have $(S, S) = \sim(\emptyset, \emptyset)$, $(\emptyset, S) = \sim(\emptyset, S)$, and $(\emptyset, \emptyset) = \sim(\emptyset, \emptyset)$. This means $K \cong \mathcal{RS}$. In general, given the set of all logic functions (propositional formula) denoted by $\mathcal{F}$, the set of all Kleene algebras by $\mathcal{A}_K$, and the collections of all rough sets over all possible information systems by $\mathcal{A}_{RS}$, the following theorem [136] is held:

$$\forall \alpha, \beta \in \mathcal{F}, \alpha \models_{\mathcal{A}_K} \beta \Leftrightarrow \alpha \models_{\mathcal{A}_{RS}} \beta. \quad (61)$$

The above can be read by: For any logic functions α and β in $\mathcal{F}$, if β is a semantic consequence of α in $\mathcal{A}_K$, then β is a semantic consequence of α in $\mathcal{A}_{RS}$. We summarize the uncertainty-aware decision-making process using Kleene's TVL in Figure 11.

3.7.2. Causes and Types of Uncertainty

Uncertainty is formalized as a logical value U (unknown) and its relationship with two classical logical values T (true) and F (false) are shown in Table 2. The stated uncertainty here refers to *unpredictability* because of a lack of information or knowledge. For example, in rough sets, due to unpredictable noises, sets are represented by approximation spaces.

3.7.3. Uncertainty Quantification

Uncertainty in TVL represents an unknown or unspecified state in the decision-making using TVL. This is related to vacuity uncertainty caused by a lack of information/knowledge or non-specificity. Since uncertainty is regarded as a logical value, uncertainty value can be quantified through logical operations of logical variables. In Kleene's TVL, the three values of T , U (uncertainty), and F are often defined by 1, 0, and -1. As seen in Table 2, uncertainty, U , can be ignored under $\wedge$ to decide T or F while it can be used to support T over F .

3.7.4. Applications of TVL on Machine/Deep Learning

As being restricted by its design, TVL can only represent qualitative uncertainties with limited granularity. Thus, TVL is only applied in primary theoretical ML/DL research. Kashkevich and Krasnoprosin [137] defined a function of TVL to solve classification problems in pattern recognition tasks. They viewed the accepted accurate classification as T , accepted incorrect classification as F , and refused classification as U . Dahl [138] leveraged TVL to construct a database used for natural

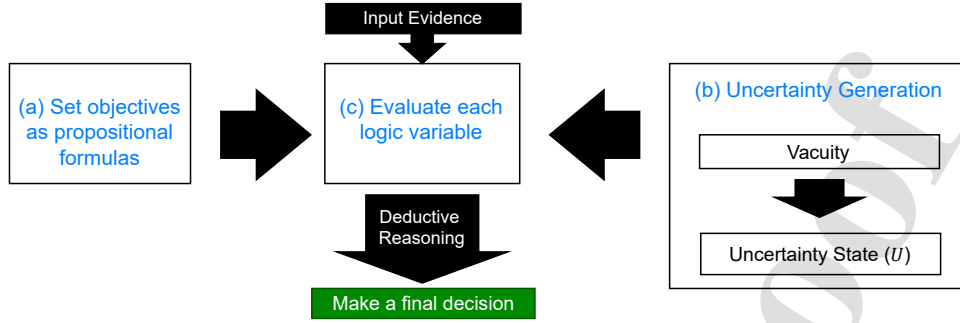


Figure 11: Decision-making using Kleene's TVL.

language consultation. Codd [139] applied TVL in Structured query language (SQLs), with the “Null” value behaving like the uncertain value U in TVL. TVL is rarely observed for its applications in recent research. Instead, as the topological generalization of TVL, rough sets are widely used in ML/DL for powerful quantitative uncertainty representations, as described in Section 4.3.

3.7.5. Algorithmic Complexity of TVL

TVL represents uncertainty as a logical value that can be derived by logical operations. The uncertainty status of a compound expression, including n simple expressions, needs $n - 1$ logical operations to be determined. Thus, the algorithmic complexity is $O(1)$ for each simple expression and $O(n)$ for a compound expression.

3.8. Subjective Logic (SL)

As a variant of DST, Jøsang [24] proposed a belief model, called *Subjective Logic* (SL) that describes the subjectivity of an opinion in terms of multiple belief masses and uncertainty.

3.8.1. Belief Formation

Since a binomial opinion is a special case of multinomial opinions where the number of belief masses is two, for brevity, we only provide the descriptions of multinomial opinions and hyper-opinions.

Multinomial Opinions: In SL, a multinomial opinion in a given proposition x is represented by $\omega_x = (b_x, u_x, a_x)$ where a domain is $\mathbb{X}$, a random variable $X \in \mathbb{X}$, $\kappa = |\mathbb{X}| > 2$ (for a binomial opinion, $\kappa = |\mathbb{X}| = 2$), and the additivity requirement of ω_x is given as $\sum_{x \in \mathbb{X}} b_x(x) + u_x = 1$ where each parameter refers to: (1) b_x : *belief mass distribution* over $\mathbb{X}$; (2) u_x : *uncertainty mass* representing *vacuity of evidence*; and (3) a_x : *base rate distribution* over $\mathbb{X}$.

The projected probability distribution of multinomial opinions is given by:

$$\mathbf{P}_X(x) = b_x(x) + a_x(x)u_x, \quad \forall x \in \mathbb{X}. \quad (62)$$

The probability distribution of a multinomial opinion follows Dirichlet distribution [24].

Hyper-opinions: Hyper-opinions represent multiple choices under a specific singleton belief x where belief mass is allowed to be assigned to a composite value $x \in \mathcal{C}(\mathbb{X})$ consisting of a

set of singleton values x 's. Belief masses assigned to composite values $x \in \mathcal{C}(\mathbb{X})$ can be used to estimate the vagueness of an opinion. *Hyperdomain*, denoted by $\mathcal{H}(\mathbb{X})$, is the reduced powerset of $\mathbb{X}$ which is the set of $\mathcal{P}(\mathbb{X})$ that excludes $\{\mathbb{X}\}$ and $\{\emptyset\}$. Hyperdomain can be defined by:

$$\text{Hyperdomain} : \mathcal{H}(\mathbb{X}) = \mathcal{P}(\mathbb{X}) \setminus \{\{\mathbb{X}\}, \{\emptyset\}\}. \quad (63)$$

Given X as a hyper variable in $\mathcal{H}(\mathbb{X})$, a hyper-opinion on X is represented by $\omega_X = (b_X, u_X, a_X)$ where each element includes: (1) b_X : *belief mass distribution* over $\mathcal{H}(\mathbb{X})$; (2) u_X : *uncertainty mass* representing *vacuity of evidence*; and (3) a_X : *base rate distribution* over $\mathbb{X}$, where $\sum_{x \in \mathcal{H}(\mathbb{X})} b_X(x) + u_X = 1$.

The projected probability distribution of a hyper-opinion can be given by:

$$\mathbf{P}_X(x) = \sum_{x_i \in \mathcal{H}(\mathbb{X})} a_X(x|x_i) b_X(x_i) + a_X(x)u_X, \quad (64)$$

$$a_X(x|x_i) = \frac{a_X(x \cap x_i)}{a_X(x_i)}, \quad \forall x, x_i \in \mathcal{H}(\mathbb{X}), \quad (65)$$

where $a_X(x|x_i)$ is the relative base rate and $a_X(x_i) \neq 0$. For the binomial or multinomial opinions, the additivity requirement is met (i.e., $\sum_{x \in \mathbb{X}} \mathbf{P}_X(x) = 1$). However, for the hyper-opinion, the additivity requirement may not be met, but $\mathbf{P}_X(x)$ follows super-additivity (i.e., $\sum_{x \in \mathcal{H}(\mathbb{X})} \mathbf{P}_X(x) \geq 1$) with a hyper-domain, $\mathcal{H}(\mathbb{X})$.

Hyper-opinions can be represented by Dirichlet PDFs and the hyper-Dirichlet distribution [140]. To do so, we can project a hyper-opinion into a multinomial opinion based on [24]. The approximation by the projection of hyper-opinions to multinomial opinions removes vague information in the representation of opinions. This allows a decision-maker to see a particular opinion without the veil of vagueness, which facilitates a more direct and intuitive interpretation of the opinion.

3.8.2. Causes and Types of Uncertainty

SL discusses three types of uncertainties [1]: *vacuity*, *vagueness*, and *dissonance*. *Vacuity* uncertainty is caused by a lack of evidence or knowledge. *Vagueness* uncertainty is caused by vague observations, leading to failure in identifying a distinctive singleton belief. *Dissonance* uncertainty is introduced due to conflicting evidence, resulting in inconclusiveness. Vacuity and dissonance can be understood as epistemic uncertainty, which can be reduced with more evidence. Vagueness is related to fuzziness, representing aleatoric uncertainty in its nature.

3.8.3. Uncertainty Quantification

Uncertainty measures across all belief masses are calculated based on the sum of uncertainty masses associated with individual belief masses, as discussed above. They include *total vacuity* (same as u_X), *total vagueness* (b_X^{TV}), and *total dissonance* (b_X^{Diss}):

$$\begin{aligned} u_X &= \sum_{x \in \mathcal{R}(\mathbb{X})} \mathbf{u}_X^F(x), \quad b_X^{TV} = \sum_{x \in \mathcal{R}(\mathbb{X})} b_X(x), \\ b_X^{Diss} &= \sum_{x_i \in \mathbb{X}} b_X^{Diss}(x_i), \end{aligned} \quad (66)$$

where $\mathbf{u}_X^F(x)$ refers to a focal uncertainty (vacuity per belief), $b_X(x)$ is a belief mass supporting x , and $b_X^{Diss}(x_i)$ indicates dissonance per singleton belief.

The uncertainty associated with each belief mass are:

1. *Belief vacuity* (a.k.a. focal uncertainty), $\mathbf{u}_X^F(x)$, is computed by:

$$\mathbf{u}_X^F(x) = \mathbf{a}_X(x)u_X, \quad (67)$$

where $x \in \mathcal{R}(\mathbb{X})$ and $\mathbf{u}_X^F(x) : \mathcal{R}(\mathbb{X}) \rightarrow [0, 1]$.

2. *Belief vagueness*, $b_X^V(x)$, is the vague belief mass associated with an individual belief $b_X(x)$ where the belief has a composite value, $x \in \mathcal{C}(\mathbb{X})$.

$$b_X^V(x) = \sum_{\substack{x_j \in \mathcal{C}(\mathbb{X}) \\ x_i \not\subseteq x}} \mathbf{a}_X(x|x_i) b_X(x_i), \quad \forall x \in \mathcal{R}(\mathbb{X}), \quad (68)$$

where $\mathbf{a}_X(x|x_i)$ is found in Eq. (64). Note that this belief vagueness can be only measured for hyper opinions (i.e., $x_i \in \mathcal{C}(\mathbb{X})$).

3. *Belief dissonance*, $b_X^{Diss}(x_i)$, estimates the difference between belief $b_X(x_i)$ and other belief masses by:

$$b_X^{Diss}(x_i) = \frac{b_X(x_i) \sum_{x_j \in \mathbb{X} \setminus x_i} b_X(x_j) \text{Bal}(x_j, x_i)}{\sum_{x_j \in \mathbb{X} \setminus x_i} b_X(x_j)}, \quad (69)$$

where the relative mass balance between belief masses, $b_X(x_j)$ and $b_X(x_i)$, is given by:

$$\text{Bal}(x_j, x_i) = 1 - \frac{|b_X(x_j) - b_X(x_i)|}{b_X(x_j) + b_X(x_i)}. \quad (70)$$

In addition, Jøsang [24] proposed a technique called *uncertainty maximization* where uncertainty refers to vacuity. Since we discussed different types of uncertainties in SL, we will call uncertainty maximization ‘vacuity maximization.’ In subjective opinions formulated by SL, when vacuity (i.e., u_X) is 0 (i.e., complete certainty), then an opinion stops updating and is ended with its final state. However, if the opinion’s dissonance is high, it leads to a situation of ‘multiple knowledge frames,’ and a decision maker still cannot make a decision even with zero uncertainty due to the same probabilities assessed for given belief masses. To mitigate this effect, we can allow an opinion to receive more information or consider other opinions

for the opinion being updated. This can be enabled by maximizing vacuity based on the minimum belief mass [24]. Given opinion $\omega_X = (\mathbf{b}_X, u_X, \mathbf{a}_X)$ where $\mathbf{P}_X(x) = \mathbf{b}_X(x) + \mathbf{a}_X(x)u_X$ in Eq. (62), the corresponding vacuity-maximized opinion is denoted by $\tilde{\omega}_X = (\tilde{\mathbf{b}}_X, \tilde{u}_X, \mathbf{a}_X)$ where $\tilde{u}_X$ and $\tilde{\mathbf{b}}_X$ are computed by:

$$\tilde{u}_X = \min_i \left[\frac{\mathbf{P}_X(x_i)}{\mathbf{a}_X(x_i)} \right], \quad \tilde{\mathbf{b}}_X(x) = \mathbf{P}_X(x) - \mathbf{a}_X(x)\tilde{u}_X, \quad \text{for } x_i \in \mathbb{X}. \quad (71)$$

A vacuity-maximized opinion is an epidemic opinion based on the uncertain likelihood of the variable’s value in an unknown past or future for a given proposition. Notice $\mathbf{a}_X$ stays the same in $\tilde{\omega}_X$, and the vacuity-maximized opinion is the same as an epidemic opinion. Remind that the purpose of updating ω_X to $\tilde{\omega}_X$ is to allow the opinion to be further updated by receiving new evidence or being combined with other opinions, which are possible only when $u_X > 0$.

When one makes a decision under uncertainty using SL, we can leverage SL’s capability to estimate multidimensional uncertainty (i.e., vagueness, vacuity, and dissonance) to make effective decisions. As in Figure 12, after estimating multiple dimensions of uncertainty, one can use the vacuity maximization technique if more evidence is needed in order to allow considering more evidence even under low vacuity, representing high certainty due to a large volume of evidence collected. Recall that SL-based opinion cannot be updated or is rarely updated significantly if its vacuity is or close to zero. One can also consider other opinions by using a variety of fusion operators in SL [24], which can generate a single opinion with the updates of corresponding belief masses and vacuity values. The generated single opinion can be also assessed based on which decision has the most utility by normalizing the opinion based on each decision (i.e., belief mass)’s utility. Most decision-making problems can be solved by these processes which allows us to make a decision with minimum uncertainty and maximum utility. However, if all decisions have the same uncertainty-aware maximum utility, one can select a decision at random, which we want to avoid.

3.8.4. Applications of SL on Machine/Deep Learning

Recently SL has been considered along with machine/deep learning. Uncertainty reasoning to solve classification tasks has been studied by leveraging SL to consider vacuity and dissonance uncertainty dimensions [5, 63]. In addition, SL-based opinion formulation is used to infer subjective opinions along with DL in the presence of adversarial attacks [64, 65, 66]. Further, SL-based opinions are considered along with deep reinforcement learning to propose uncertainty-aware decision-making [67]. Using SL combined with DL is relatively a recent move since 2018 [5, 65, 66]. As discussed in Section 4.1 SL’s belief mass is obtained from the output of the last layer in DNNs to assist a decision-making process with a level of estimated confidence based on uncertainty masses (e.g., vacuity, dissonance, vagueness). Since then, SL has been applied in graph neural networks to solve semi-supervised node classification [141]. [142] also combined SL with probabilistic soft logic (PSL) to develop a hybrid approach, called CST⁺, to pre-

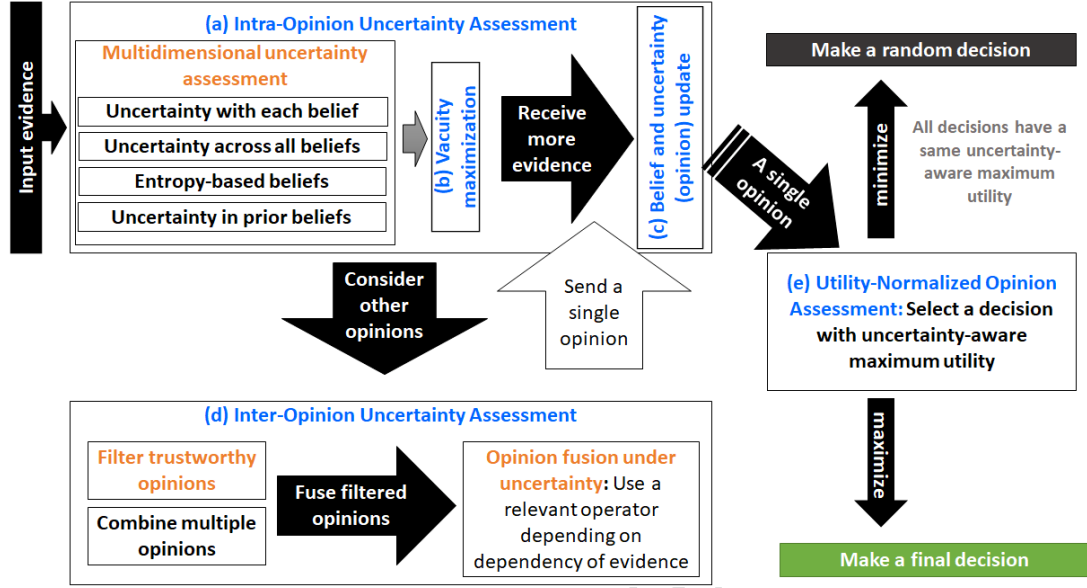


Figure 12: Uncertainty-aware decision-making process using Subjective Logic.

dict unknown opinions. Further, SL leveraged evidential neural networks (ENNs) to solve classification and out-of-distribution problems where multidimensional uncertainty estimates in SL are used to predict a class label. SL was also incorporated into deep reinforcement learning (DRL) for achieving effective decision-making [67] or detecting intent of fake news [143] using multidimensional uncertainty estimates where each class is formulated as an uncertain opinion in SL.

3.8.5. Algorithmic Complexity of SL

SL has a variety of operators applicable to aggregate two opinions. For each operator, the algorithmic complexity is $O(1)$ while aggregating one opinion with n opinions can be linear, i.e., $O(n)$.

4. Applications of Belief Theory in DL

In this section, we review several hybrid frameworks that combine belief models and neural networks, including evidential (or subjective) neural networks, fuzzy neural networks, and rough deep neural networks.

4.1. Evidential Neural Networks (ENNs)

Evidential neural networks (ENNs) [5] is a hybrid framework of subjective belief models and neural networks. They are similar to classic neural networks for classification. The main difference is that the softmax layer is replaced with an activation function in ENNs, e.g., ReLU, to ensure non-negative output in the range of $[0, +\infty]$, which is taken as the evidence vector for the predicted Dirichlet distribution, or equivalently, multinomial opinion.

4.1.1. Key Formulation of ENNs

Given the feature vector $\mathbf{x}$ of an input sample, let $f(\mathbf{x}|\theta)$ represent the evidence vector by the network for the classification, where θ is network parameters. Then the corresponding Dirichlet distribution has parameters $\alpha = f(\mathbf{x}|\theta) + 1$, where the k -th parameter α_k denotes the effective number of observations of the k -th class, and the total number of classes is K . Let $\mathbf{p} = (p_1, \dots, p_K)^T$ be the probabilities of the K predefined classes. The Dirichlet PDF (i.e., $\text{Dir}(\mathbf{p}; \alpha)$) with $\mathbf{p}$ as a random vector is defined by:

$$\text{Dir}(\mathbf{p}|\alpha) = \frac{1}{B(\alpha)} \prod_{k \in \mathbb{Y}} p_k^{(\alpha_k-1)}, \quad (72)$$

where $\frac{1}{B(\alpha)} = \frac{\Gamma(\sum_{k \in \mathbb{Y}} \alpha_k)}{\prod_{k \in \mathbb{Y}} \Gamma(\alpha_k)}$, $\alpha_k \geq 0$, and $p_k \neq 0$, if $\alpha_k < 1$. The expected value of class probabilities $\mathbf{p} = (p_1, \dots, p_K)^T$ is given by:

$$\mathbb{E}[p_k] = \frac{\alpha_k}{\sum_{j=1}^K \alpha_j} = \frac{e_k + a_k W}{\sum_{j=1}^K e_j + W}. \quad (73)$$

The observed evidence in a Dirichlet PDF $\text{Dir}(\mathbf{p}|\alpha)$ can be mapped to a multinomial opinion $(b_1, \dots, b_K, u)$ as follows:

$$b_k = \frac{e_k}{S}, \quad u = \frac{W}{S}, \quad \text{for } k = 1, \dots, K, \quad (74)$$

where $S = \sum_{k=1}^K \alpha_k$ refers to the Dirichlet strength. Without loss of generality, we set $\alpha_k = 1/K$ and the non-informative prior weight (i.e., $W = K$), which indicates that $a_k \cdot W = 1$ for each $k \in \{1, \dots, K\}$. Therefore, the output of an ENN can be applied to measure the subjective uncertainty about the predictive class variable y in different types, such as vacuity and dissonance as defined based on a multinomial opinion (See Section 3.8).

The Bayesian framework of ENNs was proposed in [23] that considers a prior distribution on the network parameters θ , denoted by $P(\theta)$. Let $P(\theta|\mathcal{D})$ be the posterior PDF, where $\mathcal{D}$ refers to the training set. Let $\text{Cat}(y|\mathbf{p})$ be the PDF of the categorical distribution about the predictive variable y , where the class probabilities $\mathbf{p}$ are the parameters. We can then show terms associated with different uncertainty, including data uncertainty (DC), subjective uncertainty (SC), and model uncertainty (MC), as follows:

$$P(y|\mathbf{x}, \mathcal{D}) = \int \int \underbrace{\text{Cat}(y|\mathbf{p})}_{DC} \underbrace{P(\mathbf{p}|\mathbf{x}, \theta)}_{SC} \underbrace{P(\theta|\mathcal{D})}_{MC} d\mathbf{p} d\theta, \quad (75)$$

where $P(\mathbf{p}|\mathbf{x}, \theta) = \text{Dir}(\mathbf{p}|\alpha)$ and $\alpha = f(\mathbf{x}, \theta)$. In this expression, data (aleatoric), subjective (distributional), and model (epistemic) uncertainty are modeled by a separate term within an interpretable probabilistic framework. The data uncertainty is described by the point-estimate categorical distribution, $\text{Cat}(y|\mathbf{p})$. The subjective (or distributional) uncertainty is described by the distribution over predictive class variables $P(\mathbf{p}|\mathbf{x}, \theta)$. The model uncertainty is described by the posterior distribution over the parameters, given the data. The relationship between uncertainties is made explicit – model uncertainty affects estimates of subjective uncertainty, which in turn affects the estimates of data uncertainty. This forms a hierarchical model with three layers of uncertainty: the posterior over classes, the per-data Dirichlet prior distribution, and the global posterior distribution over model parameters. The uncertainty due to the mismatch between testing and training distributions can be measured by two methods. First, as the Dirichlet distribution $P(\mathbf{p}|\mathbf{x}, \theta)$ is equivalent to a subjective multinomial opinion based on the mapping defined in Eq. (74), we can quantify subjective uncertainty types directly based on the Dirichlet distribution, such as vacuity and dissonance, where vacuity captures elements of distributional uncertainty. Second, the distributional uncertainty can be measured based on mutual information between the categorical label y and the class probabilities $\mathbf{p}$ as:

$$\begin{aligned} & \underbrace{I[y, \mathbf{p}|\mathbf{x}, \mathcal{D}]}_{\text{epistemic uncertainty}} \quad (76) \\ &= \underbrace{\mathcal{H}[\mathbb{E}_{P(\mathbf{p}|\mathbf{x}, \mathcal{D})}[\text{Cat}(y|\mathbf{p})]]}_{\text{entropy}} - \underbrace{\mathbb{E}_{P(\mathbf{p}|\mathbf{x}, \mathcal{D})}[\mathcal{H}[\text{Cat}(y|\mathbf{p})]]}_{\text{aleatoric uncertainty}}. \end{aligned}$$

We note that distributional uncertainty and vacuity negatively correlate if the parameters θ are deterministic. The former is maximized (and the latter is minimized) when all categorical distributions are equiprobable, which occurs when the Dirichlet distribution is flat.

Recent research works have proposed several methods to estimate model parameters θ of an ENN model based on training pairs of feature vectors and class labels, i.e., $\{\mathbf{x}^{(i)}, y^{(i)}\}_{i=1}^N$. As class labels are indirect samples of Dirichlet distributions, traditional loss functions, such as mean squared error (MSE) and cross-entropy (CE) functions, are ineffective for training ENNs to predict Dirichlet distributions. In particular, MSE and CE functions effectively train a model to predict class probabilities. However, the two very different Dirichlet distributions

may have the same projected class probabilities and hence cannot be well differentiated by MSE and CE functions. Recent works addressed this challenge and developed a regularization term enforcing that each predicted Dirichlet distribution should be close to a uniform Dirichlet prior [5, 23]. Several studies developed generative adversarial networks (GANs) to generate synthetic out-of-distribution feature vectors (OODs) and developed regularization terms enforcing small vacuity values for these OODs [144, 4]. Another research developed a new NN architecture called *posterior networks* composed of three main components: an *encoder*, which outputs a position in a latent space, a *normalizing flow*, which performs a density estimation in this latent space, and a *Bayesian loss* for the network training [145, 146].

4.1.2. Causes and Types of Uncertainty

Since ENNs provide a hybrid framework of subjective belief models and neural networks, we can estimate evidential uncertainty, such as *vacuity* (*scenario uncertainty*) and *dissonance* (*discord uncertainty*), based on a subjective opinion. Recall that vacuity is due to a lack of evidence introducing uncertainty by incomplete knowledge. Dissonance is due to conflicting evidence, resulting in multiple knowledge frames.

4.1.3. Uncertainty Quantification

ENNs estimate Dirichlet distribution parameters directly, which can be transferred to a subjective opinion. After then, we can estimate vacuity (u) and dissonance ($diss$) based on SL-based subjective opinion where there are K classes and e_k number of evidence to support each class k by:

$$u = \frac{K}{\sum_{k=1}^K e_k + K}, \quad (77)$$

$$diss = \sum_{k=1}^K \left(\frac{b_k \sum_{j \neq k} b_j \text{Bal}(b_j, b_k)}{\sum_{j \neq k} b_j} \right), \quad (78)$$

$$\text{Bal}(b_j, b_k) = \begin{cases} 1 - \frac{|b_j - b_k|}{b_j + b_k} & \text{if } b_j b_k \neq 0 \\ 0 & \text{if } \min(b_i, b_j) = 0, \end{cases} \quad (79)$$

where b_k and b_j refer to the belief masses supporting k class and j class, respectively.

4.1.4. Applications of ENNs

There is a whole class of evidential neural networks with the interpretation that evidence represents the number of nearby training samples of various classes relative to the sample under test. This includes the generative version from [144], posterior networks based on density-based pseudo-counts in [145], and Epistemic Neural networks [21] that allow a general interface to distinguish epistemic from aleatoric uncertainty. The ENNs have been applied on several applications in different domains, such as justified true belief models [147, 148], active learning on image data [149], misclassification and out-of-distribution detection on graph data [7, 4], event early detection on time series data [3], and self-training on NLP task [150].

4.2. Fuzzy Deep Neural Networks (FDNNs)

Fuzzy deep neural networks (FDNNs) are a hybrid framework of fuzzy logic systems and deep neural networks [120]. FDNNs are designed to address the drawback that deep neural networks are sensitive to the uncertainties and the ambiguities of real-world data. Multiple approaches are developed to implement an FDNN. Some models, such as Fuzzy Restricted Boltzmann Machines (FRBMs) [126], consider the concept of fuzzy numbers to represent network weights. Some models use fuzzy logic units to replace perceptrons in the network [151]. Fuzzy systems are also used to train the network parameters of a deep neural network [135]. In this section, we use Pythagorean Fuzzy Deep Boltzmann Machines (PFDBMs) [127], a recent extension of FRBMs, to demonstrate how fuzzy logic can be integrated as a part of deep neural networks, such as deep Boltzmann machines.

4.2.1. Key Formulation of PFRBMs

We will first introduce the building blocks, including deep Boltzmann machines and Pythagorean fuzzy set, and then introduce the architecture design of PFRBMs. A deep Boltzmann machine (DBM) is an extension of the restricted Boltzmann machine [152] and considers multiple hidden layers to capture more complex correlations of the activities of the preceding layers [153]. Considering a L hidden layers DBM whose set of layers is $\{\mathbf{x}, \mathbf{h}_1, \dots, \mathbf{h}_L\}$ where $\mathbf{x}$ is a set of visible units $\mathbf{x} \in \{0, 1\}^D$, and $\mathbf{h}_l$ is l -th hidden layer with a set of hidden units $\mathbf{h}_l \in \{0, 1\}^{p_l}$. DBM is an energy-based probabilistic model which defines a joint probability distribution over $\mathbf{x}$ as

$$P(\mathbf{x}; \theta) = \frac{1}{Z(\theta)} \sum_{\mathbf{h}_1} \dots \sum_{\mathbf{h}_L} e^{-E(\mathbf{x}, \mathbf{h}_1, \dots, \mathbf{h}_L; \theta)} \quad (80)$$

where $\theta = [W_1, \dots, W_L]$ is a vector of the parameters, and $E(\mathbf{x}, \mathbf{h}_1, \dots, \mathbf{h}_L; \theta)$ is the energy function [154] defined as

$$E(\mathbf{x}, \mathbf{h}_1, \dots, \mathbf{h}_L; \theta) = -\mathbf{x}^T W_1 \mathbf{h}_1 - \sum_{l=2}^L \mathbf{h}_{l-1}^T W_l \mathbf{h}_l, \quad (81)$$

and $Z(\theta)$ is the partition function defined as

$$Z(\theta) = \sum_{\mathbf{x}} \sum_{\mathbf{h}_1} \dots \sum_{\mathbf{h}_L} e^{-E(\mathbf{x}, \mathbf{h}_1, \dots, \mathbf{h}_L; \theta)} \quad (82)$$

DBM aims to maximize the joint probability $P(\mathbf{x}; \theta)$, which has the same effect as minimizing energy function $E(\cdot)$.

Pythagorean fuzzy sets (PFS) are an extension of the basic fuzzy sets in two perspectives. First, it introduces a non-membership degree besides the standard membership degree. Second, it considers the restriction that the sum of the squares of the membership degree is between 0 and 1. PFS is defined by the mathematical object:

$$\mathcal{P} = \{ \langle x, \mu_p(x), \nu_p(x) \rangle \mid x \in S \}, \quad (83)$$

where $\mu_p(x) : S \rightarrow [0, 1]$ is the membership degree (how much degree of $x \in S$) of element x to S in $\mathcal{P}$, and $\nu_p(x) : S \rightarrow [0, 1]$

is the non-membership degree (how much degree of $x \notin S$) as well. In addition, we have $\mu_p^2(x) + \nu_p^2(x) \leq 1$, and the hesitant degree, neither membership nor non-membership degree may consider as uncertainty degree. The hesitation degree (uncertainty degree) is the function that expresses a lack of knowledge of whether $x \in S$ or $x \notin S$. It can be calculated by:

$$\pi_p(x) = \sqrt{1 - \mu_p^2(x) - \nu_p^2(x)}. \quad (84)$$

Moreover, to simplify it, $\mathcal{P}(\mu_p(x), \nu_p(x))$ is called a Pythagorean fuzzy number (PFN) denoted $\beta = \mathcal{P}(\mu_\beta, \nu_\beta)$, where $\mu_\beta, \nu_\beta \in [0, 1]$ and $\mu_\beta^2 + \nu_\beta^2 \leq 1$. We can use two metrics to rank a PFN:

$$h(\beta) = \mu_\beta^2 + \nu_\beta^2, \quad s(\beta) = \mu_\beta^2 - \nu_\beta^2, \quad (85)$$

where $h(\beta)$ is the accuracy function of β and $s(\beta)$ is the score function of β . The ranking of two PFNs, $\beta_1 = \mathcal{P}(\mu_{\beta_1}, \nu_{\beta_1})$ and $\beta_2 = \mathcal{P}(\mu_{\beta_2}, \nu_{\beta_2})$, is performed by: (1) If $s(\beta_1) < s(\beta_2)$, then $\beta_1 < \beta_2$; and (2) If $s(\beta_1) = s(\beta_2)$, then (a) if $h(\beta_1) < h(\beta_2)$, then $\beta_1 < \beta_2$; and (b) if $h(\beta_1) = h(\beta_2)$, then $\beta_1 = \beta_2$.

The Pythagorean Fuzzy Restricted Boltzmann Machine (PFRBM) extends the DBM model by replacing the standard real-valued parameters with PFNs. PFRBM can handle fuzzy and/or incomplete data and the fuzzy parameters provide a better representation of the data using fuzzy probability. Figure 13 describes the framework of a PFRBM with L layers, denoted as $h_1, \dots, h_L$. Let $\tilde{\theta} = [\tilde{W}_1, \dots, \tilde{W}_L]$ be the fuzzy parameters and $\mathbf{x} = (x_1, \dots, x_D)$ be the input feature vector. The energy function and probability function of a PFRBM is shown by:

$$\tilde{E}(\mathbf{x}, \mathbf{h}_1, \dots, \mathbf{h}_L; \tilde{\theta}) = -\mathbf{x}^T \tilde{W}_1 \mathbf{h}_1 - \sum_{l=2}^L \mathbf{h}_{l-1}^T \tilde{W}_l \mathbf{h}_l, \quad (86)$$

$$\tilde{P}(\mathbf{x}; \tilde{\theta}) = \frac{1}{\tilde{Z}(\tilde{\theta})} \sum_{\mathbf{h}_1} \dots \sum_{\mathbf{h}_L} e^{-\tilde{E}(\mathbf{x}, \mathbf{h}_1, \dots, \mathbf{h}_L; \tilde{\theta})}. \quad (87)$$

Therefore, we consider the log-likelihood as the objective function,

$$\max_{\tilde{\theta}} \tilde{\mathcal{L}}(\tilde{\theta}, D) = \sum_{\mathbf{x} \in D} \log(\tilde{P}(\mathbf{x}, \tilde{\theta})). \quad (88)$$

As fuzzy optimization problems are intractable, the PFDBM is trained using a combination of gradient descent and metaheuristic techniques.

4.2.2. Causes and Types of Uncertainty

PFDBMs provide a hybrid framework of fuzzy sets and DNNs where uncertainty comes from a fuzzy set due to fuzzy and/or incomplete data, leading to unpredictability. The fuzziness has its root nature in aleatoric uncertainty.

4.2.3. Uncertainty Quantification

Unlike traditional DNNs, PFDBMs with fuzzy parameters can better represent data using a fuzzy probability to represent uncertainty. The fuzzy parameters can learn new features and investigate how much a certain (or uncertain) feature influences the output.

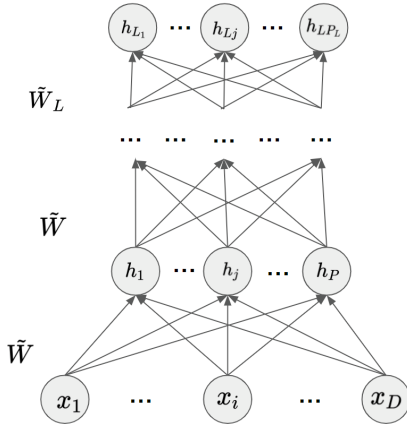


Figure 13: Pythagorean Fuzzy Deep Belief Network (PFDBN).

4.2.4. Applications of FDNNs

PFDBMs were proposed to develop an airline passenger profiling [127] and provide an early warning system for industrial accidents [132]. Besides PFDBMs, Park et al. [151] developed intra- and inter-fraction FDNNs to track lung-cancer tumor motion. Similar to [135], fuzzy logic was employed to train the learning parameters in FDNNs for traffic incident detection. In addition, some models consider fuzzy logic and deep learning in a sequential or parallel fashion. Wang et al. [121] proposed a model that uses a deep neural network with fuzzy feature points for damaged fingerprint classification. Zhang et al. [122] proposed a model utilizing fuzzy granulation and deep belief network for predicting time-series data.

4.3. Rough Deep Neural Networks (RDNNs)

Rough neural networks (RNNs) have been studied for a decade by combining a rough set or rough neuron with DNNs to process the uncertainties and high-dimensional data [155]. The methods fall into two main categories: rough neural-based and rough set-based.

Rough set theory is first introduced in [156, 157] to deal with the problems of inexact, uncertain, or vague knowledge. An information system is defined by the four-tuple $S = \langle U, A, V, f \rangle$ where U is a finite nonempty set called the universe of primitive objects and A is a finite nonempty set of attributes. Each attribute $a \in A$ is associated with a domain set V_a and $V = \bigcup_{a \in A} V_a$. The mapping $f : U \times A \rightarrow V$ is an information function. Assume S is an information system and $M \subseteq A$. Two objects $x, y \in U$ are indiscernible from each other by the set of attributes M in S if and only if for every $a \in M$, $f(x, a) = f(y, a)$. Therefore, every $M \subseteq A$ has a the indiscernibility relation. Rough set theory defines two approximations for any concept set $X \subseteq U$ and attribute set $M \subseteq A$. Using the knowledge of M , X can be approximated by the M -lower approximation $\underline{MX}$ and M -upper approximation $\overline{MX}$:

$$\begin{aligned} \underline{MX} &= \{x | [x]_M \subseteq X\} \\ \overline{MX} &= \{x | [x]_M \cap X \neq \emptyset\}, \end{aligned} \quad (89)$$

where $[x]_M$ is the equivalence classes of the M -indiscernibility relation. The M -boundary region of set X is defined by

$$BND_M(X) = \overline{MX} - \underline{MX} \quad (90)$$

where $\underline{MX}$ is the set of all objects in U , which can be certainly classified as members of X with respect to the set of attributes M . $\overline{MX}$ is the set of objects in U , which can possibly be classified as members of X with respect to the set of attributes M . The boundary region is the set of objects that cannot certainly be classified to X only by employing the set of attributes M . $BND_M(X)$ describes the vagueness of X . If $BND_M(X) = \emptyset$, then X is crisp (exact) with respect to M and if $BND_M(X) \neq \emptyset$, then X is called a rough (inexact) set.

4.4. Regression Evidential Neural Networks (RENNs)

In Section 4, we have discussed Evidential Neural Networks (ENNs) model for classification tasks. Next, we discuss the uncertainty quantification for regression tasks when considering ENNs. Some metrics used in classification cannot apply to regression, such as vacuity and dissonance from Subjective Logic [7, 158]. For regression tasks, we can directly get uncertainty information from the Gaussian process [159]. Like classification, epistemic and aleatoric uncertainties can be applied to regression tasks. Epistemic uncertainty can be calculated by the variance from dropout inference [160], or estimated via an ensemble way [161]. More specifically, Eq. (91) gives an approximation formulation when we use dropout sampling or ensemble.

$$\text{Var}(\mathbf{y}) \approx \frac{1}{T} \sum_{t=1}^T \mathbf{f}(\mathbf{x}; \theta_t)^T \mathbf{f}(\mathbf{x}; \theta_t) (\mathbf{x}_t) - E(\mathbf{y})^T E(\mathbf{y}), \theta_t \sim q(\theta), \quad (91)$$

where $q(\theta)$ is the approximated posterior and $E(\mathbf{y}) \approx \frac{1}{T} \sum_{t=1}^T \mathbf{f}(\mathbf{x}; \theta_t)$ is the predictive mean. For aleatoric uncertainty, Kendall and Gal [162] considered Gaussian likelihood as the objective function,

$$\mathcal{L}_{NN}(\theta) = \frac{1}{N} \sum_{i=1}^N \frac{1}{2\sigma(\mathbf{x}_i; \theta)^2} \|\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i; \theta)\|^2 + \frac{1}{2} \log \sigma(\mathbf{x}_i; \theta)^2 \quad (92)$$

where $\sigma(\mathbf{x}_i; \theta)$ is heteroscedastic (data-dependent) aleatoric uncertainty for each training sample and can be learned from a neural network. In addition, Kendall et al. [163] considered homoscedastic aleatoric uncertainty in a multi-task objective, which is considered as Gaussian likelihood for the two tasks respectively, Eq. (4.4) shows a two-task objective,

$$\begin{aligned} & -\log p(\mathbf{y}_1, \mathbf{y}_2 | \mathbf{f}(\mathbf{x}_i; \theta)) \propto \\ & \frac{1}{2\sigma_1^2} \|\mathbf{y}_1 - \mathbf{f}(\mathbf{x}; \theta)\|^2 + \frac{1}{2\sigma_2^2} \|\mathbf{y}_2 - \mathbf{f}(\mathbf{x}; \theta)\|^2 + \log \sigma_1 \sigma_2, \end{aligned} \quad (93)$$

where σ_1 and σ_2 are the Homoscedastic (task-dependent) aleatoric uncertainty corresponding to each task.

To address the limitation of prior network [23], Malinin et al. [164] proposed a regression prior network that considers Normal-Wishart distribution as the prior to estimating

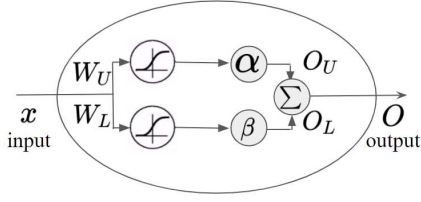


Figure 14: Rough neuron with six tunable parameters

knowledge uncertainty (distributional uncertainty) for regression tasks. Normal-Wishart distribution is similar to the Dirichlet distribution, which is the conjugate prior to the categorical distribution (used in classification task on Prior Networks [23]). The Normal-Wishart distribution is defined by:

$$\mathcal{N}\mathcal{W}(\mu, \Lambda \mid \mathbf{m}, \mathbf{L}, \kappa, \nu) = \mathcal{N}(\mu \mid \mathbf{m}, \kappa\Lambda)\mathcal{W}(\Lambda \mid \mathbf{L}, \nu), \quad (94)$$

where $\mathbf{m}$ and $\mathbf{L}$ are the prior mean and inverse of the positive-definite prior scatter matrix, while κ and ν are the strengths of belief in each prior, respectively. The parameters κ and ν are conceptually similar to the precision of the Dirichlet distribution. Usually, we use a neural network to estimate the parameters of the Normal-Wishart,

$$\begin{aligned} p(\mu, \Lambda \mid \mathbf{x}, \theta) &= \mathcal{N}\mathcal{W}(\mu, \Lambda \mid \mathbf{m}, \mathbf{L}, \kappa, \nu), \\ \{\mathbf{m}, \mathbf{L}, \kappa, \nu\} &= \mathbf{\Omega} = \mathbf{f}(\mathbf{x}; \theta), \end{aligned} \quad (95)$$

where $\mathbf{\Omega} = \{\mathbf{m}, \mathbf{L}, \kappa, \nu\}$ is the parameters of the Normal-Wishart predicted by a neural network. Then we obtain the posterior predictive,

$$\begin{aligned} p(\mathbf{y} \mid \mathbf{x}, \theta) &= \mathbb{E}_{p(\mu, \Lambda \mid \mathbf{x}, \theta)}[p(\mathbf{y} \mid \mu, \Lambda)] \\ &= \mathcal{T}\left(\mathbf{y} \mid \mathbf{m}, \frac{\kappa + 1}{\kappa(\nu - K + 1)}\mathbf{L}^{-1}, \nu - K + 1\right). \end{aligned} \quad (96)$$

And similar to the distribution uncertainty, we calculate the knowledge uncertainty based on mutual information,

$$\begin{aligned} \underbrace{I[\mathbf{y}, \{\mu, \Lambda\}]}_{\text{Knowledge Uncertainty}} &= \underbrace{\mathcal{H}\left[\mathbb{E}_{p(\mu, \Lambda \mid \mathbf{x}, \theta)}[p(\mathbf{y} \mid \mu, \Lambda)]\right]}_{\text{Total Uncertainty}} \\ &\quad - \underbrace{\mathbb{E}_{p(\mu, \Lambda \mid \mathbf{x}, \theta)}[\mathcal{H}[p(\mathbf{y} \mid \mu, \Lambda)]]}_{\text{Expected Data Uncertainty}} \end{aligned} \quad (97)$$

Similar to [164], [165] accomplished uncertainty quantification in the regression model by placing evidential priors (Normal Inverse-Gamma distribution) over the original Gaussian likelihood function and training the neural network to infer the hyperparameters of the evidential distribution. Furthermore, Russell and Reale [166] modeled multivariate uncertainty for regression problems with neural networks, incorporated both aleatoric and epistemic sources of heteroscedastic uncertainty by training a deep uncertainty covariance matrix model directly using a multivariate Gaussian density loss function, or indirectly using end-to-end training through a Kalman filter.

4.4.1. Key Formulation of RDNNs

The rough neural-based method considers a rough neuron in DNNs to improve the robustness of learning. For a traditional

neural network, if the input feature is represented by a range, such as the temperature of climate (e.g., daily maximum and minimum temperature), the neural network cannot learn a good representation, and the prediction error will be relatively large. The neural network based on a rough neuron can address this issue.

Figure 14 shows how the rough neuron is applied for rough pattern recognition. This neuron consists of an upper bound neuron with parameters $\theta_U = \{W_U, b_U, \alpha\}$, and a lower bound neuron with parameters $\theta_L = \{W_L, b_L, \beta\}$. Here W_U and b_U are the weight and bias of the upper bound, respectively, while W_L and b_L are those for the lower bound neuron, respectively. Output coefficients, $0 \leq \alpha$ and $\beta \leq 1$, determine the contribution of upper bound output O_U and lower bound output O_L to the overall neuron's output O . A rough extension of auto-encode, called *rough auto-encoder* (RAE), uses rough neurons in its hidden layer and output layer. Here W_U^k, b_U^k , and α^k are the upper bound parameters of layer k and W_L^k, b_L^k , and β^k are the lower bound parameters of layer k , respectively.

Given the RAE defined with input vector $h_0 = X$, the upper bound and lower bound outputs of the first hidden layer are shown where W_U^1 and W_L^1 are the learned parameters and $f^1 W_U^1 X + b_U^1$ can be larger than $f^1 W_L^1 X + b_L^1$. The $h_U^1(X)$ and $h_L^1(X)$ are defined by:

$$\begin{aligned} h_U^1(X) &= \max \left[f^1(W_U^1 X + b_U^1), f^1(W_L^1 X + b_L^1) \right], \\ h_L^1(X) &= \min \left[f^1(W_U^1 X + b_U^1), f^1(W_L^1 X + b_L^1) \right], \end{aligned} \quad (98)$$

where f^1 is a *sigmoid* function. The latent representation in the hidden layer is computed by:

$$h^1 = \alpha^1 h_U^1 + \beta^1 h_L^1. \quad (99)$$

For the rough decoding process in the output layer, the upper bound and lower bound outputs are computed as:

$$\begin{aligned} h_U^2 &= \max \left[f^2(W_U^2 h^1 + b_U^2), f^2(W_L^2 h^1 + b_L^2) \right], \\ h_L^2 &= \min \left[f^2(W_U^2 h^1 + b_U^2), f^2(W_L^2 h^1 + b_L^2) \right], \end{aligned} \quad (100)$$

where f^2 is considered to be a linear function. Therefore, we have the reconstructed input,

$$r = \alpha^2 h_U^2 + \beta^2 h_L^2. \quad (101)$$

4.4.2. Causes and Types of Uncertainty

In RDNNs, uncertainty is considered in a rough set introducing unpredictability and rough neuron introducing incomplete knowledge. Hence, the rough set and neuron can capture vagueness from model input and parameter uncertainty from model parameters.

4.4.3. Uncertainty Quantification

The uncertainty in RDNNs can be estimated based on the rough set theorem introduced in Section 3.7. The rough set theorem approximates an M -boundary region, which contains a set of objects that cannot be clearly classified by only employing the set of attributes and representing vagueness.

4.4.4. Applications of RDNNs

Most RDNNs are proposed to reduce uncertainty. Zhang and Wang [167] applied the fuzzy-rough neural network in vowel recognition. Khodayar et al. [168] proposed a rough extension of *stacked denoising autoencoder* (SDAE) for ultrashort-term and short-term wind speed forecasting, incorporating a rough neural network into wind uncertainties. Sinusoidal Rough-Neural Network (SR-NN) [169] is proposed to predict wind speed by using rough neurons to handle the high intermittent behavior of wind speed.

In Appendix A, Table A.1 summarizes the notations and their meanings used in each theory. Table A.2 summarizes the key components of belief models and ML/DL models to compare their features. Table A.3 summarizes the advantages and disadvantages of each theory discussed in this work.

5. Summary of the Key Findings

We summarize the key findings from our survey by answering the key research questions below:

RQ1. What are the key causes and types of uncertainty studied in belief theory and deep learning?

Answer: The majority of belief models, such as DST, TBM, IDM, SL, TVL, and Bayesian inference, consider uncertainty caused by a lack of evidence, which is called *vacuity* in SL. It is related to aleatoric uncertainty where a long-term probability can increase as more evidence is received. The second most common uncertainty type considered in belief models, such as DSMT, SL, or Fuzzy Logic, is *discord* (or dissonance), caused by disagreement or conflicting evidence from multiple sources or observers, which generates multiple knowledge frames and results in inconclusiveness in decision making. Lastly, unpredictability is introduced by unclearness or impreciseness of observations or beliefs, which are considered as *fuzziness* in TVL, *vagueness* in SL, and *imprecise beliefs* in DSMT. In DL, two types of uncertainty natures are mainly considered: epistemic uncertainty and aleatoric uncertainty. Epistemic uncertainty, also called ‘model or systematic uncertainty,’ represents the model (parameters) uncertainty due to the limited training data. Aleatoric uncertainty indicates data uncertainty introduced by the nature of randomness in data.

RQ2. What are the key differences between belief theory and deep learning about uncertainty reasoning and quantification?

Answer: We demonstrated the ontology of uncertainty in Figure 2. The source of uncertainty can be from machines, networks, environmental factors, and humans that can generate a lot of various types of uncertain data. Uncertainty has a model of reasoning and quantifying various types of uncertainties to make effective decision-making. We limited the models to belief theory and DL. Uncertainty has procedures to collect evidence, including both subjective and objective data or information. Uncertainty has multiple types including ambiguity and fuzziness which also have been studied under different taxonomies (see Figure 1), such as vagueness, imprecision, unclearness, and so forth. Uncertainty has its root nature in the

most popularly used two types of uncertainty: aleatoric and epistemic uncertainty.

RQ3. How has each belief model considered and measured uncertainty?

Answer: DST’s combination rule helps the decision by combining beliefs from multiple information channels. Imprecise Dirichlet Model (IDM) provides a belief range, rather than a single value, allowing a decision-maker to be aware of the magnitude of uncertainty. In DSMT, Shannon’s entropy and the Probabilistic Information Content (PIC) score are used to indicate uncertainty caused by a lack of evidence where a decision is made based on Generalized pignistic transformation (GPT) and Dezert-Smarandache Theory (DSMT). Bayesian inference theory uses a variance or co-variance to measure uncertainty representing unpredictability. In SL, one can use a projected belief that interprets uncertainty (i.e., vacuity) based on its prior belief (i.e., base rate). If there is very low vacuity but high dissonance, one can maximize vacuity by offsetting the amount of the smallest belief mass while increasing vacuity to have a high effect on new evidence. TVL uses an unknown status to model system uncertainty and defines a set of logical operations to decide the system status for decision-making in system operations. Fuzzy Logic uses fuzzy entropy to quantify the unpredictability and multiple knowledge frames of fuzzy events.

RQ4. How has each belief model been applied in deep learning and vice-versa for effective decision-making under uncertainty?

Answer: TVL is used to solve classification problems in pattern recognition tasks [137] and leveraged to establish a database for natural language consultation [138] in 1970s. We rarely found any recent work using TVL in ML/DL applications. DST is mainly used to fuse data from multi-sensors before conducting neural network training, or fuse predictions from two identically trained models [84, 85, 86]. To our knowledge, TBM is also used to solve classification problems but is not used with ML/DL. DSMT is used along with ML/DL to solve classification problems where it is integrated with SVM, CNN, LSTM, and RF [93, 94]. IDM is used to improve ML algorithms, such as AdaBoost [106], Decision Tree [107], or naïve classifier [108]. Fuzzy Logic is combined with DNNs, named fuzzy DNNs, to deal with ambiguity in data [120, 133, 135, 121, 122]. Bayesian inference is mainly used to infer the posterior distribution of Bayesian features [102, 103, 104]. SL’s vacuity and dissonance uncertainty dimensions are considered in evidential neural networks for uncertainty-aware decision-making in classification problems [5, 63]. Vacuity is used to detect out-of-distribution (OOD) samples while dissonance is used to detect misclassification samples. Rough set theory is combined with DNNs, named RDNNs, to deal with imprecise information and uncertainty in data (e.g., ranges as values for input and/or output variables) [170, 171, 155, 168].

RQ5. What are the key differences between belief theory and deep learning about uncertainty reasoning and quantification?

Answer: Deep learning (DL) has received high attention because of its powerful capability to deal with a large volume of high dimensional data and provide solutions to maximize decision performance. However, as DL is limited in dealing with uncertainty explicitly, it often faces the issue of unexplainability, a well-known issue of explainable AI (XAI), due to its nature of statistical inference. On the other hand, belief models provide rigorous mathematical formulation based on a limited number of parameters which can offer the capability to easily reason and quantify different types of uncertainties. This merit of *quantifiable uncertainty* in belief models can provide reasons to explain a decision made based on mathematical induction. However, belief models suffer from dealing with a large volume of data, which can be complemented by DL. Therefore, our work discussed how a belief model (e.g., SL) has been bridged with DL to improve decision-making capability based on the merits of both approaches to achieve XAI.

RQ6. How can belief model(s) be applied in DL to solve complicated decision-making problems?

Answer: There may be various ways to leverage belief models considered in DL research. One example we discussed in Section 4 is combining SL's opinions with DNNs by constructing evidential NNs (ENNs). That is, ENNs can be built to generate evidence to formulate a subjective opinion in SL, rather than using a common activation function, such as softmax, generating class probabilities. Based on the estimated evidence in ENNs, we can calculate vacuity and dissonance uncertainty by leveraging the operators in SL. Depending on the degree of the quantified uncertainty values, such as vacuity, vagueness, or dissonance, diverse algorithms can be developed for effective decision-making. Based on our prior work [63], we found vacuity is a promising uncertainty type to detect out-of-distribution (OOD) samples while dissonance is an uncertainty type that can effectively detect misclassification samples.

6. Concluding Remarks

6.1. Insights, Lessons Learned, and Limitations

- Recent efforts have been made to estimate different types of uncertainty in our prior work [1] while vacuity and vagueness have been mainly considered in the past [45, 24, 44]. However, reasoning and quantification of other dimensions of uncertainty still remain unaddressed in the literature. Furthermore, the question of how different types of uncertainty can be helpful for effective decision-making has not been addressed in the literature.
- When both the belief masses and a prior belief supporting each belief are the same, decision-making becomes more challenging because it leads to inconclusiveness. Although utility-based belief masses have been studied [172, 24], their contributions are limited to theoretical discussions based on simple examples.
- Most belief/evidence theories and their uncertainty reasoning show high maturity as they have been explored since the 1960s. However, they are mostly theoretical and have not

been thoroughly validated based on real datasets and/or applications for effective decision-making and learning.

- Bayesian theorem and inference methods are the foundations of advanced machine learning and deep learning algorithms. However, the fitting of Bayesian inference from a one-parameter model to a deep learning model with a large volume of parameters can introduce non-trivial challenges in quantifying uncertainties.
- Although belief models have considered various types of uncertainties, as described in Figure 1, the terminologies of those types are often found very similar but their distinctions have not been clarified. Although our survey paper can help readers better understand the diverse types of uncertainties, one may want to argue about our clarification of uncertainty types in Section 2. Much more effort should be made to investigate different types of uncertainties and their effect on diverse decision-making settings and applications.
- Some efforts leveraging both a belief theory and deep learning have been made, such as fuzzy deep neural networks (FDNNs) combining fuzzy logic and DNNs [120], rough deep neural networks (RDNNs) combining rough logic and DNNs [155], and evidential neural networks (ENNs) combining SL and DNNs [7]. However, although belief theories have been explored for several decades and their uncertainty research has matured more than any other field, their applications in reasoning and quantifying uncertainty in DL are still in an infant stage.
- Although we can detect misclassification and out-of-distribution to evaluate the accuracy of predictive uncertainty measured in a DL model, the metrics of predictive uncertainty have not been validated due to the challenges of determining the ground truth of measured uncertainty. To have more valid metrics of predictive uncertainty, we need to develop a way to evaluate the data generation process by considering the causes of uncertainty (e.g., how to generate data with vagueness).
- Uncertainty can be easily introduced by intelligent adversarial attacks performing deceptive poisonous or evasion attacks. Detecting such adversarial attacks with the intent to increase various types of uncertainty should first reduce noises or false information, increasing uncertainty before estimating uncertainty in data for effective decision-making.

6.2. Future Research Directions

- Uncertainty quantification research can be explored more for the studies using qualitative labels. Since belief theories, such as DST, DSMT, or TBM, can provide the capability to fuse qualitative beliefs, their applications in Natural Language Processing (NLP) [173] are promising. Other conventional NLP methods can be compared to DSMT in handling uncertainty for the qualitative beliefs and its effect on application performance.
- Belief models can be incorporated into deep learning models, such as ENNs, FDNNs, and RDNNs. However, software platforms lack accelerating such models compared to traditional deep learning processes. Furthermore, the current

design of such models is mainly limited to particular applications, which lacks the generalizability of the models to other applications. Therefore, developing a generic model based on a balance between simplicity and technical solidity is critical for efficient and effective validation.

- Each belief theory provides a way of estimating different uncertainty types. Hence, when one combines a belief model with a deep learning technique, we can investigate how such different uncertainty quantification methods can introduce different impacts on decision-making performance. For instance, we can propose a novel hybrid framework combining hyper-opinions in SL and deep neural networks to handle vague labels in training data and accurately measure their associated vagueness in vague labels.
- Existing ENN models are difficult to accurately estimate Dirichlet distribution parameters (or evidence) due to the lack of ground truth evidence. Therefore, it is critical to develop a method that can improve the accuracy of the Dirichlet distribution parameters estimation. In addition, most existing ENN models have been applied to classification and regression tasks. We can extend the ENN models to solve other complex tasks, such as time series tasks, sequence data, and multi-label classification.
- We can reduce the complexity of DSMT fusion algorithms by limiting the number of combining sources or ignoring several events in the hyper-power domain. Since the hyper-power domain becomes large with sparse distribution of belief masses, we need to improve the efficiency of DSMT.
- When diverse uncertainty reasoning and quantification in belief models are combined with deep learning, we can investigate how the different ways of measuring uncertainty can impact decision-making performance.
- IDM offers the ability to derive a belief without prior knowledge about a given proposition. In settings that do not allow any prior knowledge or information about the proposition, IDM can allow one to make decisions under uncertainty. In addition, incorporating IDM's probability range into a NN will produce more randomness in results. Particularly when data is insufficient at the beginning of training, introducing the probability range gives a better explorability in reaching optimal solutions.
- To solve sequential decision-making problems, belief models can be combined with deep reinforcement learning. In particular, as IDM does not require having prior knowledge in the decision-making process, it can be easily used for an RL agent to make decisions in the environment with no prior knowledge and to learn an optimal action via trial and error.
- Although many different types of uncertainty have been discussed in belief models, we can capture three main uncertainty types: vacuity caused by a lack of evidence, vagueness (or fuzziness) by failing to capture a singleton belief, and discord (or dissonance) by conflicting evidence. This can be further examined to propose a unified mathematical belief framework to quantify various uncertainty types and belief masses for its broader applicability.
- Most uncertainty measurement approaches are designed for singleton prediction, such as an image or node classifications.

However, they may be unable to extend for time series application because they ignore temporal dependencies in uncertainty quantification. It is critical to develop novel uncertainty metrics considering the temporal dependencies of time series data.

- Most uncertainty estimation research focuses on unstructured tasks, such as classification and regression tasks. Meanwhile, for the structured prediction tasks, such as language modeling (e.g., machine translation and named entity recognition), we can further investigate a general, unsupervised, interpretable uncertainty framework.

Acknowledgement

This work is partly supported by the Army Research Office under Grant Contract Number W91NF-20-2-0140 and NSF under Grant Numbers 2107449, 2107450, and 2107451. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes, notwithstanding any copyright notation herein.

Appendix A. Notations

The notations and their meanings used in each uncertainty reasoning and quantification theory are summarized in Table A.1. Table A.2 compares belief and ML/DL models in their uncertainty consideration and applications in various domains.

Table A.1: NOTATIONS AND THEIR MEANINGS

Notation	Meaning
TVL	
$\mathcal{K}$	Kleene algebra
$\mathcal{K}_3$	Kleenes three-valued logic
I	Information system
IND	The set of equivalence relationships
$X = (\underline{RX}, \overline{RX})$	Rough set
$\mathcal{F}$	The set of all logic functions
$\mathcal{A}_{\mathcal{K}}$	The set of all Kleene algebras
$\mathcal{A}_{RS}$	The collections of all rough sets over all possible information systems
DST	
Θ	A set of all propositions (also known as the <i>frame of discernment (FOD)</i>)
$P(\Theta)$	A power set (also known as the <i>powerset of FOD</i>)
A	Focal element in $P(\Theta)$
m	Belief mass (also known as evidence)
$Bel(A)$	Belief of A
$pl(A)$	Plausibility of A (i.e., $1 - Dis(A)$)
$Dis(A)$	Disbelief of A (i.e., $1 - pl(A)$)
IOU	Interval of Uncertainty
$U^I(A)$	Total uncertainty in DST
TBM (inherits all notations from DST)	
$m_B(A)$	Belief mass supporting propositions A when the conditional evidence supports proposition B
$\overline{B}$	A set not support B
$BetP$	Probability transformed through the pignistic probability function for decision making
X	<i>Boolean algebra</i> of the subset of Ω
DSmT (inherits all notations from DST)	
D^Θ	Hyper-power set of frame Θ
G^Θ	Any set including power set, hyper-power set, and super-power set
$qm(\theta)$	Qualitative beliefs of evidence θ
$m^I(A)$	Imprecise beliefs of evidence A
E_H	Normalized Shannon's entropy
PIC	Probabilistic information content score
GPT	Generalized pignistic transformation
$DSmP_\epsilon$	Probability transformation with a subjective probability measure ϵ
$\mathcal{M}^0(\Theta)$	DST model, the most restricted DSm model
$\mathcal{M}^I(\Theta)$	Free DSm model, without constraints on the elements
$\mathcal{M}(\Theta)$	Hybrid DSm models
L	Set of labels for qualitative beliefs
I_t	Total ignorance, as the union of all hypotheses elements
IDM	
$\Omega = [\omega_1, \dots, \omega_k]$	Set of worlds (truth) / Sample space / Exhaustive set of all possible outcomes
$\overline{P}(A n)$ & $\underline{P}(A n)$	Upper and lower posterior probabilities from IDM, where n is the number of observations towards event A
A	An observed event
N	Total number of observations for all events
s	Positive constant used in IDM
θ	Probability distribution of the observations

TABLE A.1: NOTATIONS AND THEIR MEANINGS (CONT'D)

Notation	Meaning
Bayesian Inference	
$p(\theta)$	Prior distribution of the population characteristics θ
$\mathcal{Y}$	Sample space
$p(y \theta)$	The belief that y is the dataset if θ is true
$p(\theta y)$	Posterior distribution based on the observed dataset y
Θ	Parameter space for the set of all possible parameter values
$\text{Var}^{\text{Bin}}, \text{Var}^{\text{Pois}}, \text{Var}^{\text{MC}}$	Variance of Binomial, Poisson, and Monte Carlo approximation models
τ_n^2	Posterior variance in normal model
$1/\tau_n^2$	Posterior inverse variance or posterior precision
Y	Dependent variables
X	Independent variables
$\text{dbinom}(y, n, \theta)$	Binomial distribution
$\text{dpois}(y, \theta)$	Poisson distribution
$\bar{\theta}$	Mean θ from S number of samples
Fuzzy Logic	
$\mathcal{T}$	The fuzzy logic truth value set
l_τ	Łukasiewicz logic
τ	Fuzzy set
$P(A)$	The probability of a fuzzy event A
$H^f(A)$	The entropy of a fuzzy event A
Subjective Logic	
x	A proposition
b_X	Belief mass distribution over $\mathbb{X}$
$\mathbb{X}$	A domain or a set of propositions where $x \in \mathbb{X}$ or A hypervariable in $\mathcal{R}(\mathbb{X})$
u_X	Uncertainty (vacuity) mass
a_X	Base rate (or prior belief) distribution over $\mathbb{X}$
$\mathbf{P}_X(x)$	Projected probability of belief x
$\mathbf{u}_x^f(x)$	Focal uncertainty
$b_X^{\text{Diss}}(x_i)$	Dissonance per singleton belief
b_X^{TV}	Total vagueness
b_X^{Diss}	Total dissonance
$\mathcal{R}(\mathbb{X})$	Hyper domain
$\mathbf{p}_X^H$	Hyper-probability distribution
α_X	Strength vector over κ number of x 's where $x \in \mathcal{R}(\mathbb{X})$
$\text{Dir}_X^H(\mathbf{p}_X^H; \alpha_X(x))$	Dirichlet hyper-PDF
Evidential Deep Neural Networks	
$\text{Dir}(\cdot)$	Probability density function of Dirichlet distribution
$\text{Cat}(\cdot)$	Probability density function of Categorical distribution
α	Parameter of Dirichlet distribution
$\mathbf{p}$	Class probability
e	Evidence in subjective opinion
b	Belief mass in subjective opinion
u	Uncertainty mass in subjective opinion, i.e., vacuity
$diss$	Dissonance uncertainty

TABLE A.1: NOTATIONS AND THEIR MEANINGS (CONT'D)

Notation	Meaning
Fuzzy Deep Neural Networks	
$\mathbf{x}$	A set of visible units $\mathbf{x} \in \{0, 1\}^D$
$\mathbf{h}$	A set of hidden units $\mathbf{h} \in \{0, 1\}^H$
$P(\mathbf{x}; \theta)$	Joint probability of $\mathbf{x}$
$\mathcal{P}$	Pythagorean fuzzy set (PFS)
$\mu_p(x)$	Membership function of x
$\nu_p(x)$	Non-membership function of x
$\pi_p(x)$	Hesitant function of x
β	$\beta = \mathcal{P}(\mu_\beta, \nu_\beta)$ is Pythagorean fuzzy number (PFN)
$h(\beta)$	Accuracy function of β
$s(\beta)$	Score function of β
Rough Deep Neural Networks	
θ_U	Parameters of upper bound neuron
θ_L	Parameters of lower bound neuron
W_U	Weight of the upper bound neuron
W_L	Weight of the lower bound neuron
b_U	Bias of the upper bound neuron
b_L	Bias of the lower bound neuron
α	Coefficient of the upper bound output
β	Coefficient of the lower bound output
f^1	<i>Sigmoid</i> function
f^2	<i>Linear</i> function

Table A.2: COMPARISON OF BELIEF AND ML/DL MODELS IN THEIR UNCERTAINTY CONSIDERATION AND APPLICATIONS.

Model	Uncertainty type(s)	Uncertainty cause(s)	Uncertainty metric(s)	Application domain(s)	ML/DL techniques	Algorithmic complexity
Kleene's Three-Valued Logic (TVL)	Unpredictability	Lack of information or knowledge	Unknown or unspecified state	Pattern recognition [137], natural language consultation [138], SQLs [139]	Rough DNNs [155]	$O(1)$
Dempster Shafer Theory (DST)	Epistemic uncertainty	Lack of evidence	Plausibility - Belief	Classification tasks [84, 85, 83], intrusion detection [86], fault diagnosis [87]	DBN [174], CNN [175], LSTM-RNN [176], Multilayer NNs [83]	$O(2^n)$
Transferable Belief Model (TBM)	Epistemic uncertainty	Lack of evidence	N/A	Prediction/classification tasks [22, 95, 96, 98, 97] using only TBM	No techniques combined with ML/DL	$O(2^n)$
Dezert-Smarandache Theory (DSmT)	Fusion of precise, imprecise, qualitative beliefs	Conflicting, imprecise, subjective probability	Uncertainty, Shannon's entropy, PIC	Robotics [91], biometrics [91], trust management [91], image fusion [93]	SVM OAA [93], CNN [94], LSTM [94], RF [94]	$O(2^{ \mathcal{P} })$ where $n = \mathcal{P} $
Imprecise Dirichlet Model (IDM)	Epistemic uncertainty	Lack of evidence	Eq. (49)	Classification tasks [106, 107, 108]	IDMBoost [106], Credal Decision Trees [107]	$O(1)$
Fuzzy Logic (FL)	Multiple knowledge frames	Linguistic imprecision or vagueness	Entropy	Traffic control [124, 125], surveillance and security [126, 127], text processing [128, 129], image processing [130], time-series prediction [131]	FDNNs [120, 132], FL-based DL [135, 121, 122, 6, 123]	$O(\mathcal{P})$ where $ \mathcal{P} $ is the cardinality of the fuzzy set
Bayesian Inference (BI)	Aleatoric, Epistemic	Data randomness, lack of data, label overlap	Uncertainty, Shannon's entropy, mutual information	Robotics [102], biometrics [105], face recognition [105], image classification [104]	Expectation-Maximization [104], parameter estimation [103]	$O(N)$ where N is the number of features
Subjective Logic (SL)	Vacuity, vagueness, and dissonance	Lack of evidence, vague observations, conflicting evidence	Belief vacuity, vagueness, dissonance	Adversarial attacks, trust networks	ENNs [5, 63], DNNs [64, 65, 66], DRL [67]	$O(1)$
Machine/Deep Learning	Aleatoric, Epistemic	Limited of training data, data noise, noise of measurement	Shannon's entropy, mutual information, variance, differential entropy	Computer vision [162], natural language processing [150], medical analysis [177], time series forecasting [178]	Monte Carlo-Dropout [179], prior network [23], ensemble method [180]	$O(N)$ where N is the number of Monte Carlo sample times

Table A.3: ADVANTAGES AND DISADVANTAGES OF EACH THEORY DISCUSSED IN OUR WORK

Model	Advantages	Disadvantages
TVL	Easy to implement simple tasks	Limited in quantifying uncertainty
DST	Providing a way of measuring plausibility; providing a combination rule to fuse data from multiple sources	Not applicable to consider multinomial opinions; generating counter-intuitive results using DST's combination rule with information from unreliable sources [46]
TBM	Resolving the issue of DST's combination rule	High complexity leading to a lack of applicability in diverse domains
DSmT	Providing multiple fusion rules and criteria; being applied in broad data combination topics	The specific combination rule should be decided based on the specific problem or topic
FL	Strong theoretical background and wide application	High complexity in modeling high order uncertainties
IDM	Providing a range of a prediction probability, which can be narrowed down with more evidence	possible to generate an invalid prediction range with a lack of evidence [24]
BI	Offering solid computational methods to be applicable in various problem domains incorporating prior distributions	High influence of prior distributions over posterior distributions while selecting good priors is not trivial
SL	Providing various types of uncertainty estimates and diverse operators of combining multiple subjective opinions	Not scalable for high-dimensional data problems
ENNs	Inheriting the merit of SL in quantifying multidimensional uncertainty; more efficient than Bayesian neural networks; possible to estimate uncertainty by only one forward pass	Challenging to train ENNs with a lack of ground truth evidence labels of Dirichlet distribution
FDNNs	Mitigating the sensitivity of DNNs to uncertainties of real-world data	Hard to deal with imprecise data and humans' subjective perception or bias
RDNNs	No need for any prior knowledge for inference; can estimate vagueness directly	Hard to train RDNNs well; difficult to extend RDNNs to other DL models

Appendix B. Taxonomy of Uncertainty

The existing taxonomies of uncertainty are summarized in Table B.1. Figure B.1 shows detailed taxonomies of uncertainty, as an extension of Figure 2.

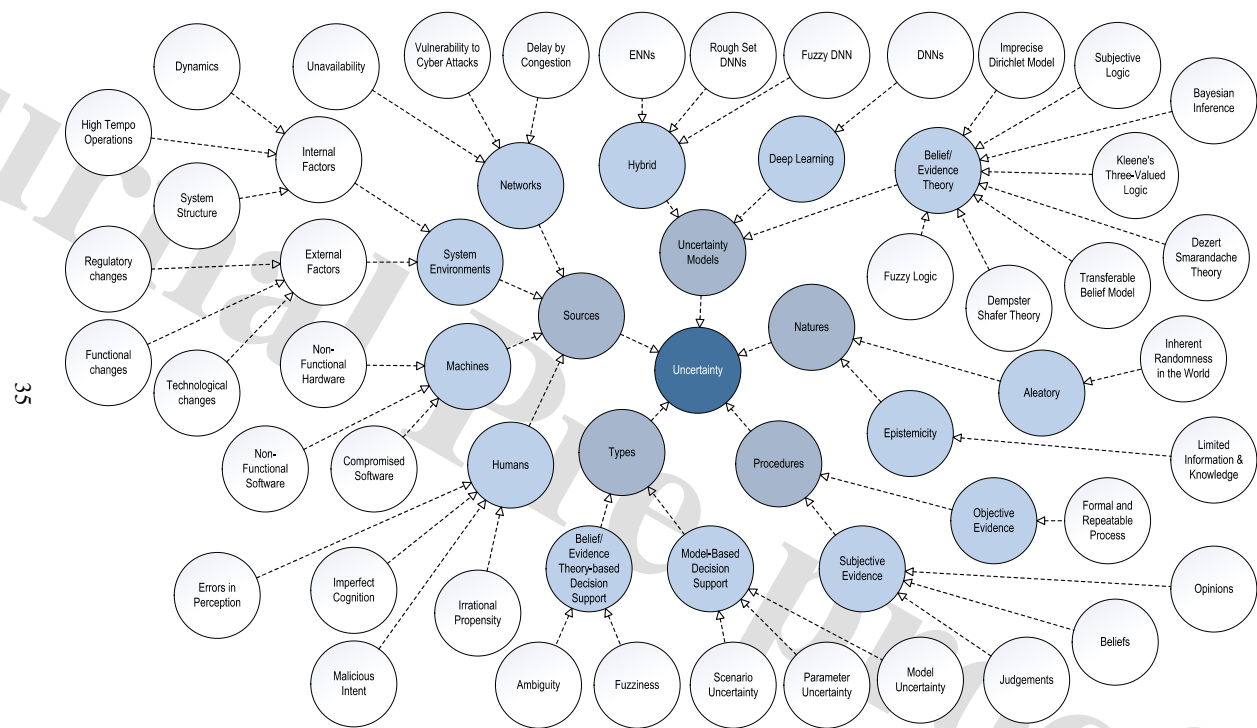


Figure B.1: A Detailed taxonomy of uncertainty, as an extension of Figure 2.

Table B.1: EXISTING TAXONOMIES OF UNCERTAINTY

Taxonomy of Spatio-Temporal Uncertainty [36]				
Intrinsic	Ontological	Caused by a lack of knowledge		
	Epistemic	Caused by imprecision, ignorance, or incompleteness		
Extrinsic	User input	Caused by errors, misinterpretation, entropy, or information truncation		
	Data conversion	Caused by changing technologies		
	Data record	Caused by ambiguity, undecidability, data conversion errors, or users' introduced errors		
Taxonomy of Ethics-based Uncertainty for Decision-Making [181]				
Objective	Epistemological	For knowledge-guided decision		
	Ontological	For quasi-rational decision		
Subjective	Moral	For rule-guided decision		
	Rule	For intuition guided decision		
Taxonomy of Ignorance and Uncertainty [37]				
Error	Delusion			
	Distortion	Inaccuracy (→ Infidelity) Confusion		
Incompleteness	Uncertainty	Stochastic Probability Vagueness (→ Fuzziness or Non-specificity)		
	Absence	N/A		
	Emergence	N/A		
	Undecidable	N/A		
	Uncomputable	N/A		
	Inconceivable	N/A		
	Contradiction	Ambiguity		
	Irrelevance	N/A Taboo		
Integrated Taxonomy of Uncertainty [182]				
Individual Difference Factors	Cognitive processing	N/A		
	Personality	Coping with uncertainty		
	Decision-making bias	N/A		
	Expertise	Experience Knowledge		
Information Flow	Source reliability	Trust		
	Raw Data	Baseline information Conflicting information Ambiguity or missing data Unreliable information		
		Taxonomy of Uncertainty from Different Sources [38]		
		The World	Natural World	Complexity Disorder Partial regularity (e.g., chaos) Stochastic regularity Fundamental dynamism
				Actors
Evidence	Data Measurement Linguistic Evidence Evidence from Actors			
	The Decision Maker			

TABLE B.1: EXISTING TAXONOMIES OF UNCERTAINTY (CONT'D)

Taxonomy of Uncertainty and Variability of Information [39]		
Metrical	Measurement	Empirical observations, interpretation of observations, and interpretation of measurements
Structural	Complexity in models and their validation	System fluctuations, parameter interactions, interpretation of models, or different choices of models
Temporal	Future and past states	Prediction, measurement, retrodiction, interpretation of data, and measurement uncertainty
Translational	Explaining uncertain results (or communications)	Different training and capability of understanding, conflicting goals and values, different views and perspectives in risk estimation

References

- [1] A. Jøsang, J. Cho, F. Chen, Uncertainty characteristics of subjective opinions, in: 2018 21st International Conference on Information Fusion (FUSION), 2018, pp. 1998–2005.
- [2] E. Hüllermeier, W. Waegeman, Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods, *Machine Learning* 110 (2021) 457–506.
- [3] X. Zhao, X. Zhang, W. Cheng, W. Yu, Y. Chen, H. Chen, F. Chen, SEED: Sound event early detection via evidential uncertainty, in: 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022.
- [4] Y. Hu, Y. Ou, X. Zhao, J.-H. Cho, F. Chen, Multidimensional uncertainty-aware evidential neural networks, In *Proceeding of the Thirty-fifth AAAI Conference on Artificial Intelligence* (2020).
- [5] M. Sensoy, L. M. Kaplan, M. Kandemir, Evidential deep learning to quantify classification uncertainty, in: *Advances in Neural Information Processing Systems*, 2018, pp. 3179–3189.
- [6] H. A. Chopade, M. Narvekar, Hybrid auto text summarization using deep neural network and fuzzy logic system, in: 2017 International Conference on Inventive Computing and Informatics (ICICI), 2017, pp. 52–56.
- [7] X. Zhao, F. Chen, S. Hu, J. H. Cho, Uncertainty aware semi-supervised learning on graph data, volume 33, 2020, pp. 12827–12836.
- [8] Y. Li, J. Chen, L. Feng, Dealing with uncertainty: A survey of theories and practices, *Transactions on Knowledge and Data Engineering* 25 (2012) 2463–2482.
- [9] H. D. Kabir, A. Khosravi, M. A. Hosen, S. Nahavandi, Neural network-based uncertainty quantification: A survey of methodologies and applications, *IEEE access* 6 (2018) 36218–36234.
- [10] R. H. Hariri, E. M. Fredericks, K. M. Bowers, Uncertainty in big data analytics: Survey, opportunities, and challenges, *Journal of Big Data* 6 (2019) 1–16.
- [11] M. Abdar, F. Pourpanah, S. Hussain, D. Rezaadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, S. Nahavandi, A review of uncertainty quantification in deep learning: techniques, applications and challenges, *Information Fusion* (2021).
- [12] D. Ulmer, A survey on evidential deep learning for single-pass uncertainty estimation, *arXiv preprint arXiv:2110.03051* (2021).
- [13] J. Gawlikowski, C. R. N. Tassi, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung, R. Roscher, et al., A survey of uncertainty in deep neural networks, *arXiv preprint arXiv:2107.03342* (2021).
- [14] A.-K. Kopetzki, B. Charpentier, D. Zügner, S. Giri, S. Günemann, Evaluating robustness of predictive uncertainty estimation: Are dirichlet-based models reliable?, in: *International Conference on Machine Learning*, PMLR, 2021, pp. 5707–5718.
- [15] V. Bengs, E. Hüllermeier, W. Waegeman, Pitfalls of epistemic uncertainty quantification through loss minimisation, in: *Advances in Neural Information Processing Systems*, 2022.
- [16] J. Pearl, *Probabilistic reasoning in intelligent systems*, volume 88, Elsevier, 2014.
- [17] S. Hanks, D. McDermott, Modeling a dynamic and uncertain world i: Symbolic and probabilistic reasoning about change, *Artificial Intelligence* 66 (1994) 1–55.
- [18] A. D. Kiureghian, O. Ditlevsen, Aleatory or epistemic? does it matter?, *Structural Safety* 31 (2009) 105–112. *Risk Acceptance and Risk Communication*.
- [19] S. Tadelis, *Game theory: an introduction*, Princeton university press, 2013.
- [20] M. Peterson, *An introduction to decision theory*, Cambridge University Press, 2017.
- [21] I. Osband, Z. Wen, M. Asghari, M. Ibrahimi, X. Lu, B. V. Roy, Epistemic neural networks, *arXiv preprint arXiv:2107.08924* (2021).
- [22] Q. Zhang, W. Hu, Z. Liu, J. Tan, TBM performance prediction with bayesian optimization and automated machine learning, *Tunnelling and Underground Space Technology* 103 (2020) 103493.
- [23] A. Malinin, M. Gales, Predictive uncertainty estimation via prior networks, in: *Advances in Neural Information Processing Systems*, 2018, pp. 7047–7058.
- [24] A. Jøsang, *Subjective Logic: A Formalism for Reasoning Under Uncertainty*, Springer Publishing Company, 2016.
- [25] E. Blasch, J. Dezert, B. Pannetier, Overview of dempster-shafer and belief function tracking methods, in: *Proceedings of SPIE*, volume 8745, Baltimore, Maryland, USA, 2013.
- [26] D. Dubois, H. Prade, Representation and combination of uncertainty with belief functions and possibility measures, *Computational Intelligence* 4 (1988) 244–264.
- [27] B. Smith, *Ontology*, in: *The furniture of the world*, Brill, 2012, pp. 47–68.
- [28] D. J. Dubois, *Fuzzy sets and systems: theory and applications*, volume 144, Academic press, 1980.
- [29] I. Linkov, D. Burmistrov, Model uncertainty and choices made by modelers: Lessons learned from the international atomic energy agency model intercomparisons, *Risk Analysis: An International Journal* 23 (2003) 1297–1308.
- [30] W. Walker, P. Harremoës, J. Rotmans, J. V. Der, M. V. Asselt, P. Janssen, M. Kraymer, Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support, *Integrated Assessment* 4 (2003) 5–17.
- [31] M. Brugnach, A. Dewulf, C. Pahl-Wostl, T. Taillieu, Toward a relational concept of uncertainty: about knowing too little, knowing too differently, and accepting not to know, *Ecology and Society* 13 (2008) 30.
- [32] H. Zimmermann, An application-oriented view of modeling uncertainty, *European Journal of Operational Research* 122 (2000) 190–198.
- [33] K. Beven, Facets of uncertainty: epistemic uncertainty, non-stationarity, likelihood, hypothesis testing, and communication, *Hydrological Sciences Journal* 61 (2016) 1652–1665.
- [34] S. A. Zarghami, J. Dumrak, Aleatory uncertainty quantification of project resources and its application to project scheduling, *Reliability Engineering & System Safety* 211 (2021) 107637.
- [35] A. Dewulf, M. Craps, R. Bouwen, T. Taillieu, C. Pahl-Wostl, Integrated management of natural resources: Dealing with ambiguous issues, multiple actors and diverging frames, *Water Science and Technology* 52 (2005) 115–24.
- [36] R. Rocha Souza, A. Dorn, B. Piringer, E. Wandl-Vogt, Towards a taxonomy of uncertainties: Analysing sources of spatio-temporal uncertainty on the example of non-standard german corpora, in: *Informatics*, volume 6, MDPI, 2019, p. 34.
- [37] M. Smithson, *Ignorance and uncertainty: Emerging paradigms*, Springer Science & Business Media, 2012.
- [38] B. E. Lovell, *A taxonomy of types of uncertainty*, Portland State University, 1995.
- [39] W. D. Rowe, Understanding uncertainty, *Risk analysis* 14 (1994) 743–750.
- [40] M. Arja, G. Sauce, B. Souyri, External uncertainty factors and lcc: a

- case study, *Building Research & Information* 37 (2009) 325–334.
- [41] A. Tversky, D. Kahneman, The framing of decisions and the psychology of choice, in: *Environmental Impact Assessment, Technology Assessment, and Risk Analysis*, 1985, pp. 107–129.
 - [42] G. Klir, B. Yuan, *Fuzzy sets and fuzzy logic*, volume 4, Prentice hall New Jersey, 1995.
 - [43] S. Kleene, On notation for ordinal numbers, *The Journal of Symbolic Logic* 3 (1938) 150–155.
 - [44] L.-A. Zadeh, Fuzzy sets, *Information and Control* 8 (1965) 338–353.
 - [45] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton, NJ, 1976.
 - [46] P. Smets, R. Kennes, The transferable belief model, *Artificial Intelligence* 66 (1994) 191–234.
 - [47] A. Jøsang, An algebra for assessing trust in certification chains, in: *Proceedings of Network and Distributed Systems Security (NDSS'99) Symposium*, 1999.
 - [48] A. Jøsang, A logic for uncertain probabilities, *International Journal of Uncertainty, Fuzziness and Knowledge-based Systems* 9 (2001).
 - [49] J. Dezert, F. Smarandache, *Advances and Applications of DSMT for Information Fusion*, American Research Press, Rehoboth, NM, USA, 2004.
 - [50] S. E. Fienberg, When did bayesian inference become “bayesian”?, *Bayesian Analysis* 1 (2006) 1–40.
 - [51] P. Walley, Inferences from multinomial data: learning about a bag of marbles, *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1996) 3–34.
 - [52] J. Wang, H. Sun, Inverse problem in DSMT and its applications in trust management, in: *The First International Symposium on Data, Privacy, and E-Commerce (ISDPE'07)*, Chengdu, China, 2007, pp. 424–428.
 - [53] R. Deepa, S. Swamynathan, *Recent Trends in Computer Networks and Distributed Systems Security Communications in Computer and Information Science*, volume 420, Springer-Verlag Berlin Heidelberg, 2014, p. 115126.
 - [54] L. A. Zadeh, The role of fuzzy logic in the management of uncertainty in expert systems, *Fuzzy Sets and Systems* 11 (1983) 197–198.
 - [55] M. Nagy, M. Vargas-Vera, E. Motta, Multi agent trust for belief combination on the semantic web, in: *4th International Conference on Intelligent Computer Communication and Processing*, 2008, pp. 261–264.
 - [56] M. Lesani, S. Bagheri, Fuzzy trust inference in trust graphs and its application in semantic web social networks, in: *World Automation Congress (WAS 2006)*, 2006, pp. 1–6.
 - [57] W. Chen, Y. Wang, S. Yang, Efficient influence maximization in social networks, in: *Proceedings of the 15th ACM SIGKDD*, New York, NY, USA, 2009, pp. 199–208.
 - [58] H. Liao, Q. Wang, G. Li, A fuzzy logic-based trust model in grid, in: *International Conference on Networks Security, Wireless Communications and Trusted Computing (NSWCTC 2009)*, Wuhan, Hubei, China, 2009, pp. 608–614.
 - [59] J. Luo, X. Liu, Y. Zhang, D. Ye, Z. Xu, Fuzzy trust recommendation based on collaborative filtering for mobile ad-hoc networks, in: *33rd IEEE Conference on Local Computer Networks (LCN 2008)*, 2008, pp. 305–311.
 - [60] D. W. Manchala, Trust metrics, models and protocols for electronic commerce transactions, in: *Proceedings of the 18th IEEE Intl Conf. on Distributed Computing Systems*, Amsterdam, Netherlands, 1998, p. 312321.
 - [61] S. Nefti, F. Meziane, K. Kasiran, A fuzzy trust model for e-commerce, in: *7th IEEE International Conference on E-Commerce Technology*, 2005, pp. 401–404.
 - [62] A. Jøsang, Artificial reasoning with subjective logic, in: *Proceedings of the second Australian workshop on commonsense reasoning*, volume 48, Citeseer, 1997, p. 34.
 - [63] X. Zhao, Y. Ou, L. Kaplan, F. Chen, J. H. Cho, Quantifying classification uncertainty using regularized evidential neural networks, *AAAI 2019 Fall Symposium Series, Artificial Intelligence in Government and Public Sector* (2019).
 - [64] A. Alim, X. Zhao, J. Cho, F. Chen, Uncertainty-aware opinion inference under adversarial attacks, in: *The 2019 IEEE International Conference on Big Data (IEEE Big Data 2019)*, 2019, pp. 6–15.
 - [65] X. Zhao, F. Chen, J. H. Cho, Deep learning based scalable inference of uncertain opinions, in: *2018 International Conference on Data Mining (ICDM)*, 2018, pp. 807–816.
 - [66] X. Zhao, F. Chen, J. H. Cho, Deep learning for predicting dynamic uncertain opinions in network data, in: *2018 International Conference on Big Data (Big Data)*, 2018, pp. 1150–1155.
 - [67] X. Zhao, S. Hu, J. Cho, F. Chen, Uncertainty-based decision making using deep reinforcement learning, in: *2019 22th International Conference on Information Fusion (FUSION)*, IEEE, Ottawa, CA, 2019, pp. 1–8.
 - [68] R. Yager, Entropy and specificity in a mathematical theory of evidence, in: *Classic Works of the Dempster-Shafer Theory of Belief Functions*, 2008, pp. 291–310.
 - [69] J. Kohlas, P.-A. Monney, *A mathematical theory of hints: An approach to the Dempster-Shafer theory of evidence*, volume 425, Springer Science & Business Media, 2013.
 - [70] K. SENTZ, S. FERSON, Combination of evidence in dempster-shafer theory (2002). URL: <https://www.osti.gov/biblio/800792>. doi:10.2172/800792.
 - [71] F. Smarandache, D. Han, A. Martin, Comparative study of contradiction measures in the theory of belief functions, in: *2012 15th International Conference on Information Fusion*, 2012, pp. 271–277.
 - [72] M. van Asselt, Perspectives on uncertainty and risk, *Dordrecht*, 2000, pp. 407–417. doi:10.1007/978-94-017-2583-5_10.
 - [73] L. Zadeh, On the validity of Dempster’s rule of combination of evidence, *Infinite Study*, 1979.
 - [74] R. Yager, On the dempster-shafer framework and new combination rules, *Information sciences* 41 (1987) 93–137.
 - [75] L. Zadeh, A simple view of the dempster-shafer theory of evidence and its implication for the rule of combination, *AI Magazine* 7 (1986) 85–85.
 - [76] C. Murphy, Combining belief functions when evidence conflicts, *Decision Support Systems* 29 (2000) 1–9.
 - [77] C. Osswald, A. Martin, Understanding the large family of dempster-shafer theory’s fusion operators-a decision-based measure, in: *2006 9th International Conference on Information Fusion*, 2006, pp. 1–7.
 - [78] R. Hartley, Transmission of information 1, *Bell System Technical Journal* 7 (1928) 535–563.
 - [79] U. Hohle, Entropy with respect to plausibility measures, in: *Proc. of 12th Int. Symp. on Multiple Valued Logic*, Paris, 1982, 1982.
 - [80] A. Josselme, C. Liu, D. Grenier, É. Bossé, Measuring ambiguity in the evidence theory, *Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 36 (2006) 890–903.
 - [81] P. Smets, Data fusion in the transferable belief model, in: *Proceedings of the 3rd International Conference Information Fusion*, Paris, France, 2000, pp. 21–33.
 - [82] G. Klir, A. Ramer, Uncertainty in the dempster-shafer theory: a critical re-examination, *International Journal of General System* 18 (1990) 155–166.
 - [83] T. Denœux, A neural network classifier based on dempster-shafer theory, *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 30 (2000) 131–150.
 - [84] R. Souza, A. Koesdwiady, F. Karray, Big-data-generated traffic flow prediction using deep learning and dempster-shafer theory, in: *2016 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2016, pp. 3195–3202.
 - [85] Z. Tong, P. Xu, T. Denœux, An evidential classifier based on Dempster-Shafer theory and deep learning, *Neurocomputing* 450 (2021) 275–293.
 - [86] Z. Tian, W. Shi, Z. Tan, J. Qiu, Y. Sun, F. Jiang, Y. Liu, Deep learning and dempster-shafer theory based insider threat detection, *Mobile Networks and Applications* (2020) 1–10.
 - [87] Z. Zhang, W. Jiang, J. Geng, X. Deng, X. Li, Fault diagnosis based on non-negative sparse constrained deep neural networks and dempster-shafer theory, *IEEE Access* 8 (2020) 18182–18195.
 - [88] X. Xu, J. Zheng, J.-b. Yang, D.-I. Xu, Y.-w. Chen, Data classification using evidence reasoning rule, *Knowledge-Based Systems* 116 (2017) 144–151.
 - [89] Q.-H. Le, T. N. Mau, R. Tansuchat, V.-N. Huynh, A multi-criteria collaborative filtering approach using deep learning and dempster-shafer theory for hotel recommendations, *IEEE Access* 10 (2022) 37281–37293.
 - [90] C. Xu, Y. Zhang, W. Zhang, H. Zu, Y. Zhang, W. He, et al., An ensemble learning method based on an evidential reasoning rule considering combination weighting, *Computational Intelligence and Neuroscience* 2022 (2022).
 - [91] F. Smarandache, J. Dezert, *Advances and applications of DSMT for in-*

- formation fusion-Collected works-volume 3, American Research Press, 2009.
- [92] F. Smarandache, Neutrosophic masses & indeterminate models: applications to information fusion, in: 2012 15th International Conference on Information Fusion, IEEE, 2012, pp. 1051–1057.
 - [93] N. Abbas, Y. Chibani, A. Martin, F. Smarandache, The effective use of the DSMT for multi-class classification, *Advances and Applications of DSMT for Information Fusion* (2015) 359.
 - [94] X. Ji, Y. Ren, H. Tang, J. Xiang, Dsmt-based three-layer method using multi-classifier to detect faults in hydraulic systems, *Mechanical Systems and Signal Processing* 153 (2021) 107513.
 - [95] F. Guil, Associative classification based on the transferable belief model, *Knowledge-Based Systems* 182 (2019) 104800.
 - [96] B. Quost, T. T. Denaux, M. Masson, Pairwise classifier combination in the transferable belief model, in: 2005 7th international conference on information fusion, volume 1, 2005, pp. 8–pp.
 - [97] A. H. Henni, R. B. Bachouch, O. Bennis, N. Ramdani, Enhanced multiple binary pir localization using the transferable belief model, *IEEE Sensors Journal* 19 (2019) 8146–8159.
 - [98] J. Honer, H. Hettmann, Motion state classification for automotive lidar based on evidential grid maps and transferable belief model, in: 2018 21st International Conference on Information Fusion (FUSION), IEEE, 2018, pp. 1056–1063.
 - [99] G. Ioannou, P. Louvieris, N. Clewley, A markov multi-phase transferable belief model for cyber situational awareness, *Ieee Access* 7 (2019) 39305–39320.
 - [100] P. D. Hoff, A first course in Bayesian statistical methods, volume 580, Springer, 2009.
 - [101] M. Tipping, Bayesian inference: An introduction to principles and practice in machine learning, in: Summer School on Machine Learning, Springer, 2003, pp. 41–62.
 - [102] B. Sofman, E. Lin, A. Bagnell, J. Cole, N. Vandapel, A. Stentz, Improving robot navigation through self-supervised online learning, *Journal of Field Robotics* 23 (2006) 1059–1075.
 - [103] S. Tripathi, R. Govindaraju, On selection of kernel parameters in relevance vector machines for hydrologic applications, *Stochastic Environmental Research and Risk Assessment* 21 (2007) 747–764.
 - [104] G. Tian, Y. Xia, Y. Zhang, D. Feng, Hybrid genetic and variational expectation-maximization algorithm for gaussian-mixture-model-based brain mr image segmentation, *IEEE Transactions on Information Technology in Biomedicine* 15 (2011) 373–380.
 - [105] H. Wang, D.-Y. Yeung, A survey on bayesian deep learning, *ACM Computing Surveys (CSUR)* 53 (2020) 1–37.
 - [106] L. Utkin, The imprecise Dirichlet model as a basis for a new boosting classification algorithm, *Neurocomputing* 151 (2015) 1374–1383.
 - [107] M. Serafin, M. C. J. C. J. G. A. Joaquín, Imprecise classification with non-parametric predictive inference, in: International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, 2020, pp. 53–66.
 - [108] G. Corani, C. P. de Campos, A tree augmented classifier based on extreme imprecise Dirichlet model, *International Journal of Approximate Reasoning* 51 (2010) 10531068.
 - [109] J. Łukasiewicz, A. Tarski, Untersuchungen über den aussagenkalkül, *CR des seances de la Societe des Sciences et des Letters de Varsovie, cl. III* 23 (1930).
 - [110] L. Zadeh, Fuzzy logic and approximate reasoning, *Synthese* 30 (1975) 407–428.
 - [111] L. Zadeh, The concept of a linguistic variable and its application to approximate reasoning, *Information sciences* 8 (1975) 199–249.
 - [112] L. Zadeh, Probability measures of fuzzy events, *Journal of mathematical analysis and applications* 23 (1968) 421–427.
 - [113] D. Wu, J. Mendel, Uncertainty measures for interval type-2 fuzzy sets, *Information Sciences* 177 (2007) 5378–5393.
 - [114] D. Wu, J. Mendel, A comparative study of ranking methods, similarity measures and uncertainty measures for interval type-2 fuzzy sets, *Information Sciences/Informatics and Computer Science, Intelligent Systems, Applications: An International Journal* 179 (2009) 1169–1192.
 - [115] D. Zhai, J. M. Mendel, Uncertainty measures for general type-2 fuzzy sets, *Information Sciences* 181 (2011) 503–518.
 - [116] R. Viertl, Statistical methods for non-precise data, CRC Press, 1995.
 - [117] A. Laurent, Generating fuzzy summaries: a new approach based on fuzzy multidimensional databases, *Intelligent Data Analysis* 7 (2003) 155–177.
 - [118] H. Bandemer, W. Näther, Fuzzy data analysis, volume 20, Springer Science & Business Media, 2012.
 - [119] P. Diamond, P. Kloeden, Metric spaces of fuzzy sets, *Fuzzy sets and systems* 35 (1990) 241–249.
 - [120] R. Das, S. Sen, U. Maulik, A survey on fuzzy deep neural networks, *ACM Computing Surveys (CSUR)* 53 (2020) 1–25.
 - [121] Y. Wang, Z. Wu, J. Zhang, Damaged fingerprint classification by deep learning with fuzzy feature points, in: 2016 9th international congress on image and signal processing, BioMedical engineering and informatics (CISP-BMEI), IEEE, 2016, pp. 280–285.
 - [122] R. Zhang, F. Shen, J. Zhao, A model with fuzzy granulation and deep belief networks for exchange rate forecasting, in: 2014 International Joint Conference on Neural Networks (IJCNN), IEEE, 2014, pp. 366–373.
 - [123] Y. Deng, Z. Ren, Y. Kong, F. Bao, Q. Dai, A hierarchical fused fuzzy deep neural network for data classification, *IEEE Transactions on Fuzzy Systems* 25 (2016) 1006–1012.
 - [124] W. Chen, J. An, R. Li, L. Fu, G. Xie, M. Z. A. Bhuiyan, K. Li, A novel fuzzy deep-learning approach to traffic flow prediction with uncertain spatial-temporal data features, *Future Generation Computer Systems* 89 (2018) 78–88.
 - [125] Y. Hernandez-Potomkin, M. Saifuzzaman, E. Bert, R. Mena-Yedra, T. Djukic, J. Casas, Unsupervised incident detection model in urban and freeway networks, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2018, pp. 1763–1769.
 - [126] C. P. Chen, C.-Y. Zhang, L. Chen, M. Gan, Fuzzy restricted boltzmann machine for the enhancement of deep learning, *IEEE Transactions on Fuzzy Systems* 23 (2015) 2163–2173.
 - [127] Y. Zheng, W. Sheng, X. Sun, S. Chen, Airline passenger profiling based on fuzzy deep machine learning, *IEEE Transactions on Neural Networks and Learning Systems* 28 (2016) 2911–2923.
 - [128] N. Shirwandkar, S. Kulkarni, Extractive text summarization using deep learning, in: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1–5.
 - [129] T. Nguyen, S. Kavuri, M. Lee, A fuzzy convolutional neural network for text sentiment analysis, *Journal of Intelligent & Fuzzy Systems* 35 (2018) 6025–6034.
 - [130] S. A. Ahmed, D. P. Dogra, S. Kar, P. P. Roy, Surveillance scene representation and trajectory abnormality detection using aggregation of multiple concepts, *Expert Systems with Applications* 101 (2018) 43–55.
 - [131] C. Luo, C. Tan, X. Wang, Y. Zheng, An evolving recurrent interval type-2 intuitionistic fuzzy neural network for online learning and time series prediction, *Applied Soft Computing* 78 (2019) 150–163.
 - [132] Y. Zheng, S. Chen, Y. Xue, J. Xue, A pythagorean-type fuzzy deep denoising autoencoder for industrial accident early warning, *IEEE Transactions on Fuzzy Systems* 25 (2017) 1561–1575.
 - [133] K. J. Holyoak, Parallel distributed processing: explorations in the microstructure of cognition, *Science* 236 (1987) 992–997.
 - [134] R. Yager, Pythagorean membership grades in multicriteria decision making, *IEEE Transactions on Fuzzy Systems* 22 (2013) 958–965.
 - [135] C. El Hatri, J. Boumhidi, Fuzzy deep learning based urban traffic incident detection, *Cognitive Systems Research* 50 (2018) 206–213.
 - [136] A. Kumar, M. Banerjee, Kleene algebras and logic: boolean and rough set representations, 3-valued, rough set and perp semantics, *Studia Logica* 105 (2017) 439–469.
 - [137] S. Kashkevich, V. V. Krasnoproshin, A two-level automated pattern recognition complex, *USSR Computational Mathematics and Mathematical Physics* 19 (1979) 227–239.
 - [138] V. Dahl, Quantification in a three-valued logic for natural language question-answering systems, in: Proceedings of the 6th international joint conference on Artificial intelligence-Volume 1, 1979, pp. 182–187.
 - [139] E. F. Codd, Missing information (applicable and inapplicable) in relational databases, *ACM Sigmod Record* 15 (1986) 53–53.
 - [140] R. K. S. Hankin, A generalization of the Dirichlet distribution, *Journal of Statistical Software* 33 (2010).
 - [141] X. Zhao, F. Chen, S. Hu, J.-H. Cho, Uncertainty aware semi-supervised learning on graph data, *Advances in Neural Information Processing Systems* 33 (2020) 12827–12836.
 - [142] A. Alim, J.-H. Cho, F. Chen, Csl+: Scalable collective subjective logic

- under multidimensional uncertainty, *ACM Trans. Intell. Syst. Technol.* 12 (2020). URL: <https://doi.org/10.1145/3426193>. doi:10.1145/3426193.
- [143] Z. Guo, Q. Zhang, X. An, Q. Zhang, A. Jøsang, L. M. Kaplan, F. Chen, D. H. Jeong, J.-H. Cho, Uncertainty-aware reward-based deep reinforcement learning for intent analysis of social media information, The 1st AAAI Workshop on Uncertainty Reasoning and Quantification in Decision Making (UDM-AAAI-23), co-located with Thirty-Seventh AAAI Conference on Artificial Intelligence (AAAI-23) (2023).
- [144] M. Sensoy, L. Kaplan, F. Cerutti, M. Saleki, Uncertainty-aware deep classifiers using generative models, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 2020, pp. 5620–5627.
- [145] B. Charpentier, D. Zügner, S. Günnemann, Posterior network: Uncertainty estimation without ood samples via density-based pseudo-counts, *Advances in Neural Information Processing Systems* 33 (2020) 1356–1367.
- [146] M. Stadler, B. Charpentier, S. Geisler, D. Zügner, S. Günnemann, Graph posterior network: Bayesian predictive uncertainty for node classification, *Advances in Neural Information Processing Systems* 34 (2021) 18033–18048.
- [147] N. Virani, N. Iyer, Z. Yang, Justification-based reliability in machine learning, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 2020, pp. 6078–6085.
- [148] C. Bhushan, Z. Yang, N. Virani, N. Iyer, Variational encoder-based reliable classification, in: *2020 IEEE International Conference on Image Processing (ICIP)*, IEEE, 2020, pp. 1941–1945.
- [149] W. Shi, X. Zhao, F. F. Chen, Q. Yu, Multifaceted uncertainty estimation for label-efficient deep learning, *Advances in Neural Information Processing Systems* 33 (2020) 17247–17257.
- [150] L. Xu, X. Zhang, X. Zhao, X. Chen, F. Chen, J. D. Choi, Boosting cross-lingual transfer via self-learning with uncertainty estimation, *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2021).
- [151] S. Park, S. J. Lee, E. Weiss, Y. Motai, Intra-and inter-fractional variation prediction of lung tumors using fuzzy deep learning, *IEEE journal of translational engineering in health and medicine* 4 (2016) 1–12.
- [152] N. Zhang, S. Ding, J. Zhang, Y. Xue, An overview on restricted boltzmann machines, *Neurocomputing* 275 (2018) 1186–1199.
- [153] R. Salakhutdinov, H. Larochelle, Efficient learning of deep boltzmann machines, in: *Proceedings of the thirteenth international conference on artificial intelligence and statistics, JMLR Workshop and Conference Proceedings*, 2010, pp. 693–700.
- [154] P. Smolensky, Information processing in dynamical systems: Foundations of harmony theory, Technical Report, Colorado Univ at Boulder Dept of Computer Science, 1986.
- [155] P. Lingras, Rough neural networks, in: *Proc. of the 6th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-based Systems*, 1996, pp. 1445–1450.
- [156] Z. Pawlak, Rough sets, *International Journal of Computer & Information Sciences* 11 (1982) 341–356.
- [157] Z. Pawlak, Vagueness and uncertainty: a rough set perspective, *Computational intelligence* 11 (1995) 227–232.
- [158] A. Josang, J. H. Cho, F. Chen, Uncertainty characteristics of subjective opinions, in: *2018 21st International Conference on Information Fusion (FUSION)*, IEEE, 2018, pp. 1998–2005.
- [159] J. Candela, L. Hansen, Learning with uncertainty-gaussian processes and relevance vector machines, Technical University of Denmark, Copenhagen (2004).
- [160] A. Kendall, V. Badrinarayanan, R. Cipolla, Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding, *arXiv preprint arXiv:1511.02680* (2015).
- [161] B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, *Advances in Neural Information Processing Systems* 30 (2017).
- [162] A. Kendall, Y. Gal, What uncertainties do we need in bayesian deep learning for computer vision?, *Advances in neural information processing systems* 30 (2017).
- [163] A. Kendall, Y. Gal, R. Cipolla, Multi-task learning using uncertainty to weigh losses for scene geometry and semantics, in: *Proceedings of the conference on computer vision and pattern recognition*, 2018, pp. 7482–7491.
- [164] A. Malinin, S. Chervontsev, I. Provilkov, M. Gales, Regression prior networks, *arXiv preprint arXiv:2006.11590* (2020).
- [165] A. Amini, W. Schwarting, A. Soleimany, D. Rus, Deep evidential regression, *Advances in Neural Information Processing Systems* 33 (2020) 14927–14937.
- [166] R. Russell, C. Reale, Multivariate uncertainty in deep learning, *arXiv preprint arXiv:1910.14215* (2019).
- [167] D. Zhang, Y. Wang, Fuzzy-rough neural network and its application to vowel recognition, *Control and Decision* 21 (2006) 221.
- [168] M. Khodayar, O. Kaynak, M. Khodayar, Rough deep neural architecture for short-term wind speed forecasting, *Transactions on Industrial Informatics* 13 (2017) 2770–2779.
- [169] H. Jahangir, M. A. Golkar, F. Alhameli, A. Mazouz, A. Ahmadian, A. Elkamel, Short-term wind speed forecasting framework based on stacked denoising auto-encoders with rough ann, *Sustainable Energy Technologies and Assessments* 38 (2020) 100601.
- [170] D. Zhang, Integrated methods of rough sets and neural network and their applications in pattern recognition, Hunan University, Hunan (2007).
- [171] R. Yasdi, Combining rough sets learning and neural learning-method to deal with uncertain and imprecise information, *Neurocomputing* 7 (1995) 61–84.
- [172] S. Yang, S. Ding, W. Chu, Trustworthy software evaluation using utility based evidence theory, *Jisuanji Yanjiu Yu Fazhan/Computer Research and Development* 46 (2009) 1152–1159.
- [173] J. O’Connor, I. McDermott, NLP, Thorsons, 2001.
- [174] G. E. Hinton, S. Osindero, Y.-W. Teh, A fast learning algorithm for deep belief nets, *Neural computation* 18 (2006) 1527–1554.
- [175] Y. Chen, Convolutional neural network for sentence classification, Master’s thesis, University of Waterloo, 2015.
- [176] A. Graves, Long short-term memory, Supervised sequence labelling with recurrent neural networks (2012) 37–45.
- [177] C. Baumgartner, K. Tezcan, K. Chaitanya, A. Hötter, U. Muehlematter, K. Schawkat, A. Becker, O. Donati, E. Konukoglu, Phiseg: Capturing uncertainty in medical image segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, 2019, pp. 119–127.
- [178] M. Biloš, B. Charpentier, S. Günnemann, Uncertainty on asynchronous time event prediction, *Advances in Neural Information Processing Systems* 32 (2019).
- [179] Y. Gal, Z. Ghahramani, Dropout as a bayesian approximation: Representing model uncertainty in deep learning, in: *International Conference on Machine Learning, PMLR*, 2016, pp. 1050–1059.
- [180] B. Lakshminarayanan, A. Pritzel, C. Blundell, Simple and scalable predictive uncertainty estimation using deep ensembles, *Advances in neural information processing systems* 30 (2017).
- [181] C. Tannert, H.-D. Elvers, B. Jandrig, The ethics of uncertainty: In the light of possible dangers, research becomes a moral duty, *EMBO reports* 8 (2007) 892–896.
- [182] L. G. Shattuck, N. L. Miller, K. E. Kemmerer, Tactical decision making under conditions of uncertainty: An empirical study, in: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 53, SAGE Publications Sage CA: Los Angeles, CA, 2009, pp. 242–246.
- [183] Y. Hu, Y. Ou, X. Zhao, J.-H. Cho, F. Chen, Multidimensional uncertainty-aware evidential neural networks, in: *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 2021, pp. 7815–7822.

- Identifying the causes and types of uncertainty in various belief models
- Uncertainty reasoning and quantification of belief models for decision-making
- Applying uncertainty quantification of belief models into deep learning
- Comparing uncertainty consideration in belief models and deep learning
- Insights, lessons learned, limitations, and future research directions

Since this paper is a survey paper, methodology, software, validation, formal analysis, and data curation are not applicable in this study. Zhen Guo, Zelin Wan, Qisheng Zhang, Xujiang Zhao, and Qi Zhang involved writing -original draft and visualization. In particular, Zhen Guo, Zelin Wan, Qisheng Zhang, and Xujiang Zhao made the same amount of contributions. Lance M. Kaplan and Audun Jøsang made contributions in terms of writing – review and editing and supervision. Dong H. Jeong, Feng Chen, and Jin-Hee Cho involved conceptualization, writing -original draft, writing – review and editing, visualization, funding acquisition, project administration, supervision, and resources. Cho particularly led this effort to instruct and supervise the entire project progress.

Declaration of interests

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Jin-Hee Cho reports financial support was provided by National Science Foundation.

Feng Chen reports financial support was provided by National Science Foundation.

Dong Hyun Jeong reports financial support was provided by National Science Foundation.