

# Uncertainty Aware AI ML: Why and How

**Lance Kaplan**  
ARL, USA

**Federico Cerutti**  
Cardiff University, UK

**Murat Sensoy**  
Ozyegin University, Turkey

**Alun Preece**  
Cardiff University, UK

**Paul Sullivan**  
Intelpoint Inc. USA

## Abstract

This paper argues the need for research to realize uncertainty-aware artificial intelligence and machine learning (AI&ML) systems for decision support by describing a number of motivating scenarios. Furthermore, the paper defines uncertainty-awareness and lays out the challenges along with surveying some promising research directions. A theoretical demonstration illustrates how two emerging uncertainty-aware ML and AI technologies could be integrated and be of value for a route planning operation.

## 1 Introduction

The recent successes of machine learning (ML) (Hinton, Osindero, and Teh 2006; Krizhevsky, Sutskever, and Hinton 2012) have sparked new interest in artificial intelligence (AI)&ML systems (LeCun, Bengio, and Hinton 2015). For the most part, the emphasis is on improving the accuracy of machine learning. However, it is known that ML techniques can only interpolate from the data that trained them. They cannot extrapolate knowledge about a test sample that is far different than the training data. Likewise, AI systems are crafted to reason over a set of concepts manually built by domain experts or learned by examples through a constrained inductive process (Russell and Norvig 2016). Thus, AI systems are unable to extend to high-order reasoning to ‘connect the dots’ the way a human can. However, humans are flawed by imperfect logical reasoning exacerbated by a number of cognitive biases, e.g. confirmation bias, group-think, loss aversion, etc.

It is crucial for AI&ML systems to work seamlessly with human agents to complement the strengths of each while abating their flaws. However, most AI&ML systems infer without any internal evaluation of the quality of the inference relative to how they were trained. This means that the system can drive a human agent, or worse yet an automated controller, to a flawed decision, e.g., driverless automobile accidents (Dickson 2018). In this paper, we argue that in many different scenarios, it is important for AI&ML systems to evaluate uncertainty in the inference based upon the similarity of the situation under test to the training data and the amount of relevant training data. Furthermore, we argue that it is critical for the AI&ML system to communicate its uncertainty along with an explanation of the source of the uncertainty to the human decision maker. We refer

to such AI&ML systems as *uncertainty-aware*. We believe that uncertainty-awareness enables the human to contextually process the inferences for the AI&ML system along with its confidence to either directly decide or to ponder further. Further ponderance includes the human using his/her creative higher-order reasoning abilities to extrapolate new inferences or seek out more information to formulate the decision.

Within the ‘Anticipatory Situational Understanding for Coalitions’ project, we view an automated decision aid as a layered system where lower-level ML methods operating on raw data feed classification results to a higher-level AI probabilistic decision logic (Harborne et al. 2018). Both the lower and higher layers of this AI&ML system should be uncertainty-aware. This paper will review some of our recent work to enable uncertainty-awareness at both layers inspired by subjective logic (Jøsang 2016). The higher layer is a subjective Bayesian network that is able to infer subjective opinions, i.e., uncertain probabilities, of latent variables given uncertain conditional probabilities. The lower layer is a neural network that is specially trained so that the output can be interpreted as a subjective opinion. Then, we show how the two layers could form an uncertainty-aware AI&ML system for a route planning scenario, and how the uncertainty is crucial for the decision maker.

The remainder of the paper is organized as follows. Section 2 enumerates various scenarios that require uncertainty-aware systems, and Section 3 details the desired attributes for uncertainty-awareness and surveys our current progress and future work to achieve uncertainty-awareness at both the ML and AI layers. Section 4 illustrates how our current results combined with explanations of uncertainty are important to the decision maker. Finally, concluding remarks are given in Section 5.

## 2 Motivational Scenario

Our focus is on scenarios where there is a significant open world element, i.e., characterised by prevalence of ‘unknown unknowns’ and ‘unknown knowns’. This kind of scenario is common in government and public sector domains. For example:

- In *securing a large-scale event*, consider a case where a known group is observed (e.g., identified from a photo on

social media or via a image processing of a CCTV feed), but that group (or type of group) has never been seen before at that particular kind of event. There is no known ‘political’ link between the group’s ideology and the event in question making their intent unpredictable using existing predictive models.

- In *narcotics policing*, consider the example of a drug user being discovered under the affects of a known substance but in an unexpected context, e.g., a drug that is in common urban usage associated with (predictable) gang activity, now discovered in use in a rural location. Again, we have key unknown knowns that make prediction and choice of remedial action challenging.
- In *public health*, consider an outbreak of a known disease (detectable by established predictive services) in an entirely unexpected location where this disease has never been seen or studied before, and where its spread cannot be reliably predicted, e.g., a tropical disease in a polar context.

To focus subsequent discussion on a single concrete example, we will consider the case of route planning in a city that has recently been taken under control by a coalition for peace-keeping. Routes might be unsafe depending on the sentiment of local groups. One of the routes goes through a market, and we have a camera there that can detect for violent activities. Furthermore, the sentiment of the local groups affects their attendance at particular public civic centers, which is directly observable by the coalition. While a Bayesian network model can probabilistically connect these observables to the safety of the various routes, there is very little historical data to learn the parameters of the network or to train a neural network to classify specific local activities at the market. This leads to uncertainty, and Section 4 details how uncertainty-aware AI&ML can help the decision maker avoid or prepare for surprises in this scenario.

### 3 Technical Challenge and Solutions

In the era of big data, many assume that the parameters of an AI&ML system can be interpreted as point estimates and the prediction accuracy of the system can generalize as conditions change. However, for the scenarios described above, training data is sparse and the situations are highly dynamic. As a result, the parameters are highly uncertain and possibly changing over time. This uncertainty propagates through the process to infer the beliefs in the state values of the latent variables crucial for decision making. When the conditions that the AI&ML system operates under become vastly different than the training conditions, these uncertainties can be high indicating the inferencing power of the system is currently poor. This section reviews current and suggests future work to understand how to extract these uncertainties.

#### 3.1 Introduction to Subjective Logic

Subjective logic is a formalism to understand reasoning under uncertain probabilistic information (Jøsang 2016). It expands the notion of a probability value to a distribution of possible probabilities. This paper considers variables such

as  $X$  that can take on values  $X = x$  such that  $x$  is in the domain  $\mathbb{X}$  of cardinality  $K = |\mathbb{X}|$ . The value of  $X$  does change over different instantiations, and there is an underlying ground truth value for the probability  $p_X(x)$  of  $X = x$  for all  $x \in \mathbb{X}$ .

A subjective opinion can be formed by directly observing  $N_{ins}$  independent instantiations of  $X$ . If over these instantiations,  $X = x$  for  $n_x$  and assuming an uninformative uniform prior, then the posterior knowledge of the ground truth outcome probability of  $X$  is known to follow the Dirichlet distribution

$$f_{\beta}(p_X|\omega_X) = \frac{1}{\beta(\alpha_X)} \prod_{x \in \mathbb{X}} p_X(x)^{\alpha_x - 1} \quad (1)$$

for  $0 \leq p_X(x) \leq 1$ , where  $\beta(\cdot)$  is the  $K$ -dimensional beta function and the Dirichlet parameters  $\alpha_x = n_x + 1$  for  $x \in \mathbb{X}$  are one particular representation of the opinion  $\omega_X$ . The opinion  $\omega_X$  in belief space is a tuple of beliefs  $b_x = \frac{n_x}{s_X}$  for  $x \in \mathbb{X}$  and uncertainty  $u_X = \frac{K}{s_X}$ , where  $s_X = \sum_{x \in \mathbb{X}} \alpha_x$  is the Dirichlet strength.

It is also convenient to represent the subjective opinion  $\omega_X$  by the mean and Dirichlet strength of the corresponding Dirichlet distribution. The mean represents the projected probability that converts the opinion into the pignistic probabilities, and is given by

$$P_X(x) = \frac{\alpha_x}{s_X} \quad \text{for } x \in \mathbb{X}. \quad (2)$$

The variances for the probabilities  $p_X(x)$ , given by

$$\sigma_X^2(x) = \frac{P_X(x)(1 - P_X(x))}{s_X + 1}, \quad (3)$$

are functions of the projected probabilities and Dirichlet strength of the subjective opinion.

Subjective logic provides a set of rules to infer the effective Dirichlet strength and projected probabilities of latent variables from the logical entailment with variables for which opinions come from direct observations. For instance, subjective opinions naturally extend to subjective conditional opinions, where for example, the opinion for  $X$  conditioned on  $Y$  and  $Z$  is interpreted as the set  $\{\omega_{X|y,z} : y \in \mathbb{Y}, z \in \mathbb{Z}\}$ , and  $\omega_{X|y,z}$  represents the effective number of times that  $X = x$  for  $x \in \mathbb{X}$  when  $Y = y$  and  $Z = z$  while jointly observing  $X$ ,  $Y$ , and  $Z$ . Overall, the efficacy of subjective logic reasoning can be evaluated via simulations to determine 1) how accurate the projected probabilities of the inferred opinions match the ground truth probabilities, and 2) how accurately the uncertainty (or Dirichlet strength) of the inferred opinions represents the ‘spread’ between the projected and ground truth probabilities.

#### 3.2 Subjective Logic Bayesian Network

The Subjective Bayesian network (SBN) was first proposed in (Ivanovska et al. 2015); it is an uncertain Bayesian network where the conditionals are subjective opinions instead of point probabilities. In other words, the conditional probabilities are known within a Dirichlet distribution. A SBN

reflects the knowledge about a Bayesian network when limited historical data is used to learn the conditionals. The inference in SBN leads to an opinion about the marginal probability of all the unobserved variables conditioned on the values of the observed variables. While different types of SBNs were discussed in (Ivanovska et al. 2015), the focus here is on the type that uses the Diriclet distribution interpretation of the subjective opinion to compute uncertainty. This section reviews subjective belief propagation (SBP) which works for singly-connected networks and is described in detail in (Kaplan and Ivanovska 2018) for SBNs with binary variables. We are currently extending the inference engine to accommodate general variables whose domain cardinality  $K > 2$ .

SBP extends the Belief Propagation (BP) inference method of Pearl (Pearl 1986). In BP,  $\pi$ - and  $\lambda$ -messages are passed from parents and children, respectively, to a node, i.e., variable. The node uses these messages to formulate the inferred marginal probability of the corresponding variable. The node also uses these messages to determine the  $\pi$ - and  $\lambda$ -messages to send to its children and parents, respectively. In SBP, the  $\pi$ - and  $\lambda$ -messages are subjective opinions characterized by a projected probability and Dirichlet strength.

The SBP formulation approximates output messages as beta-distributed random variables using the methods of moments and a first-order Taylor series approximation to determine the mean and variance of the output messages in light of the beta-distributed input messages. The details of the derivations are provided in (Kaplan and Ivanovska 2018). Given a node  $X$  with  $m$  parents  $U_i$  for  $i = 1, \dots, m$ , the subjective opinions of the  $\pi$ -messages sent to  $X$  are characterized by the projected probabilities  $\pi_{U_i, X}(x)$  and Dirichlet strengths  $s_{\pi_{U_i, X}}$ . Likewise, given that  $X$  has  $k$  children  $Y_j$  for  $j = 1, \dots, k$ , the subjective opinions of the  $\lambda$ -messages sent to  $X$  are characterised by the projected probabilities  $\lambda_{U_i, X}(x)$  and Dirichlet strengths  $s_{\lambda_{U_i, X}}$ .

In the end, SBP computes opinions for each latent variable  $X$  conditioned on the state values for all the observed variables. In (Kaplan and Ivanovska 2018), it is shown that the projected probabilities of the inferred opinions accurately reflect the probabilities that could be inferred if the ground truth conditional probabilities were known. More importantly for uncertainty-awareness, the uncertainty of the inferred opinions faithfully characterized the accuracy of the projected probabilities. Specifically, empirical results demonstrated that the divergence between the desired and actual confidence bound significance is low. Current and future work is expanding the inference to efficiently work for general (non-binary) variables under general (possibly network structures).

### 3.3 Uncertainty-Aware Neural Network

Uncertainty-aware machine learning methods should express confidence based upon the amount of training data of each class that is similar to the data sample under test. Figure 1 illustrates this concept by visualizing the training and testing data in a feature space. In this example, test samples ‘1’ and ‘2’ should have high beliefs in the ‘x’ and ‘o’ classes, respectively, along with low uncertainty as they are

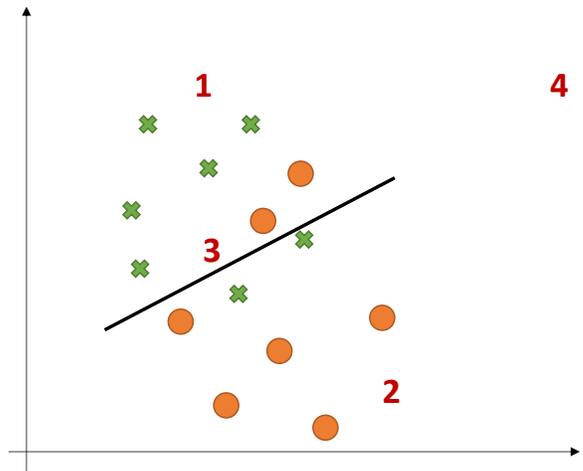


Figure 1: Hypothetical feature space of training data for two classes and four test data samples.

nearby many training samples of the same class. On the other hand, test sample ‘3’ should have beliefs around 0.5 for both the ‘x’ and ‘o’ classes along with low uncertainty as it lies near an equal and large number of training data from the two classes. In other words, the sample lies near a decision boundary. Finally, test sample ‘4’ is far from any of the training data and should have small beliefs for the two classes along with high uncertainty. Overall, the output opinion from an uncertainty-aware machine learner should represent the amount of nearby training data of the various classes.

It is a challenge to efficiently train a machine learning method to output subjective opinions. Part of the challenge is to understand how to tabulate distances in the feature space of the training data into class evidences. Furthermore, one must determine a proper feature space, which is most likely not the raw data space, that reduces the effects of noise, invariants, and irrelevant dimensions. Finally, the machine learner must regress the subjective opinion from its model parameters without explicitly storing all of the training data.

Evidential Deep Learning (EDL) aims to extend classical deep learning with the ideas from evidential reasoning and subjective logic (Dempster 2008; Jøsang 2016) to quantify uncertainty in classification tasks (Sensoy, Kandemir, and Kaplan 2018). As mentioned before, subjective logic uses Dirichlet distributions to represent subjective opinions, which encodes belief and uncertainty. For a sample, EDL tries to learn parameters of a predictive posterior as a Dirichlet density function for the classification of the sample. This is done by replacing the *softmax* layer in a classical deep classifier with an activation function that produces only non-negative outputs (e.g., *relu*, *softplus* etc.). The resulting output  $e_X = [e_1, \dots, e_n]$  is considered as the evidence for the classification of the sample over  $n$  class labels. From the calculated evidence, parameters of the corresponding Dirichlet density is calculated as  $\alpha_X = e_X + 1$ . Hence, when the

total evidence is zero, the resulting Dirichlet density corresponds to the uniform distribution, i.e., a Dirichlet density whose all parameters are one. It is mapped to a subjective opinion with zero belief and total uncertainty.

The Dirichlet distribution can be considered as generative process for the categorical distribution over possible class labels, i.e.,  $\mathbf{p}_X$ . Hence, we can calculate the expected prediction failure for the sample using the following:

$$\int \|\mathbf{y}_X - \mathbf{p}_X\|_2^2 \frac{1}{\beta(\boldsymbol{\alpha}_X)} \prod_{x \in \mathbb{X}} p_X(x)^{\alpha_x - 1} d\mathbf{p}_X \quad (4)$$

where  $\mathbf{y}_X$  is the one-hot representation of the class label (i.e., ground truth) and  $\|\mathbf{y}_X - \mathbf{p}_X\|_2^2$  is the sum square error between the truth and the categorical distribution  $\mathbf{p}_X$ . Hence, the equation calculates the expected sum square error over possible values of  $\mathbf{p}_X$  using the predicted Dirichlet density function.

The loss function consists of the expected sum square error regularized by a term that measures the similarity of the Dirichlet distribution minus the ground truth evidence to the uniform distribution. The regularization term helps to reduce confidence, i.e., increase uncertainty, in cases of conflict. These cases occur in regions of sparse training data, and the regularization term conditions EDL to recognize such cases.

In a set of experiments performed over well-known benchmarks, EDL significantly outperforms state-of-the-art Bayesian Neural Network approaches for uncertainty estimation. Furthermore, it is more robust to adversarial attacks, since it generates less evidence and increases uncertainty when encountered with adversarial examples.

### 3.4 Future Extensions

Many other AI&ML approaches can be extended to accommodate uncertainty-awareness beyond Bayesian networks and neural networks. There is a need to incorporate first-order logics into the higher layer reasoning, and it is natural to think about a Markov logic network (MLN) (Richardson and Domingos 2006) where the rule weights are learned within a distribution. However, MLNs are computationally inefficient and perhaps other probabilistic logic reasoning systems such as probabilistic soft logic (Bach et al. 2017) or ProbLog (De Raedt, Kimmig, and Toivonen 2007) should be considered first. In fact, collective subjective logic has been proposed as uncertainty-aware extension of probabilistic soft logic (Chen, Wang, and Cho 2017). Beyond other reasoning approaches, it is also important to understand how to isolate the observations contributing to the uncertainty for each decision variable and how to disseminate the reasoning chain from such observations to the decision variables to the decision maker.

In the near future, we plan to investigate the uncertainty-aware extension of ProbLog. ProbLog (De Raedt, Kimmig, and Toivonen 2007; Fierens et al. 2015)<sup>1</sup> belongs

<sup>1</sup>More information on ProbLog, including an open source implementation and an interactive online tutorial, can be found at <https://dtai.cs.kuleuven.be/problog/>.

to a family of probabilistic logic programming (PLP) languages (De Raedt and Kimmig 2015) following Sato’s distribution semantics (Sato 1995). It extends logic programming by annotating some ground facts with their probability of being true, which generalizes a single program into a distribution over programs that share their rules, but differ in their databases. More specifically, a ProbLog program consists of two parts, a set  $F$  of ground probabilistic facts  $p : : f$  where  $p$  is a probability and  $f$  a ground atom, and a set  $R$  of rules  $h : - b_1, \dots, b_n$  where  $h$  is a logical atom and the  $b_i$  are literals.<sup>2</sup> While the semantics is defined for countably infinite sets of probabilistic facts, see (Sato 1995) for details, we restrict the discussion to the finite case in the following. ProbLog considers the ground probabilistic facts as independent random variables, i.e., we obtain the following probability distribution  $P_F$  over truth value assignments to sets of ground facts  $F' \subseteq F$ :

$$P_F(F') = \prod_{f_i \in F'} p_i \cdot \prod_{f_i \in F \setminus F'} (1 - p_i) \quad (5)$$

As each logic program obtained by choosing a truth value for every probabilistic fact has a unique least Herbrand model,  $P_F$  can be used to define the *success probability*  $P(q)$  of a query  $q$ , that is, the probability that  $q$  is true in a randomly chosen such program, as the sum over all programs that entail  $q$ :

$$P(q) := \sum_{\substack{F' \subseteq F \\ \exists \theta F' \cup R \models q\theta}} P_F(F') \quad (6)$$

$$= \sum_{\substack{F' \subseteq F \\ \exists \theta F' \cup R \models q\theta}} \prod_{f_i \in F'} p_i \cdot \prod_{f_i \in F \setminus F'} (1 - p_i). \quad (7)$$

Inference in ProbLog is concerned with computing marginal probabilities of queries, i.e., ground atoms, under this distribution, potentially conditioned on a conjunction of evidence atoms. While this is a #P-hard problem in general, ProbLog relies on state-of-the-art knowledge compilation techniques to achieve scalable inference across a wide range of models.

The parameters of ProbLog programs can be learned from partial interpretations (Fierens et al. 2015), and ProbLog rules defining a target predicate can be learned from a ProbLog program specifying background knowledge (in the form of facts and/or known rules for other predicates) and ground atoms using the target predicate annotated with target probabilities (De Raedt et al. 2015).

Inspired by subjective logic, we plan to learn the parameters of the ProbLog programs such as the rule probabilities as beta distributions. Furthermore, we plan to update the inference engine to accommodate these distribution to

<sup>2</sup>For the semantics of ProbLog to be well-defined, the set of rules has to have a two-valued well-founded model for each subset of the probabilistic facts: a sufficient condition for this is for programs to be stratified, i.e., have no loops through negation. See (De Raedt and Kimmig 2015; Fierens et al. 2015) for further details.

output subjective opinions about success probability of the query. This will enable a systematic study of the strengths and weakness of the subjective ProbLog against subjective Bayesian networks.

#### 4 Theoretical Demonstration

To demonstrate the need for uncertainty-aware AI&ML, we consider a route planning operation within a purposely simplified peace keeping scenario to highlight the salient value of uncertainty-awareness. In this scenario, the GoodGuys are trying to stabilize a city. Consider that a commander of the GoodGuys needs to transport some medical supplies from headquarters to a local hospital. There are three routes through a city populated by two factions, the Capulets and the Montagues. These factions have tumultuous relationships with each other and the GoodGuys. One route (Route A) goes through the center of the Capulet territory, and another route (Route C) goes through the center of the Montague territory. The remaining route (Route B) skirts the border of Capulet and Montague territory and passes by the central marketplace.

There is a surveillance camera at the marketplace that is connected to a neural network that classifies between normal and violent activity. The commander also has information about the number of people who visit the community centers every day within the Capulet and Montague territories. Given these observations, the commander wants to decide what route to use and the level of armored escorts needed to protect the transport operation. The commander is using the Bayesian network model illustrated in Figure 2 to predict the route conditions. This model indicates that the daily dispositions of each faction towards the GoodGuys influences the attendance at the civic centers, the marketplace activity, and the safety of the three routes.

The actual conditional probabilities of the Bayesian network are not known to the GoodGuys. Instead, they lean a subjective Bayesian network using human intelligence to obtain 100 daily instantiations of the dispositions of the two factions, along with reports of route conditions and civic center attendance. The ground truth conditional probabilities for the Bayesian network is given in Table 1.<sup>3</sup> The table indicates that the Capulets usually have a favorable disposition towards the Goodguys, in contrast to the Montagues. Whenever either faction has a negative disposition, attendance at the corresponding civic center is higher than normal. It is also more likely the faction will disrupt any Goodguys traveling through their territory. Violent activities in the marketplace are only likely to occur when both factions have negative dispositions, and only then is it likely the factions would disrupt GoodGuys traveling along Route B. The commander’s knowledge of all these trends is based on the Subjective Bayesian network obtained from the 100 instantiations.

<sup>3</sup>In this example, the variables are binary simply because the current subjective belief propagation method works only for the binary case. We are currently extending the propagation method to accommodate more general multinomial variables.

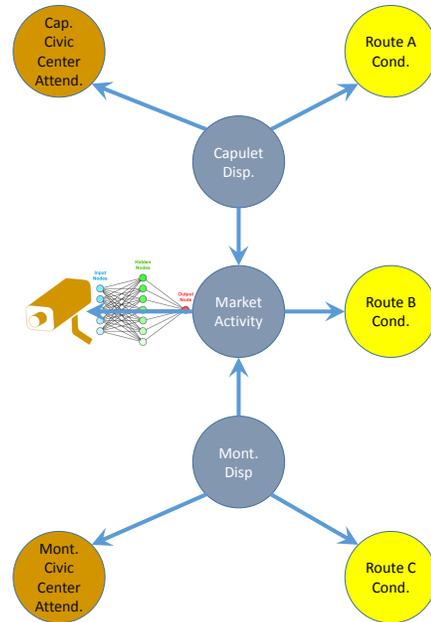


Figure 2: Bayesian network model describing the probabilistic interactions of the various variables that influences how to infer route safety in light of the observations including the civic center engagement and marketplace activity likelihood via the surveillance camera.

Table 2 provides the inferred opinions of the route conditions from the subjective Bayesian network for each of the three routes for various observations. In the absence of any observations, the uncertainty of all these opinions is relatively low and reflective of 100 effective pieces of evidence. The opinions for routes A and B are comparable with a safety and danger belief of about 0.78 and 0.20, respectively, and these beliefs are flipped for the Route C opinion. These opinions are consistent with general trends from the ground truth conditional probabilities.

When the camera is detecting normal activity at the marketplace, and the attendance at the civic centers in Capulet and Montague sections are normal and high, respectively, the route conditions opinions demonstrate low uncertainty. The uncertainty does increase slightly over the previous case of no observation because the observations do constrain the number of effective training instantiations that are appropriate for the inference. Since these observations are likely to occur, the uncertainty remains relatively small, and observations increases the beliefs that routes A and B are safe and Route C is dangerous. With these observations, the commander can choose Route A or B with limited armor support.

When the attendance at the Montague civic centers goes

Variable	States	Conditional Probabilities
Capulet Disposition	{pos, neg}	$P(CD = \text{pos}) = 0.9$
Montague Disposition	{pos, neg}	$P(MD = \text{pos}) = 0.1$
Civic Center Attendance	{norm, high}	$P(CCA = \text{norm}   CD = \text{pos}) = 0.8$
		$P(CCA = \text{norm}   CD = \text{neg}) = 0.1$
Marketplace Activity	{norm, violent}	$P(MCA = \text{norm}   MD = \text{pos}) = 0.8$
		$P(MCA = \text{norm}   MD = \text{neg}) = 0.1$
Route Condition	{safe, danger}	$P(RA = \text{safe}   CD = \text{pos}) = 0.9$
		$P(RA = \text{safe}   CD = \text{neg}) = 0.1$
		$P(RB = \text{safe}   MA = \text{norm}) = 0.9$
		$P(RB = \text{safe}   MA = \text{violent}) = 0.1$
		$P(RC = \text{safe}   MD = \text{pos}) = 0.9$
		$P(RC = \text{safe}   MD = \text{neg}) = 0.1$

Table 1: Ground truth conditional probabilities for the Bayesian network in Figure 2.

Observations	Route A			Route B			Route C		
	$b_{\text{safe}}$	$b_{\text{dang}}$	$u$	$b_{\text{safe}}$	$b_{\text{dang}}$	$u$	$b_{\text{safe}}$	$b_{\text{dang}}$	$u$
none	0.78	0.20	0.02	0.77	0.21	0.02	0.21	0.77	0.02
$CCA = \text{norm}$ $MCA = \text{high}$ $\omega_{MA}^{\text{cam}} = (0.95, 0, 0.05)$	0.92	0.06	0.02	0.91	0.07	0.02	0.13	0.84	0.03
$CCA = \text{norm}$ $MCA = \text{norm}$ $\omega_{MA}^{\text{cam}} = (0.95, 0, 0.05)$	0.90	0.08	0.02	0.91	0.07	0.02	0.54	0.37	<b>0.09</b>
$CCA = \text{norm}$ $MCA = \text{norm}$ $\omega_{MA}^{\text{cam}} = (0, 0.95, 0.05)$	0.66	0.15	<b>0.19</b>	0.02	0.50	<b>0.48</b>	0.54	0.37	<b>0.09</b>
$CCA = \text{norm}$ $MCA = \text{norm}$ $\omega_{MA}^{\text{cam}} = (0, 0, 1)$	0.88	0.09	0.03	0.79	0.08	<b>0.13</b>	0.55	0.36	<b>0.09</b>

Table 2: Opinions for the route conditions obtained via subjective belief propagation in light of various observations. Note that camera likelihood opinion  $\omega_{MA}^{\text{cam}} = (b_{\text{norm}}, b_{\text{violence}}, u)$  is the backpropagated message to the marketplace activity variable from the camera data.

normal, then uncertainty about Route C increase while the opinions of routes-A and B are unaffected. This is due to the fact that it is unlikely for the Montagues to have a positive disposition towards the GoodGuys, which reduces the amount of effective training samples to determine the condition or Route C. Again the commander can choose Route A or B with limited armor support.

Then when the camera has high confidence of violence at the marketplace, the uncertainty for the opinions of routes A and-B increases drastically. This is due the fact that it is extremely unlikely for violence to occur in the marketplace when attendance at both civic centers are at normal levels. There are very few training instantiations to help determine these opinions. On the other hand, the opinion for Route C has not changed as it still reflects the oddity of normal attendance at the Montague civic center. In this case, the commander would need to pick either Route A or Route C where

the probability of safety of Route A is higher but at more than twice the uncertainty. In either case the commander may want heavier armor support.

Finally, the neural network reports that it is uncertain about the activity class. This primary affects the uncertainty about Route B. Furthermore, resolution of this activity classification has tremendous influence over the subjective opinions of routes A and B as can be seen in the third and fourth rows of Table 2. Therefore, the AI&ML system alerts the commander that it is uncertain due to its processing of the surveillance camera data. The commander directs his team to analyze the video footage. The team determines that the people in the marketplace have synchronized in a dance routine that the neural network was never trained to recognize. The commander digs further with some experts to learn that the dance is actually an ancient ritual that two tribes in the region perform to synchronize the mind and body when the

tribes are preparing to attack a common exogenous enemy. In light of this additional context, the commander decides to provide heavy armed support on both the ground and air to escort the supply truck.

The scenario illustrates how understanding uncertainty is important for the commander to assess risk. The uncertainty can arise because the observations are due to rare events for which sparse training means few exemplars to establish opinions for the states of the decision variables. The uncertainty can also arise when the neural networks cannot interpret the input data. This case can represent a ‘black swan’ event that the neural network was never trained to understand. The synchronized dance is an example of an unknown unknown that requires human investigation.

## 5 Conclusions

AI&ML systems have shown great promise as decision making aids. However, such systems are not impervious to making mistakes. Many such mistakes occur because the systems are operating in situations that they were not trained to understand. Instead of forcing a decision, it is important for such a system to be able to characterize its uncertainty relative to the similarity of the current observations to the training data, and to explain the uncertainty to a human decision maker. This way, the human can trust the AI&ML system to crunch through the large piles of data at a rate faster than the human and alert the human when the system realizes that it no longer can provide quality inferences. By isolating the uncertainty, the system can then explain the source and nature of the uncertainty to the user. This enables the human to focus his/her reasoning strengths to the special cases that need extra attention. This paper highlights some recent work in subjective Bayesian network and evidential neural networks to eventually achieve this uncertainty-awareness. Some other future directions and challenges are discussed. We believe that the development of uncertainty-aware AI&ML systems will become an important avenue of research for the community.

## Acknowledgement

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

## References

- [Bach et al. 2017] Bach, S. H.; Broecheler, M.; Huang, B.; and Getoor, L. 2017. Hinge-loss Markov random fields and probabilistic soft logic. *Journal of Machine Learning Research* 18:1–67.
- [Chen, Wang, and Cho 2017] Chen, F.; Wang, C.; and Cho, J.-H. 2017. Collective subjective logic: Scalable uncertainty-based opinion inference. In *2017 IEEE International Conference on Big Data (Big Data)*, 7–16.
- [De Raedt and Kimmig 2015] De Raedt, L., and Kimmig, A. 2015. Probabilistic (logic) programming concepts. *Machine Learning* 100(1):5–47.
- [De Raedt et al. 2015] De Raedt, L.; Dries, A.; Thon, I.; Van den Broeck, G.; and Verbeke, M. 2015. Inducing probabilistic relational rules from probabilistic examples. In *Proceedings of 24th International Joint Conference on Artificial Intelligence*, 1835–1842.
- [De Raedt, Kimmig, and Toivonen 2007] De Raedt, L.; Kimmig, A.; and Toivonen, H. 2007. ProbLog: A probabilistic Prolog and its application in link discovery. In *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 2462–2467.
- [Dempster 2008] Dempster, A. P. 2008. A generalization of Bayesian inference. In *Classic works of the Dempster-Shafer theory of belief functions*. Springer. 73–104.
- [Dickson 2018] Dickson, B. 2018. Uber’s self-driving car accident: Did AI fail us? *PC Magazine*.
- [Fierens et al. 2015] Fierens, D.; Van den Broeck, G.; Renkens, J.; Shterionov, D.; Gutmann, B.; Thon, I.; Janssens, G.; and De Raedt, L. 2015. Inference and learning in probabilistic logic programs using weighted Boolean formulas. *Theory and Practice of Logic Programming* 15(03):358–401.
- [Harborne et al. 2018] Harborne, D.; Raghavendra, R.; Willis, C.; Chakraborty, S.; Dewan, P.; Srivatsa, M.; Tomsett, R.; and Preece, A. 2018. Reasoning and learning services for coalition situational understanding. In *Proc. of the SPIE*, volume 10635, 106350V.
- [Hinton, Osindero, and Teh 2006] Hinton, G. E.; Osindero, S.; and Teh, Y.-W. 2006. A fast learning algorithm for deep belief nets. *Neural computation* 18(7):1527–1554.
- [Ivanovska et al. 2015] Ivanovska, M.; Jøsang, A.; Kaplan, L.; and Sambo, F. 2015. Subjective networks: Perspectives and challenges. In *Proc. of the 4th International Workshop on Graph Structures for Knowledge Representation and Reasoning*, 107–124.
- [Jøsang 2016] Jøsang, A. 2016. *Subjective Logic: A Formalism for Reasoning Under Uncertainty*. Springer.
- [Kaplan and Ivanovska 2018] Kaplan, L., and Ivanovska, M. 2018. Efficient belief propagation in second-order Bayesian networks for singly-connected graphs. *International Journal of Approximate Reasoning* 93:132–152.
- [Krizhevsky, Sutskever, and Hinton 2012] Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 1097–1105.
- [LeCun, Bengio, and Hinton 2015] LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep learning. *Nature* 521(7553):436–444.
- [Pearl 1986] Pearl, J. 1986. Fusion, propagation, and struc-

turing in belief networks. *Artificial Intelligence* 29(3):241–288.

[Richardson and Domingos 2006] Richardson, M., and Domingos, P. 2006. Markov logic networks. *Machine Learning* 62(1-2):107–136.

[Russell and Norvig 2016] Russell, S. J., and Norvig, P. 2016. *Artificial Intelligence: A Modern Approach*. Pearson Education Limited.

[Sato 1995] Sato, T. 1995. A statistical learning method for logic programs with distribution semantics. In *Proceedings of the 12th International Conference on Logic Programming (ICLP-95)*.

[Sensoy, Kandemir, and Kaplan 2018] Sensoy, M.; Kandemir, M.; and Kaplan, L. 2018. Evidential deep learning to quantify classification uncertainty. <https://arxiv.org/abs/1806.01768>.