Uncertainty in Modeling and Reasoning About Beliefs

Ph.D. Comprehensive Exam

Mohammad Shayganfar - mshayganfar@wpi.edu

May, 26 2015

1 Introduction to Uncertainty in AI

In many practical problem-solving applications, the available knowledge is incomplete or inexact. In such cases, the available knowledge is inadequate to support logical reasonings. Humans use some techniques such as generalization and approximation when they are confronted with uncertain information about decisions they should make. Although these techniques are subject to error, humans still use them. Most of the time the sources of uncertainty, e,g., missing or noisy data, or incomplete knowledge of a causality, are unpredictable or even unavoidable. Therefore, when we design intelligent machines, they should be able to operate in uncertain and sometimes ambiguous environments, and yet provide the most acceptable results for us according to the given conditions. The reasoning methods under uncertainty allow AI systems to use uncertain knowledge and make decisions while minimizing risk [36].

There are some major theories of uncertainty in reasoning about beliefs in AI. In this response, first, we provide three of the most prominent theories including *Bayesian Belief Networks* (probabilistic reasoning), *Dempster-Shafer* theory (evidential reasoning), and *Fuzzy Logic* (reasoning under ambiguity) – see Sections 2.1, 2.2, and 2.3 respectively. There are other approaches such as rule-based systems which emerged from early work on practical and intuitive systems for logical inference [28]. However, we do not review these approaches due to their relatively low level of importance in the reviewed literature and time. This document continues by providing the advantages and disadvantages of the three major theories we mentioned above (see Section 3). Then, we briefly provide applications of the Bayesian Networks for robots and autonomous agents (see Section 4). This document ends with our conclusion about the theories of reasoning about beliefs under uncertainty (see Section 5).

2 Theories of Uncertainty

The major theories of uncertainty in AI each seek to address different aspects of uncertainty. The underlying principles of each of these theories are outlined in the following sections.

2.1 Bayesian Belief Networks

A *Bayesian Belief Network* [26] is a directed acyclic graph consisting of nodes and edges which provides a graphical model for reasoning under uncertainty. Each node in the network represents a random variable from the domain. The state of each node is called *belief*, which based on the prior evidence reflects the posterior probability distribution of the other values associated with that node. Each node also has an associated *Conditional Probability Table* (CPT) which represents the conditional probability of the variable given the value of its parents in the graph. Each individual edge between two variables represents the relation or conditional dependence between those two variables. Also, the explicit directions represented by arrows as directional edges indicate the notion of causality in the network (see Figure 1). They are always drawn from cause nodes to effect nodes, indicating dependencies between variables [9]. Assuming discrete variables, the strength of the relationship between variables is quantified by conditional probability distributions associated with each node.

Constructing a belief network can be divided into two different subtasks: a) specifying the causal structure among the existing variables in the network, and b) specifying the prior and conditional probabilities for these variables.

2.1.1 Network's Structure

The structure, or topology, of the network captures qualitative relationships between variables (see Figure 1). The first step in building the Bayesian network's structure is to determine a) what are the nodes/variables to represent in the structure, and b) what are their possible values? For instance, nodes with discrete values can have boolean (to represent that a proposition is true or false), ordered (e.g., enumeration), and integral values (e.g., height of a person). Then, one should determine the existing causality between nodes, i.e., to determine which node (parent) influences the other (child) and connect them through directed edges [20].

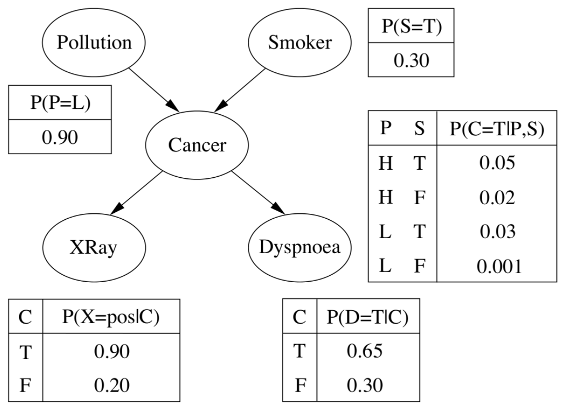

Figure 1: A Belief Network for a lung cancer problem [20].

2.1.2 Conditional Probability Table

As we mentioned earlier, after specifying the structure of a Bayesian Network, the next step is to quantify the relationships between connected nodes by specifying the conditional probability distribution for each node. These conditional probability distributions appear as *Conditional Probability Tables* (CPT) if we consider discrete variables in the structure. To calculate values in CPTs, for each node, we need to think about all possible combinations of values of parent nodes. Each row in a CPT will contain the value of a conditional probability of a node for each case of the possible combination of values for the parent node. Clearly, a node that has many parents or a node with the parents taking a large number of values, can cause very large CPTs. The size of the CPT is exponentially related to the number of parents. For instance, if the nodes of a network are boolean, a variable with n parents requires a CPT with 2^{n+1} probabilities. The probabilities in a CPT are typically acquired from experts on the subject, but they can also be learned automatically using machine learning approaches. Figure 1 shows variables, their relations, and associated CPTs for diagnosis of a lung cancer problem taken from [20].

2.1.3 Markov Property

In Bayesian networks, each variable is independent of its non-descendants given its parent variables. Therefore, there are no direct dependencies in the system being modeled other than those already explicitly shown via edges.

Meaning, there is no hidden connection between variables. This is called *Markov property* in a Bayesian network. If Bayesian networks do not adhere to Markov property, there will be redundant edges that connect independent variables together. Consequently, the network will not represent a minimal model.

2.1.4 Joint Probability Distribution

In many applications of probability, there are more than one random variable to be measured over the same sample space (e.g., existence of multiple causes for a lung cancer). A Bayesian network provides a complete description of the domain. Once we identify random variables and their probabilistic relationships, the values in a joint probability distribution can then be obtained from the probabilities relating the random variables. Therefore, all the entries in the full joint probability distribution can be calculated from the information in the network. There is also a fundamental assumption that in the underlying structure of the problem being modeled by a Bayesian network, not every single node is connected to every other one [20]. Therefore, if there is such a problem structure, then a Bayesian network can provide a compact representation of a model for that problem. In the following formula $P(x_1, x_2, \ldots, x_n)$ is an abbreviation for the conjunction of $\boldsymbol{n}$ assignments to each variable. Hence, the following formula gives the value of each variable:

$$P(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} P(x_i|parents(X_i))$$

where $parents(X_i)$ denotes the specific values of the variables in $Parents(X_i)$. As we see, given Markov property, the product of only the appropriate elements (parent nodes) of the CPTs in the network represents the value of each individual entry in the joint probability distribution. The following provides an example based on the network provided in Figure 1.

$P(X = pos \wedge D = true \wedge C = false \wedge P = high \wedge S = true)$

$= P(X = pos|D = true, C = false, P = high, S = true)$
$\times P(D = true|C = false, P = high, S = true)$
$\times P(C = false|P = high, S = true) \times P(P = high|S = true) \times P(S = true)$

$= P(X = pos|C = false) \times P(D = true|C = false) \times P(C = false|P = high, S = true) \times P(P = high) \times P(S = true)$

2.1.5 Reasoning in Bayesian Networks

Reasoning in Bayesian networks is the process of updating beliefs in the face of evidence. In other words, it is the process of efficiently deducing the belief distribution over a particular subset of random variables given that we know the states of some other variables in the network. Bayesian networks can be conditioned upon any subset of their variables, supporting any direction of reasoning. Figure 2 shows four different types of reasoning using the network shown in Figure 1. These four types are reasoning are [20]:

Diagnostic reasoning – This is the reasoning from symptoms (effects) to cause. For instance, a doctor updates her belief about a patient's cancer when she checks the X-ray results.

Predictive reasoning – This is the reasoning based on new information about the causes to new beliefs about the corresponding effects. For instance, if the patient tells his doctor the information about the polluted area he lives in, the doctor's belief about the patient having cancer increases, even without assessing the patient's symptoms.

Intercausal reasoning – This is the reasoning about the mutual causes of a common effect. For instance, suppose that there are two different causes for lung cancer, smoking and pollution (see Figure 1). Initially these two causes are independent of each other; i.e., the patient smoking or not, does not change the probability of the patient being subject to pollution. However, as soon as the patient is diagnosed with cancer, the probability of smoking or living in a polluted area increases. Now, if the doctor discovers that her patient is a smoker, then the probability of him living in polluted area decreases. Therefore, the presence of one explanatory cause for the cancer lowers the probability of the alternative cause, even though they both were independent causes. In other words, the first explanatory cause *explains away* the alternative one.

Combined reasoning – Sometimes the reasoning does not fit into one of the explained types. Thus, any of these reasoning types can be combined to solve a problem.

2.1.6 Conditional Independence

Bayesian networks which satisfy the Markov property (see Section 2.1.3) explicitly express conditional independencies in probability distributions [20]. Therefore, since a Bayesian Network is based on a joint probability distribution of a set of random variables, knowledge about the conditional indepen-

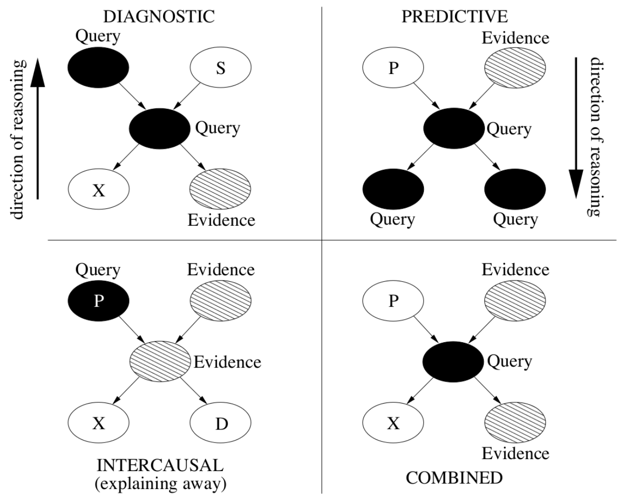

Figure 2: Types of Reasoning [20].

dence of these random variables is important for understanding reasoning based on conditional probabilities.

Two random variables A and B are *conditionally independent* given another variable C, if $p(A,B|C) = p(A|C).p(B|C)$, therefore:

$$p(A|B,C) = \frac{p(A,B|C)}{p(B|C)} = \frac{p(A|C).p(B|C)}{p(B|C)} = p(A|C)$$

And similarly, $p(B|A,C) = p(B|C)$. Figure 3(a) shows a *causal chain* between A and B and C. For instance, being a smoker can cause lung cancer which causes shortness of breath, in our example. This kind of causal chains can cause a conditional independence which can be described as: $P(C|A,B) = P(C|B)$. This means that if one already knows that C has occurred, knowing that A occurred doesnt make a difference to one's beliefs about C. Figure 3(b) shows that both variables A and C have a *common cause* called B. For instance, based on our example, lung cancer is a common cause for a positive x-ray and dyspnoea in the patient. This kind of common causes can also cause a conditional independence which, again, can be described as: $P(C|A,B) = P(C|B)$. This means that if one already knows about B, then an additional information that A provides, will not give more information about the chances of C. Figure 3(c) shows that one variable has two causes. *Common effect* produces the opposite conditional independence to that of common causes and causal chains. This means that parents are independent until the common effect provides new information. This kind of common effects can cause a conditional dependence which can be described as: $P(A|B,C) \neq P(A|B)$. In other words, if one knows about B (the effect), then finds out that for example A (one of two causes) is absent, this increases the probability of C (alternative cause).

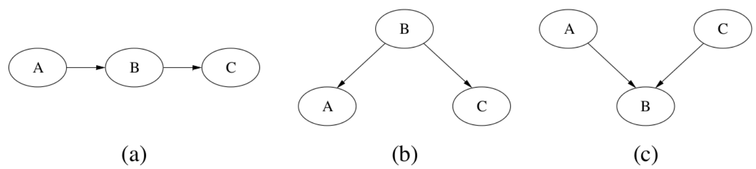

Figure 3: (a) causal chain, (b) common cause, and (c) common effect [20].

2.1.7 d-Separation

The concepts of conditional dependencies and independencies, discussed above, can apply not only between pairs of nodes, but also between sets of nodes. In general, it is possible to determine whether two sets of nodes X and Y are independent, if there is a set of evidence nodes E, given the Markov property. If the two sets of nodes X and Y are *d-separated* (directional-dependent separation) by an evidence set of nodes E, then (given the Markov property) the two sets of nodes X and Y are conditionally independent given E. d-separation is a topological criterion for Bayesian networks [28]. Figure 4 shows how the evidence set of nodes E is blocking the two sets of nodes X and Y in three different conditions. In a graph, a path is blocked given a set of nodes E, if there is a node Z on the path for which at least one of the three conditions discussed in Section 2.1.6 holds.

Analogous to our example, based on this definition the pollution and smoking variables are d-separated from x-ray and dyspnoea (blocking condition 1), x-ray is d-separated from dyspnoea (blocking condition 2), and if cancer and x-ray or dyspnoea are not observed, then smoking variable would have been d-separated from pollution (blocking condition 3).

2.2 Dempster-Shafer Theory

In [12], Dempster proposed a probabilistic framework based on lower and upper bounds on probabilities. In [31], Shafer developed a formalism for reasoning under uncertainty which uses some of Dempster's mathematical expressions with a different interpretation. Based on Shafer's formalism, each piece of evidence may support a subset containing several hypotheses.

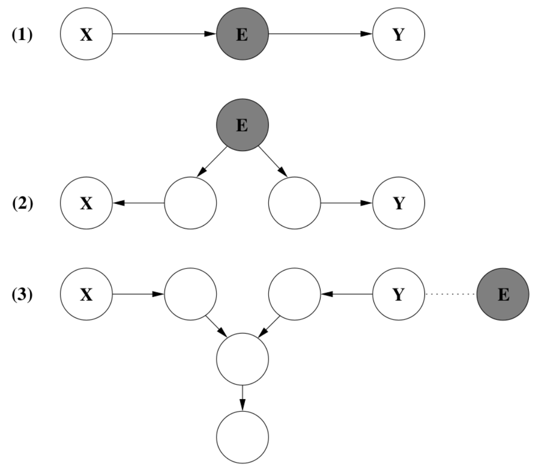

Figure 4: Three types of situations in which the path from X to Y can be blocked, given evidence E. In each case, X and Y are d-separated by E [20].

This is a generalization of the pure probabilistic framework in which every finding corresponds to a value of a variable (a single hypothesis) [13]. Therefore, Dempster-Shafer theory is the generalization of the Bayesian theory of subjective probability to combine accumulative evidence or to change prior opinions in the light of new evidence [9]. Dempster-Shafer theory is designed to deal with the distinction between uncertainty and ignorance. Rather than computing the probability of a proposition, it computes the probability that the evidence supports the proposition [28], and it does not require the assumption that $\mathit{Belief}(\mathrm{A}) + \mathit{Belief}(\neg\mathrm{A}) = 1$. Dempster-Shafer theory deals with the possible values of an unknown variable, just as the theory of probability does [36].

There are three basic functions in the Dempster-Shafer theory that we need to understand for modeling purposes, *mass function, belief function*, and *plausibility function*. Let $\Theta = \{h_1, h_2, \ldots, h_n\}$ be a finite set of hypotheses. This set of hypotheses is also called *frame of discernment*. The hypotheses represent all of the possible states of the system considered. The set of all subsets of Θ is its *power set*: 2^Θ. A subset of these 2^Θ sets may consist of a single hypothesis or of a conjunction of several hypotheses (e.g., a snowy day and a dry day). The pieces of evidence are events that occurred or may occur (e.g., high pressure shown by a barometer, or low temperature). One piece of evidence can be related to a single hypothesis or a set of hypotheses. However, it is not allowed to have different pieces of evidence lead to the same hypothesis or set of hypotheses. In fact, the relation between a piece of evidence and a hypothesis corresponds to a cause-effect chain, i.e., a piece of evidence implies a hypothesis or a set of hypotheses [19]. Moreover, it is required that all hypotheses are unique, not overlapping and mutually exclusive.

2.2.1 Mass Function

A *Basic Probability Assignment* (BPA) or *mass function* is a function $m : 2^{\Theta} \rightarrow [0,1]$ such that:

$$m(\emptyset) = 0, \text{ and } \sum_{x \in 2^{\Theta}} m(x) = 1.$$

The value 0 indicates no belief and the value 1 indicates total belief, and any value between these two indicate partial belief. As you see the mass function uses the notion of 2^{Θ} to be able to use all possible subsets of the *frame of discernment* Θ. All of the assigned probabilities sum to unity. There is no belief in an empty set. Any subset x of the frame of discernment Θ for which m(x) is non-zero is called a *focal element* and represents the exact belief in the proposition depicted by x. Thus, any subset is proposition and vice versa. Other elements in Dempster-Shafer theory are defined by mass function.

2.2.2 Belief Function

Now, we can define another important notion in Dempster-Shafer theory, the *belief function* (sometimes called a *support function*). It is the measure of total belief committed to $A \subseteq \Theta$ that can be obtained by simply adding up the mass of all the subsets of A. In other words, given the frame of discernment Θ and $A \subseteq \Theta$, the belief in A, denoted $Belief(A)$, is a number in the interval [0, 1]. Belief in a set of elements, say A, of a frame Θ, represents the total belief that one has based on the evidence obtained. Unlike probability theory, $Belief(A) = 0$ represents lack of evidence about A, while $p(\mathrm{A}) = 0$ represents the impossibility of A. However, $Belief(A) = 1$ represents certainty, that is A is certain to occur, similar to $p(\mathrm{A}) = 1$, which also represents the certainty that A is true. A belief function defined on a space Θ must satisfy the following three properties:

$$Belief(\emptyset) = 0$$

$$Belief(\Theta) = 1$$

$$Belief(A_1 \cup \ldots A_n) \geq \sum_i Belief(A_i) - \sum_{i<j} Belief(A_i \cap A_j) + \ldots + (-1)^{n+1} Belief(A_i \cap \ldots \cap A_n)$$

A belief function is a function $Belief : 2^{\Theta} \rightarrow [0,1]$ and is defined by:

$$Belief(A) = \sum_{B \subseteq A} m(B) \qquad \text{for all } A \subseteq \Theta$$

2.2.3 Plausibility Function

Plausibility in a set, say A of a frame Θ consisting of a mutually exclusive and exhaustive set of elements, represents the maximum possibility that a set A is true given all the evidences. A plausibility function *Plausible* defined on a space Θ must satisfy the following three properties:

$$Plausible(\emptyset) = 0$$

$$Plausible(\Theta) = 1$$

$$Plausible(A_1 \cap \ldots A_n) \leq \sum_i Plausible(A_i) - \sum_{i<j} Plausible(A_i \cup A_j) + \ldots + (-1)^{n+1} Plausible(A_i \cup \ldots \cup A_n)$$

A *plausibility* measure is a function $Plausible : 2^{\Theta} \rightarrow [0,1]$, and is defined by:

$$Plausible(A) = \sum_{B \cap A \neq \emptyset} m(B) \qquad \text{for all } A \subseteq \Theta$$

$Plausible(A)$ in a subset A is defined to be the sum of all mass functions for the subsets B that have non-zero intersections with A, and it represents the extent to which we fail to disbelieve A. In other words, it corresponds to the total belief that does not contradict A. The plausibility and belief functions are related to one another, and we can represent this relation as:

$$Belief(A) = 1 - Plausible(\neg A) \qquad \text{and} \qquad Plausible(A) = 1 - Belief(\neg A),$$

where $\neg A$ is A's complement. Also, $Belief(\neg A)$ is often called the *doubt* in A. It is noteworthy to mention that Dempster-Shafer theory allows the representation of *ignorance* since $Belief(A) = 0$ does not imply $Belief(\neg A) > 0$ even though $Belief(\neg A) = 1$ implies $Belief(A) = 0$. Other notable relations are:

$Belief(A) + Belief(\neg A) \leq 1$, and

$Plausible(A) + Plausible(\neg A) \geq 1$.

Here, we also note that in the case of each of the focal elements being singletons, we return back to traditional Bayesian analysis incorporating normal probability theory, since in this case $Belief(A) = Plausible(A)$ [5].

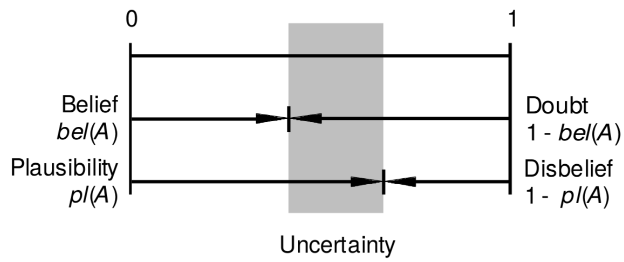

Figure 5: Measures of belief and plausibility. The uncertainty interval is shaded gray. [19].

Collectively the above measures provide Dempster-Shafer theory with an explicit measure of ignorance about A and its complement. All the above measures of confidence and the BPA are equivalent, in the sense that each of them can be expressed as a function of any one of the rest. The *uncertainty* measure is defined as the length of the interval $[Belief(A), Plausible(A)]$ where $Belief(A) \leqslant Plausible(A)$ [40], and it is also called *belief interval*. Figure 5 illustrates a graphical representation of the belief, plausibility, and doubt measures which we defined above. As it is shown and said earlier, the difference between plausibility and belief describes the evidential interval range which represents the uncertainty concerning the set A. Also, as we see in Figure 5, lack of belief does not imply disbelief, since the complements of belief and plausibility are doubt and disbelief, respectively. Furthermore, the mass assigned to Θ can be interpreted as the global ignorance, since the level of mass value is not discernible among the hypotheses.

2.2.4 Dempster's Rule of Combination

Suppose that we have two pieces of uncertain evidence relevant to the same frame of discernment Θ. Dempster-Shafer theory also provides a method to combine the measures of evidence from different sources, using Dempster's rule of combination which combines two pieces of evidence into a single new piece. The rule assumes that the sources are independent. If m_1 and m_2 are the BPA's associated with Bel_1 and Bel_2 respectively and Bel_1 and Bel_2 are independent, then Dempster's rule of combination is as follows:

$$[m_1 \oplus m_2](y) = \begin{cases} 0, & y = \emptyset \\ \dfrac{\sum\limits_{A \cap B = y} m_1(A)m_2(B)}{1 - \sum\limits_{A \cap B \neq \emptyset} m_1(A)m_2(B)}, & y \neq \emptyset \end{cases}$$

The numerator, i.e., $\sum_{A \cap B = y} m_1(A)m_2(B)$, represents the accumulated evidence for the sets A and B, which supports the given hypothesis y. The denominator in the Dempster's rule of combination, i.e., $1 - \sum_{A \cap B \neq \emptyset} m_1(A)m_2(B)$, is an important normalization factor denoted by $\mathcal{K}$ which can be interpreted as a measure of conflict between the sources [33].

2.3 Fuzzy Logic Theory

Fuzzy Logic, introduced by Zadeh in 1965 [42], provides a mathematical framework to capture uncertainty. Fuzziness manipulates uncertainty by dealing with the boundaries of a set that are not clearly defined. Fuzzy Logic is a multivalued logic, that allows intermediate values to be defined between conventional evaluations like "true" and "false". Fuzzy Logic's ultimate goal is to provide foundations for approximate reasoning using imprecise propositions based on fuzzy set theory. In order to deal with such imprecise inference, Fuzzy Logic allows the imprecise linguistic terms such as: fuzzy predicates (e.g., old, expensive), fuzzy quantifiers (e.g., many, little), and fuzzy truth values (e.g., unlikely false or unlikely true). Fuzzy Logic is a method for reasoning with logical expressions describing membership in fuzzy sets [28]. Logic as a base for reasoning can be essentially distinguished by three items: truth values, operators, and reasoning procedures (e.g., tautologies) [44]. For instance, in dual logic, truth values can be "true" (1) or "false" (0), operators can be defined using the truth tables, and modus ponens or contrapositions can be considered as tautology. In Fuzzy Logic, the truth values are no longer restricted to two values, but are expressed by the linguistic variables such as and including "true" or "false". In all forms of fuzzy reasoning, the implications can be modeled in different ways.

Figure 6 shows the Fuzzy Logic algorithm. It begins with initialization of linguistic variables (see Section 2.3.4) and construction of appropriate

```
1. Define the linguistic variables and terms (initialization)
2. Construct the membership functions (initialization)
3. Construct the rule base (initialization)
4. Convert crisp input data to fuzzy values
   using the membership functions (fuzzification)
5. Evaluate the rules in the rule base (inference)
6. Combine the results of each rule (inference)
7. Convert the output data to non-fuzzy values (defuzzification)
```

Figure 6: Fuzzy Logic algorithm.

membership functions (see Section 2.3.3) and the rule-base of the fuzzy system (see Section 2.3.6). The constructed membership functions transform the input data to fuzzy values (see Section 2.3.7). Then the inference system evaluates the constructed rules with respect to the given input value, and merges the results obtained from each rule. Finally, the overall result will be transformed to a non-fuzzy (crisp) value (see Section 2.3.9).

2.3.1 Probability vs Possibility

The theory of possibility is analogous and yet conceptually different from the theory of probability. Based on the Fuzzy Logic theory of Zadeh [42] there is a difference between the possibility of an event happening and the probability of that. The following example by Zimmermann in [44] shows the difference. Consider the statement "Hans ate X eggs for breakfast", where $X \in U = \{1, 2, \ldots, 8\}$. We may associate a probability p by observing *Hans* eating breakfast for 100 days,

$$\begin{array}{lcl} U & = & [\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array}] \\ p & = & [\begin{array}{cccccccc} .1 & .8 & .1 & 0 & 0 & 0 & 0 & 0 \end{array}] \end{array}$$

A fuzzy set expressing the degree to which *Hans* can eat X eggs for breakfast may be the following possibility distribution π,

$$\begin{array}{lcl} U & = & [\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{array}] \\ \pi & = & [\begin{array}{cccccccc} 1 & 1 & 1 & 1 & .8 & .6 & .4 & .2 \end{array}] \end{array}$$

where the possibility of $X = 3$ is 1, while the probability of *Hans* eating 3 eggs for breakfast is only 0.1. Therefore, as the example shows, a possible event does not necessarily imply that it is probable too. However, if the event is probable it must also be possible [44].

2.3.2 Fuzzy Sets

Fuzzy sets are a further development of the mathematical concept of conventional or crisp sets. A fuzzy set is a class of objects with a continuum of degrees of membership [42]. Following Zadeh [42] many sets have more than an either-or criterion for membership, such as, the set young people, which can contain people of different ages. A fuzzy set A is defined by a membership function μ_A from the universe of discourse $\mathcal{X}$ to the closed unit interval [0,1]. We interpret $\mu_A(x)$ as the degree of membership of x in A (see Section 2.3.3). Zadeh proposed this degree of membership, such that the transition from membership to non-membership is gradual rather than abrupt. Therefore, the degree of membership for all its members describes a fuzzy set.

2.3.3 Membership Functions

Membership functions are the crucial part of the Fuzzy Logic theory. In fact, the difference between crisp (i.e., classical) and fuzzy sets is established by introducing membership functions. Membership functions are mathematical tools for indicating flexible membership to a set, modeling, and quantifying the meaning of symbols. Membership functions are used in the fuzzification and defuzzification steps (see Sections 2.3.7 and 2.3.9) of a Fuzzy Logic system. A membership function is used to quantify a linguistic term (see Section 2.3.4). Therefore, the manipulation of fuzzy quantities can be accomplished by the manipulation of fuzzy set membership functions. Some of the manipulation includes set complement, intersection, and union as well as fuzzification and defuzzification (see Sections 2.3.7 and 2.3.9) [30].

Let $\mathcal{X}$ be a crisp universal set. A fuzzy subset A of $\mathcal{X}$ is characterized by a membership function; $\mu_A : \mathcal{X} \rightarrow [0, 1]$. $\mu_A(x)$ is called the *membership degree (grade)* of x in A. The degree of membership is expressed by a real number in the interval [0, 1]. The degree of membership is a precise, but subjective measure that depends on the context.

Figure 7 shows membership functions for three linguistic terms of the variable age (see also Section 2.3.4). It shows three examples of membership functions in the interval 0 to 70 years. These three functions define the degree of membership of any given age in the sets of young, mature, and old ages. Note that an important characteristic of fuzzy logic is that a numerical

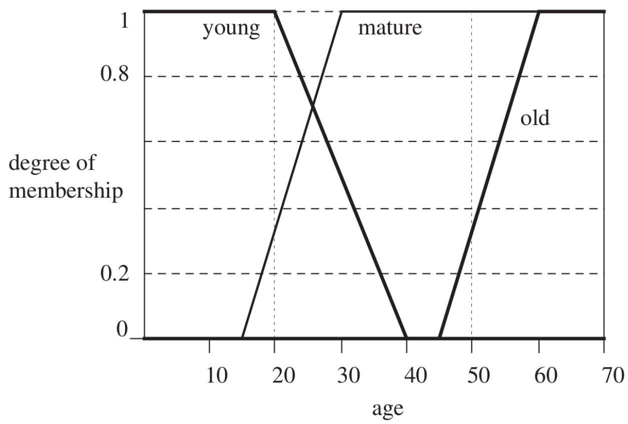

Figure 7: Membership functions for the concepts young, mature and old.

value does not have to be fuzzified using only one membership function. In other words, a value can belong to multiple sets at the same time. For instance, if someone is 20 years old her degree of membership in the set of young persons is 1.0 (maximum value), in the set of matures 0.35, and in the set of old persons 0.0 (minimum value). As another example, if someone is 50 years old the degrees of membership in the sets of young, mature, and old are 0.0, 1.0, and 0.3 respectively.

Membership functions can have different shapes and their shapes can be determined arbitrarily based on experience or sometimes by running statistical studies on data. They can be sigmoidal, hyperbolic, Gaussian or any other shape. The followings are some of the important properties of fuzzy sets and membership functions:

Height: The height of a fuzzy set A, denoted by $h(A)$, corresponds to the upper bound of the membership function. In other words, it is the largest membership degree obtained by any element in that set:

$$h(A) = \sup\{\mu_A(x) | x \in \mathcal{X}\}.$$

Support: The support of a fuzzy set A is a set of all elements x of $\mathcal{X}$ for which $(x, \mu_A(x)) \in A$ and $\mu_A(x) > 0$ holds. In other words, support is a set of all elements of $\mathcal{X}$ that have non-zero membership degrees in A.

α**-cut:** An α-cut of a fuzzy set A is the subset of elements with a membership degree greater than or equal to α. The α-cut is denoted by:

$\alpha\text{-cut}(A) = \{x \in \mathcal{X} | \mu_A(x) \geqslant \alpha\}$.

core: The core of a fuzzy set A is the crisp set that contains all the elements of $\mathcal{X}$ that have the membership degrees of **one** in A.

2.3.4 Linguistic Variables

The concept of membership functions discussed in Section 2.3.3 allows us to define fuzzy systems in natural language. Linguistic variables are the input or output variables of the system whose values are words or sentences from a natural language, instead of numerical values. In fact, just like an algebraic variable takes numbers as values, a linguistic variable takes words or sentences as values [44]. A linguistic variable is generally decomposed into a set of linguistic terms. For instance, for people's age, we usually use terms such as "old" or "young" which are called linguistic values of the age. Then, we can consider a set of decompositions for the linguistic variable age, $Age(a) = \{$very-old, old, mature, young, very-young$\}$. The members of this decomposition set are called *linguistic terms* which each cover a portion of the overall values of people's age. In other words, the values that a linguistic variable can take is called its linguistic terms.

2.3.5 Fuzzy Operators

Fuzzy operators are used in order to manipulate fuzzy sets, and to evaluate the constructed fuzzy rules (see Section 2.3.6), and ultimately to combine the results of the individual rules. The operations on fuzzy sets are different than the operations on classical sets. The definitions of operators on fuzzy sets are not the same and can be arbitrarily chosen. Zadeh in [42] defined the intersection (logical and), union (exclusive or), and complement (negation) operations for fuzzy sets as generalizations of crisp sets and of crisp statements. The operators for the complement (NOT), the intersection (AND) and union (OR) that are most commonly used are:

The membership function of the **Intersection** of two fuzzy sets A and B:

$$\mu_{A\cap B}(X) = \text{Min}(\mu_A(X), \mu_B(X)) \qquad \forall x \in X$$

The membership function of the **union** of two fuzzy sets A and B:

$$\mu_{A \cup B}(X) = \text{Max}(\mu_A(X), \mu_B(X)) \qquad \forall x \in X$$

The membership function of the **complement** of a fuzzy set A:

$$\mu_A(X) = 1 \text{ - } \mu_A(X) \qquad \forall x \in X$$

These definitions were later extended by other reasearchers, e.g., [41].

2.3.6 Fuzzy Rules

In a Fuzzy Logic system, a rule-base is constructed to determine and control the output variable. Fuzzy rules are simply comprised of IF-THEN rules which include two parts of condition and conclusion. A fuzzy rule is encoded in a statement in the following form:

IF (a statement of conditions is satisfied)
THEN (a set of consequences can be inferred).

The followings are two examples of fuzzy rules based on the age example depicted in Figure 7:

IF (age is *young*) **THEN** (run *command "talk"*)
IF (age is *old* **OR** *mature*) **THEN** (run *command "listen"*)

The rules use the input membership values as weighting factors to determine their influence on the fuzzy output sets of the final output conclusion.

2.3.7 Fuzzification

In order to map the crisp values to fuzzy ones, we need to evaluate their membership degree using membership functions (see Section 2.3.3). This process is called fuzzification and it helps us to get one fuzzy value for each crisp input. Therefore, the fuzzification process is mainly used to transform a crisp set to a fuzzy set.

2.3.8 Reasoning in Fuzzy Logic

In order to draw conclusions from a rule-base, we need a mechanism that can produce an output from a collection of IF-THEN rules. Meaning, after evaluating the result of each rule with respect to the given input value(s), the results of the rules should be combined to obtain a final result. This process in Fuzzy Logic systems is called reasoning. Fuzzy reasoning includes two distinct parts: evaluating the IF part of the rule and applying the result to the consequent (the THEN part of the rule). In fuzzy systems the evaluation is slightly different than the classical rule-based systems. In fuzzy systems the IF part of the rule is a fuzzy statement which means all the rules fire to some extent. If the IF part of the rule is true in some degree of membership, then the consequent is also true in some degree. It is noteworthy to mention that the results of individual rules can be combined in different ways. There are different types of accumulation methods that can be used to combine the results of individual rules.

2.3.9 Defuzzification

After the reasoning step, the Fuzzy Logic system provides the overall result as a fuzzy value. Then, to obtain a final crisp output value, this fuzzy result should be defuzzified which is the purpose of the defuzzifier component of a Fuzzy Logic system. Defuzzification is performed according to the membership function of the output variable. Figure 8 shows the defuzzification step in a Fuzzy Logic system.

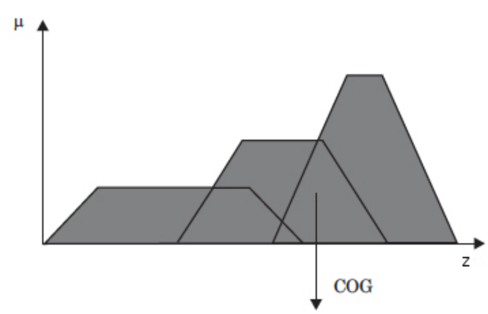

Figure 8: Defuzzification using Center of Gravity (COG) method [1].

There are different algorithms for the defuzzification step. One of the most widely used algorithms is called *centroid*, or *Center of Area*, or *Center of Gravity* (COG). This method computes the center of area of the region under the curve defined by a fuzzy set. In this method, the defuzzified values tend to move smoothly in reaction to small changes, and it is relatively easy to compute the value. In the following formula, A is a fuzzy set and z_{COG} is the final single crisp output which in this case is obtained by COG method:

$$z_{COG} = \frac{\sum_{i=1}^{n} \mu_A(z_i).z_j}{\sum_{i=1}^{n} \mu_A(z_i)}$$

In summary, Figure 9 (see also the fuzzy algorithm in Figure 6) shows the process of fuzzy logic. Firstly, a crisp set of input data are gathered and converted to a fuzzy set using fuzzy linguistic variables (see Section 2.3.4), fuzzy linguistic terms and membership functions (see Section 2.3.3). This step is known as fuzzification (see Section 2.3.7). Afterwards, an inference is made based on a set of rules (see Section 2.3.8). Lastly, the resulting fuzzy output is mapped to a crisp output using the membership functions, in the defuzzification step (see Section 2.3.9).

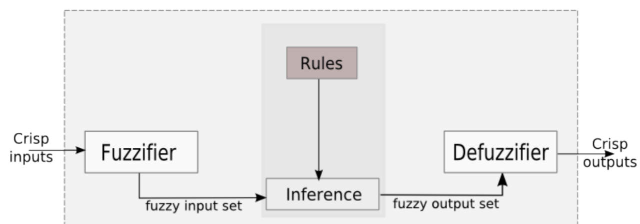

Figure 9: A Fuzzy Logic system.

3 Advantages and Disadvantages of Theories of Uncertainty

Each of the three theories of uncertainty presented herein has certain advantages and disadvantages. They are discussed in the following sections.

3.1 Advantages and Disadvantages of Belief Networks

Like any other computational formalism, Bayesian networks offer certain advantages and disadvantages. Advantages of Bayesian networks include:

- Transparent representation of causal relationships between variables which can facilitate understanding the causal relationships.
- Relatively easy recognition of dependencies and independencies between various nodes.

- The ability to handle situations where the data set is incomplete since the model accounts for dependencies between all variables.
- Capable of being readily updated when a new knowledge (evidence) becomes available.
- Both predictive/deductive and diagnostic/abductive reasonings are possible.
- Computational tractability exists for most practical applications.

The followings are the disadvantage of Bayesian networks:

- A high level of effort is required to build network models where a significant amount of probability data is required due to an increasing number of nodes and links in the structure (possible large CPT sizes).
- Computationally intensive if the conditional independencies are not properly considered among the variables.
- Challenging to obtain experts' knowledge in the form of probability to build the network.
- No feedback loops in the Bayesian network's structure, which has an acyclic nature. This structure prevents typical feedback loops in design of Bayesian network models.

3.2 Advantages and Disadvantages of Dempster-Shafer Theory

Advantages of Dempster-Shafer theory are:

- Addressing the concept of possibility.
- The ability to represent the concept of ignorance to allow one to specify a degree of ignorance in a situation, instead of being forced to supply prior probabilities.
- Consistent with classical probability theory.
- Distinguishing randomness from missing information.
- No required a priori knowledge.

- Including an evidence combination rule which provides an operator to integrate multiple pieces of information from different sources.

Disadvantages of Dempster-Shafer theory are:

- Computational complexity grows exponentially with the number of hypotheses (in original formulation).
- Small modifications in the evidence assignments may lead to a completely different conclusion, which can lead to misleading and counter-intuitive results.

3.3 Advantages and Disadvantages of Fuzzy Logic Theory

Advantages of Fuzzy Logic theory are:

- Describing algorithms in terms of a combination of numerics and linguistics.
- Capturing the concept of the ambiguity of information.
- Flexible and intuitive knowledge-base design.
- Easy computation.
- Relatively robust algorithms.

Disadvantages of Fuzzy Logic theory are:

- Determining the exact fuzzy rules and membership functions is a hard task.
- Requires manual tuning to obtain a better result.
- Requires tuning in many options in design of a system.
- The order of inference steps matters.
- After reasoning, it can be difficult to exactly interpret the membership value.
- Validation of a fuzzy knowledge-base is typically expensive.

4 Applications of Bayesian Networks

There are different applications of Bayesian networks in robots and autonomous agents such as motion control, sensory data fusion, and modeling domain knowledge. In this section, we provide some of these applications.

In [37] Toussaint proposes a new approach to robotic motion control and planning based on probabilistic inference. His method uses structured probabilistic models to represent a scenario and efficient inference techniques (belief propagation) to solve planning problems. He also uses Bayesian networks to find solutions to problems combining multiple criteria or sources of information (sensor fusion) beside using them to solve a fusion problem on the motor level. [25] uses probabilistic modeling for service robots to provide users with high-level context-aware services required in home environments, and proposes a systematic modeling approach for modeling a number of Bayesian networks. In [24] researchers have developed a system to integrate indirect measures of different sensors. This system is designed based on a Bayesian network and can be applied to use any type of sensor which provides measures of the robot's environment independent of sensor types, environment and application of the robot. In [7] authors use Bayesian networks to model interactions between humans and multiple semi-autonomous robots. They have modeled discrete operator decisions as probabilistic Bayesian network blocks with conditional dependencies on individual system states. In [27] researchers propose a novel Bayesian approach to determine the emotional state the robot shall assume according to how the interlocutor is talking to it through a vocalization channel. In [32] authors adopt Bayesian Networks and an ontology together for modeling domain knowledge and reasoning objects in a probabilistic framework for a service robot. In this work, they use objects as context information for predicting the target object being present, since it can be occluded or small in indoor environments. In [10] researchers use a Bayesian network to represent a forward model of a robot's actions to predict the consequences of the robot's actions on its own motor system and the environment. In [22] researchers propose a general object affordance model based on Bayesian networks linking actions, object features and action effects. The network is learned by the robot through interaction with the surrounding objects. [38] presents the navigation mobile robot systems for movement in non-stationary and non-structured environments, using a Bayesian approach for avoiding obstacles and dynamical stability control for motion on rough terrain. As another example, in [8] researchers explore the use of Dynamic Bayesian networks as a decision process to aid a mobile robot in planning its surveillance route. In [17] Iwahashi describes a method that enables the robot to learn a system of beliefs through multi-modal language interaction with the user. He uses the Bayesian learning method to learn the robot's decision and confidence functions. [15] presents learning imitative whole body motions in a humanoid robot using probabilistic inference in Bayesian networks. Authors provide a model for exploiting prior information about whole-body motions gathered from observing a human performance of the motion.

There are other applications such as intention recognition for intelligent human-robot interaction [34], mobile-robot localization [43] [14], and autonomous reconstruction of 3D images for indoor mobile robots [11]. Here, we also provide applications of Bayesian networks in agents.

In [2] researchers propose an efficient and scalable Bayesian network technique using relevance reasoning, that are suitable for the needs of their autonomous agents in directed emergent drama and allow for real-time decision making. They do this by extracting only a relevant part of the network and only update the values in this reduced network. In [39] Xiang extends Multiply Sectioned Bayesian networks (MSBNs) for single-agent systems into a framework for multi-agent distributed interpretation systems. In this work, each cooperative agent is represented as a Bayesian subnet that consumes its own computational resource, gathers its own evidence, and can answer queries. In [29] authors present a decision architecture of the arguing agent. They propose a novel Bayesian network based an argumentation and decision making framework that allows agents to utilize models of the other agents. In [16] authors provide a review of Bayesian networks' applications in different domains including image interpretation and modeling cognitive processes. For all of the examples they explain how the Bayesian models are implemented, their practical use and their limitations. There are also other applications for Bayesian networks for the agent, such as building internal models in multi-agent settings [23], knowledge representation of the agents [6] [18] [21], modeling theory of mind [3] [4], and modeling trustworthiness of agents [35].

5 Conclusion

In this response, we started by generally understanding the importance of reasoning under uncertainty in intelligent machines. There are some theories explaining similar or different aspects of uncertainty. Each of these theories relates to the same concept of uncertainty from slightly different angles. For instance, Bayesian networks rely on probability theory and assign one value to each existing variable and relate them based on their causal relationship. Dempster-Shafer theory as an evidential reasoning theory discusses whether there is any evidence for a particular belief rather than probability of its truth, and it is designed to deal with the distinction between uncertainty and ignorance. And finally, the Fuzzy Logic theory has the power to deal with ambiguous linguistic words and still provide reasonable results that gain enough credit to be used as an underlying concept in many industrial controllers. However, all of these theories have their own drawbacks in dealing with the key issue, uncertainty (see Section 3). As a result, it is one's experience and the nature of the application which helps one to choose from among these approaches to gain more accurate and reasonable results for the application.

In our proposed work, we believe sources of uncertainty are everywhere in a collaboration. The collaboration theories in general, and specifically SharedPlans theory, are concerned about teamwork and the involvement of others to form an intention, to generate or evolve the shared plan, or even to establish a single mutual belief. To do so, one should rely on perceiving others' verbal and nonverbal behaviors; in both cases there is a certain amount of uncertainty, ambiguity and lack of evidence. Therefore, processes involved in this domain need to be designed to address the existence of uncertainty if they are to autonomously generate and evolve the underlying structure of a collaboration procedure. Moreover, beliefs, as the foundation of the underlying emotion-driven processes, are involved in collaboration mechanisms, e.g. appraisal, motivation. Each individual belief includes a certain amount of uncertainty with respect to whichever process has formed the belief, and under what situation. For instance, sometimes the lack of evidence about a counterpart's belief about an event, or the feeling of a counterpart for a collaborative action can cause irrelevant responses based on false beliefs which could be at least mitigated by having a mechanism to deal with uncertainty in some level. As we mentioned, the existence of uncertainty appears in different constituents of a collaboration procedure and it is for us to choose where to apply the appropriate mechanism to make more stable collaborative behaviors.

References

[1] Antonio Claudio Marques Afonso, Andre Maciel Netto, and Wagner Estaquio de Vasconcelos. Fuzzy logic applied to the modeling of water dynamics in an oxisol in northeastern brazil. *Revista Brasileira de Ciencia do Solo*, 38(2):454–463, 2014.

[2] Maria Arinbjarnar and Daniel Kudenko. Bayesian networks: Real-time applicable decision mechanisms for intelligent agents in interactive drama. In *IEEE Symposium on Computational Intelligence and Games (CIG)*, 2010.

[3] Chris L. Baker, Rebecca R. Saxe, and Joshua B. Tenenbaum. Bayesian theory of mind: Modeling joint belief-desire attribution. In *Proceedings of the 33th Annual Conference of the Cognitive Science Society*, pages 2469–2474, 2011.

[4] Paul Bello. Cognitive foundations for a computational theory of mindreading. *Advances in Cognitive Systems*, 1:59–72, 2012.

[5] Malcolm Beynon, Bruce Curry, and Peter Morgan. The dempstershafer theory of evidence: an alternative approach to multicriteria decision modelling. *Omega, The International Journal of Management Science*, 28(1):37–50, 2000.

[6] Mark Bloemeke and Scott Langevin. Agent-encapsulated bayesian networks and the rumor problem. In *Proceedings of the Ninth International Conference on Autonomous Agents and Multiagent Systems (AAMAS-10)*, 2010.

[7] Frederic Bourgault, Nisar Ahmed, Danelle Shah, and Mark Campbell. Probabilistic operator-multiple robot modeling using bayesian network representation. In *AIAA Guidance, Navigation and Control Conference*, 2007.

[8] Leon Chong and Raymond Jarvis. A dynamic bayesian network approach to surveillance decision planning, 2010.

[9] Subrata Das. *Foundations Of Decision-Making Agents: Logic, Probability and Modality.* World Scientific Publishing Co., 2008.

[10] Anthony Dearden and Yiannis Demiris. Learning forward models for robots. In *Proceedings of the 19th International Joint Conference on*

Artificial Intelligence, pages 1440–1445. Morgan Kaufmann Publishers Inc., 2005.

[11] Erick Delage, Honglak Lee, and Andrew Y. Ng. A dynamic bayesian network model for autonomous 3d reconstruction from a single indoor image. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2418–2428, 2006.

[12] Arthur P. Dempster. A generalization of bayesian inference. *Journal of the Royal Statistical Society*, 30(B):205–247, 1968.

[13] Francisco J. Diez and Marek J. Druzdzel. Reasoning under uncertainty. In L. Nadel, editor, *Encyclopedia of Cognitive Science*, pages 880–886. London: Nature Publishing Group, 2003.

[14] Dieter Fox, Jeffrey Hightower, Lin Liao, Dirk Schulz, and Gaetano Borriello. Bayesian filtering for location estimation. *IEEE Pervasive Computing*, 2:24–33, 2003.

[15] David B. Grimes, Rawichote Chalodhorn, and Rajesh P. N. Rao. Dynamic imitation in a humanoid robot through nonparametric probabilistic inference. In *Proceedings of Robotics: Science and Systems (RSS06)*, 2006.

[16] Tijl Grootswagers, Bas Kooiker, and Lysanne Sloff. Applications of bayesian network models in cognitive science, 2012.

[17] Naoto Iwahashi. A method for the coupling of belief systems through human-robot language interaction. In *The 12th IEEE International Workshop on Robot and Human Interactive Communication*, pages 385–390, 2003.

[18] Eugene Santos Jr. A cognitive architecture for adversary intent inferencing: Structure of knowledge and computation. In *Proceedings of the SPIE 17th Annual International Symposium on Aerospace/Defense Sensing and Controls (AeroSense)*, 2003.

[19] Rakowsky Uwe Kay. Fundamentals of the dempster-shafer theory and its applications to system safety and reliability modelling. *Reliability : Theory & Applications*, 3(4):173–185, 2007.

[20] Kevin B. Korb and Ann E. Nicholson. *Bayesian Artificial Intelligence*. Taylor & Francis, 2003.

[21] Yves Martin and Michael Thielscher. Integrating reasoning about actions and bayesian networks. In *ICAART*, 2010.

[22] Luis Montesano, Manuel Lopes, Alexandre Bernardino, and Jose Santos-Victor. Modeling affordances using bayesian networks. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4102–4107, 2007.

[23] Soren Holbech Nielsen and Simon Parsons. An application of formal argumentation: Fusing bayesian networks in multi-agent systems. *Artificial Intelligence*, 171(1015):754–775, 2007.

[24] José Antonio López Orozco, Jesús Manuel de la Cruz García, J. Sanz, and J. Flores. Multisensor fusion of environment measures using bayesian networks. In *Fusion'98: Proceedings of the International Conference on Multisource-Multisensor Information Fusion*. CSREA Press, 1998.

[25] Han-Saem Park and Sung-Bae Cho. A modular design of bayesian networks using expert knowledge: Context-aware home service robot. *Expert Systems with Applications*, 39(3):2629–2642, 2012.

[26] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., 1988.

[27] Jose Augusto Prado, Carlos Simplicio, and Jorge Dias. Robot emotional state through bayesian visuo-auditory perception. In Luis M. Camarinha-Matos, editor, *Technological Innovation for Sustainability*, pages 165–172. Springer Berlin Heidelberg, 2011.

[28] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson Education, 2003.

[29] Sabyasachi Saha and Sandip Sen. A bayes net approach to argumentation based negotiation. In *First International Workshop on Argumentation in Multi-Agent Systems*, pages 208–222, 2004.

[30] Robert J. Schalkoff. *Intelligent Systems: Principles, Paradigms and Pragmatics*. Jones Bartlett Learning, 2011.

[31] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[32] Youn-Suk Song and Sung-Bae Cho. Automated bayesian network integration based on ontology for reasoning object existence of service robot. In *Proceedings of the 17th World Congress. The International Federation of Automatic Control*, 2008.

[33] Rajendra P. Srivastava. An introduction to evidential reasoning for decision making under uncertainty: Bayesian and belief functions perspectives. *International Journal of Accounting Information Systems*, 12(2):126–135, 2011.

[34] Karim A. Tahboub. Intelligent human-machine interaction based on dynamic bayesian networks probabilistic intention recognition. *Journal of Intelligent and Robotic Systems*, 45(1):31–52, 2006.

[35] Colin Keng-Yan Tan. A belief augmented frame computational trust model. In *FLAIRS Conference*, pages 647–652. AAAI Press, 2005.

[36] Steven Tanimoto. *The elements of artificial intelligence: an introduction using LISP*. Computer Science Press, 1987.

[37] Marc Toussaint. Bayesian inference for motion control and planning. Technical Report 2007-22, Machine Learning Group, TU Berlin, 2007.

[38] Luige Vladareanu, Gabriela Tont, Victor Vladareanu, Florentin Smarandache, and Lucian Capitanu. The navigation mobile robot systems using bayesian approach through the virtual projection method. In *Advanced Mechatronic Systems (ICAMechS)*, pages 498–503, 2012.

[39] Yang Xiang. Distributed multi-agent probabilistic reasoning with bayesian networks. In Zbigniew W. Ras and Maria Zemankova, editors, *Methodologies for Intelligent Systems*, pages 285–294. Springer Berlin Heidelberg, 1994.

[40] Ronald R. Yager. On the dempstershafer framework and new combination rules. *Information Science*, 41(2):93–137, 1987.

[41] Ronald R. Yager, Carol L. Walker, and Elbert A. Walker. Generalizing leximin to t-norms and t-conorms: the lexit and lexis orderings. *Fuzzy Sets and Systems*, 151(2):327–340, 2005.

[42] Lotfi A. Zadeh. Fuzzy sets. *Information and Control*, 8(3):338–353, 1965.

[43] Hongjun Zhou and Shigeyuki Sakane. Sensor planning for mobile robot localization using bayesian network representation and inference. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 440–446, 2002.

[44] Hans-Jürgen Zimmermann. *Fuzzy Set Theory and Its Applications*. Springer Science Business Media, 2001.