Extended Abstract: Using Defeasible Arguments to Update Quantified Beliefs ⋆

Nandi Schoots[1,2] and Sanjay Modgil[1][0000−0002−7873−8086]

[1] King's College London
[2] Imperial College London

This work investigates how an agent should update the probability estimates for their beliefs upon hearing an argument. Here, beliefs are framed in terms of the truth-statuses of logical sentences and inferences between them. Inferences between premises and conclusions are fully-instantiated inference rules and can be interpreted as a reason for why a conclusion follows from premises.

Consider an argument consisting of a premise A and a (defeasible) inference f which leads to conclusion C. How much should an agent's probability estimate of the truth-status of C change based on probability estimates of the truth-statuses of A and f? Given an answer to this question, how should the agent propagate the changes through a reasoning chain? This work investigates whether a belief update is an improvement by relating an agent's beliefs to a ground truth.

We often think of statements as being either true or false. However, for many statements, such as "It will rain tomorrow" we currently do not have the tools to assess their truth-status with complete certainty. A recent pragmatic theory of uncertainty [3] says that we should ask whether, for a specific modeller, uncertainty is reducible. When we model an agent who has beliefs about the world, we may have irreducible uncertainty about the correctness of these beliefs.

In our investigation, the ground truth consists of a hypothetical *true probability distribution* which represents irreducible uncertainty. The distribution is hypothetical because, whilst we assume that one exists, we do not expect to have access to it. We call the function the true probability distribution to emphasise that it is the ground truth in the sense that it represents an objectively best possible estimate of what statements are true in our world. We will assess an agent's probability estimates by comparing them to this best possible estimate.

We assume the hypothetical true probability distribution is a probability function over arguments. A notion that we define below. Given a set of well-formed formulas Φ and a set of inferences $\mathcal{H}$ between these formulas, we consider possible combinations of true formulas and true inferences. We define the sample space $\mathcal{W}$ of arguments as the largest subset of $2^{\Phi \sqcup \mathcal{H}}$, such that for all $W \in \mathcal{W}$:

1. If inference $(f\colon A \to B)$ holds in W then $B \in W$ or $A \notin W$. Note that $B \in W$ or $A \notin W$ is a necessary but insufficient condition for $(f\colon A \to B) \in W$;
2. For all $A \in \Phi$, the trivial inference "A follows from A" is an element of W;
3. Consider inferences $(f\colon A \to B), (g\colon B \to C) \in \mathcal{H}$. If $f, g \in W$, then $(g \circ f\colon A \to C) \in W$. If $f \notin W$ or $g \notin W$, then $g \circ f \notin W$;
4. W is consistent and closed under the common logical connectives.

⋆ Supported by UKRI studentship EP/S023356/1

We call an element $W \in \mathcal{W}$ a world, which can be interpreted as an internally consistent possible world of well-formed formulae and inferences between them.

A probability function over arguments is defined as a probability function over a sample space of arguments. Let P be a probability function over a sample space of arguments $\mathcal{W} \subset 2^{\Phi \sqcup \mathcal{H}}$, $A \in \Phi$ and $(f\colon A \to C) \in \mathcal{H}$. We denote the event $\{W \in \mathcal{W} | A \in W\}$ by A and likewise we denote $\{W \in \mathcal{W} | f \in W\}$ by f. Whenever an inference f has probability $P(\mathrm{f}) = 1$, we say the inference is deductive. Conversely, whenever an inference does not have probability 1, we say that the inference is defeasible. The strength of an argument consisting of a premise A and a (defeasible) inference f which leads to conclusion C is then defined as the joint probability of A and f.

We place an assumption on the beliefs that an agent starts out with, relating the probability estimates of the agent to the hypothetical true probability distribution. Namely, we assume that the agent's initial estimates form a justified function with respect to that function. We say a function J is justified with respect to a probability function P if $J(\mathrm{E}) \leq P(\mathrm{E})$ for all events E. The belief function in Dempster-Shafer theory is an example of a justified function.

We propose a method for updating a function which is justified with respect to a probability function over arguments. The strength of an argument which concludes C is treated as a lower bound for the probability $P(\mathrm{C})$. Furthermore we use lower bounds $J(\mathrm{A}) \leq P(\mathrm{A})$ and $J(\mathrm{f}) \leq P(\mathrm{f})$ to update J.

Using logical properties of the sample space of arguments and results from probability theory we prove that our update rules strictly improve justified functions. Lastly, we propose an anytime algorithm for propagating updates.

The main contributions of this work are: 1) a new formalism of a probability function over arguments; and 2) a mechanism for improving a justified function.

Haenni's probabilistic argumentation framework [1] investigates the probability that a conclusion can be logically deduced from premises, when probabilities for these premises are known. Our work differentiates from this work in that we consider defeasible inferences. We offer a quantitative model of defeasible reasoning by assigning probabilities to inferences. In our probability space of arguments, every world is monotonic. However, an inference f may hold in some worlds but not all, and this results in probabilities on inferences.

A recent contribution to the field of probabilistic argumentation is the epistemic approach [2], which assumes a user has access to a probability distribution over logical sentences. This distribution is used to reduce the number of arguments that need to be taken into account. This approach does not aim to answer the question 'How can we obtain or update a probability function?'

In the DeLP3E paradigm [4] an agent is assumed to have access to the 'environmental model' (which describes uncertain knowledge subject to probabilistic events) and know exactly which possible worlds there are. In our formalism we do not assume that an agent has access to all possible worlds, i.e. the agent does not have to know what sentences are true and false in each world. We consider a setting in which an agent only has access to a correlate of the true probability distribution, and aims to improve their correlate.

References

1. Haenni, R.: Probabilistic argumentation. J. Appl. Log. **7**(2), 155–176 (2009)
2. Hunter, A.: Generating Instantiated Argument Graphs from Probabilistic Information. In: 24th European Conference on Artificial Intelligence, pp. 769–776. IOS Press, Santiago de Compostela (2020)
3. Der Kiureghian, A., Ditlevsen, O.: Aleatory or epistemic? Does it matter? Structural Safety **31**(2), 105-112 (2009)
4. Simari, G.-I., Shakarian, P., Falappa, M.-A.: A quantitative approach to belief revision in structured probabilistic argumentation. Ann. Math. Artif. Intell. **76**(3-4), 375–408 (2016)