Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Markov Argumentation Random Fields

Yuqing Tang[1], **Nir Oren**[2], and **Katia Sycara**[1]
[1] Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, USA
[2] Department of Computing Science, University of Aberdeen, Aberdeen, UK

Abstract

We demonstrate an implementation of Markov Argumentation Random Fields (MARFs), a novel formalism combining elements of formal argumentation theory and probabilistic graphical models. In doing so MARFs provide a principled technique for the merger of probabilistic graphical models and non-monotonic reasoning, supporting human reasoning in "messy" domains where the knowledge about conflicts should be applied. Our implementation takes the form of a graphical tool which supports users in interpreting complex information. We have evaluated our implementation in the domain of intelligence analysis, where analysts must reason and determine likelihoods of events using information obtained from conflicting sources.

Markov Argumentation Random Fields

A longstanding goal for the AI community has been to integrate symbolic and probabilistic knowledge. The latter is suited for dealing with uncertainties, while the former allows for explicit (symbolic) knowledge representation, and helps to handle complex knowledge structure. Real world tasks, such as intelligence analysis and social network analysis, require both forms of knowledge. These tasks typically face uncertainties and conflicting information from inaccurate sensors and human. The core challenge is of two folds: 1) how to make use of knowledge on uncertainty and conflicts while incorporating probabilistic and symbolic reasoning in a mathematically sound manner, 2) how to expose reasons and rejections in an intuitive manner.

In this work, we present an approach that combines reasoning about probabilities with argumentation based non-monotonic theory, forming *Markov Argumentation Random Fields* (MARFs). Unlike classical logic which describes knowledge as what holds in all situations, therefore not allowing any conflicts, formal argumentation theory (Dung 1995) describes how a justifiable "stable" set of arguments can be extracted from a large set built on the principle of reinstatement which has been confirmed by human experiments. Underpinning by formal argumentation theory, MARFs construct possible worlds of argumentation as acceptability interpretations of a predicate language which represents knowledge about inference — argument rules — and knowledge about conflicts — defeat rules. MARFs compile such knowledge into factor graphs which specify factorized probabilistic distributions over the possible worlds of argumentation. MARFs then identify features of these possible worlds via evaluation of acceptability status of arguments and defeats. The acceptability status takes 4 values: accepted (**A**), rejected (**R**), undecided (**U**), and ignored (**I**). These features are defined in a way that the nature of these worlds can be exposed and parameterized (weighted) governed by the argumentation properties of being admissible, complete, grounded, stable and preferred (Dung 1995). As a result, MARFs are able to reasoning with knowledge about conflicts and uncertainties while at the same time exposing reasons and rejections of the inference to humans in an intuitive manner (Fig. 2).

MARFs inference is in the form of marginal and maximal probability distribution $Pr(q \mid E)$ over the acceptability status of a query q, condition on observed evidence. The list of observed evidence is of the form $E = \{e_1 = y_1, e_2 = y_2, .., e_m = y_m\}$ where each y_i is the observed acceptability of the evidence e_i $(i = 1, .., m)$. MARFs also derive the *sensitivity* of a piece of information p_i — a premise of q or a piece of information relevant to q — characterizing how the changes in p_i render the outcome of q differently.

Intelligence Analysis using MARFs

Built on the above theory, the MARF software is composed of 1) a front-end web interface, and 2) a back-end inference engine. The front-end has 1) an online knowledge editor (Fig. 1), and 2) an argumentation graph explorer (Fig. 2). Upon the request from the front-end, the back-end engine generates an argumentation graph (Tang et al. 2012) and performs inference using algorithms adapted from message-passing and MC-SAT (Richardson and Domingos 2006).

We exemplify MARFs with an intelligence analysis task — the ELICIT task (Chan and Adali 2012). The ELICIT task requires a human analyst to answer a list of questions: who, what, when, and where regarding a possible terrorist attack given a list of facts. These facts not only contain information related to the questions but also contain noise and information regarding ruling out possible answers (a form of argument defeat). In a typical task, facts are incomplete, inconsistent and ambiguous requiring the analyst to apply both logical reasoning and inconsistency resolution principles to

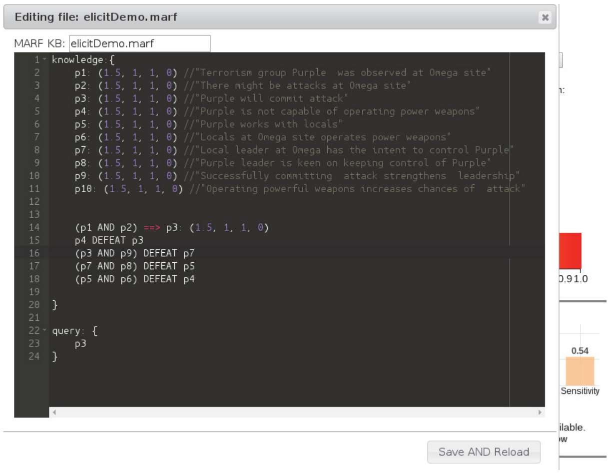

Figure 1: MARF online Knowledge editor

produce answers.

ELICIT inputs come from different sources with different nature over time. First, $p1$: "Purple arrives at Omega site" and $p2$: "There might be attack at Omega site" in Figure 2, come in. However, at the same time, vague information against Purple being the attacker also comes in, such as $p4$: "Purple is not capable of operating power weapons". Additional information that strengthens the conclusion on Purple being the attacker, such as $p9$: "Successfully committing attack strengthen group leadership", continues to come in causing MARF to maintain that the most likely culprit is Purple. However, later on, a considerable amount of information against Purple being the attacker and relevant information, such as $p7$: "Local leader at Omega has the intent to control Purple" and $p8$: "Purple leader is keen on keeping the control of Purple" keeps coming in, causing the MARF to decrease its belief on "Purple" being the attacker. Without additional information, assuming that all knowledge has equal weight $\langle 1.5, 1.0, 1.0, 0.0 \rangle$ (this weight vector is an arbitrary choice to test the system), MARF outputs an acceptability distribution over whether p_3: the attacker is "Purple" as $Pr(p_3) = \langle 0.46, 0.51, 0.03 \rangle$ along an argumentation graph visualizing how relevant information are interrelated as inference and conflicts (Fig. 2). Now assume that the analyst is more certain with regards to p_5: "Purple works with locals", and he updates the weights to $\langle 3.0, 1.0, 1.0, 0 \rangle$ with the online editor (Fig. 1). Now, the acceptability distribution for p_3 changes to $\langle 0.51, 0.46, 0.03, 0.0 \rangle$, thereby indicating that purple is more likely to be the attacker. The sensitivity of all the relevant premises, defeats and consequences of p_3 are derived and color coded in the argumentation graph (Fig. 2).

Conclusion

By combining the strengths of Markov Random Fields with Argumentation, MARFs gain the strengths of both. With MARFs, we are capable of modeling and linking reasoning and conflict patterns and analyzing them probabilistically for the applications where both symbolic argumentation theory and probabilistic inference is useful. We demonstrate such capability in intelligence analysis tasks as an example.

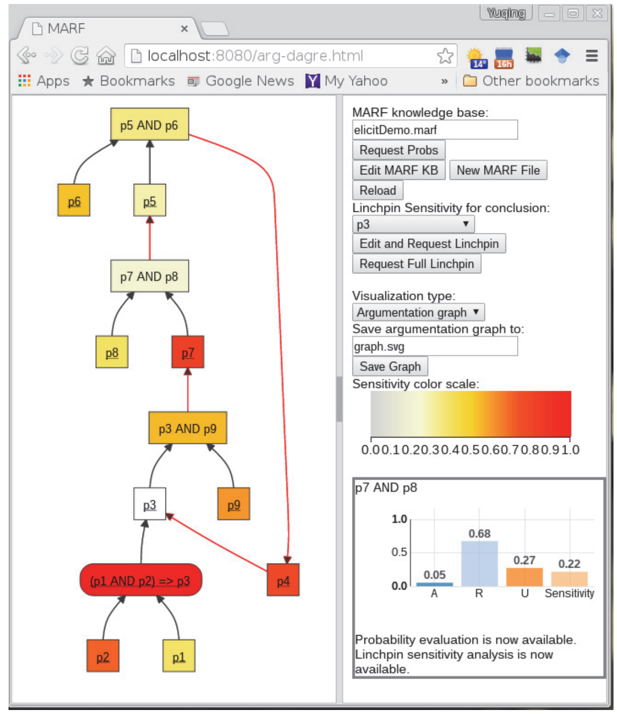

Figure 2: MARF Argumentation Explorer (nodes are color coded by sensitivity levels where the sensitivity scale low-medium-high is represented by the color scale gray-yellow-red)

Acknowledgments

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defense and was accomplished under Agreement Number W911NF-06-3-0001 and Cooperative Agreement Number W911NF-09-2-0053. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

Chan, K., and Adali, S. 2012. An agent based model for trust and information sharing in networked systems. In *Cognitive Methods in Situation Awareness and Decision Support*, 88–95. IEEE.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artif. Intell.* 77:321–357.

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

Tang, Y.; Cai, K.; McBurney, P.; Sklar, E.; and Parsons, S. 2012. Using argumentation to reason about trust and belief. *Journal of Logic and Computation* 22(5):979–1018.