Some Open Issues in Argumentation and KR

Martin Caminada

January 2018

1 Introduction

One of the advantages of working in Leon's group is the wide range of strong researchers one gets to work with. Looking back at it, the time I spent in Luxembourg was one of the most fruitful periods of my career, and it was through Leon that I learned what a strong research group should look like.

Although it was a productive time when it came to publications, there were also a few research problems that I found difficult to tackle. The first problem has to do with argumentation and explainability, especially on how to use discussion as a way to explain argument-based entailment. The second problem has to do with the philosophical concept of knowledge, and the role of argumentation theory. The third problem has to do with how to actually define progress in KR (and in formal argumentation theory in particular). In the following sections, these points will be elaborated on.

2 Argumentation and Discussion

The key idea of KR, and of symbolic AI in general, is to represent knowledge in a formal but human readable way, which can be used as a basis of well-defined reasoning. An often mentioned advantage of symbolic AI (e.g. rule-based reasoning, non-monotonic logic, formal argumentation theory) above non-symbolic AI (e.g. neural networks and deep learning) is the fact that the reasoning of non-symbolic AI tends to be opaque whereas the reasoning of symbolic AI could in principle be explained. Although explainability is potentially a huge advantage, the devil is in the details. One particularly tricky issue is: to whom should symbolic AI be explainable? The current answer that tends to be predominant in the KR community is that explainability should be aimed at other members (often mathematicians) of the KR community. After all, when one submits a KR related paper, this will most likely be reviewed by other KR researchers, who will apply their own intuitions on whether the work satisfies criteria like significance and relevance. However, in order for applications to really take off, what matters is not explainability for other KR scholars but explainability for the end users of the systems KR tries to assist implementing. No doctor will blindly follow the advice of a knowledge based system unless he has at least some idea how the advice came about and is able to somehow assess the correctness of it.

As an example, in the domain of non-monotonic reasoning, formal argumentation theory has been introduced as a way to come closer to human reasoning [6]. However, most of the argumentation semantics have originally been stated in a highly technical way.[1] A more promising approach is to express argumentation semantics in the form of

[1] like "the minimal fixpoint of the monotonic function $F : Ar \rightarrow Ar$ such that $F(\mathcal{A}rgs) = \{A \in Ar \mid A^- \subseteq \mathcal{A}rgs^+\}$"

formal discussion, such that an argument is accepted iff a proponent of this argument is able to "win" a discussion for it [2].

One of the key problems is that in order to make such a discussion feel natural, one would have to adopt the principle of unrestricted rebut [3] which has been shown to satisfy the rationality postulates [3] only under a combination of grounded semantics and rule preferences that are either completely empty or form a total pre-order (thus ruling out the possibility of a proper partial order). The challenge would be to lift these restrictions and to come up with a formal discussion protocol that is not only sound and complete regarding argument acceptance, but also yields a discussion that the user perceives as natural and convincing, while at the same time being powerful enough to work with a wide variety of knowledge bases (for instance, a knowledge base where the ruke strengths form a partial pre-order). It is this challenge that is in my view the most important open research issue in formal argumentation theory.

3 Argumentation, Knowledge and Informedness

Knowledge, from a philosophical perspective, can be defined as justified true belief. However, the most popular logical formalisms for reasoning about beliefs (modal logic with a KT45 axiomatization) cannot deal with the concept of "justified" and therefore implements the more basic notion of "true belief". However, for quite some practical purposes, even the notion of "justified belief" can be more important than that of "true belief".[2]

Justified belief, unlike true belief, allows for various grades. This would make it a suitable concept to be modelled in terms of formal argumentation, as a belief can be regarded as more justified when it is based on a superset of relevant arguments. However, when one attempts to work out this idea in more detail, some of the obvious candidates for formalisation all have their shortcomings, as is indicated by various examples that yield counter intuitive outcomes [4]. The question of how to use formal argumentation theory to model the concept of justified belief is therefore still an open (and relevant) research issue [4].

4 On the Issue of Progress in KR

Many years ago, when I was still working on my PhD thesis, one of the things that I could not get my head around was the precise nature of Assumption-Based Argumentation (ABA) [1]. Although "normal" Dung-style argumentation was based on a graph, there appeared to be no such graph in ABA. Sure, it was possible to translate ABA to standard Dung-style argumentation, but what about the nature of ABA itself? It would take quite some years before I was able to return to this question. It turned out the answer was surprisingly simple: ABA is in essence Logic Programming.

In a recently published paper [5] it is shown that there is a straightforward way to translate an ABA framework to a logic program. In essence, the idea is to translate an ABA rule $a \leftarrow b, c, \delta, \epsilon$ (with $\bar{\delta} = d$ and $\bar{\epsilon} = e$) to a logic programming rule $a \leftarrow b, c, \texttt{not}\ d, \texttt{not}\ e$. When doing so, the complete (resp. preferred, stable, grounded and ideal) assumption labellings of the original ABA framework to coincide with the 3-valued stable (resp. regular, 2-valued stable, well-founded and ideal) models of the associated logic program. As such, it turns out that ABA is essentially logic programming (using 1990s LP semantics) with a slightly non-standard notation.

[2] For instance, when holding someone to account for a particular action whose consequences turned out to be bad, what matters is not so much whether his beliefs were true, but that they were justified at the time the action was taken.

To make the point, suppose one were to define a logical system called $\mathcal{GF}$ around the values $\mathcal{G}$ and $\mathcal{F}$ and logical connectives $\mathcal{E}$, $\mathcal{O}$ and $\mathcal{N}$, and that one would start publishing about this in respectable journals. Suppose further that it is subsequently discovered that actually, there exists a trivial translation from $\mathcal{GF}$ to standard propositional logic by substituting *true* for $\mathcal{G}$, *false* for $\mathcal{F}$, $\wedge$ for $\mathcal{E}$, $\vee$ for $\mathcal{O}$ and $\neg$ for $\mathcal{N}$. Doesn't the trivial translation from $\mathcal{GF}$ to propositional logic mean that $\mathcal{GF}$ is nothing more than propositional logic with a bit of non-standard syntax? However, the recently obtained translation from ABA to LP raises precisely the same questions for ABA. For instance, one might wonder why nobody noticed this equivalence before, and whether the various ABA work that has been published over the years would also have been accepted had the authors chosen to present their work as plain logic programming.

On a more general level, there are also some tricky questions about the nature of progress in KR. Whereas in natural sciences, progress means a better understanding of the objective world, what precisely is it that constitutes progress in KR? Does progress simply mean that someone invents yet another formalism and starts to prove some properties of it? What if this formalism is actually equivalent with another formalism? What if it lacks any philosophical underpinning regarding the concepts it aims to model? What if its claim to "intuitiveness" appeals to only a select group of mathematically skilled people, with very few lay persons being able to figure out what is going on? If KR is ever to make any serious impact, some clear idea of what progress actually means would be highly desirable.

References

[1] A. Bondarenko, P.M. Dung, R.A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

[2] M.W.A. Caminada. Argumentation semantics as formal discussion. *IfCoLog Journal of Logic and its Applications*, 4(8):2457–2492, 2017.

[3] M.W.A. Caminada. Rationality postulates: Applying argumentation theory for non-monotonic reasoning. *IfCoLog Journal of Logic and its Applications*, 4(8):2707–2733, 2017.

[4] M.W.A. Caminada and Ch. Sakama. On the issue of argumentation and informedness. In *2nd International Workshop on Argument for Agreement and Assurance*, 2015.

[5] M.W.A. Caminada and C. Schulz. On the equivalence between assumption-based argumentation and logic programming. *Journal of Artificial Intelligence Research*, 60:779–825, 2017.

[6] P.M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.