Explaining presumptive arguments

This memo describes a method for generating Dung argument frameworks from Argument Interchange Format (AIF) argument graphs that maintains a direct relationship between S-nodes in the AIF argument graph and corresponding arguments in the argument framework. This allows extensions derived from evaluation of the argument framework to be simply related to subgraphs of the AIF argument graph. Since the AIF argument graph is linked data, and therefore connected (or connectable) to other sources of information, these extension AIF subgraphs can be interpreted in the context of linked data. This supports *explanation* of AIF argument graphs.

Argument framework

Dung [1] defines an argument framework as a pair $F = (A, D)$, where A is a set of arguments, and $D \subseteq A \times A$ is a binary relation on attacks between arguments. We say argument a attacks argument b if $(a, b) \in D$. The set of arguments $\{a \in A \mid a\ attacks\ b\}$ is denoted $\bar{b}$. An extension of a framework F is a set of arguments $\mathcal{E} \subset A$, and argument frameworks are evaluated by identifying which extensions are acceptable with respect to some semantics $\sigma(F) \subset 2^A$.

Argument Interchange Format

An AIF [2] argument graph G is a simple digraph (V, E) where

- $V = I \cup RA \cup CA$ is the set of nodes in G, where I are the I-nodes, RA are the RA-nodes, and CA are the CA-nodes; and
- $E \subseteq V \times V \setminus I \times I$ is the set of edges in G; and
- if $v \in V \setminus I$ then v has at least one direct predecessor and exactly one direct successor.

Constructing an argumentation framework from AIF

We construct a Dung framework $F(A, D)$ from an AIF argument graph G so that $A = S = RA \cup CA$; and $D \subseteq S \times S$ having attacks constructed according to the rules

1. Rebut: $s_a \in \bar{s_b}$, $s_b \in \bar{s_a}$ if $\delta_{ab} = 0$ and $Conc(s_a) = Conc(s_b)$
2. Undercut: $s_a \in \bar{s_b}$, $s_b \in \bar{s_a}$ if $s_a \in CA$ and $Conc(s_a) = s_b$
3. Undermine: $s_a \in \bar{s_b}$ if $s_a \in CA$ and $Conc(s_a) \cap Prem(s_b) \neq \emptyset$

where δ_{ab} is the Kronecker delta of the AIF node types; and $Conc$ and $Prem$ are functions that return the conclusions and premises of S respectively.

We are dealing with presumptive arguments here, where the canonical truth of premises and conclusions is often impossible to determine. In this context, an undercutting argument often raises questions or doubts about an inference rather that strictly defeating it; and a rebuttal often signals a difference of opinion rather than disproval of a conclusion. For these reasons, we make rules (1) and (2) symmetric so that rebuttals and undercuts are not necessarily accepted when the argument framework is evaluated. Rule (3) is asymmetric however. This makes arguments that challenge the premises of other arguments

"stronger" in the sense that they are not automatically counter-attacked. This encourages arguments rooted in evidence.

The construction creates a $1{:}\,1$ mapping between S-nodes in an AIF argument graph and arguments in an argument framework. Some evaluation of the argument framework will produce extensions $\mathcal{E}$ that can be directly mapped back to the corresponding set of S-nodes $S(\mathcal{E})$. The AIF argument subgraph corresponding to an extension can be induced from the original AIF argument graph by the set of nodes $S(\mathcal{E}) \cup Prem(S(\mathcal{E})) \cup Conc(S(\mathcal{E}))$.

Partition semantics

We can use semantic information in the AIF argument map to select or partition extensions of the argument framework. For example, we might want to label some I-nodes in the AIF argument graph as *hypotheses* $H \subset I$, and provide an explanation regarding some particular hypothesis $h \in H$. The problem here is that arguments, and their acceptability in terms of some extension semantics, are defined in terms of S-nodes rather than I-nodes. We therefore need to relate acceptable arguments to *acceptable information*.

We define the acceptable information relating to extension $\mathcal{E}$ as $I_{acc} = (Prem(\mathcal{E} \cap CA) \cap I \setminus Conc(S)) \cup Prem(S \cap RA) \cup Conc(S \cap RA)$. That is to say, we accept the premises and conclusions of RA-nodes, together with the premises of CA nodes on the condition that they are not also the conclusion of some other S-node. The set of hypotheses acceptable to an extension is $H_{acc}(\mathcal{E}) = I_{acc}(\mathcal{E}) \cap H$. Since $H_{acc} \in 2^H$, we can partition the set of extensions produced by an evaluation of the argument framework by acceptable hypotheses, so that a partition $P_{\acute{H}} = \{\mathcal{E} \,|H_{acc}(\mathcal{E}) = \acute{H}, \acute{H} \in 2^H\}$. We can then consider that the partition $P_{\{h\}}$ *supports* hypothesis h.

Notions of extension semantics can be extended to *partition semantics*. For example, if $h \in Conc(S)$ then there must be some $s \in RA$ in each extension of $P_{\{h\}}$ that has h as its conclusion, but this need not necessarily be the same RA-node in each extension. A partition *sceptically* accepts h without there necessarily being any argument that is sceptically acceptable in the extensions that make up the partition.

We designate an argument in $\mathcal{E}$ equivalent to an RA-node as *support*, and an argument equivalent to a CA-node as an *objection*. We can then define *necessary support* in a partition P as $\bigcap_{s \in P \cap RA} s$, *sufficient support* as $\bigcup_{s \in P \cap RA} s$, *necessary objections* as $\bigcap_{s \in P \cap CA} s$, and *sufficient objections* as $\bigcup_{s \in P \cap CA} s$.

Explanation graphs

By construction, $P_{\{h\}}$ contains all the acceptable arguments that support h. Note however that if $|H| > 1$ then each extension of $P_{\{h\}}$ must also include arguments that make any $\acute{h} \in H, \acute{h} \neq h$ unacceptable. In other words, there must be some $s \in CA$ in each extension of $P_{\{h\}}$ that has $\acute{h}$ as its conclusion. A partition $P_{\{h\}}$ therefore describes both the support for h and the objections to any hypotheses in competition with h. It can be considered an *explanation* for h.

We can induce an AIF *explanation graph* for $P_{\{h\}}$ in much the same way as for extensions above; but with the sets of RA-nodes and CA-nodes used in the construction drawn from a partition rather than an extension, and each collected according to either necessary or sufficient semantics as desired. For example, in explaining the case in favour of h, an explanation graph of sufficient support and necessary

objections is suitable; whereas in seeking to explain the weaknesses of h, the explanation graph of necessary support and sufficient objections is preferred.

An illustrative example

This section briefly introduces a worked example [3] that will be developed more fully separately.

The *Fortitude South* example considers arguments hypothesising about the site of the D-Day landings. There are 5 hypotheses: *C: Pas de Calais, N: Normandy, B: Brittany, P: Cotentin Peninsula, and O: Elsewhere*. Of these *B*, *P* and *O* are firmly ruled out by argument. The bulk of the remaining arguments are concerned with pitting *C* and *N* against each other as competing (but not mutually exclusive) hypotheses. This means that although $|H| = 5$, only two hypotheses can ever be acceptable so $|\acute{H}| = 2$.

The full AIF argument graph has 3 components: one rules out *O*, one rules out both *B* and *P*, and the third (and by far the largest) component concerns *C* and *N*. In general, we can evaluate each component of an AIF argument graph independently; and since the arguments strictly ruling out hypotheses are uncontroversial, we will omit them altogether from further discussion here.

We construct a Dung argumentation framework as above, and evaluate it with stable semantics [4]. This produces 48 extensions. Partitioning the 48 extensions by hypotheses gives 4 ($= 2^{|\acute{H}|}$) partitions: $P_{\{\},}\ P_{\{C\}}, P_{\{N\}}, P_{\{C,N\}}$, containing 8, 16, 8 and 16 extensions respectively.

We denote the explanation graphs for $P_{\{C\}}$ and $P_{\{N\}}$ by G_C^e and G_N^e, and consider the explanation graphs of sufficient support and necessary objections for each. The AIF argument graph G has a sub-argument, concerning evidence of troop dispositions and commands, which produces an *ORBAT* conclusion that supports *C* and objects to *N*. The *ORBAT* argument itself is acceptable to both $P_{\{C\}}$ and $P_{\{N\}}$; but in G_C^e it is connected to the graph component that contains *C*, and it is disconnected from the graph component in G_N^e that contains *N*. Both G_C^e and G_N^e have two other disconnected (from the component containing h) components in common: one relating to an argument about V-weapon sites that has been explicitly ruled out for consideration (but kept in the AIF argument graph for the purpose of documenting that fact), and one that concerns the need for a port, that is defeated by the argument (available only to the Allies) concerning Mulberry harbours. This characterizes components in an AIF explanation graph that do not contain h as acceptable but *irrelevant* information with respect to h.

Conclusion

This memo introduces a method to construct an argument framework from an AIF argument graph which allows the rich semantics of the AIF graph to be simply related to the extension semantics of the argument map. In doing so, the notion of partition semantics is introduced.

We propose that methods of scoring argument framework extensions for the purpose of decision making can form the basis of equivalent methods for scoring argument framework partitions, with the added benefit of incorporating notions such as irrelevance in extending those measures. Further work in developing partition semantics is suggested.

The method of constructing argumentation frameworks from AIF argument graphs creates an injective mapping that allows the results from analysing the argument framework to be readily interpreted in terms of AIF explanation graphs. This suggests opportunities in explainable AI.

There is also much to explore in exploiting the semantics of the AIF argument graph in partitioning the argument framework. For example, consideration of argumentation schemes such as *Argument from Consequences*, or consideration of the dialogical structure of arguments. These give context to the evaluation of an argument that may be important in decision-making.

References

1. Dung, P.M., *On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games.* Artificial intelligence, 1995. **77**(2): p. 321-357.

2. Bex, F., J. Lawrence, and C. Reed, *The Argument Interchange Format (AIF) Specification*. 2011.

3. Knox, A. *Operation Fortitude*. eleatics 2021; Available from: https://dstl.github.io/eleatics/argumentation/fortitude/.

4. Baroni, P., M. Caminada, and M. Giacomin, *An introduction to argumentation semantics.* The knowledge engineering review, 2011. **26**(4): p. 365-410.