COMMA 2010

Lower Bounds on Argument Verification

Gerard Vreeswijk, Intelligent Systems Group, Computer Science Department,
Faculty of Sciences, Utrecht University, The Netherlands.

Argumentation performed by computers: a problem

- Motivation behind computational models of argument: computers are sometimes better debaters than human beings.
- **Fact**. Argumentation software operates on formal grounds, and on formal grounds only.

  It is therefore unable to fathom the meaning of the arguments it processes.[a]
- However, the meaning of an argument tells a lot about its importance in the debate.
- A dispute conducted on formal grounds only, most likely will lead to the wrong outcome, for such a dispute misses out on important information: the meaning of the arguments has not been taken into account.

[a] This statement quickly leads to the Chinese room argument.

Example of the problem: conscientious and precise agent vs. agent replying with nonsense

- Suppose that A is a conscientious agent that is pretty sure that every argument it puts forward is double checked and makes sense.
- Suppose on the other hand that agent B is a formalist, that is programmed to play the argument game in a way that is formally correct.

  Otherwise, B does not care much about the contents of its arguments.
- We further stick to the rule that the one which has the last argument wins.
- B shall win all disputes, simply by countering all arguments of A with formal nonsense.

Solution: A must verify some of B's arguments. Question:

> What percentage of incoming arguments must A verify to keep the status assignment of arguments in line with the actual status assignment?

Same problem, but A now verifies a percentage of incoming arguments

1. Suppose agent A puts forward argument a, and asks for opposition to test a.
2. Suppose some other agent, B, responds with b to attack a.
3. Normally, argument b would defeat a so that in all argument semantics s, we would have $s(a) = 0$ and $s(b) = 1$.
4. If, however, A knows that only 80% of B's arguments are sound it may derive that a is out with probability 0.2.
5. To eliminate this uncertainty, A can decide to select b for verification.
6. Verification involves costs but may be beneficial if A has serious interests to be relatively sure about the status of a proper,

   for example, if assigning a wrong status to a would incur a heavy penalty.

Related problem: Tradeoff between cost of verification and cost of error

Problem:

> What percentage of incoming arguments must be verified if there is no prescribed error tolerance, but there is a tradeoff between cost of verification and cost of error?

This problem is more of an economic problem.
It has connections to optimisation in operations research and microeconomy.

Related work

Research on numeric status assignment to arguments.

Cayrol, Lagasquie-Schiex. **Graduality in argumentation**. In: *Journal of A.I. Research*, 2005.

Dunne, Hunter, McBurney, Parsons, Wooldridge. **Inconsistency tolerance in weighted argument systems**. In: *AAMAS* 2009.

Krause, Ambler, Elvang-Gøransson, Fox. **A logic of argumentation for reasoning under uncertainty**. In: *Computational Intelligence*, 1995.

Diego C. Martínez, Alejandro Javier García, and Guillermo Ricardo Simari. In: **An abstract argumentation framework with varied-strength attacks**. In *KR*, 2008.

Riveret, Rotolo, Sartor, Prakken, and Roth. **Success chances in argument games: a probabilistic approach to legal disputes**. In: *Conf. on Legal Knowledge and Information Systems*, 2007.

Difference with related work

- Most theories are concerned with uncertainty on attack relations, rather than with uncertainty on the arguments themselves.
- The domain of application of the above mentioned research is not argument verification, and deducing lower bounds for verification, as well as different ways to represent uncertainty in argument graphs.
- Dunne *et al.*'s AAMAS 2009 work is concerned is with the computational complexity of deriving precise lower bounds for tolerance in another setting. ⇒ computationally more involved.
- Cost-benefit analysis of (formal) argumentation:

Karunatillake and Jennings. **Is it worth arguing?**, In: *Springer Lecture Notes in AI*, 2004.

Paglieri and Castelfranchi. **Why argue? Towards a cost-benefit analysis of argumentation**. In: *Argument and Computation*, 2010.

Basic concepts

Players and agents

Definition.

- A player is a human being that uses an agent to represent his/her interests.
- An agent is a device that argues on behalf of a player by constructing arguments that are then exchanged with other agents (owned by other players).

1. The model assumes that argumentation takes place without immediate supervision of (or interference by) players.
2. The model assumes that players provide agents with the material to construct arguments.

   This is because agents cannot have knowledge of the semantics of arguments.

Soundness

Arguments themselves must satisfy certain standards in order to be entitled to play a role in the argumentation process at all.

Informal logic speaks of *admissibility*: must meet criteria of the "RSA triangle," viz. relevance, sufficiency and acceptability.

- The argument conforms to elementary syntactic standards. [For example, it is delivered in the agreed upon argument interchange format (AIF).]
- Inferences are instances of accepted rules of inference [known as *inference schemes*].
- Rules of inference have been applied correctly.
- Inferences connect logically from one sub-argument to the next.
- Variable-substitutions have been applied correctly
- ...

Verification rate

The model distinguishes two important events.

1. The event that a player decides not to accept an incoming argument at face value, and selects it to examine its structure and content.
2. The event that a player receives a sound argument.

1. It is assumed that verification is performed randomly at a certain rate θ at the control of the player, which is called the verification rate.
2. It is assumed that receiving a sound argument is Bernoulli distributed with parameter η.

   It is assumed that

   $$\Pr(\eta) = \Pr(\eta|\theta).$$

   It is possible to differentiate η per agent i, thus working with different probabilities η_i.

Justified acceptance

Definition. An argument is said to be justly accepted if it is accepted and sound.

	Sound	**Faulty**
Verify	$\eta\theta$	$(1-\eta)\theta$
Accept blindly	$\eta(1-\theta)$	$(1-\eta)(1-\theta)$

The probability that an accepted argument is justly accepted is

$$\alpha = P(\text{sound} \mid \text{accepted}) = \frac{P(\text{sound and accepted })}{P(\text{accepted})} = \frac{\eta}{1-(1-\eta)\theta}.$$

A helpful analogy, perhaps, is the customs (ymmv).

Probability that an argument, once accepted, is sound

incoming argument

argument is sound
η

argument is faulty
$1 - \eta$

accept verified argument
$\eta\theta$

accept unverified argument
$\eta(1 - \theta)$

reject verified argument
$(1 - \eta)\theta$

accept unverified argument
$(1 - \eta)(1 - \theta)$

Warning: do not confuse (as I did) the probability of justified acceptance with the probability that the status of an incoming argument is determined correctly. The latter is equal to:

$$1 - (1 - \theta)(1 - \eta).$$

Independence of justified acceptance of attackers

Assumption. The justified acceptance of an argument and its attackers (actual and forthcoming) are independent events.

This assumption of independence rests on two assumptions.

1. Players select arguments for inspection at random.
2. Soundness of attackers is independent.

The latter assumption is controversial, because one may expect that some arguments are more prone to rebuttal by unsound arguments than others. (Think of an irrefutable argument that is otherwise highly undesirable from a certain agent's point of view.)

Propagation of justified acceptance

Under the assumption of independence of justified acceptance of attackers (actual and forthcoming), the probability that a is justly accepted is given by

$$j(a) = \alpha \cdot \Pi_{b \in Att(a)}(1 - j(b))$$

- This equation is not a definition, but follows from the laws of probability.
- If $Att(a) = \emptyset$ we assume this product is 1 and $j(a) = \alpha$.

Example. Suppose $a \leftarrow b \leftarrow c \leftarrow d$.

In a Dung type argumentation system, $j(a) = j(c) = 0$ and $j(b) = j(d) = 1$. With a probability α that an argument is justly accepted, we have
$j(d) = \alpha$,
$j(c) = \alpha(1 - \alpha) = \alpha - \alpha^2$,
$j(b) = \alpha(1 - \alpha + \alpha^2) = \alpha - \alpha^2 + \alpha^3$,
and so forth.

If a is at the end of a linear defeat chain of length n, then

$$j(a) = \Sigma_{i=1}^{n}(-1)^{n+1}\alpha^n,$$

which goes to $\alpha/(\alpha + 1)$ for large n.

Probabilistic status assignment

Definition. The process of incorporating probabilities of acceptance due to not verifying unsound arguments is called a probabilistic status assignment.

Definition.

- The error of a probabilistic status assignment j on a particular argument a is defined as $|s(a) - j(a)|$, where $s : \text{Arguments} \to \{0, 1\}$ is a conventional status assignment on arguments.
- For a set of arguments, the error is defined as the supremum of the errors of the individual arguments.

RESULTS

Step 1: Probabilistic status assignment is continuous

In particular, the following two properties hold.

1. If, for at least one $b \in Att(a)$: $j(b) \rightsquigarrow 1$, then $j(a) \rightsquigarrow 0$.
2. If, for all $b \in Att(a)$, $j(b) \rightsquigarrow 0$, then $j(a) \rightsquigarrow 1$.

This claim follows from the continuity of functions of the form

$$f : [0,1]^n \rightarrow [0,1] : (x_1, \ldots, x_n) \mapsto \alpha \cdot \Pi_{i=1}^{n}(1 - x_i)$$

for every integer $n \geq 0$ and every real constant $0 \leq \alpha \leq 1$.

The two above properties inductively apply through attack chains of finite length, hence to finite (or finite parts of) Dung graphs.

Step 2: Verification rate can be used to control acceptance errors

Claim. Suppose $\eta < 1$ and the length of attack chains is bounded. Then for every $\epsilon, \eta > 0$ there exists a verification rate $\theta > 0$ such that the error on probabilistic status assignments does not exceed ϵ.

Proof. Let $\epsilon, \eta > 0$ be given. Thanks to the continuity of propagated errors and the boundedness of attack chains there exists an $0 < \alpha \leq 1$ such that the error of status assignments j on individual arguments does not exceed ϵ. Solving

$$\alpha \geq \frac{\eta}{1 - (1 - \eta)\theta}$$

for θ yields a verification rate θ. $\square$

The above claim is qualitative: given $\epsilon, \eta > 0$, it does not specify a specific value for θ. It only specifies that such a value exists.

Step 3: To find a specific lower bound for θ

Probabilistic status assignment satisfies the following two properties.

1. If $j(b) \geq 1 - \epsilon$ for at least one $b \in Att(a)$, then $j(a) \leq \alpha\epsilon$.
2. If $j(b) \leq \epsilon$ for all $b \in Att(a)$, then $j(a) \geq \alpha(1 - \epsilon)^n$.

Now,

1. Item (1), is not problematic, since the probability of acceptance remains smaller than ϵ. This probability may then be propagated further into the defeat chain.
2. Item (2), is more problematic because the error (the difference between pure defeat status and probabilistic defeat status) increases every step of the defeat chain, and there is nothing we can do about it, i.e., $\alpha(1 - \epsilon)^n$ is the greatest lower bound on acceptance in case all attackers remain undefeated with probability ϵ.

The double dip function

The attention will be focused on

$$\phi(x) = \alpha(1-x)^2.$$

- The function ϕ^2 can be used to estimate the error in an elementary defence chain of length two. A problem with ϕ^2 and iterations of it, is that it is difficult to seize its amplitude for repeated applications of ϕ:

$$\begin{aligned}
\phi^0(\alpha) &= \alpha \\
\phi^1(\alpha) &= \alpha(1-\alpha)^2 \\
\phi^2(\alpha) &= \alpha(1-\alpha(1-\alpha)^2)^2 = \alpha - 2\alpha^2 + 5\alpha^3 - 6\alpha^4 + 6\alpha^5 - 4\alpha^6 + \alpha^7 \\
&\vdots
\end{aligned}$$

- One way to predict the development of these expressions is to use generating functions $\Rightarrow$ Does not work because non-uniformity.

Lower bound on iterations of double dip

Find a simple (preferably polynomial) function on $[0, 1]$ that acts as a lower bound on iterations of ϕ^2.

It is possible to handcraft one for $n = 2$. (Can later be generalised to fixed $n > 2$.) To this end, there is the following result.

Lemma. Let $\phi(x) = \alpha(1 - x)^2$. Then for every function $f : [0, 1] \to R$,

$$x^\kappa \le f(x) \quad \Rightarrow \quad x^\kappa \le \phi^2 f(x).$$

where $\kappa = 3.81884154$.

The proof amounts to verifying that

$$\alpha(1 - \alpha(1 - \alpha^\kappa)^2)^2 \ge \alpha^\kappa,$$

for $0 \le \alpha \le 1$, which is indeed the case.

Lower bound on iterations of double dip is tight and unique

Find a simple (preferably polynomial) function on $[0, 1]$ that acts as a lower bound on iterations of ϕ^2.

Lemma. Let $\phi(x) = \alpha(1 - x)^2$. Then for every function $f : [0, 1] \to R$,

$$x^\kappa \leq f(x) \quad \Rightarrow \quad x^\kappa \leq \phi^2 f(x).$$

Remarks:

1. The inequality is tight. That is, there is a value for α for which the inequality becomes an equality. (This is for $\alpha = 3/4$.)
2. The inequality does not hold for other values of κ. (For other values of κ, the function $\alpha(1 - \alpha(1 - \alpha^\kappa)^2)^2 - \alpha^\kappa$ possess two zero points near $\alpha = 3/4$.)

Thus, there do not exist better lower estimations of the form x^κ, and κ is unique.

Main result

Theorem. Suppose the length of attack chains is bounded, and all arguments have at most two attackers. Let $\epsilon, \eta > 0$ be given, and let θ be the verification rate. If

$$\theta \geq \frac{1 - \eta(1-\epsilon)^{-1/\kappa}}{1-\eta}$$

where $\kappa = 3.81884154$, then the error of status assignments j on individual arguments is not greater than ϵ.

Two different types of attack structures are considered: attack structures that are named saturated and those that are not. An attack structure is called *saturated* if all attack chains have length $2n$, for some $n \geq 0$.

A saturated attack structure causes the largest errors in a probabilistic status assignment.

Case I. Saturated. Above lemma. *Case II. Other.* Peel off paths length $\in \{1, 2\}$.

Learning and optimisation

Learning η_i to control ϵ

Idea:

- Errors (arguments unjustly accepted) involve costs.
- Sampling is a human activity that involves costs as well.

Thus, there is a tradeoff.

– Cost of verification $+$ reliability of opponent $\Rightarrow$ verification rate.

– The probability η_i that an argument received from player i is sound may be estimated through exploration (by sampling arguments manually for verification).

– Since verification is a human activity, it is unreasonable to assume that verification costs linearly depend on the verification rate. Instead, the law of diminishing marginal returns for human labour prescribes that verification has increasing relative costs, say $C_1\theta^\beta$, for some constant $C_1 > 0$ and $\beta > 1$.

Cost of errors

It is assumed that the cost of errors is C_2 per error per argument, $C_2 > 0$. The expected cost is then

$$C(\theta) = C_1\theta^\beta + C_2\epsilon$$

Because the error ϵ can be controlled by θ, the expected cost effectively depends on θ and η. In this way, the expected cost per argument would be

$$C(\theta) = C_1\theta^\beta + C_2(1 - (\theta + \eta - \theta\eta)^\kappa) \quad (1)$$

where the right term is obtained by solving the inequality in the main theorem for ϵ, and $\kappa = 3.81884154$.

Example

$C_1 = 8$, $C_2 = 9$, $\beta = 2$ and $\eta = 0.65$.

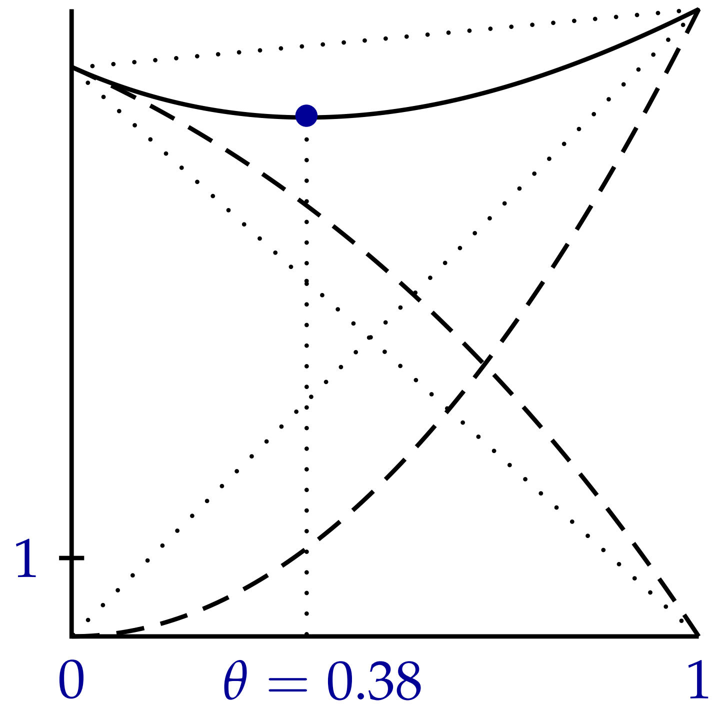

Fig. 1: Tradeoff between cost of verification and cost of errors.

- The cost function is not always convex.
- If it is, then there is a proper tradeoff, and cost is minimised for the unique stationary point for which $dC(\theta)/d\theta = 0$ ($\theta = 0.38$ in the example).
- If the law of increasing costs of human argument verification does not hold, then the cost function is never convex, and interesting tradeoffs would never occur.