Game Theoretic Machine to Machine Argumentation

Henry Hexmoor and Arnab Sinha, 2010

1.0 Introduction

Argumentation is one of the oldest research foci and one of the most enduring ones in Artificial Intelligence (Bench-Capon and Dunne, 2007; Rahwan and Smari, 2009) and in parallel in Philosophy, first in (Toulman, 1958) and now in (Pollock, 2010). Abstract argumentation has been a rich and varied new discipline that started with (Birnbaum, 1982) but widely credited to (Dung, 1995). It has been adapted to many domains including computational law (Dunne, 2003) and multiagent negotiations (Fatima, et. al., 2004). Currently, the most vigorous and prolific research group is Argugrid (www.argugrid.eu) , which is a grid based research consortium funded by EU and directed by Dr. Francesca Toni of Imperial College in London, UK.

Dung's framework prescribes a set of arguments A and an attack relation R among them. This binary relation is often denoted as → between a pair of arguments. Pollock's inference graphs (Pollock, 2010) are very similar to graphs produced by depicting Dung's attack relationships. For brevity, we list the main properties for set A without elaborate notations and detailed explanations that needlessly obscure the essence of discussion.

1. A' is *acceptable* with respect to a set of A arguments. I.e., Every argument in A' is defendable against an attack. This is assured by having arguments in the set complement A – A' protect arguments in A' by attacking possible offending arguments. This is a rather common event in society. This is how in-groups emerge (Pennington, 2002).

2. A' is *conflict-free*. I.e., there are no attack relations among any pairs of arguments in A'. This is less common than acceptability property but in-groups exhibit this phenomenon.

3. A' is *admissible*. I.e., Once A' is conflict free, if A' is acceptable, then every argument in A' must be acceptable with respect to A. Finally,

4. A' is a *stable extension* of A. I.e., Once A' is conflict free, iff it attacks every argument in the complement set A – A'. This property appears to identify a xenophobic tendency that is an unreasonable fear or hatred of the unfamiliar. As such, this notion of stability is rather an odd fit for a scientific endeavor.[1]

An interested reader may consult a large number of student theses ranging from 2000-present that offer rather voluminous, and repetitious background, e.g., (Rahwan, 2004; Karunatillake, 2006). Of note is the recent trend to employ game theory for strategic dialogue (Procaccia

[1] We have traced this notion to the long standing *Arab-Israeli conflict* example elaborated in (Birnbaum, 1982). This may account for this bizarre characterization. A similar example is *Hatfield-McCoy* feud (1878 – 1891), the account of American folklore that has become a metaphor for bitterly feuding rival parties in general.

and Rosenschein, 2005; Rahwan and Larson, 2008). A particular dialectic style is strategic *dialogue* that is a turn taking setup where two agents promote their own argumentation. We generalize this to N agents who simultaneously arrive at an argumentation arena and simultaneously engage in generalized multi-way dialectic; i.e., a form of reasoning based on the exchange of arguments and counter-arguments.[2] In section 2 we state our assumptions and goals for M2M. Our formalism is presented in section 3.

2.0 Motivation

We will extend the state of the art game theoretic argumentation protocols for machine to machine scenario.

In a departure from previous work, we wish to address argumentation among computers with no human intervention, i.e., M2M argumentation. M2M differs from previous approaches with the following assumptions:

1. No human mediation, supervision, or intervention is allowed.
2. Each agent, in parallel with peers, asynchronously will have the opportunity and capacity to initiate or further promote their own argument.
3. Agents will not collude or have private agenda.
4. All arguments are performed in public.
5. Time will be modeled in discrete units. An agent's public argumentation activity will take shorter than one time unit. Each agent may take as much time as needed to ponder about arguments.
6. There are no overall winners or losers.
7. Based on apriori pay off functions, agents will receive payoffs proportional to strength of their ending position. These values are computed in an anytime algorithmic fashion.
8. There are no fixed, externally imposed, termination conditions.
9. Termination for each agent is independently and internally determined.

2.0 Background

Argument is a social and verbal means of trying to resolve or at least to contend with, a conflict or difference that has arisen or exists between two (or more) parties. An argument necessarily involves a claim that is advanced by at least one of the parties. The study of argumentation may, informally, be considered as concerned with how assertions are proposed, discussed, and

[2] Strictly speaking, we may intend to the Sanskrit concept of उपाय upāya rather than Dialectic.

resolved in the context of issues upon which several diverging opinions may be held. Argumentation by definition refers to a coherent series of statements leading from a premise to a conclusion by analysis and evaluation of the given arguments.

Computer scientists, especially AI researchers has been motivated by the argumentative study since their existence but its seminal work began to emerge in the last ten years as an important sub-discipline of AI. There have been significant contributions resulting from this, including approaches to modeling and analysis of defeasible reasoning, formal bases for negotiation and dialogue processes in multiagent systems, and the use of argumentation theory in AI applications whose nature is not best described through traditional logics, e.g. legal reasoning, evaluation of conflicting beliefs, etc. The process of interpreting and exploiting classical treatments of Argumentation Theory in effective computational terms has led to a rich interchange of ideas among researchers from disciplines such as Philosophy, Linguistics, AI and Economics. While work over recent years has done much to consolidate diverse contributions to the field, many new concerns have been identified. (Dunne and Bench-Capon, 2006)

Majority of the models developed for automated negotiation can be classified primarily into three major approaches: game theoretic, heuristic-based and argumentation-based. Argumentation-based negotiation has a prominent advantage over the others for being more dynamic and flexible to adopt changes to its positions while negotiating. (Procaccia and Rosenschein, 2005)

Argumentation-based negotiation has undergone extensive research and is a field by itself encompassing many techniques. Among many, Abstract Argumentation is a well known framework formulated by Dung (Dung, 1995). It has a set of arguments with attack relations between them.

Game theory came was initiated with into being with (von Neumann and Morgenstern, 1944). GT has widely proliferated into mathematical and computational sciences as well as social sciences, economics, biology, and philosophy. GT approach for argumentation can be seen in extensive forms of a series of offers and counter-offers. The set of agents or players engaging in the negotiation dialogue is likened to a game having a set of actions/strategic decisions available to each and pay-offs. Each agent competes strategically to choose an action in order to secure an optimized payoff, a kind of outcome that is expressed in some other numerical way of quantifying the values of the pay-offs. (Walton, 1998).

Broadly, GT can be incorporated in argumentation in the following ways:

1. **Extensive game**: Analysis of how the agents work together in particular strategic settings and predict the outcome.
2. **Mechanism design**: Design rules of the game in such a way so as to make self-interested agents behave in prescribed manner.

Both approaches are pursued within multi-agent systems (Nisan, et. Al., 2007). On the one hand, an agent analyses its given situation before making a choice while on the other hand mechanism design may be used to design the argumentation protocol in a way to promote good argumentative behavior.

Abstract argumentation (Dung 1995) has a dialectic approach that models conflict between arguments and formalizes the acceptability of arguments. Henceforth several works e.g. (Krause *et al.*, 1995; Jakobovits and Vermeir, 1999; Besnard and Hunter, 2001; Cayrol and Legasquie-Schiex, 2005) have devised larger number of categories or continuous numerical scales for the argument set. (Matt and Toni, 2008) assesses the strength of arguments on a scale of values ranging from 0 to 1. An argument is called strong whenever the argument can be defended by one or several well-formed opinion(s) that properly withstand(s) external criticism. To assess the strength of an argument in a given dispute, a proponent's and an opponent's opinions are weighed for or against the argument. A notion of argument strength matching is defined in terms of the value of a game of strategy (Borel, 1921; von Neumann, 1928; von Neumann and Morgenstern, 1944) confronting two fictive players endorsing the roles of proponent and opponent of an argument.

An extensive form of GT is found in (Procaccia and Rosenschein, 2005) where they consider both two agents take turns in advancing their argument. This enables the use of existing methods from game theory for solving the associated extensive game. Every agent is possesses a set of a arguments. A binary attack relation exists between arguments with a payoffs for every possible valid dialogues. The agents don't argue between themselves with the intention of winning or losing. Rather than an absolute win, their argument results with a quantitative payoff, which reflects the extent of an agent's argumentation strength. Relations are depicted in the form of an interaction graph where a cyclic graph represents the game to be infinite.

This model of negotiation framework (Procaccia and Rosenschein, 2005) considers multiple agents but at any instance only two agents can participate augment. What if a third agent has a valid argument to put forward on behalf of a weaker agent? In other words situations can happen where multiple agents must argue simultaneously to agree to a common viewpoint. We envisage to incorporate trustworthiness of agents as a function of the success of agents. It can be related like an agent being more trustworthy than another if it is more successful to defend itself while augmenting.

3.0 Formalism

In this section we formalize our assumptions from section 21 into an M2M framework.

Definition 1. Let M2M = <AR,→,ARi,U> be a M2M system where AR is a set of arguments, → is a set of binary contention relationships among arguments, For i = 1, 2, …, n: $AR_i \subset AR$, and AR_i is finite. Furthermore, there exist $t_i \in AR_i$ such that t_i attacks all arguments in AR_i but is attacked by none. U is a utility function U, which assigns payoffs to each agent for each possible

terminated argumentation cycle (Nisan, et. al., 1997); that is any dialogue that ends either with a t_i. Game theory does not usually provide depictions and details of multidimensional payoffs.

The following five properties are adopted and generalized from (Procaccia and Rosenschein, 2005).

1. An argument a ∈ AR is locally-acceptable with respect to a set of arguments S iff for each argument b ∈ $\cup AR_i$: if b attacks a then b is attacked by S.
2. A conflict-free set of arguments S is locally-admissible iff each argument in S is locally-acceptable with respect to S.
3. A locally-preferred extension of A is a ⊆-maximal locally-admissible set of A.
4. A conflict-free set of arguments S is called a stable extension iff S attacks each argument in $\cup AR_i$, which does not belong to S.
5. A locally-admissible set A' of arguments is called a locally-complete extension iff each argument, which is locally-acceptable with respect to A', belongs to A.

Definition 2. An *proposal cycle* is a collection (a1, a2, ... , a_l) of l arguments proposed in a single time unit by any and all agents. It is probable for an agent to abstain from proposing an argument in order to compute outcomes of existing and anticipated arguments. Any agent can terminate the argumentation cycle at any time by issuing its special t_i argument.

Definition 3. An argumentation *resolution cycle* is a single time unit in which an agent attempts to solve the argumentation game in order to determine dominance or equilibria. Such an agent abstains from suggesting a new proposal.

4.0 Proposed Tasks

We believe an N dimensional extension of *extensive* game theoretic argumentation to a major first leap step towards full M2M argumentation. This task will require supervision of one PhD student for two years at estimated the cost of $150K.

Going beyond dialogue and designing a mechanism is at the core of our second task. This task will require supervision of one PhD student for two years at estimated the cost of $150K.

Overall, our two tasks are modestly designed despite their major impacts.

References

1. T.J.M. Bench-Capon and P.E. Dunne , 2010. Argumentation in artificial intelligence, Volume 171, Issues 10-15, July-October 2007, Pages 619-641.
2. L. Birnbaum, 1982. Argument Molecules: A Functional Representation of Argument Structures.

3. P. E. Dunne. (2003). Prevarication in dispute protocols. In G. Sartor, editor, *Proceedings of the 9th International Conference on AI and Law (ICAIL)*, New York, NY, USA, ACM Press.
4. P. M. Dung. (1995). On the Acceptability of Arguments and its Fundamental Role in Non-monotonic Reasoning, Logic Programming and N-person Games. In *Artifcial Intelligence*, 77(2):321.358.
5. S. Fatima, M. J. Wooldridge, and N. R. Jennings. (2004). An agenda based framework for multi-issues negotiation. *Artificial Intelligence Journal*, 152(1):1.45.
6. I. Rahwan, (2004). Interest-based Negotiation in Multi-Agent Systems, PhD dissertation, University of Melbourne.
7. N.C. Karunatillake (2006). Argumentation–Based Negotiation in a Social Context, PhD dissertation, University of Southampton.
8. J. von Neumann, O. Morgenstern, 1944. Theory of Games and Economic Behavior, Princeton University Press.
9. Nisan, et. al., 2007. Algorithmic Game Theory, Cambridge University Press.
10. J. Pollock, (2010). Defeasible Reasoning and Degrees of Justification, in Argument & Computation, Taylor and Francis.
11. A.D. Procaccia and J. Rosenschein (2005).In Proceedings of 3rd European Workshop on Multi-Agent Systems, pp. 312-322.
12. A. Procaccia, (2010). Extensive-Form Argumentation Games, Hebrew University of Jerusalem.
13. I. Rahwan and K. Larson (2008). Pareto Optimality in Abstract Argumentation, In Twenty-Third AAAI Conference on Artificial Intelligence.
14. I. Rahwan, and G.R. Simari, G. R. (Eds., 2009). Argumentation in Artificial Intelligence. Springer.
15. P.E. Dunne and T.J.M. Bench-Capon (2006). In Proceedings of Proceedings of COMMA on Computational Models of Argument, pp. 352.
16. D.C. Pennington, 2002. The social psychology of behaviour in small groups, Psychology Press.
17. S. Toulman, 1958. Uses of Arguments, Cambridge University Press.
18. D. N. Walton, 1998. The new dialectic: conversational contexts of argument. University of Toronto Press.