BDI and BOID Argumentation
Some examples and ideas for formalization

Guido Boella
Dipartimento di Informatica
Università di Torino
Italy
guido@di.unito.it

Leendert van der Torre
SEN3
CWI Amsterdam
The Netherlands
torre@cwi.nl

Abstract

In this discussion paper we present some preliminary ideas on the role of argumentation in the context of BDI and BOID agents, i.e. agents whose deliberation is based on beliefs, obligations, intentions and desires. More precisely, we identify several argumentation issues which do not occur in argumentation concerned with only the information or knowledge of single agent systems. We distinguish among argumentation issues for single agent deliberation, multiagent dialogues, and interaction between agents and their normative system. For each category we present some examples, and some ideas about their formalization.

1 Introduction

Formal argumentation has focussed on deliberation about information and knowledge, for example as the proof theory of default logic. We are interested in formal argumentation in the context of cognitive agents, which has been pioneered by [**parsons:jlc98 ?**]. We are interested in BDI and BOID agents, i.e. agents whose deliberation is based on beliefs, obligations, intentions and desires. In this paper we present some preliminary ideas on different issues in argumentation and dialogue in BOID setting, which do not occur in single agent argumentation concerned only with information and knowledge. In particular, we distinguish the following questions:

1. Which kinds of argumentation issues can be distinguished in BDI/BOID agents?
2. Which kinds of argumentation issues can be distinguished in dialogues among BDI/BOID agents?
3. Which kinds of argumentation issues can be distinguished in interactions between a BOID agent and its normative system?

We use rule based logics such as input/output logic, and rule based architectures such as used in BDP logic [**thomason:kr00 ?**], in the BOID architecture [Broersen *et al.*, 2002] and in our previous research [Boella and van der Torre, 2003].

The layout of this discussion papers follows the three questions above. In section 2 we discuss which kinds of arguments. In Section 3 we discuss which kinds of dialogues. In Section 4 we discuss which kind of interaction with normative systems.

2 Arguments

In this section we consider arguments provided by a BOID agent to justify its decisions in terms of its beliefs, desires, intentions and obligations.

2.1 Examples

Consider the following example of reasoning with beliefs, obligations, intentions and desires:

1. You want to go to Acapulco for holidays;
2. You have to spend little money;
3. You intend to go to conference in Acapulco;
4. You believe that combining conference and implies spending little.

Now, assume you go to Acapulco and someone asks you why did so. Now you have to reconstruct an argument, which may or may not be based on the motivation you really used to go to Mexico. In this case, you may tell that you want to go to Acapulco, and that was the reason why you did so. Alternatively, he may say that:

- You have the normative goal to spend little;
- You believe that this can be achieved by combining conference and holidays;
- You already intend to go to conference in Acapulco.
- You therefore spend your holidays following IJCAI 2003.

In the first explanation, your desire immediately implies the decision to go to Acapulco. In the second explanation, you derive from the normative goal to spend little, that you will combine conference and holidays. Note that this inference is not by application of a belief rule, only by application of the inverse of a belief rule.

This is a very familiar situation, it is planning based on abduction as it has been studied since decades. However, the new issue in this example is that this abduction to find a plan is combined with deduction to find a goal. For example, the first line may be replaced by:

3. If the budget is nearly finished, then you have the normative goal to spend little; You believe the budget is nearly finished;

The example can also be extended by replacing the third line by:

3. You intend to go to Acapulco if you submit a paper and this paper is accepted

In this case, the logic should imply that you will submit paper. Submitting a paper does not imply that you will go to Acapulco, but it is a necessary precondition to complete the above argument.

2.2 Notes on formalization

To model the example, we can formalize an argument as a set of rules. For example, in input/output logic [Makinson and van der Torre, 2000], we may define an argument for p in context c as a minimal set of rules R such that $p \in out(R, c)$. To formalize the various kinds of rules we must distinguish between the various mental attitudes. For the extension of the example with the submission of a paper, we need decision variables and a way to deal with uncertainty, because submitting paper does not necessarily imply that you go there.

It is a common misunderstanding that we need modal logic to formalize such examples. As shown by for example BDP logic [**thomason:kr00 ?**], the BOID architecture [Broersen *et al.*, 2002] and in our previous research [Boella and van der Torre, 2003], we do not have to introduce modal logic. We should distinguish between the logics of the various attitudes. For example, with rule set $R = \{(c, p)\}$ we have argument for p in context c, regardless whether R stands for beliefs or obligations, but not always an argument for $c \wedge p$. The latter may make sense for beliefs, but definitely not for obligations, desires and intentions. And in fact, this reflects the way belief rules and motivational attitude rules are used by a BOID agents. Beliefs rules are iteratively applied to compute consequences of actions from initial states. In contrast, desire and goal rules are used to value states by checking which rules are applicable in a given state but not consistent with it.

As a first sketch of formalization we list the rules involved in the example above - by $R(l_1 \wedge l_n \to l)$ we mean that the set R contains the rule having as preconditions $l_1 \wedge l_n$ and l as consequence. The set of rules involved in the example are the beliefs B, the desires D, the intentions I and obligations O:

1. $D(\top \to va \vee vk)$ I want to go to Acapulco or Kazakstan for holidays
2. $O(\top \to \neg s)$ I am obliged to spend little money
3. $I(\top \to ca)$ I intend to go to conference in Acapulco
4. $B(ca \wedge va \to \neg s)$ I believe that combining conference and implies spending little

3 Dialogues

In this section we consider the case of dialogues between two agents who are arguing on some topic in order to take a decision. In this context considering desires, intentions and obligations besides beliefs raises several new issues.

3.1 Examples

The first example shows how the information about desires, intentions and obligations which an agent puts forwards can be reused in the other agent's counter arguments.

A: I want new computer

B: we cannot afford it now

A: OK

B: I want to go to acapulco, because we have to spend budget

A: if we have to spend budget, then we can afford new computer

As in the previous example B's argument in favor of going to the IJCAI 03 conference uses the inverse of a belief rule (going to Acapulco is the means to spend budget). Moreover, the goal which is achieved by going to the conference is justified by the obligation to finish spending the budget: agent B adopts the content of the obligations as its goal.

However, by mentioning the obligation B opens the way to A's counter argument: if B had not mentioned that he wants to go to Acapulco, A would not have had a good argument to get his computer.

Note that A's counter argument implicitly presupposes that going to the conference and buying a computer cannot be done at the same time and that buying computers is more important than going to conference. This reflects the existence of a mechanism in a BOID agent to resolve conflicts between incompatible rules.

In single agent argumentation it is not possible to observe a new phenomenon which arise when two agents dispute with each other: it is possible not only that a new argument changes the beliefs of the adversary but also a new argument can lead an agent to reconsider its intentions.

In the next example we return on the decision to go to Acapulco: B tries to change A's intention by suggesting another option he did not consider:

A: I want go to Acapulco for IJCAI 03 since I want to make some holidays without spending too much

B: There is also a conference in Kazakistan and going to Kazakistan is less expensive than going to Mexico

For a BOID agent an intention is the result of the optimal decision given the facts at stake. When agent B provides new information the optimal decision for agent A changes: the goal of not spending too much is better achieved by going to Kazakistan, provided that it is still an option for enjoying holidays.

But the revision of a decision can be induced also by an agent's actions. For example providing new evidence:

A: I want to smoke

B: Smoking is not healthy

A: I know that

B: here you see some nice pictures

A: I lost my appetite

Since desires have a conditional character they can be activated by making their conditions true. In this example B tries to activate the desire to stay alive by recalling A that smoking is not healthy. Unluckily this desires has been already been considered by A in its decision to smoke. In order to make A change its decision is necessary to enable some other desire in favor of not smoking. Since people usually fear to die when they realize how painful it can be, B shows A some medical picture of lung diseases.

Similar arguments are possible also when A shows to be uncertain about what to do:

A: I desire to go home and I desire another beer

B: Do you want another beer?

A: OK

Or, conversely, a real friend:

A: I desire to go home and I desire another beer

B: Is your girlfriend not worried?

A: OK, I go home

Nail and hammer example of Parsons, Sierra etcetera (check details)

A wants to drive in nail

B has hammer

3.2 Notes on formalization

First example illustrates that communication can be strategic: by making argument you also inform the other about which desires, intentions and obligation one have.

Second example illustrates cooperation to taking a decision in a group. New information by one party can modify the decision of the group. This example presupposes some form of conflict resolution in order to take an alternative among different alternative incompatible solutions, such as the one described in [**broesen:csq ?**]).

The remaining examples illustrate manipulation by giving the other agent another option. There are two possible ways: the agent can only give another option, and trust the decision making of the other agent to reconsider - the agent does not trust the decision making of the other agent, and explains to him in detail why alternative is better (in real life, there are good reasons for the first option (if the other agent finds out himself, he thinks it is his own idea and will do it) and for the second option (most people will not succeed in finding the alternative).

The first example of the section above can be modelled by keeping two distinct sets of rules for each agent A and B and for the beliefs and obligations they share:

1. $D_B(\top \rightarrow c)$ B desires a new computer
2. $B_{AB}(c \rightarrow s)$ The computer is expensive
3. $D_{AB}(\top \rightarrow \neg s)$ We cannot afford it
4. $O_{AB}(\top \rightarrow s)$ we have to spend budget
5. $B_{AB}(a \rightarrow s)$ acapulco means spending budget, i want to go to acapulco to spend budget
6. $D_A(\top \rightarrow a)$ I want to go to acapulco

The example about influencing decisions requires the following beliefs, desires and intentions.

1. $I_A(\top \rightarrow s)$ A wants to smoke
2. $D_A(\top \rightarrow s)$ A desires to smoke (support the intention above)
3. $B_A(s \rightarrow k)$ Smoking kills
4. $B_A(p \rightarrow h)$ Pictures are horrible
5. $D_A(h \rightarrow \neg k)$ horrible means fear to die

Note that the last desire of agent A is a conditional one: unless its precondition is not true, the desire cannot be counted among the satisfied nor among the unsatisfied ones. Hence, in order to make A take this desire into account in its decision whether to smoke B must decide to show some pictures which recall A how bad is falling ill. Once agent A reconsiders its intention to smoke, it may form an intention not to smoke.

4 Normative dialogues

In this section we consider dialogues involving sanction based obligations as they are defined in [Boella and van der Torre, 2003]. The point is that agents cannot be presumed to comply with obligations: agent may or may not be respectful - i.e., they do what they are obliged to do. Selfish agents who are not respectful must be motivated to respect norms by sanctions. In [Boella and van der Torre, 2003] sanctions are not mere consequences of violations. Rather they are the actions of the normative system, whose reaction must be taken into account in the discussion.

4.1 Examples

In this example agent B tries to make A reconsider his intention to smoke by recalling him, first, that there is an obligation not to smoke; second, that the violation of the obligation is punished. The first argument is rejected by A since it is not a respectful agent, the second one is rejected on the ground that who is in charge of punishing violations at the moment is not able to do that.

A: I want to smoke

B: if you smoke you violate an obligation

A: i dont' care

B: But you got a fine of 100euro

A: The policeman is busy and hence he cannot apply the sanction

4.2 Notes on formalization

To have dialogues with normative system, it is useful to attribute mental attitudes to the normative system, thus dealing with it as normative agent. In order to take a decision, the agent A who is subject to the obligation has to consider the reaction he expects the normative agent N will have. This reaction is computed by recursively model N's decision using the beliefs, desires, intentions attributed to N. The example above involves the following mental attitudes of both agent A and N (the set of rules of the two agents are distinguished by the index A/N):

1. $I_A(\top \rightarrow c)$ I want to smoke
2. $D_N(c \rightarrow v)$ Smoking counts as a violation of some norm for the normative agent N
3. $\neg D_A(\top \rightarrow v)$ agent A does not care of being a violator
4. $D_N(v \rightarrow s)$ violators are punished with a fine
5. $I_N(\top \rightarrow b)$ the normative agent has currently other intention
6. $B_N(b \wedge s \rightarrow \bot)$ the normative agent cannot both sanction and do other things (b).

5 Conclusion

In this paper we argue that argumentation in the context of BDI and BOID agents raises new issue. First of all an argument of a BOID agent may involve reference not only to its beliefs but also to its desires, intentions and obligations. Second, in disputes between BOID agents desires, intentions and goals of both agents can be used as pros and cons; moreover, agents try not only to make the adversary change its beliefs but they can try to make him reconsider its intentions. Third, when we consider also sanction based obligations, the agents must take into account in their discussion also the beliefs, goals and intentions of the normative agent who is in charge of monitoring and sanctioning violations.

Further issues to be addressed are considering how agents face the problem of conflict resolutions among their mental attitudes when they form their optimal decisions. Finally, the scenario involving normative dialogues becomes more complex when we consider hierarchical normative systems composed of agents playing different roles.

Acknowledgement

Some of the examples discussed have come up in discussions of the second author with Yao-Hua Tan and Martin Caminada.

References

[Boella and van der Torre, 2003] G. Boella and L. van der Torre. Your wish is my command: Sanction-based obligations in a qualitative decision theory. In *Procs. of AAMAS 03 Conference*, Melbourne, 2003. ACM Press.

[Broersen *et al.*, 2002] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal generation in the boid architecture. *Cognitive Science Quarterly*, 2(3-4):428–447, 2002.

[Makinson and van der Torre, 2000] D. Makinson and L. van der Torre. Input-output logics. *Journal of Philosophical Logic*, 29:383–408, 2000.