An Argumentation-based Computational Model of Trust for Negotiation

Maxime Morge[1]

Abstract. The fact that open multiagent systems are vulnerable with respect to malicious agents poses a great challenge: the detection and the prevention of undesirable behaviours. That is the reason why techniques such as trust and reputation mechanisms have been proposed. In this paper, we explore the cognitive science background which captures the notions of trust, reputation and confidence to provide a computational trust mechanism applied to negotiations within artificial societies. For this purpose, we formalize here these notions and we apply to them a particular argumentation technology for allowing agents to initiate, evaluate, reason, decide, and propagate reputation values.

1 Introduction

In the last decades, multiagent systems have been proposed as a new paradigm of computation based on cooperation and autonomy. The decentralization and the openess are main characteristics of these systems. There is no central point of control and new agents can enter/leave the system during the execution. Therefore, these systems are vulnerable with respect to malicious agents. This fact poses a great challenge: the detection and the prevention of undesirable behaviours. That is the reason why techniques such as trust and reputation mechanisms have been proposed.

In this paper, we explore the cognitive science background which captures the notions of trust, reputation, and confidence to provide computational trust mechanisms applied to negotiations within artificial societies. For this purpose, we formalize here these notions and we apply to them the argumentation technology proposed in [6] for allowing agents to initiate, evaluate, reason, decide, and propagate reputation values.

The paper is organised as follows. In Section 2, we introduce the walk-through example to motivate our approach. Section 3 presents the background of our proposal. We focus on the notion of interpersonal confidence and we characterize it. Section 4 introduces a conceptual framework for analyzing the decision problem related to the confident behaviour of agents. Section 5 presents our computational Argumentation Framework (AF) for decision making. Section 7 outlines the social interaction amongst agents. This interaction is illustrated by an example of information-seeking dialogue to propagate reputation values. The paper is summarised in section 8 where we also discuss related and future work.

2 Walk-through example

In order to illustrate our model, we consider an e-procurement scenario where two different types of agents, referred to below as A-type and B-type agents, negotiate a service. We consider here a specific case where A-type agents look for a service S. The A-type agents requesting S are `Al` and `Alice`; the B-type agents providing S are `Bob` and `Barbara`. `Carla` is a C-type agent which is an external and neutral observer. A-type agents are responsible for selecting B-type agents. A possible contract may concerns the concrete service `S(c)` and may be between `Al` and `Bob`. `Al`'s goal consists of finding and agreeing to a service S provided by a B-type agent. According to its preferences and constraints, the undermining of the quality specified in the contract is preferred than the overcharge of the price specified in the contract. Taking into account its goal and preferences/constraints, `Al` needs to solve a decision-making problem where the decisions amount to a service and a provider for that service. On the other hand, the goal of the provider `Bob` consists of agreeing to provide a service. According to `Bob`'s preferences and constraints, the seller must pay the service he buys. Taking into account its goal and preferences/constraints, `Bob` needs to solve a decision-making problem where the decisions amount to a service and a client for that service. The two decision making processes take place in a dynamic setting, whereby information about other agents, the services they require/provide and the characteristics of these services is obtained incrementally within the dialogues. The outcome of this process may be the contract obliging `Bob` to deliver on time the service to `Al`.

Within our computational model of trust for negotiation, the reasoning of each agent is supported by argumentation. In the remainder of this paper, we will focus on the reasoning of `Al`. In the concrete use case, `Al`, as a requester, uses argumentation for collecting information on the providers. For instance, `Al` can ask `Carla` its opinion about `Bob` and can ask to justify it (by providing an argument for it). `Al`, as a requester, uses argumentation for deciding which provider it selects taking into account its preferences/constraints and possibly the inconsistency of information it has gathered. Moreover, through argumentation, the participants provide an interactive and intelligible explanation of their choices. For instance, `Al` can argue that the selection of `Bob` is justified since the latter will not overcharge the price. Thus, agents can use argumentation for reasoning about trust for negotiating.

3 Reputation model

Cognitive science provides a pertinent and well-grounded background for developing computational models of trust. In this section, we focus on the notion of interpersonal confidence. We characterize it in order to provide a reasoning tool.

According to the Merriam-Webmester, the confidence is the “ faith or belief that one will act in a right, proper, or effective way”. In

---

[1] Universitá di Pisa, Italy, email: morg@di.unipi.it

the literature, this notion is defined in many different ways by psychologists, sociologists, economists. We restrict ourself here on the *interpersonal confidence*, which is the confidence of an agent about another agent in accordance with the information the first one has about the interactions of the second one with other agents. Considering the interpersonal confidence, [7] distinguishes the beliefs about the confidence that we called *reputation*, the reasoning with these beliefs, and the confident behaviour which is deduced. Considering a reputation, [7] distinguishes a *target*, i.e. the agent which is observed and evaluated, the *observers* i.e. the agents which observe the target, and the *evaluator*, i.e. the agent which evaluates and benefits of the reputation. Actually, we consider here the case where the observation of the target can be done by several agents, possibly *propagated* (i.e. communicated) to the evaluator.

Let us consider the properties of the reasoning over reputation that we will consider in this paper. The reputation is *subjective*, since it depends on the beliefs of the evaluator. The reputation is *uncertain*, since agents needs to accept information with reserve. The reputation is *multidimensional*, since it depends on the competences, the honesty, the foreseeability, the integrity, and the good will of the target. The reputation is *dynamic*, since it depends on the interactions of the target with other agents. The reputation is *defeasible*, since it can be challenged or reinstantiated in the light of new information not previously available. In this paper, we consider that the reputation is *qualitative* and not quantitative, since an evaluator can trust more a target than another one but there is no unit to measure reputation.

Amongst the computational model of trust and reputation, [7] distinguishes:

- the *decision tools*, i.e. which compute the reputation values and the confident behaviour;
- the *evaluation tools*, i.e. decision tools with a revision mechanism to update the reputation values in the light of new information;
- the *reasoning tools*, i.e. evaluation tools with a propagation mechanism to communicate reputation values;

Obviously, we want to provide here a reasoning tool.

4 Decision analysis

Our methodology is to decompose the complex problem related to the confident behaviour of agents into elements that can be analyzed and can be brought together to create an overall representation. We use here *influence diagrams* which are simple graphical representations of decision problems [3] including the decisions to make amongst the possible courses of action (called *decision nodes*, represented by squares), the value of the specific outcomes that could result (called *value nodes*, represented by rectangles with rounded corners), and the uncertain events which are relevant information for decision making (called *chance nodes*, represented by ovals). In order to show the relationship amongst these elements, nodes are put together in a graph connected by arrows, called *arcs*. We call a node at the beginning of an arc a *predecessor* and one at the end of an arc a *successor*. The nodes are connected by arcs where predecessors are independent and affect successors. Influence diagrams which are properly constructed have no cycles. In order to capture multi-criteria decision making, it is convenient to include additional nodes (called *abstract value nodes*, represented by double line) that aggregate results from predecessor nodes. While a *concrete value* is specified for every possible combination of decisions and events that feed into this node, an abstract value is specified for every possible combination of values that feed into this node, and so the multiple attributes are represented with a hierarchy of values where the top, abstract values aggregate the lower, concrete values. We assume that influence diagrams are provided by users via a GUI which allows them to communicate user-specific preferences.

We consider here the decision problem related to the confident behaviour of `Al` (cf Fig. 1). Throughout the paper we adopt the following convention: variables are in italics and constants are in typescript font. The evaluation of the contract (`contract`) depends on the selection of the proposal from the provider x about the service y. This top main value is is split into abstract values, the provision of the service (`provision`) and the provider (`reputation`). The evaluation of the service depends on its cost (`cost`) and its quality (`qos`). The evaluation of these criteria depends on the agent knowledge, namely the information about the services provided by the interlocutors (e.g. `price` and `warranty`). The evaluation of the partners depends on its reliability (e.g. competences or honesty) to deliver the service with the same quality (respectively price) as specified in the contract, denoted `rqos` (respectively `rcost`). The evaluation of the provider is influenced by its expected behaviour ($\mathtt{will}(x, do)$) which depends on the testimonials about it previous behaviour, $\mathtt{Test}(tid, x, done)$. Each testimonial `tid` can be provided by a neutral ($\mathtt{type}(\mathtt{tid}, \mathtt{neutral})$) or a concurrent agent ($\mathtt{type}(\mathtt{tid}, \mathtt{concurrent})$). The interaction of the testimonial `tid` can be direct ($\mathtt{int}(\mathtt{tid}, \mathtt{direct})$) or observed ($\mathtt{int}(\mathtt{tid}, \mathtt{observed})$). Moreover, the testimonial can be more or less recent, $\mathtt{time}(\mathtt{tid}, t)$. The user represented by the agent `Al` also provides, through the GUI, her preferences and constraints. For instance, the undermining of the quality specified in the contract is preferred than the overcharge of the price specified in the contract. Since concurrent agent can hide or lie, we prefer testimonials from neutral agents. In this way, we consider sociological information about the roles of agents. Even if witness information is usually the most abundant, direct experience is the most reliable source of information.

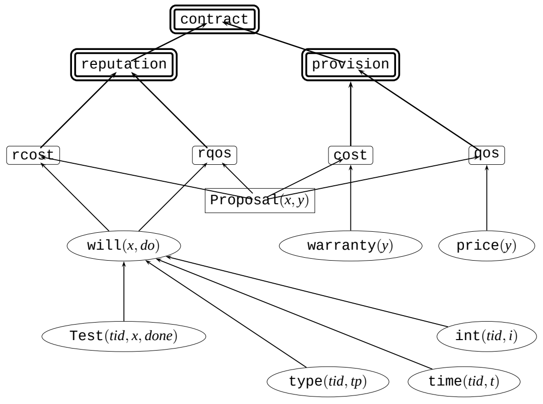

Figure 1. Influence diagram for the confident behaviour.

5 Argumentation framework

According to the approach of defeasible argumentation of [4], arguments are reasons supporting claims which can be defeated[2] by other arguments.

Definition 1 (AF) *An* argumentation framework *is a pair* $AF = \langle \mathcal{A}, \mathtt{defeats} \rangle$ *where* $\mathcal{A}$ *is a finite set of arguments and* `defeats`

---

[2] The defeat relation is called attack in [4].

is a binary relation over $\mathcal{A}$*. We say that a set* S *of arguments defeats an argument* a *if* a *is defeated by at least one argument in* S*.*

[4] also analysis when a set of arguments is collectively justified.

Definition 2 (Semantics) *A set of arguments* $\mathsf{S} \subseteq \mathcal{A}$ *is:*

- conflict-free *iff* $\forall \mathsf{a}, \mathsf{b} \in \mathsf{S}$ *it is not the case that* a *defeats* b*;*
- admissible *iff* S *is conflict-free and* S *defeats every argument* a *such that* a *defeats some arguments in* S*.*

For simplicity, we restrict ourself to admissible semantics.

5.1 Decision framework

Since we want to instantiate our `AF` for our example, we need to specify a particular framework capturing the decision problem.

Definition 3 (Decision framework) *A* decision framework *is a tuple* $\mathcal{D} = \langle \mathcal{L}, \mathcal{A}sm, \mathcal{I}, \mathcal{T}, \mathcal{P} \rangle$*, where:*

- $\mathcal{L}$ *is the* object language *which captures the statements about the decision problem;*
- $\mathcal{A}sm$*, is a set of sentences in* $\mathcal{L}$ *which are taken for granted, called* assumptions*;*
- $\mathcal{I}$ *is the* incompatibility relation*, i.e. a binary relation over atomic formulas which is asymmetric. It captures the mutual exclusion between the statements;*
- $\mathcal{T}$ *is the* theory *which gathers the statements;*
- $\mathcal{P} \subseteq \mathcal{T} \times \mathcal{T}$ *is a (partial or total) preorder over* $\mathcal{T}$*, called the* priority *relation, which captures the uncertainty of beliefs, the priority amongst goals, and the expected utilities of the decisions.*

In the object language $\mathcal{L}$, we distinguish six disjoint components:

- a set of *abstract goals* (resp. *concrete goals*), i.e. some propositional symbols which capture the abstract values (resp. concrete values) that could result;
- a set of *decisions*, i.e. some predicate symbols which capture the decision nodes;
- a set of *alternatives*, i.e. some constants symbols which capture the mutually exclusive actions for each decision;
- a set of *beliefs*, i.e. some predicate symbols which capture the chance nodes;
- the *names* of rules in $\mathcal{T}$ which are unique.

In $\mathcal{L}$, we consider strong negation (classical negation) and weak negation (negation as failure). A strong literal is an atomic first-order formula, possible preceded by strong negation $\neg$. A weak literal is a literal of the form $\sim L$, where L is a strong literal.

We explicitly distinguish *assumable* (respectively *non-assumable*) literals which can (respectively cannot) be taken for granted, meaning that they can or cannot be assumed to hold as long as there is no evidence to the contrary. Decisions (e.g. $\mathtt{Proposal}(x, y) \in \mathcal{A}sm$) as well as some beliefs (e.g. $\mathtt{int}(tid, \mathtt{observed}) \in \mathcal{A}sm$) can be taken for granted. In this way, $\mathcal{D}$ can capture incomplete knowledge.

The *incompatibility relation* captures the conflicts. We have $L\,\mathcal{I}\,\neg L$, $\neg L\,\mathcal{I}\,L$, and $L\,\mathcal{I} \sim L$ but we do not have $\sim L\,\mathcal{I}\,L$. We say that two sets of sentences Φ_1 and Φ_2 are incompatible ($\Phi_1\,\mathcal{I}\,\Phi_2$) iff there is at least one sentence ϕ_1 in Φ_1 and one sentence ϕ_2 in Φ_2 such as $\phi_1\,\mathcal{I}\,\phi_2$.

A *theory* gathers the statements about the decision problem.

Definition 4 (Theory) *A* theory $\mathcal{T}$ *is an extended logic program, i.e a finite set of rules* R*:* $L_0 \leftarrow L_1, \ldots, L_j, \sim L_{j+1}, \ldots, \sim L_n$ *with* $n \geq 0$*, each* L_i *being a strong literal in* $\mathcal{L}$*. The literal* L_0*, called the* head *of the rule, is denoted* $\mathsf{head}(R)$*. The finite set* $\{L_1, \ldots, \sim L_n\}$*, called the* body *of the rule, is denoted* $\mathsf{body}(R)$*. The body of a rule can be empty. In this case, the rule, called a* fact*, is an unconditional statement.* R*, called the unique* name *of the rule, is an atomic formula of* $\mathcal{L}$*. All variables occurring in a rule are implicitly universally quantified over the whole rule. A rule with variables is a scheme standing for all its ground instances.*

For simplicity, we will assume that the names of rules are neither in the body nor in the head of the rules thus avoiding self-reference problems. Considering a decision problem, we distinguish:

- *goal rules* of the form R: $G_0 \leftarrow G_1, \ldots, G_n$ with $n > 0$. Each G_i is a goal literal in $\mathcal{L}$. The head of the rule is an abstract goal (or its strong negation). According to this rule, the abstract goal is promoted (or demoted) by the goal literals in the body;
- *epistemic rules* of the form R: $B_0 \leftarrow B_1, \ldots, B_n$ with $n \geq 0$. Each B_i is a belief literal of $\mathcal{L}$. According to this rule, B_0 is true if the conditions $B_1, \ldots, B_n$ are satisfied;
- *decision rules* of the form R: $G \leftarrow D(a), B_1, \ldots, B_n$ with $n \geq 0$. The head of the rule is a concrete goal (or its strong negation). The body includes a decision literal ($D(a) \in \mathcal{L}$) and a set of belief literals possibly empty. According to this rule, the concrete goal is promoted (or demoted) by the decision $D(a)$, provided that conditions $B_1, \ldots, B_n$ are satisfied.

Due to our representation of decision problems, we assume that the elements in the body of rules are independent, the decisions do not influence the beliefs, and the decisions have no side effects.

In order to evaluate the previous statements, all relevant pieces of information should be taken into account, such as the uncertainty of knowledge, the priority between goals, or the expected utilities of the decisions. In this work, we consider that all rules are potentially defeasible and that the priorities are extra-logical and domain-specific features. We consider that the *priority* $\mathcal{P}$ which is a reflexive and transitive relation considering possible *ex æquo*. $R_1 \mathcal{P} R_2$ can be read "R_1 has priority over R_2". $R_1 \not\mathcal{P} \mathrm{R}_2$ can be read "R_1 has no priority over R_2", either because R_1 and R_2 are *ex æquo* or because R_1 and R_2 are not comparable. The priority over concurrent rules depends on the nature of rules. Rules are *concurrent* if their heads are identical or incompatible. We define three priority relations:

- the priority over *goal rules* comes from the *preferences* overs goals. The priority of such rules corresponds to the relative importance of the combination of (sub)goals in the body as far as reaching the goal in the head is concerned;
- the priority over *epistemic rules* comes from the *uncertainty* of knowledge. The prior the rule is, the more likely the rule holds;
- the priority over *decision rules* comes from the *expected utility* of decisions. The priority of such rules corresponds to the expectation of the conditional decision in promoting/demoting the goal literal.

In order to illustrate the previous notions, let us consider the goal rules, the epistemic rules, and the decision rules of our example which are represented in Tab. 1. According to the decision rules, the goal `rcost` (respectively `rqos`) is reached if the partner does not overcharge the price (respectively undermine the quality) of the service, $\mathtt{r}_{31}$ (respectively $\mathtt{r}_{41}$). According to the goal rules, the undermining of the quality specified in the contract is preferred than the overcharge of the price specified in the contract. Indeed, achieving the goals `rcost` and `rqos` are required to reach `reputation` (cf r_{134}). However, these constraints can be relaxed and the achievement of `rqos` can be relaxed ($r_{134}\mathcal{P}r_{13}$). According to the epistemic rules, the testimonials from neutral agents are preferred to the testimonials from concurrent agents ($\mathtt{r}_1\mathcal{P}r'_1$) and the direct interactions are more reliable than the observed interactions ($\mathtt{r}_2\mathcal{P}\mathtt{r}'_2$).

5.2 Arguments

Since we want that our AF not only suggests some actions but also provides an intelligible explanation of them, we adopt here the tree-like structure for arguments proposed in [10] and we extend it with suppositions on the missing information.

Definition 5 (Argument) *An* argument *built upon $\mathcal{D}$ is composed by a conclusion, a top rule, some premises, some suppositions, and some sentences. These elements are abbreviated by the corresponding prefixes. An argument $\mathtt{a}$ can be:*

1. *a* hypothetical argument *built upon an unconditional ground statement. If L is an assumable literal (possibly its negation), then the argument built upon a ground instance of this assumable literal is defined as follows[3]: $\mathtt{conc}(\mathtt{a}) = L$, $\mathtt{top}(\mathtt{a}) = \theta$, $\mathtt{premise}(\mathtt{a}) = \emptyset$, $\mathtt{supp}(\mathtt{a}) = \{L\}$, $\mathtt{sent}(\mathtt{a}) = \{L\}$.*
   or
2. *a* built argument *built upon a rule such that all the literals in the body are the conclusion of arguments.*
   - *(a) If f is a fact in $\mathcal{T}$ (i.e. $\mathtt{body}(f) = \emptyset$), then the* trivial *argument $\mathtt{a}$ built upon this fact is defined as follows: $\mathtt{conc}(\mathtt{a}) = \mathtt{head}(f)$, $\mathtt{top}(\mathtt{a}) = f$, $\mathtt{premise}(\mathtt{a}) = \emptyset$, $\mathtt{supp}(\mathtt{a}) = \emptyset$, $\mathtt{sent}(\mathtt{a}) = \{\mathtt{head}(f)\}$.*
   - *(b) If r is a rule in $\mathcal{T}$ with $\mathtt{body}(r) = \{L_1, \ldots, L_j, \sim L_{j+1}, \ldots, \sim L_n\}$ and there is a collection of arguments $\{\mathtt{a}_1, \ldots, \mathtt{a}_n\}$ such that, for each strong literal $L_i \in \mathtt{body}(r)$, $\mathtt{conc}(\mathtt{a}_i) = L_i$ with $i \leq j$ and for each weak literal $\sim L_i \in \mathtt{body}(r)$, $\mathtt{conc}(\mathtt{a}_i) =\sim L_i$ with $i > j$, we define the* tree *argument $\mathtt{a}$ built upon the rule r and the set $\{\mathtt{a}_1, \ldots, \mathtt{a}_n\}$ of arguments as follows: $\mathtt{conc}(\mathtt{a}) = \mathtt{head}(r)$, $\mathtt{top}(\mathtt{a}) = r$, $\mathtt{premise}(\mathtt{a}) = \mathtt{body}(r)$, $\mathtt{supp}(\mathtt{a}) = \cup_{\mathtt{a}' \in \{\mathtt{a}_1, \ldots, \mathtt{a}_n\}} \mathtt{supp}(\mathtt{a}')$, $\mathtt{sent}(\mathtt{a}) = \cup_{\mathtt{a}' \in \{\mathtt{a}_1, \ldots, \mathtt{a}_n\}} \mathtt{sent}(\mathtt{a}') \cup \mathtt{body}(r) \cup \{\mathtt{head}(r)\}$. The set of arguments $\{\mathtt{a}_1, \ldots, \mathtt{a}_n\}$ are called the set of* subarguments *of $\mathtt{a}$ (denoted $\mathtt{sbarg}(\mathtt{a})$).*

The set of arguments built upon $\mathcal{D}$ is denoted $\mathcal{A}(\mathcal{D})$.

Notice that the subarguments of a tree argument concluding the weak literals in the body of the top rule are hypothetical arguments. Indeed, the conclusion of an hypothetical argument could be a strong or a weak literal while the conclusion of a built argument is a strong literal. As in [10], we consider composite arguments, called *tree* arguments, and atomic arguments, called *trivial* arguments. Contrary to other definitions of arguments (set of assumptions, set of rules), our definition considers that the different premises can be challenged and can be supported by subarguments. In this way, arguments are intelligible explanations. Moreover, we consider *hypothetical* arguments which are built upon missing information or a decision. In this way, our framework allows to reason further by making suppositions related to the unknown beliefs and over possible decisions.

In our example, the argument `b` (respectively `a`), `Bob` concludes that `Bob` will (respectively not) overcharge the service if we suppose that the interaction was observed. The tree arguments `a` and `b` are composed of four subarguments: one hypothetical argument concluding that the interaction is observed and three trivial arguments concluding the other premises.

6 Interactions

The interactions amongst arguments may come from their conflicts, from their nature (hypothetical or built), and from the priority of rules. We examine in turn these different sources of interaction.

Since their sentences are conflicting, the arguments interact with one another. For this purpose, we define the following attack relation.

Definition 6 (Attack relation) *Let $\mathtt{a}, \mathtt{b} \in \mathcal{A}(\mathcal{D})$ be two arguments. $\mathtt{a}$* attacks *$\mathtt{b}$ iff $\mathtt{sent}(\mathtt{a})\ \mathcal{I}\ \mathtt{sent}(\mathtt{b})$.*

This relation encompasses both the direct (often called *rebuttal*) attack due to the incompatibility of the conclusions, and the indirect (often called *undermining*) attack, i.e. directed to a “subconclusion”. According to this definition, if an argument attacks a subargument, the whole argument is attacked.

Since arguments are more or less hypothetical, we define the size of their suppositions.

Definition 7 (Supposition size) *Let $\mathtt{a} \in \mathcal{A}(\mathcal{D})$ be an argument. The* size of suppositions *for $\mathtt{a}$, denoted $\mathtt{suppsize}(\mathtt{a})$, is the number of suppositions of $\mathtt{a}$: $\mathtt{suppsize}(\mathtt{a}) = |\mathtt{supp}(\mathtt{a})|$.*

The size of suppositions for an argument is the number of decision literals and assumable belief literals in the sentences of the argument.

Since arguments have different natures (hypothetical or built) and the top rules of built arguments are more or less strong, we define the strength relation as follows.

Definition 8 (Strength relation) *Let A_1 be a hypothetical argument, and A_2, A_3 be two built arguments.*

1. *A_2 is stronger than A_1 (denoted $A_2\ \mathcal{P}^{\mathcal{A}}\ A_1$);*
2. *If $(\mathtt{top}(A_2)\mathcal{P}\mathtt{top}(A_3)) \wedge \neg(\mathtt{top}(A_3)\mathcal{P}\mathtt{top}(A_2))$, then $A_2\ \mathcal{P}^{\mathcal{A}}\ A_3$;*
3. *If $(\mathtt{top}(A_2)\not\mathcal{P}\mathtt{top}(\mathrm{A}_3)) \wedge (\mathtt{suppsize}(\mathrm{A}_2) < \mathtt{suppsize}(\mathrm{A}_3))$, then $A_2\ \mathcal{P}^{\mathcal{A}}\ A_3$;*

Since $\mathcal{P}$ is a preorder on $\mathcal{T}$, $\mathcal{P}^{\mathcal{A}}$ is a preorder on $\mathcal{A}(\mathcal{T})$. Since it is preferable to consider fewer suppositions as possible, built arguments are preferred to hypothetical arguments. Moreover, we want to take into account the preferences captured by the priorities. That is the reason why we consider that an argument is stronger than another argument if the top rule of the first argument has a proper higher priority than the top rule of the second argument, or if it is not the case but the number of suppositions made in the first argument is properly smaller than the number of suppositions made in the second argument.

In order to adopt Dung’s seminal calculus of opposition, we define the defeat relation.

Definition 9 (Defeat relation) *Let $\mathtt{a}, \mathtt{b} \in \mathcal{A}(\mathcal{D})$ be two arguments. $\mathtt{a}$* defeats *$\mathtt{b}$ iff: i) $\mathtt{a\ attacks\ b}$; ii) $\neg(\mathtt{b}\ \mathcal{P}^{\mathcal{A}}\ \mathtt{a})$.*

---

[3] θ denotes that no literal is required.

$\mathcal{T}$

r$_1$(*tid*): will(*x*, *do*) ← Test(*tid*, *x*, *do*), type(*tid*, neutral), int(*tid*, *i*), time(*tid*, *t*)
r$_2$(*tid*): will(*x*, *do*) ← Test(*tid*, *x*, *do*), type(*tid*, *tp*), int(*tid*, direct), time(*tid*, *t*)

r$'_1$(*tid*): will(*x*, *do*) ← Test(*tid*, *x*, *do*), type(*tid*, concurrent), int(*tid*, *i*), time(*tid*, *t*)
r$'_2$(*tid*): will(*x*, *do*) ← Test(*tid*, *x*, *do*), type(*tid*, *tp*), int(*tid*, observed), time(*tid*, *t*)
f$_1$: Test(Carla$_1$, Bob, ¬overcharge)
f$_2$: Test(Carla$_1$, Bob, ¬underquality)
f$_3$: type(Carla$_1$, neutral)
f$_4$: time(Carla$_1$, 1)
f$_5$: Test(Alice$_1$, Bob, overcharge)
f$_6$: Test(Alice$_1$, Bob, underquality)
f$_7$: type(Alice$_1$, concurrent)
f$_8$: time(Alice$_1$, 1)
f$_9$: Test(Al$_1$, Barbara, overcharge)
f$_{10}$: Test(Al$_1$, Barbara, ¬underquality)
f$_{11}$: type(Al$_1$, neutral)
f$_{12}$: int(Al$_1$, direct)
f$_{13}$: time(Al$_1$, 3)
f$_{14}$: Test(Carla$_2$, Barbara, ¬overcharge)
f$_{15}$: Test(Carla$_2$, Barbara, underquality)
f$_{16}$: type(Carla$_2$, neutral)
f$_{17}$: time(Carla$_2$, 2)
f$_{18}$: price(d, high)
f$_{19}$: warranty(d, low)
f$_{20}$: price(c, low)
f$_{21}$: warranty(c, high)
f$_{22}$: price(e, low)
f$_{23}$: warranty(e, low)
f$_{24}$: price(f, high)
f$_{25}$: warranty(f, high)

$\mathcal{T}$

r$_{012}$: contract ← reputation, provision
r$_{134}$: reputation ← rcost, rqos
r$_{256}$: provision ← cost, qos

r$_{01}$: contract ← reputation
r$_{13}$: reputation ← rcost
r$_{25}$: provision ← cost

r$_{02}$: contract ← provision
r$_{14}$: reputation ← rqos
r$_{26}$: provision ← qos

$\mathcal{T}$

r$_{31}$: rcost ← Proposal(*x*, *y*), will(*x*, ¬overcharge)
r$_{41}$: rqos ← Proposal(*x*, *y*), will(*x*, underquality)
r$_{51}$: cost ← Proposal(*x*, *y*), price(*y*, low)
r$_{62}$: qos ← Proposal(*x*, *y*), warranty(*y*, high)

r$_{52}$: cost ← Proposal(*x*, *y*), price(*y*, high)
r$_{61}$: qos ← Proposal(*x*, *y*), warranty(*y*, low)

Table 1. The epistemic rules (at left), the goal rules (at upper right), and the decision rules (at lower right).

Let us consider our previous example. The arguments `a` and `b` attack each other. Since the top rules of `a` is r_1 and the top rule of `b` is r'_1, `a` is stronger than `b`, and so `a` defeats `b`. If we consider now the whole problem, there is an argument concluding that `contract` is reached if we consider the service `S(c)` provided by `Bob` due to the provision of the service and the reputation of the potential suppliers. This argument is in an admissible set, and so this proposal can be adopted and justified by `Al`.

7 Social interaction

The social statements are exchanged during dialogues and notified in the *dialogical commitments*. Our agent drives the interactions by the adherence to protocols.

The computation of the reputation values, the confident behaviour, the revision mechanism and the propagation mechanism are driven according to the individual/social statements concerning the goals of agents (their own goals and the goals of their interlocutors), the decisions they make, the knowledge, and preferences over them. The social statements are exchanged during dialogues and notified in the *dialogical commitments* which are internal data structures which contain propositional/action social obligations involving the agent, namely with the agent being either the debtor or the creditor. The choice amongst actions is made according to the agent's statements and the preferences over them. The dialogical commitments of `Al` include commitments involving `Al`: either `Al` is the creditor of the commitment, for instance a commitment is added when `Bob` suggests a concrete service; or `Al` is the debitor of the commitment, for instance a commitment is added when `Al` accepts it.

In this way, agents reasons and take decision about the proposals and arguments which are exchanged during the dialogues in accordance to the reputation values. For instance, `Al` can built an argument concluding that its goal related to the cost of the service is reached if the price of the concrete service $S($`c`$)$ is low. This argument is useful for `Al` to justify its choice, $S($`c`$)$, in front of `Bob`.

Dialogue protocols are required to conduct the interaction. For this purpose, the social reasoning uses a boot strap mechanism that initiates the required protocol, the role the agent will play in that protocol, and the other participants. The protocol engine determines the appropriate message to be sent given those parameters. When there is a decision to be made either between the choice of two locutions (e.g. an accept or a challenge) to be sent or the instantiation of the content of the locution (e.g. the definition of an assert), the protocol engine uses a precondition mechanism to prompt the reasoning of the agent. Upon the satisfaction of the precondition, the protocol engine sends the locution. A similar mechanism is used for incoming messages. If it is necessary to update the commitments of the agent, this can be done with the post condition mechanism which operates in a similar manner.

The agents utter messages to exchange goals, decisions, and knowledge. The syntax of messages is in conformance with a common agent communication language. We assume that each message: has an identifier, M_k; is uttered by a speaker (S$_k$); is addressed to a hearer (H$_k$); responds to a message with identifier R$_k$; is characterised by a speech act A$_k$ composed of a locution and a content. The locution is one of the following: `question`, `assert`, `accept`, `why`, `withdraw` (see Table 2 below for examples). The content is a triple consisting of: a goal G_k, a decision D_k, and a knowledge K_k[4]

Fig. 2 depicted our information-seeking protocol from the initiator viewpoint with the help of a deterministic finite-state automaton. The choice of locutions to send depends on the way the reasoning fulfills preconditions. For example, the following rule dictates whether the protocol engine sends `accept`, `assert` or `why`:

```
IF receive assert(G,D,K) from inter THEN {
  update commit(interlocutor,[G,D,K]);
  IF evaluate(G,D,K) THEN{
```

---

[4] We will use θ to denote that no goal is given and $\emptyset$ to denote that no knowledge is provided.

```
          send accept(G,D,K) to inter;
          commit(me,[G,D,K]);}
    ELSEIF send why(G,D,K) to inter;}
```

In this rule `me` denotes the reasoning agent and `inter` denotes the agent it dialogues with. `evaluate(G,D,K)` is a predicate which evaluates if the goal G is supported by an admissible argument built upon the decision D and the knowledge K. According to this rule, the dialogical commitments are updated when a proposal is received. If an admissible proposal have been suggested, then the speech act is an `accept`. Otherwise the speech act is a `why`.

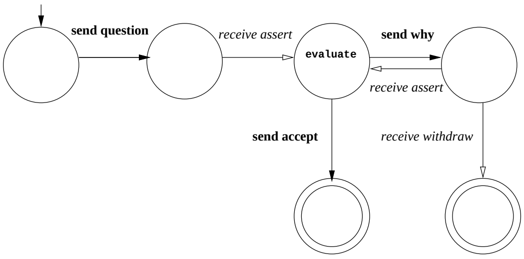

Figure 2. Information-seeking protocol for the initiator.

Tab. 2 shows the speech acts exchanged between `Al` and `Carla` playing an information-seeking dialogue. The first move is for `Al` to pose a question to `Carla`, M_0. This locution seeks the expected behaviour `Bob`. This expected behaviour is good (cf M_1) and argued by the testimonial that `Bob` did not overcharge the price specified in the contract in a previous interaction. Therefore, `Al` consider `Bob` as a trusted supplier and they can negotiate.

M_k	S$_k$	H$_k$	A$_k$	R$_k$
M_0	Al	Carla	question(rcost, Proposal(Bob, y), [will(Bob, do)])	θ
M_1	Carla	Al	assert(rcost, Proposal(Bob, y), [will(Bob, $\neg overcharge$)])	M_0
M_2	Al	Carla	why(rcost, Proposal(Bob, y), [will(Bob, $\neg overcharge$)])	M_1
M_3	Carla	Al	assert(rcost, Proposal(Bob, y), [[Test(Carla$_1$, Bob, ¬overcharge), time(Carla$_1$, 1)])	M_2
M_4	Al	Carla	accept(rcost, Proposal(Bob, y), [Test(Carla$_1$, Bob, ¬overcharge), time(Carla$_1$, 1)])	M_3

Table 2. Information seeking dialogue

8 Conclusions

In this paper, we have proposed a computational model of trust for negotiation. For this purpose, we have provided an AF for decision-making to perform the reasoning of agents about the reputations. In order to valid this approach, we use MARGO[5].

As the computational model of trust proposed by [8], our model uses the reputation values to guide the negotiation, The computational model of trust in [8] guides the negotiation by (i) choosing the partners, (ii) devising the set of negotiable issues, and (iii) determining negotiation intervals. Our computational model of trust guides the negotiation by collecting information on the partners and considering the trust issues to chooses one of the partners.

In order to compare our computational model of trust, we can consider the analysis grid of [9]. The game theoretical approach is the predominant paradigm for the design of computational model of trust. These models have given good results for simple scenarios but the reduction of reputation to a risk probability in decision making is too restrictive in scenarios as considered in the ARGUGRID[6] project. For this purpose, we have adopted a social and cognitive approach where the mental states that leads to trust another agent as well as its consequences and its propagations are essential. Our model takes into account direct experiences (direct and observed interactions), witness information (also called word-of-mouth), and sociological information (the fact that the observer is a concurrent). However we do not consider prejudice, i.e. the use of signs that identify the agent as a member of a group. In this paper, we have considered the reputation values as subjective (rather than global) and multi-context (i.e. multidimensional). Our model assumes that agents can hide or lie (level 2 in agent behaviour assumptions of [9]). Our model can consider boolean information as well as continuous measures if we adopt the extension of our AF for quantitative preferences [5]. The information is composed rather than aggregates through a dialectical process.

We have used here the argumentation-based mechanism for decision making proposed in [6]. The framework of [1, 6] incorporates abduction on missing information, while the frameworks of [2, 6] can be applied to a multi-criteria decision making. To the best of our knowledge, the framework of [6] is the only one integrating both of these proprieties required by our application. Moreover, the other existing frameworks do not come with a conceptual framework for creating a model and a representation of decision problems. By relying on [6], the decision problem related to the confident behaviour is firstly analyzed, and so treated.

ACKNOWLEDGEMENTS

This work is supported by the Sixth Framework IST programme of the EC, under the 035200 ARGUGRID project. We would like to thank the referees for their comments which helped improve this paper.

REFERENCES

[1] A.K. and P. Moraitis, 'Argumentative-based decision-making for autonomous agents', in *Proc. of AAMAS*, pp. 883–890. ACM Press, (2003).

[2] L. Amgoud, 'A general argumentation framework for inference and decision making', in *Proc. of the 21st Conference on Uncertainty in Artif. Intell.*, AUAI Press, pp. 26–33, (2005).

[3] R. T. Clemen, *Making Hard Decisions*, Duxbury. Press, 1996.

[4] P.M. Dung, 'On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games', *Artif. Intell.*, **77**(2), 321–357, (1995).

[5] *Arguing about quantitative aspects of decisions*, ed., M. Morge, ARGUGRID report, 2008.

[6] M. Morge, 'The hedgehog and the fox. an argumentation-based decision support system', in *Argumentation in Multi-Agent Systems*, volume 4946 of *Lecture Notes in Artificial Intelligence*, pp. 114–131. Springer-Verlag, (2008).

[7] G. Muller, *Utilisation de normes et de réputations pour détecter et sanctionner les contradictions*, Ph.D. dissertation, ENS Mines de Saint Etienne, Decembre 2006.

[8] S.D. Ramchurn, N.R. Jennings, C. Sierra, and L. Godo, 'Devising a trust model for multi-agent interactions using confidence and reputation', *Int. J. of Applied Artificial Intelligence*, **18**(9-10), 833–852, (2004).

[9] J. Sabater, 'Review on computational trust and reputation models', *Artif. Intell. Review*, **24**(1), 33–60, (2005).

[10] G. Vreeswijk, 'Abstract argumentation systems', *Artif. Intell.*, **90**(1-2), 225–279, (1997).

---

[5] http://margo.sourceforge.net

[6] http://www.argugrid.eu