Agents adapt to majority behaviours

Duy Hoang Pham, Guido Governatori, and Simon Raboczi
School of Information Technology and Electrical Engineering
The University of Queensland, Brisbane, Queensland, 4072, Australia
{pham,guido,raboczi}@itee.uq.edu.au

Abstract— **Agents within a group can have different perceptions of their working environment and autonomously fulfil their goals. However, they can be aware of beliefs and goals of the group as well as other members so that they can adjust their behaviours accordingly. To model these agents, we explicitly include knowledge commonly shared by the group and that obtained from other agents. By avoiding actions which violate "mental attitudes" shared by the majority of the group, agents demonstrate their social commitment to the group. Defeasible logic is chosen as our representation formalism for its computational efficiency, and for its ability to handle incomplete and conflicting information. Hence, our agents can enjoy the low computational cost while performing "reasoning about others". Finally, we present the implementation of our multi-agent system.**

I. Introduction

As can be observed from a society, every individual member can take any action driven by his/her desires. However, the individuals are often required to comply with the society "conventions". Essentially "conventions" could be norms, constraints or desires which are popularly recognised by the society. Being aware of those conventions, an individual member can strengthen his/her social relationships and coordinates better with other members. Similarly, in a multi-agent system, an agent maintains its social commitments by discovering the "common attitudes" and fulfils its own demands whilst obligating to these attitudes. Hence agents reason not only about their own beliefs and goals but also about those of other agents. Besides, agents autonomously observe and judge their surroundings by their own view resulting in partial and sometimes conflicting descriptions of the world. Consequently, modelling those agents requires representing and reasoning with incomplete and conflicting information, which is beyond the classical logics and monotonic reasoning.

Recently, defeasible logic (DL)[1], [2] has attracted considerable interest from the research community [3]–[5], especially in application to modelling rational agents [6]–[9]. DL is an elegant and computationally efficient tool [10], [11] to deal with partial and conflicting knowledge. The key advantage of DL is being able to draw a plausible conclusion from a reasonable amount of information. In addition, DL provides a compact representation and an effective way to accommodate new information.

In this paper, we propose a new modelling technique based on DL to explicitly describe the knowledge that is commonly shared by agents, and that obtained from other agents. The new model enables an agent to reason about the environment and the intentions of other agents. Social actions are achieved by balancing between the desires of an individual and the beliefs of the majority. To tackle multi-source reasoning, we extend the reasoning mechanism of DL with the notion of *superior knowledge*. This new mechanism allows an agent to integrate its mental attitudes with a more trustworthy source of information such as the knowledge shared by the majority of other agents. In the implementation of our multi-agent system, we add modal notions including Belief, Intention, and Obligation in order to have a fine grained model of "mental attitudes" and social actions.

In the rest of this paper, we firstly describe the basic concepts of DL in Section II. We then introduce our modelling technique and discuss discuss how to represent the knowledge base of agents including the meta-knowledge about agents' importance in Section III. Also, we outline the strategies to allow agents to discover approximate "common attitudes" among the group and the mechanism for reasoning with a priority source of knowledge. In Section IV, we present the architecture of our system including the extension of Rule Markup Language as a knowledge representation tool and the algorithm of the reasoning engine. Finally, we provides an overview of research works related to our system in Section V.

II. Defeasible Logic

Following the presentation in [5], a defeasible theory D is a triple $(F,R,>)$ where F is a finite set of facts, R is a finite set of rules, and $>$ is a superiority relation on R. The language of DL consists of a finite set of literals, l, and their compliments $\sim l$. A rule in R is composed of an antecedent (body) $A(r)$ and a consequent (head) $C(r)$, where $A(r)$ consists of a finite set of literals and $C(r)$ contains a single literal. $A(r)$ can be omitted from the rule if it is empty. There are three types of rules in R, namely R_s (strict rules), R_d (defeasible rules), and R_{dft} (defeaters). We use R_{sd} for the set of strict and defeasible rules, and $R[q]$ for the set of rules whose head are q. A conclusion q derived from the theory D is a tagged literal and is categorised according to how the conclusion can be proved: $+\Delta$ – definitely provable; $-\Delta$ – definitely unprovable; $+\partial$ – defeasibly provable; $-\partial$ – defeasibly unprovable.

The set of conclusions of a defeasible theory is finite [1], and can be computed in linear time [10](See [4] for a more detailed exposition of DL and [11] for existing implementations).

[1]It is the Herbrand base that can be built from the literal occurring in the rules and the facts of the theory.

Proceedings of The 2008 IEEE International Conference on Research, Innovation and Vision for the Future, 2008. RIVF'08. Ho Chi Minh City, 13–17 July.

III. MODELLING TECHNIQUE

A. Knowledge Representation

In general, each individual agent can take any action by balancing the its desires, its knowledge about the environment, and perception of other agents' behaviours. As a member of a group, each agent should be aware of the mental attitudes commonly held among the group and should avoid actions which can violate the group's desires. Similarly, individual behaviours can be significantly influenced by either members with high reputation or the majority of the group. By considering its knowledge and the "collective wisdom" of the group, an agent can adjust its behaviours accordingly. In order to capture this concept, we propose a knowledge structure for an agent, which consists of three components including *background*, *other members*, and *internal knowledge* i.e. the agent's own knowledge.

Given a group of agents, $\mathscr{A} = \{A_1, \ldots, A_{n+1}\}$, and a weight function, $w_A : \{A_1, \ldots, A_{n+1}\} \mapsto \mathbb{R}^+$, representing the importance (reliability) of an agent to the group. The knowledge structure of an agent A_{me} in $\mathscr{A}$ is depicted by a set of defeasible theories $\mathscr{T} = \{T_{bg}, T_{me}, \mathscr{T}_{other}\}$.

T_{bg} is the *background theory* representing the background knowledge. This knowledge represents information commonly shared by all agents, which motivates general (social) behaviours. Also this knowledge can present desires or restrictions popularly known among agents.

T_{me} is the *internal theory* representing the own knowledge of A_{me}, which describes its own view about the working environment. This knowledge enables A_{me} to achieve autonomously and distinctively its goals.

$\mathscr{T}_{other} = \{T_i : 1 \leq i \leq n+1 \ \& \ i \neq me\}$ where T_i is a defeasible theory that A_{me} obtains from A_i in $\mathscr{A}$. The importance of T_i is derived from that of the corresponding agent as $w_T(T_i) = w_A(A_i)$. The knowledge of *other agents* provides a rough understanding of their possible behaviours. This information could be learnt from past experience or via information exchange. However, methods to obtain this information are not the primary concern of this paper.

Our approach favours the internal view approach [12] to model agents' behaviours in the sense that an agent can adapt or react to events depending on what it knows about the environment and other agents. Moreover, our approach can be used as a tool for modelling the external view on behaviours of agents. Interactions between agents in the group can be fully investigated and validated when every individual agent is equipped with detailed knowledge of the other agents.

B. Majority Knowledge

The *majority rule* from [13] retrieves a maximal amount of consistent knowledge from a set of agents' knowledge. Conflicts between agents can be tackled by considering not only the number of agents supporting that information but also the importance of the agents. The approach provides a useful and efficient method to discover information largely held by agents. The majority knowledge can be used either to reinforce the current knowledge of an agent or to introduce new information into the agent's knowledge.

Due to possible conflicting information within a knowledge source, the majority rule cannot be directly applied to our approach. Instead, the majority rule pools potential joint conclusions derived by the defeasible reasoning process.

Given the knowledge structure of an agent, A_{me}, C_i denotes the set of conclusions derived from $T_i \in \mathscr{T}_{other}$. The support level of a conclusion c is obtained from the weight of the theory that holds c:

$$support(c, T_i) = \begin{cases} w_T(T_i) & c \in C_i \\ 0 & \text{otherwise} \end{cases}$$

The majority knowledge from the others, T_{maj}, whose elements are inferred from $\{C_1, \ldots, C_n\}$ by the majority rule, is determined by:

$$T_{maj} = \left\{ c | \sum_{T_i \in \mathscr{T}_{other}} support(c, T_i) > \frac{W - w_T(T_{me})}{2} \right\}$$

where $W = \sum_{A_i \in \mathscr{A}} w_A(A_i)$. W is the total weight of the group. Each conclusion in T_{maj} can have different support levels accumulated from individual theories. The weight of majority conclusions is a set $\{w_{maj}\}$, whose elements have values ranging from $W - w_T(T_{me})$ to $\frac{W - w_T(T_{me})}{2}$.

Proposition 1: For any literal q, it is impossible to have both $+\partial q$ and $+\Delta{\sim}q$ in T_{maj}

Due to the nature of DL proofs and conflicts between knowledge sources, there can be strict and defeasible conclusions of a literal and its complement in the inference from individual sources. However, the outcome of the majority rule is still coherent. The proof for the proposition is rather straightforward by contradiction.

As T_{maj} is derived from what the agent A_{me} knows about the other agents, T_{maj} can conflict with T_{me} (the internal knowledge of A_{me}). In the case that A_{me} joins the majority pool, the greater importance (weight) A_{me} acquires, the greater influence it has on the joint knowledge. If the weight of A_{me} is greater than $W/3$, A_{me}'s support for any conclusion c is tantamount to half of the group's support for c. A_{me} can have two strategies to handle conflicts :

1) *Adaptive strategy*: if $w_{me} \leq W/3$, A_{me} should consider conclusions from the others, since it is unlikely that A_{me} can successfully override conflicts from the joint knowledge. That means T_{maj} introduces new information to A_{me}.
2) *Dominant strategy*: otherwise, A_{me} can defeat conflicts from other agents if A_{me} joins the pool. The joint knowledge from the others just reinforces the current knowledge of A_{me}.

In both strategies, *background knowledge* commonly shared by the group is respected absolutely; i.e. in case of a conflict between a conclusion from background knowledge and either from the majority or the agent's knowledge, the conclusions, which are supported by the background part prevail.

C. Defeasible Reasoning with Superior Knowledge

In this section, we propose a simple method to integrate two independent defeasible theories. Note that a defeasible theory has finite sets of facts and rules, and a derivation from the theory can be computed in linear time [10].

Suppose that an agent considers two knowledge sources represented by defeasible theories labelled as T_{sp} – the superior theory, and T_{me} – the agent's internal theory. The agent considers that T_{sp} is more important than T_{me}. Thus, conclusions from the internal theory should be withdrawn if they conflict with the superior theory; the agent prefers the superior theory's conclusions to its own.

Owing to the transformations of the superiority relation and defeater rules [4], we can assume that the two theories contain only strict and defeasible rules. To perform the defeasible reasoning, the agent generates a superiority relation over sets of rules as in $R_s^{sp} > R_s^{me}; R_d^{sp} > R_d^{me}$. In this scheme, the subscript denotes the type of rules while the superscript indicates the type of the theory which contains the rules.

A derivation from the two theories is a finite sequence $P = (P(1), \ldots, P(n))$ of tagged literals satisfying proof conditions (which correspond to inference rules for each of the four kinds of conclusions). $P(1..i)$ denotes the initial part of the sequence P of length i. The definite conclusion, $+\Delta q$, will be derived by performing forward chaining with the strict rules in the superior theory, or in the internal theory if the complementary literals cannot be positively proved by the superior theory.

$+\Delta$: If $P(i+1) = +\Delta q$ then
(1) $q \in F$ or
(2) $\exists r \in R_s^{sp}[q]\ \forall a \in A(r) : +\Delta a \in P(1..i)$ or
(3) $\exists r' \in R_s^{me}[q]\ \forall a \in A(r') : +\Delta a \in P(1..i)$ and
$\forall r \in R_s^{sp}[\sim q]\ \exists a \in A(r) : -\Delta a \in P(1..i)$

The conclusions tagged with $-\Delta$ mean that the extended mechanism cannot retrieve a positive proof for the corresponding literals from the strict parts of both theories.

$-\Delta$: If $P(i+1) = -\Delta q$ then
(1) $q \notin F$ and
(2) $\forall r \in R_s^{sp}[q] \cup R_s^{me}[q]\ \exists a \in A(r) : -\Delta a \in P(1..i)$.

The proof for $-\Delta$ implicitly satisfies the principle of strong negation [3]. The proof, which strictly complies with that principle, requires an additional condition such that at least one strict rule from the superior theory supports the complementary literals. However, this condition is never met as it violates the coherence property of the strict rules.

A defeasible conclusion $+\partial q$ can either be drawn directly from definite conclusions, or by investigating the defeasible part of the integrated theory. In particular, it is required that a strict or defeasible rule with an "applicable" head q is in the theory (2.1). In addition, the possible "attacks" must be either unprovable (2.2 and 2.3.1) or counter-attacked by "stronger" rules (2.3.2).

$+\partial$: If $P(i+1) = +\partial q$ then either
(1) $+\Delta q \in P(1..i)$ or
(2.1) $\exists r \in R_{sd}^{sp}[q] \cup R_{sd}^{me}[q]\ \forall a \in A(r) : +\partial a \in P(1..i)$ and
(2.2) $-\Delta \sim q \in P(1..i)$ and
(2.3) $\forall s \in R_{sd}^{sp}[\sim q] \cup R_{sd}^{me}[\sim q]$ either
(2.3.1) $\exists a \in A(s) : -\partial a \in P(1..i)$ or
(2.3.2) $\exists t \in R_{sd}^{sp}[q] \cup R_{sd}^{me}[q]$ such that $t > s$ and $\forall a \in A(t) : +\partial a \in P(1..i)$

The conclusions tagged with $-\partial$ mean that the extended mechanism cannot retrieve a positive proof for the corresponding literals from the strict and defeasible rules of both theories or these conclusions are rebutted because of "stronger" conclusions. The proof for $-\partial$ derives from that of $+\partial$ by using the strong negation principle.

The extended defeasible reasoning with the superior knowledge has these properties[2]

1) If $T_{sp} \vdash +\Delta q$ then $T_{sp} + T_{in} \not\vdash +\Delta{\sim}q$ and $T_{sp} + T_{in} \not\vdash +\partial{\sim}q$. If a strict conclusion is derived from the superior theory, the extended mechanism does not provide any proof for its negation.
2) If $T_{sp} \vdash \sim\Delta{\sim}q$ and $T_{in} \vdash +\Delta q$ then $T_{sp} + T_{in} \vdash +\partial q$. The conclusions from the extended mechanism can violate defeasible conclusions obtained from the superior theory if the agent has a strong evidence of the contradiction in its internal knowledge.
3) The extended reasoning mechanism is coherent and consistent.

Given two defeasible theories T and S and a proof tag #, we use $T \vdash \#q$ to state that $\#q$ can be proved from theory T using the basic proof conditions of DL, while $T \ni S \vdash \#q$ means that there is a derivation of $\#q$ from the theory integrating T and S using the proof conditions given in this section whereas T plays the role of T_{sp} and S the role of T_{me}.

The extended mechanism computes a consistent set of conclusions with respect to the superior theory. The mechanism goes beyond the standard defeasible reasoning since it extends the superiority relation of rules to that of theories. This increases the size of theory to be investigated. Hence the complexity class of the reasoning algorithm [10] remains unchanged.

D. Reasoning Mechanism

We are now able to describe how to incorporate the majority rule into a reasoning mechanism based on the notions introduced in the previous sections. The reasoning mechanism operates in two steps. The first step is to identify the majority knowledge from the other agents. In the second step, the agent performs either adaptive or collective reasoning depending on its weight.

a) Determining the majority knowledge from other agents.
i) Draw defeasible conclusions from the others:

$$T_{bg} \ni T_i \vdash C_i : 1 \leq i \leq n and i \neq me$$

ii) Establish the majority knowledge T_{maj} over the sets of defeasible conclusions, $\{C_i : 1 \leq i \leq n\}$ (see III-B).

[2] The proof for these properties is omitted due to the space limit.

b) Reasoning strategies.
At this stage, the set of knowledge sources is reduced to the background, the majority, and the agent's own knowledge. Depending on its weight, an individual agent can either follow or reject the majority knowledge.
Adaptive reasoning.: First, the agent combines the background and its own knowledge by considering the background as the superior source. Next, the joint knowledge is used to adjust the derivation from the first step. That is, the agent withdraws conclusions which violate the joint knowledge.

$$T_{maj} \ni (T_{bg} \ni T_{me}) \vdash C^{'}_{me}$$

Dominant reasoning: Any conflict from the majority is rejected by the agent.

$$(T_{bg} \ni T_{me}) \ni T_{maj} \vdash C^{'}_{me}$$

Proposition 2: The complexity of the proposed mechanism is in the $O(n)$ class.

This property is due to the linear complexity of defeasible reasoning and the majority rule.

IV. "LIKE MAJORITY" MULTI-AGENT SYSTEM

A. MAS-LM Architecture

Our architecture for "like majority" multi-agent system, MAS-LM, has two major components as shown in Figure 1. The first component is the repository of the agents' knowledge, which is built by designers. In order to facilitate the interactions between designers and agents, RuleEditor module provides a Java user interface to create defeasible theories representing knowledge of individual agents. Once designers finish composing sets of knowledge including individual's knowledge and meta-knowledge of the agents' weights, these knowledge are stored in the RuleML-like repository (see Section IV-B). Only background knowledge and meta-knowledge are accessible to all agents.

The second component in the dashed-line box in Figure 1 presents essential modules of an individual agents. RuleLoader parses defeasible theories into Java objects, which is suitable for ReasoningEngine. KnowledgeExchange performs communication with other agents in the group. Incoming information is stored in the internal repository as Java objects.

ReasoningEngine performs decision-making process by using the extended defeasible reasoning (see Section IV-C). Decisions are stored in the internal repository for knowledge exchange or further investigation. Action module essentially provides connections between an agent and its working environment.

B. Knowledge Representation

1) Knowledge Structure with modal notions: As presented in Section III-A, an individual agent has a knowledge structure as a tuple $\mathscr{T} = \{T_{bg}, T_{me}, \mathscr{T}_{other}\}$, whose elements are defeasible theories. In order to better capture social actions, we introduce modal notions including Belief, Intention, and Obligation. These notions allow agents to explicitly reason

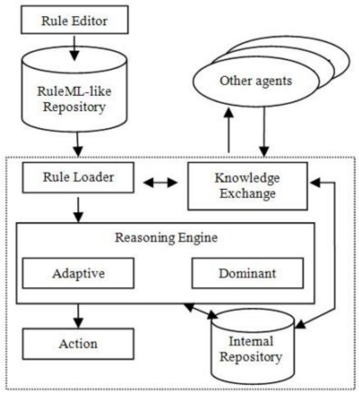

Fig. 1. MAS-LM system architecture

not only about beliefs of other agents but also about their goals resulting in a stronger social behaviours [14]. Now every $T_i \in \mathscr{T}$ $(T_i \neq T_{bg})$ has two independent sets of defeasible theories T_i^B and T_i^I that represent the set of beliefs and the set of intentions correspondingly. Meanwhile, T_{bg} has all three modal notions $T_{bg} = \{T_{bg}^B, T_{bg}^I, T_{bg}^O\}$. Essentially, beliefs represent what agents believe to be true; intentions represent what agents want to achieve; and obligations represent what agents should commit to the group.

Example 1: There is a man in a sinking boat and three agents A_1, A_2, A_3 observe the situation having the weight of $\{6, 3, 1\}$ respectively. Knowledge commonly shared among the agents is:

$$T_{bg}^B = \{R_s = \{r_1 :\to manOnSinkingBoat\}\}$$
$$T_{bg}^I = \{R_d = \{\; r_1 : manOnSinkingBoat \Rightarrow manInDanger$$
$$r_2 : manInDanger \Rightarrow rescue\}$$
$$T_{bg}^O = \{R_d = \{r_1 : risk \Rightarrow \sim rescue\}\}$$

The background knowledge state that the man is in danger, the rescue should be performed if it is safe to do so. Besides, A_3 knows about the intentions of A_1, A_2, and itself:

$$T_1^I = \{R_d = \{r_1 :\Rightarrow swim, r_2 : swim \Rightarrow risk\}\}$$
$$T_2^I = \{R_s = \{r_1 :\to throwRope, r_2 : throwRope \to \sim risk\}\}$$
$$T_{me}^I = T_3^I = \{R_d = \{r_1 :\Rightarrow surf, r_2 : surf \Rightarrow \sim risk\}\}$$

Essentially, the knowledge structure is interpreted as A_1 wants to swim directly to the sinking boat while A_2 intends to throw a rope to the boat, and A_3 plans to approach the boat by a surf board.

2) RuleML Extension as Knowledge Representation: Rule Markup Language (RuleML) is an XML based language for the representation of rules. It offers facilities to specify different types of rules from derivation rules to transformation rules to reaction rules. RuleML already supports derivation rules via *Implies* element. However, we need to define a new syntax to represent the strength of the rules and superiority relations. Following that of DR-DEVICE [15], every rule in the knowledge structure now has a *@ruletype* attribute taking one of three values: *strictrule*, *defeasiblerule* or *defeater*. Because a rule can be superior to more than one other rule, we explicitly represent the superiority relation using the predicate *Override* [16].

The conclusions from corresponding theories, represented by the *Conclusion* element, are also stored for exchanging knowledge or explaining agents' behaviours. Each conclusion includes the literal and the strength of the proof. Finally, every defeasible theory in the knowledge structure, containing a collection of rules, facts, and superiority, is represented by *DLTheory* element having three attributes, namely *source*, *weight*, and *modality* corresponding to the source name, the weight, and the modal notion of the theory.

C. Reasoning Engine

1) Social Categories: As in Section III-B, an individual agent can adapt to the majority by dropping its own beliefs and intentions in favour of those popularly recognised by the group. However, the agent can dominate the group by promoting its own intentions and rejecting contradictory beliefs and intentions from the majority of the group. In both situations, obligations from the background knowledge plays as "filter" so that any behaviour violating these obligations is cancelled by the individual agent.

We categorise our agents into two types of social behaviours entitled "majority" and "obedience". Agents in the first category totally commit to the group avoiding conflicts with the majority and the group's "common conventions", represented by T^O_{bg}. These agents collect the majority beliefs and intentions from others by running the reasoning mechanism in Section III-D over belief and intention elements of the knowledge structure respectively:

1) $T^B_{maj} \ni (T^B_{bg} \ni T^B_{me}) \vdash C'^B_{me}$
2) $T^I_{maj} \ni (T^I_{bg} \ni T^I_{me}) \vdash C'^I_{me}$
3) $\{T^O_{bg}; C'^B_{me}\} > C'^I_{me}$

In contrast, obedient agents only commit to "common conventions" and perform the reasoning process:

1) $(T^B_{bg} \ni T^B_{me}) \ni T^B_{maj} \vdash C'^B_{me}$
2) $(T^I_{bg} \ni T^I_{me}) \ni T^I_{maj} \vdash C'^I_{me}$
3) $\{T^O_{bg}; C'^B_{me}\} > C'^I_{me}$

Example 2: Reconsidering Example 1, since A_3 does not know about the belief of other agents, the majority belief equals the derivation of the background belief T^B_{bg}. That is $T^B_{maj} = \{+\Delta manOnSinkingBoat\}$.

A_3 identifies intentions of A_1 and A_2 by integrating what A_3 knows with T^I_{bg}, $T^I_{bg} \ni \mathscr{T}_i^I \vdash C^I_i : i = 1,2$:

$$C^I_1 = \{+\partial manInDanger^6, +\partial swim^6, +\partial risk^6, +\partial rescue^6\}$$
$$C^I_2 = \{+\partial manInDanger^3, +\Delta throwRope^3, +\Delta{\sim}risk^3, +\partial rescue^3\}$$

The superscript of defeasible conclusions represents the weight inherited from the corresponding knowledge source. The majority intentions from others are:

$$T^I_{maj} = \{+\partial manInDanger^9, +\partial swim^6, +\partial risk^6, +\partial rescue^9\}$$
$$w_{maj} = \{9, 6\}$$

The superscript of a majority conclusion shows the weight accumulated from that of the sources supporting the conclusion. Since A_3's weight is the smallest, A_3 adapts its intentions to the majority $I_{A_3} = T^I_{maj}$. In the final step, A_3 drops the intention of doing the rescue due to r_1 in T^O_{bg}.

Suppose the weight of the group changes from $\{6,3,1\}$ to $\{6,3,5\}$. By integrating A_3's intentions with that of the background, $T^I_{bg} \ni T^I_{me}$, A_3 derives

$$C^I_{me} = \{+\partial manInDanger^5, +\partial surf^5, +\partial{\sim}risk^5, +\partial rescue^5\}$$

Clearly, if A_3 joined the majority pool, the majority conclusions would favour those from A_3. A_3 now rejects conflicts from the majority intentions, T^I_{maj}, and persists with its own intentions with respect to group obligations. That is obedient agents only maintain their commitments to the group by eliminating intentions against obligations specified in the background knowledge.

We believe that the "majority agents" can express a strong social commitments to the group. Being aware of knowledge from others, these agents dynamically learn new "conventions" recognised by the majority and change their intentions toward this knowledge. On the other hand, "obedient agents" can introduce "new values" into the group. Thanks to their high weight, these agents could take leading actions so that other agents could follow.

2) Algorithm: As in the previous section, every agent in MAS-LM has there knowledge layers corresponding to Belief, Intention and Obligation notions. Eliminating conflicts with Obligation layer from agents' intentions is done by the standard defeasible reasoning. The key component for MAS-LM engine is the mechanism for integrating with a superior knowledge source (Section III-D), which operates on Belief and Intention layers in order to determine beliefs and intentions of the majority. The engine allows an agent either to adapt to or to override the mental attitudes of the majority by implementing adaptive or dominant strategies.

Thanks to the conflict resolution of the reasoning mechanism, the implementation of the majority rule is straightforward. Therefore, this part focuses on the implementation of the reasoning with the superior knowledge. The algorithm for the reasoning mechanism extended from [11] takes theories in RuleML format as input to create the data structure for the inference process. The inference process flattens the superiority relation between theories in order to apply Basic Defeasible Logic. Differing from [11], literals proved in the strict rules can be defeated by definite conclusions derived from the superior theory. Hence, the inference can be used for both strict and defeasible rules by separately investigating those rules. The outcome from processing the strict rules is considered as facts when examining the defeasible rules.

V. Related work

One major stream in multi-agent systems to capture social commitments is to modify the BDI architecture [17] by introducing deontological properties like laws, norms, and obligations to place constraints on agents' behaviours. The deontological properties are considered as external influences on individual's decision making and the commitment to other members. This idea is supported by several authors in [18]–[20]. Clearly modal logics are very powerful to represent these concepts. Our approach differs from BDI-like agents in the consideration of incomplete and conflicting information. The social commitment is implemented by pondering the conflicts with "desires" commonly shared by the group and "desires" shared by the majority. That is our agents demonstrate a social ability via their commitments to beliefs and goals of the group [14].

Generally, agents adjust their behaviours to the majority attitudes, which are dynamically discovered during interactions with other agents. However, if an agent has a strong belief to the contrary of the common (shared) desires, the agent can break its commitment. This exception can be against the goals of the group but offers the agent some levels of autonomy and flexibility in making a decision.

The modal logic framework in [21] combines the multi-agent epistemic logic and multi-source reasoning systems to perform knowledge integration. The reasoning process tackles with conflicts by applying two cautious strategies namely *level cutting fusion* – reject all conflicting beliefs of sources having lower reliability; and *level skipping fusion* – discard only the level containing the conflicts. This framework differs from ours not only in fusion techniques, but also in conflict handling. Thanks to the defeasible logic, the agents' knowledge can have conflicting information but the information is consistent at the end of reasoning process. Also, the conflicts in knowledge sources can be resolved by further exploiting the superiority relation.

VI. Summary

This paper has presented a modelling technique based on defeasible logics for multi-agent systems which explicitly captures the background knowledge of the group and knowledge about other agents. The technique also uses the meta-knowledge to the importance of individual agents in the system. One important feature of our technique is being able to maintain the computational efficiency of both the defeasible reasoning and the majority rule.

Being aware of beliefs and goals of other agents, an individual agent can discover mental attitudes, which are largely shared by the group, and balances those attitudes with its own. In our MAS-LM, agents are categorised into two types: majority and obedience depending on the reasoning strategies they use. An agent can either adapt to majority behaviours or dominate the group. In the second strategy, the agent can introduce a distinct behaviour that would lead other agents while committing to obligations commonly recognised by the group.

We are investigating a framework to simulate emergency situations where rescue teams are well equipped with comprehensive emergency protocols but the information is incomplete and conflicting. The simulation tool can be used to evaluate the performance of team members and the effectiveness of the protocols. To better model agents' behaviours, we need to develop a sophiticated mechanism for rule-based information exchange. With respect to Foundation of Intelligent Physical Agents (FIPA) speech acts and defeasible logics, [22] is a worth investigation.

References

[1] D. Nute, "Defeasible reasoning," in *Proceedings of 20th Hawaii International Conference on System Science*. IEEE press, 1987.

[2] ——, "Defeasible logic," in *Handbook of Logic in Artificial Intelligence and Logic Programming*. Oxford University Press, 1994, vol. 3.

[3] G. Antoniou, D. Billington, G. Governatori, and M. J. Maher, "A flexible framework for defeasible logics," in *AAAI-2000*. AAAI/MIT Press, 2000, pp. 401–405.

[4] ——, "Representation results for defeasible logic," *ACM Transactions on Computational Logic*, vol. 2, no. 2, pp. 255–287, 2001.

[5] D. Billington, "Defeasible logic is stable," *Journal of Logic and Computation*, vol. 3, pp. 370–400, 1993.

[6] G. Governatori and A. Rotolo, "Defeasible logic: Agency, intention and obligation ," in *Deontic Logic in Computer Science*, ser. LNAI 3065, 2004, pp. 114–128.

[7] G. Governatori, A. Rotolo, and V. Padmanabhan, "The cost of social agents ." ACM Press, 2006, pp. 513–520.

[8] M. Dastani, G. Governatori, A. Rotolo, and L. van der Torre, "Preferences of agents in defeasible logic ," ser. LNAI 3809. Springer, 2005, pp. 695–704.

[9] ——, "Programming cognitive agents in defeasible logic ," ser. LNAI 3835, 2005, pp. 621–636.

[10] M. J. Maher, "Propositional defeasible logic has linear complexity," *Theory and Practice of Logic Programming*, vol. 1, no. 6, pp. 691–711, 2001.

[11] M. J. Maher, A. Rock, G. Antoniou, D. Billignton, and T. Miller, "Efficient defeasible reasoning systems," *International Journal of Artificial Intelligence Tools*, vol. 10, no. 4, pp. 483–501, 2001.

[12] D. Kinny, M. Georgeff, and A. Rao, "A methodology and modelling technique for systems of bdi agents," in *MAAMAW '96*. Springer-Verlag New York, Inc., 1996, pp. 56–71.

[13] J. Lin, "Integration of weighted knowledge bases," *Artificial Intelligence*, vol. 83, pp. 363–378, 1996.

[14] C. Castelfranchi, "Modelling social action for ai agents," *Artif. Intell.*, vol. 103, no. 1-2, pp. 157–182, 1998.

[15] N. Bassiliades, G. Antoniou, and I. Vlahavas, "DR-DEVICE: A defeasible logic system for the Semantic Web," ser. LNCS, no. 3208, 2004, pp. 134–148.

[16] G. Governatori, "Representing business contracts in RuleML ," *International Journal of Cooperative Information Systems*, vol. 14, no. 2-3, pp. 181–216, 2005.

[17] A. S. Rao and M. P. Georgeff, "Modeling rational agents within a bdi-architecture," in *KR*, 1991, pp. 473–484.

[18] J. Broersen, M. Dastani, J. Hulstijn, Z. Huang, and L. van der Torre, "The boid architecture: conflicts between beliefs, obligations, intentions and desires." ACM, 2001, pp. 9–16.

[19] C. Castelfranchi, F. Dignum, C. M. Jonker, and J. Treur, "Deliberative normative agents: Principles and architecture." Springer-Verlag, 2000, pp. 364–378.

[20] F. Dignum, D. Kinny, and L. Sonenberg, "Motivational attitudes of agents: On desires, obligations, and norms." Springer-Verlag, 2002, pp. 83–92.

[21] C. Liau, "A modal logic framework for multi-agent belief fusion," *ACM Transactions on Computational Logic*, vol. 6, no. 1, pp. 124–174, 2005.

[22] G. Boella, J. Hulstijn, G. Governatori, R. Riveret, A. Rotolo, and L. van der Torre, "Fipa communicative acts in defeasible logic ," in *Procs. of the 7th International Workshop on Nonmontonic Reasoning, Action and Change*, 2007.