No Agent is an Island: A Framework for the Study of Inter-Agent Behavior

T.J.M. Bench-Capon
Department of Computer Science
The University of Liverpool dress
Liverpool, UK
044-151-3521
tbc@csc/liv.ac.uk

P.E. Dunne
Department of Computer Science
The University of Liverpool dress
Liverpool, UK
044-151-3521
ped@csc/liv.ac.uk

ABSTRACT

We describe a framework for the study of conflict in inter-agent behavior. Section 1 motivates the framework, section 2 formally defines the framework, and section 3 discusses some issues the framework can be used to explore.

Categories and Subject Descriptors

I.2.11 Intelligent agents, Multi-agent systems

General Terms

Design, Experimentation

Keywords

Multi-agent systems, Conflict, Interaction

1. INTRODUCTION

Performing any action will render other actions impossible, by falsifying preconditions for their performance. When an agent chooses to perform an action, it thereby constrains its choices and the choices of other agents. Moreover, since other agents may not be aware of its choices, its choices may frustrate the performance of actions chosen by other agents. John Donne wrote that "No man is an island, entire of itself". Jean-Paul Sartre wrote that "L'enfer est les autres", that hell is other people. The Rolling Stones sang "You can't always get what you want to". It has been long recognized, that what I choose constrains the choices of others, and what others choose constrains my choices. In response, a number of social organizations and moral codes have developed, to guide and regulate inter-personal behavior so as to promote peaceful co-existence and even mutual benefit. If we are to take the notion of societies of agents seriously, we need to recognize that this interconnectedness will hold for agents also. In this paper we develop a framework for reasoning about agent choices and the regulation of inter-agent behavior.

AAMAS'02, July 15-19, 2002, Bologna, Italy.

2. FRAMEWORK DEFINITION

We begin by introducing the notion of an *option*. For our purposes we want to keep this notion as abstract as possible. Thus we are not interested in the details of the action which can be opted for, nor in any of the mechanics of how it might be carried out. An option is simply something which an agent can do, which can have some utility for one or more agents and which can, if performed, block the selection of other options. The options which are available to the agents under consideration, and the conflicts between them are denoted by an *Option Framework*.

Definition 1. An *option framework* is a pair

$$OF = \langle OP, conflicts \rangle$$

where *OP* is a set of options, and *conflicts* is a *symmetric non-reflexive* binary relation on OP, i.e. $conflicts \subseteq OP \times OP$.

For two options op_1 and op_2, the meaning of $conflicts(op_1, op_2)$ is that op_1 and op_2 can never both be chosen together. This may be because the state of affairs realized by op_1 is incompatible with that realized by op_2, or because performing op_1 violates some precondition for performing op_2, or any other source of incompatibility. The relation must be symmetric. Suppose that performing op_1 violates a precondition for performing op_2, then if op_2 is chosen, op_1 cannot be performed, since performing it would render it impossible to perform the chosen option. Thus the notion of conflict embraces both the notion of one action physically preventing another, and the notion of the choice of an action meaning that another cannot be consistently chosen.

OP is intended to describe the totality of options available to a set of agents *A*. Each agent $a \in A$ will be capable of choosing some distinct subset of *OP*. This means that agents do not have options in common: if an option selected by an agent involves the performance of some action, that is seen as a different state of affairs from that in which another agent chooses to perform that action, and there may well be different consequences in terms of the conflict relations the option enters into, and in the benefits its performance affords to various agents. We therefore define the *option set of an agent* as in Definition 2.

Definition 2. The option set, OP_a, of an agent *a* is a subset of *OP*. For a set of agents *A*, $A = \{a_1, a_2, \ldots, a_k\}$, the option sets $\{OP_i\}$ associated with each agent induce a partition of *OP* (i.e for distinct *agents* a_i, a_j, $OP_i \cap OP_j = \varnothing$ and for all $x \in OP$, *x* is in the option set of some agent.)

We next need to provide some means of evaluating the various options, if there is to be a possibility of rational choice. We do this by introducing the notion of the *utility* of an *option for an agent*. Selection and exercise of an option may confer benefits *not only on the agent that chooses the option, but on other agents as well*. Equally it is possible that exercising an option may reduce the welfare of the agent concerned, or of other agents.

Definition 3. The utility of an option for an agent. For all $a \in A$, $op \in OP$ there is a relation *utility(a,op,z)*, where z is an integer (possibly negative).

We read *utility(a,op,z)* as *op has utility z for a.*

The task of an agent is to select a subset of its options, S_a. To be a legitimate selection this set must be consistent: it cannot contain any options that would conflict in the absence of other agents. This means that the selection of an agent cannot contain any option that conflicts with one of its own options. It can, however, include options that conflict with the options of *another* agent, in the hope that either that other agent will not choose to exercise the option, or that the conflict will be resolved favorably.

Definition 4. Selection of an Agent. A subset S_a of OP_a is *selectable* by an agent *a*, if $\forall x \forall y \in OP_a, \neg\, conflicts(x,y)$.

If a conflict occurs between options in the option set of a given agent, then the agent is free to choose whichever option it wishes, although it cannot choose both. If, however, the conflicting options are in the selections of distinct agents, one agent must be dominant with respect to the particular option. We say that the dominant agent *controls* the conflict.

Definition 5. Control of a conflict. For all conflicts between a pair of options op_1 and op_2, $conflicts(op_1, op_2)$, such that $op_1 \in OP_a$ and $op_2 \in OP_b$ for distinct agents a and b, either $controls(conflicts(op_2, op_2),a)$ or $controls(conflict(op_1, op_2),b)$. Note that if $controls(conflicts(op_1,op_2),a)$, then also $controls(conflicts(op_2, op_1),a)$. We can add that for all $op_1 \in OP_a$ and $op_2 \in OP_a$, $controls(conflicts(op_1, op_2),a)$.

In the case where two agents select options which conflict, the option selected by the agent which controls the conflict will be realized and the conflicting option will not. We call the options selected by an agent which are realized, the *realization* of the agent.

Definition 6. Realization of an agent. The realization of an agent $a \in A$, is a set R_a such that $R_a \subseteq S_a$ and no $r_1 \in R_a$ is such that there exists an $r_2 \in S_b$ for some agent $b \in A$, b distinct from a, such that $controls(conflicts(r_1, r_2),b)$.

The total utility of a given agent will depend not only on the options in its own realization, but those in the realizations of its fellow agents. It is therefore also convenient to talk of the realization of an option framework, *R*, which is the union of all the individual realizations of the agents in *A*. We call this R_A. We can say that the utility of an agent U_a is the sum of the utilities for *a* of the elements of R_A. It is also convenient to talk about the utility of a group of agents. We will write this as U_B, where $B \subseteq A$, and is the sum of the individual utilities U_a for all $a \in B$. The final notion we need to introduce here is the notion of the evaluation of an agent of a realization. This is a function of the total realization R_A, $eval_a(R_A)$, and is intended to be some measure of how content with the overall realization the agent is. This function, $eval_a$, may be defined in a number of different ways, some of which are explored in [1]. It is, however, critical, since it is this function that the agent will try to maximize when determining its selection S_a, in so far as it is in its power to do so.

Definition 7. Task of an Agent. The task of an agent *a* in the framework is to construct the selection S_a which is expected to maximize the value of $eval_a(R_A)$.

Not everything that happens does so as the result of the action of an individual agent. Some things happen as the result of nature, and other states of affairs are the product of the actions of more than one agent. In order to represent this we distinguish a partition element OP_N, ("nature"). Since Nature is not an agent, each option (or, rather *event*, since no agent chooses it) is selected with some probability. This means that for all events in OP_N there is a given probability that they will be realized unless prevented by some other option. Similarly an event may prevent an agent from realizing an action. The probability of an event may be independent or conditionally dependent on the probabilities of other events.

3. USING THE FRAMEWORK

The framework given in section 2 is, and is intended to be, very abstract. In order to represent a particular situation we must determine a number of factors, including: the information available to an agent (concerning other agents and their actions, the controls relation etc), the strategy of an agent (for example, minimax or maximin), the extent of communication between agents (can they negotiate?), the way agents evaluate realizations (do they consider the utility of others?), the way in which the controls relation is determined (e.g. an ordering on agents), and the possibility of external regulation of agent choice (to express legal or moral codes). The intention is to use the framework to perform a systematic empirical study of how these factors affect inter-agent behavior., such as:

- Under what circumstances do agents attain the realization that the agents collectively regard as subjectively preferred?
- Does forcing agents to select so as to avoid preventing of selections with higher utility always lead to a greater total utility?

Initial work will be on frameworks with two agents, but the longer term intention is to explore situation with many agents. This will allow consideration of agents acting co-operatively in groups of different sizes. Bench-Capon and Dunne (2002) gives a fuller description of the factors and issues, and some examples, intended to explore the effect of agents having different attitudes towards one another, ranging from devotion through indifference to active dislike. One example uses the well known Prisoner's Dilemma, discussed in relation to agents which consider the utility of other agents as well as themselves.

REFERENCE

[1] Bench-Capon, T., and Dunne, P., (2002). No Agent is an Island, Technical Report UCLS-02-002. Dept. of Computer Science, The University of Liverpool.