When Agents Communicate Hypotheses in Critical Situations

Gauvain Bourgne, Nicolas Maudet, and Suzanne Pinson

LAMSADE, Université Paris-Dauphine
Paris 75775 Cedex 16, France
{bourgne,maudet,pinson}@lamsade.dauphine.fr

Abstract. This paper discusses the problem of *efficient propagation of uncertain information* in dynamic environments and critical situations. When a number of (distributed) agents have only partial access to information, the explanation(s) and conclusion(s) they can draw from their observations are inevitably uncertain. In this context, the efficient propagation of information is concerned with two interrelated aspects: spreading the information as quickly as possible, and refining the hypotheses at the same time. We describe a formal framework designed to investigate this class of problem, and we report on preliminary results and experiments using the described theory.

1 Introduction

Consider the following situation: witness of a threathening and unexpected event, say a fire in a building, Jeanne has to act promptly to both escape the danger and warn other people who might get caught in the same situation. However, there are no official signs or alarms indicating where the fire actually started. Given her partial knowledge of the situation, Jeanne may build some hypotheses explaining her observations (where the fire did start in the first place, maybe why), but the conclusions she may reach would remain *uncertain* (that is, uncertainty here lies on the fact that she has incomplete knowledge of the world, rather than untrusted perceptions of this world). In addition, there is no way for Jeanne to trigger an alarm. In other words, Jeanne will try to both circulate the information in order to spread the information to colleagues, and refine the hypotheses at the same time. Typically, Jeanne faces two questions:

- What information should I transmit?
- To whom should I transmit this information?

Clearly, these two questions are interrelated. Depending on the person Jeanne selected to communicate with, she may decide to transmit different messages: the objectives being to ensure that the transmitted information can be used efficiently in the next transmission, and so on. This defines, we believe, a problem of efficient propagation of uncertain information. The purpose of this paper is to put forward a formal framework expliciting both the reasoning and communicational aspects involved in these situations. We explore some preliminary

M. Baldoni and U. Endriss (Eds.): DALT 2006, LNAI 4327, pp. 89–104, 2006.

properties of the proposed framework and interaction protocol, and illustrate our approach with a case study experimented using the described theory.

The remainder of this paper is as follows. Section 2 presents the formal reasoning machinery that we shall use in the framework: it heavily builds upon Poole's Theorist system [14]. Section 3 details the communication module, and explores specifically some properties of a protocol designed to exchange hypotheses. Section 4 describes our case study example, instantiating the proposed framework. The situation involves a number of agents trying to escape from a burning building. We give the detail of a simple example, showing how critical, in this crisis context, can be the decisions taken by agents as to whether/what communicate. Section 5 draws connections to related works, and Section 6 concludes.

2 Agents' Reasoning

This section introduces the formal machinery involved in the agents reasoning process. The described situation suggests agents able to deal with partial perception of the world, to build hypotheses from observations they make, to draw conclusions from a set of explanations, and to communicate with each other in order to exchange pieces of information. Agents reasoning process builds on Poole's framework [14,15], which allows to elegantly combine both the explanation and the prediction processes, using a single axiomatization. In what follows, by formulae we mean well-formed formulae in a standard first order language. Each agent is (a slightly modified version of) an instance of a Theorist system [14]:

$$\langle \mathcal{F}, \mathcal{H}, \Delta, O, E, \leq \rangle$$

where

- $\mathcal{F}$ a set of *facts*, closed formulae taken as being true in the domain;
- Δ a set of *defaults*, formulae taken as being true without evidence of the contrary. They are used for prediction and can be part of an explanation;
- $\mathcal{H}$ a set of formulae which act as *conjectures*, possible hypotheses common to all agents, usually a set of *abducible predicates*;
- O is a set of grounded formulae representing the *observations* made so far by the agent. Each agent believes every observation in this set to be true;
- E is the set of *prefered explanations*, it is the set of all justifiable explanations of the observation set O;
- $\leq$ is the *preference relation*, a pre-order on the explanations common to all agents.

We first recall a number of basic definitions.

Definition 1 (Scenario [15]). *A scenario of $(\mathcal{F}, A)$ is a set $\theta \cup \mathcal{F}$ where θ is a set of ground instances of elements of A such that $\theta \cup \mathcal{F}$ is consistent. θ is called the* assumption *of the explanation.*