EBDI: An Architecture for Emotional Agents

Hong Jiang*
Computer Science and Eng.
University of South Carolina
Columbia, SC 29208
jiangh@engr.sc.edu

Jose M. Vidal
Computer Science and Eng.
University of South Carolina
Columbia, SC 29208
vidal@sc.edu

Michael N. Huhns
Computer Science and Eng.
University of South Carolina
Columbia, SC 29208
huhns@sc.edu

ABSTRACT

Most of the research on multiagent systems has focused on the development of rational utility-maximizing agents. However, research shows that emotions have a strong effect on peoples' physical states, motivations, beliefs, and desires. By introducing primary and secondary emotion into BDI architecture, we present a generic architecture for an emotional agent, EBDI, which can merge various emotion theories with an agent's reasoning process. It implements practical reasoning techniques separately from the specific emotion mechanism. The separation allows us to plug in emotional models as needed or upgrade the agent's reasoning engine independently.

Categories and Subject Descriptors

I.2.0 [**Artificial Intelligence**]: General—*Cognitive simulation*; I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Intelligent agents, Multiagent systems*

General Terms

Theory

Keywords

Agent architecture, Emotional agent, Belief-Desire-Intention

1. INTRODUCTION

Most of the research into agents has focused on the development of rational utility-maximizing agents. This research assumes that decisions derive from an analysis of the future outcomes of various options and alternatives. The influence that emotions have on human decision-making is largely ignored until recently. A few projects are working on commonsense reasoning, such as the "Architectures for Commonsense Reasoning" project at the MIT Media Laboratory and the "Reasoning and Cognition" project at The University of Birmingham. However, research on applying emotions to agents' decision making are still very limited.

Taking into account the incompleteness of current emotion theories and emotional differences among individual persons, this paper presents a generic architecture, EBDI, for emotional agents, instead of trying to model emotion theories to reflect the reasoning process specifically. It separates the practical reasoning from the specific emotion menchanism. Our EBDI architecture adds the influence of primary and secondary emotions [1] into the decision making process of a traditional BDI architecture. We handle bounded resources by using primary emotions as the first filter for adjusting the priority of beliefs, thereby allowing the agents to speed up decision making. Secondary emotions are used to refine the decision when time permits. Instead of just considering beliefs ambiguously from perception as in the conventional BDI model, we consider belief candidates from perception of the environment, communication and inner-contemplation. We describe the EBDI architecture and an interpreter for it. We also present a sample EBDI agent for the Tileworld domain [5].

2. EBDI ARCHITECTURE

To incorporate emotions into agents, we need solve three problems: (1)How to measure or present emotions? (2)How do emotions affect the decision making process? (3)How to update the status of emotions? Most of the time, the details of the solutions depend on specific applications. Thus, our EBDI architecture combines these three concerns into a BDI architecture based on human's practical reasoning process, while leaving the details open to the designers.

According to [7], practical reasoning involves two important processes: deciding what state of affairs we want to achieve, deliberation, and deciding how we want to achieve this state of affairs, means-ends reasoning. In the EBDI architecture we still follow this processes. We divide the process into four components: Emotion, Belief, Desire and Intention, and we connect these four components through some main functions.

Since there is no standard definition for emotions – [3] mentions that there are as many as 92 different definitions in the literature, we did not limit the definition and the representation method of emotions.

Belief is usually defined as a conviction to the truth of a proposition. Beliefs can be acquired through perception, contemplation or communication. In the psychological sense,

*The primary author is a Ph.D student.
The full paper is available at (USC CSE Technical Reports TR-2007-002): `http://ftp.cse.sc.edu/reports/`

AAMAS'07 May 14–18 2007, Honolulu, Hawai'i, USA.

978-81-904262-7-5 (RPS) ©2007 IFAAMAS

belief is a representational mental state that takes the form of a propositional attitude. Knowledge is often defined as justified true belief, in which the belief must be considered to correspond to reality and must be derived from valid evidence and arguments. Still, we believe the component of belief in the original BDI model is enough to cover the idea of resources added by David Pereira [4], that is, the resources mentioned can actually be looked as a kind of beliefs.

Desires point to the options that are available to the agent, or the possible courses of actions available to the agent. Desires are obtained through an *option generation function*, on the basis of its current beliefs and intentions.

Intentions play a crucial role in the practical reasoning process, because they tend to lead to action. More specifically, intentions here represent the agent's current focus—those states of affairs that it has committed to trying to bring about, and are affected by current emotional status together with current desires and working intentions.

Formally, let E be the set of all possible emotions; B be the set of all possible beliefs, D be the set of all possible desires, and I be the set of all possible intentions. Beliefs can be acquired from perception, contemplation or communication.Here, we define three belief revision functions which map input from these three areas into beliefs. The Belief Revision Function through Perception (*brf-see*) generates belief candidates from the environment:

$$brf\text{-}see : Env \rightarrow B_p$$

where Env denotes the environment, $B_p \subseteq B$ is the set of possible belief candidates from perception. Belief Revision Function through Communication (*brf-msg*) generates belief candidates from the content of communication messages:

$$brf\text{-}msg : Cont \rightarrow B_m$$

where $Cont$ denotes the content of possible communication messages, B_m the set of possible belief candidates from message, and $B_m \subseteq B$. Belief Revision Function through Contemplation (*brf-in*) takes into consideration the current emotion status[1] and intentions, and revises the current beliefs based on previous beliefs and the set of belief candidates from the environment and communication messages:

$$brf\text{-}in : E \times I \times (B \cup B_p \cup B_m) \rightarrow B$$

where B is the previous belief set without being updated, B_p and B_m are new coming belief candidates.

Emotion Update Functions take into account both primary emotions and secondary emotions [1], so we have two corresponding update functions.

Primary emotions are those that we feel first, as a first response to a situation. Thus, if we are threatened, we may feel fear. When we hear of a death, we may feel sadness. They appear before conscious thought and are thus instinctive or reactive functions of the human brain. When time is limited and we do not have enough time to think about something clearly, primary emotions become extremely useful in decision making. In agents, we can use primary emotions to speed up decision making similarly. The primary emotion update function (*euf1*) can be defined as:

$$euf1 : E \times I \times (B_p \cup B_m) \rightarrow E$$

Secondary emotions appear after primary emotions, which may be caused directly by primary emotions, or come from more complex chains of thinking. For example where the fear of a threat turns to anger that fuels the body for a fight reaction. For agents, the secondary emotions come from the result of further deliberation and can replace the primary emotions. They are used to refine the decision making if time permits. The secondary emotion update function (*euf2*) is defined as:

$$euf2 : E \times I \times B \rightarrow E$$

Option Generate Function is similar to the one in BDI model, defined as:

$$options : B \times I \rightarrow D$$

Filter Function adds emotions, which are used to find the best option(s).

$$filter : E \times B \times D \times I \rightarrow I$$

From intentions to actions can be represented by the plan function (*plan*):

$$plan : I \times Ac \rightarrow \pi$$

where Ac is the set of possible actions that the agent can do, and π denotes a plan which is a sequence of actions

$$\pi = (\alpha_1, \cdots, \alpha_n)$$

where α_i is an action, and $\alpha_i \in Ac$.

[1] Frijda [2] shows emotions' influence on beliefs.

EBDI-MAIN-LOOP

```
E ← E0;                ▷ E0 are initial emotions
B ← B0;                ▷ B0 are initial beliefs
I ← I0;                ▷ I0 are initial intentions
while true
      do Bp ← brf-see(Env);
         Bm ← brf-msg(Cont);
         E ← euf1(E, I, Bp ∪ Bm);
         B ← brf-in(E, I, B ∪ Bp ∪ Bm);
         D ← options(B, I);
         I ← filter(E, B, D, I);
         E' ← E
         E ← euf2(E, I, B);
         if time permits and E ≠ E'
            then B ← brf-in(E, I, B);
                 D ← options(B, I);
                 I ← filter(E, B, D, I);
         π ← plan(I, Ac);
         execute(π)
```

Figure 1: Pseudo-code of an EBDI agent's main loop.

The interpreter that is used in the EBDI architecture is shown in figure 1. We can summarize the execution cycle as follows:

1. When there is new information from the environment via sensor or communication messages, the EBDI agent generates belief candidates;
2. These belief candidates together with current intentions trigger emotion updating, that is, the agent obtains its first feeling about the information;
3. Based on the new emotion status and the new information, together with current intentions as a guide, the agent re-evaluates its beliefs;
4. From the beliefs and intentions, the agent generates desires;
5. Under influence of the emotions, the agent chooses the best options or intentions based on current beliefs, desires and intentions.
6. From this deliberation result, the secondary emotions are triggered, and this updating is based on current intentions, beliefs and previous emotions.

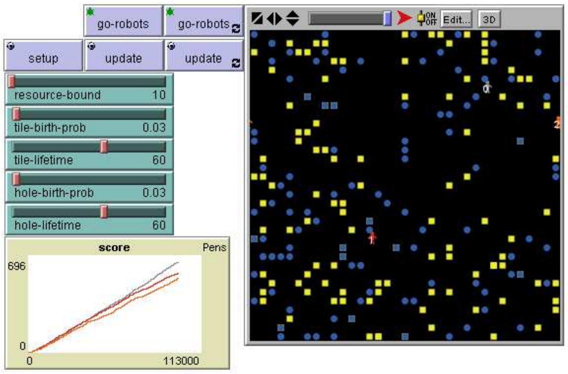

Figure 2: Simulation of Tileworld where the objects with person shape are agents, the yellow squares are tiles, and the blue circles are holes.

7. If there is no time for deeper consideration or emotion status is not changed, the agent will directly generate a detailed plan and execute it. Otherwise, the agent begins a deeper consideration and refines the decision making. It will reconsider if current beliefs are suitable, as in line 14, and reconsider the desires and intentions, as in line 15 and 16. After this reconsideration, the agent then generates a plan and executes it.

Our EBDI agent architecture thus manages to integrate emotions into the standard processing loop of a BDI agent.

3. EBDI AGENT IN TILEWORLD

We chose Tileworld [5] as a platform for experimentally investigating the behavior of various metalevel reasoning strategies,which is simulated in NetLogo [6], as in Figure 2. A Tileworld with EBDI agent is described as follows:

Environment: It is both dynamic and unpredictable. There are holes and tiles, which appear with some probability at a randomly selected location, live for awhile and then disappear. We design the size of the grid be 35×35 in the simulation.

Agents: The task for agents is to push the tiles to cover as many holes as possible. Each agent i tries to obtain the highest utility of its own $u_i = \textit{num-holes-filled}_i$, where $\textit{num-holes-filled}_i$ denotes the number of holes filled by agent i in the environment. We assume each agent can only see holes and tiles in front of it, such that it must partially rely on the information from communication, and the information can be a lie.

To make a comparison between an EBDI agent and a rational agent, we design three agents in the system: A truth telling agent always tells the truth, and a selfish lying BDI agent always lies to others so as to increase its chances of getting all the tiles into the holes. An EBDI agent will feel angry when others lie and feel thankful when others tell the truth, and correspondingly, it will decrease or increase the priority of the information from the agent who lies or tells the truth.

We compare the performance of the three agents and test it for dozens of time, all the results are similar, one of which is shown in figure 3, where we set the birth probability be

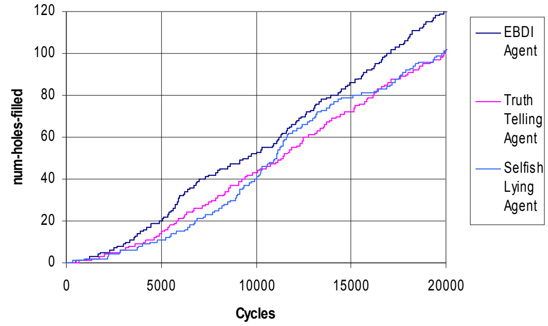

Figure 3: Testing Result

0.03 and the lifetime be 60 cycles for both tiles and holes. It shows that at the beginning, the differences between the performance are hard to tell, since the information is very limited and every agent does the task bychance; but the emotional agent gets better and better and always ends up with the highest score. Thus, the emotional agent here has better performance than rational agents because it is more adaptive in this dynamic environment.

4. CONCLUSION

We presented EBDI, a generic architecture for emotional agents which implements both practical reasoning and emotional mechanisms. EBDI can be used to build high EQ agents by selecting and implementing specific emotion theories into the architecture as needed. We showed a sample EBDI agent in Tileworld, and the test results showed that this EBDI architecture is applicable and that the emotional agent has better performance than rational agents.

5. REFERENCES

[1] A. R. Damasio. *Descartes' Error : Emotion, Reason, and the Human Brain*. Avon Books, New York, 1994.

[2] N. H. Frijda, A. S. R. Manstead, and S. Bem, editors. *Emotions and Beliefs: How Feelings Influence Thoughts*. Cambridge Univ. Press, New York, 2000.

[3] P. Kleinginna and A. Kleinginna. A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion*, 5:345–379, 1981.

[4] D. Pereira, E. Oliveira, N. Moreira, and L.Sarmento. Towards an architecture for emotional bdi agents. In *EPIA '05: Proceedings of 12th Portuguese Conference on Artificial Intelligence*, pages 40–47. Springer, 2005.

[5] M. Pollack and M. Ringuette. Introducing the tileworld: experimentally evaluating agent architectures. In *Proceedings of the Eighth National Conference on Artificial Intelligence*. AAAI Press.

[6] U. Wilensky. NetLogo: Center for connected learning and computer-based modeling, Northwestern University. Evanston, IL, 1999. `http://ccl.northwestern.edu/netlogo/`.

[7] M. J. Wooldridge. *An introduction to multiagent systems*. John Wiley & Sons, LTD, Baffins Lane, Chichester, West Sussex, England, 2001.