A Computational Model of Persistent Beliefs

Sunju Park

Artificial Intelligence Laboratory
The University of Michigan
Ann Arbor, MI 48109-2110
boxenju@eecs.umich.edu

The persistence of beliefs has been assumed in many research effotrs, either explicitly or implicitly, but a computational model is hard to find. For instance, in his well-known AOP article (Shoham 1993), Shoham suggests a formal language for beliefs and states that beliefs persist by default. The author writes $(Bel_A^3\ Bel_B^{10}\ Like(A,B)^7)$ means that at time *3* agent *A* believes that at time *10* agent *B* will believe that at time *7 A* liked *B*. Shoham, however, does not elaborate on how to formally interpret the persistence of beliefs. Moreover, in his implemented agent language, AGENT-0, both the temporal aspect of beliefs (e.g., At time *t*, I believe ...) and the nested beliefs (e.g., I believe you believe I believe ...) have been omitted.

In this abstract, we summarize our work on developing a computational model of persistent beliefs, which supports both the temporal information and the nested belief model.

First, we propose a time-interval representation for nonambiguous interpretation of persistent beliefs. The main idea is rather simple: to have explicit lower and upper time-bounds when representing facts and beliefs.

The time-interval representation clarifies the meaning of persistence without tedious elaboration of each implied belief. For example, the time-interval representation, $(Bel_A^{[10\ \infty]}\ on(paper,\ table)^{[10\ \infty]})$, elaborates the *implied* persistence of the belief, $(Bel_A^{10}\ on(paper,\ table)^{10})$, without ambiguity. It is read as "Agent *A* believes at $t \geq 10$ that $on(paper,table)$ will be true at $t \geq 10$". In addition, it can represent the history of beliefs, which was not possible in AOP.

Secondly, we have developed an algorithm for checking consistency between two beliefs. The basic idea is that *two beliefs are always compatible with each other unless all the following four conditions are satisfied.*

- Negated, same facts: Two beliefs (without any conjunction, disjunction, and deduction) can be potentially inconsistent only if they are about the same facts, one of which is negated.
- Same depth of nested beliefs: If the depth of two nested beliefs are different, they are always consistent.
- Beliefs of same agents: Two beliefs are always compatible if the agents holding two beliefs are different.
- Overlapping time-intervals: Only the overlap and subsume relations can have a potential for conflicts.

We have developed a consistency-checking algorithm between two beliefs, whose time complexity is $O(d)$, where d is the smaller nested depth between two beliefs. If we consider d as a large constant, the complexity is $O(1)$.

Thirdly, to incorporate a new set of beliefs into its old beliefs, the agent needs a belief-revision algorithm. At present, we consider two revision methods: one where new beliefs override old beliefs in the case of inconsistency, and the other where an agent chooses to believe the maximal number of consistent beliefs.

The former case is an extension of AGENT-0, since beliefs now can have temporal information and can be nested. The consistency-checking between two sets of beliefs will take $O((n+m)^2 \times d)$, where n and m represent the number of new beliefs and old beliefs, respectively.

On the other hand, the problem of finding the maximally consistent beliefs is transformed to the independent-set problem, which is NP-complete (Garey & Johnson 1979). If we assume internal consistency of the new belief set and of the belief DB, respectively, however, a polynomial-time algorithm can be possible (Park 1996).

Our research shows some promising early results. First, the interval-based representation is able to represent history and allows nonambiguous interpretation of persistent beliefs. Second, a computationally simple consistency-checking algorithm has been developed. Finally, although finding a maximally-consistent belief set is NP-complete, a polynomial-time revision algorithm is possible under the assumption of internal consistency.

In the future, we will work on relaxing our assumptions of not allowing disjunction and conjunction, and will develop a belief DB that supports basic operations, such as add, delete, update, and query (e.g., what the agent believes, believed, or will believe at time *t*).

This research has been funded in part by the Digital Libraries Initiative under CERA IRI-9411287.

References

Garey, M. R., and Johnson, D. S. 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*: W. H. Freeman.

Park, S. 1996. Belief DB: A Computational Model for Persistent Beliefs. Forthcoming.

Shoham, Y. 1993. Agent-oriented Programming. *Artificial Intelligence* 60:51-91.