Iterative Refinement of Knowledge Bases With Consistency Guarantees

Stephen F. Correl

Department of Computer Sciences
University of Texas at Austin
Austin, TX 78712 USA
correl@cs.utexas.edu

Natural kinds, such as the concepts *toy block* or *photosynthesis*, are ubiquitous in human reasoning yet lack definitional (i.e., individually necessary and jointly sufficient) properties as membership conditions. The meaning of these concepts emerges through the inter-relatedness of facts about them rather than from mathematical definition. The goal of my research is to support construction of large, multi-functional knowledge bases of assertions about such natural kinds.

This knowledge representation setting poses special problems. As new concept descriptions are expressed in terms of existing ones, it must be efficient to logically check for mismatched assertions. We want to be able to verify that the meaning of each concept does not conflict with the constraints implied by its constituents. Knowledge solicited from domain experts (or added via translation from other sources) may introduce subtle inconsistencies due to changes in terminology or variations in the level of abstraction.

This problem can be viewed in terms of two component challenges: (1) as the KB grows, a practical means of guarding against ontological drift is required; and (2) the system must be able to capture the fundamental inter-dependent nature of concept meanings as described by the user. Approaches to these challenges must be computationally tractable, they must be task-independent, and they must not rely on definitional properties of concepts.

Our approach to challenge (1) recognizes that detecting every inconsistency in a real-world KB is impractical or impossible because of the level of expressiveness required and the need for scalability. Instead, we define localized, limited consistency guarantees that are enforced as the KB is constructed. These properties are formally defined, yet intuitively meaningful to the user.

These guarantees are incrementally enforced by means of primitive editing operations that modify the KB. For each edit, a guard computation checks that its effect will not violate consistency. More advanced operations can be implemented entirely in terms of the primitives.

Though it is rigorous in terms of guarantees, this approach is practical in that it depends neither on classification by definitional properties, nor on generally computing entailment among assertions. From the user's perspective, the computation focuses on a question of immediate concern at knowledge acquisition time: is the newly added knowledge stated sensibly with respect to the concepts used to define it? We believe that general classification, while useful as an inference capability, is not practical for detecting ontological drift in the building of real-world KBs.

Our approach to problem (2) is based on the observation that new concepts are naturally built in terms of existing concepts, by referring to their internal details, and by configuring these details to build the new definition. The editing operators allow the user to refer to parts of an existing concept in order to specialize them in their new use. Although it would be simpler to restrict concept references to lambda calculus-style calls (e.g. send(x, y, m)), concepts do not seem to fit this mold in the real world. Explicit support for concept composition without requiring lambda expression interfaces is in contrast to many logic-based approaches; for example, Conceptual Graphs (Sowa 1984) rely on these specifications.

Much work in consistency of knowledge representation has focused on trading representation expressiveness to obtain completeness of inference (e.g. Borgida & Patel-Schneider 1994; Levy & Rousset 1996), rather than on trading completeness for limited consistency control. Systems which emphasize expressiveness at the expense of completeness (e.g. MacGregor 1991) are not organized around formal guarantees.

The complete thesis work will include (1) a definition of the logical basis for the KB and associated operations, (2) proofs that the guards preserve the desired guarantees, (3) an analysis of the complexity of the guard computations as the knowledge base grows, and (4) a web-based KB implementation supporting iterative, guarded definition of concepts and ground instances.

References

A. Borgida and P.F. Patel-Schneider. A semantics and complete algorithm for subsumption in the CLASSIC description logic. *JAIR*, 1:277-308, 1994.

A.Y. Levy and M-C. Rousset. Verification of knowledge bases based on containment checking. *Proc. of the 13th Nat. Conf. on Artificial Intelligence (AAAI-96)*, AAAI Press.

R.M. MacGregor. The evolving technology of classification-based knowledge representation systems. In John Sowa, ed., *Principles of Semantic Networks: Explorations in the Representation of Knowledge*. Morgan Kaufmann, 1991.

J.F. Sowa. *Conceptual structures: Information processing in mind and machine*. Addison-Wesley, 1984.