Efficient Representations for Multi-Modal Interaction

Syed S. Ali
Mathematical Sciences Dept.
University of Wisconsin – Milwaukee
syali@uwm.edu

Susan W. McRoy
EE & Computer Science Dept.
University of Wisconsin – Milwaukee
mcroy@uwm.edu

Abstract

This position paper argues in favor of *uniform, mixed-depth* knowledge representation in intelligent systems that support flexible, multi-modal, mixed-initiative, two-way interaction. Uniformity has the advantage of minimizing knowledge interchange and can be implemented efficiently. Moreover, enhancements to the reasoner improve the performance of the entire system as all components use the same reasoner. Mixed-depth representations allow "shallow" or "deep" representations as required. Some representations can always be shallow, and this eliminates unnecessary interpretation. For example, mixed-depth representations permit the representation of ill-formed, vague, or ambiguous inputs, which can be further interpreted as necessary.

Introduction

This position paper argues in favor of *uniform, mixed-depth* knowledge representation in intelligent systems that support flexible, multi-modal, mixed-initiative, two-way interaction. Multi-modal interaction combines information from many levels of abstraction and many conceptually different types of objects. To support effective communication, interactants should be able to reason about physical actions, intentions, screen objects, etc. The ability to reason about such objects improves the ability of a system to flexibly interpret and incrementally present information. While this does not replace the need for ergonomics and user studies, it supplements the ability of systems to behave well.

Following the distinction of (Bordegoni *et al.* 1997), we are concerned with *intelligent multimedia dialogue systems (IMDS)* rather than *intelligent multimedia presentation systems (IMPS).*

Intelligent multimedia presentation systems are concerned with effective presentation of a fixed content subject to constraints (for example, the user's apparent expertise and preferred presentation style). The knowledge structures built during an interaction are limited to information about what has been presented, what will be presented and what's currently visible, identifiable, or located at a particular position. Their "intelligence" lies in sophisticated domain and presentation models used to adapt presentations, as in plan-based coordination of media to meet the stated needs of the user. This intelligence can take the form of heuristics (don't overlap important windows) or first-order logic-based plan representations that take into account constraints like preconditions, applicability, or intentions. For example, based on the user's stated expertise, an IMPS might choose to present a graphic with an associated explanation. In general, IMPS are concerned with output issues and cannot monitor the effectiveness of the presentation to deal with ambiguity, misunderstanding or non-understanding.

Intelligent multimedia dialog systems are concerned with the effective management of an incremental, interactive, user-system interaction. The knowledge structures built during an interaction include those of an IMPS as well as goals, intentions, and beliefs (including beliefs about the effectiveness of the interaction) of the user and system. Content to be presented, as well as the system's model of the user (for example the user's apparent level of expertise), will change dynamically during an interaction. The "intelligence" of IMDS lies in the traditional utility of a dialog model, which includes the ability to handle fragments, anaphora, misunderstanding, non-understanding, and follow up questions. In general, an IMDS must deal with both input and output (interpretation and presentation) and must monitor the effectiveness of its actions.

The distinction between IMPS and IMDS can be likened to the task of a teacher in two different types of courses; one a 300 student course taught in an auditorium, and the other in a small seminar course.

Most work with intelligent multimedia interfaces to this point has been with IMPS, for reasons of complexity and tractability. It is clearly possible to build interfaces that perform well with little or no knowledge representation (for example, by the use of heuristic algorithms as in (Funke, Neal, & Paul 1993).) However, to build more flexible, user-friendly interfaces, a shift from IMPS to IMDS seems likely. This is because people building intelligent interfaces are increasingly including knowledge representation(s) to exploit the readily available information about the interaction and its potential to improve the interaction (for some specific systems see (Neal & Shapiro 1991; Wahlster *et al.* 1993;

Traum *et al.* 1996); an architecture for IMPS that explicitly promotes knowledge "experts" is that of (Bordegoni *et al.* 1997)). In this paper we are going to argue that an IMDS should use a knowledge representation with specific characteristics: it must be *uniform* and allow *mixed-depth* representations. These characteristics help address the complexity and tractability problems in building detailed models of dialog.

We term a knowledge representation *uniform* when it allows the representation of different kinds of knowledge in the same knowledge base using the same inference processes. For example, the representation of a spoken Huh might be represented similarly to a typed Huh? and reasoned with identically. (Minimally, modality information would differ in the representations; this could be used to track the user's preferred modality.)

A *mixed-depth representation* is one that may be shallow or deep in different places, depending on what was known or needed at the time the representation was created (Hirst & Ryan 1992). Moreover, "shallow" and "deep" are a matter of degree. Shallow representations might be a representation of the interaction such as a sequence of time-stamped events. Deep representations might be a conventional first-order (or higher-order) AI knowledge representation. Unlike quasi-logical form, which is used primarily for storage of information, mixed-depth representations are well-formed propositions, subject to logical inference. Disambiguation and interpretation, when it occurs, is done by reasoning. Most KR systems allow mixed-depth representations, uniform KR systems require them (because everything is represented in the same knowledge base).

To motivate our argument for uniform, mixed-depth representation we first argue the advantages of such representations in an ongoing interaction; we then discuss the practicable nature of the proposed representations.

The Advantages of Uniform Representations

The primary advantage of a uniform representation is that it eliminates knowledge interchange overhead. That is, there are no special-purpose reasoners with specialized knowledge representation(s), and all reasoning uses the same reasoner. We believe that this may scale better than the traditional non-uniform approach. We are not alone in advocating a uniform representation, see for example Soar (Rosenbloom, Laird, & Newell 1993).

The traditional approach to building intelligent, interactive systems is to "compartmentalize" the special-purpose reasoners with different knowledge representations appropriate to the specialized tasks. This is efficient in the initial stages of system building, however as a system matures, components with rich, detailed representations will have to communicate with components having more superficial representations. Knowledge interchange is a serious problem, even in systems that have a common knowledge representation ancestor, such as KL-ONE (Heinsohn *et al.* 1994). One common problem that arises is conflicting ontologies (Traum *et al.* 1996). For example the TRAINS-93 system has many special-purpose components where each component has its own fairly sophisticated representation (Logical Form, Episodic Logic, Conversation Representation Theory, Tyro, Event-based Temporal Logic). In later work with the TRAINS-96 system there is still the stratified architecture, however the components all have more superficial representations, and communicate with each other in KQML (Ferguson *et al.* 1996).

Uniform representations have not been used in traditional intelligent interfaces for reasons of perceived computational and management complexity. In industry, project management is achieved by the use of standards and standards committees that enforce the goal of uniformity. Thus, we feel that management complexity is tractable. Computational complexity is more problematic, as the speed of inference in a monolithic knowledge base has been shown to grow in proportion to the knowledge (Heinsohn *et al.* 1994). There are two answers to this—the first solution is to use distributed computation for reasoning and knowledge base access (Geller 1994). The second is to structure the knowledge for efficient access. This is done by adding meta-knowledge that specifies the nature of the knowledge in the uniform knowledge base. For example, knowledge associated with the current user model would be indexed by meta-facts (that say that these are facts about the user model). Using these meta-facts to index the knowledge base, reasoning about the user model can be done without search. In our current work with B2, a collaborative system that allows medical students to practice their decision-making skills, we have a single knowledge representation and reasoning component that acts as a blackboard for intertask communication and cooperation (McRoy, Haller, & Ali 1997). All knowledge is represented using a semantic network formalism (Shapiro & Rapaport 1992). We structure the knowledge by the "links" between facts in the knowledge base. Thus, for example, all knowledge about, or references to the concept "disease" share the same subnetwork corresponding to that concept. This uniformity allows a concept to be realized as a presentation in multiple ways, depending on the context. The concept "disease" could be realized as a unpleasant graphic, a textual word, or a spoken word, but all realizations would share the common underlying concept.

The Advantages of Mixed-Depth Representations

To maintain a complete model of the interaction, the interaction model must have multiple levels of information, corresponding to different levels of abstraction. Minimally these levels are analogous to the locutionary/illocutionary distinction made by Austin (Austin 1962). Figure 1 gives some simple examples (a plan-based approach using these kinds of communicative acts

Modality	Locution	Illocution
Text	Words	Questions, Answers, Requests, such as `Why?`, `Tell me a story.`
Mouse	click(x,y,type), mode(t)	Command(operation(mode(t), click(type), object-at(x,y)) such as `Select topic of gallstones`
Graphic	barchart	Compare relative magnitudes of probabilities

Figure 1: Example locutions and corresponding illocutions

for multimedia can be found in (Maybury 1993).)

The difficulty with requiring "deep" knowledge representations (i. e., completely interpreted), is that much of an ongoing interaction may be uninterpretable at the moment it occurs, or it may be subject to multiple interpretations or misinterpretation. Mixed-depth representations are more efficient than "deep" representations because computation can be postponed until needed (if ever). Additionally, such representations better tolerate lack of information or ill-formed inputs (minimally because a representation of only the locution is built).

In our current work, the mixed-depth approach allows us to use the same representation framework to produce a detailed representation of requests (which are often interpreted through plan recognition without considering the context) and to produce a partial representation of questions (which tend to require more inference). Moreover, these representations use the same knowledge representation framework that is used by the system to reason about the discourse and the domain—so that the system can reason with (and about) the utterances, if necessary.

Summary

We have argued for *uniform, mixed-depth* knowledge representation in intelligent systems that support flexible, multi-modal, mixed-initiative, two-way interaction. Uniformity has the advantage of minimizing knowledge interchange and can be implemented efficiently. Moreover, enhancements to the reasoner improve the performance of the entire system as all components use the same reasoner. Mixed-depth representations allow "shallow" or "deep" representations as required. Some representations can always be shallow, and this eliminates unnecessary interpretation. For example, mixed-depth representations permit the representation of ill-formed, vague, or ambiguous inputs, which can be further interpreted as necessary.

References

Austin, J. L. 1962. *How to Do Things with Words.* London, England: Oxford University Press. Reprinted 1975.

Bordegoni, M.; G.Faconti; Maybury, M. T.; Rist, T.; Ruggieri, S.; Trahanias, P.; and Wilson, M. 1997. A standard reference model for intelligent multimedia presentation systems. In *Proceedings of the IJCAI '97 Workshop on Intelligent Multimodal Systems.*

Ferguson, G. M.; Allen, J. F.; Miller, B. W.; and Ringger, E. K. 1996. The design and implementation of the trains-96 system: A prototype mixed-initiative planning assistant. TRAINS TN 96-5, Computer Science Dept., University of Rochester.

Funke, D. J.; Neal, J. G.; and Paul, R. D. 1993. An approach to intelligent automated window management. *International Journal of Man-Machine Studies* 38:949–983.

Geller, J. 1994. Advanced update operations in massively parallel knowledge representation. In Kitano, H., and Hendler, J., eds., *Massively Parallel Artificial Intelligence.* MIT Press. 74–101.

Heinsohn, J.; Kudenko, D.; Nebel, B.; and Profitlich, H.-J. 1994. An empirical analysis of terminological representation systems. *Artificial Intelligence* 68:367–397.

Hirst, G., and Ryan, M. 1992. Mixed-depth representations for natural language text. In Jacobs, P., ed., *Text-Based Intelligent Systems.* Lawrence Erlbaum Associates.

Maybury, M. T. 1993. Planning multimedia explanations using communicative acts. In Maybury, M. T., ed., *Intelligent Multimedia Interfaces.* AAAI/MIT Press. 59–74.

McRoy, S.; Haller, S.; and Ali, S. 1997. Uniform knowledge representation for nlp in the b2 system. *Journal of Natural Language Engineering* 3(2).

Neal, J. G., and Shapiro, S. C. 1991. Intelligent multimedia interface technology. In Sullivan, J. W., and Tyler, S. W., eds., *Intelligent User Interfaces.* New York: ACM Press. 11–44.

Rosenbloom, P.; Laird, J.; and Newell, A., eds. 1993. *The Soar Papers: Readings on Integrated Intelligence.* MIT Press.

Shapiro, S. C., and Rapaport, W. J. 1992. The SNePS family. *Computers & Mathematics with Applications* 23(2–5).

Traum, D. R.; Schubert, L. K.; Poesio, M.; Martin, N. G.; Light, M. N.; Hwang, C. H.; Heeman, P. A.; Ferguson, G. M.; and Allen, J. F. 1996. Knowledge representation in the trains-93 conversation system. *International Journal of Expert Systems* 9(1):173–223.

Wahlster, W.; Andre, E.; Finkler, W.; Profitlich, H.-J.; and Rist, T. 1993. Plan-based integration of natural language and graphics generation. *Artificial Intelligence* 63:387–427.