Using Spatial Language in Multi-Modal Knowledge Capture

Kate Lockwood

Northwestern University, Qualitative Reasoning Group
2145 Sheridan Road Room L359, Evanston Illinois
kate@cs.northwestern.edu

Abstract

The ability to understand and communicate spatial relationships is central to many human-level reasoning tasks. People often describe spatial relationships using prepositions (i.e., *in, on*, *under*). Being able to use and interpret spatial prepositions could help create interactive systems for many tasks, including knowledge capture. Here I describe my thesis work modeling the learning and use of spatial prepositions and applying this model to the task of knowledge capture.

Problem Being Addressed

Spatial relationships play an important role in many reasoning tasks such as navigation and solving physics/engineering problems. Because space is such an important component of so many tasks, humans have developed specialized language for describing spatial relationships (i.e. prepositions such as *in* and *on)*. Ideally, intelligent systems would be able to understand and use spatial language in their interactions with human users, particularly when doing visual-spatial tasks.

Unfortunately, many systems modeling the use of spatial prepositions have had serious limitations. Many systems (e.g. Regier, 1995; Gap, 1995) operate only on geometric shapes, not real-world objects. While interesting, psychology has shown repeatedly (e.g. Herskovits, 1985; Coventry & Garrod, 2004) to the functional features of objects also play a prominent role in the use of spatial prepositions. Therefore systems that operate only on geometry are not easily extended to real-world tasks. When systems do address real-world objects, they typically rely on hand-annotated databases of objects restricting their scalability (e.g. Regier & Carlson, 2001).

My thesis work seeks to create a domain-independent, scalable computational model of spatial preposition use. The utility of the model will be shown by using spatial language to resolve ambiguities in cooperative, multi-modal (sketch + text) knowledge capture tasks.

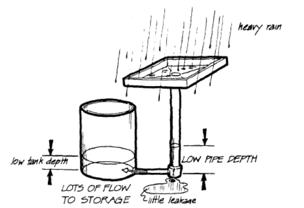
Proposed Plan for Research

My research plan can be divided into two stages, a model of the formation of spatial categories and a system for knowledge capture in a multi-modal environment. This model is being built using a series of sketches depicting real-world objects in simple spatial relationships (*in*, *on*, *above*, *below*, and *left*). All sketches are being created using sKEA (Forbus, Ferguson, & Usher, 2001) the first open-domain sketching system. sKEA operates free of domain restrictions because it does not attempt to do recognition via default, relying instead on an interface mechanism that enables users to label pieces of ink with concepts drawn from a knowledge base. (The KB is a subset of ResearchCyc, containing over 35,000 concepts.) sKEA is ideal for investigating models of spatial language because it allows us to access both perceptual information (computed from the ink in a sketch) and also functional information (based on the link to the underlying KB).

The sketches will be run through SEQL (Kuehne et al, 2000), a relational generalization algorithm. Given a set of cases represented as sets of predicate calculus facts, SEQL divides them into generalizations (categories) based on similarity using structure mapping (Gentner, 1983). The output of SEQL is a set of generalizations, each consisting of a list of facts. Each fact has an associated probability based on the number of cases it appears in (Halstead & Forbus, 2005). By altering the pre-processing of the sketch cases before generalization, I plan to show that sketches can be automatically categorized based on the spatial relationship depicted. Further, by examining the facts associated with each category/generalization I can determine which facts were key to the formation of the generalizations. These sets of facts, along with their associated probabilities, will then be converted into evidential rules for a SpaceCase, my Bayesian model which can label the spatial relationships in novel sketches with the correct preposition.

This process describes how I will bootstrap a system that learns to create and use representations of spatial prepositions to correctly label relationships in sketches. Once this system is sufficiently accurate, i.e., able to correctly label a library of new (not seen in training) sketches based on stimuli from psychological experiments, I plan to demonstrate its usefulness as a disambiguation tool during knowledge capture tasks.

For the knowledge capture component I am building a dialogue system into the Companions System architecture (Forbus & Hinrichs, 2006). This system will help the user and the Companions cooperatively execute multi-modal knowledge capture tasks. Multimodal knowledge capture is important because many textbooks are a mixture of diagrams and text that must be examined in tandem to understand the concepts being communicated. Consider the following example from *Sun Up to Sun Down* (Buckley, 1979), an introductory text on solar energy:



Suppose rain is falling on the rainwater tray. The water will flow down the pipe and start filling the tank. Some of the incoming rainwater will flow into the tank, and some will flow out of the leak …

Figure 1. Text + diagram example from *Sun Up to Sun Down*

To fully understand the concepts being communicated, one must parse the spatial information in the text and understand how it fits with the different parts of the diagram. Spatial prepositions help disambiguate the referenced items in a sketched diagram in order to create a model for the sketched objects. The dialogue system will allow the Companion to query the user to clarify ambiguous or contradictory statements. For example in Figure 1 above, if the system is told "the water level drops" this is a potentially confusing piece of information – there are several instances of water in the diagram. The system could then ask "which water are you talking about?" and display the text in question. The user could respond "the water *in* the pipe" using a spatial preposition to specify the important entity. Now the system knows that the "water level *in the pipe* drops" making its understanding more complete.

To test the system's ability to capture knowledge, I plan to feed it the complete text of *Sun Up to Sun Down* along with sketches of all the diagrams in the book. I plan to do the same for a middle-school science textbook on heat energy. I will then test the system's understanding by using the questions and quizzes in the teacher's manual to provide an external benchmark for evaluation.

Progress to Date

To date I have made progress on several fronts. First, I have shown that SEQL can be used to learn spatial language categories (Lockwood, et. al., 2006). In these experiments I ran SEQL on a set of 50 sketches depicting simple spatial relationships (*in*, *on*, *above*, *below*, and *left*) composed of only geometric shapes. By pre-processing the cases that were sent to SEQL I was able to get the sketches to automatically form categories corresponding to the spatial relationship depicted. I am in the process of creating a new library of sketches with real-world objects to repeat these experiments.

I have also implemented a first-pass of the SpaceCase, Bayesian model for preposition labeling (Lockwood, et al, 2005). In this implementation, the rules were hand-coded rather than automatically generated and were only applicable to *in* versus *on* distinctions. The finished model was able to recreate results from two human subject experiments (Feist & Gentner, 2001; Feist & Gentner, 2003). These experiments are both subsets of the experiments I lay out in my proposed plan and serve as proof-of-concept demonstrations.

Acknowledgements

This work is sponsored by a grant from the Office of Naval Research.

References

Buckley, S. (1979). *Sun Up to Sun Down*. New York: McGraw Hill.

Coventry, K.R., & Garrod, S.C. (2004). *Saying, Seeing, and Acting: The Psychological Semantics of Spatial Prepositions.* Lawrence Erlbaum Associates.

Feist, M.I., & Gentner, D., (2001). An influence of spatial language on recognition memory for spatial scenes. *Proceedings of the 23rd Annual Conference of the Cognitive Science Society.*

Feist, M.I., & Gentner, D. (2003). Factors Involved in the Use of In and On. *Proceedings of the 25th Annual Conference of the Cognitive Science Society.*

Forbus, K., Ferguson, R., & Usher, J. (2001). Towards a computational model of sketching. *IUI'01.* January 14-17, 2001. Santa Fe, New Mexico.

Forbus, K., & Hinrichs, T. (2006). Companion Cognitive Systems: A Step toward Human-Level AI. *AI Magazine,* 27(2): Summer 2006: 83-95.

Gapp, K-P. (1995). An Empirically Validated Model for Computing Spatial Relations. *19th Annual German Conference on Artificial Intelligence.*

Halstead, D., & Forbus, K. (2005). Transforming between Propositions and Features: Bridging the Gap. *Proceedings of AAAI-2005.* Pittsburgh, PA.

Kuehne, S., Forbus, K., Gentner, D., & Quinn, B. (2000). SEQL: Category learning as progressive abstraction using structure mapping. *Proceedings of the Annual Conference of the Cognitive Science Society.*

Lockwood, K., Forbus, K., & Usher, J. (2005). SpaceCase: A model of spatial preposition use. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*. Stressa, Italy.

Lockwood, K., Forbus, K., Halstead, D., & Usher, J. (2006). Automatic Categorization of Spatial Prepositions. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Vancouver, BC, Canada.

Regier, T. (1995). A Model of the Human Capacity for Categorizing Spatial Relations. *Cognitive Linguistics*, 6(1): 63-88.

Regier, T., & Carlson, L.A. (2001). Grounding Spatial Language in Perception: An Empirical and Computational Investigation. *Journal of Experimental Psychology*, 130(2): 273-298.