Representing Biologic Objects as Geometric Constraint Networks

James F. Brinkley, M.D., Ph.D.
Knowledge Systems Laboratory
Stanford University

Artificial intelligence in medicine has been largely concerned with representing and reasoning about medical knowledge for use in diagnosis. Much current work is aimed at developing causal and probabilistic models for allowing greater depth of understanding than is possible with simple rule-based systems. An important component of the causal knowledge utilized by physicians is the spatial relationships inherent in the anatomic, histologic and molecular structure of the human body. An overall theme in my work over the past twelve years has been the development and utilization of schemes for representing biologic structure. The work has included representation of organs for interpreting medical images, and representation of proteins for analyzing NMR spectroscopic data. In both cases a basic approach of representing objects as geometric constraint networks has proven to be useful for capturing much of the essential shape and variation observed.

I received the M.D. from the University of Washington in 1974, after which I worked in the University of Washington Bioengineering Department developing a system for determining organ models from three-dimensional ultrasound data. This work became the subject of my Ph.D. (in medical computer engineering at Stanford), which I completed in 1984. Since that time I have worked as a research associate in the Stanford Knowledge Systems Laboratory, applying AI methods to the problem of determining protein structure from NMR data.

A geometric constraint network may be defined as a set of physical objects, a set of geometric constraints between the objects, and a set of possible locations for the objects. The geometric constraint network may be represented as a graph much like the graphs used to represent inference networks, which are also examples of constraint networks. A solution to a geometric constraint network is one location per object such that all the constraints are satisfied. In the case of biologic objects the constraint network can represent a biologic object class (such as a normal left ventricle or a protein together with a particular set of NMR measurements). A solution to the constraint network would then be a particular instance of the object class compatible with the data (the particular left ventricle compatible with the image data, or the particular protein conformation compatible with the NMR data). The advantage of the network representation is that it appears to allow the essential shape and range of variation of an object class to be captured, which is difficult to do for biologic objects using more rigid geometric models. The interaction of the individual constraints produces a global result that is a good approximation to our intuitive notion of the shape of an object class.

The organ modelling system utilized a custom-built position locating device (developed at the University of Washington), coupled with a commercial ultrasound machine, to provide three-dimensional image information of an organ such as the fetus or the heart. In a "knowledge-driven" version of the system a generic organ model was learned from a set of training instances of an organ of the same shape class. The training instances were used to determine shape constraints between points on the surface of the organ. This model was then used to guide the analysis of ultrasound data from an organ of a similar shape class. Use of the model allowed examination of only one third the data required without the model, and showed the feasibility of plan-guided edge detection. Further work could allow the system to detect abnormal shapes by triggering examination of an alternate model when a particular shape constraint was violated.

The protein modelling system, called PROTEAN, is a large research effort under the direction of Bruce Buchanan and Oleg Jardetzky at Stanford. In this system each atom of a protein is an object in the constraint network, and the various measured and known distances between atoms are the constraints. A solution to the constraint network is a three-dimensional protein conformation that is compatible with the constraints. Because of the large number of atoms and possible locations per atom the major problem in this case is computational complexity. For this reason we are developing a method, called heuristic refinement, for solving such a constraint network at increasingly finer levels of abstraction. For example, we first determine the rough placement of the rigid structures of the protein

such as alpha helices, then refine these to their atomic components. A blackboard system is used to represent and reason about the possible actions to be taken in solving the geometric constraint satisfaction problem, since the order of operations can greatly influence the results. Results to date have shown the ability to obtain abstract pictures of protein structure. We are currently developing methods for refining these structures to the atomic level.

My longer term interest is the development of representations and algorithms for a geometric constraint satisfaction "engine" that could be used to solve many geometric constraint satisfaction problems of this type. The concept of decomposing abstract object representations into their component parts, each of which is recursively represented as a constraint network, could in theory allow the same basic representation to capture the entire range of human anatomy. A learning component could use this representation, together with the constraint satisfaction engine,

to develop generic models of biologic objects that could form a knowledge base of human biologic structure. These models could eventually provide standards of normal and abnormal shape, and would be useful for applying knowledge to the analysis (and design) of biologic objects. Experience in building such a system should also provide insight into the more general issue of multi-dimensional constraint satisfaction, of which I believe both geometric networks as well as inference networks are examples.