DECLARATIVE BIAS FOR STRUCTURAL DOMAINS

Benjamin N. Grosof
IBM T. J. Watson Labs
P. O. Box 704
Yorktown Heights, NY 10598

Stuart J. Russell
Computer Science Division
University of California
Berkeley, CA 94720

ABSTRACT

We present a formal solution to the problem of situation identification in learning of structural concepts. Structural concepts are characterized by the interrelationships and attributes of their *parts*, rather than by just their own direct attributes. Our solution extends the declarative approach to bias of (Russell and Grosof, 1987) by formalizing the beliefs about *relevancy* in a more complex form that expresses the preservation of properties under mappings, using second-order logic to express the existence of isomorphisms. Concept learning, including prediction, analogical inference and single-instance generalization, then emerges as deduction from such *isomorphic* determinations plus instance data.

Situation Identification and Declarative Bias

One of the main tasks faced by a learning agent is the *situation-identification problem*: to identify the aspects of its empirical experience that are relevant to learning a goal concept (Charniak and McDermott, 1985; Bundy et al., 1985). Such relevancy constraints constitute an important part of the bias the agent needs to focus its hypothesis formation. Without a suitably condensed intermediate description language, an agent learning to classify 100x100-binary-pixel visual scenes, for example, would confront a hypothesis space of size $2^{2^{10000}}$.

Declarative Bias (Russell and Grosof, 1987, 1989; Grosof and Russell, 1989; Russell, 1989) is an approach that enables a learning agent to *autonomously derive* these relevancies in a goal-directed manner from its background beliefs. *Determinations* are a form of axiom that express relevancy and can be used to represent appropriate description languages for instances. For example:

$$\{Make(x,u) \wedge Model(x,v)\} \;\succ\; CarValue(x,z) \qquad \stackrel{\text{def}}{\equiv}$$

$$\forall x_1 x_2 uvz. Make(x_1,u) \wedge Make(x_2,u) \wedge Model(x_1,v) \wedge Model(x_2,v)$$

$$\Rightarrow \{CarValue(x_1,z) \equiv CarValue(x_2,z)\}$$

Concept learning, including prediction, analogical inference and single-instance generalization, then emerges as deduction from such determinations plus instance data. An inference engine called IDL that performs this kind of reasoning, including a form of chaining among determinations, has been implemented on top of the MRS theorem prover (Genesereth, 1983) by Lise Getoor (Getoor, 1989). More generally, *default* determinations and other defeasible beliefs can be used to perform inference in a non-monotonic logical system, enabling the system to make inductive leaps and to shift bias when observations contradict the original bias (Grosof and Russell, 1989).

Structural Domains

Determinations work well when the relevant aspects of the situation are features expressible as predicates. "Structural" concepts, however, are characterized essentially by the interrelationships and attributes of the parts of an object, rather than by just the object's own direct attributes (Haussler, 1988; Stepp, 1987). Structural concepts are common in physical and engineering domains.

Let's consider the by-now classic example of the arch concept, in simplified form. An arch is defined as something that has three parts, of which two support the third, while those two do not support each other. Suppose we want to represent a piece of relevancy knowledge, namely that: "the support relationships among the parts of an object determine whether it is an arch". Given such knowledge, the agent could then reason from a positive instance scene g_1 of *Arch* with parts a and b supporting c, to predict that another instance scene g_2 with parts d and e supporting f is also an *Arch*, despite any differences in other attributes such as colors of the parts, geographic location of the scene, etc.. Thus from the following axioms, plus the relevancy constraint, we would like the agent to be able to infer $Arch(g_2)$:

$$\forall xy.Supports(x,y) \Rightarrow \{(x \neq y) \wedge \neg Supports(y,x)$$

$$Arch(g_1) \wedge \{\forall x.Part(x,g_1) \equiv (x = a \vee x = b \vee x = c)\}$$
$$\wedge Supports(a,c) \wedge Supports(b,c) \wedge \neg Supports(a,b) \wedge \neg Supports(b,a)$$
$$\wedge Color(a,red) \wedge Color(b,green) \wedge Loc(g_1,Paris)$$

$$\{\forall x.Part(x,g_2) \equiv (x = d \vee x = e \vee x = f)\}$$
$$\wedge Supports(d,f) \wedge Supports(e,f) \wedge \neg Supports(d,e) \wedge \neg Supports(e,d)$$
$$\wedge Color(e,gray) \wedge Color(f,brown) \wedge Loc(g_2,Stonehenge)$$

Although the relevancy constraint is easy to say in English, and seems a natural piece of commonsense knowledge, it is not possible to express it naturally in the form of a determination (unless we happen to have a *SupportRelationsAmongParts* predicate handy). No simple determination relation exists between the *Supports* and *Part* relations and the *Arch* concept. The relevant similarity between the two scenes is not that the *Supports* relation among the parts of g_1 is equivalent to that in g_2, but rather that it *corresponds* when we match a to d, b to e, and c to f. In other words, the relevant similarity is that the *Supports* relation among the parts of g_1 is equivalent to that in g_2 when the parts of g_1 are *mapped* appropriately to the parts of g_2. What we need is a notion of determination that can express preservation of properties under mapping.

Isomorphic Determinations

Our solution is to generalize the idea of determination accordingly. We define an *isomorphic determination* axiom to say that if the relation A is isomorphic under some bijective mapping E between the "parts" (variables $\underline{x}$ specified by a relation P) of two situations $\underline{w}_1$ and $\underline{w}_2$, then the situations are equivalent with regard to the relation Q as well. Or more briefly, "A, modulo P, determines Q". Isomorphic determination reduces to the old kind of determination, used in the $CarValue$ example above, in the special case when there are no part variables being mapped. Formally:

$$\{A[\underline{x},\underline{w},\underline{y}] \;/\; \underline{x} : \lambda u.P[u,\underline{w}]\} \quad \succ \quad Q[\underline{w},\underline{z}] \qquad \stackrel{\text{def}}{\equiv}$$
$$\forall \underline{w}_1 \underline{w}_2.\{\exists E.\mathcal{BIJECT}(E,P,\underline{w}_1,\underline{w}_2) \wedge \mathcal{ISOMORPH}(E,A,\underline{w}_1,\underline{w}_2)\} \Rightarrow \{\forall \underline{z}.Q[\underline{w}_1,\underline{z}] \equiv Q[\underline{w}_2,\underline{z}]\}$$

where A,P, and Q are (first-order) formulas, $\underline{r}$ denotes a tuple of individual variables, E is a (second-order quantified) predicate variable, and:

$$\mathcal{BIJECT}(E,P,\underline{w}_1,\underline{w}_2) \quad \stackrel{\text{def}}{\equiv}$$
$$\{\forall uv.E(u,v) \Rightarrow \{P[u,\underline{w}_1] \wedge P[u,\underline{w}_2]\}\} \wedge$$
$$\{\forall u.P[u,\underline{w}_1] \Rightarrow \{\exists! v.E(u,v)\}\} \wedge \{\forall v.P[v,\underline{w}_2] \Rightarrow \{\exists! u.E(u,v)\}\}$$

$$\mathcal{ISOMORPH}(E,A,\underline{w}_1,\underline{w}_2) \quad \stackrel{\text{def}}{\equiv}$$
$$\forall \underline{x}_1 \underline{x}_2.\{\bigwedge_{i=1}^{m} E(x_1^i,x_2^i)\} \Rightarrow \{\forall \underline{y}.A[\underline{x}_1,\underline{w}_1,\underline{y}] \equiv A[\underline{x}_2,\underline{w}_2,\underline{y}]\}$$

Here x_j^i means the i^{th} member out of m in the tuple $\underline{x}_j$.

In the arch example, we can capture the relevancy belief as:

$$\{Supports(x, y) \ / \ x, y : \lambda u.Part(u, w)\} \ \succ \ Arch(w)$$

From this axiom plus the others above for the arch example, the conclusion $Arch(g_2)$ is entailed, as desired.

Discussion

In taking a declarative approach to bias in structural domains, we encountered a difficulty with the tool of determinations that served well in simpler, feature-based domains. We develop a more general form, the *isomorphic* determination, to express relevancy of properties preserved under mappings. This enables us to provide a formal solution to the situation identification problem for learning of structural concepts.

Isomorphic determinations (second-order axioms), appear to be a tool to axiomatize the instance language bias when concepts are expressed using the class of first-order existential conjunctive formulae. An interesting question is the precise expressive relationship. We can view the form of the isomorphic determination as giving some insight into what is hard about learning in structural domains: implementing the finding of isomorphisms involves operations such as NP-hard graph-matching. Issues for future work include finding inference rules and implementation techniques for isomorphic determinations.

Acknowledgments

We thank Francisco Corella, Lise Getoor, Norman Haas, and Jean-Louis Lassez for useful discussions.

References

Bundy, A., Silver, B., and Plummer, D. (1985) An Analytical Comparison of Some Rule-Learning Programs. *Artificial Intelligence*, **27**.

Charniak, E., and McDermott, D. (1985) *Introduction to artificial intelligence.* Reading, MA: Addison-Wesley.

Genesereth, M. R. (1983) An Overview of Meta-Level Architecture. In *Proceedings of AAAI-83*, pp. 119-124, Austin, TX: Morgan Kaufmann.

Getoor, L. (1989) *The instance description: How it can be derived and the use of its derivation.* MS report, Computer Science Division, University of California, Berkeley, CA.

Grosof, B. N., and Russell, S. J. (1989) Shift of Bias As Non-Monotonic Reasoning. In Brazdil, P., and Konolige, K., eds., *Machine Learning, Meta-Reasoning, and Logics.* Amsterdam: Kluwer Academic Press.

Haussler, D. (1988) Learning Conjunctive Concepts in Structural Domains. Unpublished manuscript. Computer Science Dept., U.C. Santa Cruz, CA.

Russell, S. J. (1989) *The use of knowledge in analogy and induction.* London: Pitman Press.

Russell, S. J., and Grosof, B. N. (1987) A Declarative Approach to Bias in Concept Learning. *Proceedings of AAAI-1987*, pp. 505-10, Seattle, WA: Morgan Kaufmann.

Russell, S. J., and Grosof, B. N. (1989) A Sketch of Autonomous Learning using Declarative Bias. In Brazdil, P., and Konolige, K., eds., *Machine Learning, Meta-Reasoning, and Logics.* Amsterdam: Kluwer Academic Press.

Stepp, R. E. (1987) Machine Learning From Structured Objects. In *Proceedings of the Fourth International Workshop on Machine Learning*, Irvine, CA: Morgan Kaufmann.