AI Magazine Volume 14 Number 4 (1993) (© AAAI)

Book Reviews

Similarity in Cognition: A Review of *Similarity and Analogical Reasoning*

Ron Sun

■ *Similarity and Analogical Reasoning*, Stella Vosniadou and Andrew Ortony, eds., Cambridge University Press, New York, New York, 1989, 592 pp., ISBN 0-521-38935-6.

The area of similarity and analogy has been of long-standing interest to cognitive scientists and AI researchers. However, recently, there seems to be a new wave of interest, as indicated by many papers, monographs, edited books, and doctoral theses, in exploring aspects of similarity and analogical reasoning from various perspectives. Amid these numerous publications, *Similarity and Analogical Reasoning* surely stands out as the most valuable reference work on the topic, covering especially well the recent advances in the understanding of this topic, with many chapters written by leading researchers. Although it is based on a collection of papers initially presented at the Workshop on Similarity and Analogy, unlike the typical workshop proceedings, this volume is well edited and coherent in both its content and format, with a great deal of cross-references and detailed summary-comment chapters for every part of the book.

Let us look at the book in detail. Because each of these chapters has a different perspective, approach, and organization, I first discuss a number of chapters one by one.

The first part of the book contains five chapters plus a commentary at the end. They mainly deal with the interrelation between similarity and conceptual representation. The main issue here is the role of similarity in representation and in some basic cognitive tasks that build up or use representation, such as categorization, classification decision making, and inference. It was widely believed that (1) categorization is based on similarity and (2) representation is stable, and the same representation is used throughout (in similarity judgment and other tasks). In this part, the first view is challenged by Lance Rips, and the second view is challenged by Lawrence Barsalou. Another interesting issue in this regard is how inference is performed with similarity, which is addressed in this part by Edward Smith and Daniel Osherson and by Ryszard Michalski. Many psychological experiments are presented and analyzed (by Smith and Osherson and by Barsalou), and some interesting theories are proposed (by Michalski and by Barsalou).

In their chapter, Smith and Osherson provide a detailed and intriguing model of how similarity figures into the classification decision-making process. Basically, their model consists of three parts: (1) a frame-based representation of concepts, (2) a procedure for combining conceptual representation, and (3) a model of similarity between an object and a concept. Smith and Osherson adopt the contrast model of Amos Tversky (1977) for the computation of the similarity between two given concepts. They proceed to explain a number of phenomena in human decision making, such as the conjunction fallacy, the base-rate effect, and causal effects of base rate. Their model is based on computing overall similarity between the properties of the concept representation and the properties of the observed objects, using the frame representation. This work represents a step toward a detailed cognitive model of similarity and its role in commonsense reasoning (as opposed to formal reasoning) and decision making.

The next two chapters discuss the instability, or *flexibility*, in conceptual representations and its implication for similarity. Barsalou demonstrates this instability by introducing a number of experiments, which show that conceptual structures vary greatly. The experiments involve property generation and graded structures of concepts. The results show that graded structures vary widely across populations, vary within the same population, vary across contexts, and even vary in the same context with the same individual. In the experiments for property generation, the same kind of instability is observed. Barsalou identifies the source of instability: By some thought experiments and analysis of experimental results, he rules out most of the possible sources; what is left is some possibility involving contextual cues and accessibility. Based on this remaining possibility, he proposes a retrieval-based framework for conceptual representation in which the interplay of context, recent experience, and stable core information creates loosely organized clusters of conceptual knowledge. Barsalou divides this knowledge into context-independent information (CI), context-dependent information (CD), and recent context-dependent information (CD_{rec}). This framework does provide a highly plausible explanation for the variability observed in the aforementioned experiments; however, some doubts arise about the clear-cut differences between the three categories. An important point in Barsalou's work, in my opinion, is that he demonstrates the ever-present role of contexts in similarity measurements. This role is an advance over, and in sharp contrast to, the context-free view of similarity of Tversky.

The chapter by Michalski discusses the flexibility of conceptual representation from the viewpoint of cognitive economy. Because of the need for economical storage and organization of knowledge, an individual concept has a two-tiered structure: "one knows what one remembers, or what one can infer from what one remembers within a certain time constraint" (p. 122). Michalski divides conceptual representation into two components: base concept representation (BCR) and inferential concept interpretation (ICI). This division bears some resemblance to Barsalou's three-part structure (ICI covers both CD and CD_{rec} in a way), but Michalski's model emphasizes inference rather than mere association. Inferences considered in his model for ICI include deductive, analogical, and inductive, which form different layers further and further away from the base representation.

The second part of the book moves up one level and is exclusively concerned with analogy and analogical reasoning: what they are, how they are carried out computationally, how they can be implemented, and so on. In other words, the main issues here are (1) a first principle for analogical reasoning that explains what it is and how it is performed, (2) an understanding of its characteristics (for example, in relation to the generic similarity) so that proper implementation can be devised, and (3) a computational implementation that is adequate and appropriate. Dedre Gentner addresses the first issue; Gentner, David Rumelhart, and Philip Johnson-Laird, respectively, discuss the second issue; and Keith Holyoak and Paul Thagard, as well as John Anderson and Ross Thompson, cover the third issue. This part, which includes five papers and two commentaries, presents not only theories, speculations, and hypotheses but also systems, architectures, and implementations.

Dedre Gentner studies mechanisms for finding analogy and, at the same time, explores similarities between analogical events—how they are differentiated in constructing analogies. The central claim is that "an analogy is a way of focusing on relational commonalities independently of the objects in which those relations are embedded" (p. 201); moreover, "people prefer to map connected systems of relations governed by higher-order relations with inferential import, rather than isolated predicates" (p. 201), that is, the so-called *principle of systematicity*. Gentner proposes a classification of different similarities: *literal similarity*, in which both relational predicates and object attributes are mapped; *analogy*, in which only relational predicates are mapped; and *mere-appearance matches*, in which chiefly object attributes are mapped. Gentner suggests that there are three steps involved in constructing analogy: (1) accessing the potential analog, (2) performing analogical mapping, and (3) judging analogical soundness. According to Gentner, in *analogical mapping processes*, relations must match identically, objects must correspond, and attributes must be ignored. A problem is that a perfect match of relations is required by the theory. The perfect match scenario, according to Gentner, avoids "a host of difficult decisions" but, in my view, at the price of creating some unnecessary rigidity. Gentner's answer is rerepresentation, which, in turn, actually requires answering a whole new set of questions about how this manipulation can be accomplished (such as whether the variations in representation are preexisting or analogy triggered, what the number of variants that should be tried out is, and where the variants should be tried—in the target domain, the source domain, or both). My overall view is that the idea of structural mapping seems to be a sensible one, but the question is how it relates to other mechanisms and considerations.

The next two chapters are concerned mainly with implementational issues. Holyoak and Thagard present a computational model for analogical reasoning. Like Gentner's model, it divides analogy into three similar stages. However, their mechanism for accomplishing this process is unified with the basic rule-based representation and the bidirectional search inference method. Anderson and Thompson talk about analogy in a production-system architecture, PUPS, in which knowledge is represented in a schema structure containing function and form slots (for Lisp functions and their forms). It seems that the only kinds of analogy being performed by the system are mapping functions based on similarities in forms, or vice versa, all with respect to Lisp code.

Rumelhart then presents some fascinating ideas about the characteristics of analogy—and reasoning in general—in the connectionist parallel distributed processing (PDP) framework. Analogy, according to Rumelhart, is part of reasoning by similarity, which is a continuum involving at one end simple remembering and at the other analogical reasoning; in between lie such things as generalization, reminding, and reasoning by examples. Thus, PDP models, because of their similarity-based nature, offer ideal mechanisms for realizing analogy or the continuum in general. In such PDP networks, microfeatures are used to represent functions, properties, and abstract relations, and a particular situation is represented by a cluster of activated microfeatures. Memory accesses in PDP models are determined by similarity between the cue clamped to a network and the previously stored memory patterns in the network. To produce analogy in such a system, Rumelhart hypothesizes that by gradually weakening retrieval cues, releasing the most concrete ones first and progressively releasing more abstract features, the system will be able to recall previous patterns similar to the cue in more and more abstract ways. Analogical reasoning is thus achieved in such systems by mainly keeping the abstract relational microfeatures.

Rumelhart proposes another way for achieving analogical reasoning, that is, "soft clamp," in which input clamps can be overridden, and the rule of thumb is that the more concrete a feature is, the easier it can be overridden. The system finds the overall best fit by "overriding input features only when doing so will lead to a sufficiently good fit to make it worthwhile to give up the goodness contribution given by conforming to the input" (p. 306). To date, to the best of my knowledge, no one has actually figured out a way of implementing this idea based on goodness (or energy functions), although some systems for formal reasoning are based on similar ideas. In fact, there is another possibility (for achieving mapping of abstract relations) left out by Rumelhart, that is, using massively parallel partial match. For example, we can represent things in two different levels: a microfeature level as before and a concept level in which each node represents a distinct event, situation, or concept. Each node in the concept level is connected to all relevant microfeature nodes in the microfeature level (the microfeature nodes are shared by many concept nodes); when a concept node is activated, its corresponding microfeature nodes are then activated, and in turn, all concept nodes sharing some of these microfeatures are activated to various degrees depending on the amounts of overlap in their corresponding microfeature representations. In this way, analogies can be achieved without a search or a settling process; along with analogy, literal similarity and mere-appearance match are also achieved, all at the same time (see Sun [1993] for detailed discussions of such a system).

Part 3 of the book focuses on issues in developmental and instructional uses of similarity and analogy. The six chapters and one commentary emphasize the role of analogy in learning and in knowledge acquisition (including during early childhood), although analogy can help, as well as hamper, learning. The role of analogy in learning is discussed by Ann Brown and by Rand Spiro et al., and the role of analogy in knowledge acquisition is discussed by Brian Ross and by John Bransford et al.; Stella Vosniadou studies the developmental change in the use of analogy. Because part 3 of the book is of marginal interest to AI, I do not discuss it any further.

I would like to discuss the implication of the work regarding similarity and analogical reasoning, especially that contained in this volume, for AI research. There are two aspects to look into: reasoning and representation.

On the representational side, the research described in this book, especially the work by Barsalou and that by Michalski, seems to go against the traditional AI representation, for example, frames (including scripts, schemas, and their logic-based variants). In typical frame-based systems, a concept is represented by a frame that has a number of slots for different properties, each of which can take on some values. This representation is static: It has a fixed set of slots and a fixed set of values, and the representation is used for all purposes and at all times. This type of representation undermines, though not necessarily excludes, the contextual effect and the recency effect in conceptual structures, as advocated by Barsalou, and the inferential character, as advocated by Michalski.

When these effects are taken into account, the conceptual representation scheme has to be expanded to a large extent, and it is not clear what will be the best way for structuring knowledge, especially when the various inference methods identified by Michalski as essential to conceptual representation are implemented. Current research in knowledge representation is far behind in understanding these components and in integrating them into a coherent whole. It is interesting to note that although still in its infancy and somewhat simplistic in character, connectionist research might prove to have an edge in tackling these problems. The research described in this book presents a grand challenge and a future prospect for AI researchers (traditional or connectionistic) in their endeavor to find a better and more cognitively plausible representation scheme.

On the reasoning side, the book also presents major challenges to existing methodologies, especially logic-based approaches. How can the type of reasoning processes as described by Smith and Osherson be formulated in an existing framework for reasoning not simply at a competence level but also at a performance level? How can such similarity-based reasoning be integrated with other reasoning methods? Can analogical reasoning be part of a rule-based system (as in the work of Holyoak and Thagard and that of Anderson and Thompson), or does it have to be something distinct (as suggested by Gentner or by Rumelhart)? There are some major difficulties involved in the reasoning problem, not the least of which is the problem of complexity and reasoning efficiency. It will be worthwhile for future AI research to take advantage of the new insight gained from the psychological research on similarity and analogy.

The work described in this book also makes some philosophical connections. Contemporary theories of meaning view meaning (and mental content) as externally determined (for example, by social and environmental interactions) rather than internally within an individual's mind ("meanings ain't in the head") (Putnum 1975; Burge 1979). No matter what one's philosophical conviction is, one has to admit that social factors play a big role in determining the meaning and the content of a concept or a conceptual structure of an individual. When Barsalou sees the instability of conceptual representation across an individual or across a population, he attributes it to the current context and the recent experience. Another factor in this instability might be, corresponding to externalism in philosophy, the partiality of conceptual representation that inevitably exists in an individual. The full content of a concept is, therefore, determined by social (linguistic) collaboration, that is, the sum of individual representations. Neither Barsalou, Rips, nor Michalski considered such a possibility in explaining their experimental findings or intuitions regarding the instability of conceptual representation. The externalistic view might prove to be of use in guiding future research in this area.

Judged from all respects, this book is a remarkable one. It will remain an invaluable resource for researchers in AI and cognitive science alike for a long time to come.

References

Burge, T. 1979. *Individualism and the Mental.* Midwest Studies in Philosophy, volume 4. Minneapolis, Minn.: University of Minnesota Press.

Putnum, H. 1975. The Meaning of “Meaning.” In *Mind and Knowledge*, ed. K. Gunderson, 215–252. Minneapolis, Minn.: University of Minnesota Press.

Sun, R. 1993. *Integrating Rules and Connectionism for Robust Commonsense Reasoning.* New York: John Wiley and Sons.

Tversky, A. 1977. Features of Similarity. *Psychological Review* 84(4): 327–352.

Ron Sun is currently an assistant professor of computer science at the University of Alabama. Sun received a Ph.D. in computer science from Brandeis University in 1991. He is the author of more than 40 technical papers and articles, and he has written, edited, or contributed to 9 books. He received the David Marr Award in 1991. His research interest centers on the studies of intelligence and cognition, especially in the areas of reasoning, learning, and connectionist models.

Books Received

The following books were received by press time:

- Baars, Bernard J. *A Cognitive Theory of Consciousness.* New York: Cambridge University Press, 1993. 424 pp. $19.95. ISBN 0-521-42743-6.
- Chipperfield, A. J., and Fleming, P. J., editors. *Matlab Toolboxes and Applications for Control.* Los Alamitos, Calif.: IEEE Computer Society Press, 1993. 237 pp. ISBN 0-86341-290-4.
- Gardner, Daniel. *The Neurobiology of Neural Networks.* Cambridge, Mass.: The MIT Press, 1993. 227 pp. $45.00. ISBN 0-262-07150-9.
- Hassoun, Mohamad H., ed. *Associative Neural Memories: Theory and Implementation.* New York: Oxford University Press, 1993. 350 pp. $85.00. ISBN 0-19-507682-6.
- Hopgood, Adrian A. *Knowledge-Based Systems for Engineers and Scientists.* Boca Raton, Fla.: CRC Press, 1993. 387 pp. ISBN 0-8493-8616-0.
- McFarland, David, and Bösser, Thomas. *Intelligent Behavior in Animals and Robots.* Cambridge, Mass.: The MIT Press, 1993. 308 pp. $39.95. ISBN 0-262-13293-1.
- Metropolis, N., and Rota, G.-C., editors. *A New Era in Computation.* Cambridge, Mass.: The MIT Press, 1993. 241 pp. $13.95. ISBN 0-262-63154-7.
- Nigrin, Albert. *Neural Networks for Pattern Recognition.* Cambridge, Mass.: The MIT Press, 1993. 413 pp. $45.00. ISBN 0-262-14054-3.
- Slade, Stephen. *Goal-Based Decision Making: An Interpersonal Model.* Hillsdale, N.J.: Lawrence Erlbaum Associates, 1994. 285 pp. $59.95. ISBN 0-8058-1366-7.
- Van de Velde, Walter, editor. *Toward Learning Robots.* Cambridge, Mass.: The MIT Press, 1993. 165 pp. $23.95. ISBN 0-262-7