In. M. Arbib (Ed.) The Handbook of Brain Theory and Neural Networks. Cambridge, MA: The MIT Press. 1995

Analogy-Based Reasoning

Dedre Gentner and Arthur B. Markman

Introduction

Analogy is important to connectionism by reasons of affinity and difference: *affinity* because analogical reasoning lays claim to connectionist advantages, like representation completion, similarity-based generalization, graceful degradation, and learning (Barnden, 1994); *difference* because analogy is a structure-sensitive process that involves the comparison of *systems of relations* between items in a domain, whereas connectionist models have concentrated on *correlational* representations that are sensitive to the statistical structure of the input. For both reasons, modeling analogy will be a key challenge for connectionist systems.

This chapter outlines four principles of analogy that have emerged from psychological studies. These serve to constrain models of comparison and can act as benchmarks for the development of connectionist models. After summarizing the principles, we discuss connectionist models of analogy. Our discussion takes place at Marr's *informational* and *algorithmic levels*, at which cognition is explained in terms of representations and associated processes. We will not evaluate the models in terms of brain functioning, not only because this seems premature, but because we believe a computational model should first justify itself as a cognitive account.

Four Tenets of Analogical Reasoning

Analogy is the perception of relational commonalities between domains that are dissimilar on the surface. These correspondences may suggest inferences about the target domain. Analogy has been widely studied in humans, and four general principles can be derived from this work. Analogy involves (1) structured pattern matching; (2) structured pattern completion; (3) a focus on common relational structure rather than on common object descriptions; and (4) flexibility in that (a) the same domain may yield different interpretations in different comparisons, and (b) a single comparison may yield multiple distinct interpretations. Any model of analogy must account for these phenomena.

Structured Pattern Matching

The defining characteristic of analogy is the alignment of relational structure. Alignment involves finding *structurally consistent* matches (those observing parallel connectivity and one-to-one correspondence). *Parallel connectivity* requires that matching relations have matching arguments; *one-to-one correspondence* limits any element in one representation to, at most, one matching element in the other representation (Gentner, 1989; Holyoak and Thagard, 1989). For example, when comparing the atom (a *target*) to the solar system (a *base*), the sun corresponds to the nucleus and the planets to the electrons because they play similar roles in a common relational structure: e.g., **revolve**(sun, planets) and **revolve**(nucleus, electron). The sun is not matched to both the nucleus and the electron, as that violates one-to-one correspondence. Another characteristic of analogy is *relational focus*: objects correspond by virtue of playing like roles and need not be similar (e.g., the nucleus need not be hot).

There is considerable evidence that people can align two situations, preserving connected systems of commonalities and making the appropriate lower-order substitutions. For example, Clement and Gentner (1991) showed people analogous stories and asked them to state which of two assertions shared by base and target was most important to the match. Subjects chose the assertion connected to matching causal antecedents. Their choice was based on both the goodness of the local match and its connection to a larger matching system. This finding demonstrates that analogies seek *connected systems of relations* rather than isolated relational matches.

Effects of structure are also seen in perceptual comparisons. Lockhead and Pomerantz (1991) have reviewed studies demonstrating that higher-order perceptual relations, like symmetry, aid perceptual discriminations. Subjects could more easily discriminate between the stimuli) and (when they were presented in the context of a third identical element,), yielding the pair)) versus (). The added component introduced a higher-order symmetry relation in the right pattern, making the two patterns more discriminable. These configural effects are evidence for the role of relational structure in perceptual similarity.

Structured Pattern Completion

Analogical reasoning also involves the mapping of inferences from one domain to another. Thus, a partial representation of the target is completed based on its structural similarity to the base. For example, Clement and Gentner (1991) extended the findings described earlier by deleting some key matching facts from the target story and asking subjects to make a new prediction about the target based on the analogy with the base story. Consistent with the previous result, subjects mapped just those predicates that were causally connected to other matching predicates.

Flexibility (1): Same Term → Different Interpretations

Flexibility, a hallmark of connectionist models, is also a virtue of analogy. Barnden (1994) suggests that analogical reasoning may reconcile connectionism's flexibility with symbolic AI's rigidity. One way that analogical reasoning is flexible is that the same item can take part in many comparisons, and different aspects of its representation can participate in each comparison. Spellman and Holyoak (1992) compared politicians' analogies for the Gulf War. Some likened it to World War II, implying that the United States was acting to stop a tyrant, whereas others likened it to Vietnam, implying that the United States had entered into a potentially endless conflict between two other opponents. Different comparisons highlighted different features of the target, the Gulf War. Flexibility is also evident when the same base term is combined with different targets. The comparison statement, "The atom is like a solar system," conveys a central attractive force causing its satellites to revolve, whereas "The campfire is like a solar system" conveys a central source of heat and light.

Flexibility (2): Same Comparison → Multiple Interpretations

A second kind of flexibility is that a single base-target comparison may give rise to multiple distinct interpretations. For a comparison like "Cameras are like tape recorders," people can readily provide an object-level interpretation ("Both are small mechanical devices,") or a relational interpretation ("Both record events for later replay").

Despite this flexibility, people generally maintain structural consistency within an interpretation. In one study, Spellman and Holyoak (1992) asked subjects to map Operation Desert Storm onto World War II (WWII). They asked, "If Saddam Hussein corresponds to Hitler, who does George Bush correspond to?" Some subjects chose Franklin Delano Roosevelt, whereas others chose Winston Churchill. The key finding was that, when asked to make a further mapping for the United States in 1991, subjects chose structurally consistent correspondences. Those who mapped Bush to Roosevelt usually mapped the U.S.-1991 to the U.S.-during-WWII, and those who mapped Bush to Churchill mapped the U.S.-1991 to Britain-during-WWII.

An extreme case of conflicting interpretations is *cross-mapping*, in which the object similarities suggest different correspondences than do the relational similarities. For example, in the comparison between "Spot bit Fido" and "Fido bit Rover," Fido is cross-mapped. When presented with cross-mapped comparisons, people can compute both alignments. Research suggests that adding higher-order relational commonalities increases people's preference for the relational alignment (Gentner and Toupin, 1986; Markman and Gentner, 1993). This ability to compute relational interpretations (even for cross-mappings) is central to human analogizing.

This flexibility and the ability to process cross-mappings have significant implications for the comparison process because they mean that simulations cannot simply be trained to generate a particular kind of interpretation. Rather, the comparison process must be able to determine both object matches and structural matches and to attend selectively to one or the other.

Connectionism and Analogical Mapping

As connectionist techniques for structured representation have become available (Hinton, 1991; also see STRUCTURED CONNECTIONIST MODELS), several models of analogy have emerged. The first model, ACME (Holyoak and Thagard, 1989), is a localist system that combines the constraints of structural consistency, semantic matching (via a table of predicate similarities), and pragmatic centrality (by activating nodes related to goals and correspondences known in advance) in a constraint satisfaction network. The analogy's interpretation reflects the best solution satisfying these constraints. The interpretation need not maintain structural consistency, so spontaneous inferencing is not generally possible. Hummel et al. (1994) point out that the implementation of the pragmatic constraint often causes the important node(s) to map to everything in the other analog. Finally, because ACME settles on a single interpretation of an analogy, its solution to cross-mappings merges the object and relational interpretations.

To escape the difficulty with pragmatic bindings, Hummel et al. (1994) propose a distributed connectionist model—IMM—that makes use of temporal synchrony in unit firing to encode relations. In IMM, the connections between relations and their arguments are maintained by having individual units, which represent concepts, fire in phase with units that represent particular relational bindings (see STRUCTURED CONNECTIONIST MODELS). For example, to represent **kiss**(John, Mary), nodes for kiss, John, and *agent* fire in phase. Nodes for kiss, Mary, and *patient* also fire in phase. To map a base to target, trainable connections are set up between object nodes and relational binding nodes. As activation spreads through the system, corresponding relations in the base and target begin to fire in phase as connections between analogically corresponding binding nodes and object nodes are strengthened. These connection strengths determine the corresponding items in base and target. IMM finds only a single interpretation of a comparison. If there are two competing interpretations, the model gives equal connection strengths to correspondences consistent with both interpretations. Thus, instead of distinguishing between two possible interpretations, the model finds a single merged interpretation that it considers to be weak.

A model of analogy has also been constructed using *tensor product representations* (Smolensky, 1991). In a tensor product, two vectors **X** and **Y** are bound by taking the outer product of these vectors, $\mathbf{YX}^T$. The outer product normally forms a matrix, but a vector can be constructed from this matrix by concatenating its columns. Given **X**, the vector **Y** can be obtained as $\mathbf{YX}^T\mathbf{X}$ if **X** is a unit vector. Variable bindings can thus be captured by using one vector to represent a predicate and the other to represent its argument.

Tensor products have been used in a distributed connectionist model—STAR—that performs a:b::c:d analogies (Halford et al., in press). STAR represents binary relations ($\mathbf{R}(\mathbf{a}, \mathbf{b})$) using tensor products of rank 3 (which are like the binary tensor products just described except that three vectors are bound together). In this model, long-term memory consists of a matrix of tensor products corresponding to various relations the system knows about. To process an analogy, the model takes the **a** and **b** terms and probes long-term memory to find a relation between them. It then takes this relation and the **c** term of the analogy and finds a fourth term that shares that relation with the **c** term. This model successfully uses a distributed connectionist representation to perform a one-relation analogical reasoning task. Presently, STAR cannot generate multiple distinct interpretations of a comparison. If the system knows many different items that could be the answer to the analogy, the output vector is a combination of them all. In addition, this model does not make use of higher-order relational structure to constrain its matches. Efforts are being made to remedy this problem.

Related Approaches

Other connectionist models have addressed some of the challenges that analogy poses to connectionism. The first challenge is the representation of higher-order relations. A promising approach to higher-order relations is holographic reduced representations (HRR) (Plate, 1994) whereby two vectors are associated using a circular convolution of an outer product matrix, yielding a vector with the same dimensionality as the original vector. A relation can be stored by adding a vector corresponding to the relation to convolutions of vectors representing the arguments of the relation and the relational roles they fill. This representation can be augmented by adding contextual information to the fillers. Hierarchical relational systems can be represented by using the vector corresponding to a whole relation as an argument to another relation. In this system, vectors corresponding to similar relations over similar objects have a high dot product. Another advantage is that vectors corresponding to similar relational structures are detected as similar, even if they contain cross-mappings.

A second challenge is structured pattern completion. Pattern matching and completion are central features of many connectionist models. In the simple, feature-based Brain State in a Box (BSB) model of Anderson et al. (1977), partial vectors are completed based on their similarity to learned vectors. It is not clear whether the pattern completion done in these models is the same as that done in analogical mapping. In many connectionist models, completion of a partial vector is based on the geometric projection of that vector onto the space of connection weights (i.e., the distance of the vector from an attractor state). This contrasts to the structured pattern completion common in analogical reasoning that we described earlier.

A third challenge is producing two or more distinct interpretations of a single analogy. Most current connectionist models of analogy yield one interpretation. Interestingly, models of the perception of the Necker cube have been able to maintain multiple stable states in the same network. In one localist model, units represent possible interpretations of corners of the Necker cube (Rumelhart et al., 1986). Excitatory connections link nodes that yield a consistent interpretation of the cube, and inhibitory connections link nodes for inconsistent interpretations. Given different random starting configurations, the system tends to settle in one of the two states representing consistent interpretations of the cube. This network displays two human-like features: (1) if one interpretation is significantly better than the other, it usually settles on the better interpretation; and (2) if multiple interpretations of the cube are equally good, its choice depends on its starting state.

Kawamoto and Anderson's (1985) model of the Necker cube is a distributed autoassociative system that learns the consistent interpretations of the cube. After learning, the network is allowed to settle on a stable state. This stable state is then *anti-learned* by subtracting the outer product matrix of the vector representing the stable state $\mathbf{XX}^T$ from the connection matrix (thereby flattening the basin of attraction). When the network is allowed to settle again, it settles on a different stable state. These techniques may be useful for capturing humans' ability to process multiple interpretations of a single comparison.

Discussion

Analogical reasoning relies heavily on correspondences in the structural relations of two domains. This leads to the seven hallmark phenomena, summarized in Table 1, that pose a challenge to connectionist models. No existing model exhibits the kind of structural alignment and mapping that people do. Three possibilities remain. First, connectionist models may be ill-suited to model analogy and should be augmented by symbolic representations. Second, advances in models of structured representations and comparison may enable connectionist models to succeed in addressing the seven empirical challenges. Third, a hybrid connectionist/symbolic system like Barnden's *implementational connectionism* (Barnden, 1994) may be most appropriate. The empirical evidence described here will be helpful in constraining the development of these models.

Table 1. Seven Phenomena of Analogy

1. Structural consistency	Analogical mapping involves one-to-one correspondences and parallel connectivity.
2. Candidate inferences	Analogy inferences are generated via structured completion.
3. Systematicity	People prefer connected relations rather than collections of isolated relations.
4. Relational focus	Relational matches are made whether or not the objects making up the relations also match.
5. Flexibility (1)	Analogy allows multiple interpretations of a single item in different comparisons.
6. Flexibility (2)	Analogy allows multiple interpretations of a single comparison.
7. Cross-mapping	Though difficult, cross-mappings are generally interpreted relationally although the competing object similarities are perceived.

Acknowledgment. This work was supported by ONR grant N00014–89–J1272, awarded to Dedre Gentner and Kenneth Forbus.

Road Map: Connectionist Psychology
Background: Associative Networks
Related Reading: Semantic Networks

References

Anderson, J. A., Silverstein, J. W., Ritz, S. A., and Jones, R. S., 1977, Distinctive features, categorical perception and probability learning: Some applications of a neural model, *Psychol. Rev.*, 84:413–451.

Barnden, J. A., 1994, On using analogy to reconcile connections and symbols, in *Neural Networks for Knowledge Representation and Inference* (D. S. Levine and M. Aparicio, Eds.), Hillsdale, NJ: Erlbaum.

Clement, C. A., and Gentner, D., 1991, Systematicity as a selection constraint in analogical mapping, *Cognit. Sci.*, 15:89–132.

Gentner, D., 1989, The mechanisms of analogical learning, in *Similarity and Analogical Reasoning* (S. Vosniadou and A. Ortony, Eds.), New York: Cambridge University Press, pp. 199–241.

Gentner, D., and Toupin, C., 1986, Systematicity and surface similarity in the development of analogy, *Cognit. Sci.*, 10:277–300.

Halford, G. S., Wilson, W. H., Guo, J., Wiles, J., and Stewart, J. E. M, in press, Connectionist implications for processing capacity limitations in analogies, in *Advances in Connectionist and Neural Computation Theory*, vol. 2: *Analogical Connections*, (K. J. Holyoak and J. Barnden, Eds.), Norwood, NJ: Ablex.

Hinton, G. E., 1991, *Connectionist Symbol Processing*. Cambridge, MA: MIT Press. ◆

Holyoak, K. J., and Thagard, P., 1989, Analogical mapping by constraint satisfaction. *Cognit. Sci.*, 13:295–355.

Hummel, J. E., Burns, B., and Holyoak, K. J., 1994, Analogical mapping by dynamic binding: Preliminary investigations in *Advances in Connectionist and Neural Computation Theory*, vol. 2: *Analogical Connections* (K. J. Holyoak and J. A. Barnden, Eds.), Norwood, NJ: Ablex.

Kawamoto, A. H., and Anderson, J. A., 1985, A neural network model of multistable perception, *Acta Psychol.*, 59:35–65.

Lockhead, G. R., and Pomerantz, J. R., 1991, *The Perception of Structure*, Washington, DC: American Psychological Association. ◆

Markman, A. B ., and Gentner, D., 1993, Structural alignment during similarity comparisons. *Cognit. Psychol.*, 23:431–467.

Plate, T. A., 1994, Estimating structural similarity by vector dot products of Holographic Reduced Representations, in *Advances in Neural Information Processing Systems 6* (J. D. Cowan, G. Tesauro, and J. Alspector, Eds.), San Mateo, CA: Morgan Kaufmann.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., and Hinton, G. E., 1986, Schemata and sequential thought processes in PDP models, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (D. E. Rumelhart, J. L. McClelland, and PDP Research Group, Eds.), Cambridge, MA: MIT Press, pp. 7–57.

Smolensky, P., 1991, Tensor product variable binding and the representation of symbolic structures in connectionist systems, in *Connectionist Symbol Processing* (G. E. Hinton, Ed.), Cambridge, MA: MIT Press.

Spellman, B. A., and Holyoak, K. J., 1992, If Saddam is Hitler then who is George Bush? Analogical mapping between systems of social roles, *J. Pers. Soc. Psychol.*, 62:913–933.