A Model of Analogical Retrieval Using Intermediate Features

Mark Alan Finlayson & Patrick Henry Winston
{markaf, phw}@mit.edu
Computer Science and Artificial Intelligence Laboratory, MIT, 32 Vassar St., Cambridge, MA 02139 USA

Background: Two Types of Retrieval

This work applies to a model of analogical reasoning reviewed in French (2002) and outlined in *Fig.* 1. According to this model, before items from memory can be structurally mapped *à la* Structure-Mapping Theory (Gentner, 1983), those items must be retrieved via a computationally efficient process.

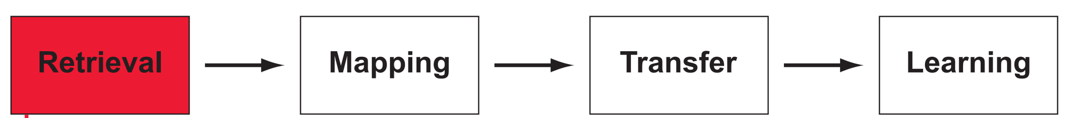


Fig 1: Process model of analogy.

Results from psychological studies indicate two main types of retrieval:

(1) Surface similarity - e.g., similar objects and first-order relations - heavily influences retrieval when the semantic domain and subject population are uncontrolled. (Gick & Holyoak, 1980; Rattermann & Gentner, 1987)

(2) Structural similarity heavily influences retrieval when we consider experts in their domain of expertise. (Chi *et al*, 1981; Schoenfeld, 1982; Shneiderman, 1977)

Symbolic Feature-Based Model

In analogy with Ullman's results, we see the following correspondences:

Surface-Similarity Dominated Retrieval	→	Retrieval based on small features
Structural-Similarity Dominated Retrieval	→	Retrieval based on intermediate features

Our model implements features in symbolic graph representations as shown in *Fig.* 3. To each representation in memory the model assigns a score which is intended to represent the preference for recall relative to a stimulus representation.

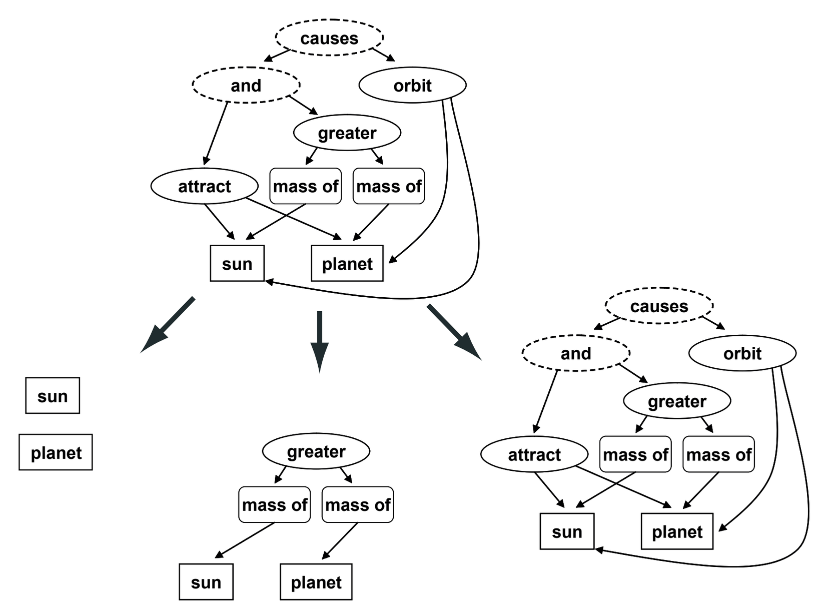


Fig 3: Breaking a representation into features of small (left), intermediate (middle) and large (right) sizes.

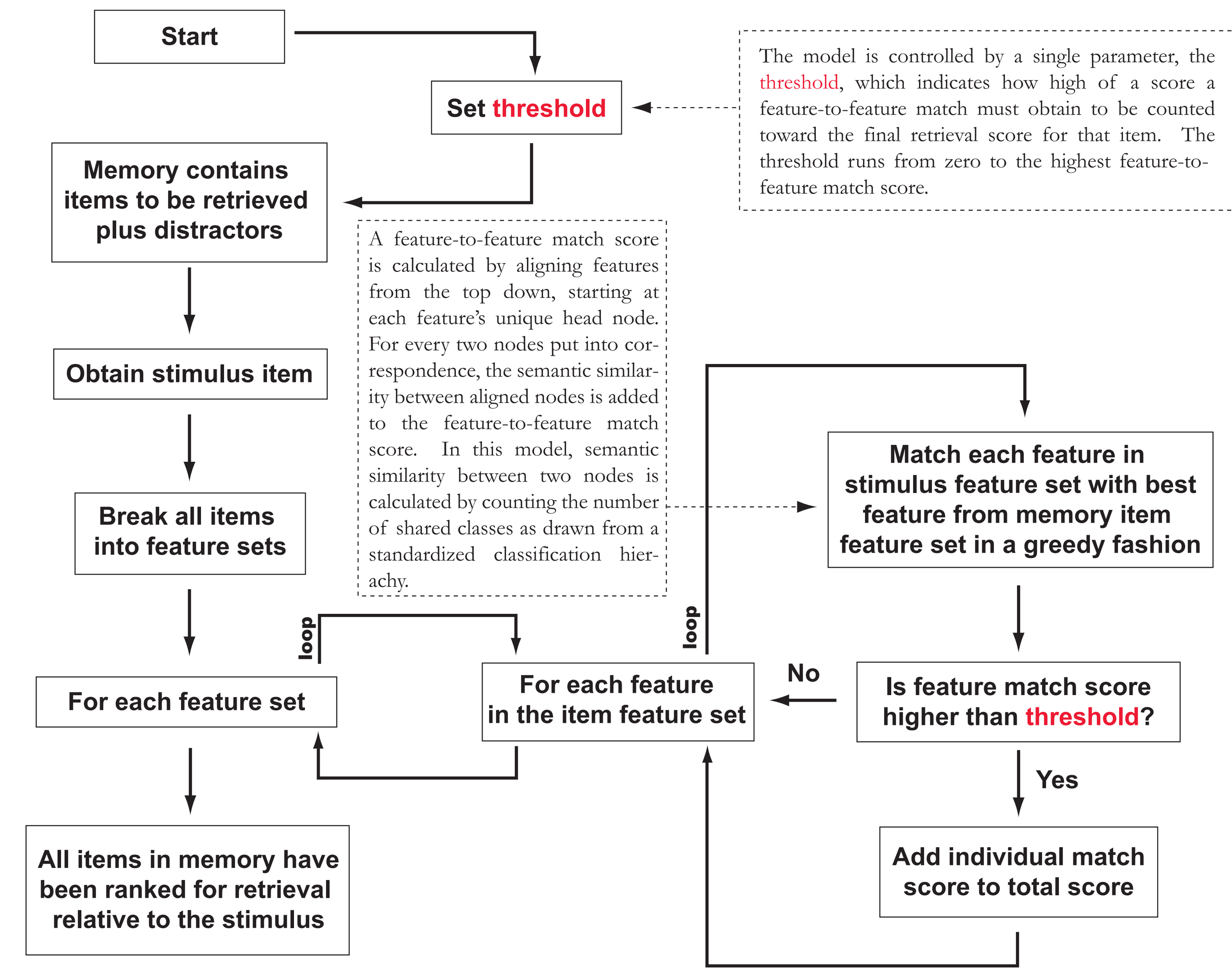


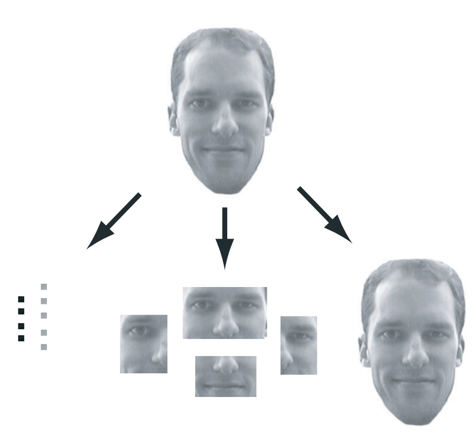
Fig 4: Flowchart of the model.

Inspiration: Image Classification

(Ullman *et al*, 2002)

Ullman's technique was intended for visual classification. Suppose that you want to identify an image such as a face by comparing it with a database of pre-classified images. Ullman's strategy was to break the image into subimages (*features*) of different sizes (*Fig.* 2) and see how often you find those subimages in a particular class.

Fig 2: Breaking an image into features of small (left), intermediate (middle) and large (right) sizes.



Intermediate Features Maximize Information

Ullman discovered that small features, like those in *Fig.* 2, are found in nearly all images, regardless of class. Large features, on the other hand, are rarely found even in the set of pictures of the same class.

In other words, small features are too general, and large features are too specific. What best identifies an image within a class are features of an intermediate size and complexity.

Results

We ran our model on a set of 56 representations of particular international conflicts. Fourteen of these representations were stimuli, and the other 42 representations were automatically generated from the stimuli to be either superficially-related to that stimulus, structurally-related, or both. Results are shown in *Fig.* 5.

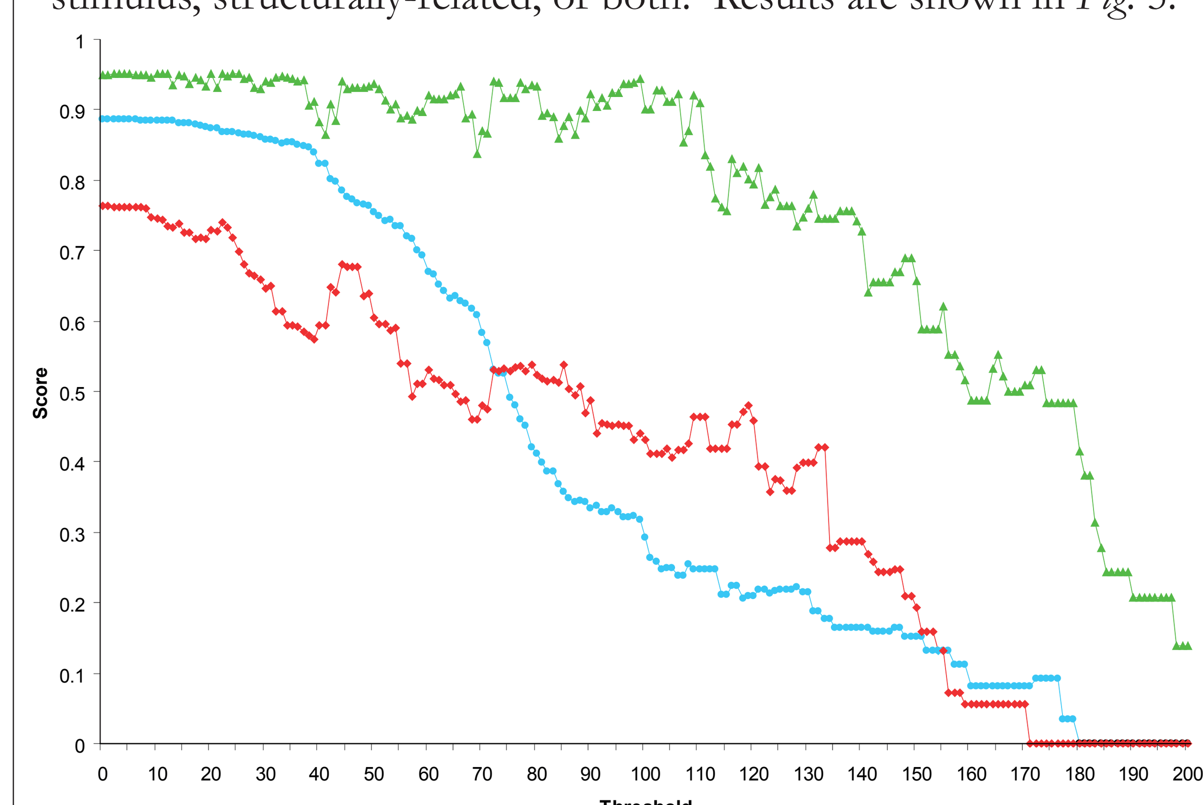


Fig 5: Results of a simulation on our dataset which show the dominance of structurally-related representations (red) over superficially-related representations (blue) at intermediate feature sizes. Representations that are both superficially- and structurally-related are also shown (green).

Contributions

- Formulated a feature-based model to account for both superficial and structural retrieval.
- Implemented the model and tested it on a dataset of our own construction.
- Demonstrated the value of intermediate features for analogical retrieval.

Future Work

Run the model on other datasets from the literature, including:

- Gentner & Forbus's Karla the Hawk dataset (Forbus *et al*, 1994)
- Thagard & Holyoak's ARCS dataset (Thagard *et al*, 1990)

Acknowledgements

This work is supported in part by the National Science Foundation under Grant Numbers 0218861 and IIS-0413206.

References

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, **5**, 121-152.

Forbus, K. D., Genter, D., & Law, K. (1994). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, **19**, 141-205.

French, R.M. (2002), The computational modeling of analogy-making. *Trends in Cognitive Sciences*, **6**, 200-205.

Gentner, D. (1983), Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, **7**, 155-170.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology*, **12**, 306-355.

Rattermann, M. J., & Gentner, D. (1987). Analogy and similarity: Determinants of accessibility and inferential soundness. In *Proceedings of The Ninth Annual Meeting of the Cognitive Science Society* (p. 23-35). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Schoenfeld, A. H. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **8**, 484-494.

Shneiderman, B. (1977). Measuring computer program quality and comprehension. *International Journal of Man-Machine Studies*, **9**, 465-478.

Thagard, P., Holyoak, K. J., Nelson, G., & Gochfeld, D. (1990). Analog retrieval by constraint satisfaction. *Artificial Intelligence*, **46**, 259-310.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, **5**, 682-687.

26th Annual Meeting of the Cognitive Science Society, August 2004, Chicago, IL