Graphs in space: A domain-general and level-spanning tool for representing structure

Fred Callaway   Cornell University   Psychology

Introduction

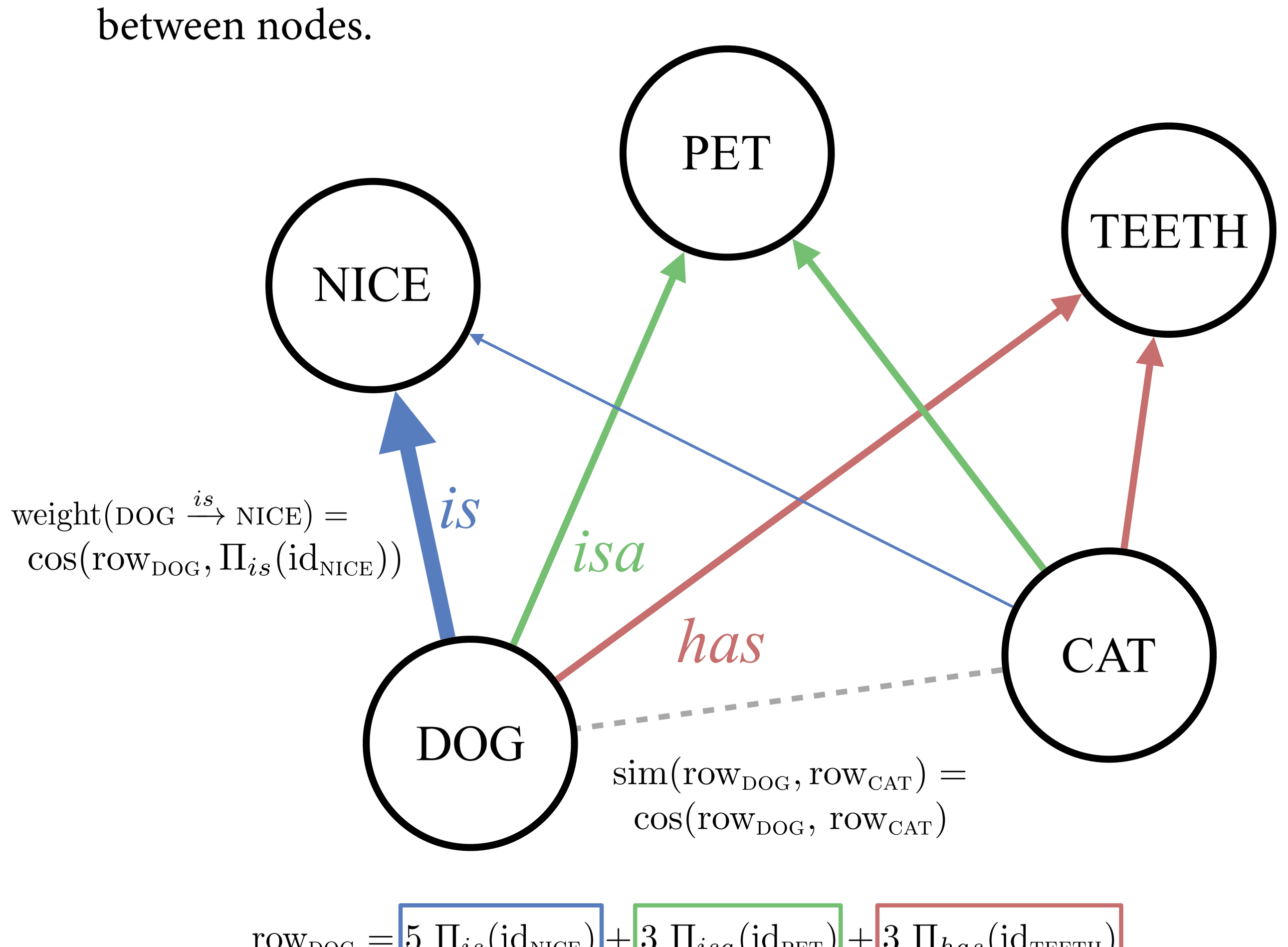
The representation of structure is a fundamental prerequisite for sophisticated cognition. Whether an agent wants to navigate a physical environment, select a socially successful mate, or read an undergraduate's half-baked research poster, she will need an internal model of the relevant system. Constructing models of these internal representations makes up a large part of research in cognitive science.

As observed by Marr (1982), cognition can—indeed must—be described at multiple levels of analysis. Thus, to fully understand how minds represent structure, we must pose explanations in terms of neurons as well as functional abstractions such as syntactic trees. This raises an important question: **Can we systematically relate functional abstractions to their neural implementations?**

We suggest that an implementation of a graph with a Vector Symbolic Architecture may serve as such a bridge. Graphs are simple, yet powerful, tools for modeling high-level representations (Tenenbaum et al. 2011), while VSAs provide a link between the symbolic representations of high-level models and the distributed representations of the brain (Gayler 1998; Kanerva 1988; Plate 1995). By uniting these frameworks, we pave the way for the potential unification of all three of Marr's levels of analysis.

VectorGraph implementation

A vector representation of a weighted, directed, labeled graph is constructed incrementally with a technique called *random indexing* (Kanerva et al. 2000). The dimensionality of the vectors is fixed ahead of time in the range of 1,000 to 10,000. New nodes and edges can be added at any point, allowing for psychologically realistic learning algorithms. The outgoing edges of a node are represented in a *row vector*, which is the sum of the unique *index vectors* of the targeted nodes, optionally permuted by an *edge-label* vector. Cosine similarity is used to measure the weight of a single edge as well as the similarity between nodes.



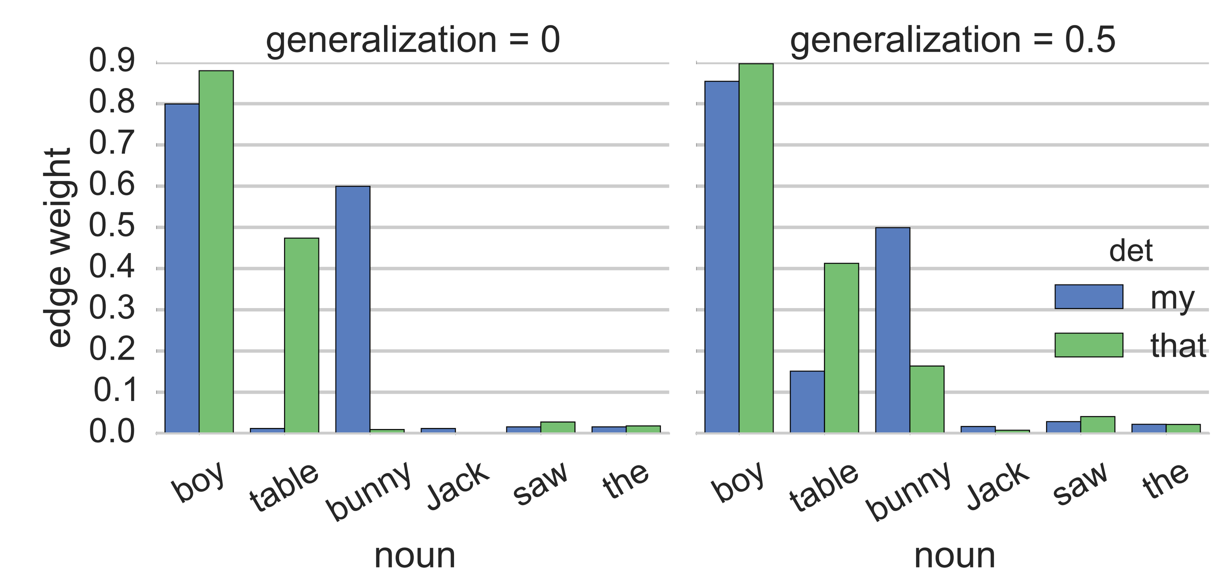
Generalization

Problem: For knowledge to be useful in an ever-changing environment, agents must be able to generalize from particular experiences to inform decisions in a novel situation.

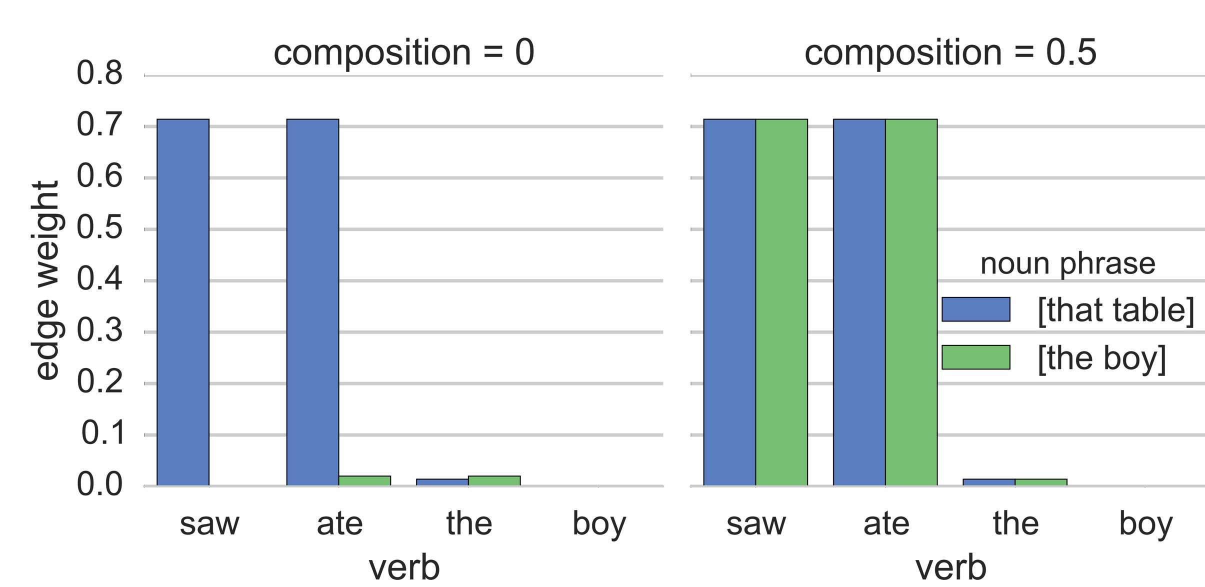
Solution: One strategy is to assume that items which are similar in many ways are likely to be similar in other ways as well.

$$\mathrm{gen}(\mathrm{row}_x) = \sum_{n \in G} \mathrm{row}_n \, \mathrm{sim}(x, n)$$

Results: A bigram model assigns a non-zero edge weight to determiner-noun pairs it has never seen before ("my table", "that bunny"), generalizing based on knowledge about other determiners and nouns.



Compositionality



Problem: The properties of many items can be inferred from the properties of their parts. Reasoning about compositional structure is essential for language, and perhaps higher-order thought in general.

Solution: A particularly easy type of compositionality can be expressed with categorical rules, such as those found in phrase structure grammars. These rules can be approximated in vector space without discrete categories.

$$\mathrm{row}_{ab} = \sum_{xy \in G} \mathrm{row}_{xy} \sqrt{\mathrm{sim}(a, b)\, \mathrm{sim}(x, y)}$$

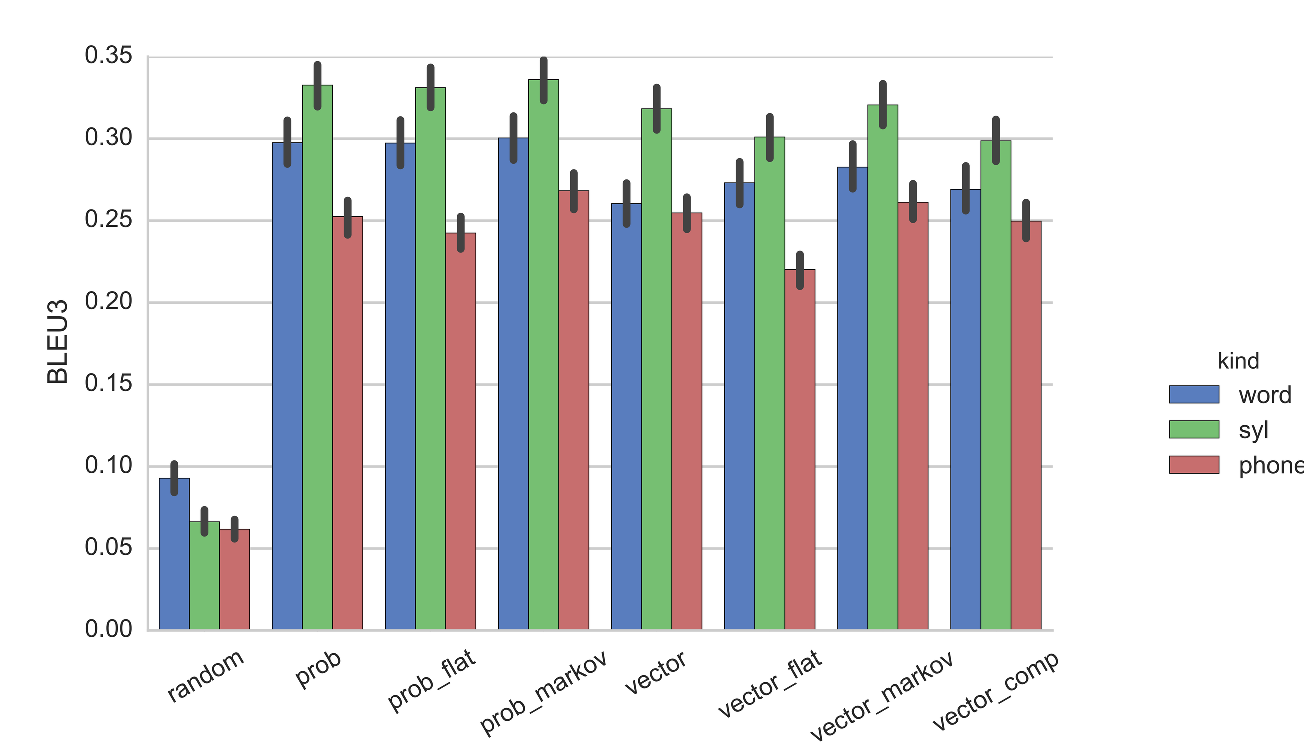
Results: This allows a chunking bigram model to predict that a newly constructed noun phrase ("the boy") will be followed by a verb, generalizing based on the properties of elements with similar constituents ("that table").

Language acquisition

Problem: Learn a language given only a small corpus of transcribed child-directed speech.

Solution: Track transitional probabilities between words and phrases (Kolodny et al. 2015; McCauley and Christiansen 2011; Solan et al. 2005). Represent language knowledge as a VectorGraph with words and phrases as nodes and transitional probabilities as edges.

Results: Much to our surprise, we did not solve Linguistics. Nümila preforms best when tracking transitional probabilities between words only (labeled *markov*), indicating that the algorithm for learning and/or using chunks is faulty. However, the model performs similarly when using a VectorGraph compared to a traditional adjacency matrix graph (labeled *prob* because edge weights are normalized). This indicates that the VectorGraph can be used in medium-scale cognitive models.



Conclusion

The computational theory of mind has provided many important insights. While neuroscience explains what brains do, and classical psychology explains what people do, computational theories may serve as the best explanations for what *minds* do. However, this powerful approach brings with it new theoretical and methodological challenges that are the subject of considerable debate.

One such challenge lies in the interpretation and comparison of models posed at different levels of analysis. For example, Bayesian models (posed at the *computational* level) are often contrasted with connectionist models (posed at the *algorithmic* level). However, given that these models are posed at different levels of analysis, it is not clear to what extent they are conflicting (Griffiths et al. 2012).

We suggest that these issues can be approached with an attempt to systematically relate models at different levels of analysis. In this way, we may begin to answer critical questions such as the degree to which distributed representations in the brain are implementations of symbols, as opposed to being importantly different, and perhaps more powerful than symbols (Chalmers 1990).

By uniting a high-level representational tool (graphs) with a level-spanning cognitive architecture (VSAs), we hope to have made a small step towards answering these questions, and constructing a unified and complete theory of cognition.

References

Chalmers, D. (1990). Why Fodor and Pylyshyn were wrong: The simplest refutation. In *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society.*

Gayler, R. (1998). Multiplicative binding, representation operators and analogy. In *Advances in analogy research: Integration of theory and data from the cognitive, computational, and neural sciences.*

Griffiths, T. L., Vul, E., and Sanborn, A. N. (2012). Bridging levels of analysis for probabilistic models of cognition. *Current Directions in Psychological Science.*

Marr, D. (1982). *Vision.* San Francisco, CA: W. H. Freeman.

McCauley, S. M., & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The CAPPUCCINO model. In *Proceedings of the 33rd annual conference of the Cognitive Science Society.*

Kanerva, P. (1988). *Sparse distributed memory.* MIT press.

Kanerva, P., Kristofersson, J., & Holst, A. (2000). Random indexing of text samples for latent semantic analysis. In *Proceedings of the 22nd annual conference of the cognitive science society.*

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022).

Acknowledgements

I thank my committee members, John Hale, Morten Christiansen, and Shimon Edelman, for their comments. I am especially grateful to my advisor, Shimon, for his tenacious support and guidance.