Generative Inferences Based on a Discriminative Bayesian Model of Relation Learning

Dawn Chen, Hongjing Lu, and Keith J. Holyoak
University of California, Los Angeles

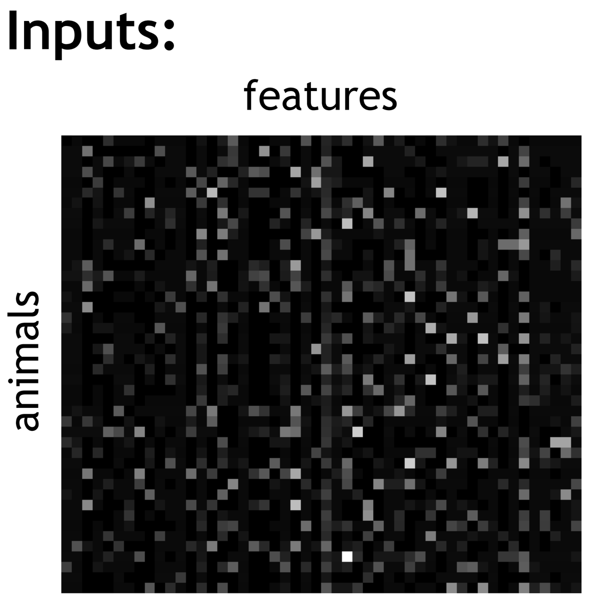
Abstract

Bayesian Analogy with Relational Transformations (BART) is a discriminative model that can learn comparative relations from non-relational inputs (Lu, Chen, & Holyoak, 2012). We extend BART by giving it the ability to generate (rather than classify) relation instances. The extended model generates human-like instantiations of a learned relation (e.g., answering the question, "What is an animal that is smaller than a dog?"). The model can also use its generative capacity to perform hypothetical reasoning, allowing it to make transitive inferences such as, "If *A* is larger than *B* and *B* is larger than *C*, then *A* is larger than *C*."

BART Model of Relation Learning

Domain: Comparative relations between animal concepts

Inputs:



The training data are animal pairs that instantiate each relation. Each animal is represented by a real-valued feature vector, derived from the frequencies with which participants at the University of Leuven generated features characterizing various animals (e.g., "eats fish," "has feathers," etc.; De Deyne et al., 2008).

Each relation (R) is represented as a function, parameterized by a *weight distribution* ($\mathbf{w}$), that takes a set of objects' feature vectors ($\mathbf{x}$) and outputs the probability that these objects instantiate the relation:

$$P(R=1|\mathbf{w},\mathbf{x}) = \left(1+e^{-\mathbf{w}^{\mathrm{T}}\mathbf{x}}\right)^{-1}$$

BART learns a relation by starting with an *empirical prior* for the weight distribution derived from initial learning of one-place predicates (e.g., *large*, *fast*) and updating the weight distribution using Bayes' rule:

$$P(\mathbf{w}|\mathbf{X},\mathbf{R}) \propto P(\mathbf{R}|\mathbf{w},\mathbf{X})P(\mathbf{w})$$

where $\mathbf{X}$ represents the feature vectors for the training pairs and $\mathbf{R}$ is a set of binary indicators denoting whether each pair instantiates the relation.

Like other discriminative models, BART focuses on learning the probabilities of concepts given features and can discriminate between novel pairs that instantiate a relation and those that do not. However, people can also complete a partial instantiation of a relation, answering questions such as, "What is an animal that is smaller than a dog?" How do we give BART this generative ability?

Extension to Generative Inference

Given the features of an object B (dog), $\mathbf{x}_B$, and the knowledge that relation R (*smaller*) holds for the object pair (A, B), the extended model generates a probability distribution for object A's features, $\mathbf{x}_A$:

$$P(\mathbf{x}_A|\mathbf{x}_B,R=1) \propto P(R=1|\mathbf{x}_A,\mathbf{x}_B)P(\mathbf{x}_A|\mathbf{x}_B)$$

The likelihood term reflects the model's desire to satisfy the relation R. It is calculated using BART's learned weights for the first and second roles of the relation ($\mathbf{w}_1$ and $\mathbf{w}_2$, respectively):

$$P(R=1|\mathbf{x}_A,\mathbf{x}_B) = \left(1+e^{-\mathbf{w}_1^{\mathrm{T}}\mathbf{x}_A-\mathbf{w}_2^{\mathrm{T}}\mathbf{x}_B}\right)^{-1}$$

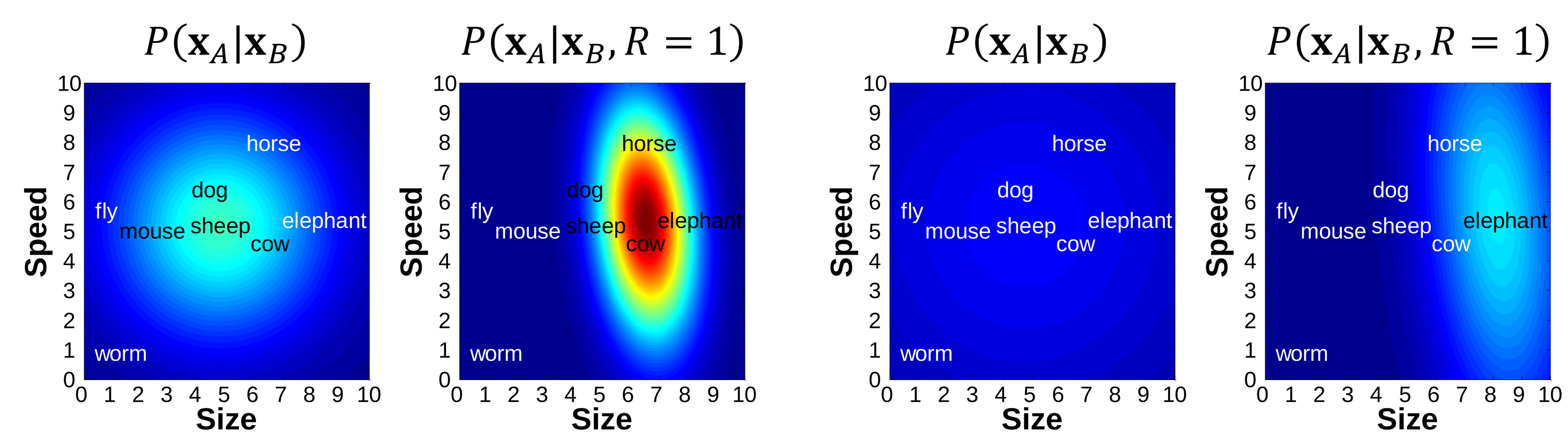
The prior term reflects the model's preference for generating A objects that are similar to B. The strength of that preference is controlled by a free parameter, σ^2:

$$P(\mathbf{x}_A|\mathbf{x}_B) = N\left(\mathbf{x}_B,\sigma^2\mathbf{I}\right)$$

The Variance Parameter

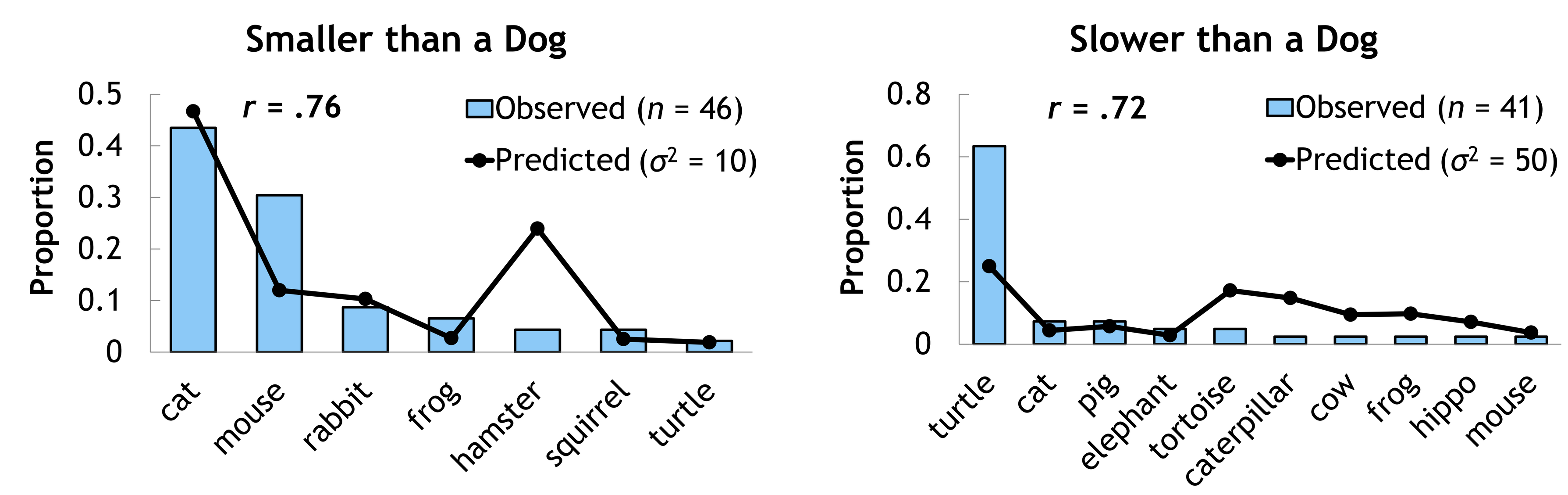
The effect of σ^2 is illustrated by having the model generate an animal that is larger than a sheep using a simpler input:

Favoring Similarity ($\sigma^2 = 7$) — Favoring "Landmarks" ($\sigma^2 = 25$)



Animal Generation

Participants on Amazon Mechanical Turk provided responses to queries of the form, "Name an animal that is smaller than a dog." These observed response proportions were compared to model predictions.

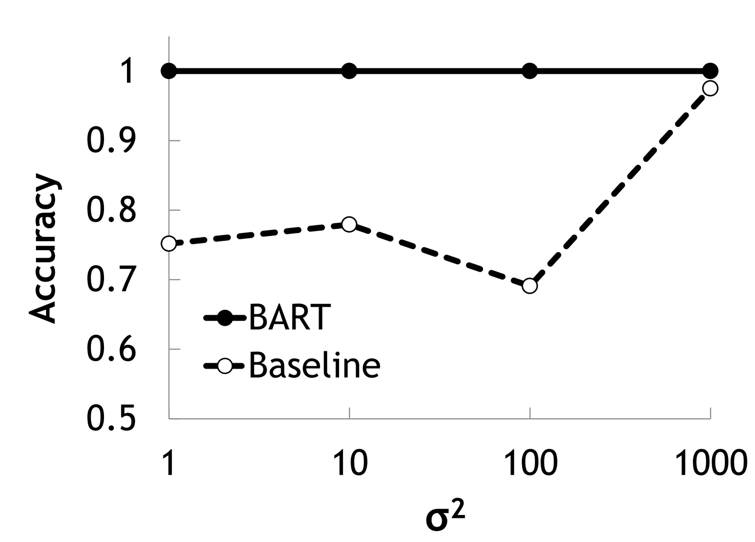


Transitive Inference

Operation of the Model: To infer "if *A* > *B* and *B* > *C*, then *A* > *C*," the model "imagines" many *A*-*B*-*C* triplets that satisfy the premises and then tests whether the conclusion is satisfied, essentially searching for a counterexample. Specifically, it repeats the following steps:

1. Sample object *B* randomly from the "animal distribution"
2. Generate object *A* to satisfy *A* > *B*
3. Generate object *C* to satisfy *B* > *C*
4. Confirm that *A* > *B* and *B* > *C* (and not vice versa)
5. If confirmed, test whether *A* > *C* (and not vice versa)

Results:



BART was compared to a baseline model that uses an uninformative prior instead of BART's empirical prior. The proportion of generated triplets that obey transitivity is shown as a function of σ^2. A lower value of σ^2 (stronger similarity constraint) is more likely to yield counterexamples to transitivity.

BART demonstrates an inductive approximation to deduction by searching for counterexamples and never finding any. The comparative relations BART has learned, albeit imperfect, pass the transitivity test.

Acknowledgements

This research was supported by grant N000140810186 from the Office of Naval Research.

References

De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel,W., Dry, M., Voorspoels, W., & Storms G. (2008). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods*, *40*, 1030-1048.

Lu, H., Chen, D., & Holyoak, K. J. (2012). Bayesian analogy with relational transformations. *Psychological Review*, *119*, 617-648.