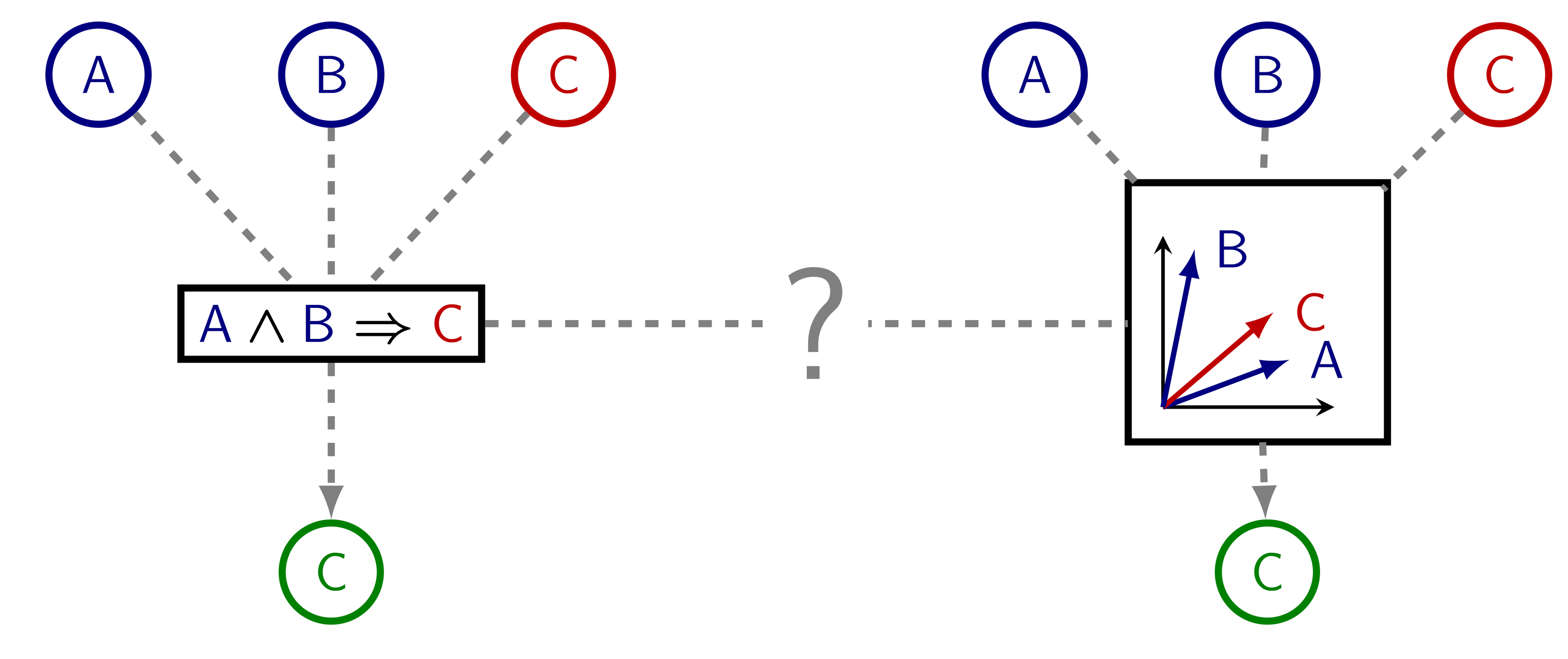
Low-dimensional Embeddings of Logic

Tim Rocktäschel,[1] Matko Bosnjak,[1] Sameer Singh[2] and Sebastian Riedel[1]

[1] University College London, [2] University of Washington

Motivation



- Captures richness
- Supports complex reasoning
- Fails to generalize

- Enables generalization
- Very efficient
- Limited reasoning

Generalized Reasoning

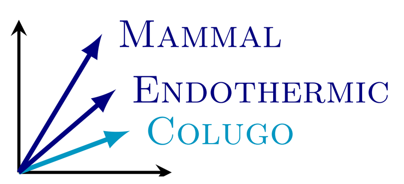
- "Colugos are arboreal gliding mammals that are found in Southeast Asia."

  MAMMAL(COLUGO)
- "All mammals are endothermic."

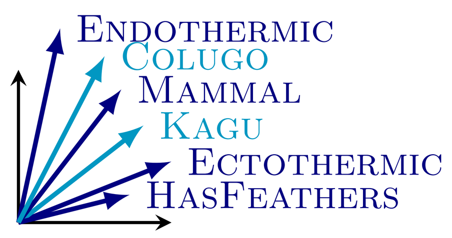
  $\forall x$: MAMMAL(x) $\Rightarrow$ ENDOTHERMIC(x)
- Reasoning...

  ENDOTHERMIC(COLUGO)

I wish I had a distributed model...



Debugging Distributed Representations



- Wrong Predictions

  MAMMAL(KAGU)

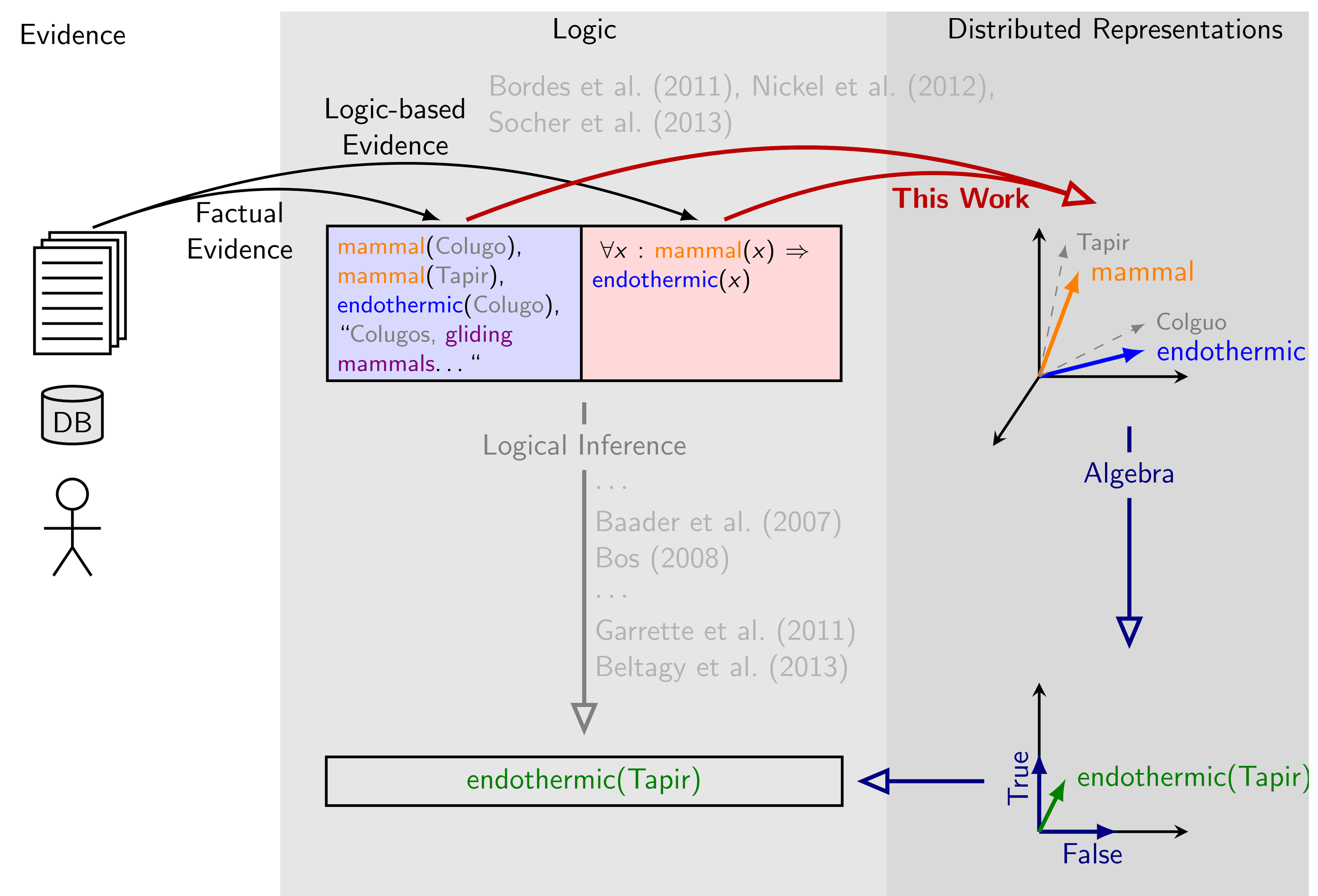
  ECTOTHERMIC(COLUGO)

I wish I could fix this with...

$\forall x$: HASFEATHERS(x) $\Rightarrow \neg$MAMMAL(x)

$\forall x$: ANIMAL(x) $\Rightarrow$ ENDOTHERMIC(x) $\oplus$ ECTOTHERMIC(x)

Overview



Propositional Logic

Logic	Logical Tensor Calculus (Grefenstette, 2013)
[true]; [false]	$\begin{bmatrix}1\\0\end{bmatrix}; \begin{bmatrix}0\\1\end{bmatrix}$
[¬]; [∧]; [⇒]	$\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}; \left[\begin{array}{cc\|cc}1 & 0 & 0 & 0\\0 & 1 & 1 & 1\end{array}\right]; \left[\begin{array}{cc\|cc}1 & 0 & 1 & 1\\0 & 1 & 0 & 0\end{array}\right]$
$[\neg\mathcal{A}]$	$[\neg][\mathcal{A}]$
$[\mathcal{A} \wedge \mathcal{B}]$	$[\wedge] \times_1 [\mathcal{A}] \times_2 [\mathcal{B}]$
$[\mathcal{A} \Rightarrow \mathcal{B}]$	$[\Rightarrow] \times_1 [\mathcal{A}] \times_2 [\mathcal{B}]$
$[\mathcal{A} \wedge \neg\mathcal{B} \Rightarrow \neg\mathcal{C}]$	$[\Rightarrow] \times_1 ([\wedge] \times_1 [\mathcal{A}] \times_2 [\neg][\mathcal{B}]) \times_2 [\neg][\mathcal{C}]$

Example

$$[\Rightarrow] \times_1 [\text{true}] \times_2 [\text{false}] =$$

$$\begin{bmatrix}1 & 0 & 1 & 1\\0 & 1 & 0 & 0\end{bmatrix} \times_1 \begin{bmatrix}1\\0\end{bmatrix} \times_2 \begin{bmatrix}0\\1\end{bmatrix} =$$

$$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} \times_2 \begin{bmatrix}0\\1\end{bmatrix} = \begin{bmatrix}0\\1\end{bmatrix} = [\text{false}]$$

Constants, Predicates, Quantifiers

	One-Hot Representation (Grefenstette, 2013)	Distributed Representation
[COLUGO]	$\begin{bmatrix}0 & 0 & 1 & 0 & \cdots & 0 & 0\end{bmatrix}^\top$	$\begin{bmatrix}? & ? & ?\end{bmatrix}^\top$
[MAMMAL]	$\begin{bmatrix}0 & 0 & 1 & 0 & \cdots & 0 & 1\\1 & 1 & 0 & 1 & \cdots & 1 & 0\end{bmatrix}$	$\begin{bmatrix}? & ? & ?\\? & ? & ?\end{bmatrix}$
[MAMMAL(COLUGO)]	[MAMMAL][COLUGO]	[MAMMAL][COLUGO]
$[\forall x \in X : F(x)]$	$\begin{cases}[\text{true}] \text{ if } \|X\| = \|\{x \mid F(x) = [\text{true}]\}\| \\ [\text{false}] \text{ otherwise}\end{cases}$	$\frac{1}{\|X\|}\sum_x [F(x)]$
$[\exists x \in X : F(x)]$	$\begin{cases}[\text{true}] \text{ if } \|\{x \mid F(x) = [\text{true}]\}\| > 0 \\ [\text{false}] \text{ otherwise}\end{cases}$	$[\neg\forall x \in X : \neg F(x)]$

Objective

$\mathcal{E}$: Set of entity (or entity-pair) vectors

$\mathcal{R}$: Set of relation matrices

$\mathfrak{K}$: Set of logical formulae Q with training signal $\gamma \in \{[\text{true}], [\text{false}]\}$
- In previous work this only contained factual statements
- In addition our objective includes first-order logic formulae!

$\mathcal{L}$: Loss function, e.g., $\|[Q] - \gamma\|_2$

$$\min_{[e]\in\mathcal{E},[r]\in\mathcal{R}} \sum_{(Q,\gamma)\in\mathfrak{K}} \mathcal{L}([Q], \gamma)$$

Toy Example

Before training

	MAMMAL	ENDOTHERMIC	VERTERBRATE
CHIMPANZEE	**1.0**	**1.0**	**1.0**
KOALA	**1.0**	**1.0**	**1.0**
COLUGO	**1.0**	?	?
KAGU	?	**1.0**	?
DODO	?	**1.0**	?

After training

	MAMMAL	ENDOTHERMIC	VERTERBRATE
CHIMPANZEE	**1.0**	**1.0**	**1.0**
KOALA	**1.0**	**1.0**	**1.0**
COLUGO	**1.0**	0.1	0.5
KAGU	0.1	**0.9**	0.3
DODO	0.1	**1.0**	0.3

After training with formulae

$\forall x \in X : [\Rightarrow] \times_1 ([\text{MAMMAL}][x]) \times_2 ([\text{ENDOTHERMIC}][x])$

	MAMMAL	ENDOTHERMIC	VERTERBRATE
CHIMPANZEE	**1.0**	**0.9**	**1.0**
KOALA	**1.0**	**0.9**	**1.0**
COLUGO	**0.9**	(0.1) 0.6	(0.5) 0.8
KAGU	0.1	**1.0**	0.2
DODO	0.0	**1.0**	0.1

Future Work

? What are the theoretical limits of embedding logical formulae in vector spaces?

? What are efficient ways of injecting quantified formulae without iterating over all elements of a domain?

? Can we provide provenance of proofs of answers?

References

Franz Baader, Bernhard Ganter, Baris Sertkaya, and Ulrike Sattler. 2007. Completing description logic knowledge bases using formal concept analysis. In IJCAI, pages 230–235.

Islam Beltagy, Cuong Chau, Gemma Boleda, Dan Garrette, Katrin Erk, and Raymond Mooney. 2013. Montague meets markov: Deep semantics with probabilistic logical form. In 2nd Joint Conference on Lexical and Computational Semantics: Proceeding of the Main Conference and the Shared Task, Atlanta, pages 11–21.

Antoine Bordes, Jason Weston, Ronan Collobert, and Yoshua Bengio. 2011. Learning structured embeddings of knowledge bases. In AAAI.

Johan Bos. 2008. Wide-coverage semantic analysis with boxer. In Johan Bos and Rodolfo Delmonte, editors, Semantics in Text Processing. STEP 2008 Conference Proceedings, Research in Computational Semantics, pages 277–286. College Publications.

Dan Garrette, Katrin Erk, and Raymond Mooney. 2011. Integrating logical representations with probabilistic information using markov logic. In Proceedings of the Ninth International Conference on Computational Semantics, pages 105–114. Association for Computational Linguistics.

Edward Grefenstette. 2013. Towards a formal distributional semantics: Simulating logical calculi with tensors. In Proceedings of the Second Joint Conference on Lexical and Computational Semantics, pages 1–10.

Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2012. Factorizing yago: scalable machine learning for linked data. In Proc. of WWW, pages 271–280.

Sebastian Riedel, Limin Yao, Andrew McCallum, and Benjamin M Marlin. 2013. Relation extraction with matrix factorization and universal schemas. In Proceedings of NAACL-HLT, pages 74–84.

Richard Socher, Danqi Chen, Christopher D. Manning, and Andrew Y. Ng. 2013. Reasoning with neural tensor networks for knowledge base completion. In NIPS, pages 926–934.