Extracting Commonsense Properties from Embeddings with Limited Human Guidance

Property Comparison from Embeddings (PCE model)

Yiben Yang, Larry Birnbaum, Ji-Ping Wang and Doug Downey

July 18, 2018

Northwestern University

Table of contents

Motivation

Commonsense Property Comparison Task

Is an **elephant** bigger or smaller than a **mouse**?
Is **Ferrari** more expensive or cheaper than **beer**?

Problem Definition

Three-way task:

$$P(\mathbf{L}|O_1, O_2, \mathit{Property}), \mathbf{L} \in \{\boxed{<}, \boxed{>}, \boxed{\approx}\}.$$

Four-way task:

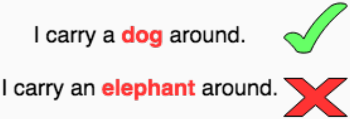
$$P(\mathbf{L}|O_1, O_2, \mathit{Property}), \mathbf{L} \in \{\boxed{<}, \boxed{>}, \boxed{\approx}, \boxed{\text{N/A}}\}.$$

Learning Commonsense Knowledge from Text?

Challenges:

- **Reporting bias** [Gordon and Van Durme 2013]: Commonsense knowledge is rarely **explicitly** stated.
- Large knowledge dimensions: Property specified by adjectives: large, heavy, fast, rigid, etc. Creating training examples and building separate models on each type of property requires **expensive** labeling efforts. Handling unseen properties during the test phase (**zero-shot** prediction)?
- Language variation: An ideal model should be able to take flexible natural language inputs.

Can we build an efficient commonsense comparison model with **word embedding** inputs only ?



Method

Categorical Linear Regressions

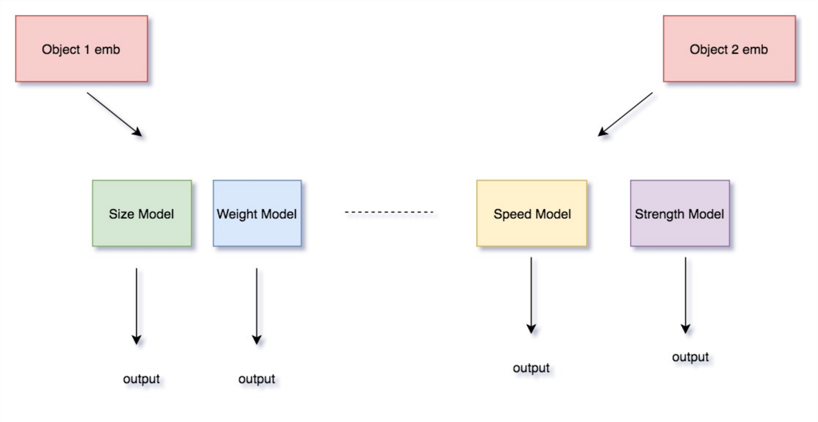
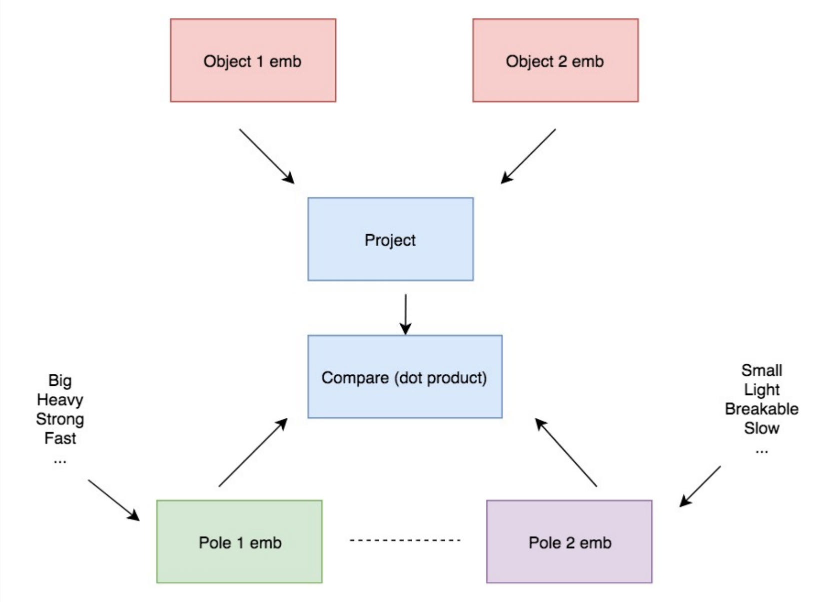


Figure 1: Creating a softmax regression model for each property.

Our PCE model



Experiment

Data

- VERB PHYSICS (5 physical properties) [Forbes and Choi 2017]
- PROPERTY COMMON SENSE (32 commonsense properties)

Results: Supervised Performance

Model	Test					
	size	weight	stren	rigid	speed	overall
Majority	0.51	0.55	0.52	0.49	0.50	0.51
F&C	0.75	0.76	0.72	0.65	0.61	0.70
PCE(LSTM)	**0.80**	**0.79**	**0.76**	**0.71**	**0.71**	**0.76**
PCE(GloVe)	0.76	0.75	0.71	0.68	0.68	0.72
PCE(Word2vec)	0.76	0.76	0.73	0.68	0.66	0.72

Table 1: Supervised accuracy on the VERB PHYSICS data set. PCE outperforms the F&C model from previous work.

Results: Zero-shot Prediction

Model	Test				
	size	weight	stren	rigid	speed
Random	0.33	0.33	0.33	0.33	0.33
Emb-Similarity	0.37	0.53	0.48	0.43	0.35
PCE	**0.74**	**0.73**	**0.70**	**0.62**	**0.58**

Table 2: Accuracy of zero-shot learning on the VERB PHYSICS data set(using LSTM embeddings).

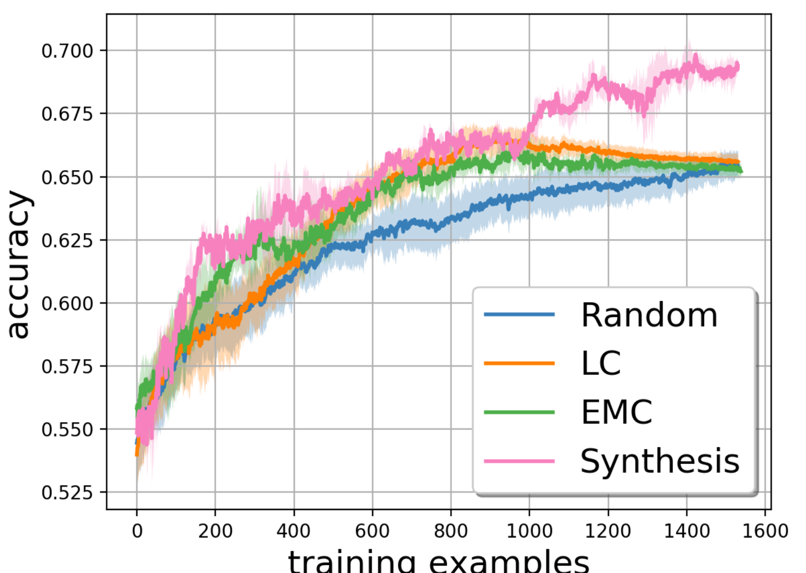
Results

Model	Test
Random	0.25
Majority Class	0.51
PCE(GloVe)	0.63
PCE(Word2vec)	**0.67**
PCE(LSTM)	**0.67**

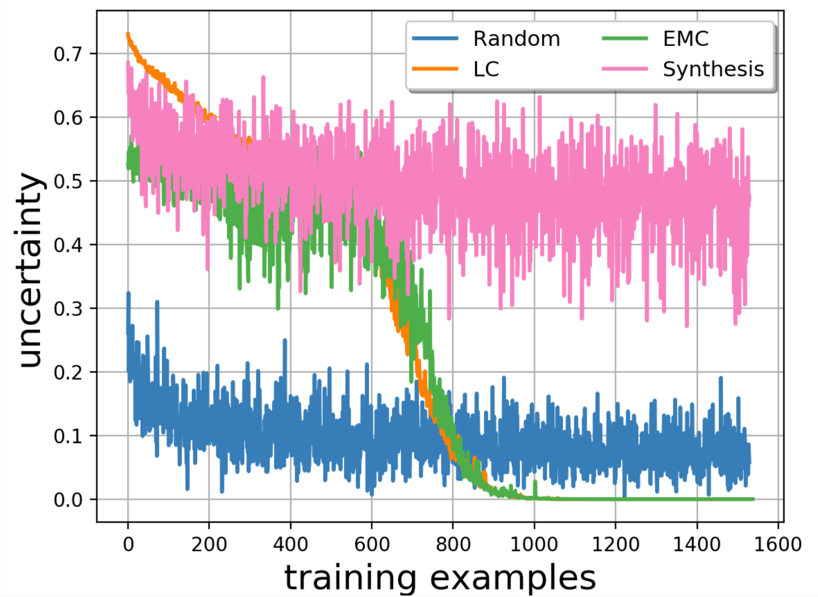
Table 3: Accuracy on the four-way task on the PROPERTY COMMON SENSE data.

Synthesis Active Learning

Want further reduce labeling effort?



Active Learning



Demo

Demo

`http://thor.cs.northwestern.edu:1959/`