



Minimally Supervised Classification to Semantic Categories using Automatically Acquired Symmetric Patterns



Roy Schwartz+, Roi Reichart* and Ari Rappoport+

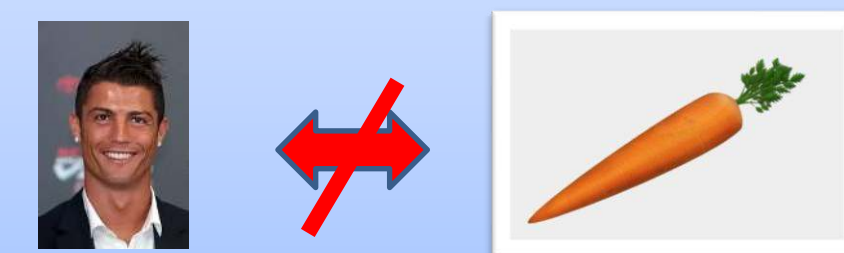
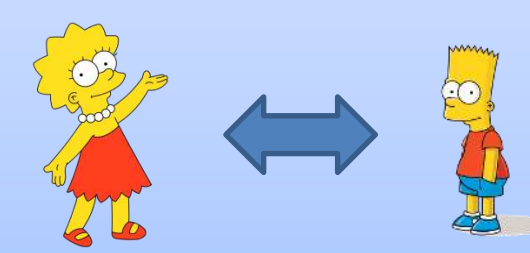
+The Hebrew University, {roys02,arir}@cs.huji.ac.il

*Technion IIT, roiri@ie.technion.ac.il

ISCOL 2014, based on a work presented at *Coling 2014*

THE GOAL

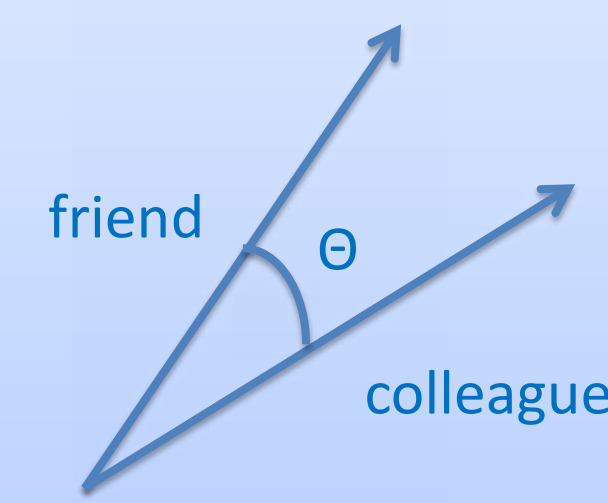
Represent Word Similarity



PREVIOUS APPROACHES

DISTRIBUTIONAL SEMANTICS HYPOTHESIS (HARRIS, 1954)

... tokens to date, friend lists and recent ...
... by my dear friend and companion, Fritz von ...
... even have a friend who never fails ...
... by my worthy friend Doctor Haygarth of ...
... and as a friend pointed out to ...
... partner, in-laws, relatives or friends speak a different ...
... petition to a friend Go to the ...
... otherwise, to a friend or family member ...
... images from my friend Rory though - ...
... great, and a friend as well as a colleague, who, ...
...

$$\begin{bmatrix} 0 \\ 0.5 \\ 0.76 \\ -0.12 \\ 0.76 \\ 0 \\ 0 \\ -0.51 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix}$$


OUR APPROACH

SYMMETRIC PATTERN CONTEXTS

friend and companion
companion and friend

from X to Y

X and Y

relatives or friends
friends or relatives

X or Y

X as well as Y

friend as well as a colleague
colleague as well as a friend

neither X nor Y

SYMMETRIC PATTERNS TO WORD SIMILARITY

S_{XY} → the number of times X,Y appeared in the same symmetric pattern

• orange ↔ apple

1. ... apples and oranges ...
2. ... oranges as well as apples ...
- ...
- K. ... neither apple nor orange ...

$$\rightarrow \text{orange} \leftrightarrow \text{apple} = \frac{K}{Z}$$

– Z: a normalization factor

• France ↔ England

1. ... England or France ...
2. ... from France to England ...
- ...
- M. ... England and France ...

$$\rightarrow \text{France} \leftrightarrow \text{England} = \frac{M}{Z}$$

AUTOMATICALLY ACQUIRED SYMMETRIC PATTERNS

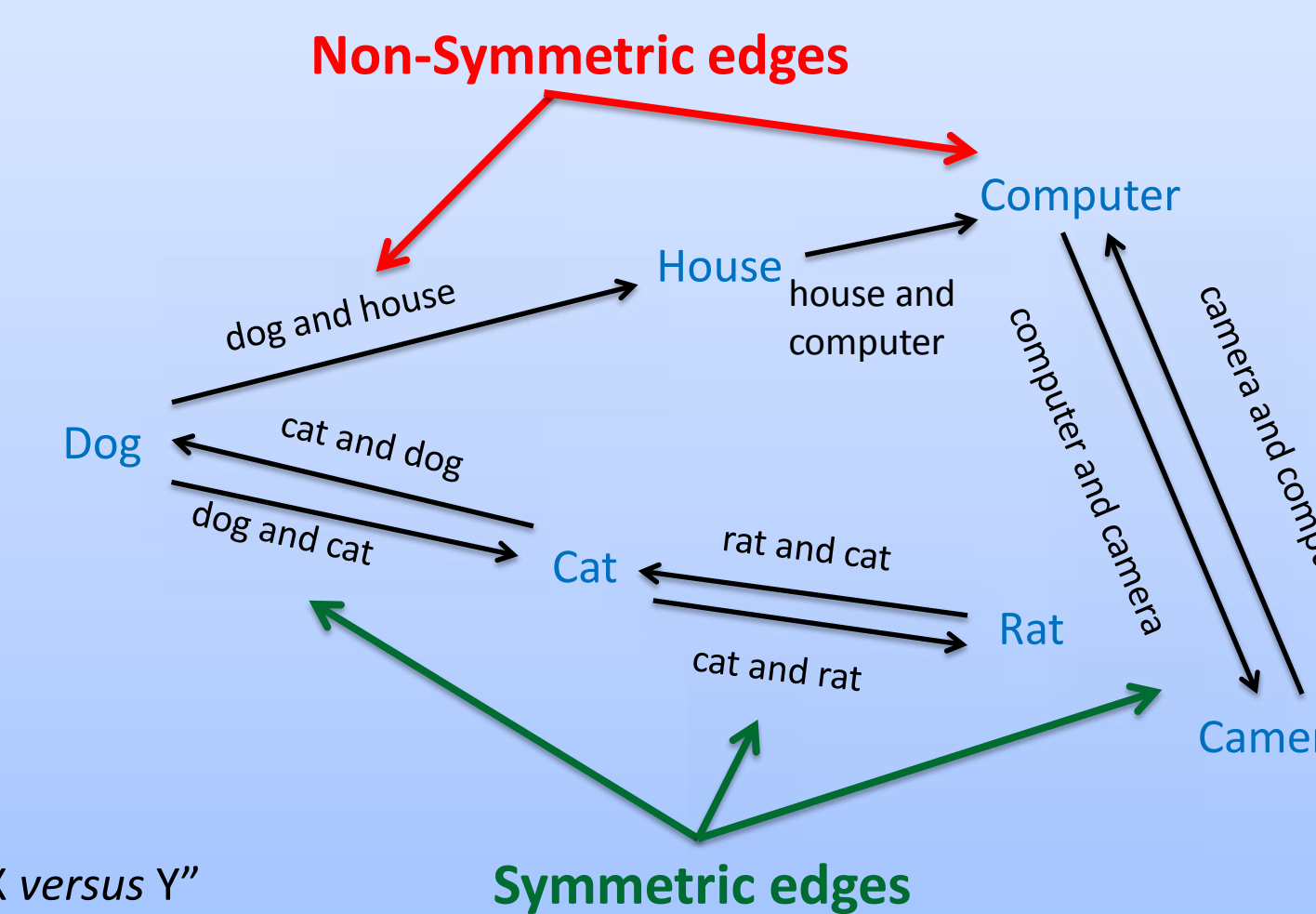
(DAVIDOV AND RAPPOPORT, 2006) ALGORITHM

• Extract List of Patterns in a Fully Unsupervised Manner

$$M = \frac{\# \text{symmetric_edges}}{\# \text{all_edges}}$$

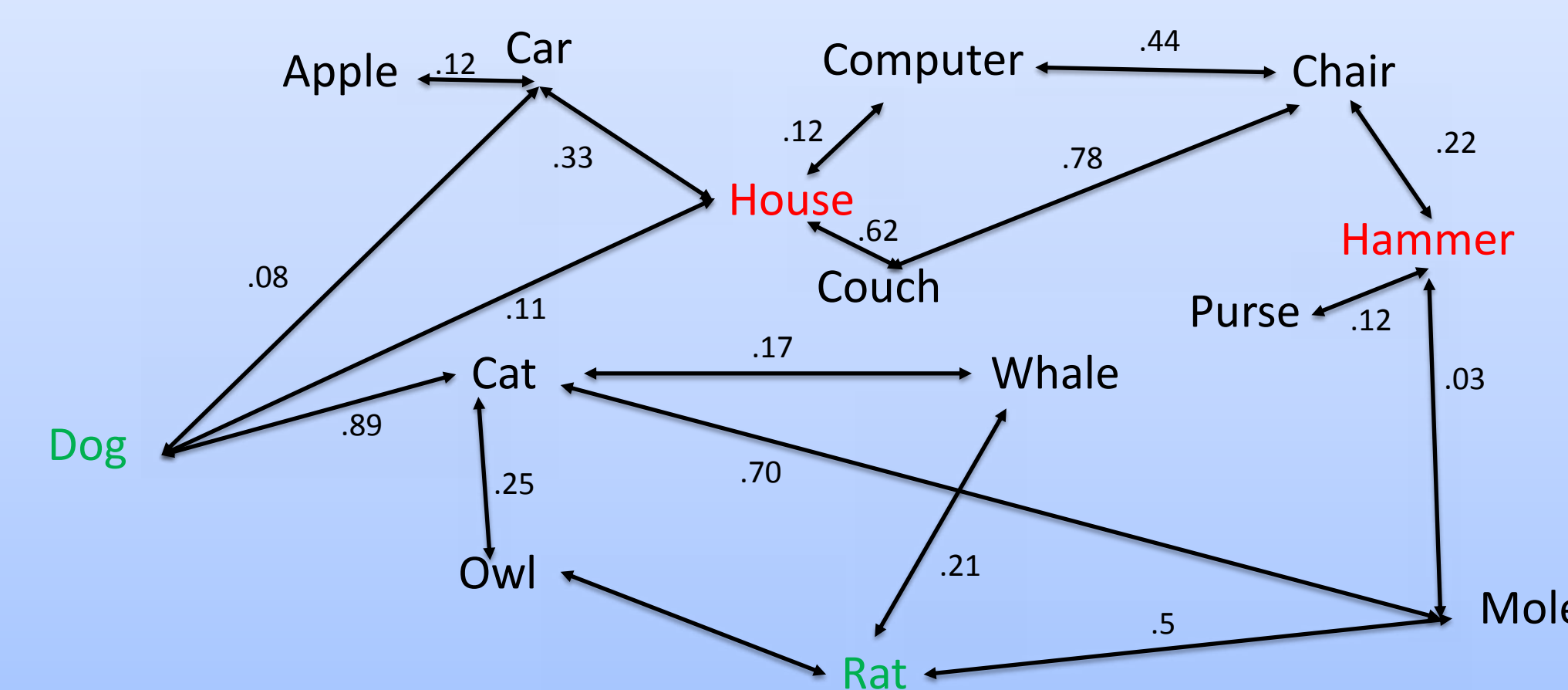
• 20 Patterns Extracted

- "X and Y", "X or Y"
- "X and the Y", "X rather than Y", "X versus Y"



MINIMALLY SUPERVISED NOUN CLASSIFICATION

EXAMPLE: ANIMATE CLASS



ALGORITHM

ITERATIVE K-NEAREST NEIGHBORS (I-K-NN)

- Run iterations of the k-NN algorithm
 - Start with a small number of labeled nodes
 - At each iteration propagate information to additional vertices by selecting words for which many of their neighbors have the same label
 - Halt when all nodes are assigned with a label

BASELINES

- Word Similarity Measures Baselines
 - SENNA word embeddings (Collobert et al., 2011)
 - Brown Clusters (Brown et al., 1992)
- Label Propagation Baselines
 - Normalized graph cut algorithm (Yu and Shi, 2003)
 - Modified Adsorption (MAD) algorithm (Talukdar and Crammer, 2009)

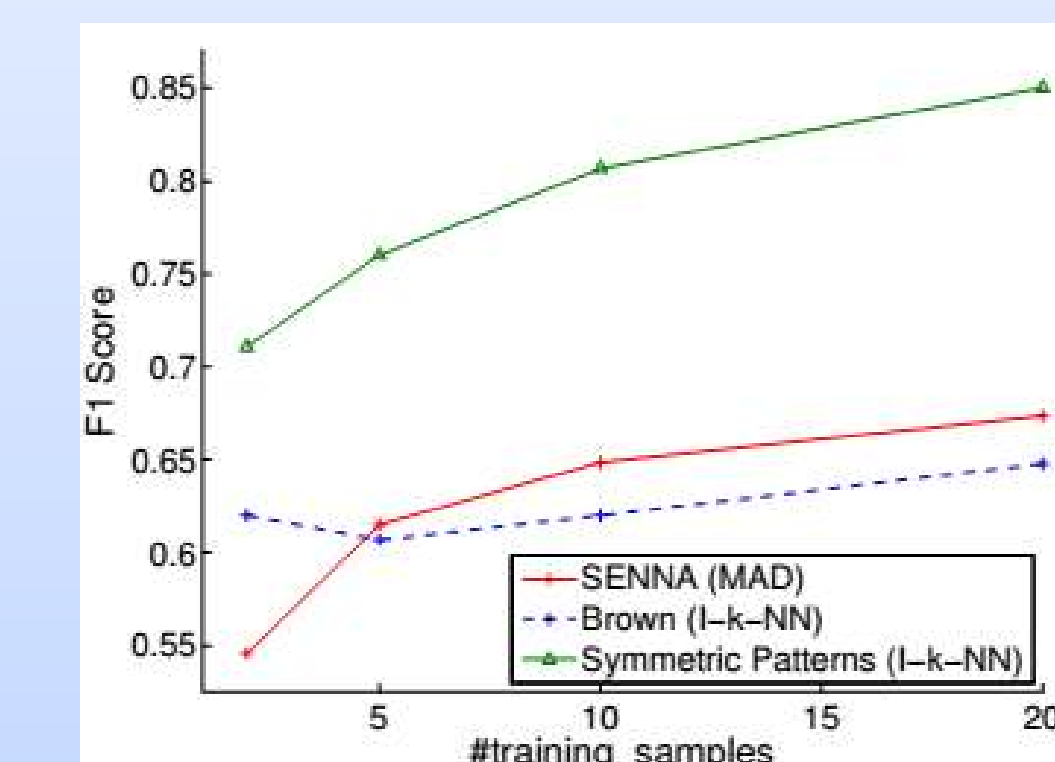
EXPERIMENTS

- A subset of the CSLB property norms dataset (Devereux et al., 2013)
 - 450 concrete nouns
 - Thirty human annotators assigned each noun with semantic categories



- Symmetric pattern similarity scores computed using the google books n-gram corpus
- Number of labeled seed words
 - 4, 10, 20, 40

RESULTS



- When using as few as **four labeled seed words**
 - Accuracy results are **82-94%**
 - F1 scores are **0.64-0.86**
- **Symmetric patterns** >> other word similarity measures across
 - semantic categories
 - label propagation algorithms
 - labeled seed set sizes
 - evaluation measures

SYMMETRIC PATTERNS

- Interpretable
- Efficient to compute
 - A count model, no vector or matrix computation
- Capture a different signal than bag-of-words or word n-gram models

FUTURE WORK

- Integrating symmetric pattern information into deep network models
 - Enhancing bag-of-words models with symmetric patterns information
 - Integrate word embeddings with symmetric patterns-based vectors