# 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

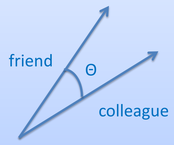
Represent Word Similarity

## 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, ...
...

$$\begin{pmatrix} 0 \\ 0.5 \\ 0.76 \\ -0.12 \\ 0.76 \\ 0 \\ 0 \\ -0.51 \\ . \\ . \\ . \end{pmatrix}$$




## Our Approach

### Symmetric Pattern Contexts

**friend** and **companion**
**companion** and **friend**

**relatives** or **friends**
**friends** or **relatives**

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

from **X** to **Y**

**X** and **Y**

**X** or **Y**

**X** as well as **Y**

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 ...

  → orange ←→ apple = $\frac{K}{Z}$

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

  → France ←→ England = $\frac{M}{Z}$

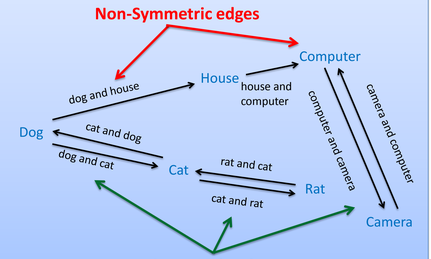
– Z: a normalization factor

## *Automatically Acquired* Symmetric Patterns

### (Davidov and Rappoport, 2006) Algorithm

- **Extract List of Patterns in a Fully Unsupervised Manner**

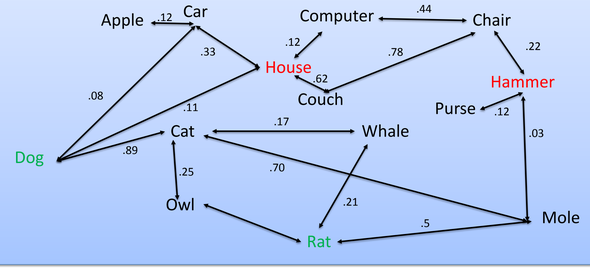
M = #symmetric_edges / #all_edges




- 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)

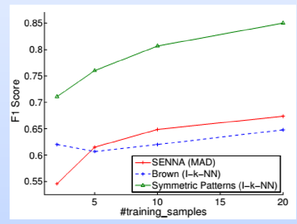
## Experiments

- 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