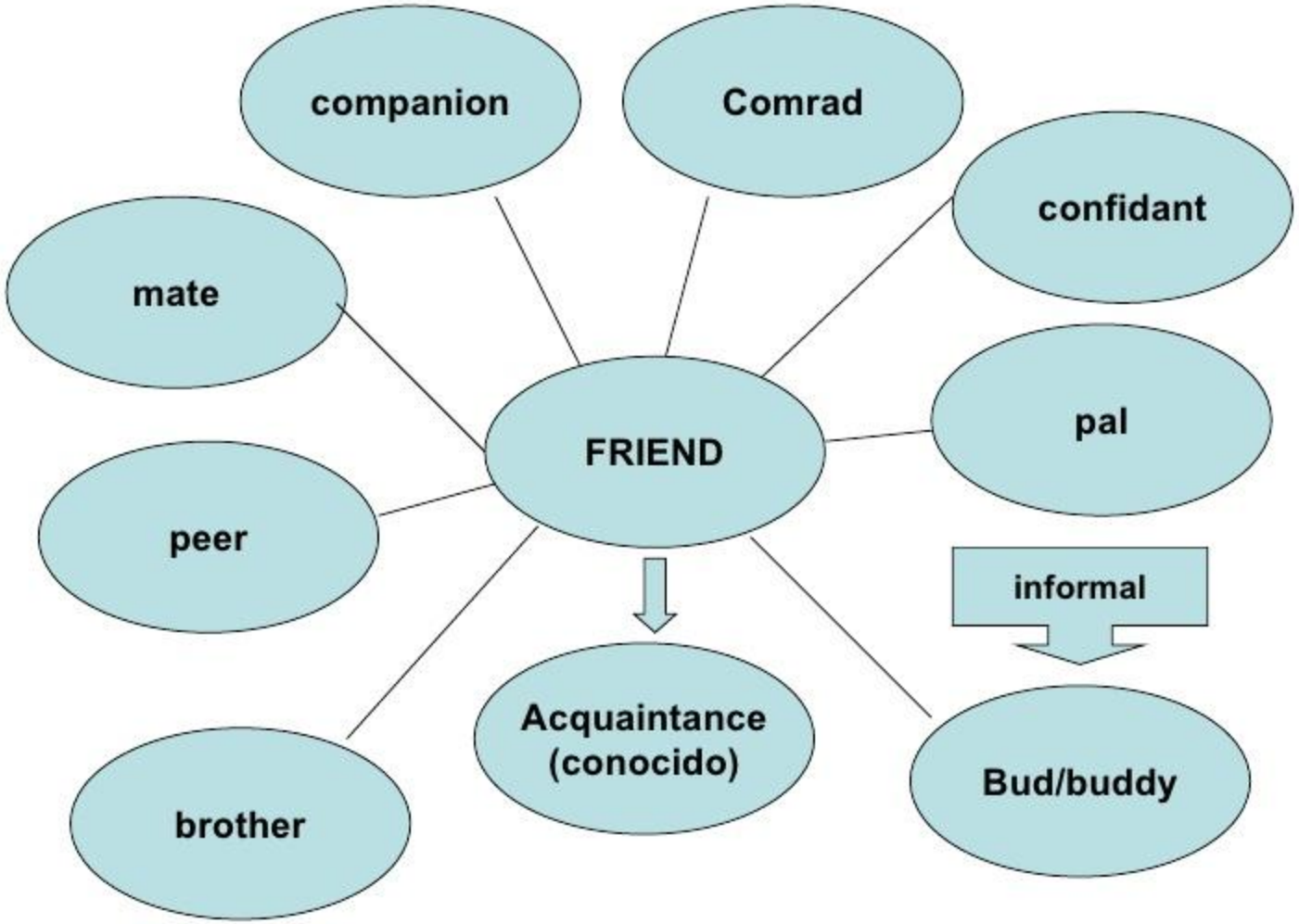
Minimally Supervised Classification to Semantic Categories using Automatically Acquired Symmetric Patterns

Roy Schwartz[+], Roi Reichart [*] and Ari Rappoport[+]

[+]The Hebrew University, [*]Technion IIT

COLING 2014



http://www.slideshare.net/halucinex/friend-word-map

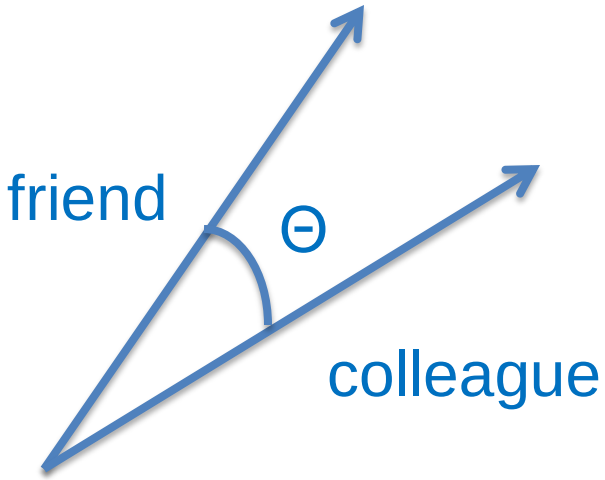
DS Hypothesis (Harris, 1954)

- ... 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, ...
- …

Examples taken from the ukwac corpus (Baroni et al., 2009)

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

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



- ... 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, ...
- …

- ... 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, ...
- …

- ... 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, ...
- …

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

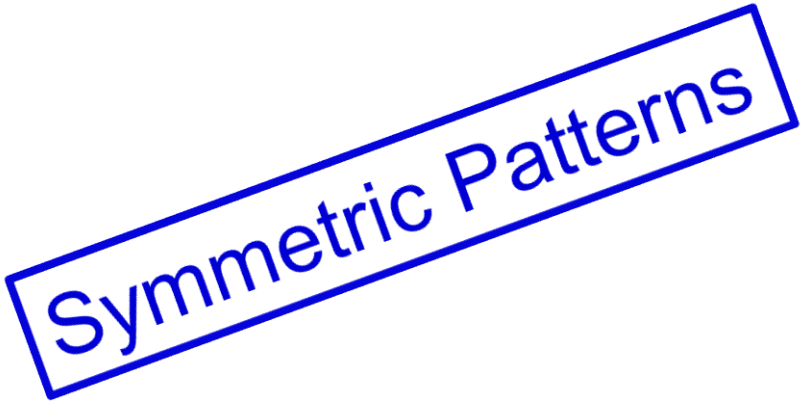
- relatives or **friends**
- **friends** or relatives

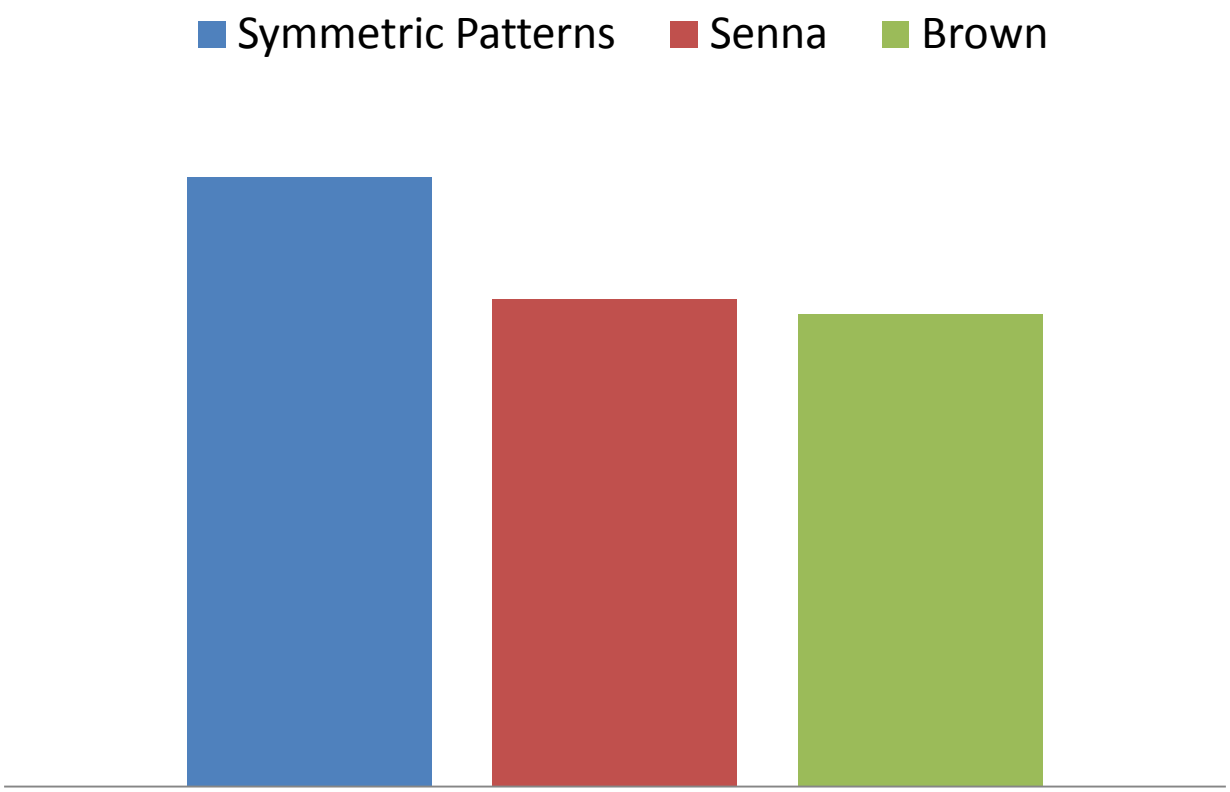
- **friend** as well as a colleague
- colleague as well as a **friend**

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

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

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




Symmetric Patterns
Senna
Brown

Overview

- **The task**
  - *Minimally supervised semantic classification*
- **The method**
  - *Automatically acquired **symmetric patterns***
- **Results**
  - *Symmetric patterns outperform **strong** baselines by > 12% accuracy*

The Task

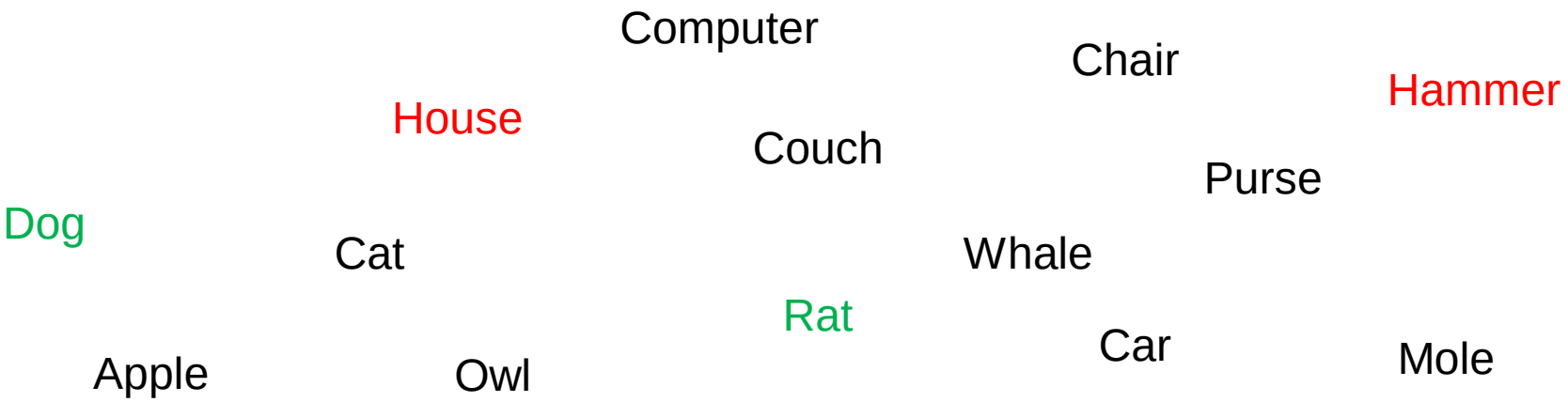
- Binary Classification of Nouns into Semantic Categories
  - *Is "dog" an animal?*
  - *Is "couch" a tool?*

- *Use **minimal supervision***

The Task

Example

- Animals



The Task

Goal

- Animals

Computer
Chair
Hammer
House
Couch
Purse
Dog
Cat
Whale
Rat
Car
Mole
Apple
Owl

Symmetric Patterns Contexts

from X to Y

X and Y

X or Y

neither X nor Y

X as well as Y

Symmetric Patterns to Word Similarity

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

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 = $\frac{K}{Z}$

– Z: a normalization factor

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

Symmetric Patterns

from X to Y

X and Y

X or Y

neither X nor Y

X as well as Y

Symmetric Patterns

from X to Y

X and Y

X or Y

neither X nor Y

X as well as Y

Automatically Extracted Symmetric Patterns

The (Davidov and Rappoport, 2006) Algorithm

- A graph-based algorithm
  - Input: a corpus of **plain text**
  - Output: a set of symmetric patterns

Automatically Extracted Symmetric Patterns

The (Davidov and Rappoport, 2006) Algorithm

- A graph-based algorithm
  - Input: a corpus of **plain text**
  - Output: a set of symmetric patterns

- The idea: search for patterns with **interchangeable** word pairs
  - For each pattern candidate, compute symmetry measure (M)
  - Select the patterns with the highest M values

Automatically Extracted Symmetric Patterns

The (Davidov and Rappoport, 2006) algorithm

- The M measure counts the proportion of pattern instances that appear in both directions ("cat and dog" + "dog and cat")
  - See paper for more details

- High M value ➔ A symmetric pattern

Automatically Extracted Symmetric Patterns

The (Davidov and Rappoport, 2006) algorithm

- The M measure counts the proportion of pattern instances that appear in both directions ("cat and dog" + "dog and cat")
  - See paper for more details
- High M value ➔ A symmetric pattern
- Twenty symmetric patterns are extracted
  - "X and Y", "X or Y"
  - "X *and the* Y", "X *rather than* Y", "X *versus* Y"

Word Similarity Measures

S_{XY} → Similarity Between Words X and Y

- Symmetric patterns
  - Extract a set of symmetric patterns from plain text
  - S_{XY} → the number of time X and Y participate in the same symmetric pattern

Word Similarity Measures

S_{XY} → Similarity Between Words X and Y

- Symmetric patterns
  - Extract a set of symmetric patterns from plain text
  - S_{XY} → the number of time X and Y participate in the same symmetric pattern

- Baselines:
  - **Senna** word embeddings (Collobert et al., 2011):
  - S_{XY} → cosine similarity between the word embeddings of X and Y

Word Similarity Measures

S_{XY} → Similarity Between Words X and Y

- Symmetric patterns
  - Extract a set of symmetric patterns from plain text
  - S_{XY} → the number of time X and Y participate in the same symmetric pattern

- Baselines:
  - **Senna** word embeddings (Collobert et al., 2011):
  - S_{XY} → cosine similarity between the word embeddings of X and Y

  - **Brown** Clusters (Brown et al., 1992):
  - S_{XY} → 1 - tree distance between X and Y clusters

Word Classification

- Reminder: Our Task
  - **Minimally-supervised** semantic word classification

Word Classification

- Reminder: Our Task
  - **Minimally-supervised** semantic word classification

- An undirected weighted graph
  - Nodes are words
  - Edge weights are word similarity scores

Word Classification

- Reminder: Our Task
  - **Minimally-supervised** semantic word classification
- An undirected weighted graph
  - Nodes are words
  - Edge weights are word similarity scores
- **Goal**: to propagate labels from a few labeled seed words to the rest of the graph

Label Propagation Algorithms

- Iterative variant of k-Nearest Neighbors
  - See paper for details

- Baselines
  - Normalized graph cut algorithm (Yu and Shi, 2003)
  - Modified Adsorption (MAD) algorithm (Talukdar and Crammer, 2009)

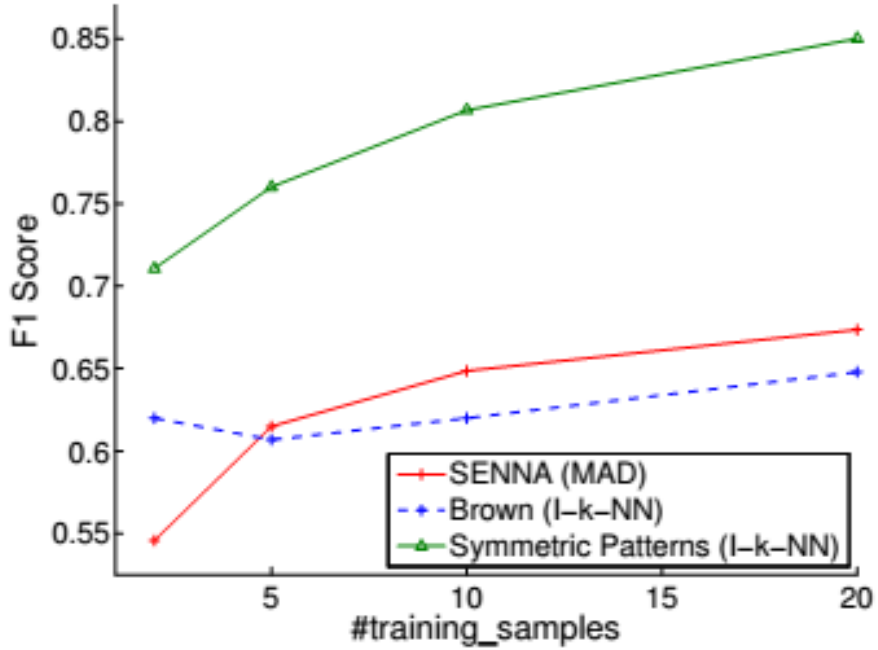
Experimental Setup

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
  - ***animals, tools, food, clothes***
- Symmetric pattern based scores computed using the google books n-gram corpus
- Number of labeled seed words
  - 4, 10, 20, 40

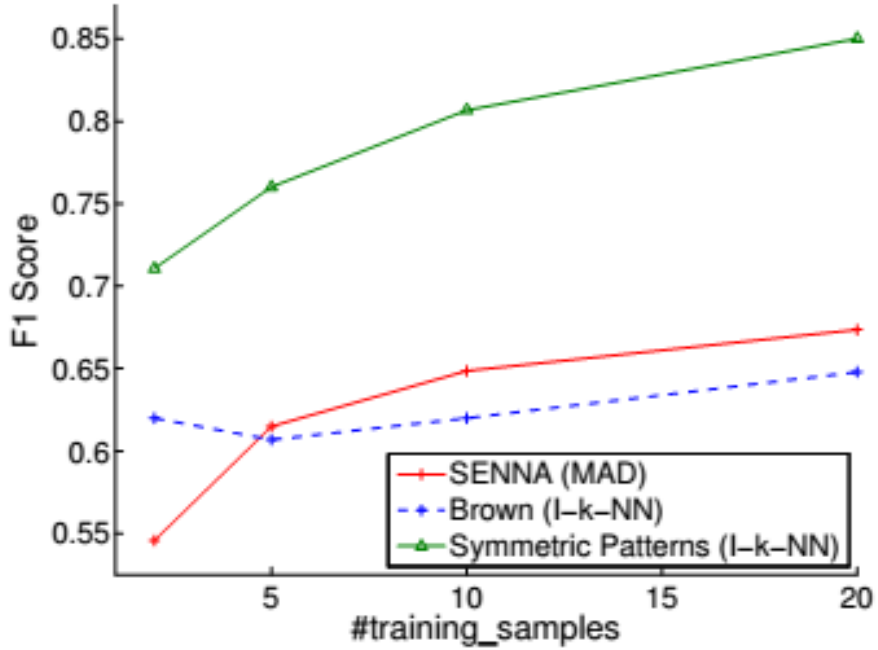
Results

Word Similarity Measures



Results

Word Similarity Measures



best **symmetric patterns** model >> any other model

12.5% accuracy, 0.13 F1 points difference

More Results

- When using as few as four labeled seed words
  - Accuracy results are 82-94%
  - F1 scores are 0.64-0.86

- Symmetric patterns are superior compared to the 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
- Captures 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

Summary

- Minimally supervised semantic classification
- Symmetric patterns form an effective manner to compute word similarity
  - And can also be extracted from plain text!
- 82%-94% accuracy using only **four** labeled examples per category
  - 12.5% accuracy improvement over strong baselines

roys02@cs.huji.ac.il

http://www.cs.huji.ac.il/~roys02/