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

AUTOMATICALLY ACQUIRED SYMMETRIC PATTERNS
(DAVIDOV AND RAPPORPORT, 2006) ALGORITHM

PREVIOUS APPROACHES

DISTRIBUTIONAL SEMANTICS HYPOTHESIS (HARRIS, 1954)

- Takes to data, friend lists and more...
- ... to me dear friend and companion, this is...
- ... even have a friend who ever that friend...
- ... and as a friend pointed out to...
- ... patterns in every relations or friends speak a different...
- ... need to be friends to the ...
... otherwise, to a friend or family member...
... images from my friend Roy though...
... great and a friend as well as a colleague...
...”

0
0.5
0.75
1.0
1.0 - 0.51

friend

D

colleague

0
0.5
0.75
1.0
1.0 - 0.51

friend

D

relatives

Y

companion

Y

friend and companion

companion and friend

relations or friends

friends or relatives

friend as well as a colleague

colleague as well as a friend

OUR APPROACH
SYMMETRIC PATTERN CONTEXTS

SYMMETRIC PATTERNS TO WORD SIMILARITY

*SXY the number of times X,Y appeared in the same symmetric pattern

- orange <-> apple
  1. ... apples and oranges...
  2. ... oranges as well as apples...
  3. ... together as well as or...
- France <-> England
  1. ... England or France...
  2. ... from France to England...
  3. ... England and France...
- France <-> England
  1. ... England or France...
  2. ... from France to England...
  3. ... England and France...

M = #symmetric_edges

W = symmetric_edges

20 Patterns Extracted

- “4” and “Y”, “5” or “Y”
- “4” and the “Y”, “c rather than Y”, “Y versus Y”

MINIMALLY SUPERVISED NOUN CLASSIFICATION EXAMPLE: ANIMATE CLASS

ALGORITHM
ITERATIVE k-NEARNEST NEIGHBORS (k-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

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

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

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

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 Adhesion (MAD) algorithm (Tulukdar and Cramer, 2009)