Co-occurrence Cluster Features for Lexical Substitutions in Context

Chris Biemann
Microsoft / Powerset
cbiemann@microsoft.com

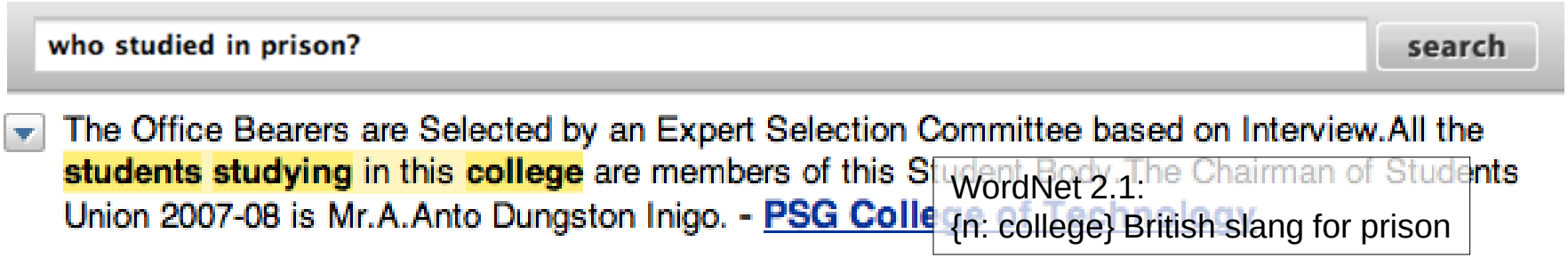
ACL Workshop on TextGraphs,
Uppsala, Sweden, July 15 2010

Outline

- Task: Lexical Substitution for Semantic Indexing
- Supervised Word Sense Disambiguation
- Clustering of Co-occurrences for Word Sense Induction
- Using Co-occurrence Clusters as Features
- Two Experiments:
  - SemEval 07 lexical sample WSD
  - TWSI Substitution Quality
- Conclusion

Lexical Substitution for Semantic Indexing

- Traditional IR: query words disambiguate each other
- Semantic IR: lexical expansion in absence of disambiguation leads to spurious matches



- Task: supply substitutions in context for indexing
- Setup: disambiguate word sense and assign substitutions accordingly

Baseline supervised WSD System

- Supervised WSD: Per target word, learn a model that assigns one of the possible senses based on features on contexts
- Baseline: 15 features (relative to target)
  - (2) word forms left and right
  - (2) POS sequences left and right bigram
  - (2) POS tags of left and right word
  - (1) POS tag of target
  - (4) two left and two right nouns
  - (2) left and right verbs
  - (2) left and right adjectives
- Classifier: Weka's AODE (handles dependent nominal features well)

Why co-occurrence cluster features ?

- Successful WSD systems model topicality via topic signatures (Martinez et al., 2008), semantic kernels and SVD (Gliozzo et al. 2006) etc.
- Co-occurrence cluster features: simple alternative: does not need predefined word sense inventory
- Approach is similar to word sense induction (e.g. (Veronis, 2004). Difference: WSI is normally used to greedily map induced senses to target senses. Here: Use output of several WSI systems as a feature

Obtaining Cooc Clusters

- Significant sentence-base co-occurrences for text corpus (log likelihood, threshold 6.63)
- Per target, cluster the open neighborhood graph (Widdows and Dorow, 2002) with Chinese Whispers (Biemann, 2006)
- Graph parameters:
  - t={50,100,200}: include most significant t neighbors for target
  - n={50,100,150,200,250}: draw edge between nodes if one is contained in the n top sig. co-occurrences of the other
- Clustering Parameter: down-weighting node influence according to degree d:
  - a) no down-weighting
  - b) weight=1/log(d+1)
  - c) weight=1/d

Cooc cluster examples “bank”

Clustering for t=50, n=200, weighting (a)

- bank0: largest, north, branches, eastern, opposite, km, east, west, branch, Thames, banks, located, Danube, town, south, situated, River, Rhine, river, western, commercial, central, southern
- bank1: right, left
- bank2: money, robbers, deposit, robberies, cash, currency, account, deposits, Bank, robbery, funds, financial, banking, loans, notes, robber, rob, accounts, credit, assets, teller, Banco, loan, investment, savings

Clustering for t=50, n=100, weighting (c)

- bank0: eastern, banks, central, river, km, western, south, southern, located, largest, east, deposits, commercial, Thames, north, west, Danube, town, situated, Rhine, River
- bank1: branches, branch
- bank2: robberies, robbers, robbery, robber
- bank3: right, left, opposite
- bank4: loans, cash, investment, teller, account, financial,loan, deposit, credit, funds, accounts, assets, savings, banking, money, rob
- bank5: Banco, currency, notes, Bank

Adding Cooc Clusters as features

- Feature=cooc cluster (1 parameterization)
- Feature Value: cluster ID with highest context overlap
- Adding these features to the baseline model

Testing feature combinations:

1. Add one at the time, rank by contribution
2. Take k top-ranked and add them together

Experiment 1: SemEval 2007 task 17

- Sense annotation: Semcor coarse-grained
- Corpus for clustering: New York Times
- Cross-validation on training (Precision in %):
  - Baseline: 87.1%
  - Single cooc features: 88.0%-88.3%
  - Best combination k=3: 88.5% (used)
- Test:

System	F1
NUS-ML	88.7% ± 1.2
top3 cluster, max recall	87.8% ± 1.2
baseline, max recall	87.3% ± 1.2
UBC-ALM	86.9% ± 1.2

TWSI example

- [189 sentences] magazine@@1

  Their first album was released by Columbia Records in 1972 , and they were voted " Best New Band " by Creem **magazine**.

  publication [42], periodical [32], journal [30], manual [9], gazette [5], newsletter [4], annual [3], digest [3], circular [2]

- [5 sentences] magazine@@2

  Instead , the film is pulled through the camera solely through the power of camera sprockets until the end , at which point springs or belts in the camera **magazine** pull the film back to the take - up side.

  cartridge [6], clip [5], chamber [3], holder [3], mag [3], ammunition chamber [2], cache [2], loading chamber [2]

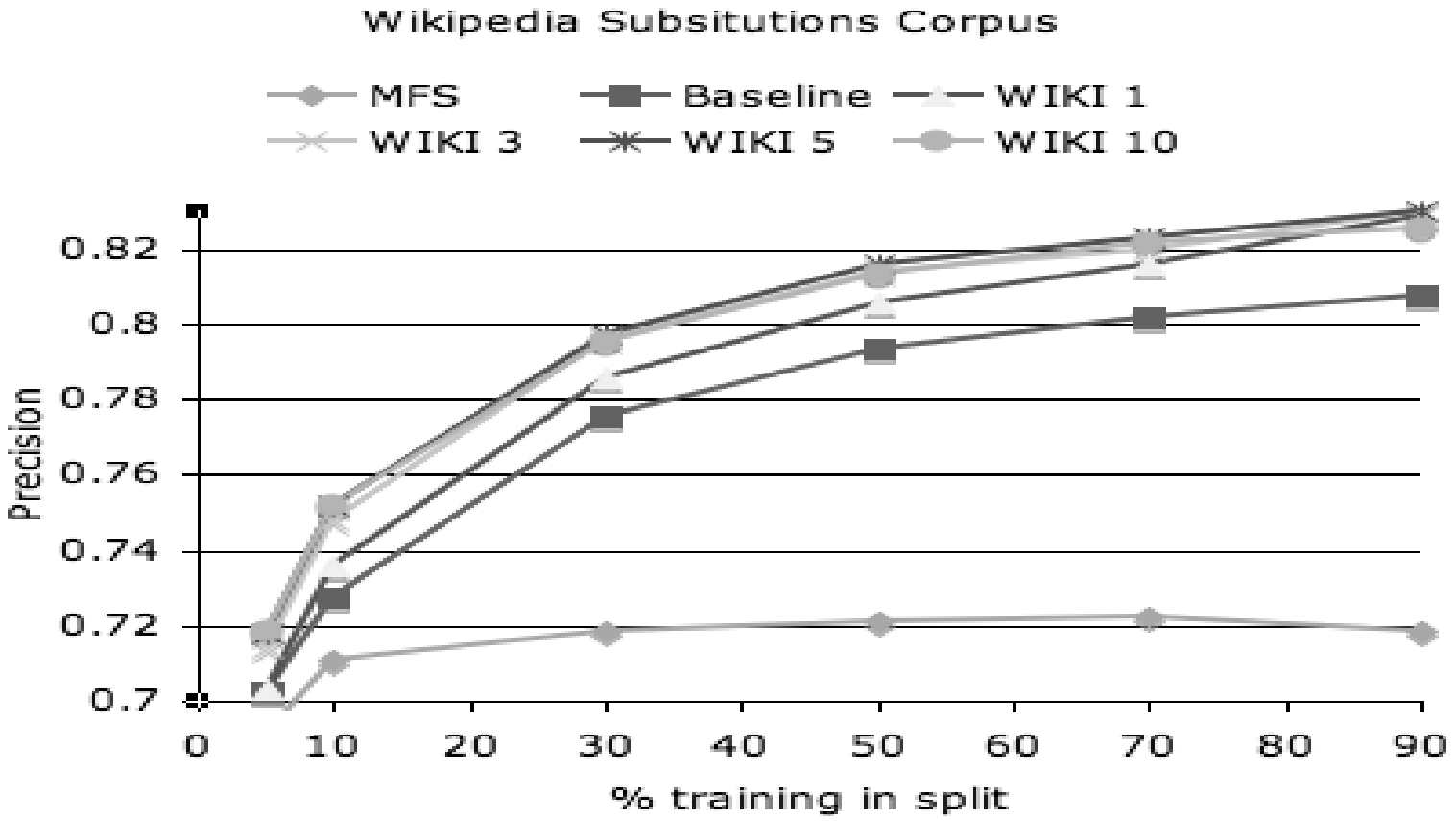
TWSI in Numbers

- Created using a bootstrapping process using AMT
- 397 words (top frequent nouns), all but 50 from trusted turkers
- $8.30 cost per word on average
- 2.1 senses / word (WordNet: 6.3)
- avg. 63 sample sentences per sense
- 51,736 sentences with target word sense labels
- avg. 17 substitutions with count>=2 per word
- avg. 4.5 substitutions with count>=10 per word

http://aclweb.org/aclwiki/index.php?title=Image:TWSI397.zip

Experiment 2a) Disambiguation

- Sense annotation: TWSI 1.0
- Corpus for clustering: Wikipedia
- Learning curve (ambiguous targets only):



Experiment 2b: Substitution Quality

	Substitutions		
	Gold	**System**	**Random**
YES	469 (93.8%)	456 (91.2%)	12 (2.4%)
NO	14 (2.8%)	27 (5.4%)	485 (97.0%)
SOMEWHAT	17 (3.4%)	17 (3.4%)	3 (0.6%)

coverage	**YES**	**NO**
100%	91.2%	5.4%
95%	91.8%	3.4%
90%	93.8%	2.9%
80%	94.8%	2.0%
70%	95.7%	0.9%

Conclusion

- Motivated WSD for Semantic IR
- Used co-occurrence clustering as features in supervised WSD task
- Showed state-of-the-art performance on standard WSD task
- Demonstrated high substitution quality using the TWSI

Cheap way to model topicality requiring only a POS-tagged corpus

THANKS FOR YOUR ATTENTION!

QUESTIONS?

Abstract

This paper examines the influence of features based on clusters of co-occurrences for supervised Word Sense Disambiguation and Lexical Substitution.

Cooccurrence cluster features are derived from clustering the local neighborhood of a target word in a co-occurrence graph based on a corpus in a completely unsupervised fashion. Clusters can be assigned in context and are used as features in a supervised WSD system.

Experiments fitting a strong baseline system with these additional features are conducted on two datasets, showing improvements.

Cooccurrence features are a simple way to mimic Topic Signatures (Martınez et al., 2008) without needing to construct resources manually. Further, a system is described that produces lexical substitutions in context with very high precision.