Knowledge-rich Word Sense Disambiguation

Recent Developments in Computational Semantics, WS 2013/14

Susanne Fertmann

January 13th 2013

Motivation

- Supervised and unsupervised systems require large amount of hand annotated data

→ **Knowledge-based** approaches exploit information from resources as WordNet

But:

- Resources as WordNet typically insufficient for high-performance WSD
- Manual extension costly

  → Automatic enrichment of WordNet with semantic relations

Overview

- Enrichment of WordNet with semantic relations from Wikipedia

  → WordNet++

  - Associate Wikipedia pages with WordNet senses
  - Transfer relations between Wikipedia pages to WordNet
  - Evaluation
- Using WordNet++ for knowledge-rich WSD
  - Integration into two simple knowledge-based algorithms
  - Evaluation against supervised and unsupervised state-of-the-art systems
- Conclusion and Summary

WordNet

- **Synsets** = set of words with a common meaning, e.g. *soda drink*

  {S: (n) pop.2, soda.2, soda pop.1, soda water.2, tonic.2}

- **Gloss**

  “a sweet drink containing carbonated water and flavoring“

Wikipedia

- Large amount of information
- Free of cost
- Wikipages about concepts or named entities (nouns)
- Wikipages can be taken as word senses
- Hyperlinks to related pages

Soda (Soft drink)

A soft drink (also called soda, pop, coke,[1] soda pop, fizzy drink, tonic, seltzer, mineral,[2] sparkling water, lolly water or carbonated beverage) is a beverage that typically contains water (often, but not always, carbonated water), usually a sweetener and usually a flavoring agent. The sweetener may be sugar, high-fructose corn syrup, fruit juice, sugar substitutes (in the case of diet drinks) or some combination of these. Soft drinks may also contain caffeine, colorings, preservatives and other ingredients.

Mapping Wikipedia to WordNet

- Links between Wikipages and WordNet senses

$$\mu : \text{SensesWiki} \rightarrow \text{SensesWN}$$

$$\mu(\text{Soda (soft drink)}) = \text{soda.2}$$

- Idea:
  - Identify **disambiguation contexts** for Wikipages and WordNet senses
  - Intersect these for mapping

- Which are the disambiguation contexts and how does the mapping work?

Disambiguation Context of a Wikipage

- **Sense labels** from Wikipage titles as “Soda (soft drink)“
  - soft, drink
- **Links** (outgoing): lemmas of the linked pages titles
  - beverage, sugar, cola, ...
- **Categories** (their syntactic heads):
  - drink, food

- Context(Soda (soft drink)) = {soft, drink, beverage, sugar, cola, drink, food, ...}

Disambiguation Context of a WordNet Sense

- **Synonyms**: {S: (n) pop.2, soda.2, soda pop.1, soda water.2, tonic.2}
  - pop, soda pop, soda water, tonic
- **Hypernyms and Hyponyms**: e.g. soft drink.2
  - soft, drink, ..
- **Sisters** (two synsets with the same hypernym), e.g. bitter lemon.1
  - bitter, lemon, …
- **Gloss**: “a sweet drink containing carbonated water and flavoring” → lemmas of context words
  - sweet, drink, contain, carbonated, water, flavoring

- Context(soda.2) = {pop, soda pop, soda water, tonic, soft, drink, bitter, lemon, sweet, drink, contain, carbonated, water, flavoring}

Mapping Algorithm

- Maps each Wikipage to a WordNet sense
- If a lemma is monosemous in Wikipedia and WordNet (e.g. antiproton):

  μ(w) = s.1
- Do the same if there is a redirection to a monosemous Wikipage
- Else: maximize conditional probability p(s|w)

$$\mu(w) = \underset{s \in Senses_{\mathrm{WN}}(w)}{\operatorname{argmax}} p(s|w) = \underset{s}{\operatorname{argmax}} \frac{p(s,w)}{p(w)}$$

$$= \underset{s}{\operatorname{argmax}}\, p(s,w)$$

Mapping Algorithm

- Maximize $$p(s,w) = \frac{score(s,w)}{\sum_{\substack{s' \in Senses_{\text{WN}}(w), \\ w' \in Senses_{\text{Wiki}}(w)}} score(s',w')},$$

score(s,w) = |Context(s) ∩ Context(w)| +1

Context(Soda (soft drink)) = {soft, drink, beverage, sugar, cola, food, water,...}

Context(soda.2) = {pop, soda pop, soda water, tonic, soft, drink, bitter, lemon, sweet, contain, water,...}

Context(soda.1) = {salt, acetate, chlocate, benzoate,..}

Mapping Algorithm

- Maximize $$p(s,w) = \frac{score(s,w)}{\sum_{\substack{s' \in Senses_{\text{WN}}(w), \\ w' \in Senses_{\text{Wiki}}(w)}} score(s',w')},$$

score(s,w) = |Context(s) ∩ Context(w)| +1

Context(Soda (soft drink)) = {soft, drink, beverage, sugar, cola, food, water,...}

Context(soda.2) = {pop, soda pop, soda water, tonic, soft, drink, bitter, lemon, sweet, contain, water,...}

Context(soda.1) = {salt, acetate, chlocate, benzoate,..}

Mapping Algorithm

- Maximize $$p(s,w) = \frac{score(s,w)}{\sum_{\substack{s' \in Senses_{\text{WN}}(w), \\ w' \in Senses_{\text{Wiki}}(w)}} score(s',w')},$$

score(s,w) = |Context(s) ∩ Context(w)| +1

Context(Soda (soft drink)) = {soft, drink, beverage, sugar, cola, food, water,...}

Context(soda.2) = {pop, soda pop, soda water, tonic, soft, drink, bitter, lemon, sweet, contain, water,...}

Context(soda.1) = {salt, acetate, chlorate, benzoate,..}

→ μ(Soda (soft drink)) = soda.2

Transferring Semantic Relations from Wikipedia to WordNet

- Collect all links w' occurring in a Wikipage w
- If w and w' are mapped to WordNet senses: transfer the corresponding edge (μ(w), μ(w')) to WordNet → WordNet++
- Example:
  - if “Soda (soft drink)” is linked with “Syrup” in Wikipedia
  - μ(Soda(soft drink)) = soda.2
  - μ(syrup) = syrup.1

    → add relation (soda.2, syrup.1) to WordNet++

- Semantic associative relations between synsets → unlabeled relations

Evaluation

- Mapping Quality
- Impact of WordNet++ for coarse-grained WSD
- Impact of WordNet++ for domain-specific WSD

Evaluation - Mapping Quality

- ~80,000 lemmas intersection between WordNet and Wikipedia
- Manual annotation of 1,000 Wikipage titles with WordNet labels

  → 505 mappings (505 titles had a corresponding WordNet sense)
- Part of it was re-annotated by a second annotator

  → high inter-annotator-agreement ($\varkappa$ = 0.9)

Results – Mapping Quality

- Baseline MFS: most frequent sense
- Different disambiguation contexts for WordNet senses (*structure* = synonyms, hypernyms, hyponyms, sisters)

	P	R	F_1	A
Structure	**82.2**	68.1	74.5	81.1
Gloss	81.1	64.2	71.7	78.8
Structure + Gloss	81.9	**77.5**	**79.6**	**84.4**
MFS BL	24.3	47.8	32.2	24.3
Random BL	23.8	46.8	31.6	23.9

- Much improvement over the baselines
- Different disambiguation contexts produce different mappings
- Richer disambiguation context helps
- No significant difference between the two baselines

Results

- ~ 81,000 Wikipedia - WordNet mappings
- ~ 1,900,000 semantic edges
  - Of which 98% were new edges (no direct edge between the senses)
  - 87% for depth 3

Impact on coarse-grained WSD

- Two simple knowledge-based algorithms:

Simplified Extended Lesk (ExtLesk): based on word overlap

Degree Centrality (Degree): graph based

Impact on coarse-grained WSD

- Two simple knowledge-based algorithms:

Simplified Extended Lesk (ExtLesk): based on word overlap

- Assign to target word w the sense whose gloss has the highest word overlap with the context of w
- Gloss = gloss of the synset of w plus glosses that have a semantical relation with s in WordNet++

Degree Centrality (Degree): graph based

Impact on coarse-grained WSD

- Two simple knowledge-based algorithms:

Simplified Extended Lesk (ExtLesk): based on word overlap

- Assign to target word w the sense whose gloss has the highest word overlap with the context of w
- Gloss = gloss of the synset of w plus glosses that have a semantical relation with s in WordNet++

Degree Centrality (Degree): graph based

- For each sense s of the target word, collect all paths connecting s to senses of other words in context
- Select sense with highest vertex degree (in- and outgoing edges)
- Suffers from noise in semantic links (Soda (soft drink) – Europe)

  → filter to delete weak semantic relations from WordNet++ (based on lemma overlap in categories) → only ~150,000 new semantic relations in WordNet++

Results

Resource	Algorithm	Nouns only P	R	F_1
WordNet	ExtLesk	83.6	57.7	68.3
	Degree	86.3	65.5	74.5
Wikipedia	ExtLesk	82.3	64.1	72.0
	Degree	**96.2**	40.1	57.4
WordNet++	ExtLesk	82.7	69.2	75.4
	Degree	87.3	72.7	**79.4**
	MFS BL	77.4	**77.4**	77.4
	Random BL	63.5	63.5	63.5

- Wikipedia better precision than WordNet for ExtLesk → high quality relations
- With filtering rule for Degree → high precision, low recall
- Degree WordNet++ beats the MFS baseline (difficult for unsupervised and knowledge-poor system)

Evaluation against state-of-the-art systems

- Dataset: SemEval-2007 (coarse grained)
- MFS as backoff

SUSSX-FR: unsupervised state-of-the-art

TreeMatch: unsupervised

NUS-PT: supervised state-of-the-art

SSI: knowledge-based

Algorithm	Nouns only P/R/F_1	All words P/R/F_1
ExtLesk	81.0	79.1
Degree	**85.5**	**81.7**
SUSSX-FR	81.1	77.0
TreeMatch	N/A	73.6
NUS-PT	82.3	**82.5**
SSI	**84.1**	**83.2**
MFS BL	77.4	78.9
Random BL	63.5	62.7

- Best WordNet++ system outperforms most systems for nouns only

  → state-of-the-art results with much simpler algorithm

Domain WSD

- Evaluation against sport and finance sections of domain corpora
- Fine-grained WSD
- Outperform other systems

  → more robust performance when evaluated across domains

Conclusions and Future Work

- Enriched WordNet integrated into (simple) knowledge-based systems yields compatible results to state-of-the-art supervised and unsupervised approaches in WSD
- Knowledge-based systems perform better in domain-specific scenarios

- Integrate WordNet++ in more sophisticated systems to see whether it improves performance
- Expansion to other languages

Summary

- Automatic enrichment of WordNet with encyclopedic knowledge (Wikipedia)
- High mapping quality
- Enables simple WSD systems to perform as well as state-of the art supervised and unsupervised systems (in coarse grained WSD)

  → Knowledge-rich WSD is competitive to supervised systems, even when using a simpler system

References

- S. Ponzetto, R. Navigli (2010): Knowledge-Rich Word Sense Disambiguation Rivaling Supervised Systems. Proceedings of ACL 2010.