Personalized PageRank over WordNet for Similarity and Word Sense Disambiguation

Eneko Agirre
e.agirre@ehu.es
(joint work with Aitor Soroa, some slides from Enrique Alfonseca)

University of the Basque Country
(Currently visiting Stanford)

Google, 2009

Summary

- Present an integrated software based on Knowledge Bases (e.g. WordNet) for:
  - Similarity of word pairs
  - Disambiguate words with respect to knowledge base concepts (aka Word Sense Disambiguation)
- Excellent results (EACL, NAACL, IJCAI 2009)
- Open source: `http://ixa2.si.ehu.es/ukb/`

Outline

1. Introduction
2. WordNet, PageRank and Personalized PageRank
3. PPR for similarity [Agirre et al.2009b]
4. PPR for WSD [Agirre and Soroa2009]
5. PPR and WSD on specific domains [Agirre et al.2009a]
6. Conclusions

Similarity

- Measuring semantic similarity and relatedness are well studied problems in lexical semantics:
  - Given two words or multiword-expressions, estimate how similar or related they are.
  - Relatedness is a more general relationship, including topical relatedness or meronymy.
  - Typically implemented as calculating a numeric value of similarity/relatedness.

Similarity examples

RG dataset		WordSim353 dataset	
cord smile	0.02	king cabbage	0.23
rooster voyage	0.04	professor cucumber	0.31
noon string	0.04	...	
...		investigation effort	4.59
glass jewel	1.78	smart student	4.62
magician oracle	1.82	...	
...		movie star	7.38
cushion pillow	3.84	...	
cemetery graveyard	3.88	journey voyage	9.29
automobile car	3.92	midday noon	9.29
midday noon	3.94	fuck sex	9.44
gem jewel	3.94	tiger tiger	10.00

Similarity

- Two main approaches:
  - Knowledge-based (Roget's Thesaurus, WordNet, etc.)
  - Corpus-based, also known as distributional similarity (co-occurrences)
- Many potential **applications**, overcome brittleness (word match), specially in very short texts, information retrieval, textual entailment, machine translation.

Similarity

- Two main approaches:
  - Knowledge-based (Roget's Thesaurus, WordNet, etc.)
  - Corpus-based, also known as distributional similarity (co-occurrences)
- Many potential **applications**, overcome brittleness (word match), specially in very short texts, information retrieval, textual entailment, machine translation.

Word Sense Disambiguation (WSD)

- Goal: determine the senses of the words in a text.
  - **"...but the location on the south bank of the Thames estuary."**
  - **"...cash includes cheque payments, bank transfers ... "**
- Dictionary (e.g. WordNet):
  - **bank#1** sloping land, especially the slope beside a body of water.
  - **bank#2** a financial institution that accepts deposits and...
  - bank#3 an arrangement of similar objects in row or in tiers.
  - bank#4 a long ridge or pile.
  - ...(10 senses total)
- Many potential applications, enable natural language understanding, link text to knowledge base, deploy semantic web.

Word Sense Disambiguation (WSD)

- Goal: determine the senses of the words in a text.
  - **“. . . but the location on the south bank of the Thames estuary.”**
  - **“. . . cash includes cheque payments, bank transfers . . . ”**
- Dictionary (e.g. WordNet):
  - **bank#1** sloping land, especially the slope beside a body of water.
  - **bank#2** a financial institution that accepts deposits and. . .
  - bank#3 an arrangement of similar objects in row or in tiers.
  - bank#4 a long ridge or pile.
  - . . . (10 senses total)
- Many potential applications, enable natural language understanding, link text to knowledge base, deploy semantic web.

Word Sense Disambiguation (WSD)

- Goal: determine the senses of the words in a text.
  - **“. . . but the location on the south bank of the Thames estuary.”**
  - **“. . . cash includes cheque payments, bank transfers . . . ”**
- Dictionary (e.g. WordNet):
  - **bank#1** sloping land, especially the slope beside a body of water.
  - **bank#2** a financial institution that accepts deposits and. . .
  - bank#3 an arrangement of similar objects in row or in tiers.
  - bank#4 a long ridge or pile.
  - . . . (10 senses total)
- Many potential applications, enable natural language understanding, link text to knowledge base, deploy semantic web.

Word Sense Disambiguation (WSD)

- Supervised corpus-based WSD performs best
  - Train classifiers on hand-tagged data (typically SemCor)
  - Data sparseness, e.g. *bank* 48 examples (25,20,2,1,0. . .)
  - Results decrease when train/test from different sources (even Brown, BNC)
  - Decrease even more when train/test from different domains
- Knowledge-based WSD
  - Uses information in a KB (WordNet)
  - Performs close to but lower than Most Frequent Sense
  - Vocabulary coverage
  - Relation coverage
  - But . . .

Word Sense Disambiguation (WSD)

- Supervised corpus-based WSD performs best
  - Train classifiers on hand-tagged data (typically SemCor)
  - Data sparseness, e.g. *bank* 48 examples (25,20,2,1,0. . .)
  - Results decrease when train/test from different sources (even Brown, BNC)
  - Decrease even more when train/test from different domains
- Knowledge-based WSD
  - Uses information in a KB (WordNet)
  - Performs close to but lower than Most Frequent Sense
  - Vocabulary coverage
  - Relation coverage
  - But . . .

Domain adaptation

Deploying NLP techniques in real applications is challenging, specially for WSD:

- Sense distributions change across domains
- Data sparseness hurts more
- Context overlap is reduced
- New senses, new terms

But...

- Some words get less interpretations in domains:
  bank in finance, *coach* in sports

Domain adaptation

Deploying NLP techniques in real applications is challenging, specially for WSD:

- Sense distributions change across domains
- Data sparseness hurts more
- Context overlap is reduced
- New senses, new terms

But. . .

- Some words get less interpretations in domains:
  bank in finance, *coach* in sports

Similarity and WSD

If using knowledge-bases, both WSD and Similarity are closely intertwined:

- Similarity between words based on similarity between senses (implicitly doing disambiguation)
- WSD uses similarity of senses to context, or similarity between senses in context

Outline

1. Introduction
2. WordNet, PageRank and Personalized PageRank
3. PPR for similarity [Agirre et al.2009b]
4. PPR for WSD [Agirre and Soroa2009]
5. PPR and WSD on specific domains [Agirre et al.2009a]
6. Conclusions

Outline

1. Introduction
2. WordNet, PageRank and Personalized PageRank
3. PPR for similarity [Agirre et al.2009b]
4. PPR for WSD [Agirre and Soroa2009]
5. PPR and WSD on specific domains [Agirre et al.2009a]
6. Conclusions

Wordnet

- Most widely used hierarchically organized lexical database for English (Fellbaum, 1998)
- Broad coverage of nouns, verbs, adjectives, adverbs
- Main unit: *synset* (concept)
  - **depository financial institution, bank#2, banking company**
    a financial institution that accepts deposits and...
- Relations between concepts:
  synonymy (built-in), hyperonymy, antonymy, meronymy, entailment, derivation, gloss
- Closely linked versions in several languages

Wordnet

Example of hypernym relations:

bank
- financial institution, financial organization
  - organization
    - social group
      - group, grouping
        - abstraction, abstract entity
          - entity

Representing WordNet as a graph:

- Nodes represent concepts
- Edges represent relations (undirected)
- In addition, directed edges from words to corresponding concepts (senses)

PageRank

- Given a graph, ranks nodes according to their relative structural importance
- If an edge from n_i to n_j exists, a vote from n_i to n_j is produced
  - Strength depends on the rank of n_i
  - The more important n_i is, the more strength its votes will have.
- PageRank can also be viewed as the result of a random walk process
  - Rank of n_i represents the probability of a random walk over the graph ending on n_i, at a sufficiently large time.

PageRank

- G: graph with N nodes $n_1, \ldots, n_N$
- d_i: outdegree of node i
- M: $N \times N$ matrix

$$M_{ji} = \begin{cases} \dfrac{1}{d_i} & \text{an edge from } i \text{ to } j \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

PageRank equation:

$$\mathbf{Pr} = c M \mathbf{Pr} + (1 - c)\mathbf{v}$$

- voting scheme
- a surfer randomly jumping to any node without following any paths on the graph

c: damping factor: the way in which these two terms are combined at each step

PageRank

- G: graph with N nodes $n_1, \ldots, n_N$
- d_i: outdegree of node i
- M: $N \times N$ matrix

$$M_{ji} = \begin{cases} \frac{1}{d_i} & \text{an edge from } i \text{ to } j \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

PageRank equation:

$$\mathbf{Pr} = cM\mathbf{Pr} + (1-c)\mathbf{v}$$

- voting scheme
- a surfer randomly jumping to any node without following any paths on the graph

c: damping factor: the way in which these two terms are combined at each step

PageRank

- G: graph with N nodes $n_1, \ldots, n_N$
- d_i: outdegree of node i
- M: $N \times N$ matrix

$$M_{ji} = \begin{cases} \frac{1}{d_i} & \text{an edge from } i \text{ to } j \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

PageRank equation:

$$\mathbf{Pr} = \boxed{cM\mathbf{Pr}} + \boxed{(1-c)\mathbf{v}}$$

- voting scheme
- a surfer randomly jumping to any node without following any paths on the graph

c: damping factor: the way in which these two terms are combined at each step

PageRank

- G: graph with N nodes $n_1, \ldots, n_N$
- d_i: outdegree of node i
- M: $N \times N$ matrix

$$M_{ji} = \begin{cases} \frac{1}{d_i} & \text{an edge from } i \text{ to } j \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

PageRank equation:

$$\mathbf{Pr} = \boxed{cM\mathbf{Pr}} + \boxed{(1-c)\mathbf{v}}$$

- voting scheme
- a surfer randomly jumping to any node without following any paths on the graph

c: damping factor: the way in which these two terms are combined at each step

PageRank

- G: graph with N nodes $n_1, \ldots, n_N$
- d_i: outdegree of node i
- M: $N \times N$ matrix

$$M_{ji} = \begin{cases} \frac{1}{d_i} & \text{an edge from } i \text{ to } j \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

PageRank equation:

$$\mathbf{Pr} = cM\mathbf{Pr} + (1-c)\mathbf{v}$$

- voting scheme
- a surfer randomly jumping to any node without following any paths on the graph

c: damping factor: the way in which these two terms are combined at each step

Personalized PageRank

$$\mathbf{Pr} = cM\mathbf{Pr} + (1-c)\mathbf{v}$$

- PageRank: $\mathbf{v}$ is a stochastic normalized vector, with elements $\frac{1}{N}$
  - Equal probabilities to all nodes in case of random jumps
- **Personalized PageRank**, non-uniform $\mathbf{v}$ [Haveliwala2002]
  - Assign stronger probabilities to certain kinds of nodes
  - Bias PageRank to prefer these nodes
- For ex. if we concentrate all mass on node i
  - All random jumps return to n_i
  - Rank of i will be high
  - High rank of i will make all the nodes in its vicinity also receive a high rank
  - Importance of node i given by the initial $\mathbf{v}$ spreads along the graph

Personalized PageRank

$$\mathbf{Pr} = cM\mathbf{Pr} + (1-c)\mathbf{v}$$

- PageRank: $\mathbf{v}$ is a stochastic normalized vector, with elements $\frac{1}{N}$
  - Equal probabilities to all nodes in case of random jumps
- **Personalized PageRank**, non-uniform $\mathbf{v}$ [Haveliwala2002]
  - Assign stronger probabilities to certain kinds of nodes
  - Bias PageRank to prefer these nodes
- For ex. if we concentrate all mass on node i
  - All random jumps return to n_i
  - Rank of i will be high
  - High rank of i will make all the nodes in its vicinity also receive a high rank
  - Importance of node i given by the initial $\mathbf{v}$ spreads along the graph

Outline

1. Introduction
2. WordNet, PageRank and Personalized PageRank
3. PPR for similarity [Agirre et al.2009b]
4. PPR for WSD [Agirre and Soroa2009]
5. PPR and WSD on specific domains [Agirre et al.2009a]
6. Conclusions

PPR for similarity [Agirre et al.2009b]

Based on [Hughes and Ramage2007]

- Given a pair of words (w1, w2),
  - Initialize teleport probability mass on either w1 or w2.
  - Run PPR
- The similarity is given by the cosine of the two PPR vectors.
- Experiment settings:
  - Damping value c = 0.85
  - Calculations finish after 30 iterations
- Variations for Knowledge Base:
  - MCR (WordNet 1.6, closely linked to Spanish WordNet) and WordNet 3.0
  - All WordNet relations, All WN+gloss relations

Datasets

Rubenstein and Goodenough (1965)

- 80 word pairs, judged by 51 human subjects
- Scale 0 to 4 based on their similarity
- Redone for a subset by Miller and Charles (1991)

WordSim353 dataset:

- Finkelstein et al. (2002)
- 353 word pairs, each with 13-16 human judgments
- Annotators were asked to rate similarity and relatedness.

Results given by rank correlation of system output with human ratings (Spearman)

Results

Competition with 1.6Twords distributional thesaurus in Google.

Method	Window size	RG dataset	WordSim353 dataset
MCR16		0.83	0.53 (0.56)
WN30		0.79	0.56 (0.58)
WN30g		0.83	**0.66 (0.69)**
CW	1	0.83	0.63
	2	0.83	0.60
	3	0.85	0.59
	4	**0.89**	0.60
	5	0.80	0.58
	6	0.75	0.58
	7	0.72	0.57
BoW	1	0.81	0.64
	2	0.80	0.64
	3	0.79	0.64
	4	0.78	0.65
Syn	G1,D0	0.81	0.62
	G2,D0	0.82	0.55
	G3,D0	0.81	0.62
CW+	4; G1,D0	0.88	**0.66**
Syn	4; G2,D0	0.87	0.64

Results

Unknown words in WordNet

Method	Spearman	Interval
WN30	0.56 (0.58)	[0.48, 0.63]
WN30 $\cup$ th	0.58	[0.51, 0.65]
WN30g	0.66 (0.69)	[0.59, 0.71]
WN30g $\cup$ th	**0.68**	[0.62, 0.73]

Results

State-of-the-art on MC (subset of RG)

Method	Source	Spearman (MC)	Pearson (MC)
(Sahami et al., 2006)	Web snippets	0.62 [0.32, 0.81]	0.58 [0.26, 0.78]
(Chen et al., 2006)	Web snippets	0.69 [0.42, 0.84]	0.69 [0.42, 0.85]
(Wu and Palmer, 1994)	WordNet	0.78 [0.59, 0.90]	0.78 [0.57, 0.89]
(Leacock et al., 1998)	WordNet	0.79 [0.59, 0.90]	0.82 [0.64, 0.91]
(Resnik, 1995)	WordNet	0.81 [0.62, 0.91]	0.80 [0.60, 0.90]
(Lin, 1998a)	WordNet	0.82 [0.65, 0.91]	0.83 [0.67, 0.92]
(Bollegala et al., 2007)	Web snippets	0.82 [0.64, 0.91]	0.83 [0.67, 0.92]
(Jiang and Conrath, 1997)	WordNet	0.83 [0.67, 0.92]	0.85 [0.69, 0.93]
(Jarmasz, 2003)	Roget's	0.87 [0.73, 0.94]	0.87 [0.74, 0.94]
(Patwardhan et al., 2006)	WordNet	n/a	0.91
(Alvarez and Lim, 2007)	WordNet	n/a	0.91
(Yang and Powers, 2005)	WordNet	0.87 [0.73, 0.91]	0.92 [0.84, 0.96]
(Hughes et al., 2007)	WordNet	0.90	n/a
Personalized PageRank	WordNet	0.89 [0.77, 0.94]	n/a
Bag of words	Web corpus	0.85 [0.70, 0.93]	0.84 [0.69, 0.93]
Context window	Web corpus	0.88 [0.76, 0.95]	0.89 [0.77, 0.95]
Syntactic contexts	Web corpus	0.76 [0.54, 0.88]	0.74 [0.51, 0.87]

Results

State-of-the-art on WordSim 353

Method	Source	Spearman
[Strube and Ponzetto2006]	Wikipedia	0.19–0.48
[Jarmasz2003]	WordNet	0.33–0.35
[Jarmasz2003]	Roget's	0.55
[Hughes and Ramage2007]	WordNet	0.55
[Finkelstein et al.2002]	Web corpus, WN	0.56
[Gabrilovich and Markovitch2007]	ODP	0.65
[Gabrilovich and Markovitch2007]	Wikipedia	0.75
Personalized PageRank	WordNet	0.66 (0.69)

Cross-lingual evaluation

Consider pairs of words from different languages.
Can we predict the similarities?

- WordNet-based method:
  - English WordNet graph, crosslingual lexical entries in synsets.
  - Personalized PageRank is calculated in the same way
- Contextual method:
  - Get the top 5 translations of the non-English word into English using the Google Machine Translation system.
  - Generate the context vectors for those 5 translations separately.
  - Add the vectors.
  - The rest of the procedure is the same.
- Evaluation:
  - RG and WordSim353
  - One of the words in each pair translated into Spanish

Cross-lingual evaluation

Dataset	Method	overall	Δ	interval
RG	MCR16	0.78	-0.05	[0.66, 0.86]
	WN30g	0.74	-0.09	[0.61, 0.84]
	Bag of words	0.68	-0.23	[0.53, 0.79]
	Context windows	**0.83**	-0.05	[0.73, 0.89]
WS353	MCR16	0.42 (0.53)	-0.11 (-0.03)	[0.34, 0.51]
	WN30g	**0.58** (0.67)	-0.07 (-0.02)	[0.51, 0.64]
	Bag of words	0.53	-0.12	[0.45, 0.61]
	Context windows	0.52	-0.11	[0.44, 0.59]

Outline

1. Introduction
2. WordNet, PageRank and Personalized PageRank
3. PPR for similarity [Agirre et al.2009b]
4. PPR for WSD [Agirre and Soroa2009]
5. PPR and WSD on specific domains [Agirre et al.2009a]
6. Conclusions

Knowledge-based WSD

- Use information in WordNet for disambiguation:
  - **"... cash includes cheque payments, bank transfers ... "**
- Traditional approach [Patwardhan et al.2007]:
  - Compare each target sense of **bank** with those of the words in the context
  - Using semantic relatedness between pairs of senses
  - Combinatorial explosion: each word disambiguated individually
    - sim(bank#1,cheque#1) + sim(bank#1,cheque#2) + sim(bank#1,payment#1) ...
    - sim(bank#2,cheque#1) + sim(bank#2,cheque#2) + sim(bank#2,payment#1) ...
    - ...
- Graph-based methods
  - Exploit the structural properties of the graph underlying WordNet
  - Find globally optimal solutions
  - Disambiguate large portions of text in one go
  - Principled solution to combinatorial explosion

Knowledge-based WSD

- Use information in WordNet for disambiguation:
  - **"... cash includes cheque payments, bank transfers ... "**
- Traditional approach [Patwardhan et al.2007]:
  - Compare each target sense of **bank** with those of the words in the context
  - Using semantic relatedness between pairs of senses
  - Combinatorial explosion: each word disambiguated individually
    - sim(bank#1,cheque#1) $+$ sim(bank#1,cheque#2) $+$ sim(bank#1,payment#1) ...
    - sim(bank#2,cheque#1) $+$ sim(bank#2,cheque#2) $+$ sim(bank#2,payment#1) ...
    - ...
- Graph-based methods
  - Exploit the structural properties of the graph underlying WordNet
  - Find globally optimal solutions
  - Disambiguate large portions of text in one go
  - Principled solution to combinatorial explosion

Using PageRank for WSD

- Given a graph representation of the LKB
- PageRank over the whole WordNet would get a context-independent ranking of word senses
- We would like:
  - Given an input text, disambiguate all open-class words in the input taking the rest as context
- Two alternatives
  1. Create a context-sensitive subgraph and apply PageRank over it [Navigli and Lapata2007, Agirre and Soroa2008]
  2. Use **Personalized PageRank** over the complete graph, initializing $\mathbf{v}$ with the context words

Using PageRank for WSD

- Given a graph representation of the LKB
- PageRank over the whole WordNet would get a context-independent ranking of word senses
- We would like:
  - Given an input text, disambiguate all open-class words in the input taking the rest as context
- Two alternatives
  1. Create a context-sensitive subgraph and apply PageRank over it [Navigli and Lapata2007, Agirre and Soroa2008]
  2. Use **Personalized PageRank** over the complete graph, initializing $\mathbf{v}$ with the context words

Using Personalized PageRank (Ppr and Ppr_w2w)

- For each word $W_i, \; i = 1 \ldots m$ in the context
  - Initialize $\mathbf{v}$ with uniform probabilities over words W_i
    Context words act as source nodes injecting mass into the concept graph
  - Run Personalized PageRank
  - Choose highest ranking sense for target word
- Problem of *Ppr*
  - Senses of the same word might be linked
  - Those senses would reinforce each other and receive higher ranks
- *Ppr_w2w* alternative:
  - Let the surrounding words decide which concept associated to W_i has more relevance
  - For each target word W_i, concentrate the initial probability mass in words surrounding W_i, but not in W_i itself
  - Run Personalized PageRank for each word in turn (higher cost)

Using Personalized PageRank (Ppr and Ppr_w2w)

- For each word $W_i,\ i = 1 \dots m$ in the context
  - Initialize $\mathbf{v}$ with uniform probabilities over words W_i
    Context words act as source nodes injecting mass into the concept graph
  - Run Personalized PageRank
  - Choose highest ranking sense for target word
- Problem of *Ppr*
  - Senses of the same word might be linked
  - Those senses would reinforce each other and receive higher ranks
- *Ppr_w2w* alternative:
  - Let the surrounding words decide which concept associated to W_i has more relevance
  - For each target word W_i, concentrate the initial probability mass in words surrounding W_i, but not in W_i itself
  - Run Personalized PageRank for each word in turn (higher cost)

Using Personalized PageRank (Ppr and Ppr_w2w)

- For each word $W_i,\ i = 1 \ldots m$ in the context
  - Initialize $\mathbf{v}$ with uniform probabilities over words W_i
    Context words act as source nodes injecting mass into the concept graph
  - Run Personalized PageRank
  - Choose highest ranking sense for target word
- Problem of *Ppr*
  - Senses of the same word might be linked
  - Those senses would reinforce each other and receive higher ranks
- *Ppr_w2w* alternative:
  - Let the surrounding words decide which concept associated to W_i has more relevance
  - For each target word W_i, concentrate the initial probability mass in words surrounding W_i, but not in W_i itself
  - Run Personalized PageRank for each word in turn (higher cost)

Experiment setting

- Two datasets
  - Senseval 2 All Words (S2AW)
  - Senseval 3 All Words (S3AW)
- Both labelled with WordNet 1.7 tags
- Create input contexts of at least 20 words
  - Adding sentences immediately before and after if original too short
- PageRank settings:
  - Damping factor (c): 0.85
  - End after 30 iterations

Results and comparison to related work (S2AW)

(Mihalcea, 2005) Pairwise Lesk between senses, then PageRank.

(Sinha & Mihalcea, 2007) Several similarity measures, voting, fine-tuning for each PoS. Development over S3AW.

(Tsatsaronis et al., 2007) Subgraph BFS over WordNet 1.7 and eXtended WN, then spreading activation.

* No statistical significance (small dataset).

Senseval-2 All Words dataset					
System	All	N	V	Adj.	Adv.
Mih05	54.2	57.5	36.5	56.7	**70.9**
Sihna07	56.4	65.6	32.3	**61.4**	60.2
Tsatsa07	49.2	–	–	–	–
Ppr	56.8	**71.1**	33.4	55.9	67.1
Ppr_w2w	**58.6**	70.4	**38.9**	58.3	70.1
MFS	60.1	71.2	39.0	61.1	75.4

Results and comparison to related work (S2AW)

(Mihalcea, 2005) Pairwise Lesk between senses, then PageRank.

(Sinha & Mihalcea, 2007) Several similarity measures, voting, fine-tuning for each PoS. Development over S3AW.

(Tsatsaronis et al., 2007) Subgraph BFS over WordNet 1.7 and eXtended WN, then spreading activation.

* No statistical significance (small dataset).

Senseval-2 All Words dataset					
System	All	N	V	Adj.	Adv.
Mih05	54.2	57.5	36.5	56.7	**70.9**
Sihna07	56.4	65.6	32.3	**61.4**	60.2
Tsatsa07	49.2	–	–	–	–
Ppr	56.8	**71.1**	33.4	55.9	67.1
Ppr_w2w	**58.6**	70.4	**38.9**	58.3	70.1
MFS	60.1	71.2	39.0	61.1	75.4

Comparison to related work (S3AW)

(Mihalcea, 2005) Pairwise Lesk between senses, then PageRank.

(Sinha & Mihalcea, 2007) Several simmilarity measures, voting, fine-tuning for each PoS. Development over S3AW.

(Navigli & Lapata, 2007) Subgraph DFS(3) over WordNet 2.0 plus proprietary relations, several centrality algorithms.

(Navigli & Velardi, 2005) SSI algorithm on WordNet 2.0 plus proprietary relations. Uses MFS when undecided.

System	All	N	V	Adj.	Adv.
Mih05	52.2	-	-	-	-
Sihna07	52.4	60.5	40.6	54.1	**100.0**
Nav07	-	61.9	36.1	**62.8**	-
Ppr	56.1	62.6	46.0	60.8	92.9
Ppr_w2w	**57.4**	**64.1**	**46.9**	62.6	92.9
MFS	62.3	69.3	53.6	63.7	92.9
Nav05	60.4	-	-	-	-

Comparison to related work (S3AW)

(Mihalcea, 2005) Pairwise Lesk between senses, then PageRank.

(Sinha & Mihalcea, 2007) Several simmilarity measures, voting, fine-tuning for each PoS. Development over S3AW.

(Navigli & Lapata, 2007) Subgraph DFS(3) over WordNet 2.0 plus proprietary relations, several centrality algorithms.

(Navigli & Velardi, 2005) SSI algorithm on WordNet 2.0 plus proprietary relations. Uses MFS when undecided.

System	All	N	V	Adj.	Adv.
Mih05	52.2	-	-	-	-
Sihna07	52.4	60.5	40.6	54.1	**100.0**
Nav07	-	61.9	36.1	**62.8**	-
Ppr	56.1	62.6	46.0	60.8	92.9
Ppr_w2w	**57.4**	**64.1**	**46.9**	62.6	92.9
MFS	62.3	69.3	53.6	63.7	92.9
Nav05	60.4	-	-	-	-

Outline

1. Introduction
2. WordNet, PageRank and Personalized PageRank
3. PPR for similarity [Agirre et al.2009b]
4. PPR for WSD [Agirre and Soroa2009]
5. PPR and WSD on specific domains [Agirre et al.2009a]
6. Conclusions

Dataset [Koeling et al.2005]

- Examples from **BNC**, **Sports** and **Finances** sections Reuters
  - 41 nouns: salient in either domain or with senses linked to these domains
  - Sense inventory: WordNet v. 1.7.1
- 300 examples for each of the **41 nouns**
  - Roughly 100 examples from each word and corpus
- Freely available

Methods

- What would happen if we apply PPR-based WSD to specific domains?
- Personalized PageRank over **context**
  - **"...has never won a league title as coach but took Parma to success..."**
- Personalized PageRank over **related words**
  - Get related words from distributional thesaurus [Koeling et al.2005]
  - **coach: manager, captain, player, team, striker, ...**
- Experiments on BNC, Sports, Finance dataset:
  - Supervised: train MFS, SVM, k-NN on SemCor examples
  - Static PageRank
  - PPRank: Personalized PageRank (same damping factors, iterations)
    - Use context
    - 50 related words [Koeling et al.2005] (BNC, Sports, Finance)

Methods

- What would happen if we apply PPR-based WSD to specific domains?
- Personalized PageRank over **context**
  - **"...has never won a league title as coach but took Parma to success..."**
- Personalized PageRank over **related words**
  - Get related words from distributional thesaurus [Koeling et al.2005]
  - **coach: manager, captain, player, team, striker, ...**
- Experiments on BNC, Sports, Finance dataset:
  - Supervised: train MFS, SVM, k-NN on SemCor examples
  - Static PageRank
  - PPRank: Personalized PageRank (same damping factors, iterations)
    - Use context
    - 50 related words [Koeling et al.2005] (BNC, Sports, Finance)

Methods

- What would happen if we apply PPR-based WSD to specific domains?
- Personalized PageRank over **context**
  - **"... has never won a league title as coach but took Parma to success..."**
- Personalized PageRank over **related words**
  - Get related words from distributional thesaurus [Koeling et al.2005]
  - **coach: manager, captain, player, team, striker, ...**
- Experiments on BNC, Sports, Finance dataset:
  - Supervised: train MFS, SVM, k-NN on SemCor examples
  - Static PageRank
  - PPRank: Personalized PageRank (same damping factors, iterations)
    - Use context
    - 50 related words [Koeling et al.2005] (BNC, Sports, Finance)

Methods

- What would happen if we apply PPR-based WSD to specific domains?
- Personalized PageRank over **context**
  - **"...has never won a league title as coach but took Parma to success..."**
- Personalized PageRank over **related words**
  - Get related words from distributional thesaurus [Koeling et al.2005]
  - **coach: manager, captain, player, team, striker, ...**
- Experiments on BNC, Sports, Finance dataset:
  - Supervised: train MFS, SVM, k-NN on SemCor examples
  - Static PageRank
  - PPRank: Personalized PageRank (same damping factors, iterations)
    - Use context
    - 50 related words [Koeling et al.2005] (BNC, Sports, Finance)

Results

	Systems	BNC	Sports	Finances
Baselines	Random	*19.7	*19.2	*19.5
	SemCor MFS	*34.9	*19.6	*37.1
	Static PRank	*36.6	*20.1	*39.6
Supervised	SVM	*38.7	*25.3	*38.7
	k-NN	42.8	*30.3	*43.4
Context	PPRank	**43.8**	*35.6	*46.9
Related words	PPRank	*37.7	**51.5**	**59.3**
	[koeling et al. 2005]	*40.7	*43.3	*49.7
Skyline	Test MFS	*52.0	*77.8	*82.3

- Supervised (MFS, SVM, k-NN) very low (see test MFS)
- Static PageRank close to MFS
- PPRank on context: best for BNC (* for statistical significance)
- PPRank on related words: best for Sports and Finance and improves over Koeling *et al.*, who use pairwise WordNet similarity.

Conclusions

- Knowledge-based method for similarity and WSD
- Based on Personalized PageRank
- Exploits whole structure of underlying KB efficiently
- Performance:
  - Similarity: best WordNet, comparable with 1.6 Tword, slightly below ESA
  - WSD: Best KB algorithm S2AW, S3AW, Domains datasets
  - WSD and domains:
    - Better than supervised WSD for domains
    - Acquisition of terms and ontology enrichment feasible
    - Interest in fields like biomedicine, where ontologies exist

Conclusions

- Knowledge-based method for similarity and WSD
- Based on Personalized PageRank
- Exploits whole structure of underlying KB efficiently
- Performance:
  - Similarity: best WordNet, comparable with 1.6 Tword, slightly below ESA
  - WSD: Best KB algorithm S2AW, S3AW, Domains datasets
  - WSD and domains:
    - Better than supervised WSD for domains
    - Acquisition of terms and ontology enrichment feasible
    - Interest in fields like biomedicine, where ontologies exist

Conclusions

- Easily ported to other languages
  - Provides cross-lingual similarity
  - Only requirement of having a WordNet
- Publicly available at `http://ixa2.si.ehu.es/ukb`
  - Both programs and data
  - Including program to construct graphs from new KB (e.g. Wikipedia)
  - GPL license, open source, free

Conclusions

- Easily ported to other languages
  - Provides cross-lingual similarity
  - Only requirement of having a WordNet
- Publicly available at `http://ixa2.si.ehu.es/ukb`
  - Both programs and data
  - Including program to construct graphs from new KB (e.g. Wikipedia)
  - GPL license, open source, free

Personalized PageRank over WordNet for Similarity and Word Sense Disambiguation

Eneko Agirre
e.agirre@ehu.es
(joint work with Aitor Soroa, some slides from Enrique Alfonseca)

University of the Basque Country
(Currently visiting Stanford)

Google, 2009

E. Agirre and A. Soroa.
2008.
Using the multilingual central repository for graph-based word sense disambiguation.
In *Proceedings of LREC '08*, Marrakesh, Morocco.

E. Agirre and A. Soroa.
2009.
Personalizing pagerank for word sense disambiguation.
In *Proceedings of EACL-09*, Athens, Greece.

E. Agirre, O. Lopez de Lacalle, and A. Soroa.
2009a.
Knowledge-Based WSD on Specific Domains: Performing better than Generic Supervised WSD.
In *Proceedings of IJCAI*, Pasadena, USA.

E. Agirre, A. Soroa, E. Alfonseca, K. Hall, J. Kravalova, and M Pas.
2009b.
A study on similarity and relatedness using distributional and WordNet-based approaches.

In *Proceedings of annual meeting of the North American Chapter of the Association of Computational Linguistics (NAAC)*, Boulder, USA, June.

L. Finkelstein, E. Gabrilovich, Y. Matias, E. Rivlin, Z. Solan, G. Wolfman, and E. Ruppin.
2002.
Placing Search in Context: The Concept Revisited.
ACM Transactions on Information Systems, 20(1):116–131.

E. Gabrilovich and S. Markovitch.
2007.
Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis.
Proceedings of the 20th International Joint Conference on Artificial Intelligence, pages 6–12.

T. H. Haveliwala.
2002.
Topic-sensitive pagerank.
In *WWW '02: Proceedings of the 11th international conference on World Wide Web*, pages 517–526, New York, NY, USA. ACM.

Thad Hughes and Daniel Ramage.

2007.
Lexical semantic relatedness with random graph walks.
In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 581–589.

M. Jarmasz.
2003.
Roget's Thesuarus as a lexical resource for Natural Language Processing.

R. Koeling, D. McCarthy, and J. Carroll.
2005.
Domain-specific sense distributions and predominant sense acquisition.
In *Proceedings of the Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing. HLT/EMNLP*, pages 419–426, Ann Arbor, Michigan.

R. Mihalcea.
2005.
Unsupervised large-vocabulary word sense disambiguation with graph-based algorithms for sequence data labeling.

In *Proceedings of HLT05*, Morristown, NJ, USA.

R. Navigli and M. Lapata.
2007.
Graph connectivity measures for unsupervised word sense disambiguation.
In *IJCAI*.

S. Patwardhan, S. Banerjee, and T. Pedersen.
2007.
UMND1: Unsupervised Word Sense Disambiguation Using Contextual Semantic Relatedness.
In *Appears in the Proceedings of SemEval-2007: 4th International Workshop on Semantic Evaluations*, pages 390–393.

R. Sinha and R. Mihalcea.
2007.
Unsupervised graph-based word sense disambiguation using measures of word semantic similarity.
In *Proceedings of the IEEE International Conference on Semantic Computing (ICSC 2007)*, Irvine, CA, USA.

M. Strube and S.P. Ponzetto.

2006.
WikiRelate! Computing Semantic Relatedness Using Wikipedia.
In *Proceedings of the AAAI-2006*, pages 1419–1424.