Knowledge-Based Word Sense Disambiguation and Similarity using Random Walks

Eneko Agirre
ixa2.si.ehu.es/eneko

University of the Basque Country
(Currently visiting at Stanford)

SRI, 2011

Summary

- Knowledge-Based random walks...
  for similarity between words
  to map words in context to KB concepts Word Sense Disambiguation
  to improve ad-hoc information retrieval
- Applied to WordNet(s), UMLS, Wikipedia
- Excellent results (EACL, NAACL, IJCAI 2009, Bioinformatics, COLING, 2010, IJCNLP, CIKM 2011)
- Open source: `http://ixa2.si.ehu.es/ukb/`

Outline

Similarity

- Given two words or multiword-expressions, estimate how similar they are.
  - cord smile
  - gem jewel
  - magician oracle
  - Features shared, belonging to the same class
- Relatedness is a more general relationship, including other relations like topical relatedness or meronymy.
  - king cabbage
  - movie star
  - journey voyage
- Typically implemented as calculating a numeric value of similarity/relatedness.

Similarity

- Given two words or multiword-expressions, estimate how similar they are.
  - cord smile
  - gem jewel
  - magician oracle
  - Features shared, belonging to the same class
- Relatedness is a more general relationship, including other relations like topical relatedness or meronymy.
  - king cabbage
  - movie star
  - journey voyage
- Typically implemented as calculating a numeric value of similarity/relatedness.

Similarity

- Given two words or multiword-expressions, estimate how similar they are.
  - cord smile
  - gem jewel
  - magician oracle
  - Features shared, belonging to the same class
- Relatedness is a more general relationship, including other relations like topical relatedness or meronymy.
  - king cabbage
  - movie star
  - journey voyage
- 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	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)
  - NLP subtasks (parsing, semantic role labeling)
  - Information retrieval
  - Question answering
  - Summarization
  - Machine translation optimizat¡ion and evaluation
  - Inference (textual entailment)

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)
  - NLP subtasks (parsing, semantic role labeling)
  - Information retrieval
  - Question answering
  - Summarization
  - Machine translation optimizat¡ion and evaluation
  - Inference (textual entailment)

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 (MFS, supervised)
  - Vocabulary coverage
  - Relation coverage

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 (MFS, supervised)
  - Vocabulary coverage
  - Relation coverage

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

- bank river
- bank money

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

Outline

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

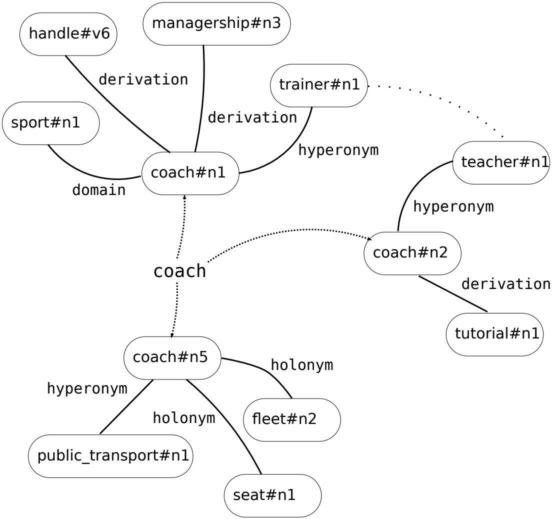
Example of hypernym relations:

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)

Wordnet


handle#v6
managership#n3
derivation
trainer#n1
sport#n1
derivation
hyperonym
teacher#n1
coach#n1
domain
hyperonym
coach#n2
coach
derivation
tutorial#n1
coach#n5
holonym
hyperonym
holonym
fleet#n2
public_transport#n1
seat#n1

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 is more commonly 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

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

PageRank equation:

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

- surfer follows edges
- surfer randomly jumps to any node (teleport)

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

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:

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

- surfer follows edges
- surfer randomly jumps to any node (teleport)

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

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:

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

- surfer follows edges
- surfer randomly jumps to any node (teleport)

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

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:

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

- surfer follows edges
- surfer randomly jumps to any node (teleport)

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

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:

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

- surfer follows edges
- surfer randomly jumps to any node (teleport)

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

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 $\mathbf{v}$ (Haveliwala 2002)
  - 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 $\mathbf{v}$ (Haveliwala 2002)
  - 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

Random walks for similarity // (with Aitor Soroa)

Based on (Hughes and Ramage, 2007)

- Given a pair of words (w_1, w_2),
  - Initialize teleport probability mass on w_1
  - Run Personalized Pagerank, obtaining to $\vec{w}_1$
  - Initialize w_2 and obtain $\vec{w}_2$
  - Measure similarity between $\vec{w}_1$ and $\vec{w}_2$ (e.g. cosine)
- Experiment settings:
  - Damping value c = 0.85
  - Calculations finish after 30 iterations
- Variations for Knowledge Base:
  - WordNet 3.0
  - WordNet relations
  - Gloss relations
  - other relations

Random walks for similarity // (with Aitor Soroa)

Based on (Hughes and Ramage, 2007)

- Given a pair of words (w_1, w_2),
  - Initialize teleport probability mass on w_1
  - Run Personalized Pagerank, obtaining to $\vec{w}_1$
  - Initialize w_2 and obtain $\vec{w}_2$
  - Measure similarity between $\vec{w}_1$ and $\vec{w}_2$ (e.g. cosine)
- Experiment settings:
  - Damping value c = 0.85
  - Calculations finish after 30 iterations
- Variations for Knowledge Base:
  - WordNet 3.0
  - WordNet relations
  - Gloss relations
  - other relations

Dataset and results

WordSim353 dataset (Finkelstein et al. 2002):

- 353 word pairs, each with 13-16 human judgments
- Annotators were asked to rate similarity and relatedness.
- Correlation of system output with human ratings (Spearman)

Method	**Source**	**Spearman**
(Agirre et al. 2009)	Combination	**0.78**
(Gabrilovich and Markovitch, 2007)	Wikipedia	0.75
WordNet 3.0 + Knownets	WordNet	0.71
WordNet 3.0 + glosses	WordNet	0.68
(Agirre et al. 2009)	Corpora	0.66
(Finkelstein et al. 2007)	LSA	0.56
(Hughes and Ramage, 2007)	WordNet	0.55
(Jarmasz 2003)	WordNet	0.35

Unknown word (Maradona).

Outline

Knowledge-based WSD *(with Aitor Soroa, Oier Lopez de Lacalle)*

- 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
(with Aitor Soroa, Oier Lopez de Lacalle)

- 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 Lapata, 2007; Agirre et al. 2008)
  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 Lapata, 2007; Agirre et al. 2008)
  2. Use **Personalized PageRank** over the complete graph, initializing $\mathbf{v}$ with the context words

Using Personalized PageRank (PPPR 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)

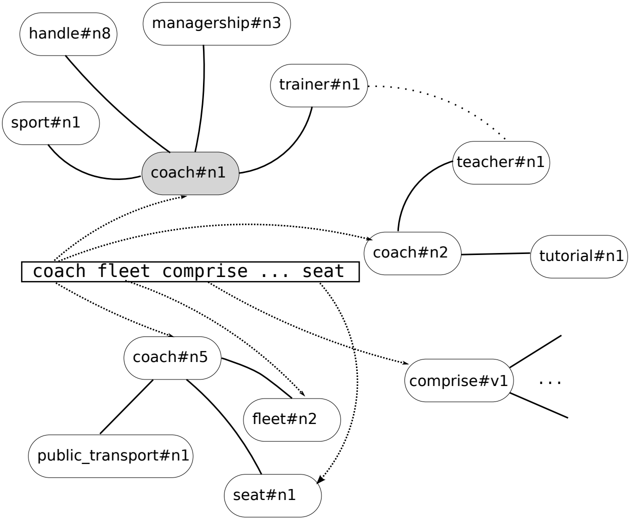
Using Personalized PageRank (PPPR and PPR_w2w)

- 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)

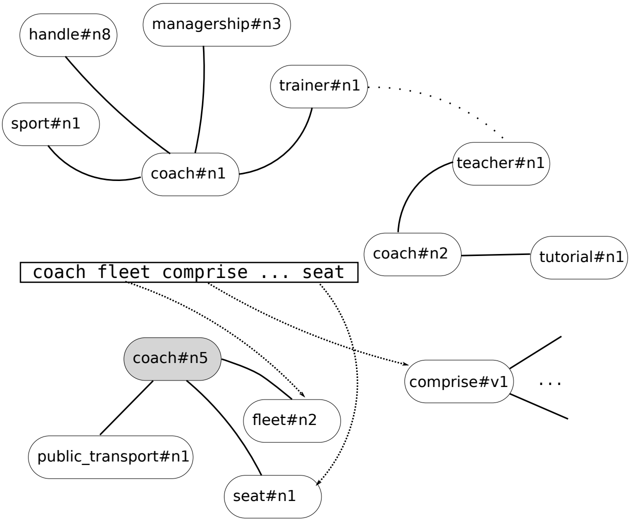
Using Personalized PageRank (PPPR and PPR_w2w)

- 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)

PPR


handle#n8
managership#n3
trainer#n1
sport#n1
coach#n1
teacher#n1
coach#n2
tutorial#n1
coach fleet comprise ... seat
coach#n5
comprise#v1
...
fleet#n2
public_transport#n1
seat#n1

PPR_w2w


handle#n8
managership#n3
trainer#n1
sport#n1
coach#n1
teacher#n1
coach#n2
tutorial#n1
coach fleet comprise ... seat
coach#n5
comprise#v1
...
fleet#n2
public_transport#n1
seat#n1

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.

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.

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

Methods

- How could we improve WSD performance without tagging new data from domain or adapting WordNet manually to the domain?
- 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
  - **coach: manager, captain, player, team, striker, . . .**

Methods

- How could we improve WSD performance without tagging new data from domain or adapting WordNet manually to the domain?
- 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
  - **coach: manager, captain, player, team, striker, ...**

Methods

- How could we improve WSD performance without tagging new data from domain or adapting WordNet manually to the domain?
- 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
  - **coach: manager, captain, player, team, striker, …**

Experiments

- Dataset with examples from **BNC**, **Sports** and **Finance** sections Reuters (Koeling et al. 2005)
  - 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
- Experiments
  - Supervised: train MFS, SVM, k-NN on SemCor examples
  - PageRank
  - Personalized PageRank (same damping factors, iterations)
    - Use context
    - 50 related words (Koeling et al. 2005) (BNC, Sports, Finance)

Experiments

- Dataset with examples from **BNC**, **Sports** and **Finance** sections Reuters (Koeling et al. 2005)
  - 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
- Experiments
  - Supervised: train MFS, SVM, k-NN on SemCor examples
  - PageRank
  - 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	PPR	**43.8**	*35.6	*46.9
Related words	PPR	*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

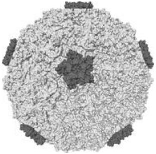
- Supervised (MFS, SVM, k-NN) very low (see test MFS)
- Static PageRank close to MFS
- PPR on context: best for BNC (* for statistical significance)
- PPR on related words: best for Sports and Finance and improves over Koeling *et al.*, who use pairwise WordNet similarity.

Outline

UMLS and biomedical text *(with Aitor Soroa and Mark Stevenson)*

- Ambiguities believed not to occur on specific domains
  - **On the Use of Cold Water as a Powerful Remedial Agent in Chronic Disease.**
  - **Intranasal ipratropium bromide for the common cold.**
- 11.7% of the phrases in abstracts added to MEDLINE in 1998 were ambiguous (Weeber et al. 2011)

- Unified Medical Language System (UMLS) Metathesaurus
- Concept Unique Identifiers (CUIs)
  - C0234192: Cold (Cold Sensation) [Physiologic Function]
  - C0009264: Cold (cold temperature) [Natural Phenomenon or Process]
  - C0009443: Cold (Common Cold) [Disease or Syndrome]



UMLS

- Thesaurus in Metathesaurus:
  Alcohol and other drugs, Medical Subject Headings, Crisp Thesaurus, SNOMED Clinical Terms, etc.
- Relations in the Metathesaurus between CUIs:
  parent, can be qualified by, related possibly sinonymous, related other
- We applied random walks over a graph of CUIs.
- Evaluated on NLM-WSD, 50 ambiguous terms (100 instances each)

KB	#CUIs	#relations	Acc.	Terms
AOD	15,901	58,998	51.5	4
MSH	278,297	1,098,547	44.7	9
CSP	16,703	73,200	60.2	3
SNOMEDCT	304,443	1,237,571	62.5	29
all above	572,105	2,433,324	64.4	48
all relations	-	5,352,190	68.1	50
combined with cooc.	-	-	**73.7**	50
(Jimeno and Aronson, 2011)	-	-	68.4	50

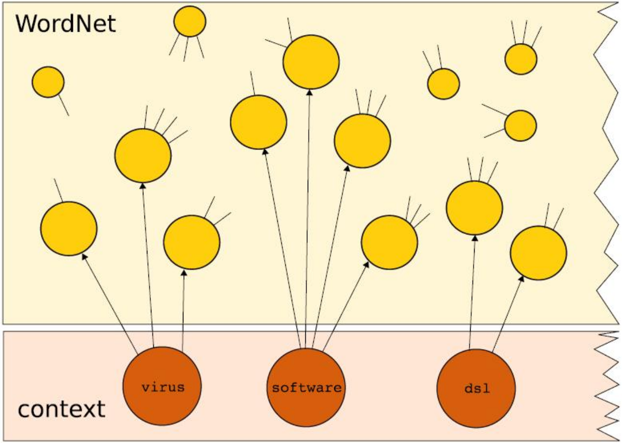
Outline

Similarity and Information Retrieval *(with Arantxa Otegi and Xabier Arregi)*

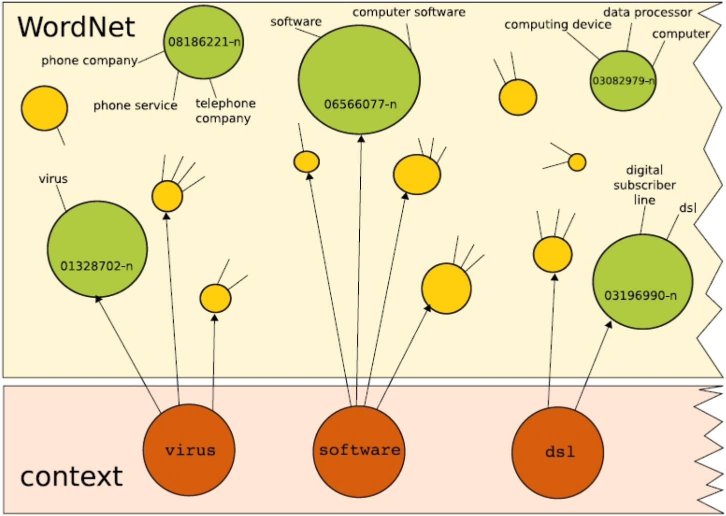
- Document expansion (aka clustering and smoothing) has been shown to be successful in ad-hoc IR
- **Use WordNet and similarity to expand documents**
- Example:
  - I can't **install DSL** because of the **antivirus program**, any hints?
  - You should turn off **virus** and anti-spy software. And thats done within each of the **softwares** themselves. Then turn them back on later after **setting up** any **DSL softwares**.
- Method:
  - Initialize random walk with document words
  - Retrieve top k synsets
  - Introduce words on those k synsets in a secondary index
  - When retrieving, use both primary and secondary indexes

Example

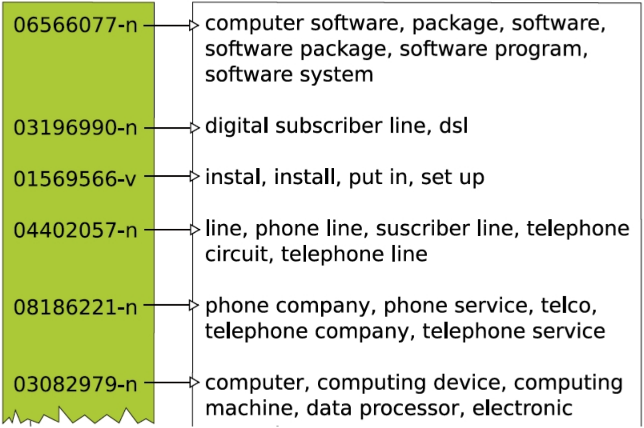
You should turn off virus and anti-spy software. And thats done within each of the softwares themselves. Then turn them back on later after setting up any DSL softwares.



Example


WordNet
08186221-n
phone company
phone service
telephone company
software
computer software
06566077-n
data processor
computing device
computer
03082979-n
virus
01328702-n
digital subscriber line
dsl
03196990-n
context
virus
software
dsl

Example



I can't install DSL because of the antivirus program, any hints?

Experiments

- BM25 ranking function
- Combine 2 indexes: original words and expansion terms
- Parameters: k_1, b (BM25) λ (indices) k (concepts in expansion)
- Three collections:
  - Robust at CLEF 2009
  - Yahoo Answer!
  - RespubliQA (IR for QA)
- Summary of results:
  - Default parameters: 1.43% - 4.90% improvement in all 3 datasets
  - Optimized parameters: 0.98% - 2.20% improvement in 2 datasets
  - Carrying parameters: 5.77% - 19.77% improvement in 4 out of 6
- Robustness
- Particularly on short documents

Experiments

- BM25 ranking function
- Combine 2 indexes: original words and expansion terms
- Parameters: k_1, b (BM25) λ (indices) k (concepts in expansion)
- Three collections:
  - Robust at CLEF 2009
  - Yahoo Answer!
  - RespubliQA (IR for QA)
- Summary of results:
  - Default parameters: 1.43% - 4.90% improvement in all 3 datasets
  - Optimized parameters: 0.98% - 2.20% improvement in 2 datasets
  - Carrying parameters: 5.77% - 19.77% improvement in 4 out of 6
- Robustness
- Particularly on short documents

Experiments

- BM25 ranking function
- Combine 2 indexes: original words and expansion terms
- Parameters: k_1, b (BM25) λ (indices) k (concepts in expansion)
- Three collections:
  - Robust at CLEF 2009
  - Yahoo Answer!
  - RespubliQA (IR for QA)
- Summary of results:
  - Default parameters: 1.43% - 4.90% improvement in all 3 datasets
  - Optimized parameters: 0.98% - 2.20% improvement in 2 datasets
  - Carrying parameters: 5.77% - 19.77% improvement in 4 out of 6
- Robustness
- Particularly on short documents

Outline

Conclusions

- Knowledge-based method for similarity and WSD
- Based on random walks
- Exploits whole structure of underlying KB efficiently
- Performance:
  - Similarity: best KB algorithm, comparable with 1.6 Tword, slightly below ESA
  - WSD: Best KB algorithm S2AW, S3AW, Domains datasets
  - WSD and domains:
    - Better than supervised WSD when adapting to domains (Sports, Finance)
    - Best KB algorithm in Biomedical texts

Conclusions

- Knowledge-based method for similarity and WSD
- Based on random walks
- Exploits whole structure of underlying KB efficiently
- Performance:
  - Similarity: best KB algorithm, comparable with 1.6 Tword, slightly below ESA
  - WSD: Best KB algorithm S2AW, S3AW, Domains datasets
  - WSD and domains:
    - Better than supervised WSD when adapting to domains (Sports, Finance)
    - Best KB algorithm in Biomedical texts

Conclusions

- Useful in applications:
  - performance gains and robustness
- 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 (WordNet, UMLS)
  - Including program to construct graphs from new KB (e.g. Wikipedia)
  - GPL license, open source, free

Conclusions

- Useful in applications:
  - performance gains and robustness
- 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 (WordNet, UMLS)
  - Including program to construct graphs from new KB (e.g. Wikipedia)
  - GPL license, open source, free

Conclusions

- Useful in applications:
  - performance gains and robustness
- 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 (WordNet, UMLS)
  - Including program to construct graphs from new KB (e.g. Wikipedia)
  - GPL license, open source, free

Future work

- Similarity: moving to sentence similarity and document similarity
- Information Retrieval: other options to combine similarity information (IJCNLP 2011)
- Domains and WSD: interrelation between domains and WSD (CIKM 2011)

Knowledge-Based Word Sense Disambiguation and Similarity using Random Walks

Eneko Agirre
ixa2.si.ehu.es/eneko

University of the Basque Country
(Currently visiting at Stanford)

SRI, 2011