

Word Sense Disambiguation - I

Pawan Goyal

CSE, IIT Kharagpur

Week 8, Lecture 3

Word Sense Disambiguation (WSD)

Sense ambiguity

- Many words have several meanings or senses
- The meaning of **bass** depends on the context
- Are we talking about music, or fish?
 - ▶ An electric guitar and **bass** player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.
 - ▶ And it all started when fishermen decided the striped **bass** in Lake Mead were too skinny.

Word Sense Disambiguation (WSD)

Sense ambiguity

- Many words have several meanings or senses
- The meaning of **bass** depends on the context
- Are we talking about music, or fish?
 - ▶ An electric guitar and **bass** player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.
 - ▶ And it all started when fishermen decided the striped **bass** in Lake Mead were too skinny.

Disambiguation

- The task of disambiguation is to determine which of the senses of an ambiguous word is invoked in a particular use of the word.

Word Sense Disambiguation (WSD)

Sense ambiguity

- Many words have several meanings or senses
- The meaning of **bass** depends on the context
- Are we talking about music, or fish?
 - ▶ An electric guitar and **bass** player stand off to one side, not really part of the scene, just as a sort of nod to gringo expectations perhaps.
 - ▶ And it all started when fishermen decided the striped **bass** in Lake Mead were too skinny.

Disambiguation

- The task of disambiguation is to determine which of the senses of an ambiguous word is invoked in a particular use of the word.
- This is done by looking at the context of the word's use.

- Knowledge Based Approaches
 - ▶ Overlap Based Approaches
- Machine Learning Based Approaches
 - ▶ Supervised Approaches
 - ▶ Semi-supervised Algorithms
 - ▶ Unsupervised Algorithms
- Hybrid Approaches

Overlap Based Approaches

- Require a **Machine Readable Dictionary** (MRD).
- Find the overlap between the features of different senses of an ambiguous word (**sense bag**) and the features of the words in its context (**context bag**).
- The features could be sense definitions, example sentences, hypernyms etc.
- The features could also be given weights.
- The sense which has the maximum overlap is selected as the contextually appropriate sense.

Lesk's Algorithm

Sense Bag: contains the words in the definition of a candidate sense of the ambiguous word.

Lesk's Algorithm

Sense Bag: contains the words in the definition of a candidate sense of the ambiguous word.

Context Bag: contains the words in the definition of each sense of each context word.

Lesk's Algorithm

Sense Bag: contains the words in the definition of a candidate sense of the ambiguous word.

Context Bag: contains the words in the definition of each sense of each context word.

*On burning **coal** we get **ash**.*

Sense Bag: contains the words in the definition of a candidate sense of the ambiguous word.

Context Bag: contains the words in the definition of each sense of each context word.

On burning **coal** we get **ash**.

Ash

- **Sense 1**
Trees of the olive family with pinnate leaves, thin furrowed bark and gray branches.
- **Sense 2**
The **solid** residue left when **combustible** material is thoroughly **burned** or oxidized.
- **Sense 3**
To convert into ash

Coal

- **Sense 1**
A piece of glowing carbon or **burnt** wood.
- **Sense 2**
charcoal.
- **Sense 3**
A black **solid combustible** substance formed by the partial decomposition of vegetable matter without free access to air and under the influence of moisture and often increased pressure and temperature that is widely used as a fuel for **burning**

In this case Sense 2 of ash would be the winner sense.

Walker's Algorithm

- A Thesaurus Based approach

Walker's Algorithm

- A Thesaurus Based approach
- **Step 1:** For each sense of the target word find the thesaurus category to which that sense belongs

Walker's Algorithm

- A Thesaurus Based approach
- **Step 1:** For each sense of the target word find the thesaurus category to which that sense belongs
- **Step 2:** Calculate the score for each sense by using the context words.

Walker's Algorithm

- A Thesaurus Based approach
- **Step 1:** For each sense of the target word find the thesaurus category to which that sense belongs
- **Step 2:** Calculate the score for each sense by using the context words.
A context word will add 1 to the score of the sense if the thesaurus category of the word matches that of the sense.

Walker's Algorithm

- A Thesaurus Based approach
- **Step 1:** For each sense of the target word find the thesaurus category to which that sense belongs
- **Step 2:** Calculate the score for each sense by using the context words.
A context word will add 1 to the score of the sense if the thesaurus category of the word matches that of the sense.
 - ▶ E.g. The money in this bank fetches an interest of 8% per annum

Walker's Algorithm

- A Thesaurus Based approach
- **Step 1:** For each sense of the target word find the thesaurus category to which that sense belongs
- **Step 2:** Calculate the score for each sense by using the context words.
A context word will add 1 to the score of the sense if the thesaurus category of the word matches that of the sense.
 - ▶ E.g. The money in this bank fetches an interest of 8% per annum
 - ▶ Target word: *bank*
 - ▶ Clue words from the context: *money, interest, annum, fetch*

Walker's Algorithm

- A Thesaurus Based approach
- **Step 1:** For each sense of the target word find the thesaurus category to which that sense belongs
- **Step 2:** Calculate the score for each sense by using the context words.
A context word will add 1 to the score of the sense if the thesaurus category of the word matches that of the sense.
 - ▶ E.g. The money in this bank fetches an interest of 8% per annum
 - ▶ Target word: *bank*
 - ▶ Clue words from the context: *money, interest, annum, fetch*

	Sense1: Finance	Sense2: Location
Money	+1	0
Interest	+1	0
Fetch	0	0
Annum	+1	0
Total	3	0

Context words add 1 to the sense when the topic of the word matches that of the sense

WSD Using Random Walk Algorithm

The church bells no longer rung on Sundays.

church

- 1: one of the groups of Christians who have their own beliefs and forms of worship
- 2: a place for public (especially Christian) worship
- 3: a service conducted in a church

bell

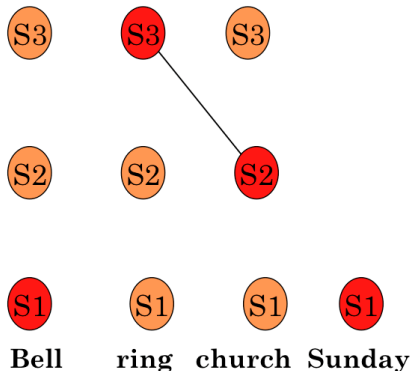
- 1: a hollow device made of metal that makes a ringing sound when struck
- 2: a push button at an outer door that gives a ringing or buzzing signal when pushed
- 3: the sound of a bell

ring

- 1: make a ringing sound
- 2: ring or echo with sound
- 3: make (bells) ring, often for the purposes of musical edification

Sunday

- 1: first day of the week; observed as a day of rest and worship by most Christians
-



Step 1: Add a vertex for each possible sense of each word in the text.

WSD Using Random Walk Algorithm

The church bells no longer rung on Sundays.

church

- 1: one of the groups of Christians who have their own beliefs and forms of worship
- 2: a place for public (especially Christian) worship
- 3: a service conducted in a church

bell

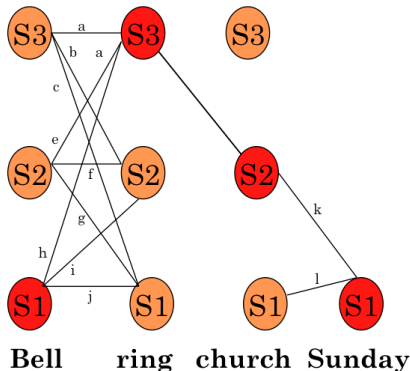
- 1: a hollow device made of metal that makes a ringing sound when struck
- 2: a push button at an outer door that gives a ringing or buzzing signal when pushed
- 3: the sound of a bell

ring

- 1: make a ringing sound
- 2: ring or echo with sound
- 3: make (bells) ring, often for the purposes of musical edification

Sunday

- 1: first day of the week; observed as a day of rest and worship by most Christians



Step 2: Add weighted edges using definition based semantic similarity (Lesk's method).

WSD Using Random Walk Algorithm

The church bells no longer rung on Sundays.

church

- 1: one of the groups of Christians who have their own beliefs and forms of worship
- 2: a place for public (especially Christian) worship
- 3: a service conducted in a church

bell

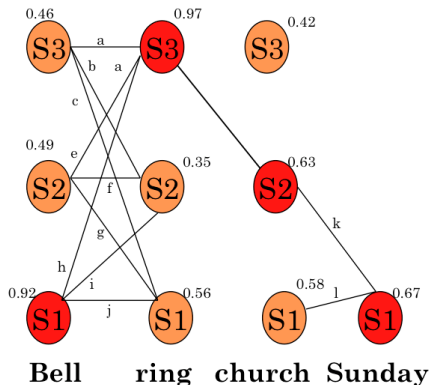
- 1: a hollow device made of metal that makes a ringing sound when struck
- 2: a push button at an outer door that gives a ringing or buzzing signal when pushed
- 3: the sound of a bell

ring

- 1: make a ringing sound
- 2: ring or echo with sound
- 3: make (bells) ring, often for the purposes of musical edification

Sunday

- 1: first day of the week; observed as a day of rest and worship by most Christians



Step 3: Apply graph based ranking algorithm to find score of each vertex (i.e. for each word sense).

WSD Using Random Walk Algorithm

The church bells no longer rung on Sundays.

church

- 1: one of the groups of Christians who have their own beliefs and forms of worship
- 2: a place for public (especially Christian) worship
- 3: a service conducted in a church

bell

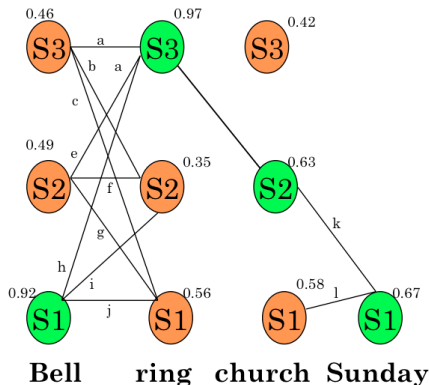
- 1: a hollow device made of metal that makes a ringing sound when struck
- 2: a push button at an outer door that gives a ringing or buzzing signal when pushed
- 3: the sound of a bell

ring

- 1: make a ringing sound
- 2: ring or echo with sound
- 3: make (bells) ring, often for the purposes of musical edification

Sunday

- 1: first day of the week; observed as a day of rest and worship by most Christians



Step 4: Select the vertex (sense) which has the highest score.

Naïve Bayes for WSD

- A Naïve Bayes classifier chooses the most likely sense for a word given the features of the context:

$$\hat{s} = \arg \max_{s \in S} P(s|f)$$

- A Naïve Bayes classifier chooses the most likely sense for a word given the features of the context:

$$\hat{s} = \arg \max_{s \in S} P(s|f)$$

- Using Bayes' law, this can be expressed as:

$$\begin{aligned}\hat{s} &= \arg \max_{s \in S} \frac{P(s)P(f|s)}{P(f)} \\ &= \arg \max_{s \in S} P(s)P(f|s)\end{aligned}$$

Naïve Bayes for WSD

- A Naïve Bayes classifier chooses the most likely sense for a word given the features of the context:

$$\hat{s} = \arg \max_{s \in S} P(s|f)$$

- Using Bayes' law, this can be expressed as:

$$\begin{aligned}\hat{s} &= \arg \max_{s \in S} \frac{P(s)P(f|s)}{P(f)} \\ &= \arg \max_{s \in S} P(s)P(f|s)\end{aligned}$$

- The 'Naïve' assumption: all the features are conditionally independent, given the sense':

$$\hat{s} = \arg \max_{s \in S} P(s) \prod_{j=1}^n P(f_j|s)$$

Training for Naïve Bayes

- ' f ' is a feature vector consisting of:
 - ▶ POS of w
 - ▶ Semantic and Syntactic features of w
 - ▶ Collocation vector (set of words around it) \rightarrow next word (+1), +2, -1, -2 and their POS's
 - ▶ Co-occurrence vector

Training for Naïve Bayes

- ' f ' is a feature vector consisting of:
 - ▶ POS of w
 - ▶ Semantic and Syntactic features of w
 - ▶ Collocation vector (set of words around it) \rightarrow next word (+1), +2, -1, -2 and their POS's
 - ▶ Co-occurrence vector
- Set parameters of Naïve Bayes using maximum likelihood estimation (MLE) from training data

$$P(s_i) = \frac{\text{count}(s_i, w_j)}{\text{count}(w_j)}$$

$$P(f_j|s_i) = \frac{\text{count}(f_j, s_i)}{\text{count}(s_i)}$$

Decision List Algorithm

- Based on ‘One sense per collocation’ property
 - ▶ Nearby words provide strong and consistent clues as to the sense of a target word
- Collect a large set of collocations for the ambiguous word
- Calculate word-sense probability distributions for all such collocations

Decision List Algorithm

- Based on 'One sense per collocation' property
 - ▶ Nearby words provide strong and consistent clues as to the sense of a target word
- Collect a large set of collocations for the ambiguous word
- Calculate word-sense probability distributions for all such collocations
- Calculate the log-likelihood ratio

$$\log\left(\frac{P(\text{Sense} - A | \text{Collocation}_i)}{P(\text{Sense} - B | \text{Collocation}_i)}\right)$$

- Higher log-likelihood \Rightarrow more predictive evidence

Decision List Algorithm

- Based on ‘One sense per collocation’ property
 - ▶ Nearby words provide strong and consistent clues as to the sense of a target word
- Collect a large set of collocations for the ambiguous word
- Calculate word-sense probability distributions for all such collocations
- Calculate the log-likelihood ratio

$$\log\left(\frac{P(\text{Sense} - A | \text{Collocation}_i)}{P(\text{Sense} - B | \text{Collocation}_i)}\right)$$

- Higher log-likelihood \Rightarrow more predictive evidence
- Collocations are ordered in a decision list, with most predictive collocations ranked highest

Decision List Algorithm

Training Data

Sense	Training Examples (Keyword in Context)
A	used to strain microscopic <i>plant</i> life from the ...
A	... zonal distribution of <i>plant</i> life
A	close-up studies of <i>plant</i> life and natural ...
A	too rapid growth of aquatic <i>plant</i> life in water ...
A	... the proliferation of <i>plant</i> and animal life ...
A	establishment phase of the <i>plant</i> virus life cycle ...
A	...
B	...
B	computer manufacturing <i>plant</i> and adjacent ...
B	discovered at a St. Louis <i>plant</i> manufacturing ...
B	... copper manufacturing <i>plant</i> found that they
B	copper wire manufacturing <i>plant</i> , for example ...
B	's cement manufacturing <i>plant</i> in Alpena ...
B	polystyrene manufacturing <i>plant</i> at its Dow ...
B	company manufacturing <i>plant</i> is in Orlando ...

Resultant Decision List

Final decision list for <i>plant</i> (abbreviated)		
LogL	Collocation	Sense
10.12	<i>plant</i> growth	⇒ A
9.68	car (within $\pm k$ words)	⇒ B
9.64	<i>plant</i> height	⇒ A
9.61	union (within $\pm k$ words)	⇒ B
9.54	equipment (within $\pm k$ words)	⇒ B
9.51	assembly <i>plant</i>	⇒ B
9.50	nuclear <i>plant</i>	⇒ B
9.31	flower (within $\pm k$ words)	⇒ A
9.24	job (within $\pm k$ words)	⇒ B
9.03	fruit (within $\pm k$ words)	⇒ A
9.02	<i>plant</i> species	⇒ A
...

Decision List Algorithm

Training Data

Sense	Training Examples (Keyword in Context)
A	used to strain microscopic <i>plant</i> life from the ...
A	... zonal distribution of <i>plant</i> life
A	close-up studies of <i>plant</i> life and natural ...
A	too rapid growth of aquatic <i>plant</i> life in water ...
A	... the proliferation of <i>plant</i> and animal life ...
A	establishment phase of the <i>plant</i> virus life cycle ...
A	...
B	...
B	computer manufacturing <i>plant</i> and adjacent ...
B	discovered at a St. Louis <i>plant</i> manufacturing ...
B	... copper manufacturing <i>plant</i> found that they ...
B	copper wire manufacturing <i>plant</i> , for example ...
B	's cement manufacturing <i>plant</i> in Alpena ...
B	polystyrene manufacturing <i>plant</i> at its Dow ...
B	company manufacturing <i>plant</i> is in Orlando ...

Resultant Decision List

Final decision list for <i>plant</i> (abbreviated)		
LogL	Collocation	Sense
10.12	<i>plant</i> growth	⇒ A
9.68	car (within $\pm k$ words)	⇒ B
9.64	<i>plant</i> height	⇒ A
9.61	union (within $\pm k$ words)	⇒ B
9.54	equipment (within $\pm k$ words)	⇒ B
9.51	assembly <i>plant</i>	⇒ B
9.50	nuclear <i>plant</i>	⇒ B
9.31	flower (within $\pm k$ words)	⇒ A
9.24	job (within $\pm k$ words)	⇒ B
9.03	fruit (within $\pm k$ words)	⇒ A
9.02	<i>plant</i> species	⇒ A
...	...	

Classification of a test sentence is based on the highest ranking collocation, found in the test sentences.

plucking flowers affects plant growth.

Decision List: Example

Example: discriminating between bass (fish) and bass (music):

Context	Sense
<i>fish in $\pm k$ words</i>	FISH
<i>striped bass</i>	FISH
<i>guitar in $\pm k$ words</i>	MUSIC
<i>bass player</i>	MUSIC
<i>piano in $\pm k$ words</i>	MUSIC
<i>sea bass</i>	FISH
<i>play bass</i>	MUSIC
<i>river in $\pm k$ words</i>	FISH
<i>on bass</i>	MUSIC
<i>bass are</i>	FISH

