Week 3 (Cont.)
Word Sense Disambiguation (WSD)

Nhung Nguyen
slides courtesy of NaCTeM

Recap

- Two approaches were introduced for learning word embeddings
  - count-based, and
  - prediction-based
- A very brief introduction to neural networks, and how they are used to learn word embeddings
- word2vec

Word senses

- We assumed each word has one meaning / sense
- In fact, a word can have several senses

mouse^{1} : .... a *mouse* controlling a computer system in 1968.
mouse^{2} : .... a quiet animal like a *mouse*
bank^{1} : ...a *bank* can hold the investments in a custodial account ...
bank^{2} : ...as agriculture burgeons on the east *bank*, the river ...

- How to define word senses? WordNet
- How to predict word senses? Knowledge-based and supervised learning-based methods

WordNet: A Database of Lexical Relations

The noun "bass" has 8 senses in WordNet.

1. bass^{1} - (the lowest part of the musical range)
2. bass^{2}, bass part^{1} - (the lowest part in polyphonic music)
3. bass^{3}, basso^{1} - (an adult male singer with the lowest voice)
4. sea bass^{1}, bass^{4} - (the lean flesh of a saltwater fish of the family Serranidae)
5. freshwater bass^{1}, bass^{5} - (any of various North American freshwater fish with lean flesh (especially of the genus Micropterus))
6. bass^{6}, bass voice^{1}, basso^{2} - (the lowest adult male singing voice)
7. bass^{7} - (the member with the lowest range of a family of musical instruments)
8. bass^{8} - (nontechnical name for any of numerous edible marine and freshwater spiny-finned fishes)

WordNet (cont.) - supersenses

Category	Example	Category	Example	Category	Example
ACT	*service*	GROUP	*place*	PLANT	*tree*
ANIMAL	*dog*	LOCATION	*area*	POSSESSION	*price*
ARTIFACT	*car*	MOTIVE	*reason*	PROCESS	*process*
ATTRIBUTE	*quality*	NATURAL EVENT	*experience*	QUANTITY	*amount*
BODY	*hair*	NATURAL OBJECT	*flower*	RELATION	*portion*
COGNITION	*way*	OTHER	*stuff*	SHAPE	*square*
COMMUNICATION	*review*	PERSON	*people*	STATE	*pain*
FEELING	*discomfort*	PHENOMENON	*result*	SUBSTANCE	*oil*
FOOD	*food*			TIME	*day*

WordNet (cont.) - sense relations

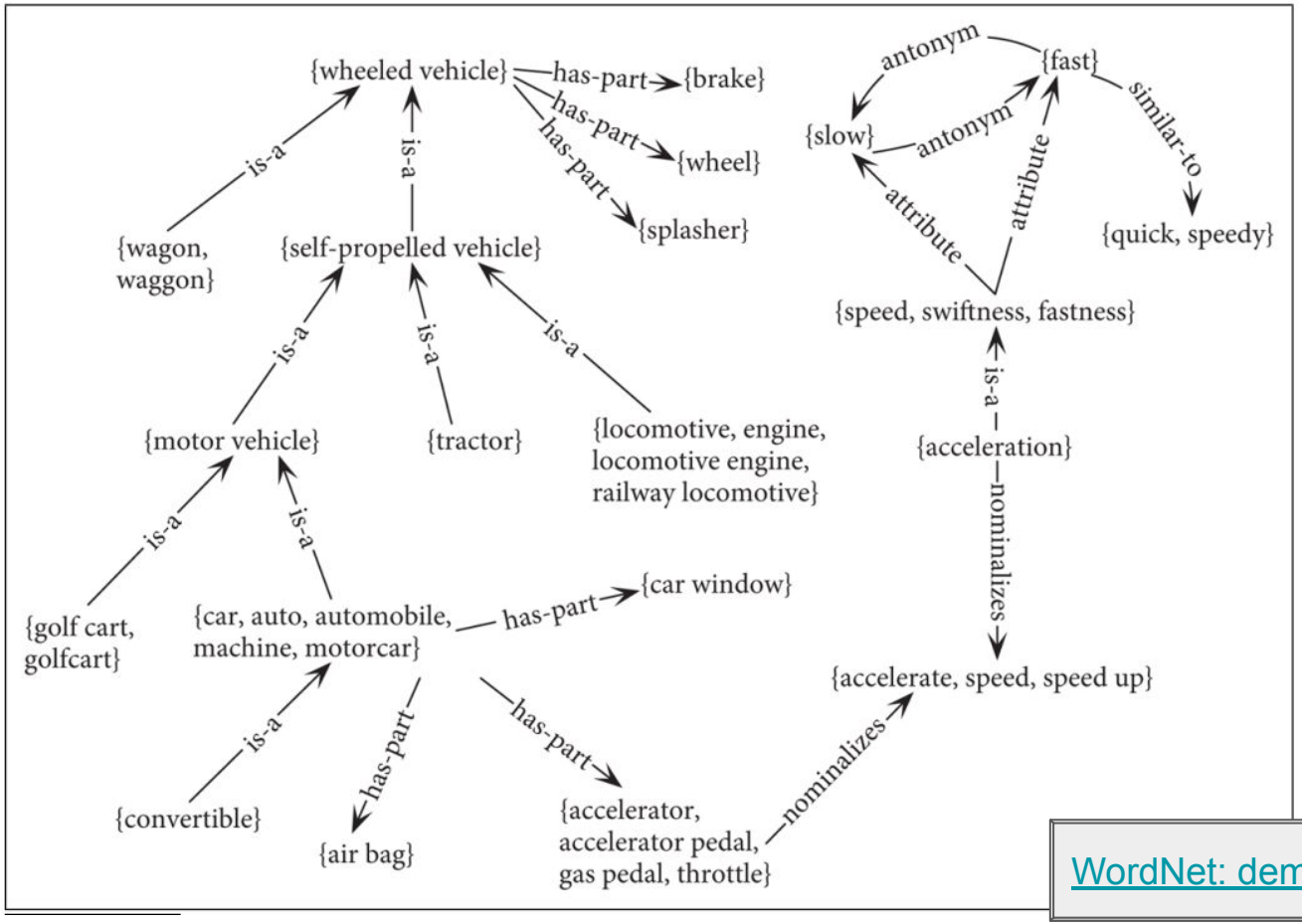
Relation	Also Called	Definition	Example
Hypernym	Superordinate	From concepts to superordinates	$breakfast^1 \rightarrow meal^1$
Hyponym	Subordinate	From concepts to subtypes	$meal^1 \rightarrow lunch^1$
Instance Hypernym	Instance	From instances to their concepts	$Austen^1 \rightarrow author^1$
Instance Hyponym	Has-Instance	From concepts to their instances	$composer^1 \rightarrow Bach^1$
Part Meronym	Has-Part	From wholes to parts	$table^2 \rightarrow leg^3$
Part Holonym	Part-Of	From parts to wholes	$course^7 \rightarrow meal^1$
Antonym		Semantic opposition between lemmas	$leader^1 \Longleftrightarrow follower^1$
Derivation		Lemmas w/same morphological root	$destruction^1 \Longleftrightarrow destroy^1$

Figure 19.3 Some of the noun relations in WordNet.

Relation	Definition	Example
Hypernym	From events to superordinate events	$fly^9 \rightarrow travel^5$
Troponym	From events to subordinate event	$walk^1 \rightarrow stroll^1$
Entails	From verbs (events) to the verbs (events) they entail	$snore^1 \rightarrow sleep^1$
Antonym	Semantic opposition between lemmas	$increase^1 \Longleftrightarrow decrease^1$

Figure 19.4 Some verb relations in WordNet.

WordNet (cont.) - view as a graph



[WordNet: demo]

Approaches to WSD

Task definition

- Input:
  - Word in a context
  - A fixed inventory of potential word senses
- Output:
  - The correct word sense in context

- Two main types:
  - Lexical sample task
  - All-words tasks: similar to part-of-speech tagging

Lesk algorithm (Lesk, 1986): WSD baseline

- Intuition: choose the sense whose dictionary gloss or definition shares the most words with the target word's neighborhood
- Formalisation
  - Target word: **w**
  - Context words of w: $\mathbf{w}_j$
  - Lexicon definition of senses: D(.)
  - Set of senses of a word: S(.)

Lesk algorithm (Cont.)

- The rule:

$$s_{optimised} = argmax_{s_k \in S(w)} (sim(D(s_k), \bigcup_{w_j \in C} \bigcup_{s_i \in S(w_j)} D(s_i)))$$

- Possible similarity measures

$$sim(X, Y) = 2\frac{|X \cap Y|}{|X| + |Y|}$$

$$sim(X, Y) = 2\frac{|X \cap Y|}{|X \bigcup Y|}$$

$$sim(X, Y) = 2\frac{|X \cap Y|}{\sqrt{|X||Y|}}$$

$$sim(X, Y) = |X \cap Y|$$

Example

Input sentence: *Waves were hitting the steep* ***bank****.*

Senses for bank:

1. sloping land (especially the slope beside a body of water)
2. a financial institution that accepts deposits and moves the money into lending activities
3. a building in which the business of banking transacted

Context definitions:

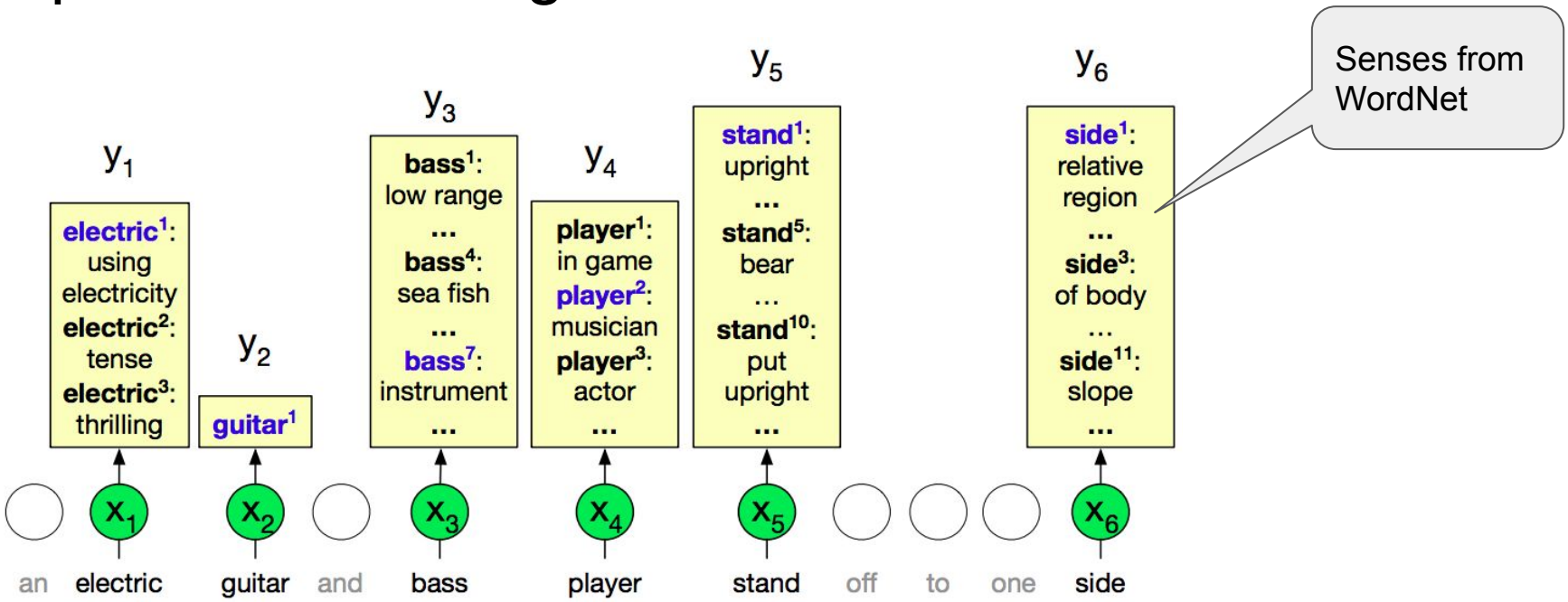
- wave – one of a series of ridges that moves across the surface of a liquid (especially across a large body of water)
- hit – hit against; come into sudden contact with
- steep – of a slope; set at a high angle

sim_1 = |body, water, slope| = 3, sim_2 = |moves| = 1, sim_3 = |∅| = 0

Lesk algorithm (Cont.)

- Pros:
  - Simple to implement
  - No training data needed
- Cons:
  - not all words have definitions in WordNet,
  - need to deal with ambiguous context words,
  - poor performance.

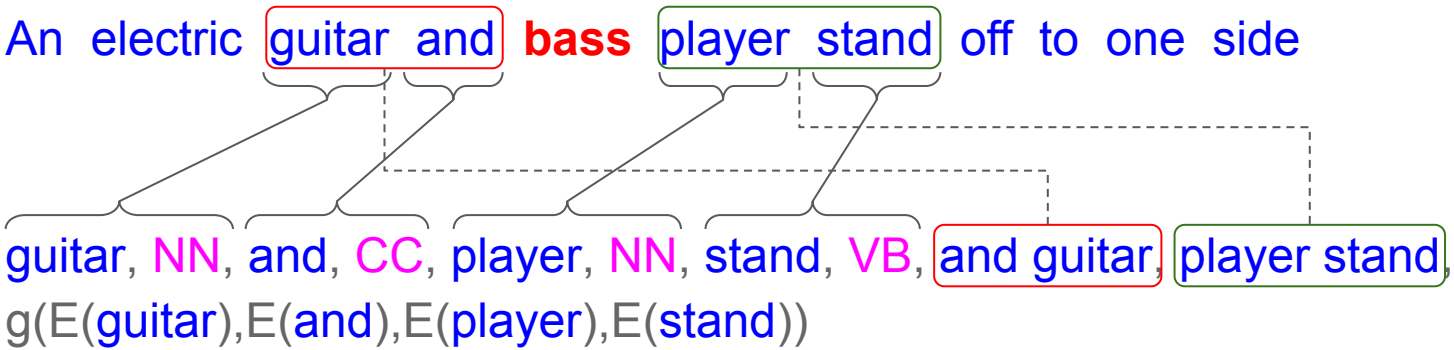
Supervised learning: all-word WSD



- SemCor corpus (https://www.kaggle.com/nltkdata/semcor-corpus)

Feature-based WSD

- Using SVM with several features
  - Part-of-speech tags
  - Collocation features of words and *n*-grams
  - Weighted average of word embeddings (of all words in a window)



Summary

- WordNet is a large database of lexical relations for English
- WSD is the task of determining the correct sense of a word in context.
- SemCor is the largest corpus with WordNet-labeled senses
- Lesk algorithm, a WSD baseline, is a knowledge-based approach
- Feature-based algorithms using parts of speech and embeddings of words in the context of the target word work well.