Lexical Semantics

Katharina Kann — CSCI/LING5832

The Big Question

What kind of thing is the meaning of a word?

The Big Question

~~What kind of thing is the meaning of a word?~~

What can you do with a word if you know its meaning?

The Big Question

Some answers:

- Pick out things or events in the world that it refers to.
- Combine it with other words to get phrase meanings.
- Work out what other kinds of words it *can* combine with.
- Measure its similarity to other words.
- Evaluate *analogy* relationships with other words.
- Find out *features* of the word (sometimes, if you know what to look for):
    - Part of speech
    - Sentiment
    - …

How Many Words?

Once when I was six years old I saw a magnificent picture in a book, …

- 15?
- 13?
- 16?

How Many Words?

Once when I was six years old I saw a magnificent picture in a book, …

- **Type:** an entry in a vocabulary
- **Token:** an instance of a type in some text

How Many Words?

Once when I was six years old I saw a magnificent picture in a book, …

- **Type:** an entry in a vocabulary
- **Token:** an instance of a type in some text

- Would you expect *years* to show up in a dictionary?

Definitions

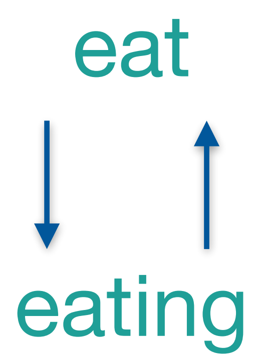
- **Lemma** or **citation form**: the form of a word you can find in a dictionary
    - Example: *year* (or *run*, *play*, *chair*, etc.)

Definitions

- **Lemma** or **citation form**: the form of a word you can find in a dictionary
  - Example: *year* (or *run*, *play*, *chair*, etc.)
- **Wordform** or **inflected form**: surface form of a word which is (potentially) different from the lemma and depends on the context
  - Example: *years* (or *running*, *runs*, etc.)

Definitions

- Common NLP tasks:
  - Lemmatization
  - Morphological inflection



English Morphology

- Morphology is the study of the ways that words are built up from smaller units called morphemes.

English Morphology

- Morphology is the study of the ways that words are built up from smaller units called morphemes.
- Two broad classes:
  - Inflectional
  - Derivational

Inflectional Morphology

- The resulting word has the same word class as the original
- And serves a grammatical/semantic purpose that is different from the original

Inflectional Morphology

- The resulting word has the same word class as the original
- And serves a grammatical/semantic purpose that is different from the original
- But it is *transparently* related to the original
  - “walk” + “s” = “walks”
  - “cat” + “s” = “cats”

Inflectional Morphology: English

- Nouns are simple
- Verbs are only slightly more complex
  - (How many forms per lemma?)
- Other languages can be quite a lot more complex

Regulars and Irregulars

- Some words don’t follow the rules
    - *mouse/mice*
    - *go/went, catch/caught*
- The terms **regular** and **irregular** are used to refer to words that follow the rules and those that don’t.

Derivational Morphology

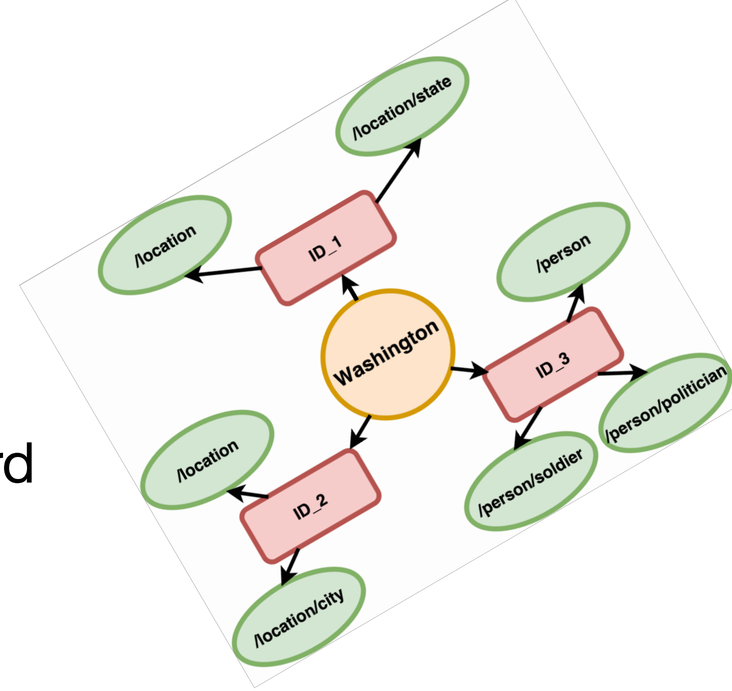
- The resulting word does **not** need to have the same word class as the original.
- English examples:
  - *friend/friendly, mother/motherly*
  - *happy/happily*
  - *employ/employee, employ/employer*

Derivational Morphology

- Not all possible operations are equally good (even if they might be understood!):
  - *clue/clueless?*
  - *clue/clueful?*
  - *clue/cluely?*

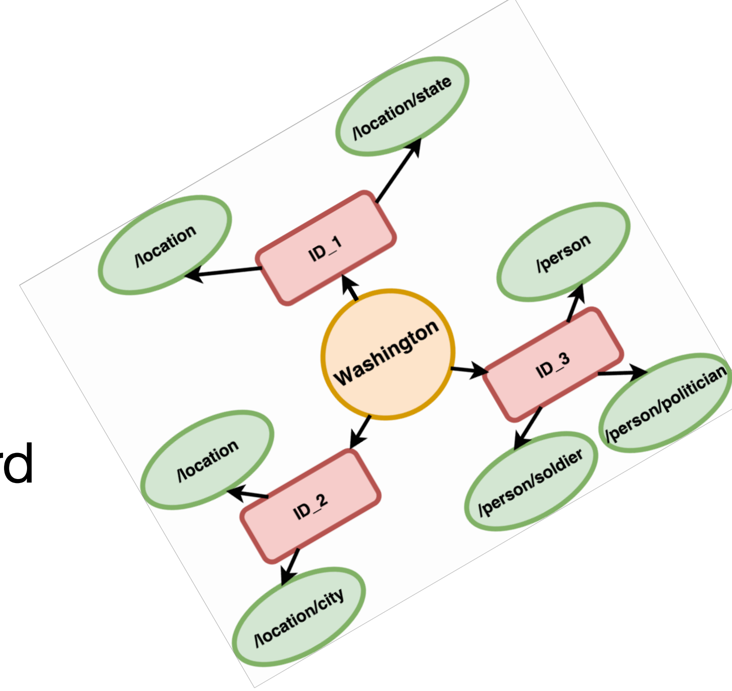
Terminology

Definitions



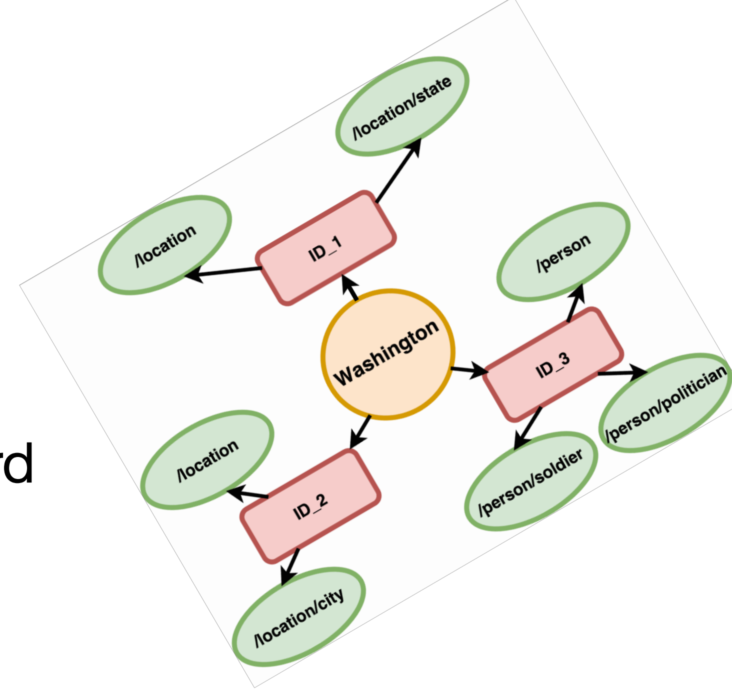
- **Word sense**: a certain meaning of a word
  - Many words have multiple senses

Definitions



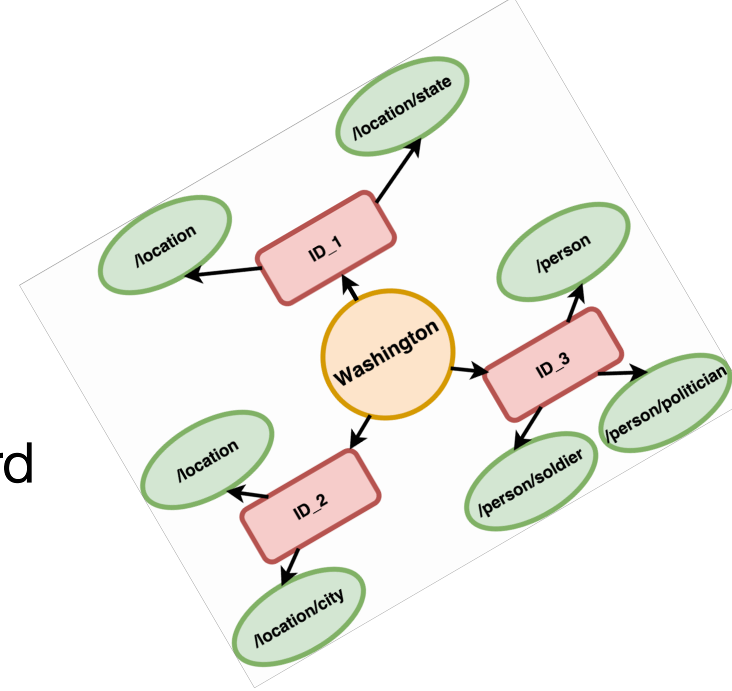
- **Word sense**: a certain meaning of a word
  - Many words have multiple senses
- **Polysemy:** senses are related
  - door ("Open the door!")
  - door ("He walked through the door.")
  - man (male human)
  - man (adult male)

Definitions



- **Word sense**: a certain meaning of a word
  - Many words have multiple senses
- **Polysemy:** senses are related
  - door ("Open the door!")
  - door ("He walked through the door.")
  - man (male human)
  - man (adult male)
- **Homonymy**: senses are unrelated
  - bank (financial institution)
  - bank (of a river)

Definitions



- **Word sense**: a certain meaning of a word
  - Many words have multiple senses
- **Polysemy:** senses are related
  - door ("Open the door!")
  - door ("He walked through the door.")
  - man (male human)
  - man (adult male)
- **Homonymy**: senses are unrelated
  - bank (financial institution)
  - bank (of a river)
- **Metonymy**: use of one aspect of a concept to refer to another
  - Example: Jane Austen, White House

Definitions

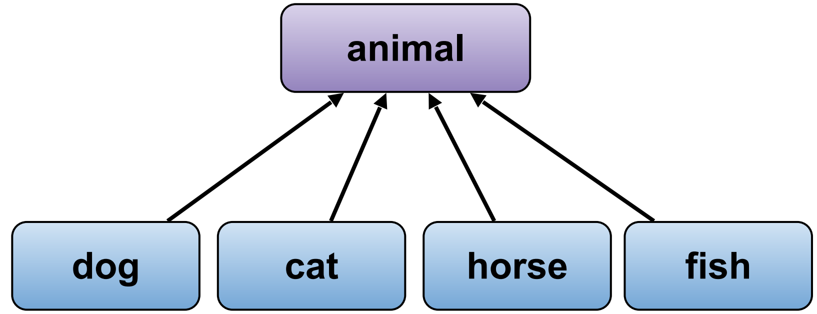
- Test if senses are related: construct a sentence which combines the two possible senses!
  - Air Canada serves breakfast.
  - Air Canada serves Montreal.
- Common NLP task: word sense disambiguation

Definitions

- **Synonyms**: Words which have the same meaning
  - Usually only in certain contexts! (Why?)
  - Example: water/H2O, hard/difficult

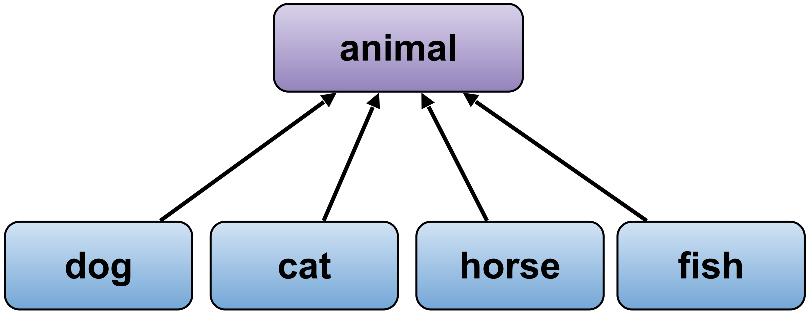
- **Antonyms**: Words with opposite meanings
  - Example: hot/cold, good/bad

Definitions



- **Hyponym**: word that denotes a more specific subclass of its hypernym
- **Hypernym**: word that denotes a superclass of its hyponym

Definitions



- **Hyponym**: word that denotes a more specific subclass of its hypernym
- **Hypernym**: word that denotes a superclass of its hyponym
- Examples:
    - Dog is a hyponym of animal
    - Pasta is a hyponym of food
    - Location is a hypernym of city

WordNet

WordNet

- A lexical database of English
- Contains:
  - Nouns
  - Verbs
  - Adjectives
  - Adverbs

Structure

- Most important relation in WordNet is synonymy
- Entries are pairs of words and senses
- Synonyms are grouped into *synsets*
    - Sets of synonyms
- Each synset is linked to other synsets by means of a small number of “conceptual relations”
- Synsets contain a brief definition and one or more short sentences illustrating the use of the synset members
- Each word-meaning pair in WordNet in unique

Relations

- Most frequent relation: hyponymy (or IS-A relation)
  - Examples:
    - bunk bed is a bed is a furniture
    - mango is a fruit is a food
- Instances are terminal nodes in their hierarchies
- Root node for nouns is {entity}

Relations

- Verb synsets are also arranged into hierarchies
- Verbs towards the bottom of the trees express increasingly specific manners characterizing an event
  - Example: communicate to talk to whisper
- Adjectives are organized in terms of antonyms

WordNet Search - 3.1

- WordNet home page - Glossary - Help

Word to search for: cat Search WordNet

Display Options: (Select option to change) Change

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **cat**, true cat (feline mammal usually having thick soft fur and no ability to roar: domestic cats; wildcats)
- S: (n) guy, **cat**, hombre, bozo, sod (an informal term for a youth or man) *"a nice guy"; "the guy's only doing it for some doll"; "the poor sod couldn't even buy a drink"*
- S: (n) **cat** (a spiteful woman gossip) *"what a cat she is!"*
- S: (n) kat, khat, qat, quat, **cat**, Arabian tea, African tea (the leaves of the shrub Catha edulis which are chewed like tobacco or used to make tea; has the effect of a euphoric stimulant) *"in Yemen kat is used daily by 85% of adults"*
- S: (n) cat-o'-nine-tails, **cat** (a whip with nine knotted cords) *"British sailors feared the cat"*
- S: (n) Caterpillar, **cat** (a large tracked vehicle that is propelled by two endless metal belts; frequently used for moving earth in construction and farm work)
- S: (n) big cat, **cat** (any of several large cats typically able to roar and living in the wild)
- S: (n) computerized tomography, computed tomography, CT, computerized axial tomography, computed axial tomography, **CAT** (a method of examining body organs by scanning them with X rays and using a computer to construct a series of cross-sectional scans along a single axis)

Verb

- S: (v) **cat** (beat with a cat-o'-nine-tails)
- S: (v) vomit, vomit up, purge, cast, sick, **cat**, be sick, disgorge, regorge, retch, puke, barf, spew, spue, chuck, upchuck, honk, regurgitate, throw up (eject the contents of the stomach through the mouth) *"After drinking too much, the students vomited"; "He purged continuously"; "The patient regurgitated the food we gave him last night"*

WordNet Search - 3.1
- WordNet home page - Glossary - Help

Word to search for: cat [Search WordNet]

Display Options: (Select option to change) [Change]

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **cat**, true cat (feline mammal usually having thick soft fur and no ability to roar: domestic cats; wildcats)
  - *direct hyponym* / *full hyponym*
  - *direct hypernym* / *inherited hypernym* / *sister term*
    - S: (n) feline, felid (any of various lithe-bodied roundheaded fissiped mammals, many with retractile claws)
      - *direct hyponym* / *full hyponym*
      - *part meronym*
      - *member holonym*
      - *direct hypernym* / *inherited hypernym* / *sister term*
        - S: (n) carnivore (a terrestrial or aquatic flesh-eating mammal) *"terrestrial carnivores have four or five clawed digits on each limb"*
          - *direct hyponym* / *full hyponym*
          - *member holonym*
          - *direct hypernym* / *inherited hypernym* / *sister term*
            - S: (n) placental, placental mammal, eutherian, eutherian mammal (mammals having a placenta; all mammals except monotremes and marsupials)
              - *direct hyponym* / *full hyponym*
              - *member holonym*
              - *direct hypernym* / *inherited hypernym* / *sister term*
                - S: (n) mammal, mammalian (any warm-blooded vertebrate having the skin more or less covered with hair; young are born alive except for the small subclass of monotremes and nourished with milk)
                  - *direct hyponym* / *full hyponym*
                  - *part meronym*
                  - *member holonym*
                  - *domain term category*
                  - *direct hypernym* / ***inherited hypernym*** / *sister term*
                    - S: (n) vertebrate, craniate (animals having a bony or cartilaginous skeleton

Click me

WordNet in Other Languages

- Open Multilingual Wordnet
- Over 20 languages:
  - Arabic
  - Chinese
  - Hebrew
  - Japanese
  - Thai
  - …

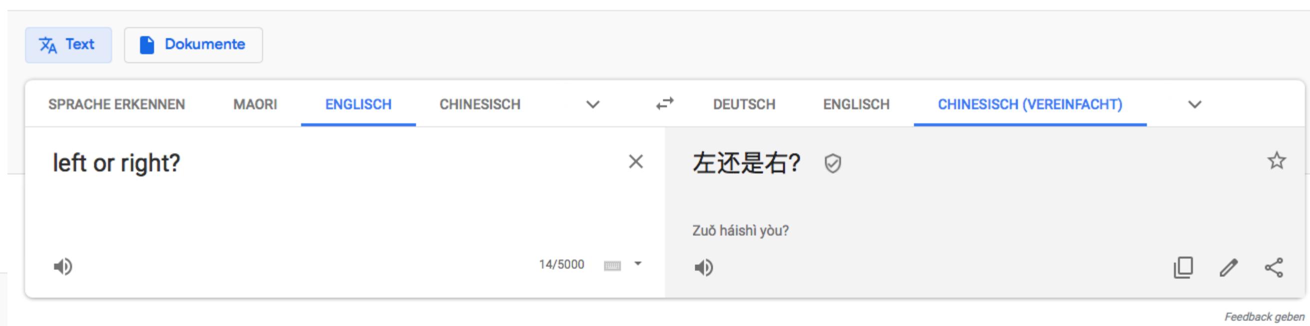
Word sense disambiguation

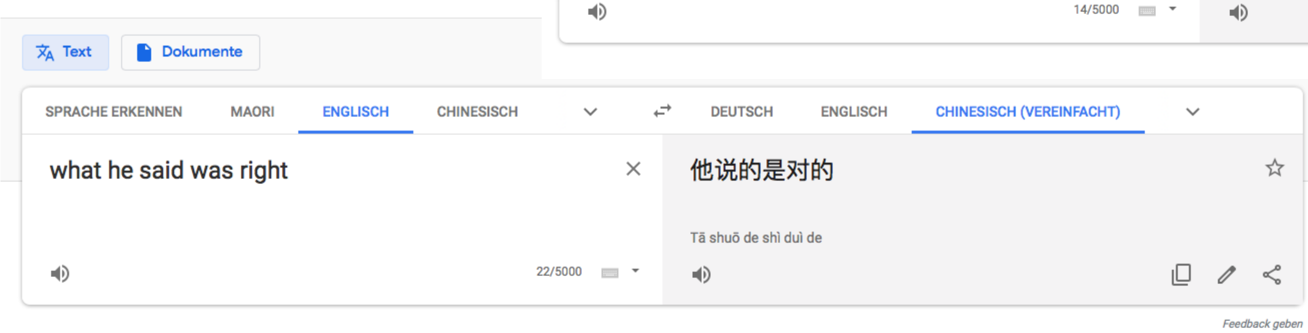
Task Definition

- Given:
  - A word in context
    - Task is *token-based*, not *type-based*!
  - A set of possible senses of that word, e.g., from WordNet
- Task: Find the sense of the word which is used in the given context!
- Examples:
  - Her company moved to **Washington**.
  - **Washington** became president in 1789.
- Similar task: entity typing

Useful for...

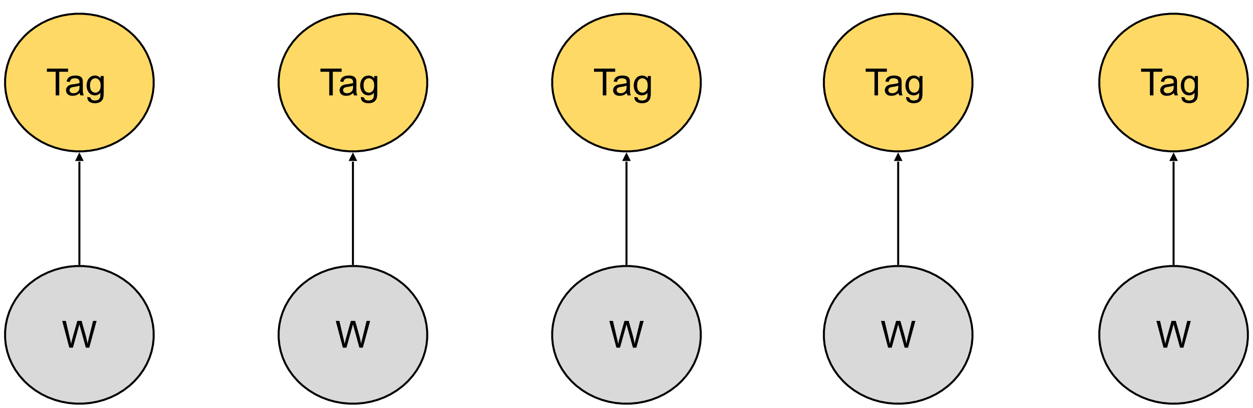
- Machine translation
- Sentiment analysis
- ...





Word Sense Disambiguation as Classification

- Requires a training corpus
  - All or some words tagged in context
- Requires a classifier



What Tags to Use?

WordNet Search - 3.1

- WordNet home page - Glossary - Help

Word to search for: cat [Search WordNet]

Display Options: (Select option to change) [Change]

Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations

Display options for sense: (gloss) "an example sentence"

Noun

- S: (n) **cat**, true cat (feline mammal usually having thick soft fur and no ability to roar: domestic cats; wildcats)
- S: (n) guy, **cat**, hombre, bozo, sod (an informal term for a youth or man) *"a nice guy"; "the guy's only doing it for some doll"; "the poor sod couldn't even buy a drink"*
- S: (n) **cat** (a spiteful woman gossip) *"what a cat she is!"*
- S: (n) kat, khat, qat, quat, **cat**, Arabian tea, African tea (the leaves of the shrub Catha edulis which are chewed like tobacco or used to make tea; has the effect of a euphoric stimulant) *"in Yemen kat is used daily by 85% of adults"*
- S: (n) cat-o'-nine-tails, **cat** (a whip with nine knotted cords) *"British sailors feared the cat"*
- S: (n) Caterpillar, **cat** (a large tracked vehicle that is propelled by two endless metal belts; frequently used for moving earth in construction and farm work)
- S: (n) big cat, **cat** (any of several large cats typically able to roar and living in the wild)
- S: (n) computerized tomography, computed tomography, CT, computerized axial tomography, computed axial tomography, **CAT** (a method of examining body organs by scanning them with X rays and using a computer to construct a series of cross-sectional scans along a single axis)

What Tags to Use?

Common tags:

- FISH
- LOCATION
- PERSON
- FOOD
- MUSIC
- CITY
- …

Possible Classifiers

Basically any!

- Logistic regression
- Neural network
- K-nearest neighbors
- Naïve Bayes
- SVMs
- …

A basic classifier: Naïve Bayes

Naïve Bayes

Given: context c, possible senses $s_i \in S$ for a word in c

Wanted: most likely sense s_{MAP}

$$s_{MAP} = \text{argmax}_{s \in S} P(s \mid c)$$

Naïve Bayes

Given: context c, possible senses $s_i \in S$ for a word in c

Wanted: most likely sense s_{MAP}

$$
\begin{aligned}
s_{MAP} &= \text{argmax}_{s \in S} P(s \mid c) \\
&= \text{argmax}_{s \in S} \frac{P(s, c)}{P(c)}
\end{aligned}
$$

Naïve Bayes

Given: context c, possible senses $s_i \in S$ for a word in c

Wanted: most likely sense s_{MAP}

$$
\begin{aligned}
s_{MAP} &= \text{argmax}_{s \in S} P(s \mid c) \\
&= \text{argmax}_{s \in S} \frac{P(s, c)}{P(c)} \\
&= \text{argmax}_{s \in S} \frac{P(c \mid s) P(s)}{P(c)}
\end{aligned}
$$

Naïve Bayes

Given: context c, possible senses $s_i \in S$ for a word in c

Wanted: most likely sense s_{MAP}

$$
\begin{aligned}
s_{MAP} &= \text{argmax}_{s \in S} P(s \mid c) \\
&= \text{argmax}_{s \in S} \frac{P(s, c)}{P(c)} \\
&= \text{argmax}_{s \in S} \frac{P(c \mid s) P(s)}{P(c)} \\
&= \text{argmax}_{s \in S} P(c \mid s) P(s)
\end{aligned}
$$

Naïve Bayes

Given: context c, possible senses $s_i \in S$ for a word in c

Wanted: most likely sense s_{MAP}

$$
\begin{aligned}
s_{MAP} &= \text{argmax}_{s \in S} P(s \mid c) \\
&= \text{argmax}_{s \in S} \frac{P(s, c)}{P(c)} \\
&= \text{argmax}_{s \in S} \frac{P(c \mid s) P(s)}{P(c)} \\
&= \text{argmax}_{s \in S} P(c \mid s) P(s)
\end{aligned}
$$

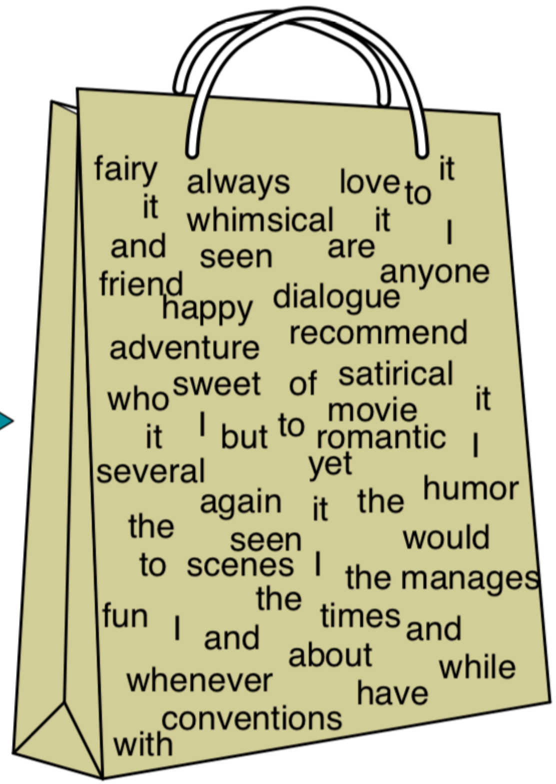
posterior

likelihood

prior

Bag of Words

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!



it	6
I	5
the	4
to	3
and	3
seen	2
yet	1
would	1
whimsical	1
times	1
sweet	1
satirical	1
adventure	1
genre	1
fairy	1
humor	1
have	1
great	1
…	…

J&M

Naïve Bayes

Given: context c, possible senses $s_i \in S$ for a word in c

Wanted: most likely sense s_{MAP}

$$
\begin{aligned}
s_{MAP} &= \text{argmax}_{s \in S} P(s \mid c) \\
&= \text{argmax}_{s \in S} \frac{P(s, c)}{P(c)} \\
&= \text{argmax}_{s \in S} \frac{P(c \mid s) P(s)}{P(c)} \\
&= \text{argmax}_{s \in S} P(c \mid s) P(s)
\end{aligned}
$$

Naïve Bayes

Given: context c, possible senses $s_i \in S$ for a word in c

Wanted: most likely sense s_{MAP}

$$
\begin{aligned}
s_{MAP} &= \text{argmax}_{s \in S} P(s \mid c) \\
&= \text{argmax}_{s \in S} \frac{P(s, c)}{P(c)} \\
&= \text{argmax}_{s \in S} \frac{P(c \mid s) P(s)}{P(c)} \\
&= \text{argmax}_{s \in S} P(c \mid s) P(s) \\
&= \text{argmax}_{s \in S} \prod_i P(x_i \mid s) P(s)
\end{aligned}
$$

Naïve Bayes assumption:

Features are independent

Naïve Bayes: “Training”

Simplest strategy: maximum likelihood estimation

- Use frequency in training data!

Naïve Bayes

Our “**training corpus**”:

...from Washington to…	=> CITY
…said Washington to…	=> PER
...leaving Washington in...	=> CITY

Naïve Bayes

Our “**training corpus**”:

...from Washington to… => CITY
…said Washington to… => PER
...leaving Washington in... => CITY

Our “**test example**”:

...leaving Washington to… => ???

Naïve Bayes

Our "**training corpus**":

...from Washington to… => CITY
…said Washington to… => PER
...leaving Washington in... => CITY

$$s_{MAP} = \text{argmax}_{s \in S} \prod_i P(x_i \mid s) P(s)$$

Our "**test example**":

...leaving Washington to… => ???

Naïve Bayes

Our “**training corpus**”:

...from Washington to… => CITY
…said Washington to… => PER
...leaving Washington in... => CITY

Our “**test example**”:

...leaving Washington to… => ???

$$s_{MAP} = \text{argmax}_{s \in S} \prod_i P(x_i \mid s) P(s)$$

$$P(CITY \mid leaving, to) = P(leaving \mid CITY) P(to \mid CITY) P(CITY)$$

Naïve Bayes

Our “**training corpus**”:

...from Washington to…   => CITY
…said Washington to…   => PER
...leaving Washington in... => CITY

Our “**test example**”:

...leaving Washington to… => ???

$$s_{MAP} = \text{argmax}_{s \in S} \prod_i P(x_i \mid s) P(s)$$

$$\begin{aligned} P(CITY \mid leaving, to) &= P(leaving \mid CITY) P(to \mid CITY) P(CITY) \\ &= 1/4 * 1/4 * 2/3 \end{aligned}$$

Naïve Bayes

Our “**training corpus**”:

...from Washington to… => CITY
…said Washington to… => PER
...leaving Washington in... => CITY

Our “**test example**”:

...leaving Washington to… => ???

$$s_{MAP} = \text{argmax}_{s \in S} \prod_i P(x_i \mid s) P(s)$$

$$\begin{aligned} P(CITY \mid leaving, to) &= P(leaving \mid CITY) P(to \mid CITY) P(CITY) \\ &= 1/4 * 1/4 * 2/3 \\ &= 2/48 = 1/24 \end{aligned}$$

Naïve Bayes

Our “**training corpus**”:

...from Washington to… => CITY
…said Washington to… => PER
...leaving Washington in... => CITY

Our “**test example**”:

...leaving Washington to… => ???

$$s_{MAP} = \text{argmax}_{s \in S} \prod_i P(x_i \mid s) P(s)$$

$$\begin{aligned} P(CITY \mid leaving, to) &= P(leaving \mid CITY) P(to \mid CITY) P(CITY) \\ &= 1/4 * 1/4 * 2/3 \\ &= 2/48 = 1/24 \end{aligned}$$

$$P(PER \mid leaving, to) = P(leaving \mid PER) P(to \mid PER) P(PER)$$

Naïve Bayes

Our “**training corpus**”:

...from Washington to… => CITY
…said Washington to… => PER
...leaving Washington in... => CITY

Our “**test example**”:

...leaving Washington to… => ???

$$s_{MAP} = \text{argmax}_{s \in S} \prod_i P(x_i \mid s) P(s)$$

$$\begin{aligned} P(CITY \mid leaving, to) &= P(leaving \mid CITY) P(to \mid CITY) P(CITY) \\ &= 1/4 * 1/4 * 2/3 \\ &= 2/48 = 1/24 \end{aligned}$$

$$\begin{aligned} P(PER \mid leaving, to) &= P(leaving \mid PER) P(to \mid PER) P(PER) \\ &= 0/2 * 1/2 * 1/3 \end{aligned}$$

Naïve Bayes

Our "**training corpus**":

...from Washington to… => CITY
…said Washington to… => PER
...leaving Washington in... => CITY

Our "**test example**":

...leaving Washington to… => ???

$$s_{MAP} = \text{argmax}_{s \in S} \prod_i P(x_i \mid s) P(s)$$

$$\begin{aligned} P(CITY \mid leaving, to) &= P(leaving \mid CITY) P(to \mid CITY) P(CITY) \\ &= 1/4 * 1/4 * 2/3 \\ &= 2/48 = 1/24 \end{aligned}$$

$$\begin{aligned} P(PER \mid leaving, to) &= P(leaving \mid PER) P(to \mid PER) P(PER) \\ &= 0/2 * 1/2 * 1/3 \\ &= 0 \end{aligned}$$

Naïve Bayes

Our "**training corpus**":

...from Washington to… => CITY
…said Washington to… => PER
...leaving Washington in... => CITY

Our "**test example**":

...leaving Washington to… => ???

$$s_{MAP} = \text{argmax}_{s \in S} \prod_i P(x_i \mid s) P(s)$$

$$\begin{aligned} P(CITY \mid leaving, to) &= P(leaving \mid CITY) P(to \mid CITY) P(CITY) \\ &= 1/4 * 1/4 * 2/3 \\ &= 2/48 = 1/24 \end{aligned}$$

$$\begin{aligned} P(PER \mid leaving, to) &= P(leaving \mid PER) P(to \mid PER) P(PER) \\ &= 0/2 * 1/2 * 1/3 \\ &= 0 \end{aligned}$$

Laplace Smoothing for Naïve Bayes

Our "**training corpus**":

...from Washington to… => CITY
…said Washington to… => PER
...leaving Washington in... => CITY

Our "**test example**":

...leaving Washington to… => ???

$$s_{MAP} = \text{argmax}_{s \in S} \prod_i P(x_i \mid s) P(s)$$

$$P(x_i \mid s) = \frac{count(x_i, s) + 1}{\sum_{x \in V} (count(x, s) + 1)}$$

Laplace Smoothing for Naïve Bayes

Our "**training corpus**":

...from Washington to… => CITY
…said Washington to… => PER
...leaving Washington in... => CITY

Our "**test example**":

...leaving Washington to… => ???

$$s_{MAP} = \text{argmax}_{s \in S} \prod_i P(x_i \mid s) P(s)$$

$$P(x_i \mid s) = \frac{count(x_i, s) + 1}{\sum_{x \in V} (count(x, s) + 1)}$$

$$\begin{aligned} P(CITY \mid leaving, to) &= P(leaving \mid CITY) P(to \mid CITY) P(CITY) \\ &= 2/9 * 2/9 * 2/3 \\ &= 2/81 \end{aligned}$$

Laplace Smoothing for Naïve Bayes

Our "**training corpus**":

...from Washington to… => CITY
…said Washington to… => PER
...leaving Washington in... => CITY

Our "**test example**":

...leaving Washington to… => ???

$$s_{MAP} = \text{argmax}_{s \in S} \prod_i P(x_i \mid s) P(s)$$

$$P(x_i \mid s) = \frac{count(x_i, s) + 1}{\sum_{x \in V} (count(x, s) + 1)}$$

$$\begin{aligned} P(CITY \mid leaving, to) &= P(leaving \mid CITY) P(to \mid CITY) P(CITY) \\ &= 2/9 * 2/9 * 2/3 \\ &= 2/81 \end{aligned}$$

$$\begin{aligned} P(PER \mid leaving, to) &= P(leaving \mid PER) P(to \mid PER) P(PER) \\ &= 1/7 * 2/7 * 1/3 \\ &= 2/147 \end{aligned}$$

In-Class Exercise

Training set:

…an apple that… => FRUIT
…that apple was… => FRUIT
…an apple store... => COMPANY
…an apple product…. => COMPANY
…tasty apple to… => FRUIT

In-Class Exercise

Training set:

…an apple that…	=> FRUIT
…that apple was…	=> FRUIT
…an apple store...	=> COMPANY
…an apple product….	=> COMPANY
…tasty apple to…	=> FRUIT

Test set:

…an apple to…	=> FRUIT or COMPANY?
…an apple product…	=> FRUIT or COMPANY?

Naïve Bayes

- Any type of features can be used.
- This can be used for a lot of NLP tasks:
  - Sentiment analysis
  - Author classification
  - …

Minimally supervised word sense disambiguation

Minimally-supervised WSD

- Famous bootstrapping algorithm: invented by Yarowsky (1995)
- Accuracy > 96%
- Uses two powerful heuristics:
  - **One sense per collocation**: nearby words provide clues to the sense of the target word, conditional on distance, order, syntactic relationship.

Minimally-supervised WSD

- Famous bootstrapping algorithm: invented by Yarowsky (1995)
- Accuracy > 96%
- Uses two powerful heuristics:
    - **One sense per collocation**: nearby words provide clues to the sense of the target word, conditional on distance, order, syntactic relationship.
    - **One sense per discourse**: the sense of a target words is consistent within a given document.

Goal: disambiguate *plant* in a corpus

Step 1: In a large corpus, identify all examples of the given polysemous word, storing their contexts as lines in an initially untagged training set.

Sense	Training Examples (Keyword in Context)
?	... company said the *plant* is still operating
?	Although thousands of *plant* and animal species
?	... zonal distribution of *plant* life . ...
?	... to strain microscopic *plant* life from the ...
?	vinyl chloride monomer *plant* , which is ...
?	and Golgi apparatus of *plant* and animal cells
?	... computer disk drive *plant* located in ...
?	... divide life into *plant* and animal kingdom
?	... close-up studies of *plant* life and natural
?	... Nissan car and truck *plant* in Japan is ...
?	... keep a manufacturing *plant* profitable without
?	... molecules found in *plant* and animal tissue
?	... union responses to *plant* closures . ...
?	... animal rather than *plant* tissues can be
?	... many dangers to *plant* and animal life
?	company manufacturing *plant* is in Orlando ...
?	... growth of aquatic *plant* life in water ...
?	automated manufacturing *plant* in Fremont ,
?	... Animal and *plant* life are delicately
?	discovered at a St. Louis *plant* manufacturing
?	computer manufacturing *plant* and adjacent ...
?	... the proliferation of *plant* and animal life
?	... ...

Goal: disambiguate *plant* in a corpus

Step 2: For each possible sense of the word, identify a relatively small number of training examples representative of that sense, for example by hand tagging a subset of the training sentences.

A ~ *life*

B ~ *manufacturing*

Sense	Training Examples (Keyword in Context)
A	used to strain microscopic *plant* **life** from the ...
A	... zonal distribution of *plant* **life** . ...
A	close-up studies of *plant* **life** and natural ...
A	too rapid growth of aquatic *plant* **life** in water ...
A	... the proliferation of *plant* and animal **life** ...
A	establishment phase of the *plant* virus **life** cycle ...
A	... that divide **life** into *plant* and animal kingdom
A	... many dangers to *plant* and animal **life** ...
A	mammals . Animal and *plant* **life** are delicately
A	beds too salty to support *plant* **life** . River ...
A	heavy seas, damage , and *plant* **life** growing on ...
A	... ...
?	... vinyl chloride monomer *plant* , which is ...
?	... molecules found in *plant* and animal tissue
?	... Nissan car and truck *plant* in Japan is ...
?	... and Golgi apparatus of *plant* and animal cells ...
?	... union responses to *plant* closures . ...
?	... ...
?	... ...
?	... cell types found in the *plant* kingdom are ...
?	... company said the *plant* is still operating ...
?	... Although thousands of *plant* and animal species
?	... animal rather than *plant* tissues can be ...
?	... computer disk drive *plant* located in ...
B	... ...
B	automated **manufacturing** *plant* in Fremont ...
B	... vast **manufacturing** *plant* and distribution ...
B	chemical **manufacturing** *plant* , producing viscose
B	... keep a **manufacturing** *plant* profitable without
B	computer **manufacturing** *plant* and adjacent ...
B	discovered at a St. Louis *plant* **manufacturing**
B	... copper **manufacturing** *plant* found that they
B	copper wire **manufacturing** *plant* , for example ...
B	's cement **manufacturing** *plant* in Alpena ...
B	polystyrene **manufacturing** *plant* at its Dow ...
B	company **manufacturing** *plant* is in Orlando ...

Goal: disambiguate *plant* in a corpus

Step 3a: Train a supervised classification algorithm on the SENSE-A/SENSE-B seed sets.

Goal: disambiguate *plant* in a corpus

Step 3a: Train a supervised classification algorithm on the SENSE-A/SENSE-B seed sets.

Step 3b: Apply the resulting classifier to the entire sample set. Take samples that are tagged as SENSE-A or SENSE-B with probability above a certain threshold, and add those examples to the growing seed sets.

Goal: disambiguate *plant* in a corpus

Step 3a: Train a supervised classification algorithm on the SENSE-A/SENSE-B seed sets.

Step 3b: Apply the resulting classifier to the entire sample set. Take samples that are tagged as SENSE-A or SENSE-B with probability above a certain threshold, and add those examples to the growing seed sets.

Step 3c: Optionally, the one-sense-per-discourse constraint can be used both to filter and augment this addition.

Goal: disambiguate *plant* in a corpus

Step 3a: Train a supervised classification algorithm on the SENSE-A/SENSE-B seed sets.

Step 3b: Apply the resulting classifier to the entire sample set. Take samples that are tagged as SENSE-A or SENSE-B with probability above a certain threshold, and add those examples to the growing seed sets.

Step 3c: Optionally, the one-sense-per-discourse constraint can be used both to filter and augment this addition.

Step 3d: Repeat Step 3 iteratively

Goal: disambiguate *plant* in a corpus

Step 4: Stop. When the training parameters are held constant, the algorithm will converge on a stable set of remaining instances.

Goal: disambiguate *plant* in a corpus

Step 5: The resulting dataset can now be used to train a classifier and to annotate new data with sense tags and probabilities.

How to evaluate?

How to evaluate?

Intrinsic evaluation:

- Sense accuracy
- On a held-out part of the corpus

How to evaluate?

Intrinsic evaluation:

- Sense accuracy
- On a held-out part of the corpus

Extrinsic evaluation:

- Use WSD in downstream tasks and evaluate final performance

How to evaluate?

Use the development set to develop your classifier (here: to pick features)!

Intrinsic evaluation:

- Sense accuracy
- On a held-out part of the corpus

Extrinsic evaluation:

- Use WSD in downstream tasks and evaluate final performance

Hypernym detection

Task Definition

- Given: a corpus
- Task: Find pairs of terms that are in a hyponym-hypernym relationship

- Examples:
  - Dog is a **<?>** of animal
  - Food is a **<?>** of pasta

Hearst Patterns (Hearst, 1992)

Idea: Pairs of words that are in hyponym-hypernym relationships tend to occur in certain lexico-syntactic patterns.

Hearst Patterns (Hearst, 1992)

Idea: Pairs of words that are in hyponym-hypernym relationships tend to occur in certain lexico-syntactic patterns.

The bow lute, such as the Bambara ndang,
is plucked and has an individual
curved neck for each string.

Hearst Patterns (Hearst, 1992)

Idea: Pairs of words that are in hyponym-hypernym relationships tend to occur in certain lexico-syntactic patterns.

```
The bow lute, such as the Bambara ndang,
is plucked and has an individual
curved neck for each string.
```

(1a) NP_0 *such as* $\{NP_1, NP_2 \ldots, (and \mid or)\}\ NP_n$

are such that they imply

(1b) *for all* NP_i, $1 \leq i \leq n$, *hyponym*(NP_i, NP_0)

Thus from sentence (S1) we conclude

hyponym("Bambara ndang", "bow lute").

Hearst Patterns (Hearst, 1992)

(2) *such NP as* {*NP* ,}* {(*or* | *and*)} *NP*

`... works by such authors as Herrick, Goldsmith, and Shakespeare.`

⟹ *hyponym*("author", "Herrick"),
hyponym("author", "Goldsmith"),
hyponym("author", "Shakespeare")

Hearst Patterns (Hearst, 1992)

(3) $NP \{, NP\}^* \{,\}$ *or other* NP

Bruises, wounds, broken bones or other injuries ...

$\Longrightarrow$ *hyponym("bruise", "injury")*, *hyponym("wound", "injury")*, *hyponym("broken bone", "injury")*

Hearst Patterns (Hearst, 1992)

(4) *NP* {, *NP*}* {,} *and other NP*

```
... temples, treasuries,and other
important civic buildings.
```

$\Longrightarrow$ *hyponym("temple", "civic building")*,
hyponym("treasury", "civic building")

Hearst Patterns (Hearst, 1992)

(5) *NP* {,} *including* {*NP* ,}* {*or* | *and*} *NP*

`All common-law countries, including Canada and England ...`

$\Longrightarrow$ *hyponym*("Canada", "common-law country"), *hyponym*("England", "common-law country")

Hearst Patterns (Hearst, 1992)

(6) *NP* {,} *especially* {*NP* ,}* {*or* | *and*} *NP*

`... most European countries, especially France, England, and Spain.`

$\Longrightarrow$ *hyponym*("France", "European country"), *hyponym*("England", "European country"), *hyponym*("Spain", "European country")

Overview: Hearst Patterns

- NP such as {NP}* {and|or} NP
- such NP as {NP ,}* {or|and} NP
- NP {, NP}* {,} or other NP
- NP {, NP}* {,} and other NP
- NP {,} including {NP, }* {or|and} NP
- NP {,} especially {NP ,}* {or|and} NP

Wrapping up

- Discussed today:
  - Some things about morphology
  - Definitions (types/tokens/hypernyms/…)
  - Word sense disambiguation
  - Naive Bayes*
  - Hypernym detection
- Next Monday: Distributional lexical semantics