Lecture 7: Distributional semantics (continued)

Outline.

Distributions as usage representations

Distributional word clustering

Selectional preferences

Distributional methods are a usage representation

- Distributions are a good conceptual representation if you believe that 'the meaning of a word is given by its usage'.
- Corpus-dependent, culture-dependent, register-dependent.
  Example: similarity between *policeman* and *cop*: 0.23

Distribution for *policeman*

policeman
0.59::ball_n+poss_rel
0.48::and_c+civilian_n
0.42::soldier_n+and_c
0.41::and_c+soldier_n
0.38::secret_a
0.37::people_n+include_v
0.37::corrupt_a
0.36::uniformed_a
0.35::uniform_n+poss_rel
0.35::civilian_n+and_c
0.31::iraqi_a
0.31::lot_n+poss_rel
0.31::chechen_a
0.30::laugh_v
0.29::and_c+criminal_n
0.28::incompetent_a
0.28::pron_rel_+shoot_v
0.28::hat_n+poss_rel
0.28::terrorist_n+and_c
0.27::and_c+crowd_n
0.27::military_a
0.27::helmet_n+poss_rel
0.27::father_n+be_v
0.26::on_p()+duty_n
0.25::salary_n+poss_rel
0.25::on_p()+horseback_n
0.25::armed_a
0.24::and_c+nurse_n
0.24::job_n+as_p()
0.24::open_v+fire_n

Distribution for *cop*

cop
0.45::crooked_a
0.45::corrupt_a
0.44::maniac_a
0.38::dirty_a
0.37::honest_a
0.36::uniformed_a
0.35::tough_a
0.33::pron_rel_+call_v
0.32::funky_a
0.32::bad_a
0.29::veteran_a
0.29::and_c+robot_n
0.28::and_c+criminal_n
0.28::bogus_a
0.28::talk_v+to_p()+pron_rel_
0.27::investigate_v+murder_n
0.26::on_p()+force_n
0.25::parody_n+of_p()
0.25::Mason_n+and_c
0.25::pron_rel_+kill_v
0.25::racist_a
0.24::addicted_a
0.23::gritty_a
0.23::and_c+interference_n
0.23::arrive_v
0.23::and_c+detective_n
0.22::look_v+way_n
0.22::dead_a
0.22::pron_rel_+stab_v
0.21::pron_rel_+evade_v

Distributions and knowledge

What kind of information do distributions encode?

- lexical knowledge
- world knowledge
- cultural biases
- boundaries between them are blurry

Distributions are partial lexical semantic representations, but useful and theoretically interesting.

Outline.

Distributions as usage representations

Distributional word clustering

Selectional preferences

Clustering

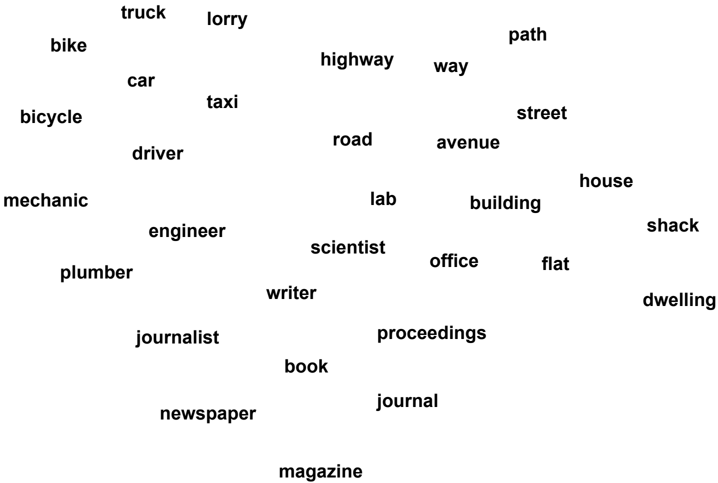
- clustering techniques group objects into clusters
- similar objects in the same cluster, dissimilar objects in different clusters
- allows us to obtain generalisations over the data
- widely used in various NLP tasks:
    - semantics (e.g. word clustering);
    - summarization (e.g. sentence clustering);
    - text mining (e.g. document clustering).

Distributional word clustering

We will:

- cluster words based on the contexts in which they occur
- assumption: words with similar meanings occur in similar contexts, i.e. are distributionally similar
- we will consider noun clustering as an example
- cluster 2000 nouns – most frequent in the British National Corpus
- into 200 clusters

Clustering nouns



Clustering nouns

Feature vectors

- can use different kinds of context as features for clustering
  - window based context
  - parsed or unparsed
  - syntactic dependencies
- different types of context yield different results
- Example experiment: use verbs that take the noun as a direct object or a subject as features for clustering
- Feature vectors: verb lemmas, indexed by dependency type, e.g. subject or direct object
- Feature values: corpus frequencies

Extracting feature vectors: Examples

tree (Dobj)	**crop (Dobj)**	**tree (Subj)**	**crop (Subj)**
85 plant_v	76 grow_v	131 grow_v	78 grow_v
82 climb_v	44 produce_v	49 plant_v	23 yield_v
48 see_v	16 harvest_v	40 stand_v	10 sow_v
46 cut_v	12 plant_v	26 fell_v	9 fail_v
27 fall_v	10 ensure_v	25 look_v	8 plant_v
26 like_v	10 cut_v	23 make_v	7 spray_v
23 make_v	9 yield_v	22 surround_v	7 come_v
23 grow_v	9 protect_v	21 show_v	6 produce_v
22 use_v	9 destroy_v	20 seem_v	6 feed_v
22 round_v	7 spray_v	20 overhang_v	6 cut_v
20 get_v	7 lose_v	20 fall_v	5 sell_v
18 hit_v	6 sell_v	19 cut_v	5 make_v
18 fell_v	6 get_v	18 take_v	5 include_v
18 bark_v	5 support_v	18 go_v	5 harvest_v
17 want_v	5 see_v	18 become_v	4 follow_v
16 leave_v	5 raise_v	17 line_v	3 ripen_v
...	...	...	...

Feature vectors: Examples

tree
131 grow_v_Subj
85 plant_v_Dobj
82 climb_v_Dobj
49 plant_v_Subj
48 see_v_Dobj
46 cut_v_Dobj
40 stand_v_Subj
27 fall_v_Dobj
26 like_v_Dobj
26 fell_v_Subj
25 look_v_Subj
23 make_v_Subj
23 make_v_Dobj
23 grow_v_Dobj
22 use_v_Dobj
22 surround_v_Subj
22 round_v_Dobj
20 overhang_v_Subj
...

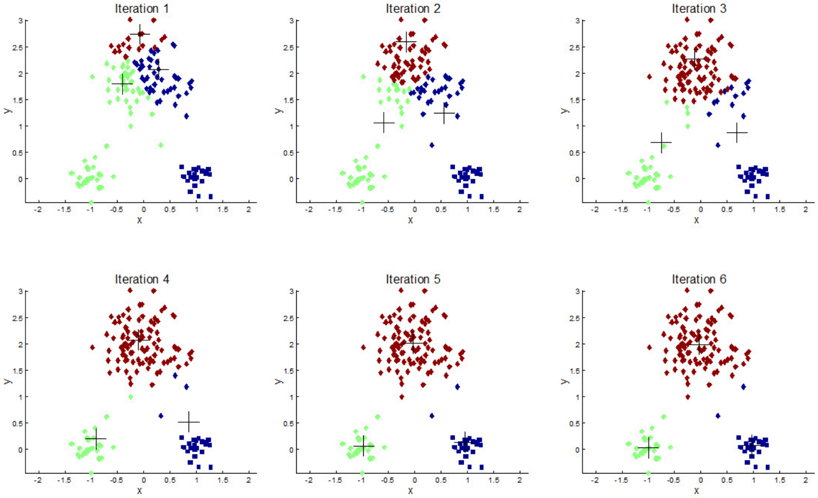
crop
78 grow_v_Subj
76 grow_v_Dobj
44 produce_v_Dobj
23 yield_v_Subj
16 harvest_v_Dobj
12 plant_v_Dobj
10 sow_v_Subj
10 ensure_v_Dobj
10 cut_v_Dobj
9 yield_v_Dobj
9 protect_v_Dobj
9 fail_v_Subj
9 destroy_v_Dobj
8 plant_v_Subj
7 spray_v_Subj
7 spray_v_Dobj
7 lose_v_Dobj
6 feed_v_Subj
...

Clustering algorithms, K-means

- many clustering algorithms are available
- example algorithm: K-means clustering
  - given a set of N data points $\{x_1, x_2, ..., x_N\}$
  - partition the data points into K clusters $C = \{C_1, C_2, ..., C_K\}$
  - minimize the sum of the squares of the distances of each data point to the cluster mean vector μ_i:

$$\arg\min_{C} \sum_{i=1}^{K} \sum_{\mathbf{x} \in C_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 \quad (1)$$

K-means clustering

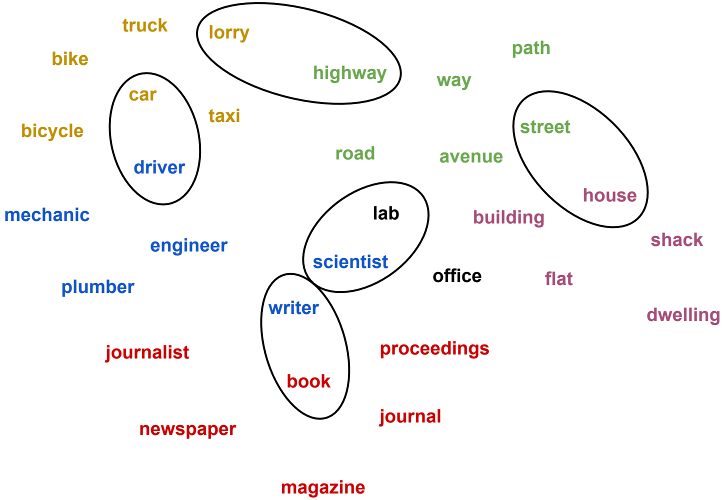


Noun clusters

tree crop flower plant root leaf seed rose wood grain stem forest garden
consent permission concession injunction licence approval
lifetime quarter period century succession stage generation decade phase interval future
subsidy compensation damages allowance payment pension grant
carriage bike vehicle train truck lorry coach taxi
official officer inspector journalist detective constable police policeman reporter
girl other woman child person people
length past mile metre distance inch yard
tide breeze flood wind rain storm weather wave current heat
sister daughter parent relative lover cousin friend wife mother husband brother father

Clustering nouns

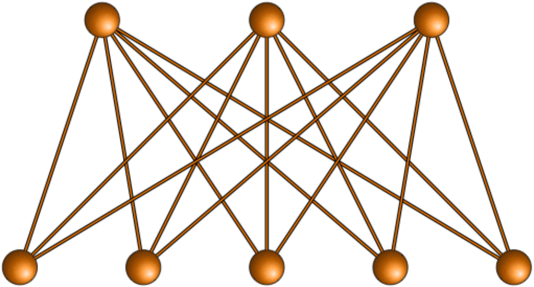
Clustering nouns



We can also cluster verbs...

sparkle glow widen flash flare gleam darken narrow flicker shine blaze bulge
gulp drain stir empty pour sip spill swallow drink pollute seep flow drip purify ooze pump bubble splash ripple simmer boil tread
polish clean scrape scrub soak
kick hurl push fling throw pull drag haul
rise fall shrink drop double fluctuate dwindle decline plunge decrease soar tumble surge spiral boom
initiate inhibit aid halt trace track speed obstruct impede accelerate slow stimulate hinder block
work escape fight head ride fly arrive travel come run go slip move

Hard and soft clustering



- Hard clustering: each data point assigned to one cluster only (as in our k-means experiment)
- Soft clustering: each data point is associated with multiple clusters with a membership probability

Uses of word clustering in NLP

- Noun and verb clustering are typically used in NLP
- for lexical acquisition tasks
  - to automatically create large-scale lexical resources to support other NLP tasks
  - e.g. selectional preferences
- dimensionality reduction for other statistical models
  - to reduce negative effects of sparse data

Outline.

Distributions as usage representations

Distributional word clustering

Selectional preferences

Subcategorization and selection

Subcategorization:

- different predicates (e.g. verbs) can take a different number and types of arguments (syntactic)

Selection:

- different predicates place different semantic constraints on the types of arguments that they can take

Subcategorization

- verbs (and other types of words) have different numbers and types of syntactic arguments:

  **Kim adored*
  **Kim gave Sandy*
  **Kim adored to sleep*
  Kim liked to sleep
  **Kim devoured*
  Kim ate

- verbs can be intransitive (1 argument: subj); transitive (2 arguments: subj and obj); ditransitive (3 arguments: subj and 2 obj).

Selectional preferences

- Selectional preferences are the semantic constraints that a predicate places onto its arguments.
- i.e. certain classes of entities are more likely to fill the predicate's argument slot than others.

The authors wrote a new paper.

Watch out, the cat is eating your sausage!

*The carrot ate the keys.

*The law sang a driveway.

Selectional preferences

- Selectional preferences are the semantic constraints that a predicate places onto its arguments.
- i.e. certain classes of entities are more likely to fill the predicate's argument slot than others.

The authors wrote a new paper.

Watch out, the cat is eating your sausage!

*The carrot ate the keys.

*The law sang a driveway.

Selectional preferences

- Selectional preferences are the semantic constraints that a predicate places onto its arguments.
- i.e. certain classes of entities are more likely to fill the predicate's argument slot than others.

The authors wrote a new paper.

Watch out, the cat is eating your sausage!

*The carrot ate the keys.

*The law sang a driveway.

Selectional preferences

- Selectional preferences are the semantic constraints that a predicate places onto its arguments.
- i.e. certain classes of entities are more likely to fill the predicate's argument slot than others.

The authors wrote a new paper.

Watch out, the cat is eating your sausage!

*The carrot ate the keys.

*The law sang a driveway.

Selectional preferences

- Selectional preferences are the semantic constraints that a predicate places onto its arguments.
- i.e. certain classes of entities are more likely to fill the predicate's argument slot than others.

The authors wrote a new paper.

Watch out, the cat is eating your sausage!

*The carrot ate the keys.

*The law sang a driveway.

Learning selectional preferences from distributions

[animate] eat [food]

[person] sing [song]

[person] read [book]

1. Need to define a set of argument classes that can fill the argument slot of the predicate

   e.g. use noun clusters for this purpose
2. Need to quantify the level of association of a particular verb with a particular noun class

Selectional preference model

Phillip Resnik, 1997. *Selectional Preference and Sense Disambiguation*

Selectional preference strength

$$S_R(v) = D_{KL}(P(c|v)||P(c)) = \sum_c P(c|v) \log \frac{P(c|v)}{P(c)}$$

D_{KL} is Kullback–Leibler divergence

Selectional association

$$A_R(v,c) = \frac{1}{S_R(v)} P(c|v) \log \frac{P(c|v)}{P(c)}$$

$P(c)$ is the prior probability of the noun class;
$P(c|v)$ its posterior probability given the verb; R is the grammatical relation

Calculating probabilities

$$P(c) = \frac{f(c)}{\sum_k f(c_k)},$$

$$P(c|v) = \frac{f(v,c)}{f(v)},$$

$$f(c) = \sum_{n_i \in c} f(n_i)$$

$f(v,c)$: frequency of verb v co-occurring with the noun class c
$f(v)$: total frequency of verb v with all noun classes
$f(c)$: total frequency of the noun class c

Selectional preferences of ***kill*** (Dobj)

0.38 girl other woman child person people
0.20 being species sheep animal creature horse baby human fish male lamb bird rabbit female insect cattle mouse monster
0.19 sister daughter parent relative lover cousin friend wife mother husband brother father
0.04 thousand citizen inhabitant resident minority youngster refugee peasant miner hundred
0.0378 gene tissue cell particle fragment bacterium protein acid complex compound molecule organism
0.0336 fleet soldier knight force rebel guard troops crew army pilot
0.0335 official officer inspector journalist detective constable police policeman reporter
0.0322 victim bull teenager prisoner hero gang enemy rider offender youth killer thief driver defender hell
0.0136 week month year
...

Selectional preferences of ***drink*** (Dobj)

0.5831 drink coffee champagne pint wine beer
0.2778 drop tear sweat paint blood water juice
0.1084 mixture salt dose ingredient sugar substance drug milk cream alcohol fibre chemical
0.0515 brush bowl bucket receiver barrel dish glass container plate basket bottle tray
0.0069 couple minute night morning hour time evening afternoon
0.0041 stability efficiency security prospects health welfare survival safety
0.0025 recording music tape song tune radio guitar trick album football organ stuff
0.0005 rage excitement panic anger terror flame laughter
0.0004 ball shot kick arrow stroke bullet punch bomb shell blow missile
0.0003 lunch dinner breakfast meal
...

Different senses of ***run***

The children **ran** to the store
If you see this man, **run**!
Service **runs** all the way to Cranbury
She is **running** a relief operation in Sudan
the story or argument **runs** as follows
Does this old car still **run** well?
Interest rates **run** from 5 to 10 percent
Who's **running** for treasurer this year?
They **ran** the tapes over and over again
These dresses **run** small

Selectional preferences of ***run*** (Subj)

0.2125 drop tear sweat paint blood water juice
0.1665 technology architecture program system product version interface software tool computer network processor chip package
0.1657 tunnel road path trail lane route track street bridge
0.1166 carriage bike vehicle train truck lorry coach taxi
0.0919 tide breeze flood wind rain storm weather wave current heat
0.0865 tube lock tank circuit joint filter battery engine device disk furniture machine mine seal equipment machinery wheel motor slide disc instrument
0.0792 ocean canal stream bath river waters pond pool lake
0.0497 rope hook cable wire thread ring knot belt chain string
0.0469 arrangement policy measure reform proposal project programme scheme plan course
0.0352 week month year
0.0351 couple minute night morning hour time evening afternoon

Selectional preferences of ***run*** (continued)

0.0341 criticism appeal charge application allegation claim objection suggestion case complaint
0.0253 championship open tournament league final round race match competition game contest
0.0218 desire hostility anxiety passion doubt fear curiosity enthusiasm impulse instinct emotion feeling suspicion
0.0183 expenditure cost risk expense emission budget spending
0.0136 competitor rival team club champion star winner squad county player liverpool partner leeds
0.0102 being species sheep animal creature horse baby human fish male lamb bird rabbit female insect cattle mouse monster
...

Selectional preferences of ***cut*** (Dobj)

0.2845 expenditure cost risk expense emission budget spending
0.1527 dividend price rate premium rent rating salary wages
0.0832 employment investment growth supplies sale import export production consumption traffic input spread supply flow
0.0738 potato apple slice food cake meat bread fruit
0.0407 stitch brick metal bone strip cluster coffin stone piece tile fabric rock layer remains block
0.0379 excess deficit inflation unemployment pollution inequality poverty delay discrimination symptom shortage
0.0366 tree crop flower plant root leaf seed rose wood grain stem forest garden
0.0330 tail collar strand skirt trousers hair curtain sleeve
0.0244 rope hook cable wire thread ring knot belt chain string
...

Uses of selectional preferences

Widely used in NLP as a source of lexical information:

- Word sense induction and disambiguation
- Parsing (resolving ambiguous attachments)
- Identifying figurative language and idioms
- Paraphrasing and paraphrase detection
- Natural language inference (e.g. in the entailment identification task)
- etc.

We looked at verbs, other parts of speech (e.g. adjectives) can have preferences too.