IN2110: Språkteknologiske metoder
Ord: Leksikalsk semantikk og ordvektorer

Eivind Alexander Bergem
Samia Touileb

Språkteknologigruppen (LTG)

19. februar 2019

Last week – Erik Velldal

- Evaluation of classifiers.
- Unsupervised machine learning for class discovery: Clustering.
- k-means clustering.

Today

- Focus on *words* rather than documents.
- Distributional models of word meaning (lexical semantics).
- Vector Semantics.
- Words and Vectors.
- Example tasks for evaluating word vectors.

Lexical Semantics

Lexical Semantics = linguistic study of word meaning.

How should we represent the meaning of a word?

What do we want a word meaning model to do for us?

We want a model of word meaning to tell us (e.g.):

- words have similar meanings (*cat* is similar to *dog*).
- words are antonyms (*cold* is the opposite of *hot*).
- words have positive or negative connotations (*happy* and *sad* respectively).
- the meaning of *sell*, *pay*, *buy* are different perspectives on the same underlying purchasing event.

What do words mean?

lemma (citation form)

Dictionary

`mouse (N)`

1. `any of numerous small rodents...`
2. `a hand-operated device that controls a cursor...`

sense

definition

mice? = wordform

What do words mean?

- A sense or "concept" is the meaning component of a word.
- Important component of word meaning = relationships between word senses.

Synonymy

- Synonyms have the same meaning in some or all contexts.
  - *couch & sofa, big & large, automobile & car, vomit & throw up, water & $H_2 0$*
- Synonymy between senses: "a word sense whose meaning is (nearly) identical to a sense of another word". Chapter 6, p.3.
- Synonymy between words (more formal): "two words are synonyms if they are substitutable one for the other in any sentence without changing the *truth conditions* of the sentence, the situations in which the sentence would be true." Chapter 6, p.3.
- No examples of perfect synonymy.
- The Linguistic Principle of Contrast:
  - Difference in form $\rightarrow$ difference in meaning.
  - *water & $H_2 0$*: $\mathrm{H_2 0}$ appropriate in scientific contexts, inappropriate in hiking guide.

Antonymy

Senses that are opposites with respect to one feature of meaning. Otherwise, very similar!

- *dark/light* , *short/long* , *fast/slow* , *rise/fall* , *hot/cold* , *up/down* , *in/out*

More formally, antonyms can:

- define a binary opposition, or be at opposite ends of a scale
  - long/short , fast/slow
- be reversives:
  - rise/fall, up/down

Synonym VS Antonyms

How to automatically distinguish synonyms from antonyms ?

Similarity

Words don't have many synonyms, but have lots of similar words.

From synonymy to similarity:

- relations between senses → relations between words.

Words with similar meanings. Not synonyms, but sharing some element of meaning:

Examples

- alligator, crocodile
- love, affection
- cat, dog

Word relatedness

Word relatedness (or word association): words are related if they do not share features, but commonly "participate" in a shared event.

Similarity VS Relatedness

- car , bicycle: similar.
- car , gasoline: related, **not** similar.

Word Relatedness

- car , tyre
- car , motorway
- car , crash
- coffee , cup

Words can be related in any way, e.g. via a semantic frame or semantic field.

Values of word similarity

How to get values for word similarities?

How to automatically differentiate between word similarity and word relatedness?

- Ask human judges!

SimLex-999 dataset (Hill et al., 2015):

- gold standard resource for evaluating distributional semantic models.
- quantifies similarity rather than relatedness.

word1	**word2**	**similarity**
vanish	disappear	9.8
behave	obey	7.3
belief	impression	5.95
muscle	bone	3.65
modest	flexible	0.98
hole	agreement	0.3

Semantic field

Words that:

- cover a particular semantic domain.
- bear structured relations with each other.

Examples

- **hospitals**: surgeon, scalpel, nurse, anesthetic, hospital
- **restaurants**: waiter, menu, plate, food, chef
- **houses**: door, roof, kitchen, family, bed

Semantic Frames and Roles

Closely related to semantic fields.

"A semantic frame is a set of words that denote perspectives or participants in a particular type of event." Chapter 6, p.4.

Frames have semantic roles, and words in a sentence can take on these roles.

- buy, sell, pay.
- buyer, seller, goods, money.

Semantic frames makes it possible for systems to recognize paraphrases.

- Sam bought the book from Ling.
- Ling sold the book to Sam.

Taxonomic Relations

Word senses can be related taxonomically.

hyponyms and hypernyms:

- a word (sense) is a hyponym of another word (sense) if the first is more specific, denoting a subclass of the other.
- a word (sense) is a hypernym of another word (sense) if the first one is more general.

Examples

- car is a hyponym of vehicle.
- mango is a hyponym of fruit.
- vehicle is a hypernym of car.
- fruit is a hypernym of mango.

Taxonomic Relations

Hypernymy can be defined in terms of entailment:

- sense A is a hyponym of B if everything that is A is also B.
- being A entails being B.

Hyponymy and hypernymy are transitive: A hyponym of B, B hyponym of C, then A is hyponym of C.

Hyponyms and hypernyms structure is the IS-A hierarchy: A IS-A B, or B subsumes A.

Hyponym and hypernym too similar: easily confused.

The words subordinate and superordinate are used instead.

Superordinate	vehicle	fruit	furniture
Subordinate	car	mango	chair

Connotation

Words have affective meanings or connotations.

Connotations are aspects of a word's meaning related to a writer/reader's emotions, sentiment, opinions, or evaluations.

- positive connotations (happy).
- negative connotations (sad).
- positive evaluation (great, love) – sentiment.
- negative evaluation (terrible, hate) – sentiment.

In affective meaning (Osgood et al., 1957) – words vary along 3 dimensions:

- **valence**, **arousal**, and **dominance** represented by numbers.
- word meaning can be represented as a vector, a list of numbers, point in a dimensional space.

A computational model of word meaning

Concepts or word senses

- Have a complex many-to-many association with words (homonymy, multiple senses)
- Have relations with each other
  - Synonymy , Antonymy, Similarity, Relatedness, Superordinate/subordinate, Connotation

How to build a computational model that successfully deals with the different aspects of word meaning?

How to define word meaning for a computational model?

Distributional hypothesis

- A perfect model that can deal with all the aspects of word meaning is very difficult to find!
- Linguistic and philosophical works from the 1950's inspired the current best model: vector semantics
- "The meaning of a word is its use in the language" (Ludwig Wittgenstein, 1953, PI 43).
- The linguistic distributionalists Joos (1950), Harris (1954), Firth (1957):
  - words are defined by their environments or distributions (the words around them).
  - a word's distribution is the set of contexts in which it occurs: the neighboring words or grammatical environment.
- "If A and B have almost identical environments we say that they are synonyms." (Zellig Harris, 1954).

→ two words that occur in very similar distributions (context, similar words) tend to have the same meaning.

What does ongchoi mean?

Suppose you see these sentences:

- Ongchoi is delicious **sautéed with garlic**.
- Ongchoi is superb **over rice**.
- Ongchoi **leaves** with salty sauces.

And you've also seen these:

- ...spinach **sautéed with garlic over rice**.
- Chard stems and **leaves** are **delicious**.
- Collard greens and other **salty** leafy greens.

Conclusion:

- Ongchoi is a leafy green like spinach, chard (bladbete), or collard greens (en type grønnkål).

How to do it computationally?

Computationally:

- count words in the context of *ongchoi*:
  - find words like *sauteed*, *eaten*, *garlic*.
  - these words occur around *spinach*/*collard green*.

→ there is a similarity between *ongchoi* and *spinach*/*collard green*.

Vector semantics combines two intuitions:

- distributional intuition.
- vector intuition (Osgood et al., 1957, slide 22)

Model of word meaning

Build a model of meaning focused on similarity:

- Each word = a vector.
- Similar words are “nearby in space”.

Words as a vectors

A word vector is called an "embedding" (embedded into a space).

The standard way to represent meaning in NLP.

Fine-grained model of meaning for similarity:

We focus on 2 kinds of embeddings:

- Tf-idf:
  - Sparse vectors and common baseline model.
  - Words are represented by a simple function of the counts of nearby words.
- Word2vec (next week with Fredrik):
  - Dense vectors.
  - Representation is created by training a classifier to distinguish nearby and far-away words.

Vectors and documents VS Words and vectors

Distributional models of meaning (vectors) generally based on a co-occurrence matrix.

Term-document matrix VS word-word matrix:

	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	0	7	13
good	114	80	62	89
fool	36	58	1	4
wit	20	15	2	3

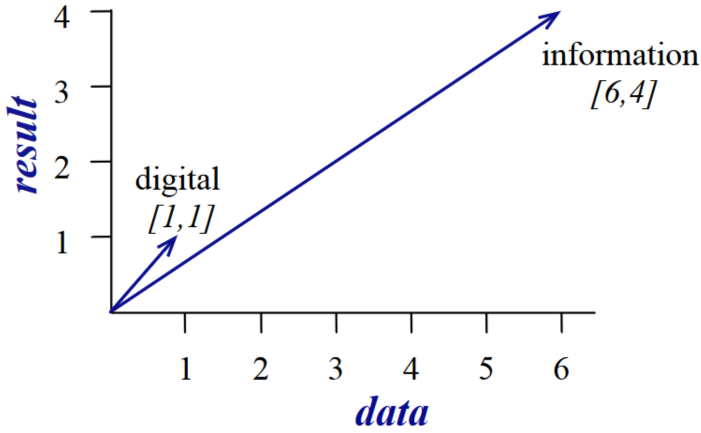
	As You Like It	Twelfth Night	Julius Caesar	Henry V
battle	1	0	7	13
good	114	80	62	89
fool	36	58	1	4
wit	20	15	2	3

Word-Word matrix

word-word matrix or term-term matrix or term-context matrix.

Two words are similar in meaning if their context vectors are similar.

	aardvark	...	computer	data	pinch	result	sugar
apricot	0	...	0	0	1	0	1
pineapple	0	...	0	0	1	0	1
digital	0	...	2	1	0	1	0
infomation	0	...	1	6	0	4	0


4
3
2
1
result
information
[6,4]
digital
[1,1]
1
2
3
4
5
6
data

Based on all we've seen so far, how to compute word similarity?
(Eivind will tell you more about it!)