An Extended NMF Algorithm for Word Sense Discrimination

Tim Van de Cruys

University of Groningen

BENELEARN 2008
May 20, 2008
Spa

Semantic similarity

- Most work on semantic similarity relies on the DISTRIBUTIONAL HYPOTHESIS (Harris 1954)
  - Take a word and its contexts:
    - tasty *klemenrak*
    - sour *klemenrak*
    - a bottle of *klemenrak*
    - *klemenrak* gone bad
- By looking at a word's context, one can infer its meaning
- Computationally: vector space model

Semantic similarity

- Most work on semantic similarity relies on the DISTRIBUTIONAL HYPOTHESIS (Harris 1954)

- Take a word and its contexts:
  - tasty *klemenrak*
  - sour *klemenrak*
  - a bottle of *klemenrak*
  - *klemenrak* gone bad

$\Rightarrow$ **FOOD/DRINK**

- By looking at a word's context, one can infer its meaning
- Computationally: vector space model

Semantic similarity

- Most work on semantic similarity relies on the DISTRIBUTIONAL HYPOTHESIS (Harris 1954)

- Take a word and its contexts:
  - tasty *klemenrak*
  - sour *klemenrak*
  - a bottle of *klemenrak*
  - *klemenrak* gone bad

- By looking at a word's context, one can infer its meaning
- Computationally: vector space model

Two kinds of context

1. 'Bag of words' context
   - a window around the word is used as context
   - e.g. a fixed numbers of words, the paragraph in which a word appears, ...
   - often used with some form of dimensionality reduction
   - 'topical' similarity
2. Syntactic context
   - a corpus is parsed, dependency triples are extracted
   - e.g. <apple, obj, eat>, <apple, adj, red>
   - typically does not use any form of dimensionality reduction
   - tighter, synonym-like similarity

Ambiguity

- **Problem**: ambiguity
  - Compare:
    a trendy bar
    ↔ *an iron bar*
    ↔ *today's air pressure: 1.013 bar*
  - Different meanings, but they are considered the same entity by a naive algorithm
- Main research question: can 'bag of words' context and syntactic context be combined to differentiate between various senses of a word?

Technique

- Given a non-negative matrix V, find non-negative matrix factors W and H such that:

$$V_{nxm} \approx W_{nxr} H_{rxm} \tag{1}$$

- Choosing $r \ll n, m$ reduces data
- Constraint on factorization: all values in three matrices need to be *non-negative values* (≥ 0)
- Constraint brings about a *parts-based* representation: only additive, no subtractive relations are allowed

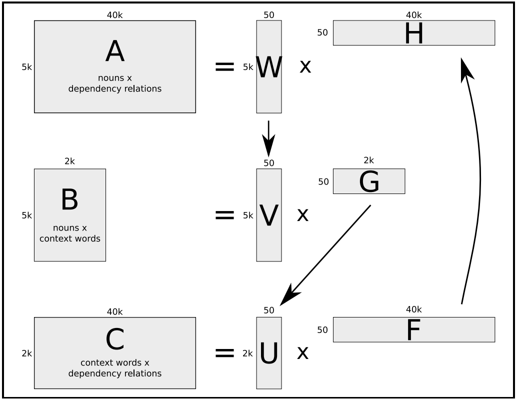
Results

- Context vectors (5k nouns $\times$ 2k co-occurring nouns)
- NMF is able to capture 'semantic' dimensions
- Examples:
  - *bus* 'bus', *taxi* 'taxi', *trein* 'train', *halte* 'stop', *reiziger* 'traveler', *perron* 'platform', *tram* 'tram', *station* 'station', *chauffeur* 'driver', *passagier* 'passenger'
  - *bouillon* 'broth', *slagroom* 'cream', *ui* 'onion', *eierdooier* 'egg yolk', *laurierblad* 'bay leaf', *zout* 'salt', *deciliter* 'decilitre', *boter* 'butter', *bleekselderij* 'celery', *saus* 'sauce'

Methodology

- Goal: classification of nouns according to both 'bag of words' context and syntactic context
- $\Rightarrow$ Construct three matrices capturing co-occurrence frequencies for each mode
  - nouns cross-classified by dependency relations
  - nouns cross-classified by (bag of words) context words
  - dependency relations cross-classified by context words
- $\Rightarrow$ Apply NMF to matrices, but interleave the process
- Result of former factorization is used to initialize factorization of the next one

Graphical Representation



Sense subtraction

- ‘switch off’ one dimension of an ambiguous word to reveal other possible senses
- Matrix H gives the importance of each dependency relation given a dimension
- ‘subtract’ dependency relations that are responsible for a given dimension from the original noun vector
  - $\overrightarrow{v}_{new} = \overrightarrow{v}_{orig}(\overrightarrow{1} - \overrightarrow{h}_{dim})$
  - each dependency relation is multiplied by a scaling factor, according to the load of the feature on the subtracted dimensions

Combination with clustering

- A simple clustering algorithm (K-means) assigns ambiguous nouns to its predominant sense
- Centroid of the cluster is fold into topic model
- The dimensions that define the centroid are subtracted from the ambiguous noun vector
- Adapted noun vector is fed to the clustering algorithm again

university of groningen

Experimental Design

- Approach applied to Dutch, using Twente Nieuws Corpus ($\pm$ 500M words)
- Corpus parsed with Dutch dependency parser ALPINO
- three matrices constructed with:
  - 5k nouns $\times$ 40k dependency relations
  - 5k nouns $\times$ 2k context words
  - 40k dependency relations $\times$ 2k context words
- Factorization to 50 dimensions

Example dimension: transport

1. **nouns**: *auto* 'car', *wagen* 'car', *tram* 'tram', *motor* 'motorbike', *bus* 'bus', *metro* 'subway', *automobilist* 'driver', *trein* 'trein', *stuur* 'steering wheel', *chauffeur* 'driver'
2. **context words**: *auto* 'car', *trein* 'train', *motor* 'motorbike', *bus* 'bus', *rij* 'drive', *chauffeur* 'driver', *fiets* 'bike', *reiziger* 'reiziger', *passagier* 'passenger', *vervoer* 'transport'
3. **dependency relations**: viertraps$_{adj}$ 'four pedal', verplaats_met$_{obj}$ 'move with', toeter$_{adj}$ 'honk', tank_in_houd$_{obj}$ [parsing error], tank$_{subj}$ 'refuel', tank$_{obj}$ 'refuel', rij_voorbij$_{subj}$ 'pass by', rij_voorbij$_{adj}$ 'pass by', rij_af$_{subj}$ 'drive off', peperduur$_{adj}$ 'very expensive'

Pop: most similar words

pop music $\leftrightarrow$ *doll*

1. *pop*, *rock*, *jazz*, *meubilair* 'furniture', *popmuziek* 'pop music', *heks* 'witch', *speelgoed* 'toy', *kast* 'cupboard', *servies* '[tea] service', *vraagteken* 'question mark'
2. *pop*, *meubilair* 'furniture', *speelgoed* 'toy', *kast* 'cupboard', *servies* '[tea] service', *heks* 'witch', *vraagteken* 'question mark' *sieraad* 'jewel', *sculptuur* 'sculpture', *schoen* 'shoe'
3. *pop*, *rock*, *jazz*, *popmuziek* 'pop music', *heks* 'witch', *danseres* 'dancer', *servies* '[tea] service', *kopje* 'cup', *house* 'house music', *aap* 'monkey'

Barcelona: most similar words

Spanish city ↔ *Spanish football club*

1. *Barcelona*, *Arsenal*, *Inter*, *Juventus*, *Vitesse*, *Milaan* 'Milan', *Madrid*, *Parijs* 'Paris', *Wenen* 'Vienna', *München* 'Munich'
2. *Barcelona*, *Milaan* 'Milan', *München* 'Munich', *Wenen* 'Vienna', *Madrid*, *Parijs* 'Paris', *Bonn*, *Praag* 'Prague', *Berlijn* 'Berlin', *Londen* 'London'
3. *Barcelona*, *Arsenal*, *Inter*, *Juventus*, *Vitesse*, *Parma*, *Anderlecht*, *PSV*, *Feyenoord*, *Ajax*

Clustering example: werk

1. *werk* 'work', *beeld* 'image', *foto* 'photo', *schilderij* 'painting', *tekening* 'drawing', *doek* 'canvas', *installatie* 'installation', *afbeelding* 'picture', *sculptuur* 'sculpture', *prent* 'picture', *illustratie* 'illustration', *handschrift* 'manuscript', *grafiek* 'print', *aquarel* 'aquarelle', *maquette* 'scale-model', *collage* 'collage', *ets* 'etching'
2. *werk* 'work', *boek* 'book', *titel* 'title', *roman* 'novel', *boekje* 'booklet', *debuut* 'debut', *biografie* 'biography', *bundel* 'collection', *toneelstuk* 'play', *bestseller* 'bestseller', *kinderboek* 'child book', *autobiografie* 'autobiography', *novelle* 'short story',
3. *werk* 'work', *voorziening* 'service', *arbeid* 'labour', *opvoeding* 'education', *kinderopvang* 'child care', *scholing* 'education', *huisvesting* 'housing', *faciliteit* 'facility', *accommodatie* 'acommodation', *arbeidsomstandigheid* 'working condition'

Methodology

- Comparison to EuroWordNet senses
- using Wu & Palmer's Wordnet similarity measure
- Calculate precision and recall
  - Precision: Percentage of correct clusters to which senses are assigned
  - Recall: Percentage of senses in EuroWordnet that have a corresponding cluster

Results

		threshold θ		
		.40 (%)	.50 (%)	.60 (%)
kmeans$_{nmf}$	prec.	78.97	69.18	55.16
	rec.	63.90	55.95	44.77
CBC	prec.	44.94	38.13	29.74
	rec.	69.61	60.00	48.00
kmeans$_{orig}$	prec.	86.13	74.99	58.97
	rec.	60.23	52.45	41.80

Conclusion

- Combining bag of words data and syntactic data is useful
  - bag of words data (factorized with NMF) puts its finger on topical dimensions
  - syntactic data is particularly good at finding similar words
  - a clustering approach allows one to determine which topical dimension(s) are responsible for a certain sense
  - and adapt the (syntactic) feature vector of the noun accordingly
  - subtracting the more dominant sense to discover less dominant senses
- Algorithm scores better with regard to precision; lower with regard to recall

Future Work

- Evaluate the method with other evaluation frameworks (focus on ambiguous nouns, Cornetto Database)
- Work out proper probabilistic framework for 'subtraction' of dimensions
- Use the results of the method to learn selectional preferences, in order to improve parser performance