Non-negative Tensor Factorization for Selectional Preference Induction

Tim Van de Cruys

University of Groningen

EACL workshop
GEometric **M**odels of Natural Language **S**emantics
March 31, 2009
Athens, Greece

Distributional similarity

Distributional similarity models are able to infer (lexical) semantics from text:

- Semantically similar words (top similar words)
  - **Athene** ‘Athens’: *Kopenhagen* ‘Copenhagen’, *Rome*, *Stockholm*, *Toronto*, *Istanbul*, *Praag* ‘Prague’, *Moskou* ‘Moscow’, *Helsinki*, *Madrid*
  - **liefde** ‘love’: *vriendschap* ‘friendship’, *passie* ‘passion’, *verlangen* ‘desire’, *angst* ‘fear’, *verdriet* ‘sadness’, *emotie* ‘emotion’, *schoonheid* ‘beauty’, *geloof* ‘faith’

Distributional similarity

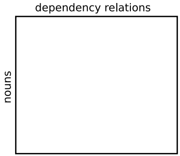
Distributional similarity models are able to infer (lexical) semantics from text:

- Semantic dimensions (top words on dimension)
    - (**transport**): *bus* 'bus', *taxi* 'taxi', *trein* 'train', *halte* 'stop', *reiziger* 'traveler', *perron* 'platform', *tram* 'tram', *station* 'station', *chauffeur* 'driver', *passagier* 'passenger'
    - (**food**): *bouillon* 'broth', *slagroom* 'cream', *ui* 'onion', *eierdooier* 'egg yolk', *laurierblad* 'bay leaf', *zout* 'salt', *deciliter* 'decilitre', *boter* 'butter', *bleekselderij* 'celery', *saus* 'sauce'

university of groningen

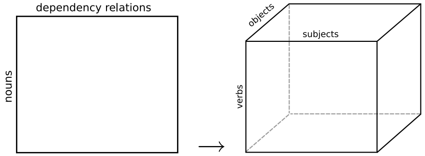
Two-way vs. three-way

- all methods use two way co-occurrence frequencies $\longrightarrow$ matrix
- suitable for two-way problems
  - words $\times$ documents
  - nouns $\times$ dependency relations
- not suitable for *n*-way problems
  - words $\times$ documents $\times$ authors
  - verbs $\times$ subjects $\times$ direct objects



Two-way vs. three-way

- all methods use two way co-occurrence frequencies $\longrightarrow$ matrix
- suitable for two-way problems
  - words $\times$ documents
  - nouns $\times$ dependency relations
- not suitable for *n*-way problems $\longrightarrow$ tensor
  - words $\times$ documents $\times$ authors
  - verbs $\times$ subjects $\times$ direct objects



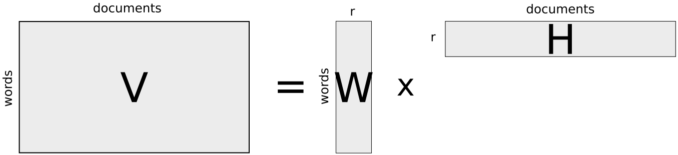
Technique

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

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

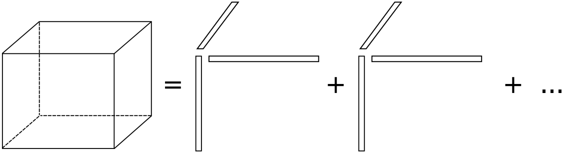
- Choosing $r \ll n, m$ reduces data
- Constraint on factorization: all values in three matrices need to be *non-negative values* (≥ 0)
- Constraint brings about a *parts-based* representation: only additive, no subtractive relations are allowed

Graphical Representation

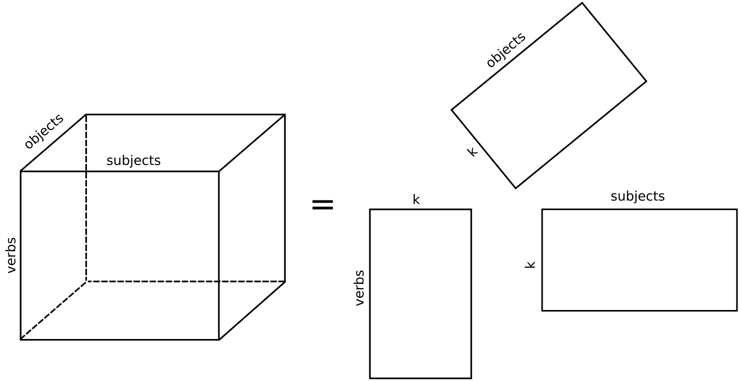


Technique

- Idea similar to non-negative matrix factorization
- Calculations are different
- $\min_{x_i \in \mathbb{R}^{D1}_{\geq 0}, y_i \in \mathbb{R}^{D2}_{\geq 0}, z_i \in \mathbb{R}^{D3}_{\geq 0}} \parallel T - \sum_{i=1}^{k} x_i \circ y_i \circ z_i \parallel_F^2$



Graphical representation



Methodology

- Three-way extraction of selectional preferences
- Approach applied to Dutch, using TWENTE NIEUWS CORPUS (500M words of newspaper texts)
- parsed with Dutch dependency parser ALPINO
- three-way co-occurrence of verbs with subjects and direct objects extracted
- adapted with extension of pointwise mutual information
- Resulting tensor 1K verbs × 10K subjects × 10K direct objects
- reduction to k dimensions ($k = 50, 100, 300$)

Police action

subjects	su_s	**verbs**	v_s	**objects**	obj_s
politie 'police'	.99	*houd_aan* 'arrest'	.64	*verdachte* 'suspect'	.16
agent 'policeman'	.07	*arresteer* 'arrest'	.63	*man* 'man'	.16
autoriteit 'authority'	.05	*pak_op* 'run in'	.41	*betoger* 'demonstrator'	.14
Justitie 'Justice'	.05	*schiet_dood* 'shoot'	.08	*relschopper* 'rioter'	.13
recherche 'detective force'	.04	*verdenk* 'suspect'	.07	*raddraaier* 'instigator'	.13
marechaussee 'military police'	.04	*tref_aan* 'find'	.06	*overvaller* 'raider'	.13
justitie 'justice'	.04	*achterhaal* 'overtake'	.05	*Roemeen* 'Romanian'	.13
arrestatieteam 'special squad'	.03	*verwijder* 'remove'	.05	*actievoerder* 'campaigner'	.13
leger 'army'	.03	*zoek* 'search'	.04	*hooligan* 'hooligan'	.13
douane 'customs'	.02	*spoor_op* 'track'	.03	*Algerijn* 'Algerian'	.13

Legislation

subjects	su_s	**verbs**	v_s	**objects**	obj_s
meerderheid ‘majority’	.33	*steun* ‘support’	.83	*motie* ‘motion’	.63
VVD	.28	*dien_in* ‘submit’	.44	*voorstel* ‘proposal’	.53
D66	.25	*neem_aan* ‘pass’	.23	*plan* ‘plan’	.28
Kamermeerderheid	.25	*wijs_af* ‘reject’	.17	*wetsvoorstel* ‘bill’	.19
fractie ‘party’	.24	*verwerp* ‘reject’	.14	*hem* ‘him’	.18
PvdA	.23	*vind* ‘think’	.08	*kabinet* ‘cabinet’	.16
CDA	.23	*aanvaard* ‘accepts’	.05	*minister* ‘minister’	.16
Tweede Kamer	.21	*behandel* ‘treat’	.05	*beleid* ‘policy’	.13
partij ‘party’	.20	*doe* ‘do’	.04	*kandidatuur* ‘candidature’	.11
Kamer ‘Chamber’	.20	*keur_goed* ‘pass’	.03	*amendement* ‘amendment’	.09

Exhibition

subjects	su_s	verbs	v_s	objects	obj_s
tentoonstelling 'exhibition'	.50	*toon* 'display'	.72	*schilderij* 'painting'	.47
expositie 'exposition'	.49	*omvat* 'cover'	.63	*werk* 'work'	.46
galerie 'gallery'	.36	*bevat* 'contain'	.18	*tekening* 'drawing'	.36
collectie 'collection'	.29	*presenteer* 'present'	.17	*foto* 'picture'	.33
museum 'museum'	.27	*laat* 'let'	.07	*sculptuur* 'sculpture'	.25
oeuvre 'oeuvre'	.22	*koop* 'buy'	.07	*aquarel* 'aquarelle'	.20
Kunsthal	.19	*bezit* 'own'	.06	*object* 'object'	.19
kunstenaar 'artist'	.15	*zie* 'see'	.05	*beeld* 'statue'	.12
dat 'that'	.12	*koop_aan* 'acquire'	.05	*overzicht* 'overview'	.12
hij 'he'	.10	*in huis heb* 'own'	.04	*portret* 'portrait'	.11

Quality count

- 44 dimensions contain similar, framelike semantics
- 43 dimensions contain less clear-cut semantics
  - single verbs account for one dimension
  - verb senses are mixed up
- 13 dimensions based on syntax rather than semantics
  - fixed expressions
  - pronomina

Methodology

- pseudo-disambiguation task to test generalization capacity (standard automatic evaluation for selectional preferences)

s	v	o	s'	o'
jongere 'youngster'	*drink* 'drink'	*bier* 'beer'	*coalitie* 'coalition'	*aandeel* 'share'
werkgever 'employer'	*riskeer* 'risk'	*boete* 'fine'	*doel* 'goal'	*kopzorg* 'worry'
directeur 'manager'	*zwaai* 'sway'	*scepter* 'sceptre'	*informateur* 'informer'	*vodka* 'wodka'

- 10-fold cross validation (± 300,000 co-occurrences)

Models

- Evaluation of 4 different models
- 2 matrix models
  $\rightarrow$ 1K verbs $\times$ (10K subjects + 10K direct objects)
  - singular value decomposition ($\mathbb{R}$)
  - non-negative matrix factorization ($\mathbb{R}_{\geq 0}$)
- 2 tensor models
  $\rightarrow$ 1K verbs $\times$ 10K subjects $\times$ 10K direct objects
  - parallel factor analysis ($\mathbb{R}$)
  - non-negative tensor factorization ($\mathbb{R}_{\geq 0}$)

Results

	dimensions		
	50 (%)	100 (%)	300 (%)
SVD	69.60 $\pm$ 0.41	62.84 $\pm$ 1.30	45.22 $\pm$ 1.01
NMF	81.79 $\pm$ 0.15	78.83 $\pm$ 0.40	75.74 $\pm$ 0.63
PARAFAC	85.57 $\pm$ 0.25	83.58 $\pm$ 0.59	80.12 $\pm$ 0.76
NTF	**89.52** $\pm$ 0.18	**90.43** $\pm$ 0.14	**90.89** $\pm$ 0.16

Conclusion

- novel method able to investigate three-way co-occurrences
- capable of automatically inducing selectional preferences
  - three-way methods improve on two-way methods
  - non-negativity constraint improves on unconstrained models
  - non-negative tensor factorization outperforms other models

Future work

- Further investigation of three-way selectional preference induction
  - more thorough quantitative evaluation (comparison to other methods)
  - include extra dependency relations
- Improvement of tensor factorization model
  - model minimizes sum of squared distance
  - kullback-leibler divergence usually better suited for language data
- Apply non-negative tensor factorization model to other three-way co-occurrences in NLP (word sense discrimination)