Multi-way Tensor Factorization for Unsupervised Lexical Acquisition

Tim Van de Cruys, Laura Rimell,
Thierry Poibeau & Anna Korhonen

University of Cambridge, UK

COLING
10 December 2012
Mumbai, India

Two subtasks

- Verb subcategorization frame induction
  - Induce number and type of arguments taken by a verb automatically from corpus data
  - Classify these into frames which describe syntactic behaviour
- Selectional preference induction
  - Predict likelihood of given lexical items occurring in an argument slot

Subcategorization frame induction

(1) [Our October review]$_{\text{NCSUBJ}}$ [shows]$_{\text{VERB}}$ [you]$_{\text{DOBJ}}$ [what's in store in next month's magazine]$_{\text{CCOMP}}$.

(2) [The blood test]$_{\text{NCSUBJ}}$ [showed]$_{\text{VERB}}$ [high blood calcium]$_{\text{DOBJ}}$.

- Example lexicon for verb show:
  - NCSUBJ-DOBJ (transitive)
  - NCSUBJ-DOBJ-CCOMP (direct object + *what*-question)

Challenge

- Adjuncts – such as temporal, locative, or manner modifiers – frequently appear in potential argument slots, giving rise to false analyses

(3) a. The doctor depended *on accurate information*.
b. The doctor worked *in the morning*.

- Co-occurrence frequency is a good cue for whether a verb takes a particular argument as part of an SCF
- But: many adjuncts are highly frequent

Selectional preference induction

(4) [Stalin]$_{\text{SUBJ}}$, who must have been well informed through his network of spies, [showed]$_{\text{VERB}}$ [no emotion]$_{\text{DOBJ}}$.

- the verb *show* with frame SUBJ-DOBJ is likely to prefer an inanimate direct object

Challenge

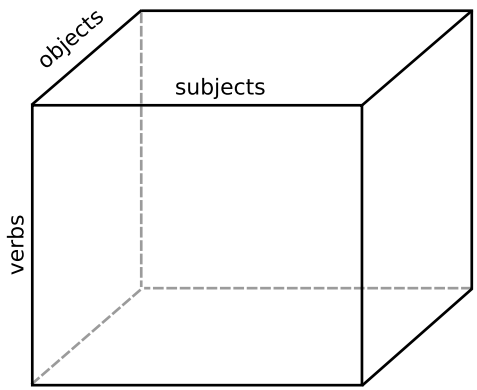
- Most work on selectional preference induction learns preferences:
  - for single argument slots
  - without taking the subcategorization frames into account
- We want to induce selectional preferences:
  - for all arguments to the verb *at the same time*
  - taking subcategorization into account

Approach

- Joint induction of subcategorization frames and selectional preferences
- Automatically learn *whether* a verb subcategorizes for particular argument slot, together with *which* lexical items occur in the slot
- Tensor factorization: generalization of matrix factorization
- Enables us to capture latent structure from multi-way co-occurrence frequencies

Tensor

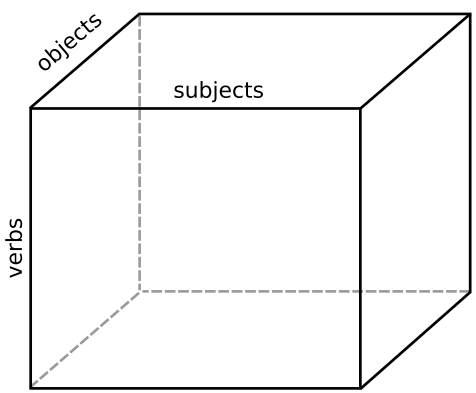
- Tensor is multi-way generalization of matrix
- Order of tensor = number of 'modes', i.e. indices needed to identify a cell
- E.g. 3-way tensor: verbs, subjects, objects



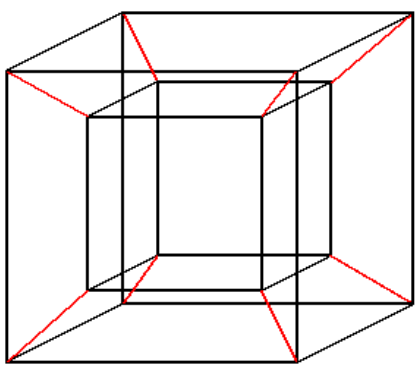
3 modes

Tensor

- Tensor is multi-way generalization of matrix
- Order of tensor = number of 'modes', i.e. indices needed to identify a cell
- E.g. 3-way tensor: verbs, subjects, objects



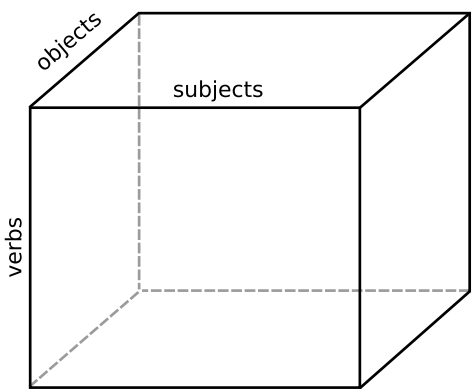
3 modes



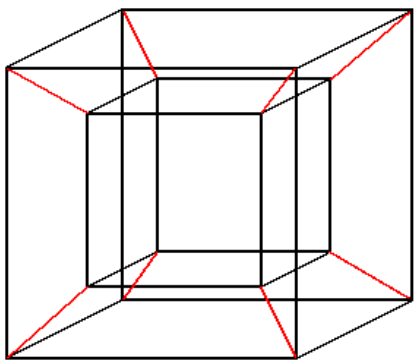
4 modes

Tensor

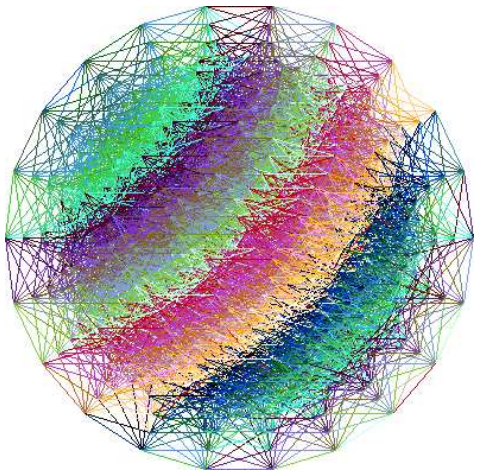
- Tensor is multi-way generalization of matrix
- Order of tensor = number of 'modes', i.e. indices needed to identify a cell
- E.g. 3-way tensor: verbs, subjects, objects



3 modes



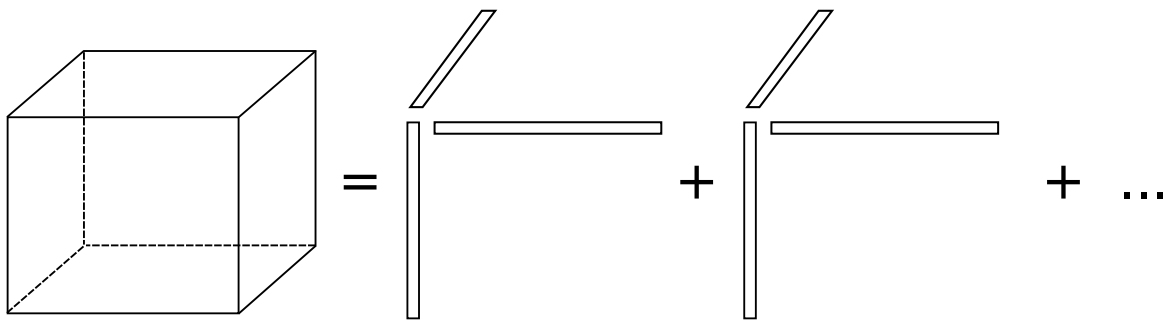
4 modes



10 modes

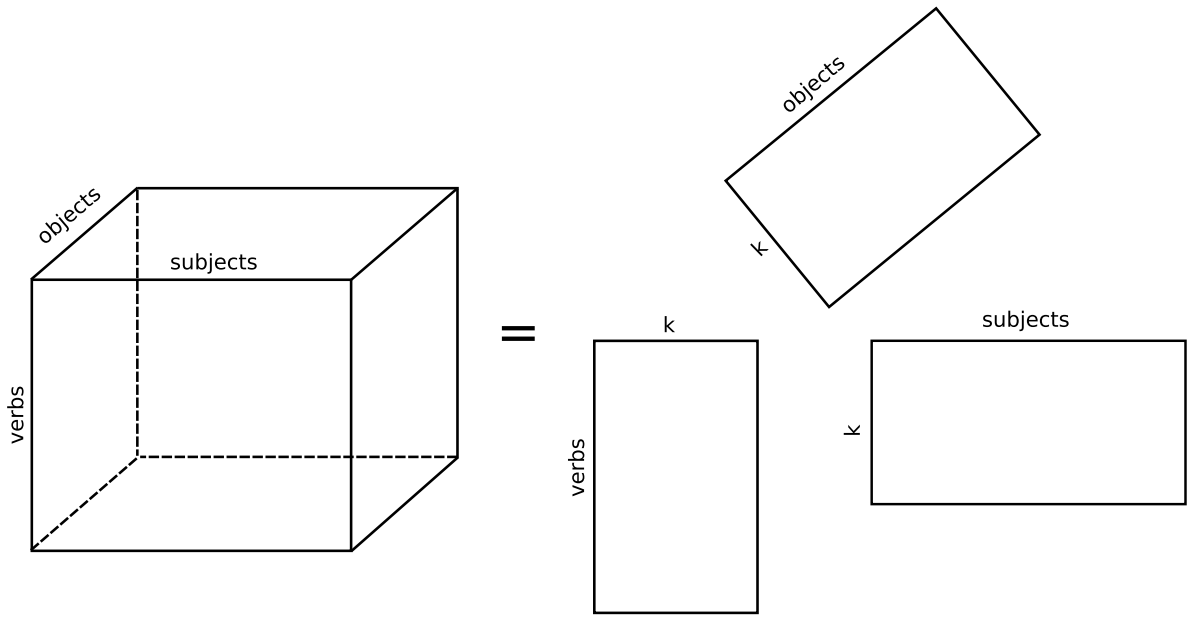
Non-negative Tensor Factorization

- Idea similar to non-negative matrix factorization
- Calculations are different
- $\min_{\mathbf{x}_i \in \mathbb{R}^{D1}_{\geq 0}, \mathbf{y}_i \in \mathbb{R}^{D2}_{\geq 0}, \mathbf{z}_i \in \mathbb{R}^{D3}_{\geq 0}} \| \ T - \sum_{i=1}^{k} \mathbf{x}_i \circ \mathbf{y}_i \circ \mathbf{z}_i \ \|_F^2$



Non-negative Tensor Factorization

- For an n-mode tensor, factorization results in n matrices, indicating the loadings of each mode on the factorized dimensions



Tensor construction

- An 9-mode tensor is constructed
  - One mode for verbs
  - One mode for eight different types of grammatical relations

GR	relation between verb and ...	example
NCSUBJ	non-clausal subject	**Edward** *eats*.
DOBJ	NP immediately right of the verb	*Edward eats* **an apple**.
OBJ2	second NP in a double object construction	*Edward gave her* **an apple**.
IOBJ	PP when PP complement is an NP	*Edward gave the apple* **to the girl**.
PCOMP	PP when PP complement is itself a PP	*Edward leads* **by between 1 and 8 points**.
XCOMP	non-finite VP complement	*Edward thought* **about eating**.
CCOMP	finite clausal complement	*Edward thought* **about him eating**.
NCMOD	non-clausal modifier	*Edward ate* **yesterday**.

- 8 different GR types selected based on frequency and relevance for subcategorization

Tensor construction

- Lexical head of the argument is used as a feature within each mode
- For a particular verb in a particular sentence, not every GR type will be instantiated, so we introduce an empty *void* (–) feature

(5) [Our October review]$_{\text{NCSUBJ}}$ [shows]$_{\text{VERB}}$ [you]$_{\text{DOBJ}}$ [what's in store in next month's magazine]$_{\text{CCOMP}}$.

$\rightarrow \langle show_V, review_N, you_P, –, –, –, –, be_V, – \rangle$

Tensor construction

- Size of 9-mode tensor: 1993 x 2351 x 1980 x 81 x 679 x 51 x 1396 x 561 x 920
- Factorized to 150 latent dimensions

Corpus Data

- Data from five large cross-domain corpora
- Only verbs with at least 500 occurrences, for a total of 1993 verbs
- Raw data tokenized, POS-tagged, lemmatized, and parsed with RASP
- RASP uses a tag-sequence grammar and is unlexicalized, so that the parser's lexicon does not interfere with SCF acquisition

SCF induction: evaluation framework

- Evaluation against gold standard of Korhonen et al. (2006) (manual annotation of 200-250 sentences for 183 verbs)
- Coarse-grained gold standard of 183 verbs, with an average of 5.5 SCFs per verb
- Evaluated in terms of precision/recall

Mapping Dimensions to SCFs

- Use scores for verbs and arguments appearing in each tensor mode for each dimension
- We also have a score for *void* argument
- Each dimension mapped to a single SCF by considering all GR slots with *void* $< \theta_{\text{void}}$
- For each dimension, all verbs with a score above θ_{verb} in this dimension are considered to take this SCF
- Simplified example:

verb		dobj		obj2		ccomp	
ask	0.004	someone	0.3	something	0.5	—	0.5
tell	0.003	you	0.25	secret	0.3	warn	0.00001
say	0.0001	her	0.22	—	0.06	do	0.00001

SCF induction: results

	P	R	F
Baseline	86.3	23.5	36.9
POS features	53.1	83.3	64.8
Final system	61.0	78.5	68.7

- Baseline: assign two most frequent SCFs (transitive and intransitive) to all verbs
- POS features: only part of speech, no lexical items
- Direct comparison with supervised methods is difficult (different data and frame inventories), but best current methods reach ceiling at $\pm$ 70% F-measure

SP induction: evaluation framework

- standard pseudo-disambiguation task
- for tuple from held-out corpus, construct all random instances with one or several arguments substituted
  - $\langle show_V, rabbit_N, you_P, –, –, –, be_V, –, – \rangle$
- Compute selectional preference values according to model
- Model needs to prefer actual tuple over random instances

SP induction: results

	accuracy (%)
baseline	29.21 ± .08
NMF	69.71 ± .28
NTF	**77.78** ± .17

- Baseline: random model
- NMF: standard non-negative matrix factorization model (similar to Rooth et al. (1999))

Transitive frames

dim 29 buy, sell, use, collect, produce, handle, remove, purchase, obtain, eat, ...

DOBJ: inanimate objects, goods (*thing, material, food, ...*)

dim 38 kill, love, see, like, marry, know, meet, visit, help, say, ...

DOBJ: animate objects (pronouns, *man, child, woman, ...*)

dim 44 examine, identify, see, consider, assess, investigate, discuss, study, determine, explore, ...

DOBJ: abstract objects (*effect, extent, nature, difference, ...*)

Frames with PP argument

dim 91	go, come, return, move, walk, get, run, rush, travel, fly, ...
dim 122	talk, speak, listen, write, belong, happen, appeal, come, say, lie, ...
dim 123	look, stare, smile, laugh, shout, gaze, glance, glare, grin, scream, ...

Conclusion

- Novel method for fully unsupervised lexical acquisition
- Joint induction of subcategorization frames and selectional preferences
- Application of non-negative tensor factorization to multi-way co-occurrence tensor of verbs and their arguments allows us to cluster them according to their syntactic and semantic behaviour

Future Work

- More data to improve performance on rarer GRs/SCFs
- Extend model to acquire finer-grained SCFs
- Produce a comprehensive lexical resource that supplements SCFs with selectional preference and verb class information