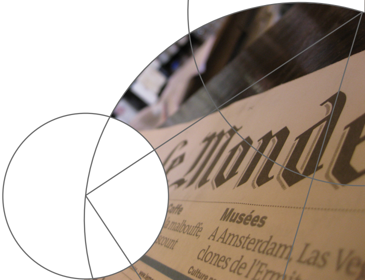
A Tensor-based Factorization Model of Semantic Compositionality

Tim Van de Cruys
CNRS & IRIT, France

18 January 2013



Introduction

- principle of semantic **compositionality** [Frege 1892]

meaning of a complex expression	=	meaning of its parts	+	the way those parts are combined

- fundamental principle that allows people to understand sentences they have never heard before
- word interactions are important for the construction of meaning

(1) Jack is listening to a **record**
(2) Jill updated the **record**

Introduction

- principle of semantic **compositionality** [Frege 1892]

meaning of a complex expression	=	meaning of its parts	+	the way those parts are combined

- fundamental principle that allows people to understand sentences they have never heard before
- word interactions are important for the construction of meaning

(1) Jack is listening to a **record**
(2) Jill updated the **record**

Introduction

- principle of semantic **compositionality** [Frege 1892]

meaning of a complex expression	=	meaning of its parts	+	the way those parts are combined

- fundamental principle that allows people to understand sentences they have never heard before
- word interactions are important for the construction of meaning

(1) Jack is listening to a **record**
(2) Jill updated the **record**

Introduction

- principle of semantic **compositionality** [Frege 1892]

meaning of a complex expression	=	meaning of its parts	+	the way those parts are combined

standard distributional model

- fundamental principle that allows people to understand sentences they have never heard before
- word interactions are important for the construction of meaning

(1) Jack is listening to a **record**
(2) Jill updated the **record**

Introduction

- principle of semantic **compositionality** [Frege 1892]

meaning of a complex expression	=	meaning of its parts	+	the way those parts are combined

standard distributional model

tensor-based factorization model

- fundamental principle that allows people to understand sentences they have never heard before
- word interactions are important for the construction of meaning

(1) Jack is listening to a **record**

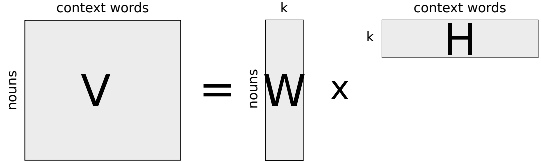
(2) Jill updated the **record**

Introduction

- model for joint composition of verb with subject and direct object
- allows us to compute semantic similarity between simple transitive sentences
- key idea: compositionality is modeled as a multi-way interaction between latent factors, automatically constructed from corpora
- implemented using tensor algebra

Step 1: construction of latent noun factors

- Construction of a latent model for nouns using non-negative matrix factorization

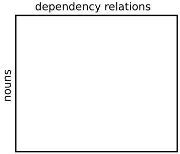


Step 1: example noun factors (*k*=300)

dim 60	**dim 88**	**dim 89**	**dim 120**
rail	journal	filename	bathroom
bus	book	null	lounge
ferry	preface	integer	bedroom
train	anthology	string	kitchen
freight	author	parameter	WC
commuter	monograph	String	ensuite
tram	article	char	fireplace
airport	magazine	boolean	room
Heathrow	publisher	default	patio
Gatwick	pamphlet	int	dining

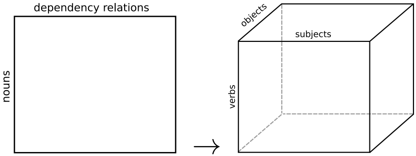
Step 2: Modeling multi-way interactions

- Standard distributional similarity methods model two-way interactions $\rightarrow$ matrix
  - words $\times$ context words
  - words $\times$ dependency relations
- not suitable for multi-way interactions
  - nouns $\times$ adjectives $\times$ context words
  - **verbs** $\times$ **subjects** $\times$ **objects**



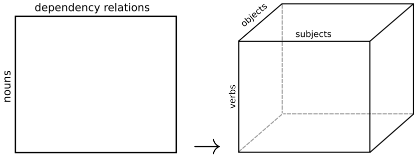
Step 2: Modeling multi-way interactions

- Standard distributional similarity methods model two-way interactions $\rightarrow$ matrix
  - words $\times$ context words
  - words $\times$ dependency relations
- not suitable for multi-way interactions $\rightarrow$ tensor
  - nouns $\times$ adjectives $\times$ context words
  - **verbs** $\times$ **subjects** $\times$ **objects**



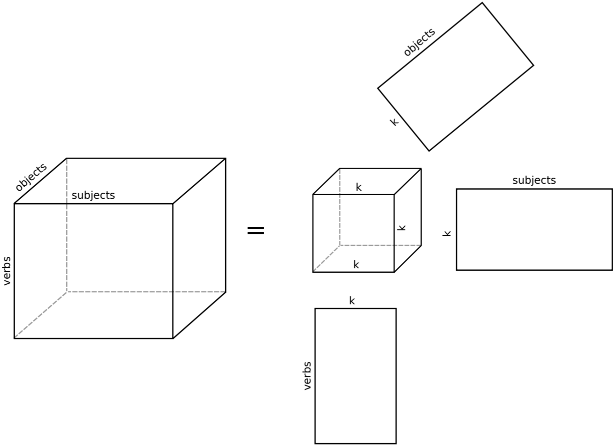
Step 2: Modeling multi-way interactions

- Standard distributional similarity methods model two-way interactions $\rightarrow$ matrix
  - words $\times$ context words
  - words $\times$ dependency relations
- not suitable for multi-way interactions $\rightarrow$ tensor
  - nouns $\times$ adjectives $\times$ context words
  - **verbs** $\times$ **subjects** $\times$ **objects**



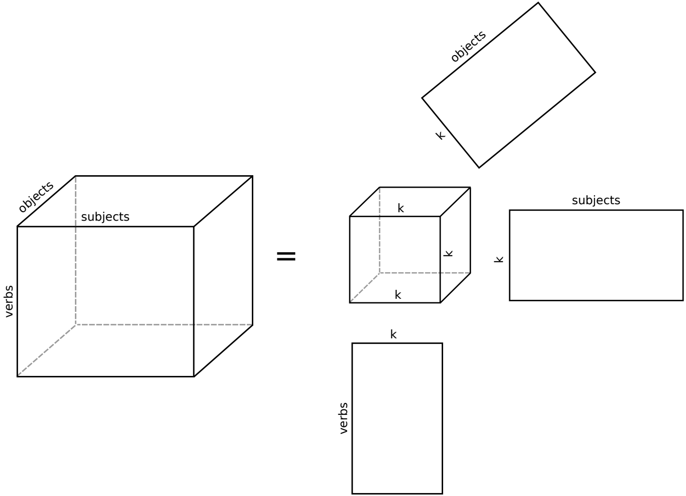
$\rightarrow$ build a latent model of multi-way interactions

Step 2: Non-negative Tucker decomposition

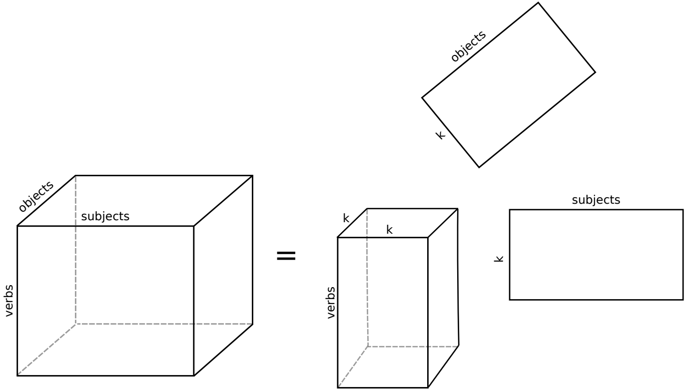


$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}$$

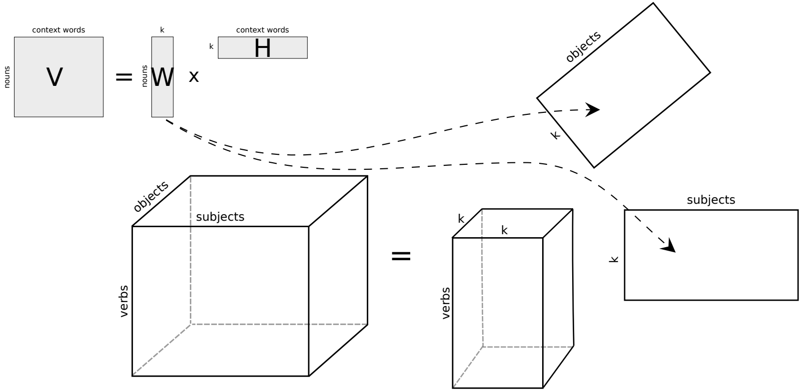
Step 2: Reconstructing a Tucker model from two-way factors



Step 2: Reconstructing a Tucker model from two-way factors

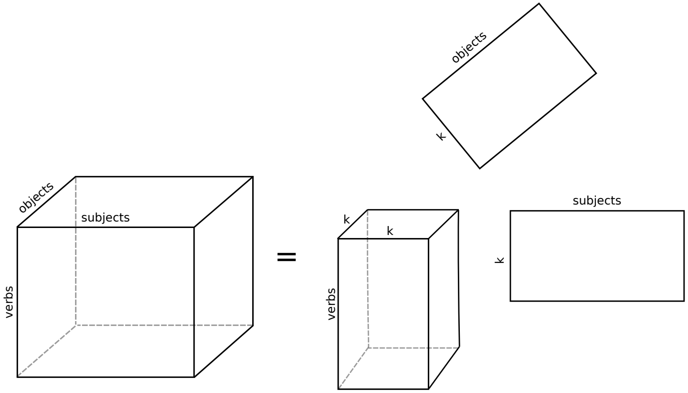


Step 2: Reconstructing a Tucker model from two-way factors

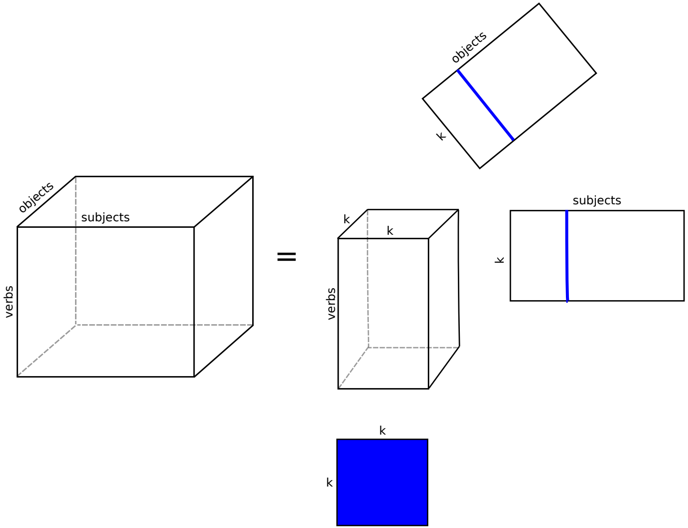


$$\mathcal{G} = \boldsymbol{\mathcal{X}} \times_2 \mathbf{W}^T \times_3 \mathbf{W}^T$$

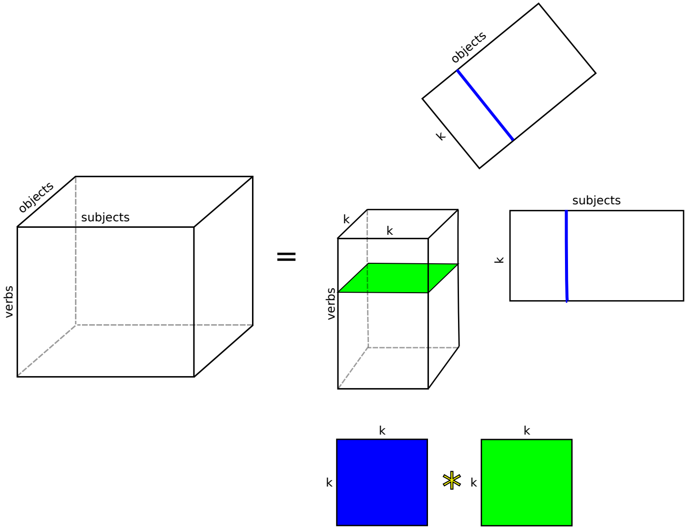
Step 3: composition of *svo* triples



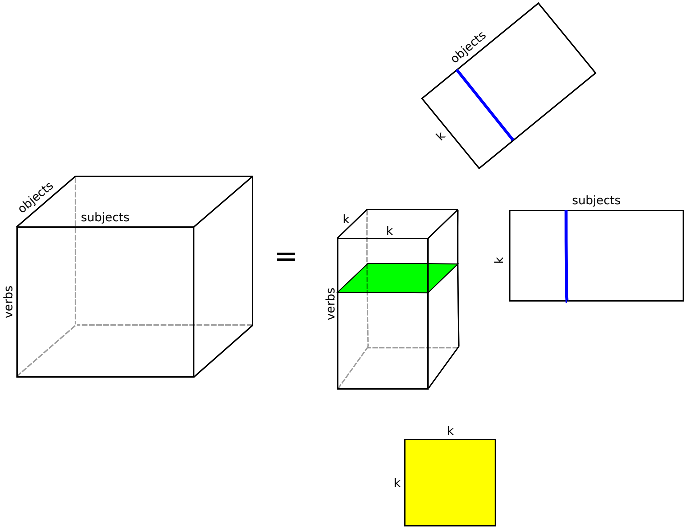
Step 3: composition of *svo* triples



Step 3: composition of *svo* triples



Step 3: composition of *svo* triples



Example

- *athlete runs race*
  - $\mathbf{Y}_{\langle athlete,race \rangle} = \mathbf{v}_{athlete} \circ \mathbf{u}_{race}$
  - $\mathbf{Z}_{run,\langle athlete,race \rangle} = \mathbf{G}_{run} * \mathbf{Y}_{\langle athlete,race \rangle}$
  - *finish* (.29), *attend* (.27), *win* (.25)
- *user runs command*
  - $\mathbf{Y}_{\langle user,command \rangle} = \mathbf{v}_{user} \circ \mathbf{u}_{command}$
  - $\mathbf{Z}_{run,\langle user,command \rangle} = \mathbf{G}_{run} * \mathbf{Y}_{\langle user,command \rangle}$
  - *execute* (.42), *modify* (.40), *invoke* (.39)
- *man damages car*
  - *crash* (.43), *drive* (.35), *ride* (.35)
- *car damages man*
  - *scare* (.26), *kill* (.23), *hurt* (.23)

Implementational details

- matrix $\mathbf{V}$: 10K *nouns* $\times$ 2K *contexts*
- tensor $\boldsymbol{\mathcal{X}}$: 1K *verbs* $\times$ 10K *subject* $\times$ 10K *objects*
- *noun*, *subject* and *object* modes contain exactly the same features
- frequency information extracted from UKWAC (web corpus of 2 billion words), parsed with MaltParser
- weighted with (extended) pointwise mutual information
- k=300 latent factors

Evaluation: methodology

- sentence similarity task for transitive sentences
- 25 participants, 2500 similarity judgements
- correlation of model's judgements with human judgements calculated using Spearman's ρ

p	**target**	**subject**	**object**	**landmark**	**sim**
19	meet	system	criterion	visit	1
21	write	student	name	spell	6

- comparison with multiplicative [Mitchell and Lapata 2008] and categorical [Grefenstette and Sadrzadeh 2011] model

Evaluation: results

model	contextualized	non-contextualized
baseline	.23	
multiplicative	**.32**	.34
categorical	**.32**	.35
latent	**.32**	**.37**
upper bound	.62	

Conclusion

- novel method for the computation of compositionality within a distributional framework
- compositionality is modeled as a multi-way interaction between latent factors, automatically constructed from corpora
- evaluation on similarity task for transitive sentences, in which it exceeds the state of the art

Future work

- extend framework to incorporate more compositional phenomena
- explore the possibilities of a model in which all three modes are represented by latent factors
- evaluate method using different evaluation frameworks (lexical substitution and paraphrasing)

Marco Baroni and Roberto Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1183–1193, Cambridge, MA.

B. Coecke, M. Sadrzadeh, and S. Clark. 2010. Mathematical foundations for a compositional distributed model of meaning. *Lambek Festschrift, Linguistic Analysis, vol. 36*, 36.

Gottlob Frege. 1892. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, 100:25–50.

Edward Grefenstette and Mehrnoosh Sadrzadeh. 2011. Experimenting with transitive verbs in a discocat.*Proceedings of the GEMS 2011 Workshop on GEometrical Models of Natural Language Semantics*, pages 62–66, Edinburgh, UK.

Jeff Mitchell and Mirella Lapata. 2008. Vector-based models of semantic composition. *proceedings of ACL-08: HLT*, pages 236–244.

Tim Van de Cruys. 2010. A non-negative tensor factorization model for selectional preference induction. *Journal of Natural Language Engineering*, 16(4):417–437.

Tim Van de Cruys, Thierry Poibeau, and Anna Korhonen. 2011. Latent vector weighting for word meaning in context. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pages 1012–1022, Edinburgh, Scotland, UK.