A Structured Vector Space Model for Word Meaning in Context

Katrin Erk
Sebastian Pado

EMNLP 2008

Presentation by Simon Ahrendt

Outline

- Introduction
- The Structured Vector Space (SVS) Model
  - General Formulation
  - Parametrization
- Evaluation
  - Mitchell & Lapata (2008) Dataset
  - SemEval (lexical substitution) Dataset
- Conclusion / Discussion

Introduction

- Vector space models are popular for modelling word meaning in many NLP applications
- Meaning in context differs from one occurence to another (WSD):

  catch a ball ↔ *attend a ball*

- Problem of **taking syntax into account:**

  a horse draws ↔ *draw a horse*

- Single vectors can not represent phrases
- Both issues are adressed in this paper

Structured Vector Space Model

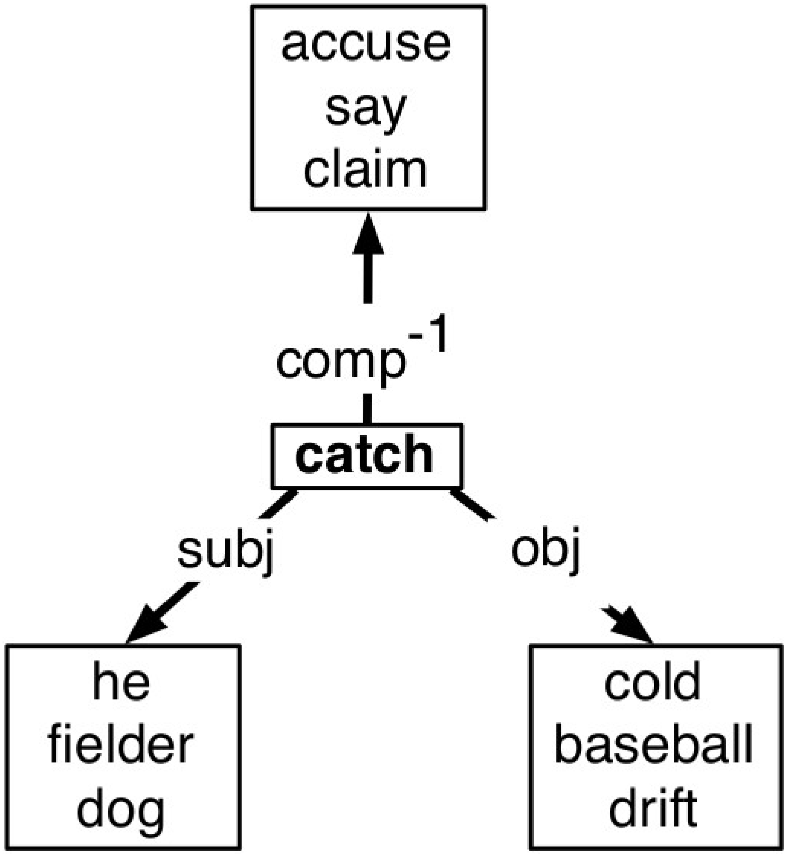
- Intuition:

  Syntax-aware expectations affect word meanings in context

- Word meaning:
  - Not only a **single vector *v*** but also
  - **Selectional preferences *R*, R^{-1}** mapping syntactic relations to a preference vector

$$w = (v, R, R^{-1})$$

Example: Word Meaning



The SVS Model

- Intuition:

  Syntax-aware expectations affect word meanings in context

- Meaning in context:
  - Combining the **lemma vector** with the **selectional preference vector**

The SVS Model

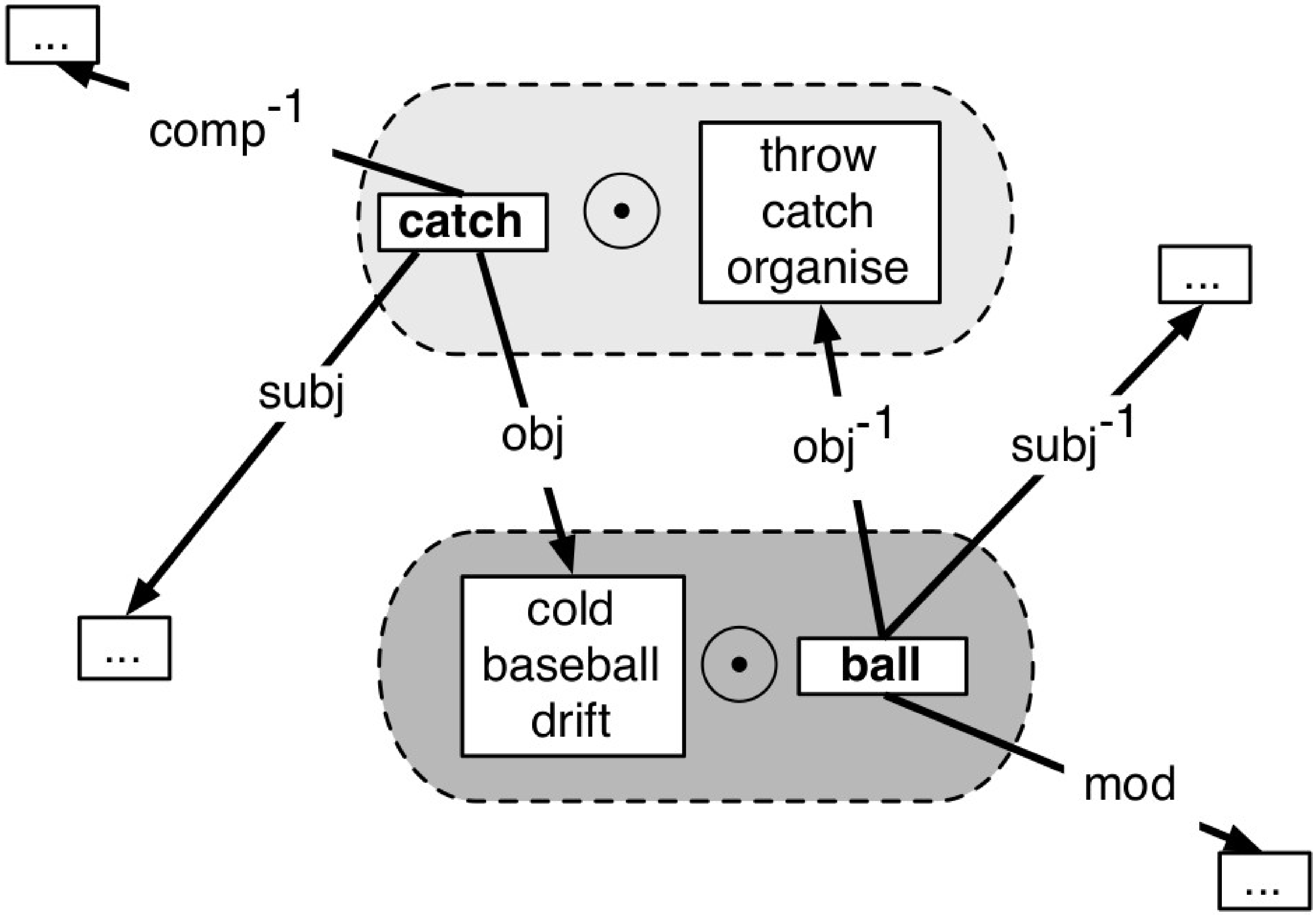
- Intuition:

  Syntax-aware expectations affect word meanings in context

- Meaning in context:
  - Combining the **lemma vector** with the **selectional preference vector**

$$
\begin{aligned}
a' &= \left(v_a \odot R_b^{-1}(r), R_a - \{r\}, R_a^{-1}\right) \\
b' &= \left(v_b \odot R_a(r), R_b, R_b^{-1} - \{r\}\right)
\end{aligned}
$$

Example: Contextualized Meaning



Parametrization

- Vector space(s):
  - BOW-space (following M&L 2008)
    - → Based on BNC, 5 words on either side, 2000 context words as dimensions
  - SYN-space
    - → Context words only count if they are syntactically related
- Set of relations:
  - Dependency relations
- Vector composition function:
  - Component-wise multiplication

Parametrization (2)

- Selectional preference vectors:
  - Weighted centroid of seen fillers:

$$R_b(r)_{\text{SELPREF}} = \sum_{a:f(a,r,b)>0} f(a,r,b) \cdot \vec{v}_a$$

  - One varation with a threshold for fillers
  - Another variation with potentiation:

$$R_b(r)_{\text{SELPREF-POW}} = \langle v_1^n, \ldots, v_m^n \rangle$$

→ To reduce noise from infrequent dimensions, e.g. because we are looking at the centroid

Evaluation: General

- Task: Predicting the appropiateness of paraphrases given a predicate-argument pair
- E.g. *draw a horse* → *pull a horse*?
- 2 Datasets:
  - The synthetic M&L 2008 dataset
  - A subset of the SemEval lexical substitution dataset, with a broader range of paraphrases

Evaluation: Baselines

- Target only:
  - Using only the target vector (e.g. *draw*)
- Selpref only:
  - Using only the selpref-vector
- Mitchell & Lapata 2008:
  - Component-wise multiplication

Evaluation A: M&L Dataset

- 120 items, each consisting of
  - A intransitive verb (target)
  - A subject noun
  - 2 landmarks (paraphrase candidates)
  - Human judgements of the appropiateness of the landmarks (ranging from 1 to 7)
- The models use cosine to compute similarity to the landmarks

Evaluation A: Results

Model	high	low	ρ
	Bow space		
Target only	0.32	0.32	0.0
Selpref only	0.46	0.4	0.06**
M&L	0.25	0.15	0.20**
SELPREF	0.32	0.26	0.12**
SELPREF-CUT, θ=10	0.31	0.24	0.11**
SELPREF-POW, n=20	0.11	0.03	**0.27****
Upper bound	–	–	0.4
	Syn space		
Target only	0.2	0.2	0.08**
Selpref only	0.27	0.21	0.16**
M&L	0.13	0.06	**0.24****
SELPREF	0.22	0.16	0.13**
SELPREF-CUT, θ=10	0.2	0.13	0.13**
SELPREF-POW, n=30	0.08	0.04	0.22**
Upper bound	–	–	0.4

Evaluation A: Results

Model	lex. vector	obj^{-1} selpref
SELPREF	0.23 (0.09)	0.88 (0.07)
SELPREF-CUT (10)	0.20 (0.10)	0.72 (0.18)
SELPREF-POW (30)	0.03 (0.08)	0.52 (0.48)

Evaluation B: Ranking Paraphrases

- Task: Ranking paraphrase candidates for words in sentential contexts
- Dataset: Constructed from SemEval lexical substitution dataset, with gold standard substitutes

Sentence	Substitutes
By asking people who **work** there, I have since determined that he didn't. (# 2002)	be employed 4; labour 1
Remember how hard your ancestors **worked**. (# 2005)	toil 4; labour 3; task 1

Evaluation B: Ranking Paraphrases

- Only a single context word is used for prediction
- Three test sets:
  - Target intransitive verbs, noun subjects (48)
  - Target transitive verbs, noun objects (213)
  - Target nouns, as objects of verbs (102)
- „Out of ten“ precision metric:

$$\mathrm{P}_{\mathrm{OOT}} = 1/|I| \sum_i \frac{\sum_{s \in M_i \cap G_i} f(s,i)}{\sum_{s \in G_i} f(s,i)}$$

Evaluation B: Results

Model	V-SUBJ	V-OBJ	N-OBJ
Target only	47.9	47.4	49.6
Selpref only	54.8	51.4	55.0
M&L	50.3	52.0	53.4
SELPREF-POW, n=30	**63.1**	**55.8**	**56.9**

Conclusion / Discussion

- It is important to consider syntactic information for semantic representations
- Erk & Pado propose a framework to incorporate this information in VSM's
- Evaluating on paraphrase detection shows the benefits of this approach
- Possible question:

  Is it enough to contextualize word meanings or do we need more to represent phrases?