

A Structured Vector Space Model for Word Meaning in Context

von Katrin Erk und Sebastian Padó (2008)

präsentiert von Frederik Arnold

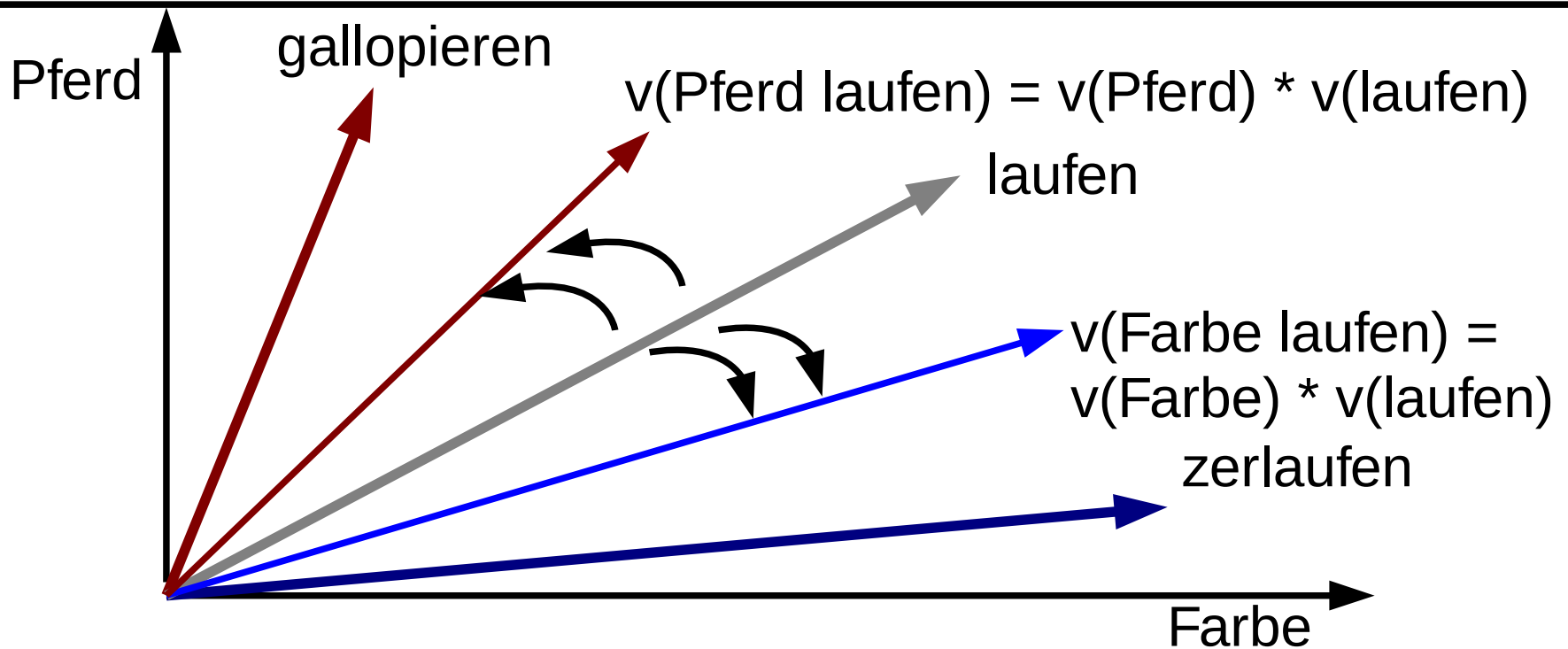
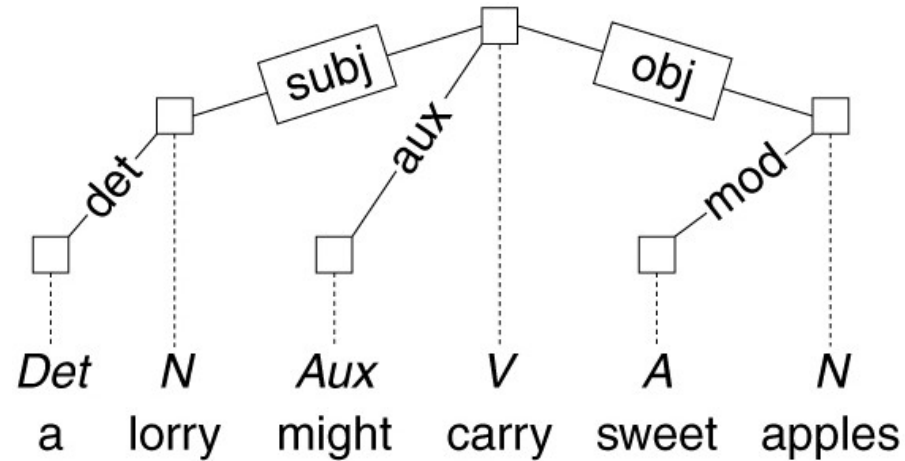
Was bisher geschah ...

- Wortbedeutung als Vektor aus Kookkurrenz mit Nachbarwörtern
- Vergleich verschiedener Modelle der Vektorkomposition (M&L 2008)
 - Multiplikationsmodell schneidet am besten ab
- Problem:
 - Syntaktische Relationen werden nicht beachtet
 - a horse draws vs. draw a horse
 - Skalierbarkeit
 - kann ein einzelner Vektor alle Informationen eines Satzes kodieren?

Was bisher geschah ...

- Dependenzbasierter Vektorraum
 - mehr als nur „einfaches“ Auszählen der Wörter rechts und links vom Zielwort
 - Zielwort und Kontext müssen in einer bestimmten Verbindung stehen
- Syntaktische Information soll bei der Erstellung des Vektorraumes eine Rolle spielen

Was bisher geschah ...



Ein bisschen Geschichte

- Schütze (1998)
 - So viel Kontext wie möglich
- Tensorprodukt (Smolensky, 1990)
 - Zwei Vektoren a und b werden zu einem dritten Vektor c kombiniert
 - jedes Wort fügt neue Dimensionen hinzu
 - Modell wird unbrauchbar in der praktischen Anwendung
- Kernelmethoden (Moschitti and Quarteroni, 2008)
 - anderer Fokus

And now ...

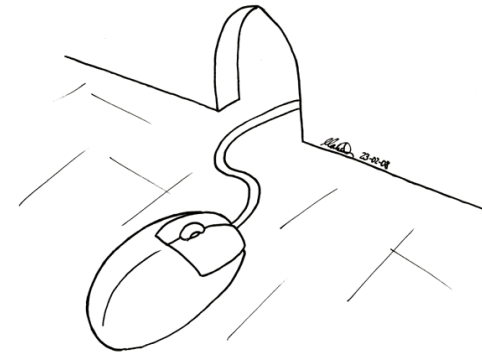
füttern

die Parkuhr



- bezahlen
- Geld
- Strafzettel

die Maus



© www.toonsup.com/ron

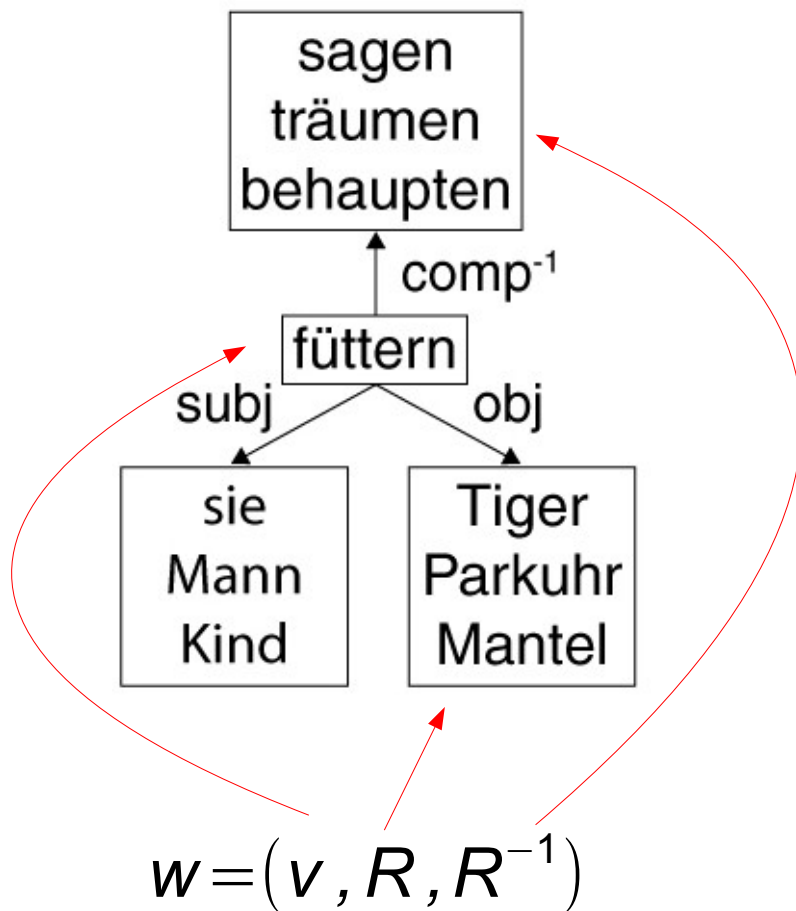
- klicken
- streicheln
- süß

A structured vector space model und die Selektionspräferenzen

- Worte/Phrasen lösen Erwartungen aus
- Motiviert auf linguistischer wie auch kognitiver Ebene
 - es existieren verschiedene Studien
 - McRae et al., 1998, Ferretti et al. 2003
- SVS Modell soll diese Erwartungen modellieren
- Wort wird dargestellt durch ein Tripel:
 - Vektor des Wortes selber
 - Zwei Vektoren, die Listen von Selektionspräferenzen repräsentieren

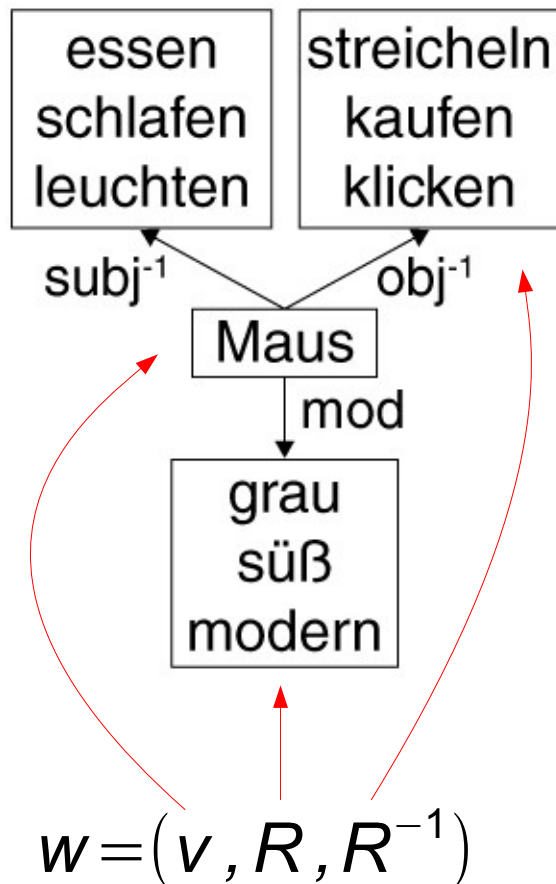
SVS

Darstellung eines Wortes



- Die schwarzen Pfeile verbinden das Wort mit
 - bevorzugten Subjekten (subj)
 - bevorzugten Objekten (obj)
 - Verben zu denen das Wort als Komplement erscheint (comp⁻¹)
- Kasten in der Mitte ist der Vektor des Wortes selber
- Die Listen von Wörtern werden durch jeweils einen Vektor dargestellt

Darstellung eines Wortes



- Die schwarzen Pfeile verbinden das Wort mit
 - bevorzugten Modifikatoren (mod)
 - Verben von denen „Maus“ Subjekt ist (subj^{-1})
 - Verben von denen „Maus“ Objekt ist (obj^{-1})

Selektionspräferenzen

SELPREF

$$R_b^{-1}(r) = \sum_{a: f(a, r, b)} f(a, r, b) * \vec{v}_a$$

Strand an v(liegen), v(lesen), v(schlafen), v(spielen), ...

- r : Stellt eine syntaktische Beziehung dar
- $f(a, r, b)$: Häufigkeit mit der a und b , wenn sie in Beziehung r zueinander stehen, auftreten

Selektionspräferenzen

SELPREF - Beispiel

$$R_b^{-1}(r) = \sum_{a: f(a,r,b)} f(a,r,b) * \vec{v}_a$$

$$\begin{aligned} R_{Strand}(an) &= f(\text{liegen}, an, Strand) * v(\text{liegen}) + \\ &\quad f(\text{lesen}, an, Strand) * v(\text{lesen}) + \\ &\quad f(\text{schlafen}, an, Strand) * v(\text{schlafen}) \\ &= 1 * \langle 0, 1, 0, 2, 1, 0, 0 \rangle + \\ &\quad 1 * \langle 2, 1, 0, 1, 1, 2, 2 \rangle + \\ &\quad 1 * \langle 1, 2, 1, 2, 1, 0, 1 \rangle = \\ &\quad \langle 3, 4, 1, 5, 3, 2, 3 \rangle \end{aligned}$$

	Bibliothek	Strand	Handtuch	Bett	Sofa	Buch	Bus
liegen	0	1	0	2	1	0	0
lesen	2	1	0	1	1	2	2
schlafen	1	2	1	2	1	0	1

Selektionspräferenzen

SELPREF-CUT

$$R_b(r) = \sum_{a: f(a,r,b) > \theta} f(a,r,b) * \vec{v}_a$$

- Nur Platzhalter ab einer bestimmten Häufigkeit
- Rauschen soll gemindert werden durch Eliminierung von seltenen Platzhalter

Selektionspräferenzen

SELPREF-POW

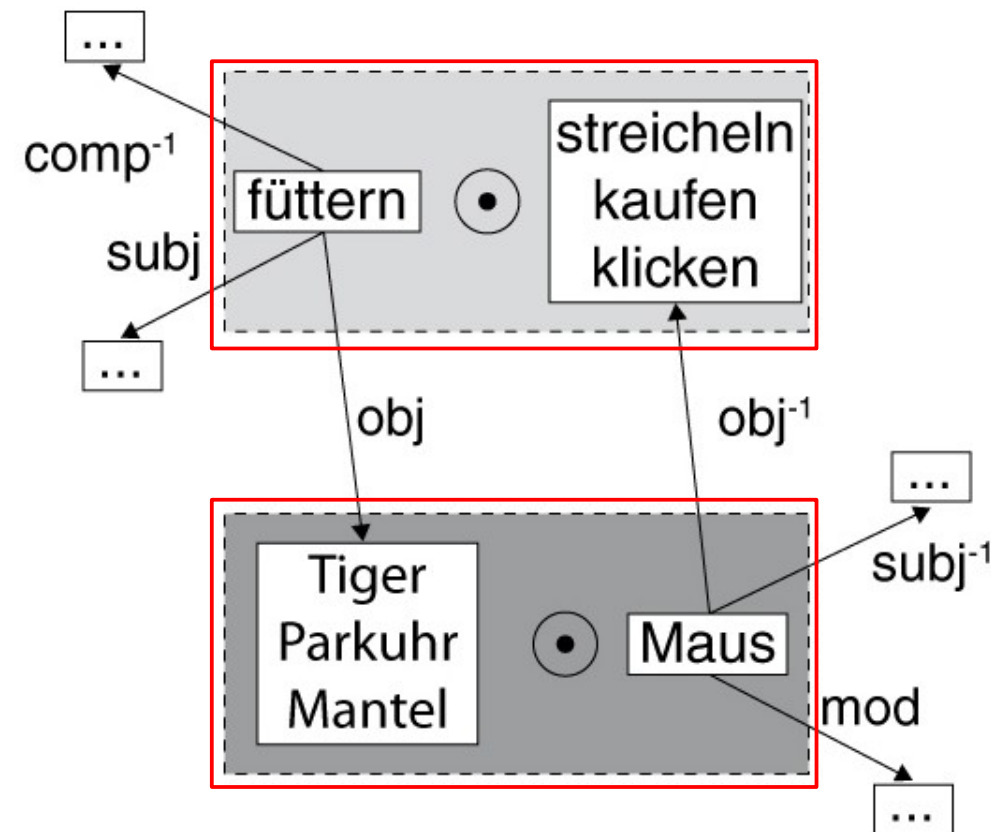
Wenn $R_b(r)_{SELPREF} = \langle v_1, \dots, v_m \rangle$

Dann: $R_b(r) = \langle v_1^n, \dots, v_m^n \rangle$

- Jede Komponente des Vektors wird mit n potenziert
- Größere Zahlen sollen mehr Gewicht bekommen als kleinere Zahlen

SVS

Vektorkomposition



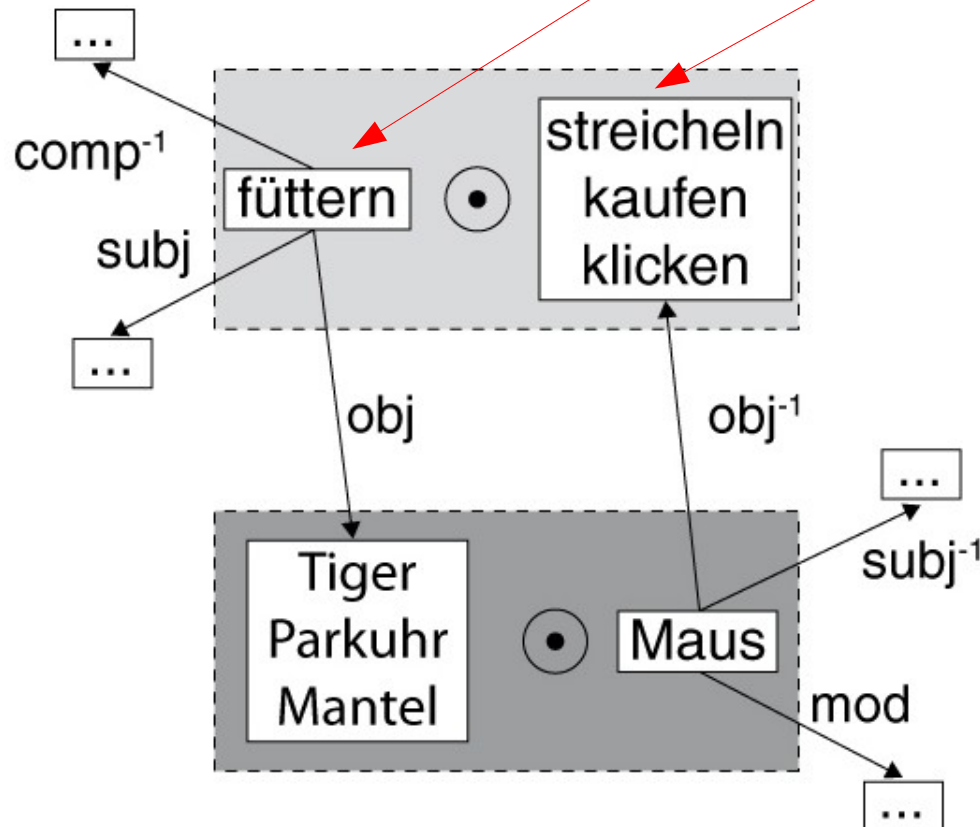
- hier betrachten wir „füttern“ im Kontext von „Maus“
- Vektor von „füttern“ wird kombiniert mit den inversen Objektpräferenzen von „Maus“
- Vektor von „Maus“ wird kombiniert mit den Objektpräferenzen von „füttern“

⊙ : Multiplikation oder Addition

SVS

Vektorkomposition

$$a' = (v_a \odot R_b^{-1}(r), R_a - \{r\}, R_a^{-1})$$



- a' ist die Bedeutung von a im Kontext von b
- a ist „füttern“, b ist „Maus“

Evaluation allgemein

Das SVS Modell **vs.** **das menschliche Sprachgefühl und Mitchell und Lapata**

In folgenden Disziplinen

1. Experiment

- Mitchell und Lapatas Datensatz (M&L, 2008)
 - 2 Vektorräume (BOW und SYN)

2. Experiment

- SemEval-1 lexical substitution dataset (McCarthy and Navigini, 2007)

Im Kampf um die beste
Einschätzung
von Ähnlichkeit im Kontext !

Experiment 1

Datensatz

- 120 Paare
 - jeweils intransitives Verb und Nomen
 - nur Paare mit hoher Varianz
- aus dem British National Corpus
- für jedes Paar
 - zwei Synonyme des Verbs
 - jeweils eins kompatibel und eins nicht
- Bewertung durch Menschen auf einer Skala von 1-7

	Nomen	Referenz	Hoch	Tief
Das	Handy	brummt	vibriert	knurrt
Der	Bär	brummt	knurrt	vibriert

Experiment 1

Vektorräume

BOW

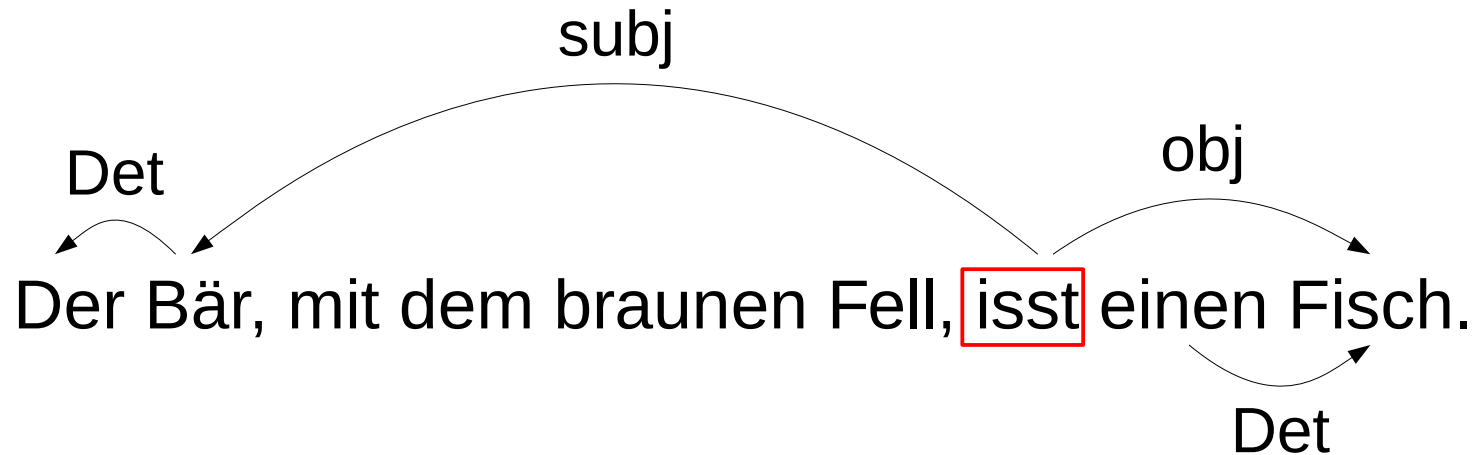
- bag-of-words
- Kookkurrenzen in einem Fenster von 10 Wörtern
- aus dem British National Corpus (BNC)
- 2000 häufigsten Kontextwörter als Dimensionen

SYN

- Dependenzbasierter Vektorraum
- Wort und Kontext müssen verbunden sein durch einen gültigen Pfad
- aus dem British National Corpus (BNC)
- Minipar

Experiment 1

SYN erklärt



	Bär	braunen	Fell	einen	Fisch
v(essen)	1	0	0	0	1

Experiment 1

Spearsman's ρ

- Ein Maß für die Ähnlichkeit
- Erzeugt Werte zwischen 1 und -1
 - in unserem Fall: je näher an 1, desto höher die Ähnlichkeit
- Obere Grenze von $\rho = 0.4$
 - ergibt sich aus den menschlichen Bewertungen
 - Vergleichswert für die Modelle

Experiment 1

Auswertung - BOW

<i>Model</i>	ρ
Target only	0.0
Selpref only	0.06
M&L	0.20
SELPREF	0.12
SELPREF-CUT, $\theta=10$	0.11
SELPREF-POW, $n=30$	0.27
Upper bound	0.4

- M&L deutlich besser als beide Baselines
- Selpref only alleine hat Ähnlichkeit zur menschlichen Bewertung
- Alle SELPREF Varianten schneiden gut ab
- SELPREF-POW schneidet am besten ab

Experiment 1

Auswertung - SYN

<i>Model</i>	ρ
Target only	0.08
Selpref only	0.16
M&L	0.24
SELPREF	0.13
SELPREF-CUT, $\theta=10$	0.13
SELPREF-POW, $n=30$	0.22
Upper bound	0.4

- SELPREF und SELPREF-CUT niedriger als Selpref only
- M&L schneidet am besten ab
 - Unterschied zu POW nicht signifikant

Experiment 1

Auswertung

	BOW space	SYN space
<i>Model</i>	ρ	ρ
Target only	0.0	0.08
Selpref only	0.06	0.16
M&L	0.20	0.24
SELPREF	0.12	0.13
SELPREF-CUT, $\theta=10$	0.11	0.13
SELPREF-POW, $n=30$	0.27	0.22
Upper bound	0.4	0.4

- SYN space allgemein näher dran an der menschlichen Bewertung
- Unterschied von SELPREF-POW zwischen beiden spaces und zu M&L statistisch nicht signifikant

Experiment 1

Auswertung

- SVS berechnet Bedeutung anders als M&L

M&L:
 $v(\text{läuft}) * v(\text{Pferd})$

SVS:
 $v(\text{läuft}) * v(\text{Dinge, die ein Pferd tut})$

Model	lex. vector
SELPREF	0.23
SELPREF-CUT, $\theta=10$	0.20
SELPREF-POW, $n=30$	0.03

- Vektoren der Nomen zeigen wenig Ähnlichkeit zu Vektoren der inversen Selektionspräferenzen der Nomen
 - durch CUT und POW wird dies noch weiter reduziert
 - die Vorhersagen von SVS unterscheiden sich von denen von M&L

Experiment 1

Auswertung

- Berechnet das Modell unterschiedliche Vorhersagen für unterschiedliche syntaktische Strukturen?

Model	obj ⁻¹ selpref
SELPREF	0.88
SELPREF-CUT, $\theta=10$	0.72
SELPREF-POW, $n=30$	0.52

- Ähnlichkeit ist sehr hoch für SELPREF
- Reduzierung des Rauschens verringert auch die Ähnlichkeit
 - Potenzierung verringert Rauschen

Experiment 1

Bis hier erstmal

- Ein, für uns, neues Modell um Bedeutung im Kontext zu berechnen
- SVS berechnet andere Werte als die direkte Multiplikation
- Syntaktische Relationen werden mit einbezogen
- Verminderung von Rauschen verbessert die Ergebnisse
- SVS schneidet gut ab

Experiment 2

Der Datensatz

- SemEval-1 lexical substitution dataset
- 200 Zielwörter, jeweils 10 Instanzen
 - aus Sharoff's English Internet Corpus
- von bis zu 6 Teilnehmern wurden für jede Instanz von jedem Wort kontextuel passende Synonyme gesucht

Experiment 2

lexsub - Beispiel

Satz	Ersetzungsmöglichkeiten
Teacher education students will acquire the knowledge and skills required to [. . .]	gain 4 amass 1 receive 1 obtain 1
Ontario Inc. will [. . .] acquire the remaining IXOS shares [. . .]	buy 3 procure 2 purchase 1 gain 1 get 1 obtain 1

Experiment 2

lexsub - variiert

- nur ein Kontextwort
 - steht in direkter Prädikat-Argument Relation zum Zielwort
- 3 Arten von Tupeln
 - intransitives Verb (Zielwort) mit Subjekt (V-SUBJ)
 - transitives Verb (Zielwort) mit Objekt (V-OBJ)
 - Nomen (Zielwort), die als Objekte von Verben auftreten (N-OBJ)
- Daten wurden nicht nach höchster Varianz ausgewählt

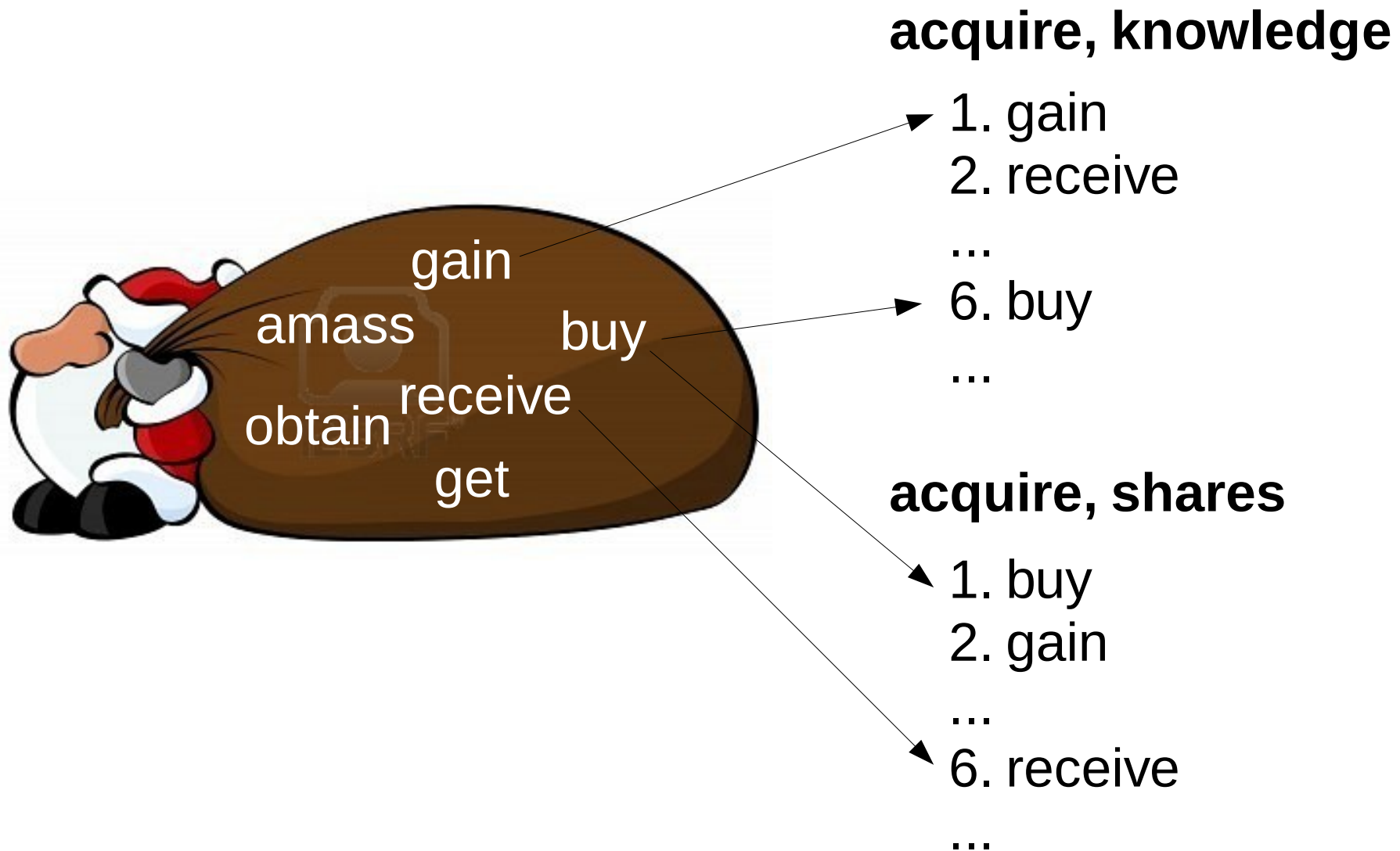
Experiment 2

Aufgabe

- Die Synonyme, aller Instanzen für das Zielwort, werden gemischt
- Die Synonyme sollen nun geordnet werden
 - M&L ordnet sie nach ihrer Ähnlichkeit zur direkten Nomen-Verb Kombination
 - SVS ordnet die Synonyme nach ihrer Ähnlichkeit anhand eben vorgestellter Kombinationen
 - V-SUBJ: Vektor des Verbs + die inversen Selektionspräferenzen für das Subjekt
 - V-OBJ: Vektor des Verbs + die inversen Selektionspräferenzen für das Objekt

Experiment 2

Aufgabenbeispiel



Experiment 2

Auswertung

<i>Model</i>	<i>V-SUBJ</i>	<i>V-OBJ</i>	<i>N-OBJ</i>
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

- Target only hat das schlechteste Ergebnis
- M&L deutlich besser als Target only
- Selpref only aber besser als M&L
- SELPREF-POW signifikant besser als M&L und die Baselines

Fazit

- Ein, für uns, neues Modell um Bedeutung im Kontext zu berechnen
- SVS schneidet gut ab
 - auf dem M&L Datensatz nicht signifikant besser
 - auf dem Lexical Substitution Datensatz signifikant besser
- Reduzierung von Rauschen verbessert das Ergebnis
- SVS berechnet andere Werte als M&L
- Syntaktische Relationen beeinflussen das Ergebnis, wenn vorher Rauschen beseitigt wurde

Ausblick

- Mehr als nur ein Wort als Kontext
- Subjekt und Objekt eines Verbes gleichzeitig miteinbeziehen
- Wie gut sind die gewählten Parameter für andere Datensätze?
- Wie beeinflusst die Art die Vektoren zu kombinieren das Ergebnis?

