Incrementality in Compositional Distributional Semantics

M. Sadrzadeh, EECS, QMUL

SemDial 2018
joint work with M. Purver, J. Hough, R. Kempson

SYCO2, Glasgow
December 2018
NLP in one slide

Formal Grammar \quad \text{structure preserving map} \quad \text{Semantic Calculus}
NLP in one slide

Formal Grammar \quad \text{structure preserving map} \quad \text{Models of First Order Logic}
NLP in one slide

Formal Grammar → structure preserving map → Distributions of Linguistic Data
Distributional Semantics

sugar, a sliced lemon, a tablespoonful of their enjoyment. Cautiously she sampled her first well suited to programming on the digital for the purpose of gathering data and apricot pineapple computer. information preserve or jam, a pinch each of, and another fruit whose taste she likened In finding the optimal R-stage policy from necessary for the study authorized in the

<table>
<thead>
<tr>
<th></th>
<th>computer</th>
<th>data</th>
<th>pinch</th>
<th>result</th>
<th>sugar</th>
</tr>
</thead>
<tbody>
<tr>
<td>apricot</td>
<td>0</td>
<td>0</td>
<td>2.25</td>
<td>0</td>
<td>2.25</td>
</tr>
<tr>
<td>pineapple</td>
<td>0</td>
<td>0</td>
<td>2.25</td>
<td>0</td>
<td>2.25</td>
</tr>
<tr>
<td>digital</td>
<td>1.66</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>information</td>
<td>0</td>
<td>0.57</td>
<td>0</td>
<td>0.47</td>
<td>0</td>
</tr>
</tbody>
</table>

$$PPMI(w, c) = \max (\log_2 \frac{P(w, c)}{P(w)P(c)}, 0)$$

Speech and Language Processing, Jurafsky and Martin
import spacy
nlp = spacy.load('en_core_web_md')
tokens = nlp('dog cat car')
for token1 in tokens:
    for token2 in tokens:
        print(token1.text, token2.text, token1.similarity(token2))

dog  dog  1.0
dog  cat  0.80168545
dog  car  0.35629162
cat  dog  0.80168545
cat  cat  1.0
cat  car  0.31907532
car  dog  0.35629162
car  cat  0.31907532
car  car  1.0
## Distributional Semantics

<table>
<thead>
<tr>
<th></th>
<th>dog</th>
<th>cat</th>
<th>car</th>
</tr>
</thead>
<tbody>
<tr>
<td>dog</td>
<td>1</td>
<td>0.80</td>
<td>0.35</td>
</tr>
<tr>
<td>cat</td>
<td></td>
<td>1</td>
<td>0.31</td>
</tr>
<tr>
<td>car</td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
A Recap on Category Theory

A monoidal functor is strongly monoidal, whenever the above morphism and natural transformation are invertible.

\[ F(A) \otimes F(B) \Rightarrow F(A \otimes B) \]
NLP in one slide

Formal Grammar → structure preserving map → Distributions of Linguistic Data

???
NLP in one slide

Formal Grammar

structure preserving map

Distributions of Linguistic Data

Type Grammars

structure preserving map

Multilinear Algebra
CCG

Types

\[ A \]
\[ X/Y \]
\[ X \parallel Y \]

Rules

\[ X/Y \ Y \Rightarrow X \]
\[ Y \ X \parallel Y \Rightarrow X \]
Multilinear Algebraic Semantics

A \mapsto \mathcal{A}

\mathcal{A} = \{e_i\}_i \quad \exists \; T_i = \sum_i C_i e_i
Multilinear Algebraic Semantics

\[ A/B \mapsto \mathcal{A} \otimes B \]

\[ \mathcal{A} = \{ e_i \}_i \quad \mathcal{B} = \{ e_j \}_j \]

\[ \mathcal{A} \otimes \mathcal{B} \ni T_{ij} = \sum_{ij} C_{ij} e_i \otimes e_j \]
Multilinear Algebraic Semantics

\[ A/(B/C) \mapsto \mathcal{A} \otimes (B \otimes C) \quad \mathcal{A} = \{e_i\}_i \quad B = \{e_j\}_j \quad C = \{e_k\}_k \]

\[ \mathcal{A} \otimes B \otimes C \ni T_{ijk} = \sum_{ijk} C_{ijk} e_i \otimes e_j \otimes e_k \]
Higher order tensors

Multilinear Algebraic Semantics

\[ A \otimes B \otimes \cdots \otimes Z \ni T_{ij \ldots w} = \sum_{ij \ldots w} C_{ij \ldots w} e_i \otimes e_j \otimes \cdots \otimes e_w \]
Multilinear Algebraic Semantics

Matrix Multiplication

\[ A/B \quad B \implies A \implies (A \otimes B) \quad B \implies A \]

\[ T_{ij} \quad T_j \quad \text{tensor contract} \quad T_i \]

\[ (\sum_{ij} C_{ije_i} \otimes e_j)(\sum_i C_{je_j}) = \sum_i C_{ij}C_{je_i}\langle e_j | e_j \rangle \]
Multilinear Algebraic Semantics

Higher order tensor contraction

\[
\cdots \mapsto A \otimes B \otimes \cdots \otimes M \quad M \otimes N \otimes P \otimes \cdots \otimes W
\]

\[
T_{ij\ldots m} \quad T_{mnp\ldots w} \quad \text{tensor contract} \quad T_{ij\ldots np\ldots w}
\]

\[
\left( \sum_{ij\ldots m} C_{ij\ldots m} e_i \otimes e_j \otimes \cdots \otimes e_m \right) \left( \sum_{mn\ldots w} C_{mn\ldots w} e_m \otimes e_n \otimes \cdots \otimes e_w \right)
\]

\[
= \sum_{ij\ldots n\ldots w} C_{ij\ldots m} C_{mn\ldots w} e_i \otimes e_j \otimes \cdots \otimes e_n \otimes \cdots \otimes e_w \langle e_m | e_m \rangle
\]
Dogs  Chase  White  Cats

\[ N \quad (S \otimes N) \otimes N \quad N \otimes N \quad N \]

\[ S \otimes N \]

\[ S \]
Dogs  Chase  White  Cats

$T_i$  $T_{ijk}$  $T_{kl}$  $T_l$

$T_k$

$T_{ij}$

$T_j$
Types ...

\[ XY^l \]
\[ Y^r X \]

Rules

\[ XY^l Y \leq X \]
\[ YY^r X \leq X \]

\[ NPNP^l \]
\[ NP^r S \]

\[ NP^r S NP^l \]

\[ NPNP^l NP \leq NP \]

\[ NPNP^r S \leq S \]
Catgorical Semantics

Formal Grammar \rightarrow \text{structure preserving map} \rightarrow \text{Distributions of Linguistic Data}

Pregroup Grammars \rightarrow \text{monoidal functor} \rightarrow \text{FVect}
Categorial Grammars + Distributional Semantics

Coecke, Sadrzadeh, Clark, 2010
Grefenstette and Sadrzadeh 2011, 2015
Maillard, Clark, Grefenstette, 2014
Krishnamurti and Mitchell, 2014
Baroni and Zamparelli 2010
Wijnholds (and Moortgat) 2015-16
Language Processing

Complete Sentences
Naturally Occurring Dialogue

- “A: mary likes …” “B: john?”
- “A: mary likes …” “B: who?”
- “A: mary likes …” “B: nobody really”
Naturally Occurring Dialogue

A: Ray destroyed . . .
B: . . . the fuchsia. He never liked it. The roses he spared . . .
A: . . . this time.
A: You are going to write the letter?
B: Only if you post it!

Howes et al, 2011, Poesio and Reiser 2010
Computational Dialogue Systems

A: I want to book a ticket ...
B: ... from where?
A: London
B: ... to where?
A: to Paris.

Purver and Kempson 2011
Purver, Eshghi, Hough 2017
Psycholinguistic Analysis

A: The footballer dribbled . . .
B (thinking) it means controlling the ball
A: . . . the ball across the pitch

A: The baby dribbled . . . the milk all over the floor.

Pickering and Frisson 2001
Cognitive Neuroscience

Predictive Processing: agents incrementally generate expectations and judge the degree to which they are met.

Frisson and Frith 2001
Clarke 2015
• Incremental Language Processing

Dynamic Syntax
+
Type Theoretic Semantics


Hough 2015, Purver et al 2014.
Recent Contribution

Dynamic Syntax + Distributional Semantics

Sadrzadeh, Purver, Hough, Kempson

SemDial 2018
Outline

- Dynamic Syntax: DS
- CDS for DS
- Some Examples
- Some Experimental Results
Dynamic Syntax

Trees decorated with semantic formulae and applications

\[ O(X_3, O(X_1, X_2)) \]

\[ X_3 \]

\[ O(X_1, X_2) \]

\[ X_1 \]

\[ X_2 \]
Dynamic Syntax

and with …

- Ty: types of formulae
- ?: requirements for further development
- <->: node currently under development
- links: connect trees of arguments of conjunctives etc
“mary ...”

\[ \text{Ty}(e), \text{Fo}(\text{mary}) \quad \text{Ty}(\langle e, t \rangle), \Diamond \]
"mary likes ..."

\[
?Ty(t) \\
\text{Ty}(e), \text{Fo}(\text{mary}) \quad ?Ty(\langle e, t \rangle) \\
\text{Ty}(\langle e, \langle e, t \rangle \rangle), \text{Fo}(\lambda y \lambda x. \text{like}(x, y))
\]
“mary likes john”

\[
?Ty(t) \\
Ty(e), Fo(mary) \quad ?Ty(\langle e, t \rangle), \diamondsuit
\]

\[
Ty(e), Fo(john) \quad Ty(\langle e, \langle e, t \rangle \rangle), Fo(\lambda y \lambda x. \text{like}(x, y))
\]
Dynamic Syntax

"mary likes john"

\[
Ty(t), Fo(\textit{like}(\textit{mary}, \textit{john})), \\
T_y(e), Fo(\textit{mary}) \quad \text{T}_y(\langle e, t \rangle), Fo(\lambda x.\text{like}(x, \textit{john}))
\]

\[
T_y(e), Fo(\textit{john}) \quad \text{T}_y(\langle e, \langle e, t \rangle \rangle), Fo(\lambda y \lambda x.\text{like}(x, y))
\]
Mary who sleeps snores.
Multilinear Algebraic Semantics for DS

\[ O(X_3, O(X_1, X_2)) \]

\[ X_3 \quad O(X_1, X_2) \]

\[ X_1 \quad X_2 \]

\[ X_1 \mapsto T_{i_1 i_2 \cdots i_n} \]
\[ X_2 \mapsto T_{i_n i_{n+1} \cdots i_{n+k}} \]
\[ X_3 \mapsto T_{i_{n+k} i_{n+k+1} \cdots i_{n+k+m}} \]

\[ \in V_1 \otimes V_2 \otimes \cdots V_n \]
\[ \in V_n \otimes V_{n+1} \otimes \cdots V_{n+k} \]
\[ \in V_{n+k} \otimes V_{n+k+1} \otimes \cdots V_{n+k+m} \]
Multilinear Algebraic Semantics for DS

\[ O(X_3, O(X_1, X_2)) \]

\[ X_3 \]

\[ O(X_1, X_2) \]

\[ X_1 \]

\[ X_2 \]

\[ O(X_1, X_2) \rightarrow T_{i_1 i_2 \cdots i_n} T_{i_n i_{n+1} \cdots i_{n+k}} \]

\[ \in V_1 \otimes V_2 \otimes \cdots \otimes V_{n-1} \otimes V_{n+1} \otimes \cdots \otimes V_{n+k} \]
Multilinear Algebraic Semantics for DS

- $\text{Ty}(t) \rightarrow S$
- $\text{Ty}(e) \rightarrow W$
- $\text{?X} \rightarrow \text{sum or direct sum of the words and phrase with semantics in } X \text{ and their probabilities}$
- $\rightarrow \text{a neutral element such as the identity in } X$
- $\rightarrow \text{a tensor full of 1’s in } X$
“mary ...”

\[ W \ni T_{i}^{mary} \]

\[ ?W \otimes S, \diamond \]
“mary likes ...”

\[
\begin{align*}
?S \\
W \ni T_i^{mary} & ?W \otimes S \\
?W, \lozenge \quad W \otimes S \otimes W & \ni T_{ijk}^{like}
\end{align*}
\]
"mary likes john"
Incremental Utterances

“Babies …”

\[ T^{\text{babies}}_i T^{+}_{ij} \]
Incremental Utterances

"Babies ..."

\[ T_{ij}^+ = \]
Incremental Utterances

“Babies …”

\[ T^+_{ij} = T_{\text{vomit}} + T_{\text{score}} + T_{\text{dribble}} + T_{\text{control baby}} + T_{\text{control milk}} + T_{\text{control footballer}} + T_{\text{control ball}} \]
Incremental Utterances

"Babies …"

\[ T_{i}^{\text{babies}} T_{ij}^{+} \]
Incremental Utterances

“Babies vomit”

\[ T_{\text{babies}} T_{\text{vomit}} \]
Incremental Utterances

“Babies score”

$T_{i \text{babies}}T_{ij \text{score}}$

Babies vomit

Babies score
Incremental Utterances

“Footballers …”

$T_{footballers}T^+$
Incremental Utterances

“Footballers vomit”

$T_{footballers} T_{vomit}$
Incremental Utterances

“Footballers score”

$T_i^\text{footballers} T_{ij}^\text{score}$
Dataset

- Kartsaklis D., MS, Pulman S.: *Separating disambiguation from composition in compositional distributional semantics*.

- Chose ambiguous verbs and two landmark meanings from Pickering and Frisson 2001

- Picked subjects and objects for landmarks using most frequently occurring ones in the BNC
Dataset

- Pairs of subjects and complete sentences:
  (footballers ..., footballers dribble milk)
  (footballers ..., footballers dribble ball)

- Pairs of subject+verb and complete sentences:
  (footballers dribble ..., footballers dribble milk)
  (footballers dribble ..., footballers dribble ball)

- Pairs of complete sentences:
  (footballers dribble milk , footballers dribble ball)
  (babies dribble milk , babies dribble ball)
Data

Vectors: 300 Dim from Word2Vec,
Tensors: the G&S EMNLP 2011 method
Data

Just subject

footballers control

0.086

footballers …

0.049

footballers drip
Data

Just Subject

footballers dribble ball

0.0046

footballers ...

0.0019

footballers dribble milk
Data

Subject + Verb

footballers dribble ball

footballers dribble ...

0.22

footballers dribble milk

0.02
Complete Sentences

0.22 < 0.36
Data

Complete Sentences

0.34 > 0.32
Accuracy Results

Accuracy

- G&S
- copy-subj
- copy-obj
- add

Partiality

- Subj
- Subj+ Verb
- Subj+ Verb+Obj
Work in Progress

Implement the plausibilities model of Clark 2013, Polajnar et al 2015
... under way ...

Extend it to experimental expectation predication
...

Incremental Understanding of Dialogue Content
Categorical Semantics

CCC

CCC + biproducts

functor

V?

FVect

FVect
A: Thank …

B: … you!