A Structured Vector Space Model for Hidden Attribute Meaning in Adjective-Noun Phrases

Matthias Hartung  Anette Frank
Computational Linguistics Department
Heidelberg University

COLING 2010
Beijing, August 24
Background: Learning Concept Descriptions

- ontology learning: describe and distinguish concepts by properties and relations
  - **motorcycle**: ride, rider, sidecar, park, road, helmet, collision, vehicle, car, moped, ...
    - Baroni et al. (2010)
  - **car**: acceleration, performance, front, engine, backseat, chassis, speed, weight, color, condition, driver, buyer, ...
    - Poesio & Almuhareb (2005)

- common denominator: learn “prototypical”, “static” knowledge about concepts from text corpora
Focus of this Talk

Concept Modification in Linguistic Contexts

▶ What are the **attributes** of a concept that are highlighted in an adjective-noun phrase?
▶ well-known problem in formal semantics: *selective binding*
  ▶ *fast car* ⇔ \( \text{SPEED(car)} = \text{fast} \)
  ▶ *red balloon* ⇔ \( \text{COLOR(balloon)} = \text{red} \)
  ▶ *oval table* ⇔ \( \text{SHAPE(table)} = \text{oval} \)

(cf. Pustejovsky 1995)

▶ attribute selection as a **compositional process**
Previous Work: Attribute Learning from Adjectives

   - goal: learn binary **noun-attribute relations**
   - detour via adjectives modifying the noun
   - for each adjective: look up attributes from WordNet

   - goal: learn binary **adjective-attribute relations**
   - pattern-based approach:
     
     the ATTR of the * is|was ADJ

**Problem:** The **ternary attribute relation**

\[
\text{ATTRIBUTE(noun)} = \text{adjective}
\]

is missed by both approaches; e.g.: *hot summer* vs. *hot soup*
Learning Ternary Attribute Relations

“Naive” Solution: Pattern-based Approach

- the ATTR of the N is|was ADJ
- challenge: overcome **sparsity issues**

A Structured VSM for Ternary Attribute Relations

- represent adjective and noun meanings independently in a **structured vector space model**
- semantic vectors capture binary relations $r' = \langle \text{noun}, \text{attr} \rangle$ and $r'' = \langle \text{adj}, \text{attr} \rangle$
- use **vector composition** to approximate the ternary attribute relation $r$ from $r'$ and $r''$:

$$v(r) \approx v(r') \otimes v(r'')$$

ex.: $v(\langle \text{speed}, \text{car}, \text{fast} \rangle) \approx v(\langle \text{car}, \text{speed} \rangle) \otimes v(\langle \text{fast}, \text{speed} \rangle)$
Outline

Introduction

A Structured VSM for Attributes in Adjective-Noun Phrases
  Building the Model
  Vector Composition
  Attribute Selection

Experiments and Evaluation

Conclusions and Outlook
Building Vector Representations for Adjectives

<table>
<thead>
<tr>
<th>enormous</th>
<th>COLOR</th>
<th>DIRECT.</th>
<th>DURAT.</th>
<th>SHAPE</th>
<th>SIZE</th>
<th>SMELL</th>
<th>SPEED</th>
<th>TASTE</th>
<th>TEMP.</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>45</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>21</td>
</tr>
</tbody>
</table>
Building Vector Representations for Adjectives

<table>
<thead>
<tr>
<th></th>
<th>COLOR</th>
<th>DIRECT.</th>
<th>DURAT.</th>
<th>SHAPE</th>
<th>SIZE</th>
<th>SMELL</th>
<th>SPEED</th>
<th>TASTE</th>
<th>TEMP.</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>enormous</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>45</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>21</td>
</tr>
</tbody>
</table>

- 10 manually selected attributes: color, direction, duration, shape, size, smell, speed, taste, temperature, weight

- vector component values: raw corpus frequencies obtained from lexico-syntactic patterns

Almuhareb (2006)
Building Vector Representations for Adjectives

<table>
<thead>
<tr>
<th>enormous</th>
<th>COLOR</th>
<th>DIRECT.</th>
<th>DURAT.</th>
<th>SHAPE</th>
<th>SIZE</th>
<th>SMELL</th>
<th>SPEED</th>
<th>TASTE</th>
<th>TEMP.</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>45</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>

- 10 manually selected attributes: `color`, `direction`, `duration`, `shape`, `size`, `smell`, `speed`, `taste`, `temperature`, `weight`

Almuhareb (2006)

- vector component values: raw corpus frequencies obtained from lexico-syntactic patterns

(A1) `ATTR` of DT? NN `is|was` JJ
(A2) DT? RB? JJ `ATTR`
(A3) DT? JJ or JJ `ATTR`
(A4) DT? NN’s `ATTR` `is|was` JJ
(A5) `is|was|are|were` JJ `in|of` `ATTR`
Building Vector Representations for Nouns

- 10 manually selected attribute nouns: *color, direction, duration, shape, size, smell, speed, taste, temperature, weight*

- Vector component values: raw corpus frequencies obtained from lexico-syntactic patterns

| enormous | 1 | 1 | 0 | 1 | 45 | 0 | 4 | 0 | 0 | 21 |
| ball     | 14| 38| 2 | 20| 26 | 0 | 45| 0 | 0 | 20 |

(N1) **NN** with\ without DT? RB? JJ? ATTR
(N2) DT **ATTR** of DT? RB? JJ? **NN**
(N3) DT **NNʼs** RB? JJ? **ATTR**
(N4) **NN** has\ had a\ an RB? JJ? **ATTR**
Vector Composition

- component-wise multiplication \( \circ \)
- vector addition \( \oplus \)

Mitchell & Lapata (2008)
Vector Composition

- component-wise multiplication \( \odot \)
- vector addition \( \oplus \)

Mitchell & Lapata (2008)

<table>
<thead>
<tr>
<th></th>
<th>COLOR</th>
<th>DIRECT.</th>
<th>DURAT.</th>
<th>SHAPE</th>
<th>SIZE</th>
<th>SMELL</th>
<th>SPEED</th>
<th>TASTE</th>
<th>TEMP.</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>enormous ball</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>45</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>38</td>
<td>2</td>
<td>20</td>
<td>26</td>
<td>0</td>
<td>45</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
</tbody>
</table>
Vector Composition

- component-wise multiplication $\odot$
- vector addition $\oplus$

Mitchell & Lapata (2008)

<table>
<thead>
<tr>
<th></th>
<th>COLOR</th>
<th>DIRECT</th>
<th>DURAT.</th>
<th>SHAPE</th>
<th>SIZE</th>
<th>SMELL</th>
<th>SPEED</th>
<th>TASTE</th>
<th>TEMP.</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>enormous ball</td>
<td>14</td>
<td>38</td>
<td>2</td>
<td>20</td>
<td>45</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>enormous $\odot$ ball</td>
<td>14</td>
<td>38</td>
<td>0</td>
<td>20</td>
<td>1170</td>
<td>0</td>
<td>180</td>
<td>0</td>
<td>0</td>
<td>420</td>
</tr>
<tr>
<td>enormous $\oplus$ ball</td>
<td>15</td>
<td>39</td>
<td>2</td>
<td>21</td>
<td>71</td>
<td>0</td>
<td>49</td>
<td>0</td>
<td>0</td>
<td>41</td>
</tr>
</tbody>
</table>
Vector Composition

- component-wise multiplication $\odot$
- vector addition $\oplus$

Mitchell & Lapata (2008)

<table>
<thead>
<tr>
<th></th>
<th>COLOR</th>
<th>DIRECT.</th>
<th>DURAT.</th>
<th>SHAPE</th>
<th>SIZE</th>
<th>SMELL</th>
<th>SPEED</th>
<th>TASTE</th>
<th>TEMP.</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>enormous</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>45</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>ball</td>
<td>14</td>
<td>38</td>
<td>2</td>
<td>20</td>
<td>26</td>
<td>0</td>
<td>45</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td>enormous $\odot$ ball</td>
<td>14</td>
<td>38</td>
<td>0</td>
<td>20</td>
<td><strong>1170</strong></td>
<td>0</td>
<td>180</td>
<td>0</td>
<td>0</td>
<td><strong>420</strong></td>
</tr>
<tr>
<td>enormous $\oplus$ ball</td>
<td>15</td>
<td>39</td>
<td>2</td>
<td>21</td>
<td><strong>71</strong></td>
<td>0</td>
<td>49</td>
<td>0</td>
<td>0</td>
<td><strong>41</strong></td>
</tr>
</tbody>
</table>

- **expectation:** vector multiplication comes closest to the linguistic function of intersective adjectives!
Attribute Selection

- goal: make attributes explicit that are most salient in the compositional semantics of adjective-noun phrases
- achieved so far: ranking of attributes according to their prominence in the composed vector representation
- **attribute selection:** distinguish meaningful from noisy components in vector representations
  - MPC Selection
  - Threshold Selection
  - Entropy Selection
  - Median Selection
MPC Selection

Functionality:

- selects the **most prominent component from each vector** (in terms of absolute frequencies)

<table>
<thead>
<tr>
<th>enormous</th>
<th>COLOR</th>
<th>DIRECT.</th>
<th>DURAT.</th>
<th>SHAPE</th>
<th>SIZE</th>
<th>SMELL</th>
<th>SPEED</th>
<th>TASTE</th>
<th>TEMP.</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>45</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>21</td>
</tr>
</tbody>
</table>

Drawback:

- inappropriate for vectors with more than one meaningful dimension
Threshold Selection

Functionality:

▶ selects all components exceeding a frequency threshold $\theta$ (here: $\theta \geq 10$)

<table>
<thead>
<tr>
<th></th>
<th>COLOR</th>
<th>DIRECT.</th>
<th>DURAT.</th>
<th>SHAPE</th>
<th>SIZE</th>
<th>SMELL</th>
<th>SPEED</th>
<th>TASTE</th>
<th>TEMP.</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ball</td>
<td>14</td>
<td>38</td>
<td>2</td>
<td>20</td>
<td>26</td>
<td>0</td>
<td>45</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
</tbody>
</table>

Drawbacks:

▶ introduces an additional parameter to be optimized
▶ difficult to apply to composed vectors
▶ unclear whether method scales to vectors of higher dimensionality
Entropy Selection

Functionality:

▶ select all informative components
▶ information theory: gain in entropy $\equiv$ loss of information
▶ retain all (combinations of) components that lead to a gain in entropy when taken out

<table>
<thead>
<tr>
<th></th>
<th>COLOR</th>
<th>DIRECT.</th>
<th>DURAT.</th>
<th>SHAPE</th>
<th>SIZE</th>
<th>SMELL</th>
<th>SPEED</th>
<th>TASTE</th>
<th>TEMP.</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>enormous ball</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>45</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>38</td>
<td>2</td>
<td>20</td>
<td>26</td>
<td>0</td>
<td>45</td>
<td>0</td>
<td>0</td>
<td>20</td>
</tr>
</tbody>
</table>

Drawback:

▶ yields no attribute for vectors with broad and flat distributions (noun vectors, in particular)
Median Selection

Functionality:

- tailored to noun vectors, in particular
- select all components with values above the median

<table>
<thead>
<tr>
<th>COLOR</th>
<th>DIRECT.</th>
<th>DURAT.</th>
<th>SHAPE</th>
<th>SIZE</th>
<th>SMELL</th>
<th>SPEED</th>
<th>TASTE</th>
<th>TEMP.</th>
<th>WEIGHT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ball</td>
<td>14</td>
<td>38</td>
<td>2</td>
<td>20</td>
<td>26</td>
<td>0</td>
<td>45</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Drawback:

- depends on the number of dimensions
Taking Stock...

Introduction

A Structured VSM for Attributes in Adjective-Noun Phrases
Building the Model
Vector Composition
Attribute Selection

Experiments and Evaluation

Conclusions and Outlook
Experimental Setup

Experiments:

1. attribute selection from adjective vectors
2. attribute selection from noun vectors
3. attribute selection from composed adjective-noun vectors

Methodology:

- vector acquisition from ukWaC corpus (Baroni et al. 2009)
- gold standards for comparison:
  - Experiment 1: compiled from WordNet
  - Experiments 2/3: manually established by human annotators
- evaluation metrics: precision, recall, $f_1$-score
Experiment 1: Attribute Selection from Adjective Vectors

Data Set

- all adjectives extracted by patterns (A1)-(A5) occurring at least 5 times in ukWaC (3505 types in total)

Gold Standard

- 1063 adjectives that are linked to at least one of the ten attributes we consider in WordNet 3.0


- patterns (A1)-(A3) only
- manually optimized thresholds for attribute selection
- frequency scores acquired from the web
**Experiment 1: Results**

<table>
<thead>
<tr>
<th></th>
<th>Almuhareb (reconstr.)</th>
<th>VSM (TSel + Target Filter)</th>
<th>VSM (ESel + Target Filter)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F</td>
</tr>
<tr>
<td>A1</td>
<td>0.183</td>
<td>0.005</td>
<td>0.009</td>
</tr>
<tr>
<td>A2</td>
<td>0.207</td>
<td>0.039</td>
<td>0.067</td>
</tr>
<tr>
<td>A3</td>
<td>0.382</td>
<td>0.020</td>
<td>0.039</td>
</tr>
<tr>
<td>A4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table: Attribute Selection from Adjective Vectors**

- re-implementation yields performance comparable to Almuhareb’s original system
- performance increase of 13 points in precision over Almuhareb; recall is still poor
- best parameter settings:
  - entropy selection method
  - target filtering (intersect extractions of two patterns in order to remove noisy or unreliable vectors)
Experiment 2: Attribute Selection from Noun Vectors

Creation of an Annotated Data Set

- random sample from the balanced set of 402 (216) nouns compiled by Almuhareb (2006)
- three human annotators
- task: remove all attributes that are not appropriate for any sense of a given noun
- adjudication of disagreements by majority voting

Resulting Gold Standard

- 100 nouns with 4.24 attributes on average
- inter-annotator agreement: $\kappa = 0.69$
Experiment 2: Results

<table>
<thead>
<tr>
<th></th>
<th>MPC P R F</th>
<th>ESel P R F</th>
<th>MSel P R F</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1</td>
<td>0.22 0.06 0.10</td>
<td>0.29 0.04 0.07</td>
<td>0.22 0.09 0.13</td>
</tr>
<tr>
<td>N2</td>
<td>0.29 0.18 0.23</td>
<td>0.20 0.06 0.09</td>
<td>0.28 0.39 0.33</td>
</tr>
<tr>
<td>N3</td>
<td>0.34 0.05 0.09</td>
<td>0.20 0.02 0.04</td>
<td>0.25 0.08 0.12</td>
</tr>
<tr>
<td>N4</td>
<td>0.25 0.02 0.04</td>
<td>0.29 0.02 0.03</td>
<td>0.26 0.02 0.05</td>
</tr>
<tr>
<td>all</td>
<td>0.29 0.18 0.22</td>
<td>0.20 0.06 0.09</td>
<td>0.28 0.43 0.34</td>
</tr>
</tbody>
</table>

Table: Attribute Selection from Noun Vectors

- MPC: relatively precise, poor in terms of recall
- ESel: counterintuitively fails to increase recall
- MSel: best recall, most suitable for this task

Problems:
- vectors with broad, flat distributions
- binary attribute-noun relation often not overtly realized
Experiment 3: Attribute Selection from Composed Adjective-Noun Vectors

Creation of an Annotated Data Set

- partially random sample from 386 property-denoting adjectives × 216 nouns
- three human annotators (same as in Experiment 2)
- task: remove all attributes not appropriate for a given pair (not provided by the noun or not selected by the adjective)
- adjudication of disagreements by majority voting

Resulting Gold Standard

- 76 pairs with 1.13 attributes on average, 24 “empty” pairs
- inter-annotator agreement: $\kappa = 0.67$
Experiment 3: Baselines

- **BL-P:** purely pattern-based method searching for patterns that make ternary attribute relations explicit
  
  the ATTR of the N is|was ADJ

- **BL-A:** take individual adjective vector as surrogate for composition

- **BL-N:** take individual noun vector as surrogate for composition
## Experiment 3: Results

<table>
<thead>
<tr>
<th></th>
<th>MPC</th>
<th></th>
<th>ESel</th>
<th></th>
<th>MSeI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F</td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Adj ⊙ N</td>
<td>0.60</td>
<td>0.58</td>
<td>0.59</td>
<td>0.63</td>
<td>0.46</td>
</tr>
<tr>
<td>Adj ⊕ N</td>
<td>0.43</td>
<td>0.55</td>
<td>0.48</td>
<td>0.42</td>
<td>0.51</td>
</tr>
<tr>
<td>BL-Adj</td>
<td>0.44</td>
<td>0.60</td>
<td>0.50</td>
<td>0.51</td>
<td>0.63</td>
</tr>
<tr>
<td>BL-N</td>
<td>0.27</td>
<td>0.35</td>
<td>0.31</td>
<td>0.37</td>
<td>0.29</td>
</tr>
<tr>
<td>BL-P</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table: Attribute Selection from Composed Adjective-Noun Vectors

- complete failure of BL-P
- modelling ternary relations by composing vector representations of reduced complexity is feasible, but: choice of composition method matters
- ESel most suitable wrt. precision (partly due to its ability to return “empty” selections)
- robustness of MPC mainly due to the large proportion of pairs in the test set that elicit one attribute only
Conclusions and Outlook

- **structured VSM** as a framework for inferring hidden attributes in the compositional semantics of adjective-noun phrases
- **vector composition** as a hinge to model ternary attribute relations from individual vectors capturing adjective and noun meanings, thus avoiding sparsity issues
- Attribute selection from adjectives: increase of 13 points in precision above pattern-based approach of Almuhareb (2006)

Future work:
- scale approach to higher dimensionality
- address problems with infrequent and unreliable vectors (particularly nouns)
References


Thanks... 

...for your attention.
Questions?