Classifying Adjectives for Attribute Learning: an Empirical Investigation

Matthias Hartung  Anette Frank

Computational Linguistics Department
University of Heidelberg

CTF 2009, Düsseldorf
Classifying Adjectives for Attribute Learning: Outline

1. Background & Motivation

2. Annotation Experiment
   - Initial Classification Scheme
   - Task Description
   - First Results
   - Results after Re-Analysis

3. Outlook: Alternative Approach
   - Foundations of Vector Space Models (VSMs)
   - Towards Attribute Learning in VSMs

4. Conclusions
Goals

- semantic interpretation of adjective-noun phrases in terms of paraphrases
- focus of today’s talk: **Is it possible to classify adjectives into attribute-denoting ones and ”others” ?**

Examples

- oval table ⇒ table has an oval **SHAPE**
- fast car ⇒ car that drives fast
- dangerous disease ⇒ disease that infects/kills many people
Motivation

Adjectives as Gateways to Conceptual Representation

Figure: Frame Representation of Geometric Forms (Barsalou, 1992)
Prior Work: Using Attributes for Clustering Nouns into Concepts

Search for Attribute-Denoting Nouns
- pattern-based strategy: the ATTR of the CONCEPT
- main problem: overgeneration of potential attributes

Detour via Adjectives
- Which adjectives act as modifiers of the respective noun and which attributes are they related to?
- best results by combination of attribute nouns and adjectives
- Hypothesis: filtering adjectives that do not denote attributes might increase performance, i.e. yield cleaner concepts

[Almuhareb, 2006]
Taking Stock...

1. Background & Motivation

2. Annotation Experiment
   - Initial Classification Scheme
   - Task Description
   - First Results
   - Results after Re-Analysis

3. Outlook: Alternative Approach
   - Foundations of Vector Space Models (VSMs)
   - Towards Attribute Learning in VSMs

4. Conclusions
Annotation Experiment

<table>
<thead>
<tr>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is it feasible, in principle, to separate adjective-denoting adjectives from &quot;others&quot;?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Initial Classification Scheme: BEO Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basic Adjectives, e.g.: <em>red carpet</em></td>
</tr>
<tr>
<td>Event-related Adjectives, e.g.: <em>fast horse</em></td>
</tr>
<tr>
<td>Object-related Adjectives, e.g.: <em>political debate</em></td>
</tr>
</tbody>
</table>

[Raskin & Nirenburg, 1998; Boleda, 2007]
BEO Classes (1)

Event-related Adjectives
- there is an event the referent of the noun takes part in
- adjective functions as a modifier of this event

Examples
- **good** knife ⇒ knife that **cuts** **well**
- **fast** horse ⇒ horse that **runs** **fast**
Event-related Adjectives: Some more examples...

- fast horse
- eloquent person
- interesting book
- oral contraceptive

Tests from the literature

- this is a ADJ ENT ⇒ this ENT is ADJ for/at/... EVENT
- this is a ADJ ENT ⇒ this ENT EVENT ADV/ADJ
- this is a ADJ ENT ⇒ this ENT is ADJ to EVENT
### Object-related Adjectives

- adjective is morphologically related to a noun reading \( N/ADJ \)
- \( N/ADJ \) refers to an entity that acts as a semantic dependent of the head noun \( N \)

### Examples

- **environmental destruction\(_N\)**
  \[ \Rightarrow \text{destruction}\(_N\) [of] the environment\(_{N/ADJ}\) \]
  \[ \Rightarrow \text{destruction}(e, \text{AGENT}: x, \text{PATIENT}: \text{environment}) \]

- **political debate\(_N\)**
  \[ \Rightarrow \text{debate}\(_N\) [on] politics\(_{N/ADJ}\) \]
  \[ \Rightarrow \text{debate}(e, \text{AGENT}: x, \text{TOPIC}: \text{politics}) \]
BEO Classes (2) – continued

Object-related Adjectives: Some more examples...
- economic crisis
- political debate
- rural visitors
- stony bridge

Tests from the literature
- an ADJ ENT ⇒ ENT on/of/from/... N/ADJ
- an ADJ ENT ⇒ ENT is made of N/ADJ
**BEO Classes (3)**

**Basic Adjectives**
- adjective denotes a value of an attribute exhibited by the noun
- adjective denotes either a discrete value of the attribute or a predication over a range of potential values (depending on the concept being modified)

**Examples**
- *red carpet* ⇒ COLOR(carpet)=red
- *young bird* ⇒ AGE(bird)=[?,?]

red carpet ⇒ COLOR(carpet)=red
young bird ⇒ AGE(bird)=[?,?]
### Basic Adjectives: Some more examples...

- white snake ⇒ COLOR(snake)=white
- high bridge ⇒ HEIGHT(bridge)=high
- long train ⇒ LENGTH(train)=long
- oval table ⇒ SHAPE(table)=oval

### Tests from the literature

- an ADJ ENT ⇒ the ENT has a ADJ ATTRIB
- the ENT is ADJ ⇒ the ENT has a ADJ ATTRIB
- an ATTRIB ENT ⇒ the ATTRIB of the ENT is ADJ
Annotation Experiment: Task Description and Methodology

Data Set
- list of 200 high-frequency adjectives from the British National Corpus
- random extraction of five example sentences from the written part of the BNC for each of the 200 adjectives

Methodology
- three annotators
- task: label each of the 1000 items with BASIC, EVENT, OBJECT or IMPOSSIBLE
- instructions: short description of the classes plus examples
BEO Classification: Fundamental Ambiguities

### EVENT vs. BASIC

- fast horse $\Rightarrow \ ?\text{VELOCITY}(\text{horse})=\text{fast}$
- good knife $\Rightarrow \ ?\text{QUALITY}(\text{knife})=\text{good}$
- eloquent person $\Rightarrow \ ?\text{ELOQUENCE}(\text{person})=\text{true}$
- difficult problem $\Rightarrow \ ?\text{DIFFICULTY}(\text{problem})=\text{true}$

### Additional Instructions: Differentiation Criteria

- ENT’s property of being ADJ is due to ENT’s ability to EVENT.
- If ENT was unable to EVENT, it would not be an ADJ ENT.
Taking Stock...

1. Background & Motivation

2. Annotation Experiment
   - Initial Classification Scheme
   - Task Description
   - First Results
   - Results after Re-Analysis

3. Outlook: Alternative Approach
   - Foundations of Vector Space Models (VSMs)
   - Towards Attribute Learning in VSMs

4. Conclusions
Tri-partite Classification: Annotator Agreement

<table>
<thead>
<tr>
<th></th>
<th>Annotator 1</th>
<th>Annotator 2</th>
<th>Annotator 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotator 1</td>
<td>—</td>
<td>0.762</td>
<td>0.235</td>
</tr>
<tr>
<td>Annotator 2</td>
<td>0.762</td>
<td>—</td>
<td>0.285</td>
</tr>
<tr>
<td>Annotator 3</td>
<td>0.235</td>
<td>0.285</td>
<td>—</td>
</tr>
</tbody>
</table>

Table: Agreement figures in terms of Fleiss’ $\kappa$

- overall agreement: $\kappa = 0.4$
- rather poor agreement; but: mainly due to one ”outlier” among the annotators
- Which ones were the most problematic cases?
Tri-partite Classification: Annotator Agreement (category-wise)

<table>
<thead>
<tr>
<th></th>
<th>BASIC</th>
<th>EVENT</th>
<th>OBJECT</th>
<th>IMPOSSIBLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \kappa )</td>
<td>0.368</td>
<td>0.061</td>
<td>0.700</td>
<td>0.452</td>
</tr>
</tbody>
</table>

**Table:** Category-wise \( \kappa \)-values for all annotators

- Separating the OBJECT class is quite feasible
- Can poor overall agreement be traced back to the ambiguities between BASIC and EVENT class?
Tri-partite Classification: Cases of Disagreement

<table>
<thead>
<tr>
<th></th>
<th>BASIC</th>
<th>EVENT</th>
<th>OBJECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2:1 agreement</td>
<td>283</td>
<td>21</td>
<td>66</td>
</tr>
<tr>
<td>3:0 agreement</td>
<td>486</td>
<td>5</td>
<td>62</td>
</tr>
</tbody>
</table>

Table: Cases of Agreement vs. Disagreement

<table>
<thead>
<tr>
<th></th>
<th>1 voter</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BASIC</td>
<td>EVENT</td>
<td>OBJECT</td>
<td></td>
</tr>
<tr>
<td>1 voter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 voters</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BASIC</td>
<td>–</td>
<td>172</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>EVENT</td>
<td>18</td>
<td>–</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>OBJECT</td>
<td>54</td>
<td>10</td>
<td>–</td>
<td></td>
</tr>
</tbody>
</table>

Table: Distribution of Disagreement Cases over Classes

- Figures corroborate that the BASIC/EVENT ambiguity is the primary source of disagreement!
- What makes this distinction so hard to draw?
Be that as it may, it is safe to say that no matter which rules a karateka fights under, he will get a **fair** deal.

→ annotators’ votes: 2 BASIC, 1 EVENT

Any changes should only be introduced after **proper** research and costing, and after an initial experiment.

→ annotators’ votes: 2 BASIC, 1 EVENT
Ambiguous Corpus Examples:

- **Strong** instructions went out to fields reviewing their progress and preparing proposals that there should be as little change as possible from that which had been originally approved.
  → annotators’ votes: 2 EVENT, 1 BASIC

- Matthew thought his mother sounded very young, her voice **bright** with some emotion he could not quite define but which made him feel instantly — paternally — protective.
  → annotators’ votes: 2 BASIC, 1 EVENT
People have substantial difficulties in distinguishing BASIC from EVENT adjectives!

Do these classes share some commonalities that make them more alike than different?

**Re-analysis**: abstract away from subtle differences by separating only two classes:
- adjectives denoting **properties** (BASIC & EVENT)
- adjectives denoting **relations** (OBJECT)

Expectation: re-analysis of the annotated data with regard to a **bi-partite** classification scheme should yield an improvement in annotator agreement!
Taking Stock...

1. **Background & Motivation**

2. **Annotation Experiment**
   - Initial Classification Scheme
   - Task Description
   - First Results
   - Results after Re-Analysis

3. **Outlook: Alternative Approach**
   - Foundations of Vector Space Models (VSMs)
   - Towards Attribute Learning in VSMs

4. **Conclusions**
Bi-partite Classification: Annotator Agreement (category-wise)

<table>
<thead>
<tr>
<th></th>
<th>BASIC+EVENT</th>
<th>OBJECT</th>
<th>IMPOSSIBLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>0.696</td>
<td>0.701</td>
<td>-0.003</td>
</tr>
</tbody>
</table>

**Table:** Category-wise $\kappa$-values for all annotators

- overall agreement: $\kappa = 0.69$ (substantial agreement)
- two-way classification into properties and relations seems to be reasonable
- difference between BASIC and EVENT-related properties is very fine-grained and difficult for humans to assess!
- Are there different types of properties?
The notion of **foundation** (Guarino, 1992)

A concept $\alpha$ is called *founded* if there exists a concept $\beta$ such that any instance $\chi$ of $\alpha$ is necessarily associated to an instance $\psi$ of $\beta$ which is not related to $\chi$ by a part-of relation.

Applying the notion of foundation to properties yields (in Guarino’s terminology):

- **attributes**: properties that are *inherent* to an entity
- **roles**: properties that are *dependent* on a property of some other entity or event
Attributes vs. Roles (1)

**Example**

![Diagram showing the relationship between attributes and roles]

*Figure: The *speed* role of cars*
Attributes vs. Roles (2)

Hypothesis

Attributes and roles denote **different types of properties**, e.g.:
- **Attributes**: size, shape, weight, duration, color, ...
- **Roles**: speed, temperature, taste, difficulty, color, ...

Assessment: So what?

- "ontological difference" might explain the difficulties in the BASIC/EVENT distinction to a certain extent
- **but**: does not provide any additional distinctive features that are "overtly" observable
Features for Classification: Overview

<table>
<thead>
<tr>
<th>Pattern</th>
<th>EVENT</th>
<th>OBJECT</th>
<th>BASIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENT is ADJ for/at/... EVENT</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ENT's property of being ADJ is due to ENT's ability to EVENT</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>If ENT was unable to EVENT, it would not be an ADJ ENT</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A N &lt;&amp;gt; R(N, N/ADJ)</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>predicative use</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>ENT is ADJ for an ENT</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>ENT has an ADJ ATTRIB</td>
<td>+/−</td>
<td>+/−</td>
<td>+</td>
</tr>
<tr>
<td>gradability</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>comparability</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>N/ADJ is an attribute of ENT</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
</tbody>
</table>

- distinction between BASIC/EVENT vs. OBJECT should be feasible with a **pattern-based approach**
- tests for BASIC/EVENT distinction rely on infrequent patterns or semantic distinctions that are difficult to decide
- argument in favour of a **semantic** model rather than a pattern-based approach for the distinction *between* BASIC and EVENT
Taking Stock...

1. Background & Motivation

2. Annotation Experiment
   - Initial Classification Scheme
   - Task Description
   - First Results
   - Results after Re-Analysis

3. Outlook: Alternative Approach
   - Foundations of Vector Space Models (VSMs)
   - Towards Attribute Learning in VSMs

4. Conclusions
A VSM for Adjective-Noun Phrases

Foundations of Vector Space Semantics

- representation of word meaning as vectors in a high-dimensional space
- dimensions of the space: contexts in which the word occurs (cf. ”distributional hypothesis”; Firth, 1957)
- ”geometric metaphor”: words that are represented by points in space that are close to each other are similar in meaning (Sahlgren, 2006)
- can be automatically induced form corpora
A VSM for Adjective-Noun Phrases: Our Proposal

<table>
<thead>
<tr>
<th></th>
<th>speed</th>
<th>color</th>
<th>price</th>
<th>beauty</th>
<th>height</th>
<th>danger</th>
</tr>
</thead>
<tbody>
<tr>
<td>fast</td>
<td>81</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>expensive</td>
<td>2</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>dangerous</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>drive</td>
<td>66</td>
<td>2</td>
<td>47</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>buy</td>
<td>3</td>
<td>13</td>
<td>73</td>
<td>3</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>paint</td>
<td>0</td>
<td>54</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>car</td>
<td>34</td>
<td>20</td>
<td>63</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>building</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>36</td>
<td>8</td>
</tr>
</tbody>
</table>

Properties of our "Toy Space"

- **dimensions**: selection of nouns denoting attributes and roles
- **targets**: adjectives, nouns and verbs are modelled in one and the same space
- **cooccurrence values**: raw frequencies or association measures (e.g. PMI variants, log likelihood, ...)
A VSM for Adjective-Noun Phrases: Hypothesis I

**Compositional Semantics**

The *compositional semantics* of an adjective-noun compound can be modelled by some linear combination of its constitutive vectors (cf. Mitchell & Lapata, 2008):

\[
[\text{fast car}] = \vec{\text{fast}} \oplus \vec{\text{car}}
\]

**Example:**

<table>
<thead>
<tr>
<th></th>
<th>speed</th>
<th>color</th>
<th>price</th>
<th>beauty</th>
<th>height</th>
<th>danger</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>fast</code></td>
<td>81</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><code>car</code></td>
<td>34</td>
<td>20</td>
<td>63</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td><code>fast \oplus car</code></td>
<td>115</td>
<td>21</td>
<td>67</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
Attribute or Role Detection

The appropriate attributes or roles that are denoted by an adjective-noun phrase $A \ N$ can be discovered from the most prominent dimension in the combined vector $\vec{A} \oplus \vec{N}$.

Example:

<table>
<thead>
<tr>
<th></th>
<th>speed</th>
<th>color</th>
<th>price</th>
<th>beauty</th>
<th>height</th>
<th>danger</th>
</tr>
</thead>
<tbody>
<tr>
<td>fast</td>
<td>81</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>car</td>
<td>34</td>
<td>20</td>
<td>63</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>fast $\oplus$ car</td>
<td>115</td>
<td>21</td>
<td>67</td>
<td>1</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>
A VSM for Adjective-Noun Phrases: Hypothesis III

Semantic Similarity

Similar distributions of targets over all dimensions indicate semantic similarity:

- adjectives of the same scale (e.g. fast, slow, ...)
- verbs of the same class (e.g. drive, run, ...)
- across POS categories: verbs that are closely associated with a particular dimension

Example:

<table>
<thead>
<tr>
<th></th>
<th>speed</th>
<th>color</th>
<th>price</th>
<th>beauty</th>
<th>height</th>
<th>danger</th>
</tr>
</thead>
<tbody>
<tr>
<td>fast</td>
<td>81</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>slow</td>
<td>54</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>expensive</td>
<td>2</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Attribute vs. Role Distinction

Let \( \vec{A} \oplus \vec{N} \) be a vector composition, for which there exists a vector composition \( \vec{V} \oplus \vec{N} \) that exhibits a similar distribution over all dimensions in an attribute space \( VS_{attr} \). If \( V \) is not an important dimension of \( N \) in an object space \( VS_{obj} \), then \( A \) is considered to denote an attribute of \( N \).

Example:

1. Which verbs are strongly associated with the most relevant dimension?

<table>
<thead>
<tr>
<th></th>
<th>speed</th>
<th>color</th>
<th>price</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>grey ⊕ cat</td>
<td>2</td>
<td>18</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>grey ⊕ building</td>
<td>4</td>
<td>27</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>paint ⊕ cat</td>
<td>2</td>
<td>59</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>paint ⊕ building</td>
<td>4</td>
<td>68</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

2. Do these verbs indicate a valid role?

<table>
<thead>
<tr>
<th></th>
<th>paint</th>
<th>boil</th>
<th>increase</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>5</td>
<td>8</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>building</td>
<td>14</td>
<td>0</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>car</td>
<td>8</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
A VSM for Adjective-Noun Phrases: First Results

Hypothesis II: Adjectives from the same scale

<table>
<thead>
<tr>
<th>Association Measure</th>
<th>Purity Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>rawFreq</td>
<td>0.73</td>
</tr>
<tr>
<td>condP</td>
<td>0.94</td>
</tr>
<tr>
<td>PMI</td>
<td>0.95</td>
</tr>
<tr>
<td>NPMI</td>
<td>0.91</td>
</tr>
<tr>
<td>MI</td>
<td>0.76</td>
</tr>
</tbody>
</table>

Table: Experimental Results for 12 adjectives and 142 dimensions

\[
Purity = 1 - \frac{\sum_{f \in F} \frac{1}{\log(f+1)}}{|C|}
\]

- C: ranks of correct adjectives on the respective scale
- F: ranks of false adjectives on the respective scale
## Conclusions

### Adjective Classification
- Separating property-denoting and relation-denoting adjectives is feasible from a theoretical perspective.
- Subclassification of property-denoting adjectives (attributes and roles) is difficult to grasp, even for human annotators.
- Classification scheme is difficult to use with corpus data.

### Vector Space Modelling
- Fits nicely with "bigger plan": paraphrasing adjective-noun phrases.
- Promising first results for the task of determining adjectival scales (without labelling them as yet).
- Explore vector space semantics for modelling attribute/role distinction.
- Evaluate VSM against sparseness of pattern-based approaches.
Thanks for your Attention!

Questions? Suggestions?