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

Background

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 $\Rightarrow$ table has an oval SHAPE
- fast car $\Rightarrow$ car that drives fast
- dangerous disease $\Rightarrow$ disease that infects/kills many people

Motivation

Adjectives as Gateways to Conceptual Representation

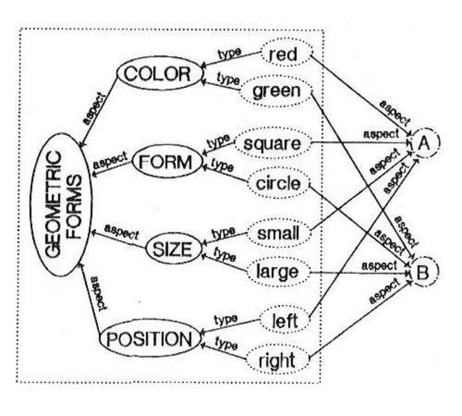

Figure: Frame Representation of Geometric Forms (Barsalou, 1992)

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...

Annotation Experiment

Goal

Is it feasible, in principle, to separate adjective-denoting adjectives from "others" ?

Initial Classification Scheme: BEO Classification

- **B**asic Adjectives, e.g.: *red carpet*
- **E**vent-related Adjectives, e.g.: *fast horse*
- **O**bject-related Adjectives, e.g.: *political debate*

[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*

BEO Classes (1) – continued

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`

BEO Classes (2)

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$_N$
  $\Rightarrow$ destruction$_N$ [of] the environment$_{N/ADJ}$
  $\Rightarrow$ destruction(e, AGENT: x, PATIENT: environment)
- political debate$_N$
  $\Rightarrow$ debate$_N$ [on] politics$_{N/ADJ}$
  $\Rightarrow$ debate(e, AGENT: x, TOPIC: 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` $\Rightarrow$ `ENT` on/of/from/... `N/ADJ`
- an `ADJ ENT` $\Rightarrow$ `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* $\Rightarrow$ $\textsc{color}$(carpet)=red
- *young bird* $\Rightarrow$ $\textsc{age}$(bird)=[?,?]

BEO Classes (3) – continued

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 ⇒ ?VELOCITY(horse)=fast
- good knife ⇒ ?QUALITY(knife)=good
- eloquent person ⇒ ?ELOQUENCE(person)=TRUE
- difficult problem ⇒ ?DIFFICULTY(problem)=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...

Tri-partite Classification: Annotator Agreement

	Annotator 1	Annotator 2	Annotator 3
Annotator 1	—	0.762	0.235
Annotator 2	0.762	—	0.285
Annotator 3	0.235	0.285	—

Table: Agreement figures in terms of Fleiss' κ

- 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)

	BASIC	EVENT	OBJECT	IMPOSSIBLE
κ	0.368	0.061	0.700	0.452

Table: Category-wise κ-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

	BASIC	EVENT	OBJECT
2:1 agreement	283	21	66
3:0 agreement	486	5	62

Table: Cases of Agreement vs. Disagreement

		1 voter		
		BASIC	EVENT	OBJECT
2 voters	BASIC	–	172	16
	EVENT	18	–	1
	OBJECT	54	10	–

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 ?

Play the Annotation Game ! (1)

Ambiguous Corpus Examples:

- `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.`
  $\longrightarrow$ annotators' votes: 2 `BASIC`, 1 `EVENT`
- `Any changes should only be introduced after` **proper** `research and costing, and after an initial experiment.`
  $\longrightarrow$ annotators' votes: 2 `BASIC`, 1 `EVENT`

Play the Annotation Game ! (2)

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.`
  $\longrightarrow$ 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.`
  $\longrightarrow$ annotators' votes: 2 BASIC, 1 EVENT

Distinguishing BASIC from EVENT Adjectives

- 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...

Bi-partite Classification: Annotator Agreement (category-wise)

	BASIC+EVENT	OBJECT	IMPOSSIBLE
κ	0.696	0.701	-0.003

Table: Category-wise κ-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 ?**

Founded vs. Inherent Properties ?

The notion of foundation (Guarino, 1992)

A concept α is called *founded* if there exists a concept β such that any instance χ of α is necessarily associated to an instance ψ of β which is not related to χ 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

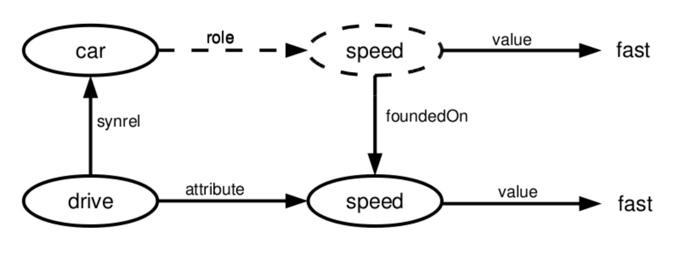

Figure: The SPEED role of cars

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

Pattern	EVENT	OBJECT	BASIC
ENT is ADJ for/at/... EVENT	+	−	−
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	+	−	−
A N ↔ R(N, N/ADJ)	−	+	−
predicative use	+	−	+
ENT is ADJ for an ENT	+	−	+
ENT has an ADJ ATTRIB	+/−	+/−	+
gradability	+	−	+
comparability	+	−	+
N/ADJ is an attribute of ENT	−	−	+

- 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...

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

	speed	color	price	beauty	height	dan-ger
fast	81	1	4	0	0	0
expensive	2	1	10	0	0	0
dangerous	0	0	0	0	2	3
drive	66	2	47	0	2	1
buy	3	13	73	3	0	1
paint	0	54	0	0	0	0
car	34	20	63	1	4	4
building	1	3	6	3	36	8

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):

$$[\![fast\ car]\!] = \vec{fast} \oplus \vec{car}$$

Example:

	speed	color	price	beauty	height	danger
fast	81	1	4	0	0	0
car	34	20	63	1	4	4
fast $\oplus$ *car*	115	21	67	1	4	4

A VSM for Adjective-Noun Phrases: Hypothesis II

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:

	speed	color	price	beauty	height	danger
fast	81	1	4	0	0	0
car	34	20	63	1	4	4
fast $\oplus$ *car*	115	21	67	1	4	4

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:

	speed	color	price	beauty	height	danger
fast	81	1	4	0	0	0
slow	54	0	1	0	0	0
expensive	2	1	10	0	0	0

A VSM for Adjective-Noun Phrases: Hypothesis IV

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 ?

	speed	color	price	...
grey $\oplus$ *cat*	2	18	3	...
grey $\oplus$ *building*	4	27	10	...
paint $\oplus$ *cat*	2	59	3	...
paint $\oplus$ *building*	4	68	10	...

2. Do these verbs indicate a valid role ?

	paint	boil	increase	...
cat	5	8	0	...
building	14	0	8	...
car	8	0	4	...

A VSM for Adjective-Noun Phrases: First Results

Hypothesis II: Adjectives from the same scale

Association Measure	Purity Score
rawFreq	0.73
condP	0.94
PMI	**0.95**
NPMI	0.91
MI	0.76

Table: Experimental Results for 12 adjectives and 142 dimensions

Purity Score

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

- 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 ?