Computational Models for Attribute Meaning in Adjectives and Nouns

Matthias Hartung
Computational Linguistics Department
Heidelberg University

September 30, 2011
Arlington, VA

Outline

Motivation

Relevance of Adjectives for Various NLP Tasks:

- ontology learning: attributes, roles, relations
- sentiment analysis: attributes
- coreference resolution: attributes
- information extraction: attributes, paraphrases
- information retrieval: paraphrases
- ...

Adjective Classification

Initial Classification Scheme: BEO

- We adopt an adjective classification scheme from the literature that reflects the different aspects of adjective semantics we are interested in:
  - **basic** adjectives $\longrightarrow$ *attributes*
    e.g.: *grey donkey*
  - **event-related** adjectives $\longrightarrow$ *roles*, *paraphrases*
    e.g.: *fast car*
  - **object-related** adjectives $\longrightarrow$ *relations*, *paraphrases*
    e.g.: *economic crisis*

(Boleda 2007; Raskin & Nirenburg 1998)

BEO Classification Scheme (1)

Basic Adjectives

Adjective denotes a value of an attribute exhibited by the noun:

- point or interval on a *scale*
- element in the set of discrete possible values

Examples

- *red carpet* $\Rightarrow$ $\textsc{color}$(carpet)=red
- *oval table* $\Rightarrow$ $\textsc{shape}$(table)=oval
- *young bird* $\Rightarrow$ $\textsc{age}$(bird)=[?,?]

BEO Classification Scheme (2)

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*
- *interesting book* ⇒ *book that is interesting to* **read**

BEO Classification Scheme (3)

Object-related Adjectives

- adjective is morphologically derived from a noun *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$_N$ [about] politics$_{N/ADJ}$
  $\Rightarrow$ debate(e, AGENT: x, TOPIC: politics)

Annotation Study

	BASIC	EVENT	OBJECT
κ	0.368	0.061	0.700

Table: Category-wise κ-values for all annotators

- BEO scheme turns out infeasible; overall agreement: $\kappa = 0.4$ (Fleiss 1971)
- separating the OBJECT class is quite feasible
- fundamental ambiguities between BASIC and EVENT class:
  - *fast car* $\equiv$ SPEED*(car)*=*fast*
  - *fast car* $\equiv$ *car that drives fast*

Re-Analysis of the Annotated Data

- `BASIC` and `EVENT` adjectives share an important commonality that blurs their distinctness !
- Re-analysis: **binary classification scheme**
  - adjectives denoting **properties** (`BASIC` & `EVENT`)
  - adjectives denoting **relations** (`OBJECT`)
- overall agreement after re-analysis: $\kappa = 0.69$

	`BASIC+EVENT`	`OBJECT`
κ	0.696	0.701

Table: Category-wise κ-values after re-analysis

Automatic Classification: Features

Group	Feature	Pattern

I	as
	comparative-1
	comparative-2
	superlative-1
	superlative-2
		`as` **JJ** `as`
		JJR `NN`
		`RBR` **JJ** `than`
		JJS `NN`
		`the RBS` **JJ** `NN`

II	extremely
	incredibly
	really
	reasonably
	remarkably
	very
		`an extremely` **JJ** `NN`
		`an incredibly` **JJ** `NN`
		`a really` **JJ** `NN`
		`a reasonably` **JJ** `NN`
		`a remarkably` **JJ** `NN`
		`DT very` **JJ**

III	predicative-use static-dynamic-1 static-dynamic-2	`NN (WP|WDT)? is|was|are|were RB?` **JJ** `NN is|was|are|were being` **JJ** `be RB?` **JJ** `.`
IV	one-proform	`a/an RB?` **JJ** `one`
V	see-catch-find	`see|catch|find DT NN` **JJ** *they saw the sanctuary desolate* *Baudouin's death caught the country unprepared*
VI	morph	adjective is morphologically derived from noun *economic* ← *economy*

Classification Results: Our Data

	PROP			REL			
	P	R	F	P	R	F	Acc
all-feat	**0.96**	**0.99**	**0.97**	0.79	0.61	0.69	**0.95**
all-grp	**0.96**	**0.99**	**0.97**	**0.85**	0.61	**0.71**	**0.95**
no-morph	0.95	0.96	0.95	0.56	0.50	0.53	0.91
morph-only	**0.96**	0.78	0.86	0.25	**0.67**	0.36	0.77
majority	*0.90*	*1.00*	*0.95*	*0.00*	*0.00*	*0.00*	*0.90*

- high precision for both classes
- recall on the `REL` class lags behind
- `morph`-feature is particularly valuable for `REL` class, but not very precise on its own

Classification Results: WordNet Data

	PROP			REL			
	P	R	F	P	R	F	Acc
all-feat	0.85	0.82	0.83	0.70	0.75	0.72	0.79
all-grp	**0.91**	0.80	**0.85**	**0.71**	**0.86**	**0.77**	0.82
no-morph	0.87	0.80	0.83	0.69	0.79	0.73	0.79
morph-only	0.80	**0.84**	0.82	0.69	0.64	0.66	0.77
majority	*0.64*	*1.00*	*0.53*	*0.00*	*0.00*	*0.00*	*0.64*

- `REL` class benefits from more balanced training data
- strong performance of *morph-only* baseline
- best performance due to a combination of `morph` and other features

Automatic Classification: Most Valuable Features

Group	Feature	Pattern
I	as	as **JJ** as
	comparative-1	**JJR** NN
	comparative-2	RBR **JJ** than
	superlative-1	**JJS** NN
	superlative-2	the RBS **JJ** NN
II	extremely	an extremely **JJ** NN
	incredibly	an incredibly **JJ** NN
	really	a really **JJ** NN
	reasonably	a reasonably **JJ** NN
	remarkably	a remarkably **JJ** NN
	very	DT very **JJ**
III	predicative-use	NN (WP|WDT)? is|was|are|were RB? **JJ**
	static-dynamic-1	NN is|was|are|were being **JJ**
	static-dynamic-2	be RB? **JJ** .
IV	one-proform	a/an RB? **JJ** one
V	see-catch-find	see|catch|find DT NN **JJ** *they saw the sanctuary desolate* *Baudouin's death caught the country unprepared*
VI	morph	adjective is morphologically derived from noun *economic ← economy*

Adjective Classification: Resume

- (automatically) separating property-denoting and relational adjectives is feasible
- largely language-independent feature set; results expected to carry over to different languages
- robust performance even without morphological resources
- classification on the type level; class volatility still acceptable
- open: attribute meaning evoked by a property-denoting adjective in context

Taking Stock...

Attribute Selection: Definition and Motivation

Characterizing Attribute Meaning in Adjective-Noun Phrases:

What are the **attributes** of a concept that are highlighted in an adjective-noun phrase ?

- *hot debate* $\longrightarrow$ EMOTIONALITY
- *hot tea* $\longrightarrow$ TEMPERATURE
- *hot soup* $\longrightarrow$ TASTE or TEMPERATURE

Goal:

- model attribute selection as a **compositional process** in a distributional VSM framework
- two model variants:
  1. pattern-based VSM
  2. combine dependency-based VSM with LDA **topic models**

Attribute Selection: Pattern-based VSM

	COLOR	DIRECT.	DURAT.	SHAPE	SIZE	SMELL	SPEED	TASTE	TEMP.	WEIGHT
enormous	1	1	0	1	45	0	4	0	0	21
ball	14	38	2	20	26	0	45	0	0	20
enormous $\times$ *ball*	14	38	0	20	**1170**	0	180	0	0	420
enormous $+$ *ball*	15	39	2	21	**71**	0	49	0	0	41

Main Ideas:

- reduce ternary relation ADJ-ATTR-N to binary ones
- vector component values: raw corpus frequencies obtained from lexico-syntactic patterns such as
  - (A1) `ATTR of DT? NN is|was JJ`
  - (N2) `DT ATTR of DT? RB? JJ? NN`
- reconstruct ternary relation by vector composition ($\times$, $+$)
- select most prominent component(s) from composed vector by entropy-based metric

Pattern-based Attribute Selection: Results

	MPC			ESel		
	P	R	F	P	R	F
$Adj \times N$	0.60	0.58	**0.59**	**0.63**	0.46	0.54
$Adj + N$	0.43	0.55	0.48	0.42	0.51	0.46
BL-Adj	0.44	0.60	0.50	0.51	0.63	0.57
BL-N	0.27	0.35	0.31	0.37	0.29	0.32
BL-P	0.00	0.00	0.00	0.00	0.00	0.00

Table: Attribute Selection from Composed Adjective-Noun Vectors

Remaining Problems of Pattern-based Approach:

- restriction to 10 manually selected attribute nouns
- rigidity of patterns entails sparsity

Using Topic Models for Attribute Selection

	ATTRIBUTE$_1$	ATTRIBUTE$_2$	ATTRIBUTE$_3$	⋮	⋮	⋮	ATTRIBUTE$_{n-2}$	ATTRIBUTE$_{n-1}$	ATTRIBUTE$_n$
enormous	?	?	?	?	?	?	?	?	?
ball	?	?	?	?	?	?	?	?	?
enormous $\times$ *ball*	?	?	?	?	?	?	?	?	?
enormous $+$ *ball*	?	?	?	?	?	?	?	?	?

Goals:

- combine pattern-based VSM with LDA topic modeling (cf. Mitchell & Lapata, 2009)
- challenge: reconcile TMs with categorial prediction task
- raise attribute selection task to large-scale attribute inventory

Using LDA for Lexical Semantics

LDA in Document Modeling (Blei et al., 2003)

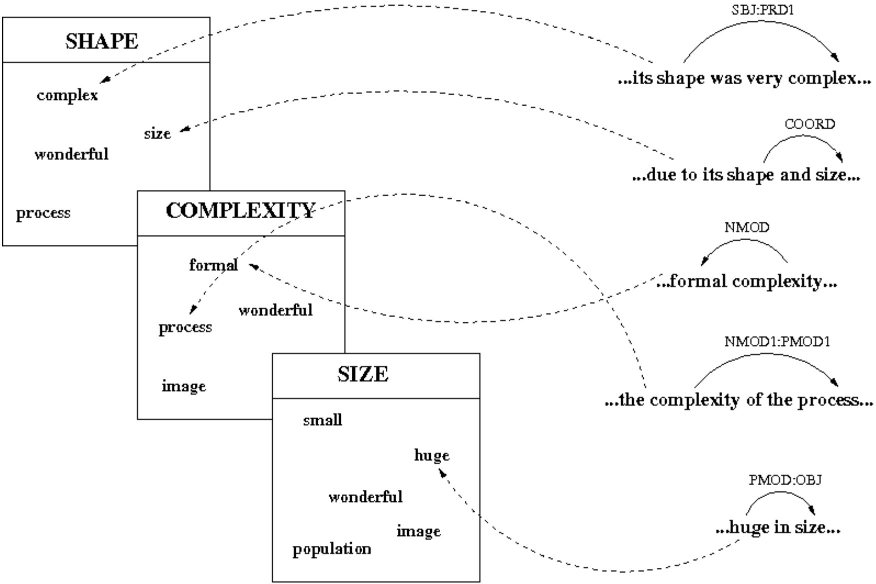
- hidden variable model for *document modeling*
- decompose collections of documents into *topics* as a more abstract way to capture their *latent semantics* than just BOWs

Porting LDA to Attribute Semantics

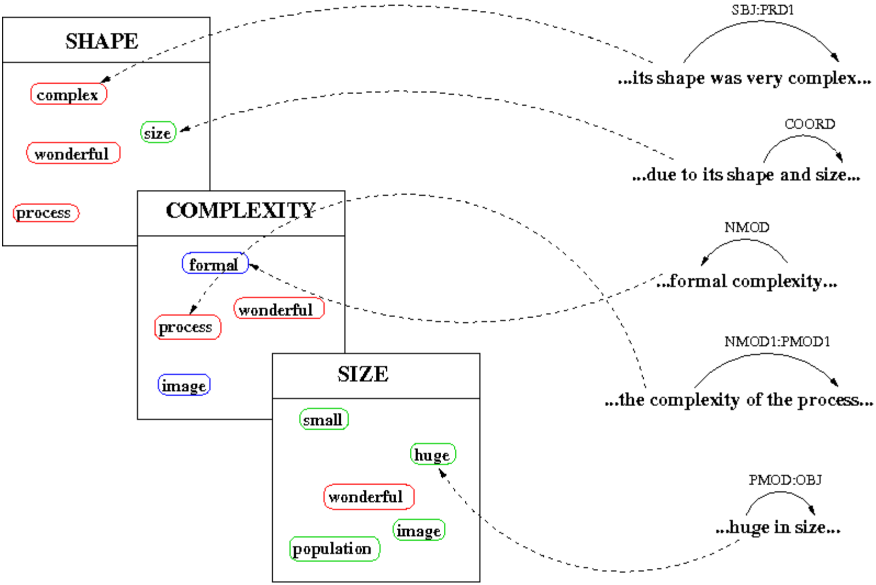
- "How do you modify LDA in order to be predictive for *categorial* semantic information (here: attributes) ?"
- build pseudo-documents[1] as distributional profiles of attribute meaning
- resulting topics are highly "attribute-specific"

[1] cf. Ritter et al. (2010), Ó Séaghdha (2010), Li et al. (2010)

C-LDA: "Pseudo-Documents" for Attribute Modeling



C-LDA: “Pseudo-Documents” for Attribute Modeling



Integrating C-LDA into the VSM Framework

	COLOR	DIRECT.	DURAT.	SHAPE	SIZE	SMELL	SPEED	TASTE	TEMP.	WEIGHT
hot	18	3	1	4	1	14	1	5	174	3
meal	3	5	119	10	11	5	4	103	3	33
hot $\times$ meal	0.05	0.02	0.12	0.04	0.01	0.07	0.00	**0.51**	**0.52**	0.10
hot $+$ meal	21	8	**120**	14	11	19	5	**108**	**177**	36

Table: VSM with C-LDA probabilities (scaled by 10^3)

Setting Vector Component Values:

$$v_{\langle w,a\rangle} = P(w|a) \approx P(w|d_a) = \sum_t P(w|t)P(t|d_a)$$

Attribute Selection with C-LDA: Results

	×			+		
	P	R	F	P	R	F
C-LDA	0.58	**0.65**	**0.61**	0.55	**0.66**	**0.61**
DepVSM	0.48	0.58	0.53	0.38	0.65	0.48
PattVSM	0.63	0.46	0.54	**0.71**	0.35	0.47

Table: Attribute selection over 10 attributes, × vs. +

- C-LDA: highest f-scores and recall over × and +
- baselines are competitive, but below LDA models
- C-LDA significantly outperforms PattVSM at a high margin (additive setting: +0.14 f-score)

Large-Scale Attribute Selection

Automatic Construction of Labeled Data from WordNet

```
Sense 1
hot (vs. cold)
       => temperature

Sense 3
hot (vs. cold)
       => emotionality, emotionalism
```

Large-Scale Attribute Selection

Automatic Construction of Labeled Data from WordNet

```
Sense 1
hot (vs. cold)
       => temperature

Sense 3
hot (vs. cold)
       => emotionality, emotionalism
```

```
1. hot -- (used of physical heat; having a high or higher than desirable tempera
ture or giving off heat or feeling or causing a sensation of heat or burning; "h
ot stove"; "hot water"; "a hot August day"; "a hot stuffy room"; "she's hot and
tired"; "a hot forehead")
2. hot, raging -- (characterized by violent and forceful activity or movement; v
ery intense; "the fighting became hot and heavy"; "a hot engagement"; "a raging
battle"; "the river became a raging torrent")
3. hot -- (extended meanings; especially of psychological heat; marked by intens
ity or vehemence especially of passion or enthusiasm; "a hot temper"; "a hot top
ic"; "a hot new book"; "a hot love affair"; "a hot argument")
```

Large-Scale Attribute Selection

Automatic Construction of Labeled Data from WordNet

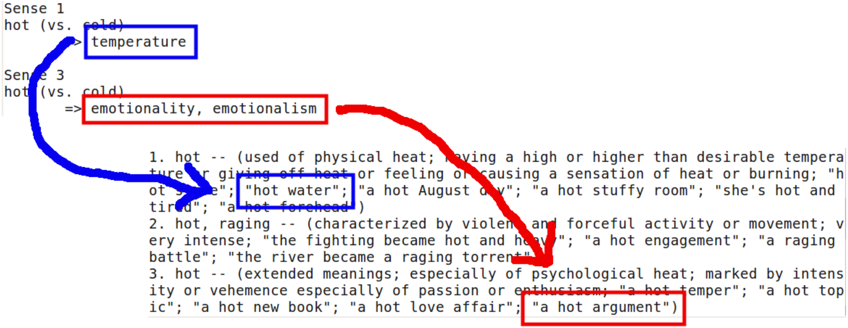
Sense 1
hot (vs. cold)
=> temperature

Sense 3
hot (vs. cold)
=> emotionality, emotionalism

1. hot -- (used of physical heat; having a high or higher than desirable temperature or giving off heat or feeling or causing a sensation of heat or burning; "hot stove"; "hot water"; "a hot August day"; "a hot stuffy room"; "she's hot and tired"; "a hot forehead")
2. hot, raging -- (characterized by violent and forceful activity or movement; very intense; "the fighting became hot and heavy"; "a hot engagement"; "a raging battle"; "the river became a raging torrent")
3. hot -- (extended meanings; especially of psychological heat; marked by intensity or vehemence especially of passion or enthusiasm; "a hot temper"; "a hot topic"; "a hot new book"; "a hot love affair"; "a hot argument")

Large-Scale Attribute Selection

Automatic Construction of Labeled Data from WordNet



Resulting Gold Standard:

- 345 phrases, each labeled with one out of 206 attributes

Large-Scale Attribute Selection: Results

	all		property	
	$\times$	$+$	$\times$	$+$
C-LDA	0.04	0.02	0.18	0.10
DepVSM	0.02	0.02	0.12	0.07

Table: Results on Large-Scale Attribute Selection (f-score)

- large-scale attribute selection is extremely difficult; very poor performance on the entire data set
- replication of the experiment on a subset of the data:
    - training attributes limited to 73 *property attributes*, test set restricted accordingly (113 adjective-noun phrases)
    - C-LDA gains more than $+0.10$ and significantly outperforms DepVSM in $\times$ setting

Large-Scale Attribute Selection: Results

	all		property	
	$\times$	$+$	$\times$	$+$
C-LDA	0.04	0.02	0.18	0.10
DepVSM	0.02	0.02	0.12	0.07

Table: Results on Large-Scale Attribute Selection (f-score)

- large-scale attribute selection is extremely difficult; very poor performance on the entire data set
- replication of the experiment on a subset of the data:
  - training attributes limited to 73 *property attributes*, test set restricted accordingly (113 adjective-noun phrases)
  - C-LDA gains more than $+0.10$ and significantly outperforms DepVSM in $\times$ setting

Large-Scale Attribute Selection: Negative Examples

	prediction	correct
serious book	DIFFICULTY	MIND
blue line	COLOR	UNION
weak president	POSITION	POWER
fluid society	REPUTE	CHANGEABLENESS
short flight	DISTANCE	DURATION
rough bark	TEXTURE	EVENNESS
faint heart	CONSTANCY	COWARDICE

Table: Sample of false predictions of C-LDA$_{\times}$

Error Analysis:

- "near misses": *weak president*, *rough bark*, *short flight*
- idiomatic expressions: *blue line*, *faint heart*, *fluid society*
- debatable WordNet labels: *serious book*

Large-Scale Attribute Selection: Positive Examples

	prediction	correct
thin layer	THICKNESS	THICKNESS
heavy load	WEIGHT	WEIGHT
shallow water	DEPTH	DEPTH
short holiday	DURATION	DURATION
attractive force	MAGNETISM	MAGNETISM
short hair	LENGTH	LENGTH

Table: Sample of correct predictions of C-LDA$_{\times}$

"Difficult" cases effectively modeled by C-LDA:

- ambiguous, context-dependent adjectives: *short holiday* vs. *short hair* vs. *short flight*
- cases that resist pattern-based modeling, e.g.: *thin layer* – `?the thickness of the * is thin`

Attribute Selection: Resume

- feasible task for a small set of 10 attributes
- pattern-based VSM yields highest precision
- sparsity can be largely mitigated by combination of dependency-based model and LDA
- large-scale attribute selection turns out extremely hard

Taking Stock...

Attribute-based VSMs for Similarity Prediction

Task:

- predict degree of *similarity* for pairs of adjective-noun phrases
- "common" distributional models: *sources* of similarity are usually disregarded
- attribute-based distributional meaning representations (AMRs): predict degree of similarity **and** its source

Example:

elderly lady vs. *old woman*

- high degree of similarity
- primary source of similarity: shared feature AGE

Similarity Prediction Experiment: Models and Data

Attribute-specific Model:

- **C-LDA:** *attributes* as interpreted dimensions of meaning for adjectives and nouns

Latent Model:

- **M&L:** 5w+5w context windows, 2000 most frequent *context words* as dimensions (Mitchell & Lapata, 2010)

Testing Data:

- human similarity judgements for 108 adj-noun phrases collected by Mitchell & Lapata (2010)
- evaluation: measure correlation between model similarity scores and human judgements in terms of Spearman's ρ

Similarity Prediction: Results

		$+$	$\times$	ADJ-only	N-only
262 attrs	C-LDA	0.19	0.15	0.17	0.11
	M&L	0.21	**0.34**	0.19	**0.27**
33 attrs	C-LDA	**0.23**	0.21	**0.27**	0.17
	M&L	0.21	**0.34**	0.19	**0.27**

- M&L$_{\times}$ performs best in both training scenarios
- C-LDA benefits from confined training data
- individual adjective and noun vectors produced by M&L and C-LDA show diametrically opposed performance

Outlook

- **improve noun representations by "space travel":**
  - enrich uninformative noun vectors in attribute space by their nearest neighbors in latent word space
- **expand and improve large-scale data set:**
  - semi-automatic acquisition of similar adj-noun phrases evoking the same attribute
  - manually determine ambiguous phrases (cf. *short flight*)
  - manually correct debatable labels and "near misses"
- **cover relational adjectives:**
  - parallels to SemEval Shared Task on Paraphrasing Noun Compounds (Nakov et al., 2010)