A Semi-supervised Type-based Classification of Adjectives: Distinguishing Properties and Relations

Matthias Hartung Anette Frank

Computational Linguistics Department
Heidelberg University

LREC 2010, Valletta

Motivation: Using Adjectives for Ontology Learning (1)

1. Learning Ontological Knowledge from Adjectives:

- **attributes**
  grey donkey ≡ COLOR(donkey)=grey
- **roles**, i.e. "founded" attributes (cf. Guarino, 1992)
  fast car ≡ SPEED(car)=fast
- **relations**
  economic crisis ≡ AFFECT(crisis, economy)

Different types of adjectives require different ontological representations !

Motivation: Using Adjectives for Ontology Learning (2)

2. Using Adjectives for Clustering Nouns into Concepts:

Clustering Features (pattern-based):

- attribute nouns:
  `the ATTR of the NOUN`
- adjectives denoting properties of the noun:
  `the ADJ NOUN`

Results:

- best results by combination of attribute and adjective features
- problem: attributive position is too unrestrictive for identifying property-denoting adjectives

(Almuhareb, 2006)

Adjective Classification for Ontology Learning

- **Hypothesis:** Classification is a prerequisite for ontology learning from adjectives.
- We adopt an adjective classification scheme from the literature that reflects the ontological information we are interested in:
  - attributes ≡ **basic** adjectives
    e.g.: *grey donkey*
  - roles ≡ **event-related** adjectives
    e.g.: *fast car*
  - relations ≡ **object-related** adjectives
    e.g.: *economic crisis*

(Boleda 2007; Raskin & Nirenburg 1998)

Overview

BEO Classification Scheme (1)

Basic Adjectives

- adjective denotes a value of an attribute exhibited by the noun
- values are either discrete or predications over a range of several values (depending on the concept being modified)

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* $\Rightarrow$ *knife that* **cuts** *well*
- *fast* *horse* $\Rightarrow$ *horse that* **runs** *fast*
- *interesting* *book* $\Rightarrow$ *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: 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

BASIC vs. EVENT

- fast horse
  - BASIC reading: SPEED(horse)=fast
  - EVENT reading: *horse that* **runs** *fast*
- good knife
  - BASIC reading: QUALITY(knife)=good
  - EVENT reading: *knife that* **cuts** *well*

Additional Instructions: Differentiation Patterns

If one of the following patterns holds for an ambiguous item, this indicates a property that is **founded** on an 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.

Category-wise Annotator Agreement

	BASIC	EVENT	OBJECT
κ	0.368	0.061	0.700

Table: Category-wise κ-values for all annotators

- overall agreement: $\kappa = 0.4$ (Fleiss 1971)
- separating the OBJECT class is quite feasible
- Can poor overall agreement be traced back to the ambiguities between BASIC and EVENT class ?

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

BASIC/EVENT ambiguity is the **primary source of disagreement** !

Re-Analysis of the Annotated Data

- People have substantial difficulties in distinguishing BASIC from EVENT adjectives !
- 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 for all annotators (after re-analysis)

Overview

Methodology

- task: automatically classify adjectives according to their denotation: **properties** (`ATTR`) vs. **relations** (`REL`)
- features: set of **lexico-syntactic patterns** capturing systematic differences of these adjective classes in certain grammatical constructions
- overcome feature sparsity:
  - classification on the **type level**
  - **semi-supervised** approach: acquire enough training material on the type level by heuristic annotation projection

Features for Classification

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*

Table: Set of features used for classification

Experimental Settings

Data Set

- manually annotated seed data ($\mathbf{A_s}$): 164 property-denoting, 18 relational adjective types
- heuristic annotation projection:
  - extract 5.000 sentences per type from ukWaC corpus ($\mathbf{A_{acq}}$)
  - for every adjective **token** in $\mathbf{A_{acq}}$: project unanimous class label from the corresponding **type** in $\mathbf{A_s}$

Evaluation

- several feature configurations:
  - `all-feat`: all features individually
  - `all-grp`: all features, collapsed into groups
  - `no-morph`: all features individually, without `morph` feature
- 10-fold cross validation
- baseline: label all instances with majority class (`ATTR`)

Experimental Results

	ATTR			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
Baseline	*0.90*	*1.00*	*0.95*	*0.00*	*0.00*	*0.00*	*0.90*

Table: Precision, recall and accuracy scores for Boosted Learner (10-fold cross-validation)

- high precision for both classes
- recall on the `REL` class lags behind
- `morph`-feature is highly valuable for `REL` class
- boosting benefits from collapsing sparse features into groups

Selective Evaluation of Class Volatility

Type	ATTR Tokens	REL Tokens	IMPOSS Tokens
beautiful (ATTR)	50	0	0
black (ATTR)	35	7	8
bright (ATTR)	45	1	4
heavy (ATTR)	42	0	8
new (ATTR)	50	0	0
civil (REL)	0	49	1
commercial (ATTR)	5	44	1
cultural (REL)	2	48	0
environmental (REL)	0	48	2
financial (REL)	0	46	4

Table: Volatility of prototypical class members

- average class volatility on the token level: 8.6%
- rough estimate of the error introduced by raising the classification task to the type level

Conclusions

Prospects of adjective classification for ontology learning:

- attribute/role distinction on the basis of adjectives alone is difficult even for human judges
- property-denoting and relational adjectives can be automatically distinguished at high precision for both classes
  - even with small and skewed training data
  - even in the absence of a morphological lexicon (see paper)

What else ?

- classification on the type level is justified by tolerable degree of class volatility
- shallow feature set should be easily applicable to specialized domains and adaptable to different languages

Thank you for your attention !
Any questions ?