Exploring Supervised LDA Models for Assigning Attributes to Adjective-Noun Phrases

Matthias Hartung   **Anette Frank**
Computational Linguistics Department
Heidelberg University

EMNLP 2011
Edinburgh, July 28

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

Goals and Challenges:

- model attribute selection as a **compositional process** in a distributional VSM framework
- data sparsity: combine VSM with LDA **topic models**
- assess model on a **large-scale** attribute inventory

Attribute Selection: Previous Work (I)

Almuhareb (2006):

- goal: learn binary **adjective-attribute relations**
- pattern-based approach:

`the ATTR of the * is|was ADJ`

Problems:

- semantic contribution of the noun is neglected
- severe sparsity issues
- limited coverage: 10 attributes

Attribute Selection: Previous Work (II)

Pattern-based VSM: Hartung & Frank (2010)

	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 × *ball*	14	38	0	20	**1170**	0	180	0	0	420
enormous + *ball*	15	39	2	21	**71**	0	49	0	0	41

- 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`
- **remaining problems:**
  - restriction to 10 manually selected attribute nouns
  - rigidity of patterns still entails sparsity

Attribute Selection: New Approach

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

Goals:

- combine attribute-based VSM of Hartung & Frank (2010) with LDA topic modeling (cf. Mitchell & Lapata, 2009)
- challenge: reconcile TMs with categorial prediction task
- raise attribute selection task to large-scale attribute inventory

Outline

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)

Two Variants of LDA-based Attribute Modeling

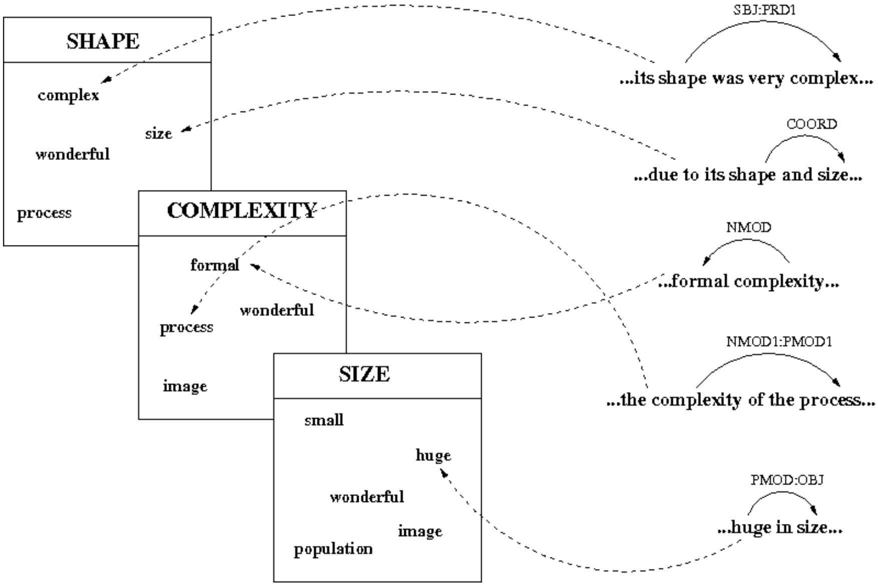
Controled LDA (C-LDA):

- documents are heuristically equated with attributes
- full range of topics available for each document
- generative process: standard LDA (Blei et al., 2003)

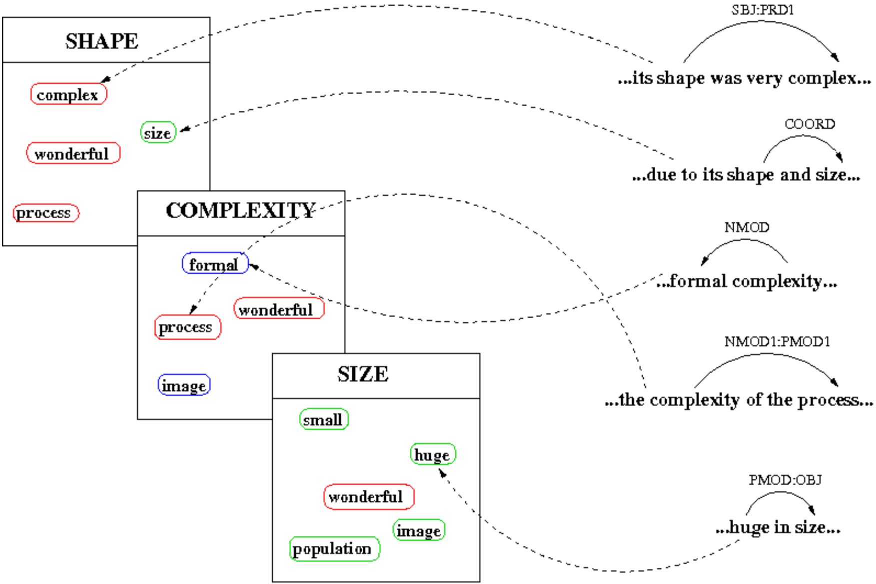
Labeled LDA (L-LDA; Ramage et al., 2009)

- documents are explicity labeled with attributes
- 1:1-relation between labels and topics
- only topics corresponding to attribute labels are available for each document

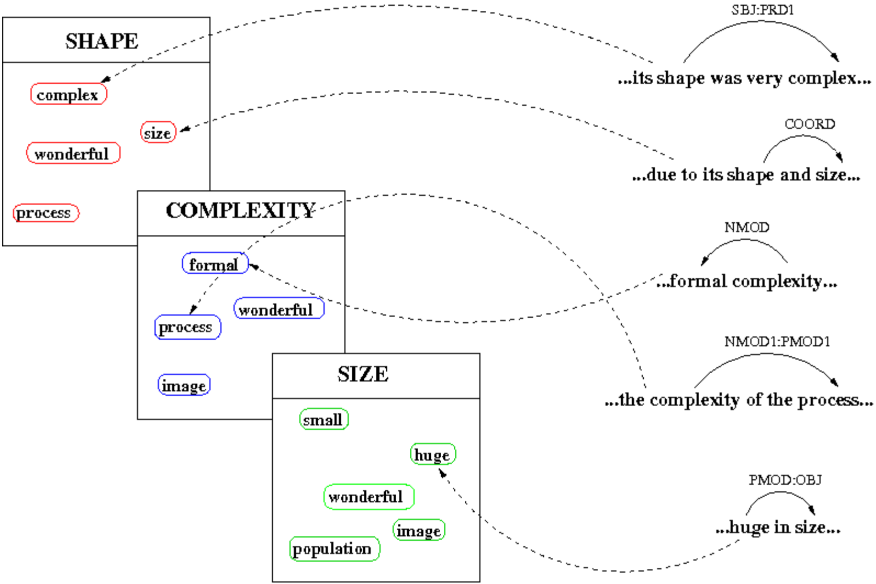
C-LDA: "Pseudo-Documents" for Attribute Modeling



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



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



Integrating Attribute Models into the VSM Framework (I)

	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:

- **C-LDA:**

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

- **L-LDA:**

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

Integrating Attribute Models into the VSM Framework (I)

	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:

- **C-LDA:**

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

- **L-LDA:**

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

Integrating Attribute Models into the VSM Framework (I)

	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:

- **C-LDA:**

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

- **L-LDA:**

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

Integrating Attribute Models into the VSM Framework (I)

	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 × 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:

- **C-LDA:**

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

- **L-LDA:**

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

Integrating Attribute Models into the VSM Framework (II)

Vector Composition Operators:

- component-wise multiplication ($\times$)
- component-wise addition ($+$)

(Mitchell & Lapata, 2010)

Attribute Selection from Composed Vectors:

Entropy Selection (ESel):

- select flexible number of most *informative* vector components
- "empty selection" in case of very broad, flat vectors

(Hartung & Frank, 2010)

Taking Stock...

Experimental Setup

Experiments:

1. attribute selection over 10 attributes
2. attribute selection over 206 attributes

Methodology:

- gold standards for evaluation:
  - Experiment 1: 100 adj-noun phrases, manually labeled by human annotators
  - Experiment 2: compiled from WordNet
- baselines:
  - **PattVSM:** pattern-based VSM of Hartung & Frank (2010)
  - **DepVSM:** dependency-based VSM (constructed from pseudo-documents without feeding them to LDA machinery)
- evaluation metrics: precision, recall, f_1-score

Experiment 1: Results

	×			+		
	P	R	F	P	R	F
C-LDA	0.58	**0.65**	**0.61**^{L,P}	0.55	**0.66**	**0.61**^{D,P}
L-LDA	**0.68**	0.54	0.60^{D}	0.53	0.57	0.55^{D,P}
DepVSM	0.48	0.58	0.53^{P}	0.38	0.65	0.48^{P}
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 +
- statistically significant differences between C-LDA and L-LDA for ×, not for +
- baselines are competitive, but below LDA models
- both LDA models significantly outperform PattVSM at a high margin (additive setting: +0.14/+0.08 f-score)

Experiment 1: Different Topic Settings for C-LDA

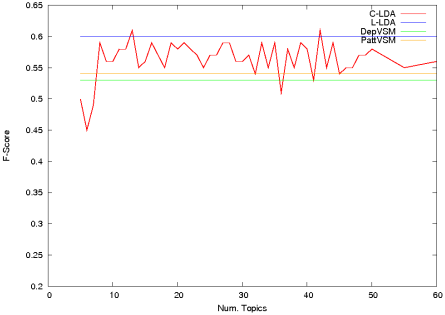


Figure: C-LDA$_\times$, different topic numbers

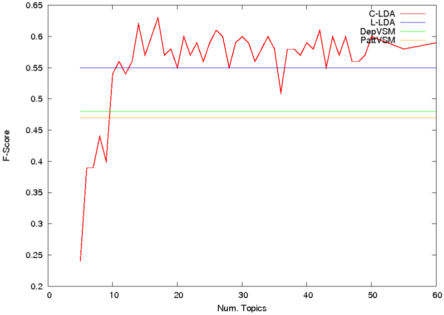


Figure: C-LDA$_+$, different topic numbers

- very few performance drops below the baselines
- C-LDA almost constantly outperforms L-LDA in the $+$ setting
- L-LDA turns out more robust in the $\times$ setting, but can still be outperformed by C-LDA in individual configurations

Experiment 1: Smoothing Power of LDA Models

	×			+		
	P	R	F	P	R	F
C-LDA	**0.39**	**0.31**	**0.35**	**0.43**	**0.33**	**0.38**
L-LDA	0.30	0.18	0.23	0.34	0.16	0.22
DepVSM	0.20	0.10	0.13	0.16	0.17	0.17
PattVSM	0.00	0.00	0.00	0.13	0.04	0.06

Table: Performance on sparse vectors (× vs. +)

- focused evaluation on subset of 22 adjective-noun phrases affected by "zero vectors" in the PattVSM model
- C-LDA provides best smoothing power across all settings, outperforming PattVSM by orders of magnitude
- higher figures for + in general, as the models can recover from sparsity by using only one vector in this setting

Experiment 2: Large-Scale Attribute Selection

Automatic Construction of Labeled Data

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

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

Experiment 2: Large-Scale Attribute Selection

Automatic Construction of Labeled Data

```
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")
```

Experiment 2: Large-Scale Attribute Selection

Automatic Construction of Labeled Data

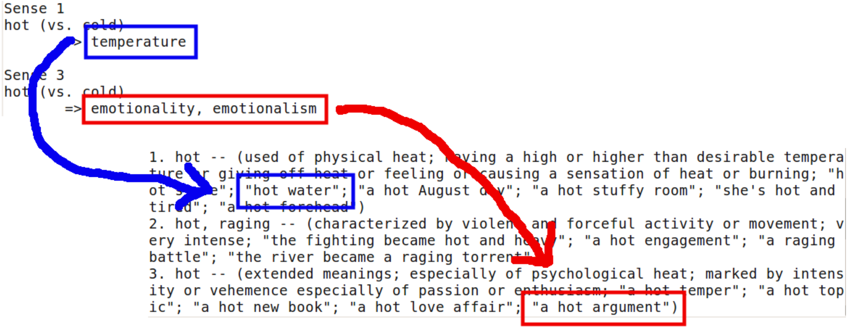
Sense 1
hot (vs. cold)
=> temperature

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")

Experiment 2: Large-Scale Attribute Selection

Automatic Construction of Labeled Data



Resulting Gold Standard

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

Experiment 2: Results

	all		property	
	$\times$	$+$	$\times$	$+$
C-LDA	0.04	0.02	$0.18^{L,D}$	0.10^{D}
L-LDA	0.03	0.04	0.15	0.15
DepVSM	0.02	0.02	0.12	0.07

Table: Results in Experiment 2 (f-score)

- large-scale attribute selection is extremely difficult; very poor performance of all models 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)
  - all models gain (more than) $+0.10$ in $\times$ setting
  - largest improvement for C-LDA

Experiment 2: Results

	all		property	
	$\times$	$+$	$\times$	$+$
C-LDA	0.04	0.02	$0.18^{L,D}$	0.10^{D}
L-LDA	0.03	0.04	0.15	0.15
DepVSM	0.02	0.02	0.12	0.07

Table: Results in Experiment 2 (f-score)

- large-scale attribute selection is extremely difficult; very poor performance of all models 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)
  - all models gain (more than) $+0.10$ in $\times$ setting
  - largest improvement for C-LDA

Experiment 2: Performance of Individual Attributes

	all			property		
	P	R	F	P	R	F
WIDTH	0.67	1.00	0.80	1.00	0.50	0.67
WEIGHT	0.80	0.57	0.67	0.50	0.57	0.53
MAGNETISM	0.50	1.00	0.67			
SPEED	0.50	0.50	0.50	1.00	0.50	0.67
TEXTURE	0.33	1.00	0.50	0.33	1.00	0.50
DURATION	0.50	0.50	0.50	1.00	1.00	1.00
TEMPERATURE	0.30	0.75	0.43	0.43	0.75	0.55
AGE	0.33	0.50	0.40			
THICKNESS	1.00	0.25	0.40	0.50	0.13	0.20
DEGREE	1.00	0.20	0.33			
LENGTH	0.17	1.00	0.29	0.50	1.00	0.67
DEPTH	1.00	0.14	0.25	1.00	0.86	0.92
ACTION	0.17	0.50	0.25			
LIGHT	0.33	0.17	0.22	0.20	0.17	0.18
POSITION	0.14	0.25	0.18	0.20	0.25	0.22
SHARPNESS				1.00	1.00	1.00
SERIOUSNESS				0.50	1.00	0.67
COLOR	0.13	0.25	0.17	0.29	0.50	0.36
LOYALTY				1.00	1.00	1.00
average	0.49	0.54	0.51	0.63	0.63	0.63

Table: C-LDA$_{\times}$, best attributes (F>0)

Complete Setting:

- large-scale approach is not a complete failure, but effective for a subset of attributes
- 50% of attributes from Exp. 1 successfully modeled

Property Setting:

- further improvement on average
- decrease of individual property attributes: some non-property attributes bear discriminative power as well

Experiment 2: Performance of Individual Attributes

	all			property		
	P	R	F	P	R	F
WIDTH	0.67	1.00	0.80	1.00	0.50	0.67
WEIGHT	0.80	0.57	0.67	0.50	0.57	0.53
MAGNETISM	0.50	1.00	0.67			
SPEED	0.50	0.50	0.50	1.00	0.50	0.67
TEXTURE	0.33	1.00	0.50	0.33	1.00	0.50
DURATION	0.50	0.50	0.50	1.00	1.00	1.00
TEMPERATURE	0.30	0.75	0.43	0.43	0.75	0.55
AGE	0.33	0.50	0.40			
THICKNESS	1.00	0.25	0.40	0.50	0.13	0.20
DEGREE	1.00	0.20	0.33			
LENGTH	0.17	1.00	0.29	0.50	1.00	0.67
DEPTH	1.00	0.14	0.25	1.00	0.86	0.92
ACTION	0.17	0.50	0.25			
LIGHT	0.33	0.17	0.22	0.20	0.17	0.18
POSITION	0.14	0.25	0.18	0.20	0.25	0.22
SHARPNESS				1.00	1.00	1.00
SERIOUSNESS				0.50	1.00	0.67
COLOR	0.13	0.25	0.17	0.29	0.50	0.36
LOYALTY				1.00	1.00	1.00
average	0.49	0.54	0.51	0.63	0.63	0.63

Table: C-LDA$_\times$, best attributes (F$>$0)

Complete Setting:

- large-scale approach is not a complete failure, but effective for a subset of attributes
- 50% of attributes from Exp. 1 successfully modeled

Property Setting:

- further improvement on average
- decrease of individual property attributes: some non-property attributes bear discriminative power as well

Experiment 2: Qualitative Analysis (I)

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$_{\times}$ in Experiment 2

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

Experiment 2: Qualitative Analysis (II)

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$_\times$ in Experiment 2

“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 layer is thin*

Conclusions

Achieved so far:

- LDA-based attribute models: correspondence between latent topics and ontological attributes
- integration of attribute models into VSM framework improves performance on attribute selection task over 10 attributes
- first approach to large-scale attribute selection: highly challenging endeavor, feasible only for a subset of attributes

Open Issues:

- reasons for unequal performance of individual attributes still widely unclear
- individual quality of noun vectors lags behind adjectives; cf. Hartung & Frank (2011) for details

References (I)

- Almuhareb, Abdulrahman (2006): Attributes in Lexical Acquisition. Ph.D. Dissertation, Department of Computer Science, University of Essex.
- Baroni, Marco, Silvia Bernardini, Adriano Ferraresi and Eros Zanchetta (2009): The WaCky Wide Web. A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation 43(3)*: 209-226.
- Blei, David, Andrew Ng and Michael Jordan (2003): Latent Dirichlet Allocation. *Journal of Machine Learning Research 3*: 993-1022.
- Hartung, Matthias & Anette Frank (2011): Assessing Interpretable Attribute-related Meaning Representations for Adjective-Noun Phrases in a Similarity Prediction Task. Proceedings of the Workshop on Geometrical Models of Semantics (GEMS), Edinburgh, UK.
- Hartung, Matthias & Anette Frank (2010): A Structured Vector Space Model for Hidden Attribute Meaning in Adjective-Noun Phrases. Proceedings of COLING, Beijing, China: 430-438.
- Li, Linlin, Benjamin Roth & Caroline Sporleder (2010): Topic models for word sense disambiguation and token-based idiom detection. Proceedings of ACL: 1138-1147.

References (II)

- Mitchell, Jeff & Mirella Lapata (2009): Language Models Based on Semantic Composition. Proceedings of EMNLP, Singapore: 430-439.
- Mitchell, Jeff & Mirella Lapata (2010): Composition in Distributional Models of Semantics. *Cognitive Science 34(8)*: 1388-1429.
- Ó Séaghdha, Diarmuid (2010): Latent Variable Models of Selectional Preference. Proceedings of ACL: 435-444.
- Ramage, Daniel, David Hall, Ramesh Nallapati and Christopher D. Manning (2009): Labeled LDA. A Supervised Topic Model for Credit Attribution in Multi-labeled Corpora. Proceedings of EMNLP, Singapore: 248-256.
- Ritter, Alan, Mausam & Oren Etzioni (2010): A Latent Dirichlet Allocation Method for Selectional Preferences. Proceedings of ACL: 424-434.

Thanks...

...for your attention.
Questions ?

Please consider also to attend our presentation at the GEMS 2011 workshop:

Assessing Interpretable, Attribute-related Meaning Representations for Adjective-Noun Phrases in a Similarity Prediction Task

Sunday, July 31, 2:30 PM

Backup Slides

C-LDA: Generative Process

For each topic $k \in \{1, \ldots, K\}$:

  Generate $\beta_k \sim Dir_V(\eta)$

For each document d:

  Generate $\theta_d \sim Dir(\alpha)$

  For each n in $\{1, \ldots, N_d\}$:

    Generate $z_{d,n} \sim Mult(\theta_d)$ with $z_{d,n} \in \{1, \ldots, K\}$

    Generate $w_{d,n} \sim Mult(\beta_{z_{d,n}})$ with $w_{d,n} \in \{1, \ldots, V\}$

(Blei et al., 2003)

L-LDA: Generative Process

For each topic $k \in \{1, \ldots, K\}$:
  Generate $\beta_k = (\beta_{k,1}, \ldots, \beta_{k,V})^T \sim Dir(\cdot \mid \eta)$
For each document d:
  For each topic $k \in \{1, \ldots, K\}$
    Generate $\Lambda_k^{(d)} \in \{0, 1\} \sim Bernoulli(\cdot \mid \Phi_k)$
  Generate $\alpha^{(d)} = L^{(d)} \times \alpha$
  Generate $\theta^{(d)} = (\theta_{l_1}, \ldots, \theta_{l_{M_d}})^T \sim Dir(\cdot \mid \alpha^{(d)})$
  For each i in $\{1, \ldots, N_d\}$:
    Generate $z_i \in \{\lambda_1^{(d)}, \ldots, \lambda_{M_d}^{(d)}\} \sim Mult(\cdot \mid \theta^{(d)})$
    Generate $w_i \in \{1, \ldots, V\} \sim Mult(\cdot \mid \beta_{z_i})$

Comments:

Generating document's labels $\Lambda_k^{(d)}$ for each topic k results in:

- vector of document labels $\lambda^{(d)} = \{k | \Lambda_k^{(d)} = 1\}$
- document-specific label projection matrix $L^{(d)}_{\lambda^{(d)} \times K}$ with

$$L_{ij}^{(d)} = \begin{cases} 1 & \text{if } \lambda_i^{(d)} = j \\ 0 & \text{otherwise} \end{cases}$$

(Ramage et al., 2009)

L-LDA: Generative Process

For each topic $k \in \{1, \ldots, K\}$:
    Generate $\beta_k = (\beta_{k,1}, \ldots, \beta_{k,V})^T \sim Dir(\cdot \mid \eta)$
For each document d:
    For each topic $k \in \{1, \ldots, K\}$
        Generate $\Lambda_k^{(d)} \in \{0, 1\} \sim Bernoulli(\cdot \mid \Phi_k)$
    Generate $\alpha^{(d)} = L^{(d)} \times \alpha$
    Generate $\theta^{(d)} = (\theta_{l_1}, \ldots, \theta_{l_{M_d}})^T \sim Dir(\cdot \mid \alpha^{(d)})$
    For each i in $\{1, \ldots, N_d\}$:
        Generate $z_i \in \{\lambda_1^{(d)}, \ldots, \lambda_{M_d}^{(d)}\} \sim Mult(\cdot \mid \theta^{(d)})$
        Generate $w_i \in \{1, \ldots, V\} \sim Mult(\cdot \mid \beta_{z_i})$

Comments:

Use matrix $L^{(d)}$ to project the Dirichlet topic prior α to a lower-dimensional vector $\alpha^{(d)}$ whose topic dimensions correspond to the document labels.

(Ramage et al., 2009)

L-LDA: Generative Process

For each topic $k \in \{1, \ldots, K\}$:
    Generate $\beta_k = (\beta_{k,1}, \ldots, \beta_{k,V})^T \sim Dir(\cdot \mid \eta)$
For each document d:
    For each topic $k \in \{1, \ldots, K\}$
        Generate $\Lambda_k^{(d)} \in \{0, 1\} \sim Bernoulli(\cdot \mid \Phi_k)$
    Generate $\alpha^{(d)} = L^{(d)} \times \alpha$
    Generate $\theta^{(d)} = (\theta_{l_1}, \ldots, \theta_{l_{M_d}})^T \sim Dir(\cdot \mid \alpha^{(d)})$
    For each i in $\{1, \ldots, N_d\}$:
        Generate $z_i \in \{\lambda_1^{(d)}, \ldots, \lambda_{M_d}^{(d)}\} \sim Mult(\cdot \mid \theta^{(d)})$
        Generate $w_i \in \{1, \ldots, V\} \sim Mult(\cdot \mid \beta_{z_i})$

Comments:

Use lower-dimensional vector $\alpha^{(d)}$ to generate topic proportions $\theta^{(d)}$ for the respective document d.

(Ramage et al., 2009)

Experiment 1: Attribute Selection over 10 Attributes

Creation of an Annotated Data Set

- partially random sample of adjective-noun phrases from 386 property-denoting adjectives $\times$ 216 nouns
- three human annotators

Resulting Gold Standard

- 76 phrases with 1.13 attributes on average, 24 "empty" phrases
- inter-annotator agreement: $\kappa = 0.67$

(Hartung & Frank, 2010)

L-LDA: Alternative Setting