Topic-Based and Cross-Lingual Scaling of Political Text

Federico Nanni, Goran Glavas and Simone Paolo Ponzetto

Data and Web Science Group
University of Mannheim

{federico,goran,simone}@informatik.uni-mannheim.de

Why is it Important?

1. Scholars want to **estimate** substantive quantities of interest
2. With the web we have **abundance** of textual **sources**
3. Computational methods are therefore **needed**

Established Approaches

Supervised -> Wordscores (Laver, Benoit, and Garry, 2003)

Unsupervised -> Wordfish (Slapin and Proksch, 2008)

Limitations

Limitations

1. No automatic topic-based comparison
2. No semantic analyses - only word frequency
3. No cross-lingual comparison
4. The evaluation is not straightforward

Today's Talk

1. Detecting topics in political texts
2. Topic-based scaling
3. Evaluation of scaling of political texts
4. Cross-lingual scaling

Detecting Topics

Detecting Topics is Complicated!

Detecting Topics is Complicated!

Supervised approaches (Naive-Bayes, SVM, etc):

A. Define a set of topics in advance
B. Prepare training data for each topic
C. Classify new text in topics
D. Evaluate the results

Detecting Topics is Complicated!

Supervised approaches (Naive-Bayes, SVM, etc):

A. Define a set of topics in advance
B. Prepare training data for each topic
C. Classify new text in topics
D. Evaluate the results

Difficult step!

Detecting Topics is Complicated!

Unsupervised approaches (topic models, clustering):

A. Choose a number of topics in advance
B. Group texts by topics
C. Evaluate the results

Detecting Topics is Complicated!

Unsupervised approaches (topic models, clustering):

A. Choose a number of topics in advance

B. Group texts by topics

C. Evaluate the results ← Difficult step!

Unsupervised Topic Detection

First step: text segmentation!

Unsupervised Topic Detection

First step: text segmentation!

Linear text segmentation aims at representing texts as sequences of semantically coherent segments.

Glavaš, G., Nanni, F. and Ponzetto, S.P. Unsupervised Text Segmentation Using Semantic Relatedness Graphs.* SEM 2016.

Text Segmentation on Choi Dataset

Method	3-5	
	P_k	WD
Choi (2000)	12.0	–
Brants et al. (2002)	7.4	–
Fragkou et al. (2004)	5.5	–
Misra et al. (2009)	23.0	–
GRAPHSEG	$\boxed{5.6}$	8.7
Misra et al. (2009)*	2.2	–
Riedl and Biemann (2012)*	1.2	1.3

Glavaš, G., Nanni, F. and Ponzetto, S.P. Unsupervised Text Segmentation Using Semantic Relatedness Graphs.* SEM 2016.

Text Segmentation on Manifestos

Method	P_k	WD
Random baseline	40.60	49.17
Riedl and Biemann (2012)	33.39	38.31
GRAPHSEG	28.09	34.04
Riedl and Biemann (2012)*	32.94	37.59
GRAPHSEG*	$\boxed{28.08}$	34.00

Table 3: Performance on the Manifesto dataset (*domain-adapted variant).

Unsupervised Topic Detection - Too difficult!

1. Text segmentation already complicated
2. Results difficult to evaluate

Supervised Topic Detection

ClassyMan

We developed a **7 classes** coarse-grained classifier using CMP categories.

Zirn, C., Glavaš, G., Nanni, F., Eichorst, J. and Stuckenschmidt, H., Classifying Topics and Detecting Topic Shifts in Political Manifestos. PolText 2016.

ClassyMan

We developed a **7 classes** coarse-grained classifier using CMP categories.

Features:

1. Bag of words of each sentence
2. Topic of the previous sentence
3. Semantic similarity between previous and current sentence
4. Level relevance of each word in sentence for each class

→ Linear Support Vector Machine

ClassyMan on Manifestos

Topic	P	R	F_1
External Rel.	83.7	86.6	85.1
Freedom & Dem.	68.0	59.9	63.7
Pol. system	69.7	65.7	67.6
Economy	73.9	77.4	75.6
Welfare & QoL	72.8	72.8	72.8
Fabric of Soc.	74.8	76.0	75.4
Soc. Groups	71.2	67.9	69.5
Micro-avg.	74.9	74.9	74.9

ClassyMan on Political Speeches

We collected all US campaign speeches for the 2008, 2012 and 2016 presidential elections.

Work at paragraph level.

Gold Standard. Contains altogether 9 speeches from 2008, 2012, and 2016 elections (around 1k paragraphs).

ClassyMan on Political Speeches

Model	F_1
ClassyMan	
Standard SVM	
ClassySpeech	

ClassyMan on Political Speeches

Model	F_1
ClassyMan	36.2
Standard SVM	
ClassySpeech	

ClassyMan on Political Speeches

Model	F_1
ClassyMan	36.2
Standard SVM	71.2
ClassySpeech	$\boxed{78.6}$

Detecting Topics - Still an Open Problem

1. Difficult to obtain in-domain training data
2. Time-expensive to evaluate
3. Cross-domain not always works

Today's Talk

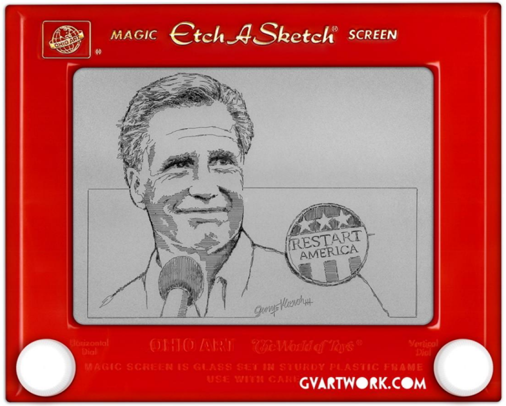
1. Detecting topics in political texts ✓
2. Topic-based scaling
3. Evaluation of scaling of political texts
4. Cross-lingual scaling

Mitt Romney's Etch-a-Sketch

Nanni, F., Zirn, C., Glavaš, G., Eichorst, J. and Ponzetto, S.P. TopFish: topic-based analysis of political position in US electoral campaigns. PolText 2016.

Mitt Romney's Etch-a-Sketch

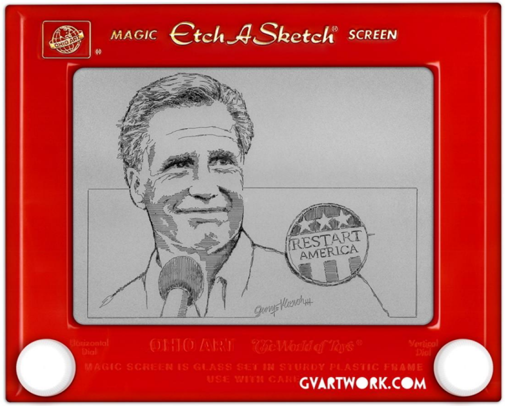
"Everything changes. It's **almost like an Etch-a-Sketch**. You can kind of shake it up and restart all over again."



Nanni, F., Zirn, C., Glavaš, G., Eichorst, J. and Ponzetto, S.P. TopFish: topic-based analysis of political position in US electoral campaigns. PolText 2016.

Mitt Romney's Etch-a-Sketch

“Everything changes. It’s **almost like an Etch-a-Sketch**. You can kind of shake it up and restart all over again.”



Starting Hypothesis: For certain **topics** these changes might be **more prominent** than for others

Topic-Based Scaling

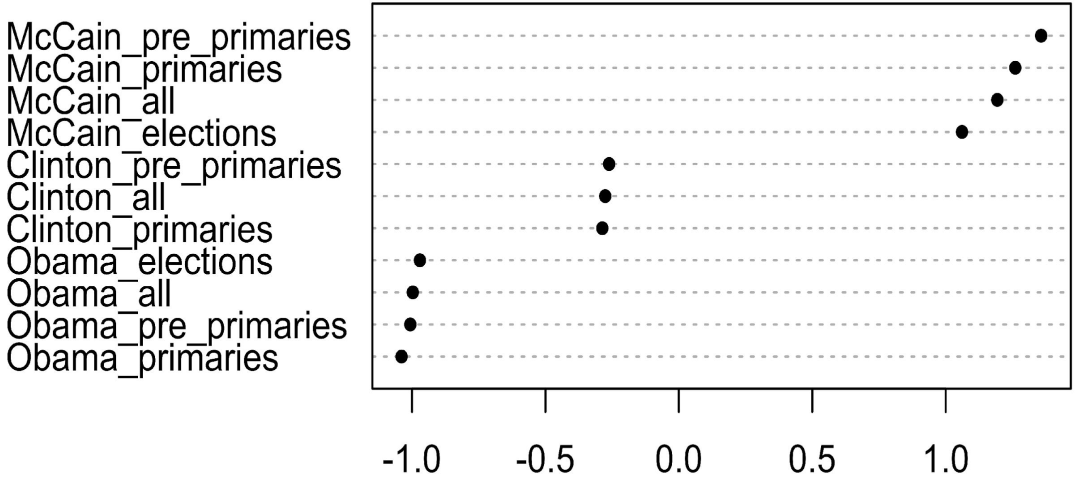
Using data from 2008 and 2012 elections.

For each candidate, we concatenate all paragraphs with the **same topic and same phase** (pre-primaries, primaries, elections).

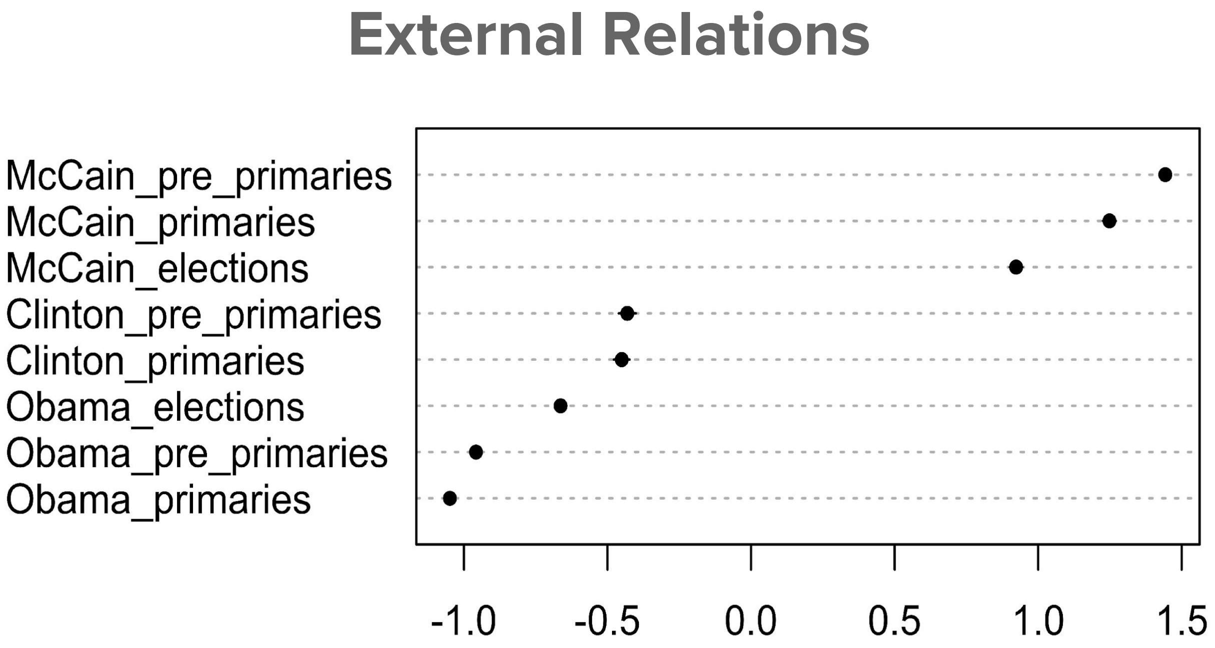
Finally, we feed each phase-topic slice to the **Wordfish** tool.

Coarse-Grained Analysis (2008)

First we analyzed the **general positions**.

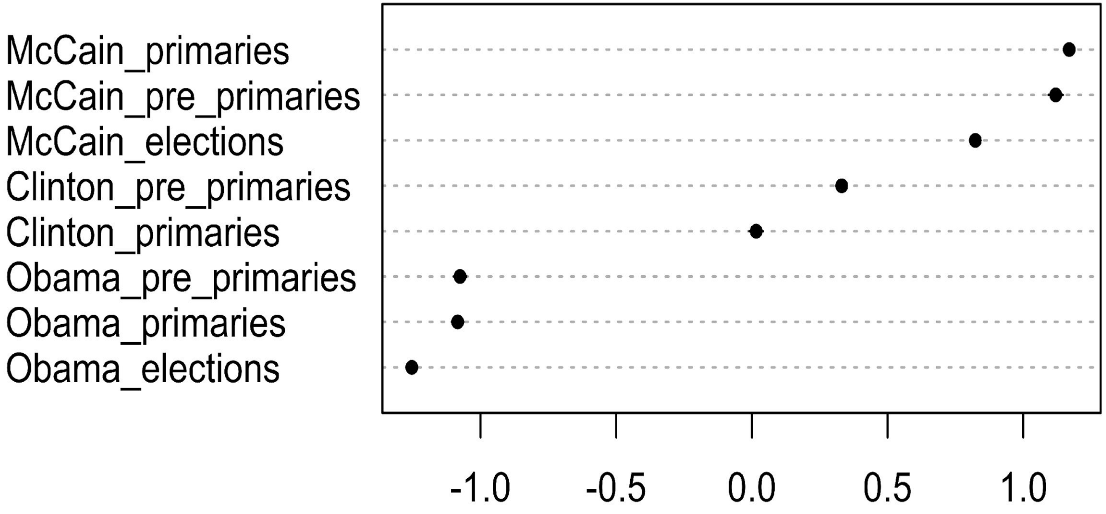


Fine-Grained Analysis (2008)



Fine-Grained Analysis (2008)

Welfare and Quality of Life



Today's Talk

1. Detecting topics in political texts ✓
2. Topic-based scaling ✓
3. Evaluation of scaling of political texts
4. Cross-lingual scaling

Evaluating Text Scaling

Evaluating Text Scaling

Check if they replicate experts opinion (example the Chapel Hill Expert Survey).

Evaluating Text Scaling

Check if they replicate experts opinion (example the Chapel Hill Expert Survey).

Evaluation metrics :

a. Pairwise Accuracy (PA), i.e., the percentage of pairs with parties in the same order as in the gold standard

b. Spearman and Pearson correlation between the two sets of positions

Today's Talk

1. Detecting topics in political texts ✓
2. Topic-based scaling ✓
3. Evaluation of scaling of political texts ✓
4. Cross-lingual scaling

Cross-Lingual Scaling

Glavaš, G., Nanni, F., Ponzetto, S. P. Unsupervised Cross-Lingual Scaling of Political Texts. EACL 2017.

Cross-Lingual Scaling

We need:

a. To capture semantic similarities between texts (i.e. parties)
b. To map them across languages

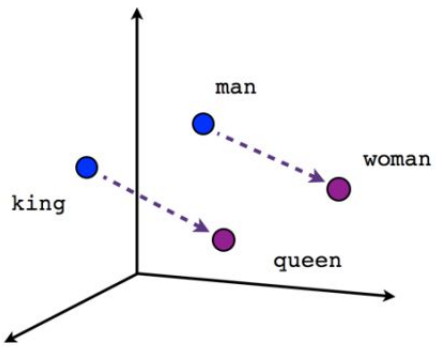
Glavaš, G., Nanni, F., Ponzetto, S. P. Unsupervised Cross-Lingual Scaling of Political Texts. EACL 2017.

How to Capture Semantic

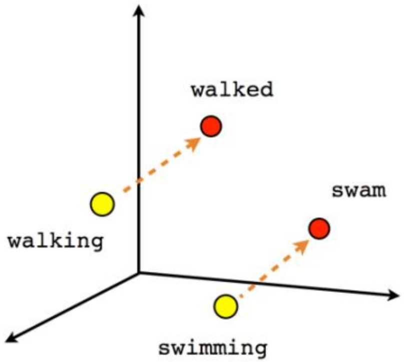
We use pre-trained word embeddings (Mikolov et al. 2013).

How to Capture Semantic

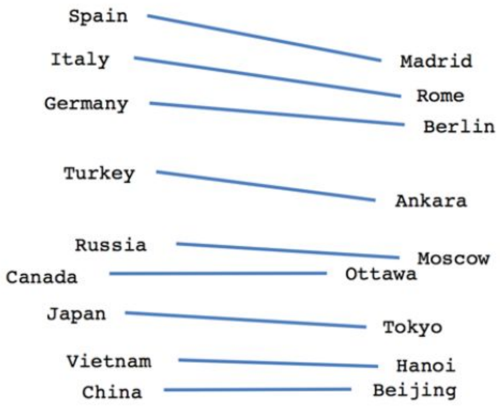
We use pre-trained word embeddings (Mikolov et al. 2013).



Male-Female



Verb tense



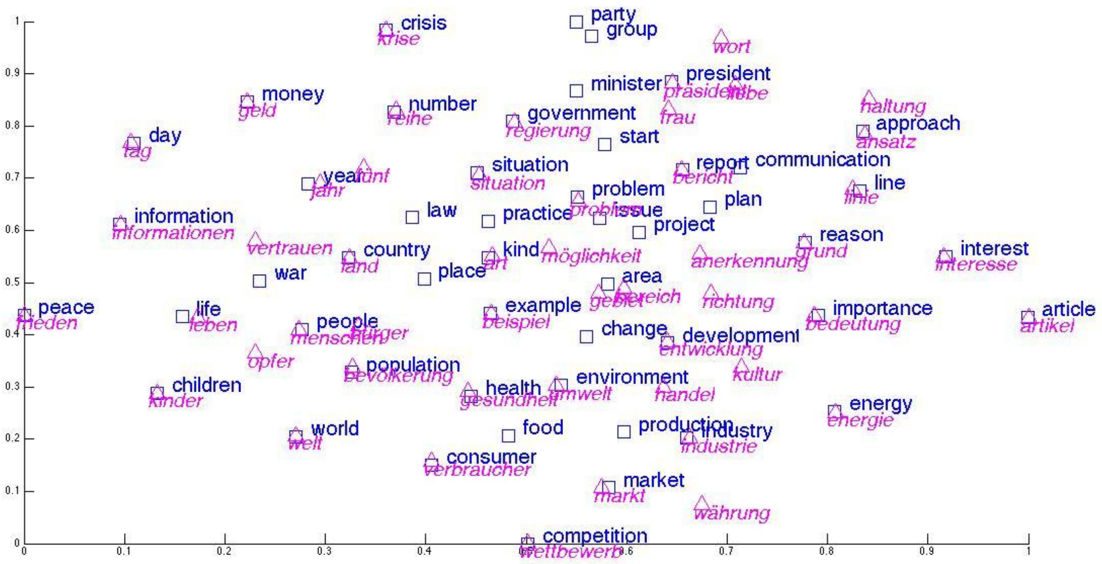
Country-Capital

Cross-Lingual Mapping

Linear translation matrix model to map them across languages.

Cross-Lingual Mapping

Linear translation matrix model to map them across languages.



Measuring Semantic Similarities

A. **Alignment similarity:** We pair words between the two documents that have the most similar embedding vectors.

B. **Aggregation similarity:** We compare the aggregate semantic vectors of entire input texts.

Graph-Based Scaling

1. Identifying the two most dissimilar pivot texts
2. Harmonic function label propagation for propagating position scores
3. Re-scaling of pivot texts

Evaluation

	Monolingual			Cross-lingual		
	PA	r_P	r_S	PA	r_P	r_S
Random						
Wordfish						
AL-HFLP						
AGG-HFLP						

Table 2: Scaling performance for the left-to-right ideological positioning.

Evaluation

	Monolingual			Cross-lingual		
	PA	r_P	r_S	PA	r_P	r_S
Random	49.7	-.03	.00			
Wordfish	55.0	.21	.20			
AL-HFLP	61.3	.35	.31			
AGG-HFLP	**67.0**	**.53**	**.46**			

Table 2: Scaling performance for the left-to-right ideological positioning.

Evaluation

	Monolingual			Cross-lingual		
	PA	r_P	r_S	PA	r_P	r_S
Random	49.7	-.03	.00	49.7	-.03	.00
Wordfish	55.0	.21	.20	–	–	–
AL-HFLP	61.3	.35	.31	57.3	.20	.25
AGG-HFLP	**67.0**	**.53**	**.46**	**63.3**	**.34**	**.39**

Table 2: Scaling performance for the left-to-right ideological positioning.

Wrap-Up and Future Work

Wrap-Up and Future Work

Topic-based analyses could improve text scaling, however:

a. Unsupervised approaches are difficult to evaluate

b. Supervised approaches need in-domain training data

Wrap-Up and Future Work

Topic-based analyses could improve text scaling, however:

a. Unsupervised approaches are difficult to evaluate

b. Supervised approaches need in-domain training data

Evaluation of text scaling is **important**!

Wrap-Up and Future Work

Topic-based analyses could improve text scaling, however:

a. Unsupervised approaches are difficult to evaluate
b. Supervised approaches need in-domain training data

Evaluation of text scaling is **important**!

Cross-lingual scaling is possible.

Wrap-Up and Future Work

Topic-based analyses could improve text scaling, however:

a. Unsupervised approaches are difficult to evaluate
b. Supervised approaches need in-domain training data

Evaluation of text scaling is **important**!

Cross-lingual scaling is possible.

Next step: Cross-lingual topic detection and scaling.

Questions?

Federico Nanni

Data and Web Science Group
University of Mannheim

federico@informatik.uni-mannheim.de

Questions?

Federico Nanni

Data and Web Science Group
University of Mannheim

federico@informatik.uni-mannheim.de

Additional Info

Unsupervised Topic Detection

Lauscher, Anne, et al. "Entities as Topic Labels: Combining Entity Linking and Labeled LDA to Improve Topic Interpretability and Evaluability." IJCoL-Italian Journal of Computational Linguistics 2.2 (2016).

Why Topic Models are Awesome

They are able to identify the most important topics in a collection of documents.

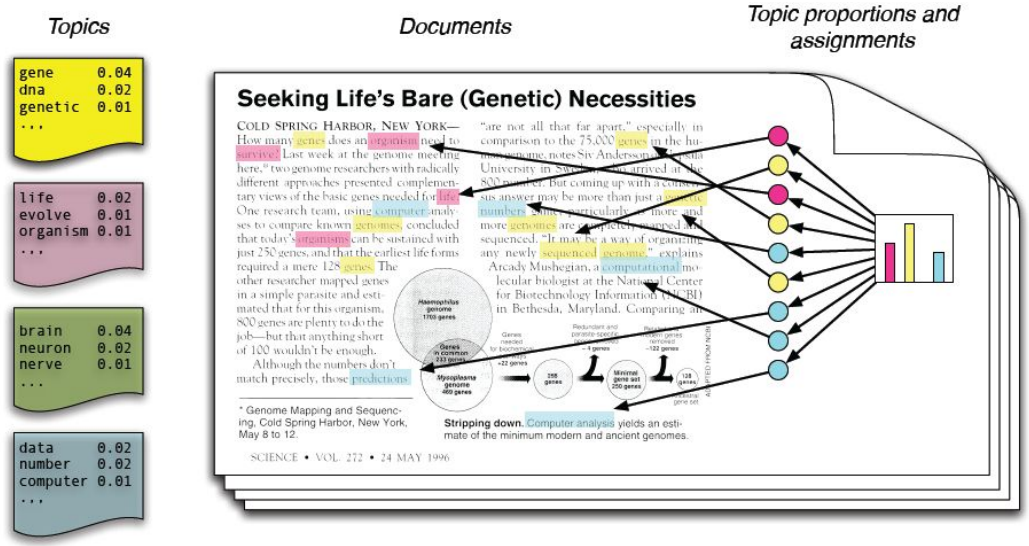


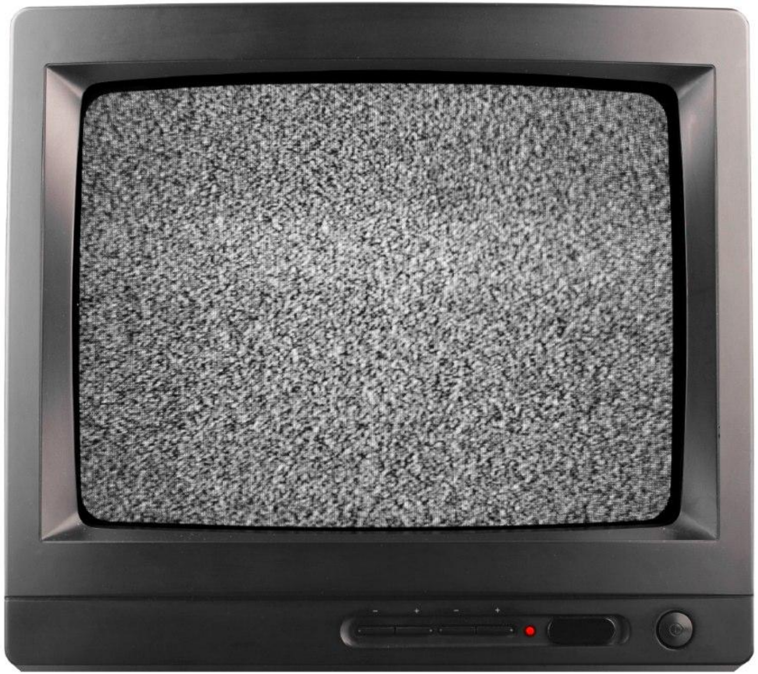
Figure from Blei (2012)

Why Topic Models are NOT Awesome

The topics obtained are difficult to interpret and evaluate.

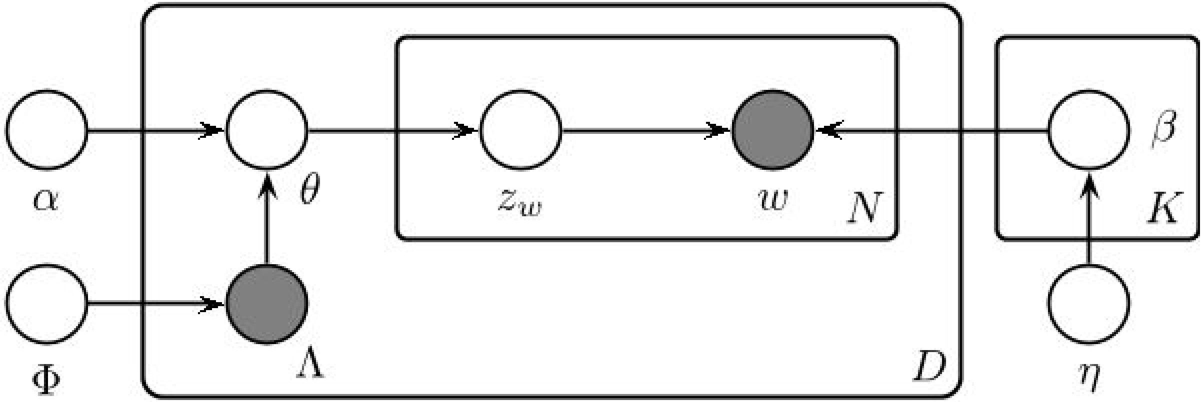
Why Topic Models are NOT Awesome

The topics obtained are difficult to interpret and evaluate.



Labeled LDA

Each **document** is described with one or more **labels**, each label is associated with a specific **topic** (Ramage et al., 2009).

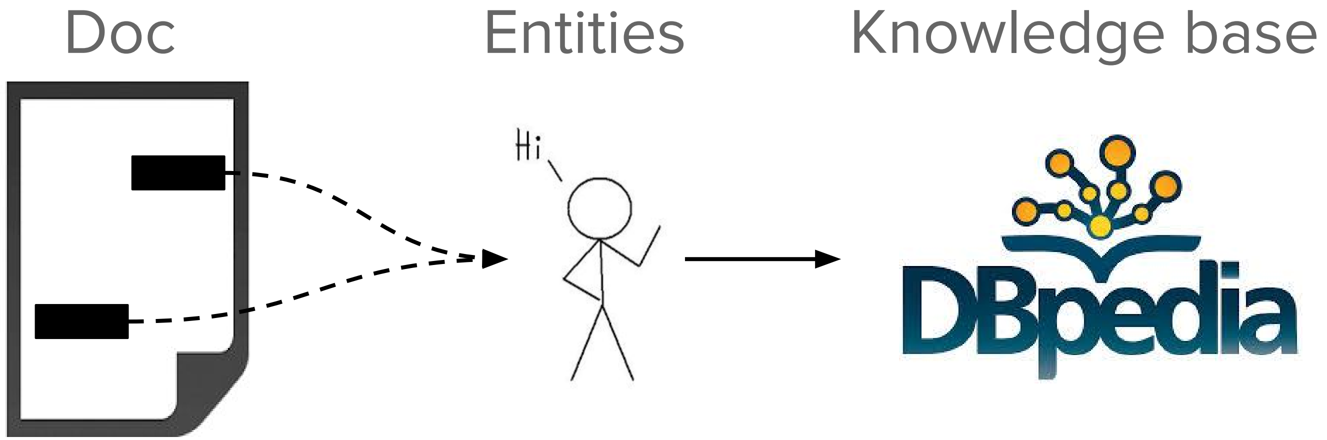


How to Automatically Obtain Labels?

Different approaches:

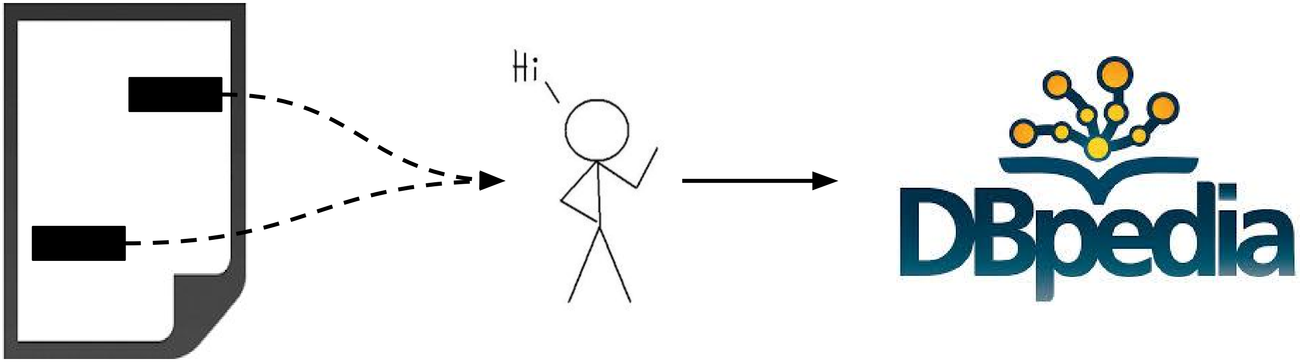
- Keyphrase digger (FBK - Trento)
- Labeling the obtained topics (Hulpus et al., 2014)

Our Approach: Entities



Our Approach: Entities

Doc | Entities | Knowledge base



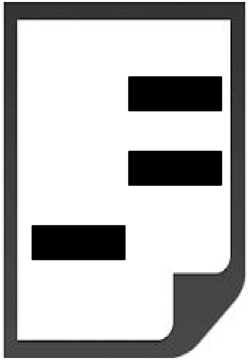
"Word sense disambiguation and entity linking together!"
Babelfy

TAGME

DBpedia Spotlight

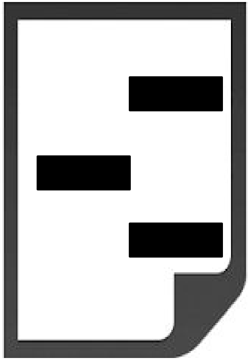
Our Approach: Entity Ranking

Doc1



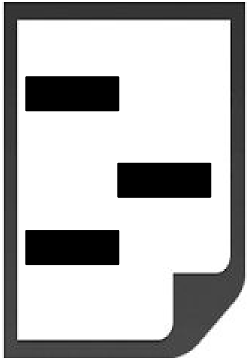
Entity1
Entity2
Entity3

Doc2



Entity1
Entity2
Entity3

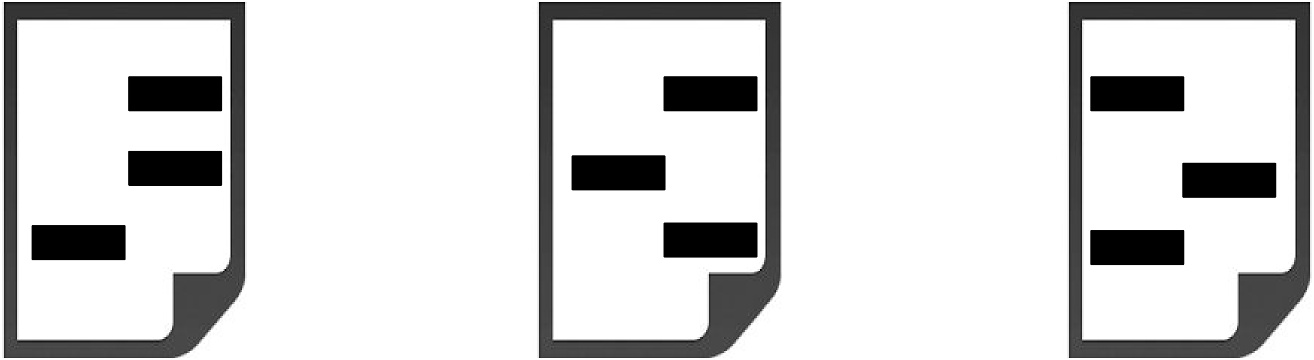
Doc3



Entity1
Entity2
Entity3

Our Approach: Entities as Topic Labels

Doc1 | Doc2 | Doc3



Doc1	Doc2	Doc3
Entity1	Entity1	Entity1
Entity2	Entity2	Entity2
Entity3	Entity3	Entity3

Labeled LDA!

Europarl Corpus

Examined most relevant topics addressed by each party in the European Parliament’s fifth term (1999-2004).

Les Verts (France)

	Label: Consumer (47%)	**Label: GMO** (34%)
Topic words	product directive consumer safety law market	human health agreement food measure sustainable

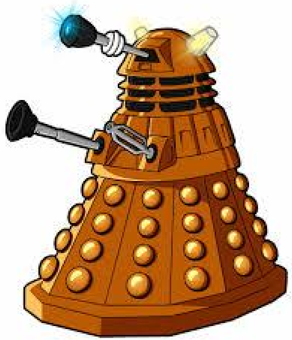
Europarl Corpus

Examined most relevant topics addressed by each party in the European Parliament's fifth term (1999-2004).

Les Verts (France)

	Label: Consumer (47%)	**Label: GMO** (34%)
Topic words	product directive consumer safety law market	human health agreement food measure sustainable

EVALUATE!
EVALUATE!
EVALUATE!



How to Evaluate it?

1. Label selection

Les Verts (France)

Consumer **GMO**

How to Evaluate it?

1. Label selection
2. Label ranking

Les Verts (France)

Consumer (**47%**) GMO (**34%**)

How to Evaluate it?

1. Label selection
2. Label ranking
3. Label-topic relation

Les Verts (France)

Consumer (47%)	GMO (34%)
↕	↕
product	human
directive	health
consumer	agreement
safety	food
law	measure
market	sustainable

Label Selection and Ranking

Country Weather
Limiting magnitude
Marginal cost
Economic growth
Brand management
Barriers to entry
Fax
Three Mile Island accident
George W. Bush
Cambridge Energy Research Associates
Macroeconomics
Customer relationship management
Loving Every Minute (album)
Wave function
Drought

Is the country facing sustained tight markets?:
In general, capacity margins have been falling over the last few years as utilities have refraind from building baseload capacity, and others have focused on developing primarily peaking capacity. The last large building boom of coal plants ended in early 1970's and nuclear boom dropped sharply after Three Mile Island incident in 1979. During 1990's, projected capacity margins have fallen from the 15 - 20 % toward 10% and below. Outside of the California situation, NYC also poses a potential risk for this summer.
If general, however, by 2002 - 2003, the amount of capacity proposed in each region more than covers normal load growth for meeting peak hour demand. Remaining question relates to performance of existing coal and nuclear stacks, also what happens during periods of persistent drought.

Will voaltility and prices remain high or lessen?:
To the extent that more capacity becomes merchant oriented focusing in marginal cost economics, and transmission congestion persists, increased volatility will continue, especially in ISO/pool type environments. Prices will reflect primary fuel dynamics, especially the interplay between gas and oil. The case for lower prices will reflect an overbuild scenario beyond this year, coupled with slowing economic activity and low incidence of extreme weather events.

How fast can generation be added and what returns should be expected?
The variability in development of greenfield capacity depends on time to permit at state or local levels. Construction time is fairly constant. In general 18 months is reasonable time frame from concept to first fire. To the extent that power plants are project financed with minimum 30 equity, returns should be consistent with other comparable project financed opportunities available to fund managers. We do not expect any more fully debt financed facilities in the near term.

Can companies make major profits owning generation long term?

Label Selection and Ranking

Country Weather
Limiting magnitude
Marginal cost
Economic growth
Brand management
Barriers to entry
Fax
Three Mile Island accident
George W. Bush
Cambridge Energy Research Associates
Macroeconomics
Customer relationship management
Loving Every Minute (album)
Wave function
Drought

Is the country facing sustained tight markets?:
In general, capacity margins have been falling over the last few years as utilities have refraind from building baseload capacity, and others have focused on developing primarily peaking capacity. The last large building boom of coal plants ended in early 1970's and nuclear boom dropped sharply after Three Mile Island incident in 1979. During 1990's, projected capacity margins have fallen from the 15 - 20 % toward 10% and below. Outside of the California situation, NYC also poses a potential risk for this summer.
If general, however, by 2002 - 2003, the amount of capacity proposed in each region more than covers normal load growth for meeting peak hour demand. Remaining question relates to performance of existing coal and nuclear stacks, also what happens during periods of persistent drought.

Will voaltility and prices remain high or lessen?:
To the extent that more capacity becomes merchant oriented focusing in marginal cost economics, and transmission congestion persists, increased volatility will continue, especially in ISO/pool type environments. Prices will reflect primary fuel dynamics, especially the interplay between gas and oil. The case for lower prices will reflect an overbuild scenario beyond this year, coupled with slowing economic activity and low incidence of extreme weather events.

How fast can generation be added and what returns should be expected?
The variability in development of greenfield capacity depends on time to permit at state or local levels. Construction time is fairly constant. In general 18 months is reasonable time frame from concept to first fire. To the extent that power plants are project financed with minimum 30 equity, returns should be consistent with other comparable project financed opportunities available to fund managers. We do not expect any more fully debt financed facilities in the near term.

Can companies make major profits owning generation long term?

Mean Average Precision

Topic-Label Relation

GMO

?

product	human
directive	health
consumer	agreement
safety	food
law	measure
market	sustainable

Topic-Label Relation

Consumer **GMO**

?

product
directive
consumer
safety
law
market

Evaluation - Label Identification

	Mean number of labels selected	Precision on user selection	Number of documents without annotation
EuroParl	4.04	0.93	4
EnronCorpus	3.88	0.97	10
ClintonCorpus	4.25	0.98	1

Evaluation - Label Ranking and Selection

	@2			@3			@4			MAP
	P	R	F_1	P	R	F_1	P	R	F_1	
EuroParl	0.65	0.32	0.44	0.52	0.39	0.45	0.45	0.56	0.50	0.51
EnronCorpus	0.76	0.35	0.48	0.61	0.41	0.49	0.48	0.53	0.50	0.40
ClintonCorpus	0.65	0.28	0.39	0.53	0.34	0.41	0.50	0.52	0.51	0.48

Evaluation - Topic-Label Relation

	Label Mode		Term Mode	
	Accuracy	# Skipped	Accuracy	# Skipped
EuroParl	0.59	2	0.38	5
EnronCorpus	0.80	4	0.75	14
ClintonCorpus	0.71	6	0.56	2

Conclusion

Entity-labels **improve** topic interpretability.

However it is necessary to **always evaluate** them.

Code available:

1. Pipeline: https://github.com/anlausch/TMELPipeline
2. Eval. platform: https://github.com/anlausch/TMEvaluationPlatform

	Monolingual			Cross-lingual		
	PA	r_P	r_S	PA	r_P	r_S
Random	49.7	-.03	.00	49.7	-.03	.00
Wordfish	55.0	.21	.20	–	–	–
AL-HFLP	61.3	.35	.31	57.3	.20	.25
AGG-HFLP	**67.0**	**.53**	**.46**	**63.3**	**.34**	**.39**

Table 2: Scaling performance for the left-to-right ideological positioning.

	Monolingual			Cross-lingual		
	PA	r_P	r_S	PA	r_P	r_S
Random	49.1	.00	.00	49.1	.00	.00
Wordfish	59.7	.18	.33	–	–	–
AL-HFLP	**62.3**	**.25**	**.39**	**64.3**	**.54**	**.40**
AGG-HFLP	60.3	.24	.30	59.3	.48	.31

Table 3: Scaling performance for the positioning regarding European integration.