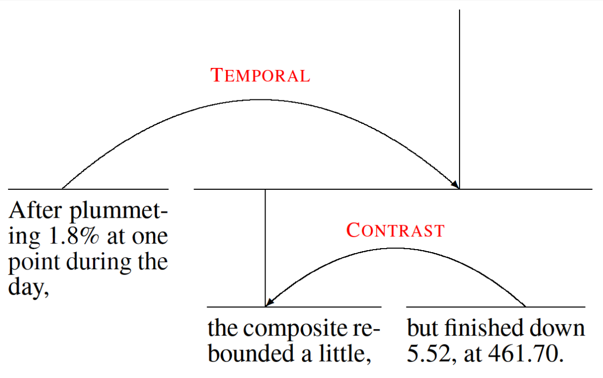
Rhetorical Relations for Information Retrieval

Christina Lioma*, Birger Larsen**, Wei Lu†

*University of Copenhagen, Denmark
**Royal School of Library and Information Science, Denmark
†Wuhan University, China

What are rhetorical relations?



How text components are linked to each other (*discourse structure*). Semantic functions: temporal, contrast, condition, cause ...

Why use rhetorical relations for IR?

Not new

Notes on Semantic Discourse Structure, KSJ 1967

> *"...understanding the message of a text involves some knowledge of the way concepts may be or are usually combined..."*

- Goal: Retrieval methods that bring us closer to this understanding

How do we use rhetorical relations for IR?

Assumption: we can identify rhetorical relations in documents
Aim: plug them into the ranking function

How do we use rhetorical relations for IR?

Assumption: we can identify rhetorical relations in documents
Aim: plug them into the ranking function

- Standard query likelihood: Probability of generating q from a model induced by d

$$p(q|d)$$

How do we use rhetorical relations for IR?

Assumption: we can identify rhetorical relations in documents
Aim: plug them into the ranking function

- Standard query likelihood: Probability of generating q from a model induced by d

$$p(q|d)$$

- Query likelihood with rhetorical relations: Probability of generating q from a model induced by d and by its rhetorical relations $rh \in d$

$$p(q|d, rh) \cdot p(rh|d)$$

Query likelihood with rhetorical relations

Breaking it down

$$p(q|d, rh) \cdot p(rh|d)$$

$p(rh|d)$: probability of generating rh from a model induced by d

Simple mixture

$$p(q|d, rh) = (1 - \alpha) \cdot p(q|d) + \alpha \cdot p(q|rh)$$

$p(q|rh)$: probability of generating q from a model induced by rh

How do we operationalise this model?

Keep it simple

$$p(q|rh): \sum_{\text{query terms}} \frac{\text{f(query term in rhet. relation)}}{\text{rhet. relation length}}$$

$$p(rh|d): \sum_{\text{rhet. relation terms}} \frac{\text{f(rhet. relation term in document)}}{\text{document length}}$$

f: frequency
Add-one smoothing

Overview

Aim: evaluate ranking model that uses rhetorical relations

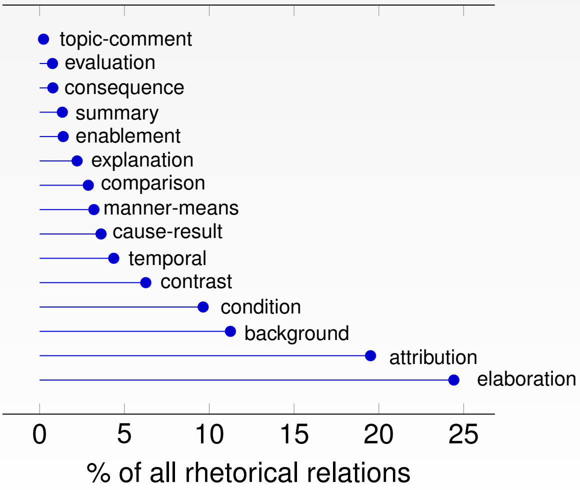
- Task: re-ranking top search results
- Collection: Clueweb09 cat. B
- Queries: TREC Web track 2009 (queries 1-50) and 2010 (queries 51-100)
- Initial ranking: INDRI, query likelihood with Dirichlet smoothing (tuned μ, 5-fold validation)

Overview

Aim: evaluate ranking model that uses rhetorical relations

- Task: re-ranking top search results
- Collection: Clueweb09 cat. B
- Queries: TREC Web track 2009 (queries 1-50) and 2010 (queries 51-100)
- Spam removal: Cormack et al. 2010, default settings
- Rhetorical relations detection: SPADE (Soricut & Marcu 2003), 15 types
- Initial ranking: INDRI, query likelihood with Dirichlet smoothing (tuned μ, 5-fold validation) [baseline]
- Re-ranking: our model

Distribution



Retrieval performance per rhetorical relation

rhetorical relation	Web 2009 (queries 1-50)					
	MAP		BPREF		NDCG	
none (baseline)	0.1625		0.3230		0.3893	
attribution	**0.1654***	+1.8%	**0.3275****	+1.4%	**0.3927****	+0.9%
background	**0.1646**	+1.3%	**0.3291****	+1.9%	**0.3910**	+0.4%
cause-result	**0.1626**	+0.1%	**0.3255****	+0.8%	**0.3900**	+0.2%
comparison	0.1610	-0.9%	**0.3251***	+0.6%	0.3877	-0.4%
condition	**0.1632**	+0.5%	**0.3258****	+0.9%	**0.3903**	+0.3%
consequence	0.1602	-1.4%	**0.3250**	+0.6%	0.3874	-0.5%
contrast	0.1549*	-4.6%	**0.3269****	+1.2%	**0.3897**	+0.1%
elaboration	0.1556*	-4.2%	**0.3292****	+1.9%	0.3866	-0.7%
enablement	0.1601	-1.4%	**0.3240**	+0.3%	0.3869*	-0.6%
evaluation	**0.1632**	+0.5%	**0.3242**	+0.4%	0.3886	-0.2%
explanation	0.1546	-4.9%	**0.3259***	+0.9%	0.3813	-2.1%
manner-means	0.1623	-0.1%	**0.3253***	+0.7%	0.3884	-0.2%
summary	**0.1626**	+0.1%	**0.3241**	+0.3%	0.3879	-0.4%
temporal	0.1615	-0.6%	**0.3262****	+1.0%	0.3887	-0.2%
topic-comment	**0.1673**	+3.0%	**0.3375**	+4.5%	**0.3976***	+2.1%

Best rhetorical relation per document

Treat as learning problem:

- Split dataset into 2 randomised samplings (50% - 50%)
- Use observations from one to make inferences about the other (Bayesian posterior inference)
- Repeat 5 times

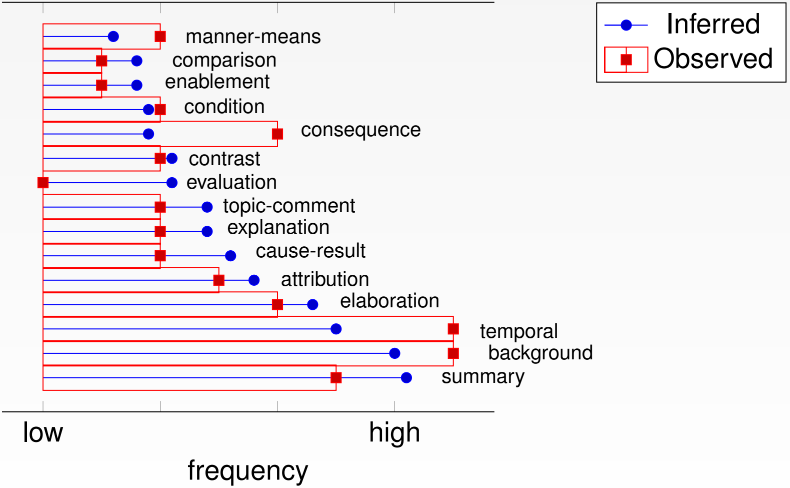
Best rhetorical relation per document

Treat as learning problem:

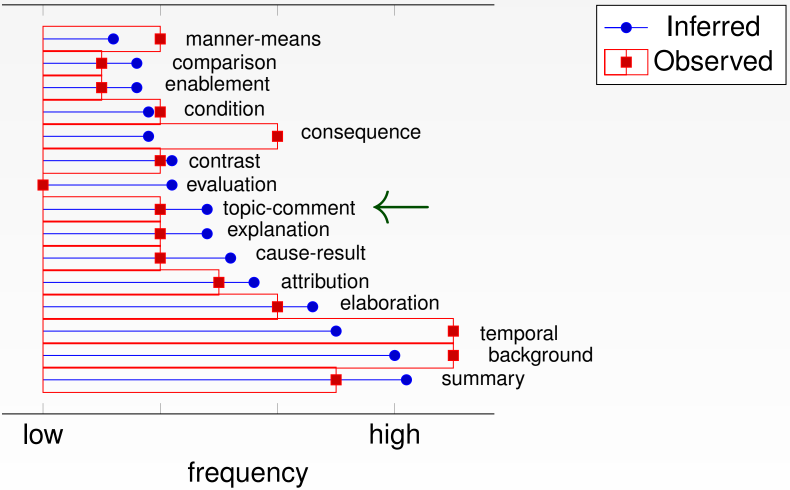
- Split dataset into 2 randomised samplings (50% - 50%)
- Use observations from one to make inferences about the other (Bayesian posterior inference)
- Repeat 5 times

rhetorical relation	Web 2009 (queries 1-50)					
	MAP		BPREF		NDCG	
none (baseline)	0.1625		0.3230		0.3894	
$\text{optimal}_{\textit{inferred}}$ (1)	**0.1879****	+15.6%	**0.3503****	+8.5%	**0.4224****	+8.5%
$\text{optimal}_{\textit{inferred}}$ (2)	**0.1948****	+19.9%	**0.3585****	+11.0%	**0.4202****	+7.9%
$\text{optimal}_{\textit{inferred}}$ (3)	**0.1984****	+22.1%	**0.3532****	+9.3%	**0.4169****	+7.1%
$\text{optimal}_{\textit{inferred}}$ (4)	**0.1952****	+20.1%	**0.3479****	+7.7%	**0.4282****	+10.0%
$\text{optimal}_{\textit{inferred}}$ (5)	**0.1950****	+20.0%	**0.3528****	+9.2%	**0.4287****	+10.1%
$\text{optimal}_{\textit{observed}}$	**0.2157**	+32.7%	**0.3660**	+13.3%	**0.4412**	+13.3%

Best rhetorical relation per document



Best rhetorical relation per document



Findings

- Different documents have different discourse structure - no globally good rhetorical relations
- Good IR potential for rhetorical relations, despite
  - out-of-the-box discourse parsing
  - simple integration into ranking

Findings

- Different documents have different discourse structure - no globally good rhetorical relations
- Good IR potential for rhetorical relations, despite
  - out-of-the-box discourse parsing
  - simple integration into ranking

What next?

- Discourse parsing: under the hood
- Nucleus vs. satellite rhetorical relations
- Nested rhetorical relations
- Faster discourse parsing (now 19 secs per document)

Findings

- Different documents have different discourse structure - no globally good rhetorical relations
- Good IR potential for rhetorical relations, despite
  - out-of-the-box discourse parsing
  - simple integration into ranking

What next?

- Discourse parsing: under the hood
- Nucleus vs. satellite rhetorical relations
- Nested rhetorical relations
- Faster discourse parsing (now 19 secs per document)

Thank you

Appendix A: Examples of rhetorical relations found in Clueweb

Rhetorical relation	Example sentences with rhetorical relations italicised and bold
attribution	... the islands now known as the Gilbert Islands were settled ***by Austronesian-speaking people*** ...
background	... many whites had left the country ***when Kenyatta divided their land among blacks*** ...
cause-result	... ***I plugged "wives" into the search box and came up with the following results*** ...
comparison	... so for humans, ***it is stronger than coloured*** to frustrate these unexpected numbers ...
condition	... Conditional money ***based upon care for the pet*** ...
consequence	... voltage drop with the cruise control switch ***could cause erratic cruise control operation*** ...
contrast	... ***Although it started out as a research project***, the ARPANET quickly developed into ...
elaboration	... order accutane ***no prescription required*** ...
enablement	... The project will also ***offer exercise programs and make eye care services accessible*** ...
evaluation	... such advances will be reflected in an ever-***greater proportion of grade A recommendations*** ...
explanation	... the concept ***called as "evolutionary developmental biology" or shortly "evo-devo"*** ...
manner-means	... Fill current path ***using even-odd rule, then paint the path*** ...
summary	... Safety Last, Girl Shy, Hot Water, The Kid Brother, Speedy ***(all with lively orchestral scores)*** ...
temporal	... Take time out ***before you start writing*** ...
topic-comment	... ***Director Mark Smith expressed support for greyhound adoption*** ...

Appendix B: Discourse parsing cost

SPADE processing speed: approximately 19 seconds per document (including the initial grammatical parsing), on a machine of 9 GB RAM, 8 core processor at 2.27GHz.