

Automatic Summarization Project

- Deliverable 3 -

Anca Burducea
Joe Mulvey
Nate Perkins

May 19, 2015

Outline

Deliverable 2 Summary

System overview

Sentence scoring

- Topic orientation

- Other methods

- Score combination

Topic clustering

Information ordering

Results and conclusions

Deliverable 2 Summary

- ▶ MEAD style approach
- ▶ TF-IDF sentence scoring + redundancy reduction
- ▶ ROUGE scores

	R	P	F
ROUGE-1	0.25909	0.30675	0.27987
ROUGE-2	0.06453	0.07577	0.06942
ROUGE-3	0.01881	0.02138	0.01992
ROUGE-4	0.00724	0.00774	0.00745

Outline

Deliverable 2 Summary

System overview

Sentence scoring

- Topic orientation

- Other methods

- Score combination

Topic clustering

Information ordering

Results and conclusions

D2 system

- ▶ score all sentences – CS
- ▶ choose highest scored sentences – CS
- ▶ order sentences – IO

D3 system

- ▶ score all sentences – CS
- ▶ **cluster sentences by their similarity – CS**
- ▶ **choose highest scored sentences from each cluster – CS**
- ▶ order sentences using block ordering – IO

New features

- ▶ experimented with different methods for sentence scoring
- ▶ added option for combining scores
- ▶ added topic clustering
- ▶ added information ordering

Outline

Deliverable 2 Summary

System overview

Sentence scoring

- Topic orientation

- Other methods

- Score combination

Topic clustering

Information ordering

Results and conclusions

Sentence scoring - Topic orientation

- ▶ TAC topic as query (e.g. "Columbine Massacre")
- ▶ use TF*IDF-like measure over sentences and query

$$\text{idf}(w) = \log\left(\frac{N+1}{0.5+\text{sf}(w)}\right)$$

$$\text{rel}(s|q) = \sum_{w \in q} \log(\text{tf}_{w,s} + 1) * \log(\text{tf}_{w,q} + 1) * \text{idf}_w$$

Sentence scoring - Topic orientation

ROUGE scores:

	R	P	F
ROUGE-1	0.20103	0.21993	0.20954
ROUGE-2	0.04781	0.05200	0.04968
ROUGE-3	0.01533	0.01669	0.01593
ROUGE-4	0.00689	0.00751	0.00716

Sentence scoring - Other methods

We tried other sentence scoring methods:

- ▶ LLR
- ▶ sentence position
- ▶ document headline similarity
- ▶ number of NERs

Sentence scoring - Other methods

We tried other sentence scoring methods:

- ▶ LLR
- ▶ sentence position
- ▶ document headline similarity
- ▶ number of NERs

... but all had low(er) scores (than our D2 results) by themselves.

Sentence scoring - Score combination

- ▶ scale all scores to $[0,1]$ range
- ▶ linearly combine different scoring methods using weights

Sentence scoring - Score combination

- ▶ scale all scores to $[0,1]$ range
- ▶ linearly combine different scoring methods using weights
e.g. $0.5 * \text{TF*IDF-score} + 0.5 * \text{headline-similarity-score}$

Outline

Deliverable 2 Summary

System overview

Sentence scoring

- Topic orientation

- Other methods

- Score combination

Topic clustering

Information ordering

Results and conclusions

Topic clustering

- ▶ cluster sentences into at most 5 clusters using cosine similarity
- ▶ remove sentences that are too similar (>0.5) within each cluster
- ▶ select highest ranked sentences accross all topic clusters

Outline

Deliverable 2 Summary

System overview

Sentence scoring

- Topic orientation

- Other methods

- Score combination

Topic clustering

Information ordering

Results and conclusions

Information ordering

- ▶ similar to Barzilay et al, 2002
- ▶ sentences A and B belong to the same topic block if $\text{sim}(A,B) > 0.6$
- ▶ for all sentence pairs (A_i, B_j) , with A_i from cluster(A) and B_j from cluster(B):

$$\text{sim}(A,B) = \frac{\#AB+}{\#AB}$$

$\#AB - \#(A_i, B_j)$ coming from same document

$\#AB+ - \#(A_i, B_j)$ coming from same document & same topic

Information ordering

- ▶ similar to Barzilay et al, 2002
- ▶ sentences A and B belong to the same topic block if $\text{sim}(A,B) > 0.6$
- ▶ for all sentence pairs (A_i, B_j) , with A_i from cluster(A) and B_j from cluster(B):

$$\text{sim}(A,B) = \frac{\#AB+}{\#AB}$$

$\#AB - \#(A_i, B_j)$ coming from same document

$\#AB+ - \#(A_i, B_j)$ coming from same document & same topic

- ▶ tweak: within the same topic segment = within a sentence window (of 5)

Outline

Deliverable 2 Summary

System overview

Sentence scoring

- Topic orientation

- Other methods

- Score combination

Topic clustering

Information ordering

Results and conclusions

Final system

- ▶ sentence scoring
 $0.7 * \text{TF*IDF} + 0.3 * \text{sentence position}$
- ▶ topic clustering
- ▶ block ordering

Results

ROUGE scores:

	R	P	F
ROUGE-1	0.25467	0.28628	0.26853
ROUGE-2	0.06706	0.07494	0.07052
ROUGE-3	0.02043	0.02219	0.02119
ROUGE-4	0.00642	0.00673	0.00655

Results - Comparison

ROUGE R scores:

	LEAD	D2	D3
ROUGE-1	0.19143	0.25909	0.25467
ROUGE-2	0.04542	0.06453	0.06706
ROUGE-3	0.01196	0.01881	0.02043
ROUGE-4	0.00306	0.00724	0.00642

Summary example: D2

- Japan, where whale meat is part of the traditional cuisine, reluctantly accepted a 1986 moratorium on commercial whaling by the International Whaling Commission (IWC).
- "The humpback whale was almost hunted into extinction.
- We're very, very keen to see firstly, no reopening of commercial whaling, and very importantly, no scientific whaling in the future," he said.
- Opponents of the plan have claimed that Japan is seeking to double to 800 the number of minke whales it will slaughter each year, and to add 50 humpback whales and 50 fin whales.

Summary example: D3

- ▶ International Whaling Commission, or IWC, banned commercial whaling in 1986, but grants limited permits to countries such as Japan that maintain whaling programs for scientific purposes.
- ▶ Japan, where whale meat is part of culinary culture, reluctantly halted commercial whaling in line with a 1986 IWC moratorium, but the next year resumed catches under a loophole that allows "research whaling".
- ▶ An animal rights group on Friday lost a bid to sue a Japanese whaling company for allegedly killing hundreds of whales inside an Australian whale sanctuary.
- ▶ "Whaling is also part of the Japanese culture," he said.

Future improvements

- ▶ improve redundancy elimination inside topic clustering
- ▶ anaphora resolution
- ▶ remove temporal expressions