



Redundancy reduction in Multi-Document Summarization


Iris Hendrickx

Universidade de Lisboa, 11 December 2009





Overview

- Background: Daeso project
 - Multi-document summarizer
 - Evaluation corpus
 - First experiments
- 

Background: Daeso Project

Detecting and Exploiting Semantic Overlap

- Duration: 2006 - 2009
- Part of Dutch-Belgian STEVIN programme
- Project members:
 - Emiel Kraemer, Erwin Marsi, **Tilburg University**, The Netherlands
 - Walter Daelemans, Iris Hendrickx, **University of Antwerp**, Belgium
 - Jacub Zavrel, **Textkernel**, Dutch company (3 months)
 - Maarten de Rijke, **University of Amsterdam**, The Netherlands (3 months)

Semantic overlap

Susan Boyle, the most unlikely reality TV star ever, shocks Simon Cowell on 'Britain's Got Talent' *(Daily News)*

The unassuming 47 year old Susan Boyle wowed the judges with her performance in the auditions for the latest season of the British talent show. *(Broadway World)*



Daeso Aims

Development of:

- **Resources:** 1M monolingual parallel corpus (treebank)
- **Tools:** for annotating the corpus and for automatically detecting semantic overlap
- **Application** to primarily summarization, but also to QA and IE

DAESO corpus

- **Corpus:** monolingual comparable treebank of 1 million words of Dutch text
 - Book translations
 - Autocue-subtitles
 - News articles
 - News headlines
- **Annotation:** semantic relations at tree-node level (500K)
- **Types of relations:** generalize, specify, intersect, restate, equals
- **Basis** for semantic overlap detection tools and text-2-text generation tools (sentence fuser)

Example tree alignment

Algraeph: example.pgc ()

Dagelijks koffie vermindert risico op Alzheimer en Dementie .

Drie koppen koffie per dag reduceert kans op Parkinson en Dementie .

Align Comment

Dagelijks koffie

Drie koppen koffie per dag

none equals
 restates generalizes
 specifies intersects

Mark Selected Alignments option is True 1 of 1 graph pairs

Antwerp Team:

Multi-document summarization: create one summary for a group of documents about the same topic.

Query-based summarization: user has a question and wants a summary that answers that question

- Development of Summarizer for Dutch
- Creation of a evaluation corpus of Automatic summarization

MEAD

Publicly available Toolkit for Automatic Summarization and Evaluation (Radev et al, 2004)

Basic method:

- Input: group of documents
- Compute for every sentence an importance weight
- Sort sentences on their weight
- Start summary with sentence with highest weight
- Take the next important sentence
 - measure similarity with sentences already in summary
 - **if little overlap, then add sentence to summary**
- Repeat until maximum summary size is reached.

Optimization

- **We adapted & optimized MEAD for Dutch**
- Two baseline systems (random / lead)
- Features: position, length, centroid, title-similarity
- Effect of several features individually
- Several combinations of features
- Query-based versus general summarization

Evaluation Corpus

Modeled like **DUC* 2006** data set

- 30 query-based document clusters,
- Documents from Dutch news sources
- Each cluster contains 5-25 relevant documents and a query description
- Each cluster has 5 X 2 summaries (250/100 words)
- and 5 extracts of 10 sentences.

*DUC: Document understanding conference

funded by Antwerp BOF 2008

First experiments

Starting point:

- Dutch multi-document summarizer based on MEAD toolkit

Research question:

- Can we improve the summarizer with a smarter method for redundancy detection?

Presented at ACL 2009 workshop on Language Generation and Summarisation

(SOD) Word Aligner

- Computes similarity between a pair of sentences
- Word alignment uses Cornetto* information (*synonyms, hyperonymes, least common subsumer*)
- Words are weighted with IDF
- Alignments as a weighted bipartite multigraph to combine the evidence

*(Dutch WordNet)

Example of word alignment

- **Three troopers** that got **heavily wounded** on friday at **an attack** in Afganistan returned to the Netherlands on monday.
- **Three of the five soldiers** that got **wounded** on friday at **a suicide attack** in Afganistan returned to the Netherlands on monday.

Redundancy detection

- **Cosine:** compute cosine similarity score between two sentences represented as tf*idf weighted word vectors
- **MMR:** maximum margin relevance (Carbonell and Goldstein, 1998)
- **Novelty:** boost sentences following important sentence
- **SOD Word Aligner**

Experimental setup

- **Data set:** split in 10/20 clusters for development/testing
- **First:** optimize automatic summarizer for Dutch news text: feature and parameter **optimization**
- **Next:** Optimized system is evaluated with different **rerankers**

Experimental setup

- Evaluation follows DUC 2006:

- **Automatic evaluation with ROUGE**

(scores system-generated summary by comparing it to 5 manually written summaries)

- **Manual evaluation** of content and linguistic quality
(average of 2 students)

Rouge scores on test set

	Rouge-2	Rouge-SU4
Rand baseline	0.101	0.153
Lead baseline	0.139	0.179
Optim-cosine	0.152	0.193
Optim-mmr	0.149	0.191
Optim-sod	0.150	0.193

Manual evaluation

Scores range from 5 (very good) to 1 (very poor)

	Gram	Redun	Refs	Focus	Struct	respon
Rand	4.08	3.9	2.58	2.6	2.	2.25
Lead	3.05	3.6	3.25	2.88	2.38	2.4
Cosine	3.9	3.18	2.65	3.15	2.43	2.75
MMR	3.98	3.13	2.55	3.13	2.38	2.7
SOD	4.05	3.13	2.85	3.23	2.5	2.7

Discussion

- MEAD works well for Dutch
- Our attempt did not improve upon the other rerankers


possible explanations:

- Often word overlap is enough to detect similarity
- Alignment tool makes errors
- Reranking is last step and cannot correct sentence ranking in general



Future plans

Multi-document summarization

- Re-run with improved version of SOD reranker
 - > Using tree structure
 - Query-relevance estimated with SOD aligner
 - Sentence fusion for summarization
- 



The End

- Dutch summarization web demo at:
 - <http://www.clips.ua.ac.be/~iris/sumdemo.html>
- 