Polarity Filtering for Sentiment Summarization

Frank Schilder Jochen L. Leidner Jack G. Conrad Ravikumar Kondadadi
Thomson Reuters Corporation, Research & Development, 610 Opperman Drive, St. Paul, MN 55123, USA.
firstname.lastname@thomsonreuters.com

THOMSON REUTERS

Introduction

Problem. Multi-document summarization saves time when dealing with large document quantities, but state of the art only handles *fact* summarization. What about *sentiment* (attitudes held by somebody about something)?
Solution. Query-based *Sentiment summarization*, a form of (multi-document) summarization whereby the query specifies the polarity of interest that informs the summarizer.

Data

- AQUAINT2 and BLOG06 corpora
- 22 topics comprising questions and documents, output from TREC's 2008 QA track

Method

- **Summarize**: Modified *FastSum*, fast query-based multi-document summarizer based solely on word-frequency features of clusters, documents and topics ranked by regression SVM [2,3]
- **Tag**: Gazetteer-based polarity tagger using data from *General Inquirer* and other sources to mark up sentiment-bearing words
- **Classify**: Classification of sentences based on majority of polarities
- **Filter**: Elimination of candidate sentences with polarities that do not match desired polarity as expressed in question

System

- Based on Apache UIMA (Unstructured Information Management Architecture) framework, an open, industrial-strength platform for unstructured information analysis and search
- Jericho HTML Parser applied to extract boilerplate blog language (e.g. *Response by*)
- Main processing phases:
  1. question sentiment and target analyzer
  2. blog entry sentence analyzer
  3. FastSum-based answer integration module with sentiment filter

Results

- Overall, our best sentiment summarization run ranked fourth in Pyramid F-score (among the 36 evaluated automatic runs)

Run	Pos. Entries	Neg. Entries	Type	Snippet Use	Pyramid F-Score
TOC1 (full gazetteer)	6,569	8,180	automatic	No	0.176
TOC2 (pruned)	2,959	4,920	automatic	No	0.150

- Associated ROUGE-2 score very low
- Attempt to improve performance by manually cleaning/pruning polarity gazetteer (run TOC2) showed negative effect compared to original gazetteer (TOC1)
- System's ability to perform polarity tagging not assessed at TAC
  →performed component evaluation based on internal annotation effort (528 sentences from BLOG06 as test set)

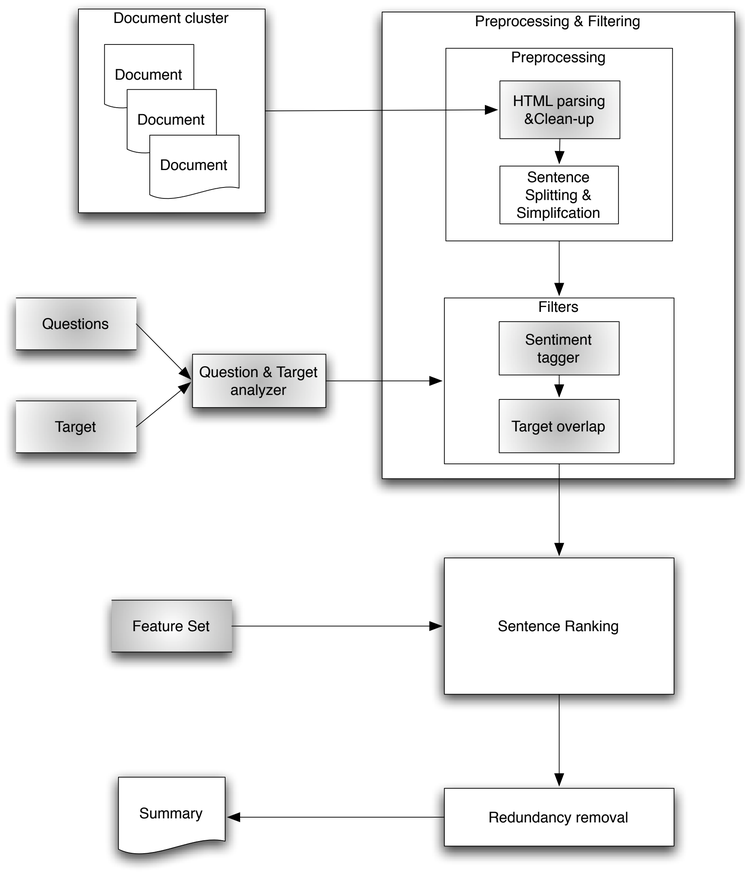
Summary, Conclusions and Future Work

- We present a first sentiment summarization system for news and blogs
- Problem modeled as gazetteer-based polarity classification of question and sentences, ranking SVM-based summarizer (FastSum) + sentiment filter
- Good pyramid F-score (4th rank in TAC), but mediocre ROUGE score
- Plan to evaluate more context-aware methods (using phrases and n-grams)
- Interested to evaluate discriminative power approach, but not possible with NIST Opinion Pilot Task 2008 (→mixed positive and negative questions in the same topic)

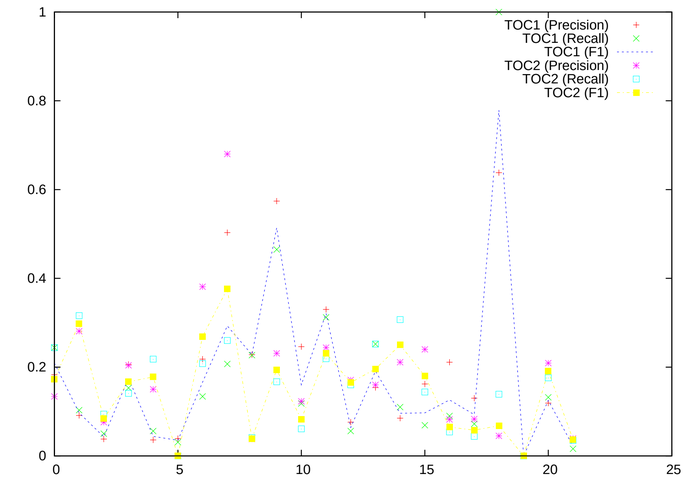
References

[1] L.-W. Ku, Y.-T. Liang, and H.-H. Chen. 2006. Opinion extraction, summarization and tracking in news and blog corpora. *Proc. Spring Sympos. Comp. Approaches to Analyzing Weblogs (AAAI-CAAW-06)*, Palo Alto, CA, USA.
[2] F. Schilder and R. Kondadadi. 2008. FastSum: Fast and Accurate Query-based Multi-Document Summarization. *Proc. ACL-HLT*, Columbus, OH, USA.
[3] F. Schilder, R. Kondadadi, Jochen L. Leidner, and Jack G. Conrad. 2008. **Thomson Reuters at TAC 2008: Aggressive Filtering with FastSum for Update and Opinion Summarization** *TAC 2008 Workshop Notes*, Gaithersburg, MD, USA.
[4] S. Somasundaran, T. Wilson, J. Wiebe, and V. Stoyanov. 2007. QA with attitude: Exploiting opinion type analysis for improving question answering in on-line discussions and the news. *Proc. Int. Conf. Weblogs and Social Media (ICWSM)*, Boulder, CO, USA.

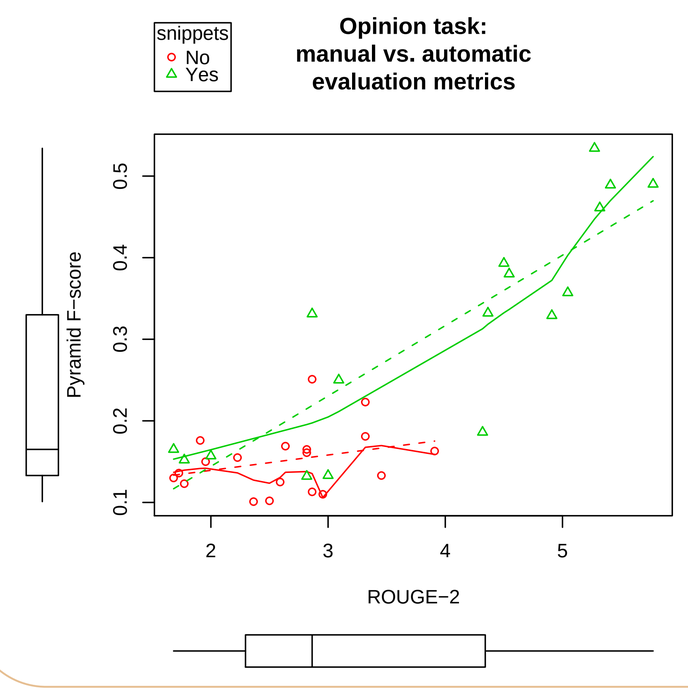
Architecture



Performance per Query



ROUGE-Pyramid Correlation



Component Evaluation

Polarity	Precision	Recall	F-score (F1)
Positive	46.97%	51.67%	49.21%
Negative	59.52%	*33.78%*	43.10%
Neutral	61.99%	73.10%	67.09%
Overall	58.10%	58.10%	**58.10%**

Acknowledgments

We would like to thank our colleagues Dan Dyke for annotating our development data, to Steve Rank and Kajsa Anderson for IT support, to Marc Light for sharing data sampling scripts, to James Allan for discussions, and to Khalid Al-Kofahi and Peter Jackson for supporting our TAC participation.