Getting Emotional About News Summarization

Alistair Kennedy Anna Kazantseva Diana Inkpen
Stan Szpakowicz

School of Electrical Engineering and Computer Science
University of Ottawa

Canadian AI, 2012

- Summarization of news has focussed on facts
  - Other domains, such as blogs have worked on sentiment/emotion more
- The emotion of a story is also important to its meaning
- Make summaries more emotional, could make summaries:
  - More interesting to read and so score higher in readability
  - Contain more relevant information – Pyramid Score
- Will it work? – Interesting negative result

- Text Analysis Conference (TAC)
- Query-driven multi-document summarization
- Guided Summarization – 5 categories of news
- Each containing its own topic statement and a list of aspects

- Accidents/Natural Disasters
- Attacks
- Health and Safety
- Endangered Resources
- Investigations and Trials

- *e.g. Plane Crash Indonesia*
- *e.g. Amish Shooting*
- *e.g. Internet Security*
- *e.g. Tuna Fishing*
- *e.g. Michael Vick Dog Fight*

- Update Summarization – two data sets A and B
  - Summarize A normally – Summarize B to only contain information not found in A
- Tuning Data – TAC 2010
  - Human written "model summaries" – 4 per topic
  - Source documents to be summarized – 10 per topic
- Testing Data – TAC 2011
  - Source documents to be summarized – 10 per topic

	Tuning 2010	Testing 2011
Accidents	7	9
Attacks	7	9
Health	12	10
Resources	10	8
Trial	10	8
Total	46	44

- Pyramid Evaluation
  - Human annotators find Summary Content Units (SCUs) in model summaries
  - Annotate automatically generated summaries with these SCUs
  - Rank based on SCU recall
  - We used a corpus of SCU annotated sentences to evaluate our sentence ranker
- Readability
  - Evaluates summaries for grammaticality, non-redundancy, referential clarity, focus, and structure/coherence
- ROUGE
  - Measures bigram overlap between model and automatic summaries
  - Two versions used ROUGE-2 and ROUGE-SU4
- Responsiveness
  - Overall summary quality

Emotional Corpus

- NRC Emotion Lexicon v0.5 [Mohammad and Turney(2012)]

- Emotion: 2283 words
  - Joy: 353
  - Sadness: 600
  - Fear: 749
  - Surprise: 275
  - Disgust: 540
  - Anger: 647
  - Trust: 641
  - Anticipation: 439
  - No emotion: 4808
- Sentiment: 2821 words
  - Positive: 1183
  - Negative: 1675
  - No sentiment: 4270

Measuring Relevant Emotions

- Are some emotions more common in summaries than source documents?
- Calculate *Emotional Density* (ED)

$$ED(E_i) = \frac{count(E_i)}{count(E_{1..N}) + count(\neg E)}$$

- ED can be calculated for each emotion E_i or no emotion $\neg E$
- ED can be calculated for model summaries and for source documents: $ED_M(E_i)$ and $ED_D(E_i)$
- For each news category calculate an emotional ratio: $\frac{ED_M(E_i)}{ED_D(E_i)}$

Discovering Significant Emotions: TAC 2010

	Emotional Ratio				
	Accidents	Attacks	Health	Resources	Trial
Joy	1.070	0.801	1.127	1.202	0.797
Sad	**1.349**	**1.220**	1.171	0.906	**1.561**
Fear	1.079	**1.242**	1.163	1.120	**1.157**
Surprise	1.036	0.996	0.973	**0.622**	**1.372**
Disgust	0.998	1.201	1.158	1.197	**1.453**
Anger	1.254	**0.593**	1.271	1.070	**1.458**
Trust	0.842	**0.593**	0.790	1.073	0.818
Anticipation	0.966	**0.590**	**0.726**	1.021	**0.841**
None	**0.917**	0.908	0.971	0.968	**0.686**
Positive	1.039	0.908	0.932	**1.305**	0.999
Negative	**1.195**	**1.323**	**1.271**	1.123	**1.522**
None	**0.924**	**0.885**	0.951	**0.901**	**0.807**

Discovering Significant Emotions: TAC 2010

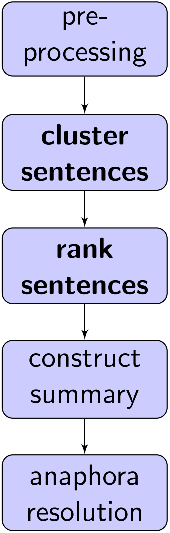
(Continued)

Maximize these emotions for each news category:

- Accidents: Sadness
- Attacks: Sadness, Fear & Anger
- Health: None, but strongly Negative
- Resources: None, but strongly Positive
- Trials: Sadness, Fear, Surprise, Disgust & Anger

Our System

Overview



- Two main components
  - Sentence Clustering: clusters related sentences
  - Sentence Ranker: ranks sentences based on their relatedness to the query
- Use Emotion to improve sentence ranking:
  - Baseline summarizer – no emotion
  - Emotionally Aware summarizer – use emotion words for query expansion

Our System

Clustering

- Objective: identify sub-topics in each collection of documents
- Representation: BOW vectors with stop-words removed, weighted by tf.idf
- Clustering algorithm: Affinity Propagation [Givoni and Frey(2009)]
  - Sentences are clustered into clusters of topics based on vocabulary
  - Each cluster has an exemplar - the most representative sentence
- Output: topical clusters.

Our System

Sentence Ranking

- *Roget's Thesaurus* based sentence ranking [Kennedy and Szpakowicz(2010)]
- For each word q in query Q, find the most related word w in a sentence S

$$score(S) = \sum_{q \in Q} max(SemDist(w, q) : w \in S)$$

- *SemDist* gives a relatedness score from 0..18
- Create summaries out of the top ranked sentences selecting at most one per cluster.

Our System

The Query for Baseline and Emotional Summaries

- What belongs in the query?
- Baseline Summarizer
  - use topic statement as query
- Emotionally Aware Summarizer
  - use topic statement as query
  - use emotional words – given much lower weight than topic words
    - only use exact matches
    - i.e. $SemDist(w, q) > 0 \iff w = q$
- These parameters were discovered using the TAC 2010 data
  - include topic statement, but leave aspects out

Intermediate Results Ranking Sentences

- Evaluate Sentence Ranking component on tuning (TAC 2010) data
- Macro-average precision (MAP)

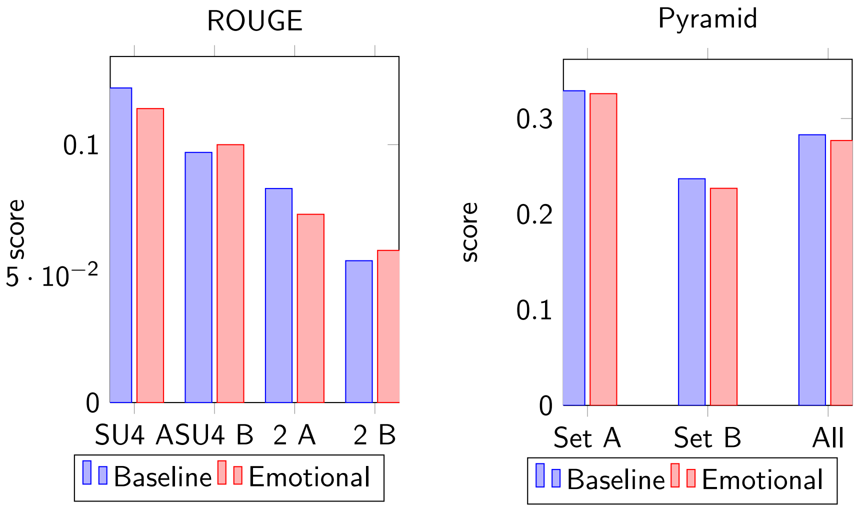
Category	Baseline	Emotion	*p*-value
Accidents	**0.603**	**0.637**	**0.008**
Attacks	0.519	0.552	0.087
Health	**0.422**	**0.476**	**0.014**
Resources	0.479	0.485	0.562
Trial	0.559	0.591	0.065
All	**0.506**	**0.539**	**0.000**

Evaluation on TAC 2011 data

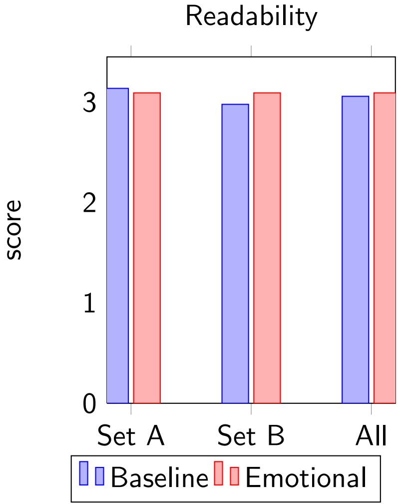
Emotional Ratio

	Emotional Ratio $\frac{\mathit{emotionCount}(\mathit{emotionalSummaries})}{\mathit{emotionCount}(\mathit{baselineSummaries})}$				
	Accidents	Attacks	Health	Resources	Trial
Joy	1.000	1.667	0.913	2.833	1.00
Sad	**3.847**	**1.900**	1.920	0.923	**2.296**
Fear	2.167	**2.182**	2.038	0.857	**1.596**
Surprise	2.364	1.125	1.000	1.400	**2.727**
Disgust	3.125	2.500	2.154	1.200	**2.368**
Anger	2.200	**1.921**	2.059	0.923	**1.837**
Trust	1.278	1.190	0.895	2.136	0.581
Anticipation	0.905	1.417	1.047	2.500	1.500
None	0.953	0.888	1.072	1.094	0.911
Positive	1.143	1.286	0.949	**2.310**	1.00
Negative	2.267	1.878	**2.244**	1.077	1.816
None	0.923	0.932	0.950	1.012	0.931

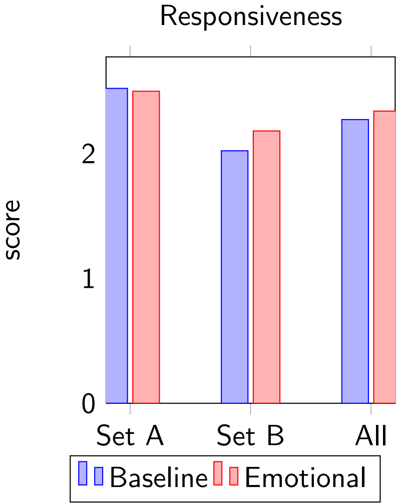
Results


ROUGE
score
0.1
$5 \cdot 10^{-2}$
0
SU4 A
SU4 B
2 A
2 B
Baseline
Emotional
Pyramid
score
0.3
0.2
0.1
0
Set A
Set B
All
Baseline
Emotional

Results



Readability
score
0
1
2
3
Set A
Set B
All
Baseline
Emotional



Responsiveness
score
0
1
2
Set A
Set B
All
Baseline
Emotional

Associated Emotions: 2010 vs 2011

Category	Emotions – 2010	Emotions – 2011
Accidents	**Sadness**	**None**
Attacks	**Sadness**, Fear & Anger	Fear & Anger
Health	None – strongly Negative	None – strongly Negative
Resources	None – strongly **Positive**	None – strongly **Negative**
Trials	Sadness, Fear, **Surprise**, **Disgust** & Anger	Sadness, Fear & Anger

Conclusion

- What worked
  - Created summaries with more emotional words
  - Some improvement for sentence ranking on the tuning data
  - Did not hurt TAC evaluation
- What did not work
  - No meaningful improvement on TAC evaluation
  - Some emotions from tuning data were not correct for the testing data
- Are ROUGE, Pyramids, etc really the right evaluation for such work?
  - Evaluate for emotional content instead?
  - Is this the right way to be using emotion?
- Future directions for research
  - Summarizing reviews, short stories, etc. instead of news
  - Make non-emotional summaries – would still need emotional awareness

Bibliography

Inmar E. Givoni and Brendan J. Frey.
A Binary Variable Model for Affinity Propagation.
Neural Computation, 21:1589–1600, 2009.

Alistair Kennedy and Stan Szpakowicz.
Evaluation of a Sentence Ranker for Text Summarization Based on Roget's Thesaurus.
In *Text, Speech and Dialogue (TSD), 13th International Conference*, pages 101–108, Brno, Czech Republic, 2010.

Saif M Mohammad and Peter D Turney.
Crowdsourcing a Word-Emotion Association Lexicon.
To Appear in Computational Intelligence, 2012.