Multi-Document Summarisation Using Generic Relation Extraction

Ben Hachey

Capital Markets CRC /
Macquarie University

5 September 2008

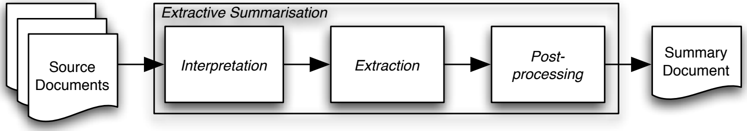
Outline

1. Background
   - Sentence extraction as set cover
   - Using generic IE to represent conceptual content
2. Experiments
   - Experimental Setup
   - Generic IE representations compared
3. Discussion
   - Complementarity of representations
   - Conclusions and future work

Outline

1. Background
   - Sentence extraction as set cover
   - Using generic IE to represent conceptual content
2. Experiments
   - Experimental Setup
   - Generic IE representations compared
3. Discussion
   - Complementarity of representations
   - Conclusions and future work

Overview of Extractive Summarisation



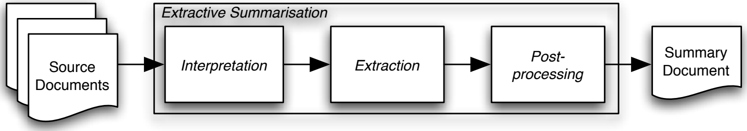
Description of sub-tasks

Input: Collection of NL documents on a given topic

1. Interpretation: Map to semantic representations
2. Extraction: Choose important, unique sentences
3. Post-processing: Maximise coherence of summary (e.g., entity re-writing, sentence ordering)

Output: Concise overview of source document content

Overview of Extractive Summarisation



Description of sub-tasks

Input: Collection of NL documents on a given topic

1. **Interpretation: Map to semantic representations**
2. Extraction: Choose important, unique sentences
3. Post-processing: Maximise coherence of summary (e.g., entity re-writing, sentence ordering)

Output: Concise overview of source document content

Extraction as Set Cover (Filatova & Hatz., 2004)

Extraction paradigm

- Filatova describes general extraction model
- Based on textual-conceptual mapping and set cover approx. algorithm

	c_1	c_2	c_3	c_4	c_5
t_1	1	1	0	1	1
t_2	1	0	0	1	0
t_3	0	1	0	0	1
t_4	1	0	1	1	1

Text-concept matrix

Extraction as set cover

- *Summary should select textual units such that there is maximal coverage of the conceptual units*
- Reducible to set cover problem for which there are polynomially-bounded approximation algorithms

Extraction as Set Cover (Filatova & Hatz., 2004)

Extraction paradigm

- Filatova describes general extraction model
- Based on textual-conceptual mapping and set cover approx. algorithm

	c_1	c_2	c_3	c_4	c_5
t_1	1	1	0	1	1
t_2	1	0	0	1	0
t_3	0	1	0	0	1
t_4	1	0	1	1	1

Text-concept matrix

Extraction as set cover

- *Summary should select textual units such that there is maximal coverage of the conceptual units*
- Reducible to set cover problem for which there are polynomially-bounded approximation algorithms

Problem

How represent conceptual content of sentences?

Outline

1. Background
   - Sentence extraction as set cover
   - Using generic IE to represent conceptual content
2. Experiments
   - Experimental Setup
   - Generic IE representations compared
3. Discussion
   - Complementarity of representations
   - Conclusions and future work

Words, Filatova Events and Generic Relations

Baseline word representation (TF)

Conceptual units are words

Filatova event representation (EV)

Conceptual units are $<Ent_i, Connector_j, Ent_k>$ triples where:

- $<Ent, Ent>$: all pairs in same sentence with connector
- *Connector*: all intervening verbs or action nouns

Words, Filatova Events and Generic Relations

Baseline word representation (TF)

Conceptual units are words

Filatova event representation (EV)

Conceptual units are $<Ent_i, Connector_j, Ent_k>$ triples where:

- $<Ent, Ent>$: all pairs in same sentence with connector
- *Connector*: all intervening verbs or action nouns

Generic relation representation (RL)

Conceptual units are $<Ent_i, Connector_j, Ent_k>$ triples where:

- $<Ent, Ent>$: $|\text{InterveningWords}| \leq 2$ **or** $|\text{DependencyRels}| \leq 1$
- *Connector*: topics from LDA (word and dependency path feats)

An example sentence and its representations

Example Sentence

Bush$_{\text{PER}}$ worked for Amoco$_{\text{ORG}}$ in Denver$_{\text{LOC}}$ and later started JNB$_{\text{ORG}}$.

Filatova event representation (*EV*)

<P_bush,worked,O_amoco>,
<P_bush,worked,L_denver>,
<O_amoco,started,O_jnb>,
<L_denver,started,O_jnb>
<P_bush,started,O_jnb>,

Baseline word representation (*TF*)

jnb, amoco, denver, bush, worked, started, later, for, in, and

An example sentence and its representations

Example Sentence

Bush$_{\text{PER}}$ worked for Amoco$_{\text{ORG}}$ in Denver$_{\text{LOC}}$ and later started JNB$_{\text{ORG}}$.

Filatova event representation (*EV*)

<P_bush,worked,O_amoco>,
<P_bush,worked,L_denver>,
<O_amoco,started,O_jnb>,
<L_denver,started,O_jnb>
<P_bush,started,O_jnb>,

Baseline word representation (*TF*)

jnb, amoco, denver, bush, worked, started, later, for, in, and

Shortcomings

- *EV*: Does not capture non-verbal relations
- *EV*: Very noisy
- *TF*, *EV*: Both ignore latent similarities
- *TF*: Very shallow

An example sentence and its representations

Example Sentence

Bush$_{\text{PER}}$ worked for Amoco$_{\text{ORG}}$ in Denver$_{\text{LOC}}$ and later started JNB$_{\text{ORG}}$.

Filatova event representation (*EV*)

<P_bush,worked,O_amoco>,
<P_bush,worked,L_denver>,
<O_amoco,started,O_jnb>,
<L_denver,started,O_jnb>
<P_bush,started,O_jnb>,

Baseline word representation (*TF*)

jnb, amoco, denver, bush, worked, started, later, for, in, and

Relation representation (*RL*)

<P_bush,rd302,O_amoco>,
<P_bush,rd188,O_amoco>,
.
.
<O_amoco,rd094,L_denver>,
<O_amoco,rd505,L_denver>,
.
.

Semantic Representations: Weighting

Baseline word representation (*TF*)

Weighted by $w = \sqrt{(1 + log\,(tf_{i,j})) * log\left(\frac{N}{df_i}\right)}$

Filatova event representation (*EV*)

Weighted by $w_{ev} = w_{ne} * w_{cn}$, where

- w_{ne} is the normalised entity pair count
- w_{cn} is the normalised connector count (in the context of the given pair)

Generic relation representation (*RL*)

Same as *EV*, but w_{cn} term is based on latent topics from LDA

Outline

Data

Data

- 30 multi-document summarisation tasks from DUC 2001 shared task, each of which:
  - is collected by a human and focused on particular topic
  - comprises approximately 10 news stories
  - has reference summaries of 50, 100, 200 and 400 words (Total of 3 human summaries for each length)
- Preprocessing
  - Sentence and token identification: LT TTT
  - Dependency parsing: Minipar
  - NER: C&C trained on MUC-7 data
  - Non-named entitiess: 10 most frequent nouns

Evaluation

Evaluation

- Evaluation: Two recall-oriented metrics
    - ***Rouge-1: Unigram overlap with reference***
    - Rouge-SU4: Skip bigram overlap with reference
- Significance testing
    - Paired Wilcoxon signed ranks
    - Across data set sub-tasks

Outline

1. Background
   - Sentence extraction as set cover
   - Using generic IE to represent conceptual content
2. Experiments
   - Experimental Setup
   - Generic IE representations compared
3. Discussion
   - Complementarity of representations
   - Conclusions and future work

Can Summarisation be Improved with GRE?

Results: Words (TF), Events (EV) and Relations (RL)

1	50	100	200	400
TF	0.0797	0.1113	0.1742	0.2467
EV	**0.1360**	**0.1776**	0.2315	**0.3019**
RL	**0.1360**	0.1766	**0.2412**	0.3014

Q: Can extractive summarisation be improved using GRE?

- Yes, *RL* significantly better than *TF* ($p \leq 0.001$).
- Rouge scores for *RL* and *EV* statistically indistinguishable.

How do Entity Pair Representations Compare?

Results: IE Representations without Connectors

1	50	100	200	400
ER	**0.1497**	**0.1929**	**0.2527**	**0.3123**
EE	0.1442	<u>0.1705</u>	<u>0.2288</u>	0.3061

Q: How do entity pair representations compare to each other?

- *ER* is better than *EE* ($p \leq 0.05$) for lengths 100 and 200.
- Same relative results for Rouge-SU4.

How do Entity Pair Representations Compare?

Results: IE Representations without Connectors

1	50	100	200	400
ER	**0.1497**	**0.1929**	**0.2527**	**0.3123**
EE	0.1442	0.1705	0.2288	0.3061

Q: How compare to respective event and relation representations?

- *EV* and *RL* indistinguishable from *EE* and *ER*.
- Mixed result for *EV* and *RL*, however..

Outline

1. Background
   - Sentence extraction as set cover
   - Using generic IE to represent conceptual content
2. Experiments
   - Experimental Setup
   - Generic IE representations compared
3. Discussion
   - Complementarity of representations
   - Conclusions and future work

Error Analysis: Rodney King Document Set

Example Human Summary for Rodney King Document Set

The most important of the many cases of police brutality reported in southern California 1989-1992, was the beating of Rodney King by four Los Angeles officers on March 3, 1991. An investigating commission outlined steps for improvement of the police department and called for the resignation of Chief Gates. Gates did not resign until the following year after the acquittal of the four officers caused massive rioting. Other cases of police brutality arose in Minneapolis, Chicago and Kansas City. Operation Rescue claimed that its non-violent anti-abortion demonstrators were seriously injured by excessive police tactics in more than 50 cities.

RL, EV better on fact-oriented tasks

- Relations and events central to document set
- *TF*, *EV* and *RL* score 0.016, 0.060 and 0.094 respectively

Error Analysis: Tuberculosis Document Set

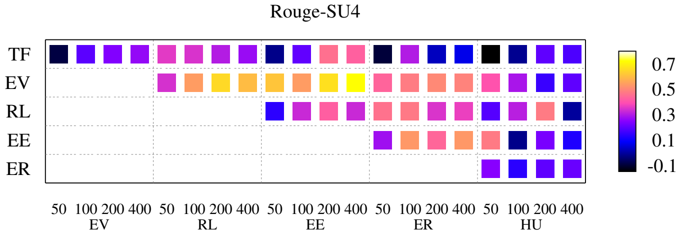
Example Human Summary for Tuberculosis Document Set

The occurrences of tuberculosis increased in the 1980s after a three decade decline. By 1990 it was the world's deadliest infectious disease, killing three million annually. The tuberculosis was fueled by AIDS patients who were vulnerable when their lowered immune system allowed the latent bacteria to develop into active tuberculosis. They then transmitted it to others. Tuberculosis ran rampant in sub-Saharan Africa, and increased in Latin America and Southeast Asia. In the United States the highest rates of infection were in the Northeast. Prisoners are highly susceptible to the disease. Airtight buildings with bad ventilation spawns tuberculosis.

TF better on description-oriented tasks

- Description and analysis central to document set
- *TF*, *EV* and *RL* score 0.046, 0.023 and 0.035 respectively

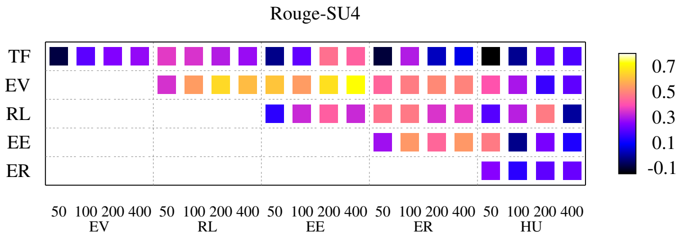
Complementarity Analysis



Spearman's ρ: Comparison to performance bounds

- IE reps have low correlation with annotator agreement
  Task difficulty is not an underlying cause
- IE reps have low correlation with *TF*
  TF *has high potential for combination with IE reps*.

Complementarity Analysis



Spearman's ρ: Comparison between IE representations

- *ER* and *RL* show potential for combination $[0.348, 0.476]$
  ER *is not a simpler version of* RL
- High correlation between *EV* and *EE* $[0.541, 0.725]$
  Low potential for combination of EV *and* EE

Outline

1. Background
   - Sentence extraction as set cover
   - Using generic IE to represent conceptual content
2. Experiments
   - Experimental Setup
   - Generic IE representations compared
3. Discussion
   - Complementarity of representations
   - Conclusions and future work

Conclusions and Future Work

Conclusions

- Generic relations are an effective representation for summarisation
  - significantly better than *tf*idf*
  - as good as generic events (comparable but less general)
- Representations are complementary

Future Work

- System combination (mean ranks, weighted mean ranks)
- Or, representations tailored to summary types

The End

Thank you

Extra Slides

...

Comparison to Supervised Extraction

Word and event features in supervised extraction

		TF	EV	RL
Unsupervised	(Current work)	0.174	0.232	0.241
Supervised	(Wong et al., 2008)	0.352	0.344	—

Events in supervised extraction

- Events do not improve supervised extraction
- Best overall score Rouge-1: 0.396

Event & Relation Weighting

Event (*EV*) and relation (*RL*) weighting combines scores for the entity pair and for the 'connector':

$$w_{ev} = w_{ne} * w_{cn}$$

Where entity pair $< i, k >$ counts are normalised by the total number of pair instances:

$$w_{ne} = \frac{Count(<i,k>)}{Count(<*,*>)}$$

Event Connector Weighting

Number of times connector *j* occurs in context of pair $< i, k >$, normalised by connector total:

$$w_{cn} = \frac{Count^{<i,k>}(j)}{Count^{<i,k>}(*)}$$

Relation Connector Weighting

Mean probability of latent topic *j* from LDA over the instances associated with pair $< i, k >$:

$$w_{cn} = \frac{\sum^{<i,k>} Pr(j)}{|<i,k>|}$$