Modelling Compression with Discourse Constraints

James Clarke and Mirella Lapata

School of Informatics
University of Edinburgh

EMNLP 2007, Prague

Outline

1. Sentence Compression
   - Definition and Overview
   - Compression beyond Sentences
2. Compression Model
   - ILP framework
   - Constraints
3. Experiments
   - Evaluation
   - Results

Outline

What is Sentence Compression?

The task

To produce a summary of a single sentence by:

- using less words than the original
- preserving the most important information
- remaining grammatical

What is Sentence Compression?

The task

To produce a summary of a single sentence by:

- using less words than the original
- preserving the most important information
- remaining grammatical

Simplification: Given an input sentence of words $W = w_1, w_2, \ldots, w_n$, a compression is formed by dropping any subset of these words (Knight and Marcu 2002).

Why Sentence Compression?

Applications

- concise summary generation (Jing 2000, Lin 2003)
- subtitle generation for TV programmes (Vandeghinste et al. 2004)
- document display on small screens (Corston-Oliver 2001)
- audio scanning devices for the blind (Grefenstette 1998)

Why Sentence Compression?

Applications

- concise summary generation (Jing 2000, Lin 2003)
- subtitle generation for TV programmes (Vandeghinste et al. 2004)
- document display on small screens (Corston-Oliver 2001)
- audio scanning devices for the blind (Grefenstette 1998)

Paradox: applications act on whole documents but compression by definition operates on isolated sentences.

Previous Work

Sentence-based models

Most use a parallel corpus with features defined over:

- words (Hori and Furui 2004)
- parse trees (Knight and Marcu 2000, Jing 2000, Riezler et al 2003, McDonald 2006, Galley and McKeown 2007)
- semantic concepts (Jing 2000)

Previous Work

Sentence-based models

Most use a parallel corpus with features defined over:

- words (Hori and Furui 2004)
- parse trees (Knight and Marcu 2000, Jing 2000, Riezler et al 2003, McDonald 2006, Galley and McKeown 2007)
- semantic concepts (Jing 2000)

Caveat: context influences what information is important; the resulting compressed document should be coherent.

This Work

We aim to:

- build a compression model that is contextually aware
- apply this model to entire documents

We need to:

- represent the flow of discourse in text
- process documents automatically and robustly

We focus on:

- representations of local coherence
- prerequisite for global coherence
- amenable to shallow processing

Outline

1. Sentence Compression
   - Definition and Overview
   - Compression beyond Sentences
2. Compression Model
   - ILP framework
   - Constraints
3. Experiments
   - Evaluation
   - Results

Discourse Representation

Centering Theory (Grosz et al. 1995)

- Entity-orientated theory of local coherence (Grosz et al. 1995)
- Entities in an utterance are ranked according to salience
- Each utterance has one center ($\approx$ topic or focus)
- Coherent discourses have utterances with common centers

Discourse Representation

Centering Theory (Grosz et al. 1995)

- Entity-orientated theory of local coherence (Grosz et al. 1995)
- Entities in an utterance are ranked according to salience
- Each utterance has one center ($\approx$ topic or focus)
- Coherent discourses have utterances with common centers

Lexical Chains (Halliday and Hasan 1976)

- Representation of lexical cohesion (Halliday and Hasan 1976)
- Degree of semantic relatedness among words in document
- Dense and long chains signal the main topic of the document
- Coherent texts have more related words than incoherent ones

Example Discourse

1 Bad weather dashed hopes of attempts to halt the flow during what was seen as a lull in the lava's momentum.

2 Some experts say that even if the eruption stopped today, the pressure of lava piled up behind for six miles would bring debris cascading down on to the town anyway.

3 Some estimate the volcano is pouring out one million tons of debris a day, at a rate of 15ft per second, from a fissure that opened in mid-December.

4 The Italian Army yesterday detonated 400lb of dynamite 3,500 feet up Mount Etna's slopes.

Centering Algorithm

1 Bad weather dashed hopes of attempts to halt the flow during what was seen as a lull in the lava's momentum.
2 Some experts say that even if the eruption stopped today, the pressure of lava piled up behind for six miles would bring debris cascading down on to the town anyway.

Centering Algorithm

1 Bad weather dashed hopes of attempts to halt the flow during what was seen as a lull in the lava's momentum.

2 Some **experts** say that even if the **eruption** stopped **today**, the **pressure** of **lava** piled up behind for six **miles** would bring **debris** cascading down on to the **town** anyway.

1. Extract entities from U_2.

Centering Algorithm

1 Bad weather dashed hopes of attempts to halt the flow during what was seen as a lull in the lava's momentum.

2 Some **experts** say that even if the **eruption** stopped **today**, the **pressure** of **lava** piled up behind for six **miles** would bring **debris** cascading down on to the **town** anyway.

1. Extract entities from U_2.
2. Rank the entities in U_2 according to their grammatical role. (subject $>$ objects $>$ others)

Centering Algorithm

1 Bad **weather** dashed **hopes** of **attempts** to halt the **flow** during what was seen as a **lull** in the **lava**'s **momentum**.
2 Some **experts** say that even if the **eruption** stopped **today**, the **pressure** of **lava** piled up behind for six **miles** would bring **debris** cascading down on to the **town** anyway.

1. Extract entities from U_2.
2. Rank the entities in U_2 according to their grammatical role. (subject $>$ objects $>$ others)
3. Find highest ranked entity in U_1 which occurs in U_2. Set entity to be center of U_2.

Centering Algorithm

1. Bad weather dashed hopes of attempts to halt the flow during what was seen as a lull in the **lava**'s momentum.
2. Some experts say that even if the eruption stopped today, the pressure of **lava** piled up behind for six miles would bring debris cascading down on to the town anyway.

1. Extract entities from U_2.
2. Rank the entities in U_2 according to their grammatical role. (subject $>$ objects $>$ others)
3. Find highest ranked entity in U_1 which occurs in U_2. Set entity to be center of U_2.

Annotated Discourse

1 Bad weather dashed hopes of attempts to halt the flow during what was seen as a lull in the **lava**'s momentum.

2 Some experts say that even if the eruption stopped today, the pressure of **lava** piled up behind for six miles would bring **debris** cascading down on to the town anyway.

3 Some estimate the volcano is pouring out one million tons of **debris** a day, at a rate of 15ft per second, from a fissure that opened in mid-December.

4 The Italian Army yesterday detonated 400lb of dynamite 3,500 feet up Mount Etna's slopes.

Lexical Chain Algorithm

1	–	–	–
2	–	–	–
3	–	–	–
4	–	–	–
5	–	–	–
6	–	–	–
7	–	–	–
8	–	–	–

Lexical Chain Algorithm

	Lava	Weight	Time
1	X	–	X
2	X	–	–
3	–	–	X
4	X	X	X
5	X	X	–
6	–	–	X
7	X	–	–
8	–	–	–

1. Compute chains for document (Galley and McKeown 2003).

Lexical Chain Algorithm

	Lava	Weight	Time
1	X	–	X
2	X	–	–
3	–	–	X
4	X	X	X
5	X	X	–
6	–	–	X
7	X	–	–
8	–	–	–

1. Compute chains for document (Galley and McKeown 2003).

- **Lava** : {lava, lava, lava, magma, lava}
- **Weight** : {tons, lbs}
- **Time** : {day, today, yesterday, second}

Lexical Chain Algorithm

	Lava	Weight	Time
1	X	–	X
2	X	–	–
3	–	–	X
4	X	X	X
5	X	X	–
6	–	–	X
7	X	–	–
8	–	–	–
Score	5	2	4

1. Compute chains for document (Galley and McKeown 2003).
2. $Score(Chain) = Sent(Chain)$

Lexical Chain Algorithm

	Lava	Weight	Time
1	X	–	X
2	X	–	–
3	–	–	X
4	X	X	X
5	X	X	–
6	–	–	X
7	X	–	–
8	–	–	–
Score	5	2	4

1. Compute chains for document (Galley and McKeown 2003).
2. $Score(Chain) = Sent(Chain)$
3. $Score(Chain) < Avg(Score)$.

Lexical Chain Algorithm

	Lava	Time
1	X	X
2	X	–
3	–	X
4	X	X
5	X	–
6	–	X
7	X	–
8	–	–
Score	5	4

1. Compute chains for document (Galley and McKeown 2003).
2. $Score(Chain) = Sent(Chain)$
3. $Score(Chain) < Avg(Score)$.
4. Mark terms in chains as topic.

- **Lava** : {lava, lava, lava, magma, lava}
- **Time** : {day, today, yesterday, second}

Annotated Discourse

1 Bad **weather** dashed hopes of attempts to halt the **flow** during what was seen as a lull in the **lava**'s momentum.

2 Some experts say that even if the eruption stopped **today**, the pressure of **lava** piled up behind for six **miles** would bring **debris** cascading down on to the **town** anyway.

3 Some estimate the volcano is pouring out one million tons of **debris** a **day**, at a **rate** of 15**ft** per **second**, from a fissure that opened in mid-December.

4 The Italian Army **yesterday** detonated 400lb of dynamite 3,500 feet up Mount Etna's slopes.

Outline

Integer Linear Programming

Properties:

- linear objective function
- decision variables (variables under our control)
- constraints over decision variables

Advantages:

- find the global minimum or maximum value of objective function (Germann et al 2001, McDonald 2007)
- incorporate global constraints over the output space (Roth and Yih 2004, Riedel and Clarke 2006)
- ensure compressions are structurally and semantically valid

Compression Model

Integer Linear Programming Formulation

- trigram language model and significance score:

$$\mathbf{c}^* = \operatorname*{argmax}_{c} \sum_{i=1}^{n} P(w_i | w_{i-1}, w_{i-2}) + \sum_{i=1}^{n} I(w_i)$$

- requires no parallel corpus
- compresses sentences sequentially

Compression Model

Integer Linear Programming Formulation

- trigram language model and significance score:

$$\mathbf{c}^* = \operatorname*{argmax}_{c} \sum_{i=1}^{n} P(w_i | w_{i-1}, w_{i-2}) + \sum_{i=1}^{n} I(w_i)$$

- requires no parallel corpus
- compresses sentences sequentially

Decision Variables

$$y_i = \begin{cases} 1 & \text{if } w_i \text{ is in the compression} \\ 0 & \text{otherwise} \end{cases} \quad (1 \leq i \leq n)$$

Outline

1. Sentence Compression
   - Definition and Overview
   - Compression beyond Sentences
2. Compression Model
   - ILP framework
   - Constraints
3. Experiments
   - Evaluation
   - Results

Modifier Constraints

Ensure the relationships between head words and their modifiers remain grammatical.

1. If a modifier is in the compression, its head word must be included:

$$y_{head} - y_{modifer} \geq 0$$

2. Do not drop *not* if the head word is in the compression (same for words like *his*, *our* and genitives).

$$y_{head} - y_{not} = 0$$

Sentential Constraints

Take overall sentence structure into account.

1. If a verb is in the compression then so are its arguments, and vice-versa:

$$y_{subject/object} - y_{verb} = 0$$

2. The compression must contain at least one verb.

$$\sum_{i \in verbs} y_i \geq 1$$

Discourse Constraints

Take overall document into account and preserve its coherence.

1. Do not drop centers and their references.

$$y_{center} = 1$$

2. Do not drop words in topical lexical chains.

$$y_{topical} = 1$$

3. Do not drop personal pronouns.

$$y_{personal\ pronoun} = 1$$

Compressed Document

1. Weather dashed hopes to halt the flow.
2. Experts say that, the pressure bring cascading down to the town.
3. Some estimate at a rate of 15ft from a fissure opened in mid-December.
4. The Italian Army detonated 400lb of dynamite.

Compressed Document

1. Weather dashed hopes to halt the flow in the lava's momentum.
2. Some experts say that, the pressure of lava would bring debris cascading down.
3. The volcano is pouring out million tons of debris a day.
4. The Italian Army yesterday detonated 400lb of dynamite.

Outline

Evaluation

Motivation

Assume the compressed document is a replacement for original:

1. is the compressed document readable?
2. Is the key information from original preserved in compression?

Question-answering paradigm

- How many questions can we answer accurately by reading the compressed document?
- Questions derived from source document.
- Two annotators created Q&A pairs.
- Fact-based questions requiring unambiguous answers.

Experimental Setup

- Created document-based compression corpus (available from `http://homepages.inf.ed.ac.uk/s0460084/data/`).
- Six documents with five to eight questions per document.
- Three conditions: gold standard, McDonald (2006), Discourse ILP.
- Sixty participants over the web.
- Rate readability on seven point scale.
- Answer questions one at a time using compressed document.

Experimental Setup

- Created document-based compression corpus (available from `http://homepages.inf.ed.ac.uk/s0460084/data/`).
- Six documents with five to eight questions per document.
- Three conditions: gold standard, McDonald (2006), Discourse ILP.
- Sixty participants over the web.
- Rate readability on seven point scale.
- Answer questions one at a time using compressed document.

Mcdonald (2006): discriminative, state-of-the-art model, with large sentence-based feature space.

Example Questions and Answers

1 Weather dashed hopes to halt the flow in the lava's momentum.
2 Some experts say that, the pressure of lava would bring debris cascading down.
3 The volcano is pouring out million tons of debris a day.
4 The Italian Army yesterday detonated 400lb of dynamite.

Q: What is posing a threat to the town?	**A:** lava
Q: What hindered attempts to stop the lava flow?	**A:** bad weather
Q: What did the Army do to stop the lava flow?	**A:** detonate explosives

Outline

Results

Model	CompR	Readability	Q&A
McDonald 2006	60.1%	2.65	54.4%
Discourse ILP	65.4%	3.00	67.8%
Gold Standard	70.3%	5.27	82.2%

Results

Model	CompR	Readability	Q&A
McDonald 2006	60.1%	2.65	54.4%
Discourse ILP	65.4%	3.00	67.8%
Gold Standard	70.3%	5.27	82.2%

- On readability Discourse ILP and McDonald are not sig. different

Results

Model	CompR	Readability	Q&A
McDonald 2006	60.1%	2.65	54.4%
Discourse ILP	65.4%	3.00	67.8%
Gold Standard	70.3%	5.27	82.2%

- On readability Discourse ILP and McDonald are not sig. different
- Both models are sig. worse than gold standard

Results

Model	CompR	Readability	Q&A
McDonald 2006	60.1%	2.65	54.4%
Discourse ILP	65.4%	3.00	67.8%
Gold Standard	70.3%	5.27	82.2%

- On readability Discourse ILP and McDonald are not sig. different
- Both models are sig. worse than gold standard
- On Q&A task Discourse ILP is sig. better than McDonald

Results

Model	CompR	Readability	Q&A
McDonald 2006	60.1%	2.65	54.4%
Discourse ILP	65.4%	3.00	67.8%
Gold Standard	70.3%	5.27	82.2%

- On readability Discourse ILP and McDonald are not sig. different
- Both models are sig. worse than gold standard
- On Q&A task Discourse ILP is sig. better than McDonald
- Both models sig. worse than gold standard

Conclusions

Contributions:

- discourse-based sentence compression model
- formulated within the ILP framework using global constraints
- unsupervised, relatively simple and intuitive model
- document-based evaluation using a Q&A task-based paradigm
- performance gains over supervised discourse agnostic system

Future work:

- interface compression model with sentence extraction
- study the effect of global discourse structure (Daumé III and Marcu 2002)
- explore the effect of discourse for other models

Q&A Task

- Each question presented in turn. No corrections allowed.
- Answers marked consistently across all three systems.

Q: What is posing a threat to the town?		
A: Lava	Volcano	Lava from Mount Etna
Q: What hindered attempts to stop the lava flow?		
A: Bad weather	Snow and winds	The weather - snow
Q: What did the Army do to stop the lava flow?		
A: Detonate explosives	Used explosives	Detonate dynamite