Neural Network-Based Abstract Generation for Opinions and Arguments

Lu Wang

Northeastern University
College of Computer and Information Science

Wang Ling

Google DeepMind

Opinions

- What do you think?

[source: www.cartoonbank.com]

Opinions

- Mundane tasks
  - Which movie to watch tonight?
  - Which hotel should I book?

- Fundamental societal issues
  - Abolish death penalty?
  - Gun laws?

- Public Deliberation

[source: www.cartoonbank.com]

The Need of Opinion Summarization

- Massive amount of opinionated text
  - Reviews
  - Comments
  - Blogs
  - Online debates

Opinion Mining and Summarization

- Opinion Mining and Retrieval
  - Document retrieval (TREC blog track, 2006)
  - Opinion target and phrase extraction (Choi, Breck, and Cardie, 2006; Yang and Cardie, 2013)

Opinion Mining and Summarization

- Opinion Mining and Retrieval
  - Document retrieval (TREC blog track, 2006)
  - Opinion target and phrase extraction (Choi, Breck, and Cardie, 2006; Yang and Cardie, 2013)
- Opinion Summarization
  - Product reviews (Hu and Liu, 2004; Lerman et al., 2009), Editorials (Paul et al., 2010), and community question answering (Wang et al., 2014)

Opinion Mining and Summarization

- Opinion Mining and Retrieval
  - Document retrieval (TREC blog track, 2006)
  - Opinion target and phrase extraction (Choi, Breck, and Cardie, 2006; Yang and Cardie, 2013)
- Opinion Summarization
  - Product reviews (Hu and Liu, 2004; Lerman et al., 2009), Editorials (Paul et al., 2010), and community question answering (Wang et al., 2014)
  - But they are all extractive summaries.

Our Goal

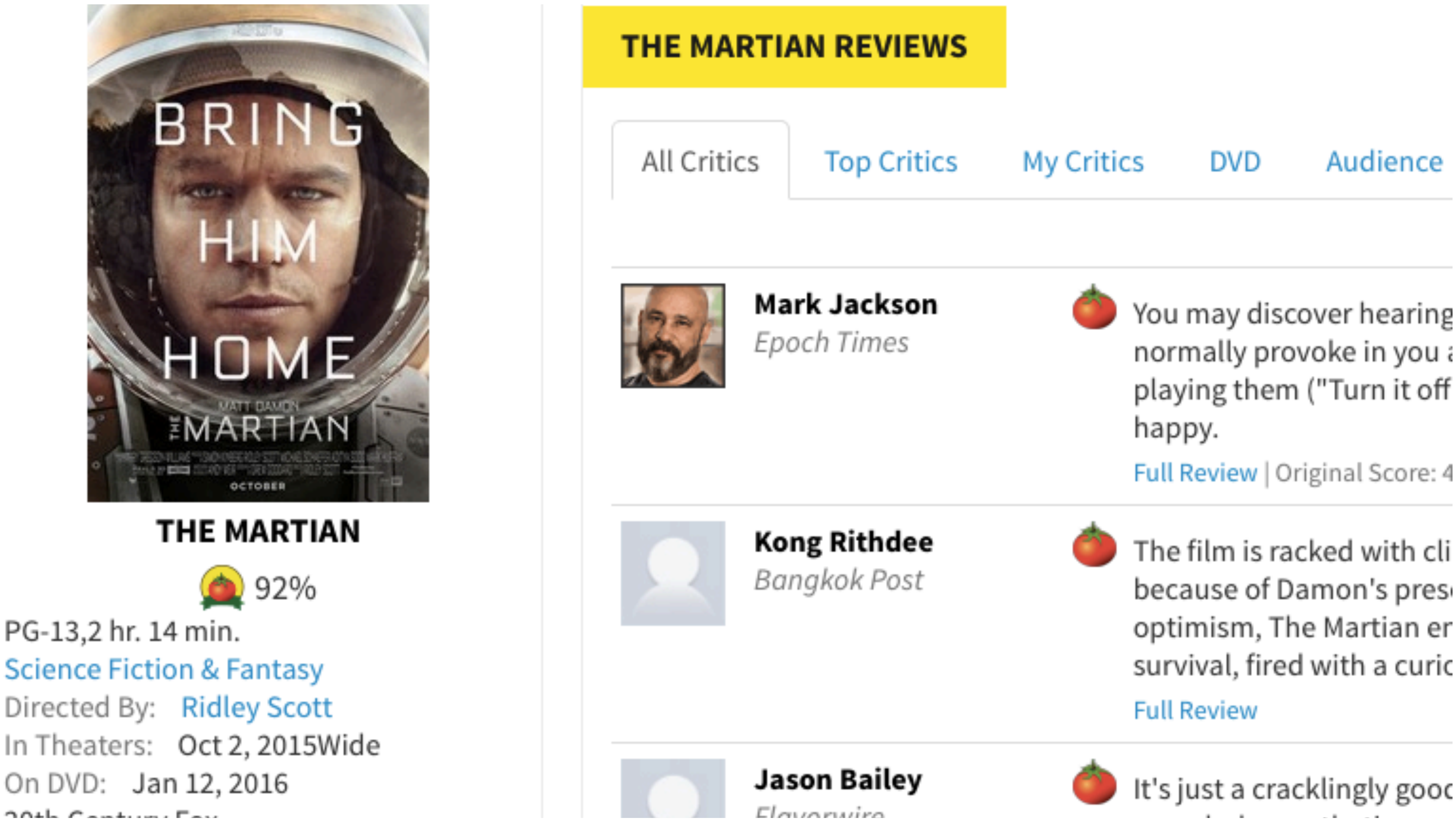
- **<u>Input:</u>** a set of text documents containing opinions about the same topic

- **<u>Output:</u>** one sentence abstractive summary that describes the opinion consensus of the input

Two Domains

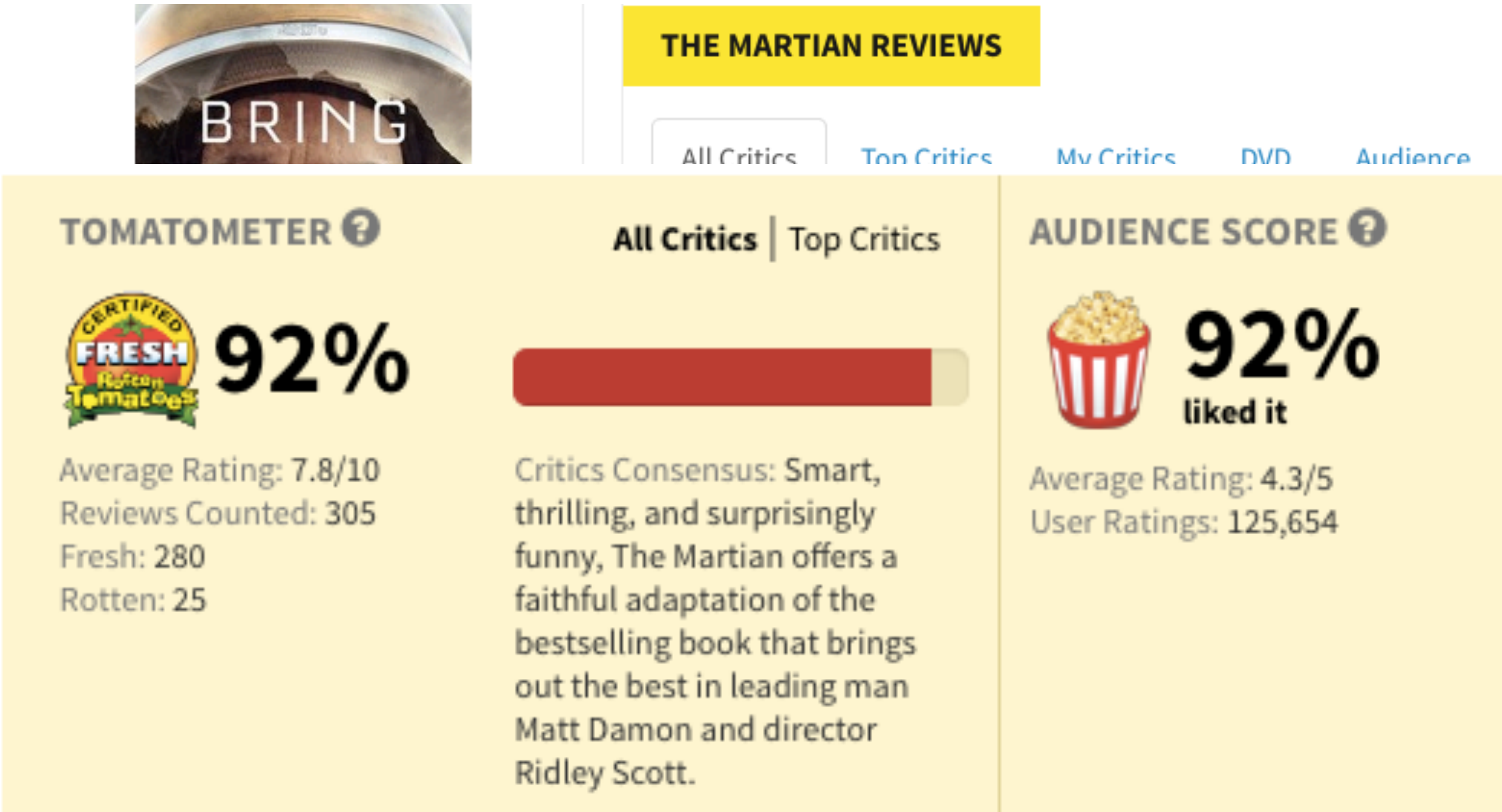
- Movie reviews
- Online arguments on controversial issues

Movie Reviews



[source: http://www.rottentomatoes.com]

Movie Reviews



Movie Reviews

- **<u>Reviews on "The Martian"</u>**
  - … an intimate sci-fi epic that is smart, spectacular and stirring.
  - The Martian is a thrilling, human and moving sci-fi picture that is easily the most emotionally engaging film Ridley Scott has made...
  - It's pretty sunny and often funny, a space oddity for a director not known for pictures with a sense of humor.
  - The Martian highlights the book's best qualities, tones down its worst, and adds its own style...
- **<u>Opinion Consensus:</u>** Smart, thrilling, and surprisingly funny, The Martian offers a faithful adaptation of the bestselling book that brings out the best in leading man Matt Damon and director Ridley Scott.

Online Arguments

THIS HOUSE SUPPORTS THE DEATH PENALTY

Capital punishment is the sentence of death, or practice of execution, handed down as punishment for a criminal offence. It can only be used by a state, after a proper legal trial. The United Nations in 2008 adopted a resolution (62/149) calling for a moratorium on the use of the death penalty, however fifty-eight countries, including the United States and China, still exercise the death penalty. As such, the topic remains highly controversial. Abolitionist groups and international organizations argue that it is cruel and inhumane, while

Read more

discuss this

POINTS FOR

POINTS AGAINST

It helps the victims' families achieve closure.

The death penalty deters crime.

Execution prevents the accused from committing further crimes.

[source: http://idebate.org]

Online Arguments

THIS HOUSE SUPPORTS THE DEATH PENALTY

Capital punishment is the sentence of death, or practice of execution, handed down as punishment for a criminal offence. It can only be used by a state, after a proper legal trial. The United Nations in 2008 adopted a resolution (62/149) calling for a moratorium on the use of the death penalty, however fifty-eight countries, including the United States and China, still exercise the death penalty. As such, the topic remains highly controversial. Abolitionist groups and international organizations argue that it is cruel and inhumane, while

Read more

discuss this

Claims

POINTS FOR

POINTS AGAINST

It helps the victims' families achieve closure.

The death penalty deters crime.

Execution prevents the accused from committing further crimes.

Online Arguments

Claim

Arguments (or premises)

The death penalty deters crime.

POINT

The state has a responsibility to protect the lives of innocent citizens, and enacting the death penalty may save lives by reducing the rate of violent crime.

The reasoning here is simple- fear of execution can play a powerful motivating role in convincing potential murderers not to carry out their acts. While the prospect of life in prison may be frightening, surely death is a more daunting prospect. Thus, the risk of execution can change the cost-benefit calculus in the mind of murderers-to be so that the act is no longer worthwhile for them[1].

Numerous studies support the deterrent effect

Online Arguments

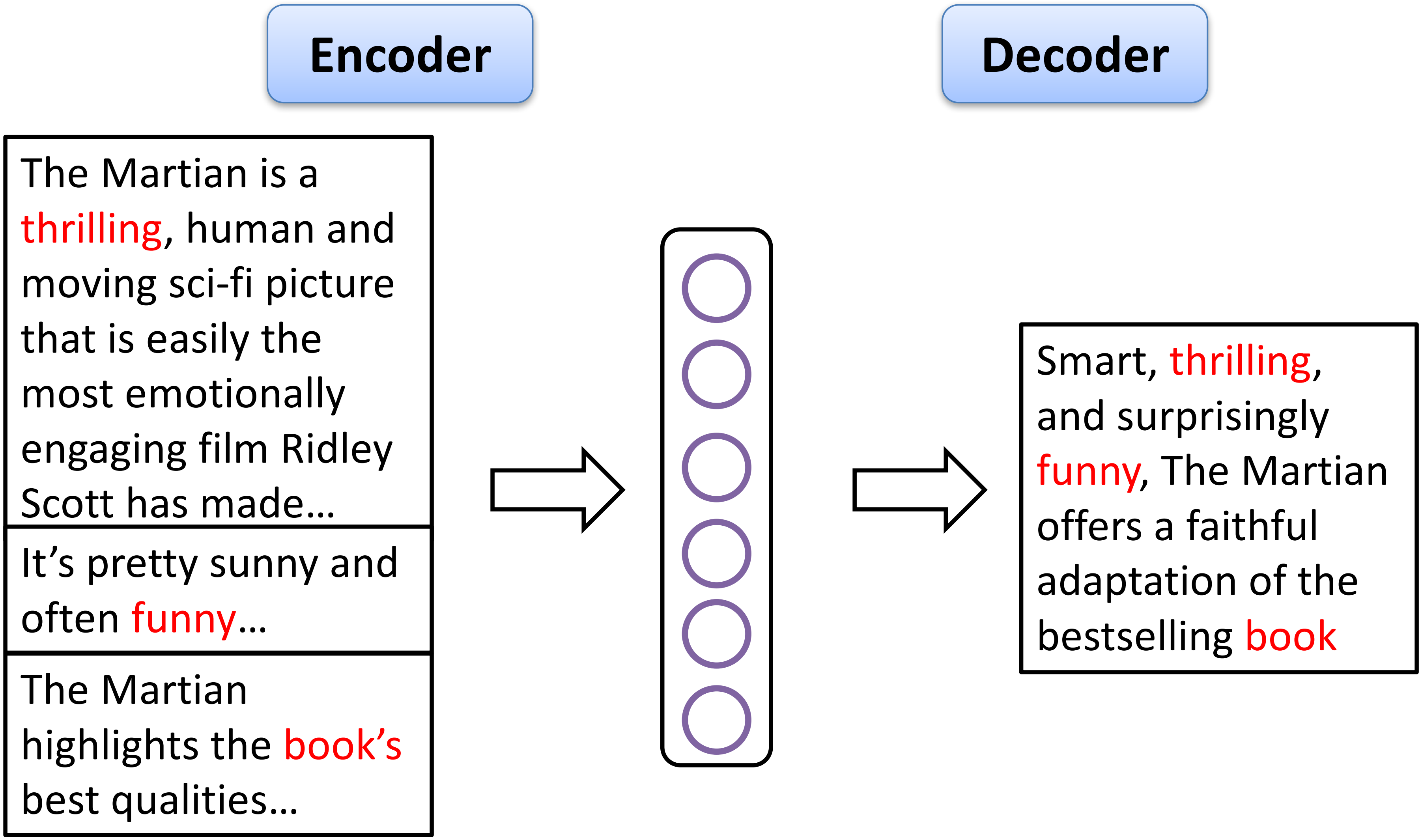
- **<u>Arguments on topic "death penalty":</u>**
  - The state has a responsibility to protect the lives of innocent citizens, and enacting the death penalty may save lives by reducing the rate of violent crime.
  - A 1985 study by Stephen K. Layson at the University of North Carolina showed that a single execution deters 18 murders.
  - Reducing the wait time on death row prior to execution can dramatically increase its deterrent effect in the United States.
- **<u>Claim (summary):</u>** The death penalty deters crime.

Datasets

- Movie reviews from Rotten Tomatoes
  - 3,731 movies with >246k critics

- Online arguments from idebate.org
  - 2,259 claims with >17k arguments (676 debate topics)

Encoder-Decoder Framework



Sequence-to-sequence Learning

- Neural machine translation
  - Source language to target language
  - Kalchbrenner and Blunsom (2013), Sutskever et al. (2014), Bahdanau et al. (2015)

Sequence-to-sequence Learning

- Neural machine translation
  - Source language to target language
  - Kalchbrenner and Blunsom (2013), Sutskever et al. (2014), Bahdanau et al. (2015)
- Sentence compression and summarization
  - News articles
  - Filippova et al. (2015)
  - Rush et al. (2015)

Decoder

- Input: a set of words x
- Output: a summary y

$$\log P(y|x) = \sum_{j=1,\ldots,|y|} \log P(y_j | y_1, \ldots, y_{j-1}, x)$$

Decoder

- Input: a set of words x
- Output: a summary y

$$\log P(y|x) = \sum_{j=1,\ldots,|y|} \log P(y_j|y_1, ..., y_{j-1}, x)$$

- The probability is estimated by

$$p(y_j|y_1, ..., y_{j-1}, x) = softmax(\mathbf{h}_j)$$

  – h_j is the Recurrent Neural Networks (RNNs) state variable at timestamp j

Decoder

- Specifically, g is the recurrent update function

$$\mathbf{h}_j = g(\mathbf{y}_{j-1}, \mathbf{h}_{j-1}, \mathbf{s})$$

Previously generated word

Previous state

Input text representation

- Implemented using Long Short-Term Memory (LSTM) network

Encoding the Input Text

- Challenge:
  - It’s unclear which part of the input should be used for summarization.

Encoding the Input Text

- Challenge:
  - It’s unclear which part of the input should be used for summarization.

- Attention model on input text
  - Learn to detect summary-worthy content

Encoding the Input Text

$$\mathbf{h}_j = g(\mathbf{y}_{j-1}, \mathbf{h}_{j-1}, \mathbf{s})$$

Previously generated word

Previous state

Input text representation

Attention Model

$$\mathbf{h}_j = g(\mathbf{y}_{j-1}, \mathbf{h}_{j-1}, \mathbf{s})$$

Previously generated word

Previous state

Input text representation

Represent S as weighted sum

$$\sum_{i=1,\ldots,|x|} a_i \mathbf{b}_i$$

context dependent representation of word x_i

Attention Model

$$\mathbf{h}_j = g(\mathbf{y}_{j-1}, \mathbf{h}_{j-1}, \mathbf{s})$$

Previously generated word

Previous state

Input text representation

Represent S as weighted sum

$$\sum_{i=1,\ldots,|x|} a_i \mathbf{b}_i$$

how likely the input word is to be used to generate the next word in summary (*soft alignment*)

context dependent representation of word x_i

Attention Model

$$\mathbf{h}_j = g(\mathbf{y}_{j-1}, \mathbf{h}_{j-1}, \mathbf{s})$$

$$\Downarrow$$

$$\sum_{i=1,\ldots,|x|} a_i \mathbf{b}_i$$

how likely the input word is to be used to generate the next word in summary

context dependent representation of word x_i

Attention Model

$$\mathbf{h}_j = g(\mathbf{y}_{j-1}, \mathbf{h}_{j-1}, \mathbf{s})$$

$$\sum_{i=1,\ldots,|x|} a_i \mathbf{b}_i$$

how likely the input word is to be used to generate the next word in summary

context dependent representation of word x_i

b_i: *Implemented by bidirectional LSTM*

Attention Model

$$\mathbf{h}_j = g(\mathbf{y}_{j-1}, \mathbf{h}_{j-1}, \mathbf{s})$$

$$\sum_{i=1,\ldots,|x|} a_i \mathbf{b}_i$$

how likely the input word is to be used to generate the next word in summary

context dependent representation of word x_i

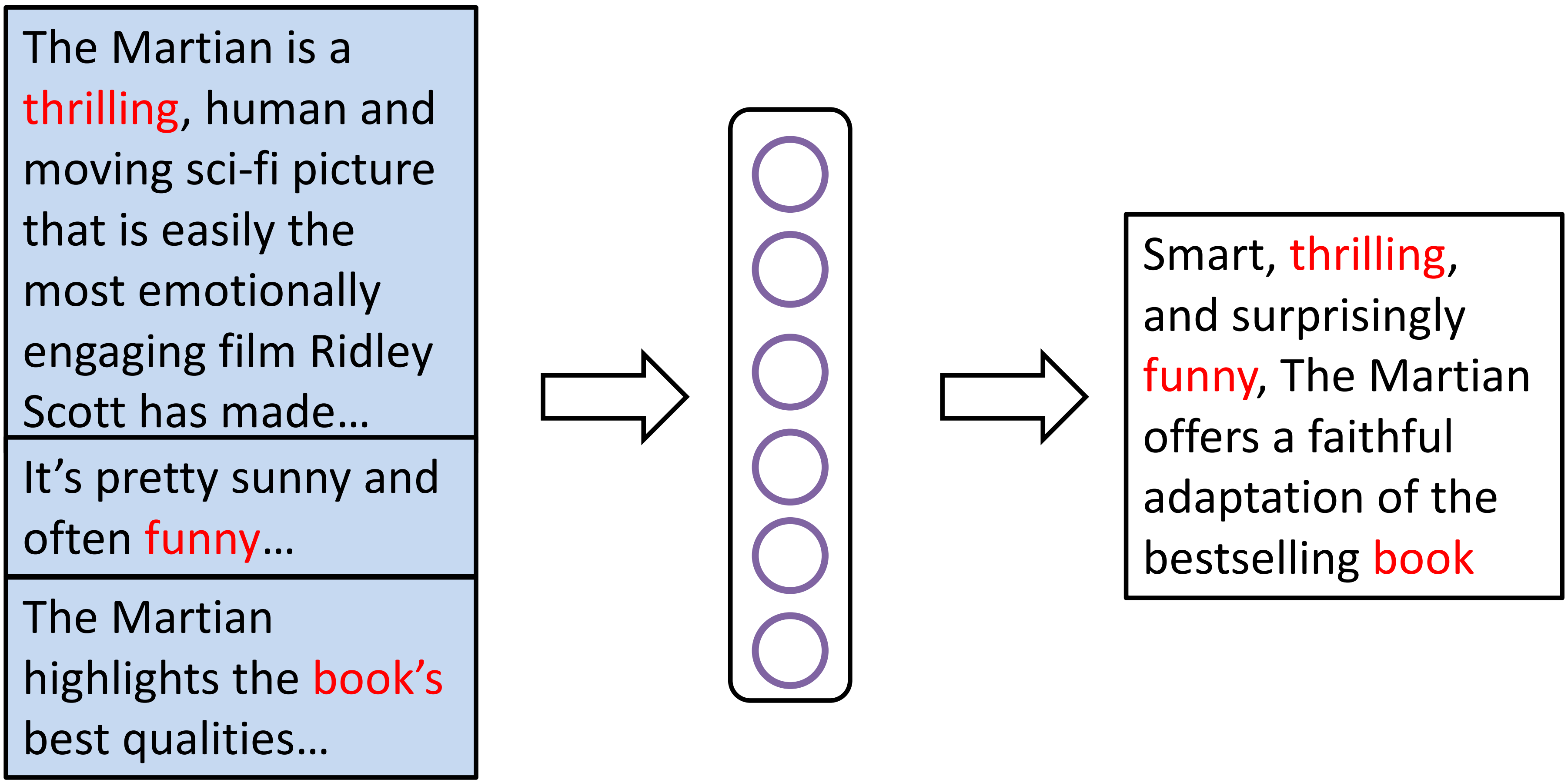
$$a_i = softmax(v(\mathbf{b}_i, \mathbf{h}_{j-1}))$$

v: implemented by feedforward neural network

b_i: *Implemented by bidirectional LSTM*

Attention Over Multiple Inputs

- Challenge: many input text units (lots of redundancy)



Attention Over Multiple Inputs

- Challenge: many input text units (lots of redundancy)
- Solution: sub-sampling from the input!

Importance Estimation

- Assigning an importance/salience score to each text input
  - (**0.9**) ….an intimate sci-fi epic that is smart, spectacular and stirring.
  - (**0.1**) The Martian proves to be an entertaining popcorn movie…

Importance Estimation

- Ridge regression based scorer (with preference regularization)
- Features
  - Centroidness (representativeness)
  - Average TF-IDF scores
  - Number of sentiment words

Post-processing

- We use beam search decoder
- Re-rank the n-best summaries according to similarity with all the input text units

Experimental Setup

- Word representation
  - Random initialization
  - Pre-trained word vectors from Google news (Mikolov et al., 2013)
- Additional features for each word
  - TF-IDF score, POS tag, dependency relation…
  - Attached to word vectors
- Sub-sampling:
  - Training: Sample K text units based on importance scores each iteration
  - Testing: Top K important text units

Automatic Evaluation

- Metrics
  - ROUGE: n-grams recall of the summaries with gold-standard abstracts as reference. (Lin and Hovy, 2003)
  - BLEU: n-grams precision (Papineni et al, 2002)
  - METEOR: recall-based, but considers synonyms and paraphrases (Denkowski and Lavie, 2014)

Automatic Evaluation

- Comparison
  - Baseline: longest text unit
  - LexRank (Erkan and Radev, 2004): PageRank-based centroidness estimation algorithm
  - OPINOSIS (Ganesan et al. 2010): Abstractive summarization system based on sentence merging and compression

	RottenTomatoes			*Idebate*		
	Length	**BLEU**	**METEOR**	**Length**	**BLEU**	**METEOR**
Longest	47.9	8.25	**8.43**	44.0	6.36	10.22
LexRank	16.7	19.93	5.59	26.5	13.39	9.33
Abstractive	22.0	19.72	6.07	23.2	15.09	**10.76**

Our Systems:

Words	15.7	19.88	6.07	14.4	22.55	7.38
Words (pre-trained)	15.8	23.22	6.51	13.9	23.93	7.42
Words+fea	17.5	19.73	6.43	13.5	23.65	7.33
Words+fea (pre-trained)	14.2	**24.88**	6.00	13.0	**25.84**	7.56

	RottenTomatoes			*Idebate*		
	Length	**BLEU**	**METEOR**	**Length**	**BLEU**	**METEOR**
Longest	47.9	8.25	**8.43**	44.0	6.36	10.22
LexRank	16.7	19.93	5.59	26.5	13.39	9.33
Abstractive	22.0	19.72	6.07	23.2	15.09	**10.76**

Our Systems:

Words	15.7	19.88	6.07	14.4	22.55	7.38
Words (pre-trained)	15.8	23.22	6.51	13.9	23.93	7.42
Words+fea	17.5	19.73	6.43	13.5	23.65	7.33
Words+fea (pre-trained)	14.2	**24.88**	6.00	13.0	**25.84**	7.56

	RottenTomatoes			*Idebate*		
	Length	**BLEU**	**METEOR**	**Length**	**BLEU**	**METEOR**
Longest	47.9	8.25	**8.43**	44.0	6.36	10.22
LexRank	16.7	19.93	5.59	26.5	13.39	9.33
Abstractive	22.0	19.72	6.07	23.2	15.09	**10.76**
Our Systems:						
Words	15.7	19.88	6.07	14.4	22.55	7.38
Words (pre-trained)	15.8	23.22	6.51	13.9	23.93	7.42
Words+fea	17.5	19.73	6.43	13.5	23.65	7.33
Words+fea (pre-trained)	14.2	**24.88**	6.00	13.0	**25.84**	7.56

The average length of human abstracts are about 11.5 and 24.6 for the two datasets.

Our Systems:

	RottenTomatoes		*Idebate*	
	BLEU	**METEOR**	**BLEU**	**METEOR**
Words	19.88	6.07	22.55	7.38
Words (pre-trained)	23.22	6.51	23.93	7.42
Words+fea	19.73	6.43	23.65	7.33
Words+fea (pre-trained)	**24.88**	6.00	**25.84**	7.56

- Pre-trained word embedding improves performance
- Additional features do not always improve performance, but helps with convergence

Human Evaluation

- Randomly select 40 movie summaries
- Each is evaluated by 5 human judges
  - Informativeness
  - Grammaticality
  - Compactness
  - Overall ranking

	Informative	**Grammatical**	**Compact**	**Best %**
LexRank	3.4	4.5	4.3	11.5%
Abstractive	2.8	3.1	3.3	5.0%
Our System	**3.6**	**4.8**	4.2	18.0%
Human Abstract (reference)	4.2	4.8	4.5	65.5%

Sample Summaries

- ***Movie: The Neverending Story***
- **Human**: A magical journey about the power of a young boy's imagination to save a dying fantasy land, The Neverending Story remains a much-loved kids adventure.
- **LexRank**: It pokes along at times and lapses occasionally into dark moments of preachy philosophy, but this is still a charming, amusing and harmless film for kids.
- **Opinosis**: The Neverending Story is a silly fantasy movie that often shows its age .
- **Our System**: The Neverending Story is an entertaining children's adventure, with heart and imagination to spare.

Sample Summaries

- **Topic:** *This House would detain terror suspects without trial.*
- Arguments:
  - Governments must have powers to protect their citizens against threats to the life of the nation.
  - Everyone would recognise that rules that are applied in peacetime may not be appropriate during wartime.
  - …

Sample Summaries

- **Human**: Governments must have powers to protect citizens from harm.
- **LexRank**: This is not merely to directly protect citizens from political violence, but also because political violence handicaps the process of reconstruction in nation-building efforts.
- **Our System**: Governments have the obligation to protect citizens from harmful substances.

Conclusion and Future Work

- We presented a neural network-based abstract generation framework for summarizing opinionated text.
- For future work
  - Towards generating paragraphs with discourse structure
  - Adding semantic information
    - “a poignant coming-of-age tale marked by a breakout lead performance from Cate Shortland” (for movie “Lore”)

Thank you!