香港城市大學
City University of Hong Kong

Abstractive Summarization with the Aid of Extractive Summarization

Yangbin Chen, Yun Ma, Xudong Mao, Qing Li

1

Outline

- Background

- Motivation

- Proposed Model

- Experiment

- Conclusion & Future Work

Background

- **Automatic Summarization** aims to condense a document in a grammatically and logically coherent way, meanwhile preserving the important information.

- Applications: Robotic news writer[1], News digest[2], Search engine[3], etc.

- **Extractive Summarization** extracts salient sentences from the source document and group them to produce a summary.

- **Abstractive Summarization** produces generalized summaries conveying information in a concise way and eliminating the limitations to the original words and sentences.

1 https://www.leiphone.com/news/201707/4a3BM1YK3fVSwiCL.html
2 https://readhub.me/
3 https://www.google.com/

香港城市大學
City University of Hong Kong

Extractive Summary

Advantage: keep the most important information
Disadvantage: incoherent or not concise

Jack White announces short run of first-ever all-acoustic shows

billboard

By Ray Waddell and Billboard staff, Billboard
Updated 1413 GMT (2213 HKT) April 15, 2015

(Billboard) — Fresh off his scorching performance at Coachella Saturday night (and days before his next one on the festival's second weekend), rocker Jack White announced he'll take a hiatus from touring.

White will wrap his touring efforts in support of "Lazeretto" with a brief, first-ever acoustic tour that will hit "the only five states left in the U.S. that he has yet to play," according to White's website.

Rounding out the acoustic quartet on tour will be Fats Kaplin, Lillie Mae Rische and Dominic Davis. The shows will be unannounced until day-of-show, with tickets priced at $3 and limited to one ticket per person, to be purchased only at the venue on a first-come, first-served basis.

Billboard: Jack White on Not Being a 'Sound-Bite Artist,' Living in the Wrong Era and Why Vinyl Records Are 'Hypnotic'

The purposely vague announcement surely has fans (and journalists) scouring the Internet for White's touring history.

Unclear is whether White includes his work with The White Stripes, The Raconteurs and Dead Weather in his touring history, or just his solo road work. Presumably, he's including all of his touring, with all bands, as Billboard could find only 29 states in which he has performed as Jack White. Tour dates with White Stripes add another 12 states.

That leaves nine states for which we could not find a show for White: Hawaii (where a show is scheduled for tomorrow, April 15), Arkansas, Idaho, Utah, Wyoming, Vermont, Iowa, and North and South Dakota.

Billboard: Jack White Plays The Hits, Declares 'Music Is Sacred' at Coachella

Through the process of elimination (surely he has played Boise, Little Rock, and Salt Lake?), our guess as to which five "states" White will play on the brief acoustic run: South and North Dakota, Wyoming, Vermont and ... Puerto Rico? If that's the case, this tour is in for some long jumps, with Puerto Rico to Vermont being a potential beast. (Though shipping acoustic instruments and ribbon mics will be a lot less taxing than a full electrified stage setup.)

The shows will be unannounced until day-of-show, with tickets priced at $3 and limited to one ticket per person, to be purchased only at the venue on a first-come, first-served basis.

Billboard: Jack White Plays The Hits, Declares 'Music Is Sacred' at Coachella

Through the process of elimination (surely he has played Boise, Little Rock, and Salt Lake?), our guess as to which five "states" White will play on the brief acoustic run: South and North Dakota, Wyoming, Vermont and ... Puerto Rico? If that's the case, this tour is in for some long jumps, with Puerto Rico to Vermont being a potential beast.

Abstractive Summary

Advantage: concise and more readable
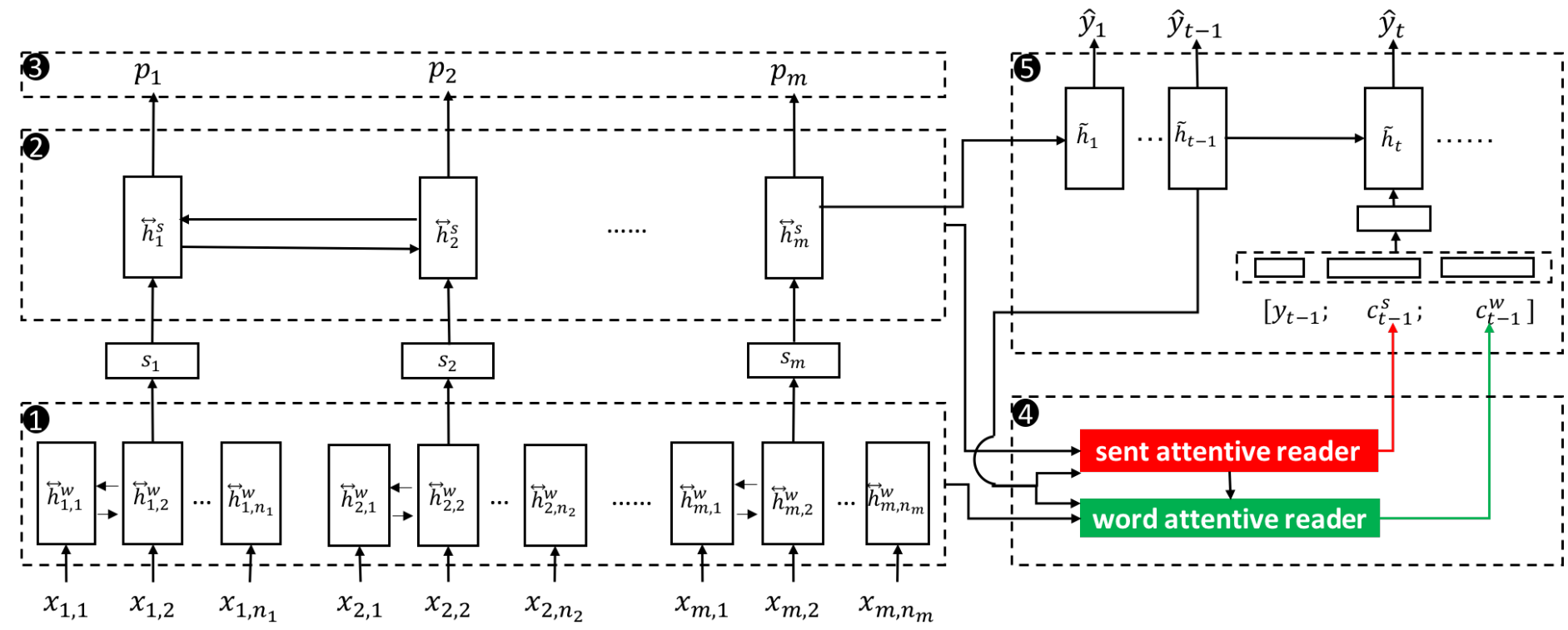**Disadvantage**: mistake grammars or facts

Places and times of shows are currently a mystery.

Motivation

- To combine the advantages of both extractive summarization and abstractive summarization.

- To solve the long-term dependency problem of RNN-based language model for document representation.

Proposed Model

- **Framework**

1. Word-level encoder
2. Sentence-level encoder
3. Sentence extractor
4. Hierarchical attention
5. Decoder

Proposed Model

- **Neural Summarization Model**

❶ Word-level encoder

$$\overrightarrow{\boldsymbol{h}}_{i,j}^{w} = GRU(\boldsymbol{x}_{i,j}, \overrightarrow{\boldsymbol{h}}_{i,j-1}^{w})$$

$$\overleftarrow{\boldsymbol{h}}_{i,j}^{w} = GRU(\boldsymbol{x}_{i,j}, \overleftarrow{\boldsymbol{h}}_{i,j+1}^{w})$$

$$\overleftrightarrow{\boldsymbol{h}}_{i,j}^{w} = [\overrightarrow{\boldsymbol{h}}_{i,j}^{w}; \overleftarrow{\boldsymbol{h}}_{i,j}^{w}]$$

❷ Sentence-level encoder

$$\boldsymbol{s}_i = tanh(\mathbf{W} \cdot \frac{1}{n_i} \sum_{j=1}^{n_i} \overleftrightarrow{\boldsymbol{h}}_{i,j}^{w} + \boldsymbol{b})$$

$$\overrightarrow{\boldsymbol{h}}_i^s = GRU(\boldsymbol{s}_i, \overrightarrow{\boldsymbol{h}}_{i-1}^s)$$

$$\overleftarrow{\boldsymbol{h}}_i^s = GRU(\boldsymbol{s}_i, \overleftarrow{\boldsymbol{h}}_{i+1}^s)$$

$$\overleftrightarrow{\boldsymbol{h}}_i^s = [\overrightarrow{\boldsymbol{h}}_i^s; \overleftarrow{\boldsymbol{h}}_i^s]$$

❸ Sentence extractor

$$p_i = P(q_i = 1 | \overleftrightarrow{\boldsymbol{h}}_i^s)$$

$$= \sigma(\mathbf{W}^{extr} \cdot \overleftrightarrow{\boldsymbol{h}}_i^s + b^{extr})$$

$$E_{se} = -\frac{1}{m} \sum_{i=1}^{m} q_i log p_i + (1 - q_i) log(1 - p_i)$$

Proposed Model

- **Neural Summarization Model**

❹ Hierarchical attention

$$\boldsymbol{\alpha}_{t,i} = \frac{\boldsymbol{e}_{t,i}^s}{\sum_{k=1}^m \boldsymbol{e}_{t,k}^s}$$

$$\boldsymbol{e}_{t,i}^s = exp\{\mathbf{V}^{sT} \cdot tanh(\mathbf{W}_1^{dec} \cdot \tilde{\boldsymbol{h}}_t + \mathbf{W}_1^s \cdot \overleftrightarrow{\boldsymbol{h}}_i^s + \boldsymbol{b}_1^s)\}$$

$$\boldsymbol{\beta}_{t,i,j} = \boldsymbol{\alpha}_{t,i} \frac{\boldsymbol{e}_{t,i,j}^w}{\sum_{l=1}^{n_i} \boldsymbol{e}_{t,i,l}^w}$$

$$\boldsymbol{e}_{t,i,j}^w = exp\{\mathbf{V}^{wT} \cdot tanh(\mathbf{W}_2^{dec} \cdot \tilde{\boldsymbol{h}}_t + \mathbf{W}_2^w \cdot \overleftrightarrow{\boldsymbol{h}}_{i,j}^w + \boldsymbol{b}_2^w)\}$$

$$\boxed{E_a = \frac{1}{m} \sum_{i=1}^m (p_i - \frac{1}{T} \sum_{t=1}^T \boldsymbol{\alpha}_{t,i})^2}$$

❺ Decoder

$$\boldsymbol{c}_t^s = \sum_{i=1}^m \boldsymbol{\alpha}_{t,i} \cdot \overrightarrow{\boldsymbol{h}}_i^s$$

$$\boldsymbol{c}_t^w = \sum_{i=1}^m \sum_{j=1}^{n_i} \boldsymbol{\beta}_{t,i,j} \cdot \overleftarrow{\boldsymbol{h}}_{i,j}^w$$

$$\tilde{\boldsymbol{h}}_t = GRU(\tilde{\boldsymbol{h}}_{t-1}, f_{in}(\hat{\boldsymbol{y}}_{t-1}, \boldsymbol{c}_{t-1}^s, \boldsymbol{c}_{t-1}^w))$$

$$P(\hat{\boldsymbol{y}}_t | \hat{\boldsymbol{y}}_1, ..., \hat{\boldsymbol{y}}_{t-1}, \boldsymbol{x}) = g(f_{out}(\tilde{\boldsymbol{h}}_t, \boldsymbol{c}_t^s, \boldsymbol{c}_t^w))$$

$$\boxed{E_y = \frac{1}{T} \sum_{t=1}^T -log(y_t)}$$

Loss function: $\quad E = E_y + \lambda \cdot E_{se} + \gamma \cdot E_a$

Experiment

- **Dataset [1]**

Dataset	Train	Valid	Test	S.S.N.	S.S.L.	T.S.L.
CNN	83,568	1,220	1,093	30.17	24.02	40.10
DailyMail	193,968	12,147	10,350	25.60	28.84	59.30

S.S.N.: the average number of sentences in the source document
S.S.L.: the average length of the sentences in the source document
T.S.L.: the average length of the sentences in the target summary

Experiment

- **Implementation details**
  - Vocabulary size: 50K
  - Word embedding size: 300
  - Hidden size: 200
  - λ: 100
  - γ: 0.5
  - Optimizer: Adagrad (learning_rate 0.15)
  - Mini-batch size: 16
  - Beam search size: 5
  - Maximum number of sentences per document: 35
  - Maximum words per sentence of source document: 50
  - Maximum words of target summary: 100

Experiment

- **Result #1: Comparison with baselines**

Method	Rouge-1	Rouge-2	Rouge-L
seq2seq+attn[2]	33.6	12.3	31.0
words-lvt2k-hieratt[3]	35.4	13.3	32.6
Our Method	**35.8**	**13.6**	**33.4**

Experiment

- **Result #2: Evaluation of proposed components**

Method	Rouge-1	Rouge-2	Rouge-L
Our Method	**35.8**	**13.6**	**33.4**
w/o extr	34.3	12.6	31.6
w/o attn	34.7	12.8	32.2
w/o extr+attn	34.2	12.5	31.6

Experiment

■ **Result #3: Case study**

Source Document
a flaw in robots designed to perform surgery has been found that lets them be easily hacked, according to researchers. the experts were able to take control of a so-called tele-robot during surgery by exploiting a simple programming trick. this enabled them to change the speed of the arms of the robot and change their orientation, making it impossible for the machines to carry out a procedure as directed. researchers at the @entity19 studied the @entity20 (shown) they found that robots designed for surgery could be 'easily' hacked in to. this is because they are operated over public networks which allowed the researchers to access them and stop them working the research was carried out by scientists ……
Golden Summary
researchers at the @entity19 studied so-called tele-robots. they found robots designed for surgery could be hacked and manipulated. this is because robots being tested were operated over public networks. it allowed the researchers to access them and stop them working.
Our Method
researchers at the @entity19 studied the @entity20. they found that robots designed for surgery could be 'easily' hacked in to. it enabled them to change the speed of the arms of the robot and change their orientation, making it impossible for the machines to carry out a procedure as directed. this is because they are operated over public networks which allowed the researchers to access them and stop them working.
Seq2seq+attn
researchers at the @entity19 studied the @entity20 (shown) robots designed for surgery could be 'easily' hacked in to. this is because they are operated over public networks which allowed the researchers to access them and stop them working.

Conclusion

- A joint learning with Extractive model improves the performance of the Abstractive model.

- The hierarchical encoder speeds up the training process.

- The attention constraint strengthen the consistency between the two tasks.

Future Work

- Analyze how the multi-task model influences all single tasks and the feasibility to strengthen them all.

- Incorporate the state-of-the-art models for certain components.

- Try semi-supervised and unsupervised methods.

- Add human evaluations.

Reference

[1] Jianpeng Cheng and Mirella Lapata. 2016. Neural summarization by extracting sentences and words. *arXiv preprint arXiv:1603.07252* .

[2] Rush, A. M., Chopra, S., & Weston, J. (2015). A neural attention model for abstractive sentence summarization. *arXiv preprint arXiv:1509.00685*.

[3] Sumit Chopra, Michael Auli, and Alexander M Rush. 2016. Abstractive sentence summarization with attentive recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* pages 93–98.

[4] Romain Paulus, Caiming Xiong, and Richard Socher. 2017. A deep reinforced model for abstractive summarization. In *Proceedings of the 2018 International Conference on Learning Representations.*

[5] Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get to the point: Summarization with pointer- generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics* (Volume 1: Long Papers), volume 1, pages 1073–1083.

[6] Wan-Ting Hsu, Chieh-Kai Lin, Ming-Ying Lee, Kerui Min, JingTang, and Min Sun. 2018. A Unified Model for Extractive and Abstractive Summarization using Inconsistency Loss. accepted by *ACL 2018*.

Thank you!

Q & A

Email: robinchan1992@gmail.com
Github: https://github.com/Codelegant92

Supplemental Materials

DeepRL[4]	39.87	15.82	36.90
Pointer-generator[5]	39.53	17.28	36.38
Unified Model[6]	**40.68**	**17.97**	**37.13**

Supplemental Materials

Extractive summarization

AFL star blames vomiting cat for speeding

Adelaide Crows defender Daniel Talia has kept his driving license, telling a court he was speeding 36km over the limit because he was distracted by his sick cat.

The 22-year-old AFL star, who drove 96km/h in a 60km/h road works zone on the South Eastern expressway in February, said he didn't see the reduced speed sign because he was so distracted by his cat vomiting violently in the back seat of his car.

In the Adelaide magistrates court on Wednesday, Magistrate Bob Harrap fined Talia $824 for exceeding the speed limit by more than 30km/h.

He lost four demerit points, instead of seven, because of his significant training commitments.

Adelaide Crows defender Daniel Talia has kept his driving license, telling a court he was speeding 36km over the limit because he was distracted by his sick cat. In the Adelaide magistrates court on Wednesday, Magistrate Bob Harrap fined Talia $824 for exceeding the speed limit by more than 30km/h.

Supplemental Materials

Abstractive summarization

AFL star blames vomiting cat for speeding

Adelaide Crows defender Daniel Talia has kept his driving license, telling a court he was speeding 36km over the limit because he was distracted by his sick cat.

The 22-year-old AFL star, who drove 96km/h in a 60km/h road works zone on the South Eastern expressway in February, said he didn't see the reduced speed sign because he was so distracted by his cat vomiting violently in the back seat of his car.

In the Adelaide magistrates court on Wednesday, Magistrate Bob Harrap fined Talia $824 for exceeding the speed limit by more than 30km/h.

He lost four demerit points, instead of seven, because of his significant training commitments.

Adelaide Crows defender Daniel Talia admits to speeding but says he didn't see road signs because his cat was vomiting in his car. 22-year-old Talia was fined $824 and four demerit points, instead of seven, because of his 'significant' training commitments.

Supplemental Materials

- BLEU
- ROUGE
- METEOR
- …
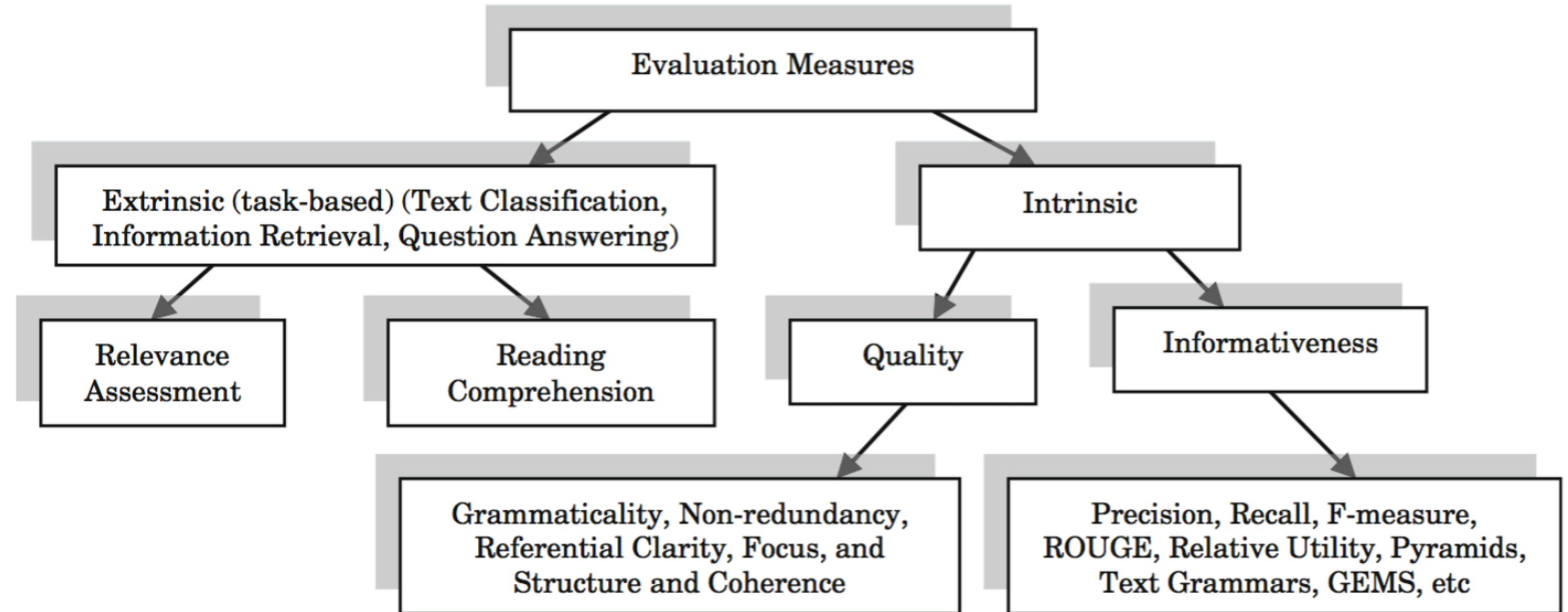
Supplemental Materials

ROUGE(Recall-Oriented Understudy for Gisting Evaluation)

- Task: Machine Translation, Summarization, Dialogue System, etc.

- ROUGE-N, ROUGE-L, ROUGE-S