Learning Comment Generation by Leveraging User-Generated Data

Zhaojiang Lin, Genta Indra Winata, Pascale Fung
ICASSP 2019 Presentation
May 16th, 2019

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

CAiRE
Centre for Artificial Intelligence Research

Outline

1. **Introduction**

2. **Existing works:**

   a. **Information-Retrieval based Approaches**

   b. **Generation based Approaches**

3. **Proposed Method**

4. **Conclusion**

Automatic Comment Generation

Apple's iPhone 8 event is happening in Sept

Apple has sent out invites for its next big event on September 12th, where the company is expected to reveal the next iPhone, along with updates to the Apple Watch, Apple TV, and iOS software. Apple is expected to announce three new iPhones at the event: a next-generation iPhone 8 model with an OLED display and a 3D face-scanning camera; and updated versions of the iPhone 7 and 7 Plus.

Comment 1: Remember a year of iPhone 5 rumors followed by the announcement of the iPhone 4S? I will be highly entertained if Apple does something similar.

Comment 2:
Looking forward to this event!

Why this task is important?

Motivation:

- **Improve the engagement of articles**

  - Provide extended information and rich personal views*

  - Engage readers to share their opinions on the topic

- **One of important skills for social chatbots**

*Lianhui Qin, Lemao Liu, Wei Bi, Yan Wang, Xiaojiang Liu, Zhiting Hu, Hai Zhao, and Shuming Shi, "Automatic article commenting: the task and dataset," in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 2018, pp. 151–156, Association for Computational Linguistics.

Main Challenges

Challenges

- The **ability** to **comment** on an article requires natural language understanding to **conceptualize the idea of the article** and **provide a relevant response**.
- How to distinguish **good comments** and **bad comments**.
- **One** to **many** generation problem.

Information-Retrieval based Approaches

Comments Rankers:

- Keyword-based similarity measures: TF-IDF (Chen et. al, 2017)
- Neural-based ranker: CNN (Qin et. al, 2018), Expected Value LSTM (Rao and Daumé III, 2018)

Goal: Choose a comment from a pool of human-generated comments.

- Most of the time, they are are grammatical
- Contain important keywords found in the article
- Retrieved comments are not proper
- Less robust when applied to new domains

Generation-based Approaches

Generation base commenting:

- Seq2Seq with Attention (Qin et. al, 2018)
- A gated attention neural-based generation to address the problem of contextual relevance by choosing news context (Zheng, et al., 2018).
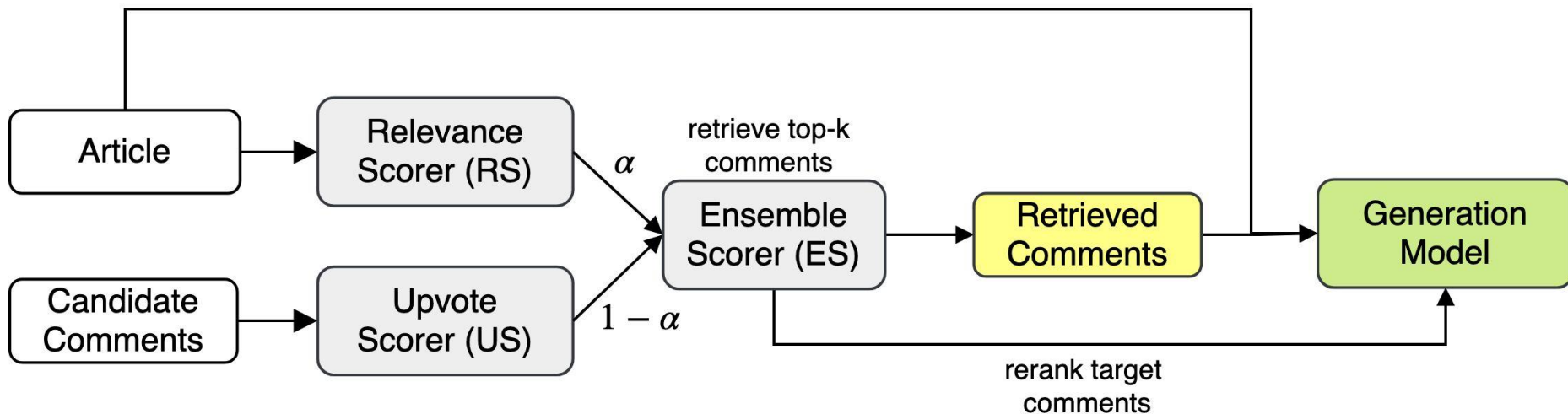
Generate a sequence using Seq2Seq-based model.

- Can construct new and unseen comments
- More robust and scalable
- Generic and boring comments
- May be ungrammatical
- Irrelevant to the articles

Proposed Method

Framework

Good vs. Bad comments!

Good comments

Apple's iPhone 8 event is happening in Sept

Apple has sent out invites for its next big event on September 12th, where the company is expected to reveal the next iPhone, along with updates to the Apple Watch, Apple TV, and iOS software. Apple is expected to announce three new iPhones at the event: a next-generation iPhone 8 model with an OLED display and a 3D face-scanning camera; and updated versions of the iPhone 7 and 7 Plus.

Comment 1: Remember a year of iPhone 5 rumors followed by the announcement of the iPhone 4S? I will be highly entertained if Apple does something similar.

Comment 2: Looking forward to this event!

Comment 3: I like the cat. it is so cute !

Comment 4: LOL。。。!!!

Bad comments

Apple's iPhone 8 event is happening in Sept

Apple has sent out invites for its next big event on September 12th, where the company is expected to reveal the next iPhone, along with updates to the Apple Watch, Apple TV, and iOS software. Apple is expected to announce three new iPhones at the event: a next-generation iPhone 8 model with an OLED display and a 3D face-scanning camera; and updated versions of the iPhone 7 and 7 Plus.

Comment 1: Remember a year of iPhone 5 rumors followed by the announcement of the iPhone 4S? I will be highly entertained if Apple does something similar.

Comment 2: Looking forward to this event!

Comment 3: I like the cat. it is so cute ! (irrelevant)

Comment 4: LOL。。。!!! (unreadable)

Using upvotes

Apple's iPhone 8 event is happening in Sept

Apple has sent out invites for its next big event on September 12th, where the company is expected to reveal the next iPhone, along with updates to the Apple Watch, Apple TV, and iOS software. Apple is expected to announce three new iPhones at the event: a next-generation iPhone 8 model with an OLED display and a 3D face-scanning camera; and updated versions of the iPhone 7 and 7 Plus.

Comment 1: Remember a year of iPhone 5 rumors followed by the announcement of the iPhone 4S? I will be highly entertained if Apple does something similar.

50 upvotes

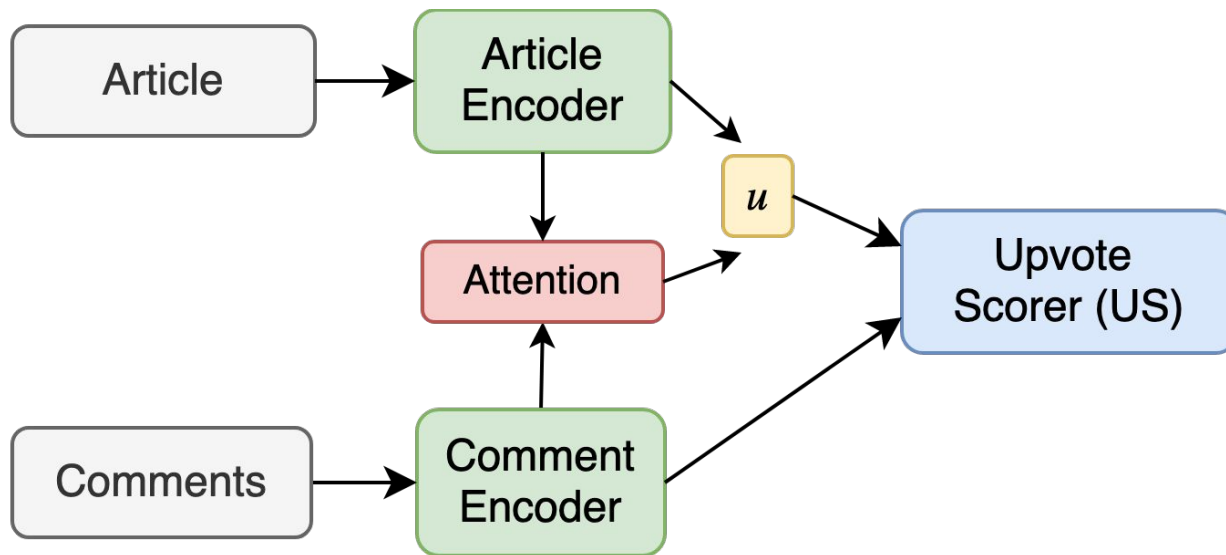**Comment 2:** Looking forward to this event!

10 upvotes

Comment 3: I like the cat. it is so cute !

0 upvotes

Comment 4: LOL。。。!!!

0 upvotes

Upvote Scorer (US)

Let a represents an article, c represents the comment.

Encoders

$$\{h_1^a, ..., h_n^a\} = \mathbf{BiLSTM}(\{v_1^a, ..., v_n^a\})$$

$$\{h_1^c, ..., h_m^c\} = \mathbf{BiLSTM}(\{v_1^c, ..., v_m^c\})$$

Attention mechanism

$$u = \sum_{i=1}^{n} \frac{exp(e_i)}{\sum_{k=1}^{n} exp(e_k)} h_i^a \qquad \text{where,} \qquad e_i = (h_i^a)^T \mathbf{W}_a h_m^c$$

Upvote score:

$$S_u = \sigma(\mathbf{F}([u, h_m^c]))$$

Relevance Scorer (RS)

Score is computed by dot product of TF-IDF weighted vector between an article with title a and comment c.

We normalize the scores (all scores are divided by the maximum score).

Relevance score: $S_r = \text{Normalize}((v^a)^T v^c)$

Ensemble Scorer (ES)

We combine the relevance score S_r and upvote score S_u by linear interpolation as an ensemble score.

$$S = \alpha S_r + (1 - \alpha)S_u \text{ , where } \alpha \in [0, 1]$$

Two-step retrieval

Given an article as query.

1. Retrieve top-5 relevant articles from a pool of articles.
2. Take all comments from retrieved articles as comment candidates.
3. Rerank comment candidates using ES.
4. Take top-k (k = 1 in our case) as retrieval result.

Generation model

We use Pointer Generator model (See, et. al., 2017) to learn how to copy important keywords from articles and retrieved comments, make generated comments more relevant to the article.

Input   : Concatenation of articles and retrieved comments
Target : Comments from a given article with the highest ES score

Abigail See, Peter J Liu, and Christopher D Manning, "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)*, 2017, vol. 1, pp. 1073–1083.

Rerank target comments

One article might have more than 100 comments with huge variance, A generation model trained by using all of comments only able to generate "safe responses" which frequently appear (e.g., I don't know, I am speechless) , which is commonly happened in one to many problem.

Solution :
1) Rerank target comments by using our ES, choose the top-1 as training target. (reduce to one to one problem)
2) Or use reinforcement learning to reward high score comments(time consuming and hard to tune)
3) CVAE to maximize the variational lower bound(not suitable, because we don't want to model to bad comments)

Result

Table 1. Results on information retrieval and generation approaches. Higher scores are better.

	Model	BLEU-1	CIDEr	ROUGE_L	METEOR
Retrieval	TF-IDF + CNN [1]	**35.55**	0.25	21.92	14.25
	TF-IDF + RS	34.67	**1.37**	**23.67**	14.80
	TF-IDF + ES	34.53	1.19	23.59	**14.85**
Generation	Seq2Seq-Attn [1]	38.80	1.41	23.53	6.08
	pointer-generator + coverage	40.84	1.29	25.44	6.49
	pointer-generator + coverage + upvote	62.03	3.55	28.24	11.00
	pointer-generator + coverage + RS	56.39	3.95	26.47	11.88
	pointer-generator + coverage + US	64.22	4.17	**30.87**	12.11
	pointer-generator + coverage + ES	**65.70**	**4.35**	30.53	**12.62**

Result

Title: 流浪狗被人残忍踢断下巴无法进食，路人心疼不已送它去医院。 (The stray dog was brutally kicked off the chin and could not eat. A passerby sent it to the hospital.)		
	Model	**Generated Comment**
Body: 一条流浪狗在路边的垃圾堆翻找着吃的东西，嘴里发出叫声，似乎很痛苦，路过的男子见状将其抱回。他发现流浪狗狗的下巴断了，似乎是人为踢断的，非常可怜，可以想象它有多么的痛苦。狗狗很饿，在垃圾堆里找吃的，但是它的下巴变成这样是无法吃东西的，不知道是什么人这么狠毒。这名男子非常心疼，马上带狗狗去了宠物医院。经过医生的治疗，狗狗的下巴正在恢复中，现在能够吃东西了，不过只能吃一些流体的、软的食物,回到咬骨头的时期恐怕是需要一段时间了。 (A stray dog was searching for food in the rubbish heap on the roadside. The cry seemed to be very painful, and the passing man saw it and took it home. He found the dog's chin was broken. It seemed to be kicked by someone. Very poor. We can imagine how painful it is. The dog was hungry, looking for food in the rubbish heap, but it was unable to eat due to the broken chin. We don't know who is so vicious. The man was very distressed, immediately took the dog to the pet hospital. After the doctor's treatment, dog's jaw is recovering, and now it can eat, but it can only eat some fluid, soft food, it may take a while to go back to the time of biting bones.)	TF-IDF + CNN [1]	想起来我曾经养的一条小狗...后来, 因为吃死老鼠被药死了。我一直都记着它那双会说话的眼睛。 (I remembered a puppy I used to raise... Later, It was dead from food poisoning because of a dead mouse. I always remember its talking eyes.)
	TF-IDF + ES	听听都害怕, 踢断下巴 (I feel scared to hear that the chin was kicked off.)
	Seq2Seq-Attn	好人有好报 (A good person will be rewarded)
	pointer-generator + coverage + RS	狗狗的下巴断了，似乎是无法吃东西的，但是只能想象它有多么的痛苦。 (The dog's chin was broken. It seemed that it was unable to eat. But we can imagine how painful it is.)
	pointer-generator + coverage + ES	可怜的狗狗! (Poor dog!)

Conclusion

- A novel framework for comment generation to **leverage user-generated data** and **generate relevant comments** according to the **user preference**.
- **Incorporate retrieved comments** helps the generation model.
- We efficiently alleviate the **one-to-many generation problem (mode collapse issue)**, and produce **more meaningful comments**.

Thanks!
Q & A