Automatic Article Commenting: the Task and Dataset

Lianhui Qin[1], Lemao Liu[2], Wei Bi[2], Yan Wang[2], Xiaojiang Liu[2],
Zhiting Hu, Hai Zhao[1], Shuming Shi[2]
Shanghai Jiao Tong University[1], Tencent AI Lab[2]

Overview

- Proposes the new task of automatic article commenting, which needs:
  - understanding the given article
  - formulating opinions and arguments
  - organizing natural language for expression
- Releases a large-scale Chinese corpus
  - 200K articles + 4.5M human comments
  - A human-annotated test set
- Develops a general approach to enhancing popular automatic evaluation metrics
- Compares basic generation/retrieval approaches empirically

Title: 苹果公司iPhone 8 发布会定在9月12号举行(Gloss: Apple's iPhone 8 event is happening on Sept. 12th)

CMT: 苹果公司正式向媒体发布邀请函，宣布将于9月12日召开苹果新品发布会，该公司将发布下一代iPhone，随之更新的还有苹果手表，苹果TV，和iOS软件。这次发布会将带来三款新iPhones：带有边缘OLED显示屏和3D人脸扫描技术的下一代iPhone8；现代iPhone 7、iPhone 7Plus的更新版。另外，据说苹果还会发布一款新的4K版AppleTV。
(Gloss: 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 widely expected to announce three new iPhones at the event: a next-generation iPhone 8 model with an edge-to-edge OLED display and a new 3D face-scanning camera; and updated versions of the current iPhone 7 and 7 Plus with wireless charging.Additionally, the company is rumored to have a new 4K Apple TV in the works)

Article Commenting Dataset

Score	Criteria	Example Comments
5	Rich in content; attractive; deep insights; new yet relevant viewpoints	还记得那年iphone 4发布后随之而来的关于iPhone 5 传闻吗？如果苹果今年也是这样我会觉得很滑稽。 (Gloss: 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.)
4	Highly relevant with meaningful ideas	可以直接说：“我们相约在那个公园”。 (Gloss: Could have just said "Meet us at the Park.)"
3	Less relevant; pale in meaning	很期待这个发布会啊！ (Gloss: Looking forward to the announcements.)
2	Fluent and grammatical but irrelevant	我喜欢这只猫，它很可爱！！ (Gloss: I like the cat. it is so cute !)
1	Hard to read	LOL。。。！！！

Table 1: A data example of an article and selected comments. A brief version of human judgment criteria is also listed.

	Train	Dev	Test
#Articles	191,502	5,000	1,610
#Cmts/Articles	27	27	27

Table 2: Data statistics.

- Collected from Tencent News (news.qq.com)
- Tokenized; improper data filtered out
- Comments in test set are rated by human

Quality Weighted Automatic Metrics

- Human comments are of varying quality
- Enhances the reference-based metrics to account for the different reference quality scores
  - Enhanced BLEU, METEOR, ROUGE, CIDEr
  - E.g., Quality weighted METEOR

$$\text{W-METEOR}(\mathbf{c}, \mathcal{R}) = (1-BP)\max_j s^j F_{mean,j}$$

Title	Baby重回《跑男》 (Gloss: AngelaBaby is coming back to <Running Man>)
CMT	Baby, Baby, 我爱你。 (Gloss:Baby, Baby, I love you.)
Scores	Human: **3** Normalized-METEOR: **4.8** (METEOR: 0.62) Normalized-W-METEOR: **3.8** (W-METEOR: 0.34)
Title	三兄弟在车祸中受伤。 (Gloss: Three siblings injured in car crash.)
CMT	祝愿三兄弟无恙。 (Gloss:I hope all is well for the three guys.)
Scores	Human: **3** Normalized-METEOR: **3.9** (METEOR: 0.40) Normalized-W-METEOR: **3.2** (W-METEOR: 0.19)

Table 3: Examples showing different scores of METEOR and W-METEOR.

The dataset is available on `http://ai.tencent.com/upload/PapersUploads/article_commenting.tgz`