Supplementary Page

This is the supplementary page for our latest paper: **Generative Adversarial Network for Abstractive Text Summarization**. Exhibiting competitive ROUGE scores with the state-of-the-art methods on the CNN/ Daily Mail dataset, experimental results show that our model is able to generate more abstractive, readable and diverse summaries.

1 Our Model

1.1 Generative Model

The generator takes the source text $x = \{w_1, w_2, ..., w_n\}$ as input, and predict the summary $\hat{y} = \{\hat{y_1}, \hat{y_2}, ..., \hat{y_m}\}$. Here, the n is the length of the source text x and m is the length of the predicted summary. We use a bi-directional LSTM encoder to convert the input text x into a sequence of hidden states $h = \{h_1, \ldots, h_n\}$. We then use a single LSTM decoder, computing the decoding hidden states s_t from the embedding vectors of y_t.

To prevent the model from attending over the same part of the input on different decoding steps, we first calculate the attention distribution at each decoding time step t as in (Bahdanau, Cho, and Bengio 2014):

$$e_i^t = v^T tanh(W_h h_i + W_s s_t + b_{attn}) \tag{1}$$

$$a^t = softmax(e^t) \tag{2}$$

where v, W_h, W_s and b_{attn} are learnable parameters. The attention distribution helps the decoder to decide which specific part of the input should be emphasized when producing the next word. Then we can calculate the context vector c_t^e as the weighted sum of encoder hidden states and the vocabulary distribution over all words in vocabulary:

$$c_t^e = \sum_{i=1}^{n} a_i^t h_i \tag{3}$$

$$P_{vocab}(\hat{y}_t) = softmax(V^{'}(V[s_t, c_t] + b) + b^{'}) \tag{4}$$

where $V^{'}$, V, b, $b^{'}$ are learnable parameters. providing us with the final probability to predict word from vocabulary. Following the work of [SLM17], we incorporate a switching pointer-generator network to use either word generator from fixed vocabulary or pointer copying rare or unseen from the input sequence. The generation probability p_{gen} for each decoding step is calculated as:

$$p_{gen} = \sigma(w_h^T h_t + w_s^T s_t + w_x^T x_t + b_{ptr}) \tag{5}$$

Finally, we can get the probability $P(\hat{y}_t)$ of each token in the target:

$$P(w) = p_{gen} P_{vocab}(w) + (1 - p_{gen}) \sum_{i:w_i=w} a_i^t \tag{6}$$

1.2 Discriminative Model

The discriminator is a binary classifier and aim at distinguishing the input sequence as originally generated by humans or synthesized by machines. We encode the input sequence with a CNN as it shows great effectiveness in text classification [Kim14]. We use multiple filters with varying window sizes to obtain multiple features and then apply a max-over-time pooling operation over the features. These pooled features are passed to a fully connected softmax layer whose output is the probability of being original.

1.3 Updating Model Parameters

In the adversarial process, using the discriminator as a reward function can further improve the generator iteratively by dynamically updating the discriminator. Once we obtain more realistic and high quality summaries generated by generator G, we re-train the discriminator as:

$$\min_{\phi} -\mathbf{E}_{Y \sim p_{data}}[logD_{\phi}(Y)] - \mathbf{E}_{Y \sim G_{\theta}}[log(1 - D_{\phi}(Y))]$$

When the discriminator D is obtained and fixed, we are ready to update the generator G. The loss function of our generator G consists two parts: the loss computed by policy gradient (denoted by J_{pg}) and the maximum-likelihood loss (denoted by J_{ml}). Formally, the objective function of G is $J = \beta J_{pg} + (1-\beta)J_{ml}$, where β is the scaling factor to balance the magnitude difference between J_{pg} and J_{ml}.

1.3.1 Loss by Policy Gradient

The generator G is considered as a stochastic parameterized policy and the score assigned by the discriminator is used as reward. The generator is trained to maximize its expected end reward:

$$L_{pg}(\theta) = \sum_{Y_{1:T-1}} G_{\theta}(Y_{1:T}|X)R_D^{G_{\theta}}((Y_{1:T-1}, X), y_T) \tag{7}$$

where $Y_{1:T} = y_1, ..., y_T$ represents the generated target sequence, X is the given source sequence. $R_D^{G_{\theta}}$ is the action-value function of a target sentence. For example, after generating the first $T-1$ words, following the policy G_{θ}, we take the action Y_T. We use the REINFORCE algorithm and take the estimated probability of being human generated by the discriminator D as the reward:

$$R_D^{G_{\theta}}((Y_{1:T-1}, X), y_T) = D(X, Y_{1:T}) - b(X, Y_{1:T}) \tag{8}$$

$b(x, y)$ denotes the baseline value to reduce the variance of the reward while keeping it unbiased. Here we adjust it as the standard score among all data. Since the Discriminator can only score a fully generated sequence, following [YZWS16], we use Monte Carlo Tree Search (MCTS) to evaluate the reward for an intermediate state. The gradient of the objective function $J(\theta)$ is calculated as:

$$\bigtriangledown_{\theta} J_{pg} = \frac{1}{T} \sum_{t=1}^{T} \sum_{y_t} R_D^{G_{\theta}}((Y_{1:t-1}, X), y_t) \cdot \bigtriangledown_{\theta}(G_{\theta}(y_t|Y_{1:t-1}, X))$$

$$= \frac{1}{T} \sum_{t=1}^{T} \mathbf{E}_{y_t \in G_{\theta}}[R_D^{G_{\theta}}((Y_{1:t-1}, X), y_t) \bigtriangledown_{\theta} \log p(y_t|Y_{1:t-1}, X)] \tag{9}$$

1.3.2 Loss of maximum-likelihood function

The most widely used and traditional method to train a sequence generation decoder is to minimize the a maximum-likelihood loss during each decoding step. Suppose $y^* = y_1^*, y_2^*, ..., y_{n'}^*$ is the groud-truth summary sequence for source text sequence x. The maximum-likelihood training loss is:

$$L_{ml}(\theta) = -\sum_{t=1}^{T} logp(y_t^*|y_1^*, y_2^*, ..., y_{t-1}^*, x) \tag{10}$$

2 Experiment

2.1 Implementation Details

We use two 128-dimensional LSTMs for the bidirectional encoder and one 256-dimensional LSTM for the decoder. We use a vocabulary of 50k words for both source and target. We don't pretrain the word embeddings, they are learned from scratch with 100 dimensional embedding size. During training and at test time, the article is truncated to 400 tokens and length of the summary is limited to 100 tokens. When pretrain the generator,we use Adagrad with a batch size of 16 and a learning rate 0.15. For adversarial training, we change the learning rate to 0.001 and the batch size to 36. For discriminator, we use filter windows of 3, 4, 5 with 100 feature maps each, dropout rate of 0.5. [Kim14]. At test time, the width of beam search is set to 5.

2.2 Trigram Avoidance

Sequence-to-sequence models often include repetitive phrases. Repetition problem is especially pronounced when generating longer documents and summaries. We adopt the repeated trigram avoidance method proposed by [PXS17]. According to their observation, ground-truth summaries almost never contain the same trigram twice. During beam search, we force the decoder to never output token which would create a trigram that already exists in the previous sequence of the current beam.

Both the coverage model which is used by [SLM17] and trigram avoidance increase Rouge by 3%. However, in experiment, trigram is less training time effective than coverage to reach a steady result and a shorter training time is very important to our model.

2.3 Quotation Weight Alleviation

According to our observation, an appropriate summary contains few quotations from conversations. Thus we lower the weight of the direct of conversation in source text during decoding. This method only increases Rouge by 0.02% which is very trivial. But it is very helpful to boost the readability of summary.

3 Experimental Results

3.1 Rouge Evaluation

Our results are firstly evaluated by the full-length F-1 score of the ROUGE-1, ROUGE-2 and ROUGE-L metrics with the Porter stemmer option. **One thing to note, for DeepRL (Paulus, Xiong, and Socher 2017), we select their RL+ML model** which obtains second highest ROUGE score but produces summaries of highest readability, hence is more relevant for our summarization task. Although their RL model achieves the highest score, it can be detrimental to the model quality.

3.2 Human Evaluation

We perform human evaluation additionally to ensure an increase in human readability and quality. We randomly select 100 test examples from the CNN / Daily Mail dataset. For each example, we show the human evaluator the ground-truth summary as well as summaries generated from different models. The human evaluator is unaware of the source of these summaries and needs to rank (1-5) each summary based on their readability. Rank 1 indicates the lowest level of readability.

3.3 Novel n-grams

We also report the rate of novel n-grams (those that dont appear in the article) in summaries, indicating the degree of abstraction. Compared with the pointer-generator network [SLM17], our model has a better performance in generating novel n-grams and go beyond the simple sentence extraction. Our model only copies 30% the original article. The generated summaries remove unnecessary phrases, and truncate long sentences into meaningful short version.
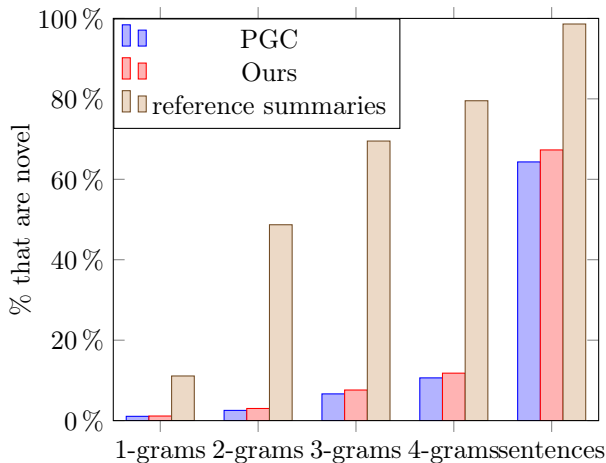
Figure 1: Novel n-grams in summaries

4 Real Summarization Examples

4.1 Example 1

Source Text

a former kansas elementary school teacher was arrested this week and charged with child sex crimes . kourtnie a. sanchez , 25 , was arrested monday on charges including electronic solicitation , unlawful sexual relations , solicitation of unlawful sexual relations and three counts of promoting obscenity to a minor . sanchez - who is believed to be a mother-of-three - was a student teacher at marshall elementary school in eureka during the fall 2014 semester . prosecutors allege her victims were aged between 15 and 17 , and that the alleged incidents occurred over five months in 2014 . however it is unclear how many victims there were . scroll down for video . charged : kourtnie a. sanchez , 25 , was arrested monday on charges including electronic solicitation , unlawful sexual relations , solicitation of unlawful sexual relations and three counts of promoting obscenity to a minor . obscene : sanchez was a student teacher at marshall elementary school in eureka during the fall 2014 semester , when the alleged incidents ' occurred . married : sanchez - whose maiden name is olsen - is married and is believed to have three children . she will next appear in court on may 8 . selfie : the former student teacher is seen here in photos from her facebook account , which has now been deactivated . she left her position in december 2014 . brief : sanchez looked sullen during her first court appearance on tuesday march 31 , 2015 . sanchez - whose maiden name was kourtnie olsen - was never employed by the district because she was a student teacher through a college program . however she still coached some junior high sports for the school for the past two years . she coached junior high girls ' volleyball and basketball , even though she was not a certified teacher , so was not permitted to coach , according to kake.com . the school said she left in december following a ' situation ' , but would not elaborate . sanchez looked sullen during her first court appearance on tuesday . the history log of her snapchat - a cellphone app that allows users to send photos and videos to each other that expire after seven seconds - have been listed as part of the complaint . she made a fast exit as soon the brief hearing had finished . scene : sanchez was a student teacher at marshall elementary school in eureka , kansas , and also coached volleyball and basketball at eureka junior high school (pictured) . eureka resident andrew bratton called the allegations atrocious . ' anybody who would mess with kids , i have a pretty low opinion of , ' bratton . ' if there 's a word lower than low , that would be my opinion of it . ' sanchez is currently out of jail on a $50,000 bond . her preliminary hearing is scheduled for may 8 .

Reference Summary

kourtnie a. sanchez , 25 , was arrested monday in eureka , kansas .

she has been charged with unlawful sexual relations and three counts of promoting obscenity to a minor , among other charges .

alleged incidents ' occurred over five months in 2014 .

sanchez was a student teacher at marshall elementary school and was a sports coach at the

nearby junior high .

ABS Summary

UNK a. sanchez , 25 , was arrested monday on charges including electronic solicitation , unlawful sexual relations , solicitation of unlawful sexual relations , solicitation of unlawful sexual relations , solicitation of unlawful sexual relations .

sanchez , whose maiden name is olsen - is married and is believed to have three children .

sanchez , whose maiden name is olsen - is married and is believed to have three children .

PGC Summary

kourtnie a. sanchez , 25 , was arrested monday on charges including electronic solicitation , unlawful sexual relations , solicitation of unlawful sexual relations and three counts of promoting obscenity to a minor .

sanchez - whose maiden name is olsen - is married and is believed to have three children .
 she will next appear in court on may 8 .

Our Summary

kourtnie a. sanchez , 25 , was arrested monday on charges including electronic solicitation , unlawful sexual relations and three counts of promoting obscenity to a minor .

sanchez was a student teacher at marshall elementary school in eureka during the fall 2014 semester .

prosecutors allege her victims were aged between 15 and 17 , and that the alleged incidents occurred over five months in 2014 .

4.2 Example 2

Source Text

a father-of-three and popular radio host in berkeley , california , was killed in a hit-and-run in the early hours of saturday morning . wesley burton , who worked at kpfa , was driving home from work when a white dodge charger crashed into his silver mercury . the crash occurred near the berkeley-oakland city line and police say the hit-and-run driver fled the on foot . wesley burton , a father-of-three and popular radio host at kpfa in berkeley , california , was killed in a hit-and-run in the early hours of saturday morning as he drove home from work . ' help us regain our peace ' : burton 's wife lucrecia has made an emotional plea for anyone with information about her husband 's killer to come forward . emergency personnel attempted life-saving measures , but burton was pronounced dead at the scene - just blocks from his home . the 36-year-old had worked at kpfa for 20 years , co-hosting the sideshow radio and after hours music shows as well as working as a sound engineer on countless other shows . his wife lucrecia has made a tearful plea for anyone with information to come forward and speak to the police . ' we lost our rock . he was our stability , our strength , ' she told ktvu . ' help us regain our peace . help us get our answers ... our questions answered . ' she also works at the radio station and met burton there 16 years ago . burton had three children - santiago , enrique , and samaya aged between 4 and 9 . burton had three children - santiago , enrique , and samaya aged between 4 and 9 and after growing up without a father his dream had been to raise his own kids . wesley burton , who worked at kpfa , was driving home from work when a white dodge charger crashed into his silver mercury . an emotional lucrecia explained burton did n't have a father growing up and his dream had been to raise his own children . ' making sure that they had a father and so he was very passionate about that , ' she said . a gofundme account has been set up to help burton 's wife pay funeral costs and other family expenses . it has so far raised over $25,000 . police are urging anyone with information to call the traffic investigation unit on (510) 777-8570 . oakland crime stoppers is offering a $10,000 reward for information leading to an arrest .

Reference Summary

wesley burton , a father-of-three and popular radio host at kpfa in berkeley , california , was killed in a hit-and-run on saturday .

he was driving home from work when a white dodge charger crashed into his silver mercury .

wife lucrecia has made an emotional plea for anyone with information about her husband 's killer to come forward . burton had three children aged between 4 and 9 and after growing up

without a father his dream had been to raise his own kids .

ABS Summary

wesley burton , a popular radio host at UNK in berkeley , was killed in a hit-and-run in the early hours of saturday morning .

his wife UNK has made an emotional plea for anyone with information about her husband 's killer to come forward .

PGC Summary

wesley burton , a father-of-three and popular radio host in berkeley , california , was driving home from work when a white dodge charger crashed into his silver mercury .

the crash occurred near the berkeley-oakland city line and police say the hit-and-run driver fled the on foot .

emergency personnel attempted life-saving measures , but burton was pronounced dead at the scene - just blocks from his home .

Our Summary

wesley burton was driving home from work when a white dodge charger crashed into his silver mercury .

crash occurred near the berkeley-oakland city line and police say the hit-and-run driver fled the on foot .

burton was pronounced dead at the scene - just blocks from his home .

4.3   Example 3

Source Text

an 11-year-old chinese schoolboy has become mute after he drank a glass of water which had been laced with perfume and chalk dust as part of a prank . pupil xiao gao , from fujian province in south eastern china , has not said a word in five days after a classmate who had allegedly bullied him gave him the water , reports people 's daily online . but medical experts are at a loss to explain his sudden inability to talk and say that the combination of water , perfume and dust - while nasty - should not have caused that type of damage . pupil xiao gao , aged 11 , (pictured) from fujian in south east china , has not spoken a word in five days after a classmate he ' does not get on with ' gave him the water which had been laced with perfume and chalk dust . gao , who is in year two at haidu number 8 middle school , is now awaiting further medical examination as well as an appointment with a psychologist . school officials told reporters that after gao lost his voice the classmate who gave him the water ' became very afraid ' and told teachers it was meant as a joke and that she did not think he would drink it . she said she only added perfume and chalk dust to the water - and not nail varnish as was rumoured . she said when gao drank the water his classmates told him to spit it out but by then he had already swallowed it . gao , who is a pupil at haidu number 8 middle school , (pictured) is now awaiting further medical examination . gao 's form teacher was informed of his condition and immediately notified his family before taking him to hospital . the unnamed girl has now been excluded from the school . a doctor at fujian provincial hospital told reporters that in all the years he has practised medicine he has never seen a case like this before . qiu bin gaosu said as perfume consists mainly of alcohol , perfume fragrance and methanol , consuming a small amount would not impact the body , although a large amount would irritate the throat and stomach . he said consuming small amounts of chalk dust , which is nontoxic , should also not have this sort of effect . xiao gao 's cousin told reporters that a number of doctors have seen gao and they found no physical abnormalities . gao claims the girl who gave him the water let him drink it and then said ' you will die after drinking this ' . relatives told reporters that gao 's mother was ill and his father worked at a vegetable market . they said the boy was well-behaved and doubted the condition was a hoax - particularly as he had been unable to speak for five days . his form teacher , mr lin , said gao was a little introverted , quiet and very honest and was often bullied by other children in his class . he said : ' whenever a teacher saw any bullying we would chastise the offender and teach them about why their behaviour was not acceptable . ' gao 's classroom - a doctor said that in all the years he has practised he has never seen a case like this before . through written messages gao told reporters that he was thirsty so he drank the water his classmate handed him - but after swallowing a couple of mouthfuls he realised

something was n't right . he claims it was then the girl took away the water and said ' you will die after drinking this . ' gao told reporters that he did n't get along with the classmate who handed him the water . the story has gone viral in china and has been reported on all the main news websites . the people 's daily reported that the local public security bureau has sta rted an investigation into the incident .

Reference Summary

schoolboy xiao gao , 11 , drank water spiked with perfume and chalk dust .
he has not spoken since the prank which went wrong five days ago .
doctors ca n't explain sudden voice loss and think it may be psychological .

ABS Summary

xiao gao , from fujian province , has not said a word in five days .
but medical experts are at a loss to explain his sudden inability to talk .
but medical experts are at a loss to explain his sudden inability to talk .

PGC Summary

pupil xiao gao has not said a word in five days after a classmate he ' does not get on with ' gave him the water which had been laced with perfume and chalk dust .
but medical experts are at a loss to explain his sudden inability to talk and say that the combination of water , perfume and dust - while nasty - should not have caused that type of damage .

Our Summary

xiao gao , 11 , drank a glass of water which had been laced with perfume and chalk dust .
medical experts are at a loss to explain his sudden inability to talk and say that the combination of water , perfume and dust - while nasty - should not have caused that type of damage .
gao , who is in year two at haidu number 8 middle school , is now awaiting further medical examination as well as an appointment with a psychologist .

References

[Kim14]   Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of EMNLP*, pages 1746–751, 2014.

[PXS17]   Romain Paulus, Caiming Xiong, and Richard Socher. A deep reinforced model for abstractive summarization. *arXiv preprint arXiv:1705.04304*, 2017.

[SLM17]   Abigail See, Peter J Liu, and Christopher D Manning. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368*, 2017.

[YZWS16] Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu Seqgan. Sequence generative adversarial nets with policy gradient. arxiv preprint. *arXiv preprint arXiv:1609.05473*, 2(3):5, 2016.