Customized Tweet Generator from Uploaded Images

Jia-Jiun Ku, Li-Pang Huang
Department of Computer Science, University of Virginia, Charlottesville, VA 22904
[jk2mf, lh5jv]@virginia.edu

Abstract

Social media has become more and more popular for celebrities in recent years because, through social media platforms, people can post whatever they want to voice their opinions to the public on the internet without any obstacle. Unlike traditional media, such as TV news, radio, or newspaper, where people play a passive role and cannot control what traditional media reports, social media has the feature that people can transmit the information directly and immediately. However, to keep the fad in the top rank, celebrities need to keep posting to maintain the rank. This becomes hard for not only celebrities but also people since people are not always sitting in front of the laptop to make a post. Although it is more convenient to post through the mobile app, the keyboard is not suitable for editing long paragraphs plus the screen of the mobile phone is relatively small for editing a post. To solve the problem, we introduce our model for generating customized posts from the image. The model is based on the previous post style to generate the new post. After the user uploads the image, it will generate the post as well as the hashtags so that the user only needs to edit the post from the generating one which can save time for not only organizing the post but also editing the text. Code, dataset, and documentation are available at https://github.com/HuangLiPang/Customized-Tweet-Generator.

1. Introduction

Maintaining the trend on the internet is never an easy task. With the growing usage of social media, it facilitates people to run their channel and commerce themselves on the platform. There are a variety of platforms that support functions, such as Twitter, Facebook, Instagram, etc. No matter which platform people are using, it requires people to spend a huge amount of time thinking about the story posting to catch followers' eyes. Furthermore, some stories are timeliness which people have to post them at the right moment. However, it is hard to post a well-organized story without a laptop. Despite the mobile phone is prevailing right now, the editor on the mobile device is not efficient for editing a long story. Besides, the screen is relatively small compared to a laptop which limits people from arranging a good story.

Due to the scenario described above, we introduce a customized post generator to ease the pain of maintaining the fad on the internet. In our work focus on providing a customized post generator from the images uploaded by users. In contrast to the template, the model takes only images as input and generates a customized post with hashtags. This can ease the work for organizing and editing the post. We use the Taiwanese president's official tweets as an example to train the model. The dataset is crawled from Twitter API and we only use the tweet with photos as the dataset. As a result, Our contribution is to introduce a model which can generate customize and personalized post from images uploaded by users and we show that this greatly enhances the quality of results.

In the remaining of this paper, Section 2 introduces related work, especially the technique we applied in our model; Section 3 describes how we prepared the data for our model; Section 4 describes the structure of our model and the detail about training the model; Section 5 compares the differences between our work and assignment 3; Section 6 and Section 7 concludes the result of our model as well as future work.

2. Related Work

There are several works related to generating descriptions from images, such as [2, 5] where the authors used a method built by the closest transitional figures to fetch image descriptions. The other works related to image description where the model has trained all parameters together for one loss function in neural networks are [7, 4, 3]. These previous works all focus on engendering image descriptions for general usage. However, our target is to train a personal model that can generate posts post based on uploaded images with someone's writing style, not just general image captioning.

In addition, our goal is not only related to producing image descriptions but also predicting the sentence from the

Figure 1: Here we show one of tweet from Ing-Wen Tsai's Twitter with four images in our dataset that we are using in our work. Our goal is to produce her style tweet from images she uploads. We collected president Tsai's tweets from Twitter's API, including posts, images, and hashtags

previous inputs. In our work, we applied the method from [6] and also assignment 3. We also refer to Gmail's Smart Compose [1]. Since Smart Compose can predict the reply sentence from the mail, it also has several extensions using linear interpolation between recurrent neural network language and an n-gram model for personalized style responses.

3. Data Collection

Since our goal is to develop a personalized model for generating a post with the user's writing style, the dataset must contain one person's posts and images. Thus, we used a dataset from Twitter because Twitter provides an easy-to-use API for crawling personal posts with images. However, not all tweets include images. Some people only use text for their tweets. Consequently, we decide to pick a celebrity who has his/her tweets with images frequently. After searching celebrities' posts and compared how many images in their tweets, we decided to use tweets from Ing-Wen Tsai who is the first female president of Taiwan. We collected all the tweets from 2010 to today from her twitter account, which amounts to 3200 tweets in total, and picked the tweets with images for our experiment.

4. Model

Generating tweets from an image is practically the same as image captioning, except it needs fine-tuning with one user's tweets to imitate their style of writing. Therefore, in our model, we generated personalized tweets by using a standard encoder-decoder framework. The input is the image uploaded by the user and the output is a stylish post.

Figure 2: Not all of tweets from Ing-Wen Tsai's Twitter includes images in our dataset, so we have to filter out those tweet without images.

In the framework, the encoder consists of ResNet, a pre-trained convolutional neural network (CNN), and a fully connected layer, so the encoder can extract features from the input image. After obtaining features, the fully connected layer of the encoder takes these features as input and produces an encoded vector. In the meantime, we also filtered out urls and special characters in previous posts and convert them to a dictionary so we can use the dictionary and the vector as inputs for the decoder. The decoder is composed of a recurrent neural network (RNN) which can generate sentences from the vector recursively. As a result, the final sentences will be the tweet of the user's style.

To train and evaluate the model, we used Multi30k dataset for first phase training and fine-tuned the model with about 2000 posts from the president Tsai's Twitter that had images as our training and validation data.

5. Experiments

Since we used assignment 3 as our reference and the purposes of assignment 3 and our work are very similar to each other, the only difference being that assignment 3 is just general image captioning while our model is fine-tuned for generating tweets, we will use the assignment 3 as the control group in the experiment to show that our model has a better output of the personalized post. Figure 4 shows the training loss(ghost line) and the smoothed training loss.

6. Results

We constructed our encoder-decoder framework using both Pytorch and Pytorch Lightning. Besides, we use some part of the code from assignment 3 as our starting implementation as well. Initially, we trained the model with the Multi30k dataset for image captioning. Then, we filtered out all tweets downloaded from Twitter with images and loaded them into a dataset. Next, we fine-tuned the model on these tweets. After fine-tuning our model, we used the pretrained 152-layer version ResNet as our base

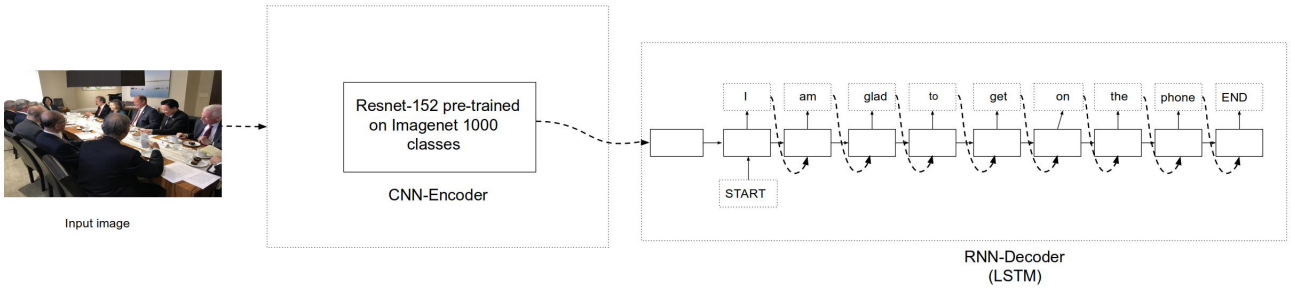

Figure 3: Here we show a flow chart explaining our model components. The input is an image. The output is a post.

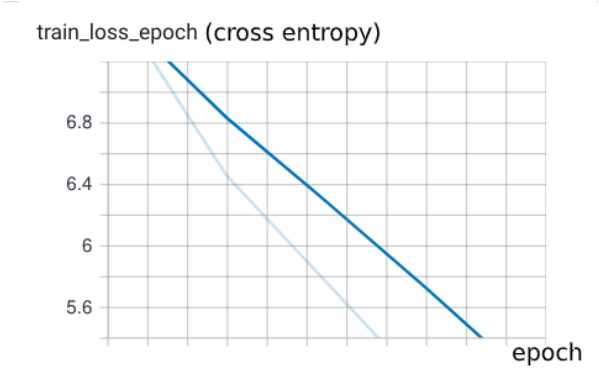

Figure 4: Traning loss and smoothed traning loss

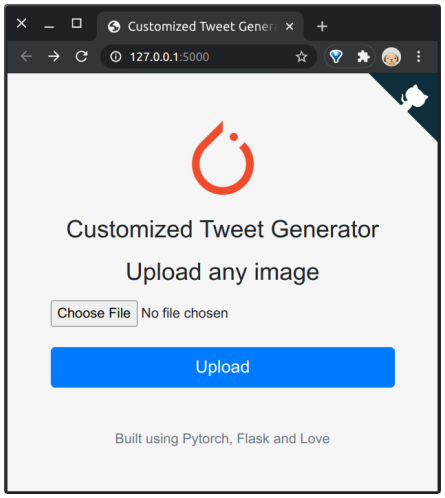

Figure 5: This figure shows the demo application that the user can upload the image and get the customized post.

convolutional neural network and trained our full model using stochastic gradient descent and Adam.

Through the demo application in figure 5, the user can upload the image and get the post. We used 3 images from president Tsai's twitter to show the demo. Figure 6b is from this tweet, figure 7b is from this tweet, and the figure 8 is from her cover photo. As can be seen, the style is very similar to the original post.

7. Future Works

In this work, our goal is to develop a model which take images as input and generate personalized post as output. Since we use the assignment 3 as the initial scratch, the work is limited by the structure of image captioning models. The future directions is described as follow: (i) generating a tweet from multiple images, (ii) creating a model that can produce a complete and structured paragraph instead of only a sentence, and (iii) generating the tweet with hashtags as well will be much more personalized for the user.

References

[1] M. X. Chen, B. N. Lee, G. Bansal, Y. Cao, S. Zhang, J. Lu, J. Tsay, Y. Wang, A. M. Dai, Z. Chen, T. Sohn, and Y. Wu. Gmail smart compose: Real-time assisted writing, 2019.

[2] A. Farhadi, M. Hejrati, M. A. Sadeghi, P. Young, C. Rashtchian, J. Hockenmaier, and D. Forsyth. Every picture tells a story: Generating sentences from images. In *European conference on computer vision*, pages 15–29. Springer Berlin Heidelberg, 2010.

[3] A. Karpathy and L. Fei-Fei. Deep visual-semantic alignments for generating image descriptions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3128–3137, 2015.

[4] J. Mao, W. Xu, Y. Yang, J. Wang, Z. Huang, and A. Yuille. Deep captioning with multimodal recurrent neural networks (m-rnn). *arXiv preprint arXiv:1412.6632*, 2014.

[5] V. Ordonez, G. Kulkarni, and T. L. Berg. Im2text: Describing images using 1 million captioned photographs. In *Advances*

(a) The original post

(b) The customized post

Figure 6: Compared the original post with the customized post.

in Neural Information Processing Systems, pages 1143–1151, 2011.

[6] I. Sutskever, J. Martens, and G. E. Hinton. Generating text with recurrent neural networks. In *ICML*, 2011.

[7] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. Show and tell: A neural image caption generator. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

(a) The original post

(b) The customized post

Figure 7: Compared the original post with the customized post.

Figure 8: Post generated from the cover photo