Image Captioning for Social Network Posts

Joanne Jiang, Zhongyao Ma, Yangyiying Pu, Chengwu Shen, Huidi Wang

Problem Statement

1 Business Context

Social media has seen an unprecedented rise of users in the past few years. According to sources[1], Facebook, Instagram and Pinterest combined to report over 3.5 billion daily active users. It implies a huge business opportunity and our idea was indeed generated from one of the most fundamental actions on social media - posting. When users post contents, including both pictures and pure texts, they often include hashtags in the middle of their posts. Hashtags are created to help users discover similar posts and in some cases assist users with organizing posts. Posts with hashtags generally have a higher viewability since the posts will be viewed by whoever clicks on the hashtag. As a result, users who want to gain more followers or increase the number of "likes" always add a list of hashtags so it will attract more attention.

2 Project Goal

Our project aims to provide a solution to users who want to become social media influencers or simply want to share with specific groups with an easy-to-use product that can generate a list of hashtags based on the picture in less than 1 seconds. Users will no longer need to manually type out the hashtags. Instead, all they need to do is to upload the picture they would like to post to our product and select from the generated hashtags. The goal of our project is to not only save time for users by automatically generating hashtags but also target the most related audience by providing the most relevant hashtags.

Technical Solution

1 Hashtag Selection

We pick more than 100 popular categories for Instagram hashtags, including nature, food, sports, etc., in order to capture a broad range of frequently used hashtags. Also, posts with these selected hashtags tend to include new hashtags as well. In this way, our models are capable of learning from and hopefully predict a diverse collection of hashtags for pictures.

2 Data Collection and Preprocessing

In order to collect posts from Instagram, our team finds an Apify API[2], which enables us to pull targeted posts from Instagram through a Python script. Our script automatically loads a target number of posts for each hashtag, separately saving the pictures and writing the hashtags to a dictionary and connecting them through a unique id. We scrape more than 4000 posts in total and do image pre-processing. Next, we select the most popular 500 hashtags as the pool of candidates, to make sure the final prediction is comprehensive. Then, we segment the hashtags to regular phrases using the

---

[1] Stout, Dustin W., et al. "Social Media Statistics: Top Social Networks by Popularity." *Dustin Stout*, 16 July 2019, dustinstout.com/social-media-statistics/.

[2] Web Scraping, Data Extraction and Automation. https://apify.com

wordsegment[3] package. We filter out unpopular hashtags and maintain around 3000 posts with images and hashtags.

3 Modeling

Our model integrates three techniques: image classification, sentence encoding and nearest neighbor search. First, we apply ResNet-50[4], a pre-trained CNN model on ImageNet, for the standard image classification task (0.02s / image). Next, the popular hashtags from our candidate pool are encoded as 512-dimensional vectors using a sentence embedding technique, and we construct a K-Dimensional binary search tree on the vectors, where each tree node represents an axis-aligned hyperrectangle. Then, each classification is also encoded as a 512-dimensional vector (0.001s / image), and its top N nearest neighbors in the K-D tree are the hashtag predictions.

The prediction is accurate when our predictions intersect with the true hashtag set. The overall accuracy is 43% when we keep the top 30 nearest hashtags. Since there is a limited coverage of our model, we set a threshold on the lowest Euclidean distance allowed for the neighbor search. In this way, our results can be more reliable. In order to improve the prediction accuracy, we also consider a weighted Euclidean distance and apply a logistic regression model to estimate the weights. We use one-hot encoding to transform multi-class classification to binary classification. Logistic regression models have better performance and accuracy but longer runtime. Given that runtime is critical for customer experience, we choose to use K-D tree as our final model.

4 User Interface

We build an User Interface with Python and Flask to connect the backend model and the frontend web application. We simulate the process of users uploading pictures from local and model predicting a list of hashtags.

Project Impact

With our product, users are able to add hashtags to their posts more easily. No matter where the user wants to post a new picture, Facebook, Instagram, or Pinterest, they can use our product to facilitate the process. Since our product can help users to post with hashtags more efficiently, it is likely that users see their posts getting more likes and attracting more followers. This is because that the system generated hashtags are highly relevant to the post and therefore, people who are interested in the specific hashtag will have a higher chance to be interested in the post as well. Our product essentially helps users to increase their influence on social media.

Moreover, when more users start to use our product, we can consider monetizing the traffic. The easiest way would be to add advertisements on the website. Advertising technology is one of the fastest growing industries and it also has tremendous potential. If we can build a well-designed ad infrastructure and adjust the user interface accordingly, it should yield really exciting returns.

[3] Python Word Segmentation. http://www.grantjenks.com/docs/wordsegment/
[4] ResNet-50. https://arxiv.org/abs/1512.03385