Multimodal Image inspired hashtag generator

IRE Team 13

1 Introduction

The problem is to generate the hashtags for a given multi-modal post that could be a text or an image or set of images/video. The generated hashtags should be relevant to the existing hashtags that may present along with the post. For example, in the Figure 1 below, we are given an image, text content and certain hashtags (#dog, #birthday) as the inputs [1]. The goal of this project is to utilize this data and suggest few other hashtags like #celebration, #party or #puppy for the same post. Utilizing the multi-modal nature of a post is essential to suggest relevant hashtags as predicting the hashtag without visual information is a difficult task. In the same Figure 1, it is difficult to predict hashtags such as #dog, given only the text content.

Figure 1: Predicting the correct hashtag: #dog is difficult without the image [7]

The project involves three major components:

1. **Identifying prior art**: Some amount of work has already been done on utilizing images and multi-modal content for hashtag generation. [4] and [7] utilize CNN and attention to make predictions for hashtags. [3] on the other hand uses zero-short learning for the same. Most of these approaches evaluate on different benchmark datasets like YFCC100M [6] and HARRISON dataset [5] or their own custom datasets. Understanding these approaches and incorporating their ideas in our project is the first task.

2. **Data Collection**: Identify a specific topic and collect data for the same by scraping Instagram. The data collection process would consist of 1) generating or creating a list of seed hashtags, 2) Segmenting it to generate more hashtags (Using Viterbi algorithm) and 3) Use this larger set to scrape more hashtags from Instagram. This process will be repeated until there is enough data for training and testing.

[1] There might be cases when any of these inputs is missing. The system should be able to take care of this as well

3. **Modeling**: The problem would most likely require us to utilize deep learning methods like CNNs and LSTMs. The task is slightly similar to image captioning tasks that mostly utilize pretrained Imagenet models and an LSTM for generating the captions. We would also be using methods similar to the ones presented in prior works that combine visual (CNN) and textual (LSTM and Word2Vec) information via attention modules. Evaluation would be done on our dataset and if time permits then on one of the benchmark datasets as well.

2 Literature Review

Some of the prior work done in this domain are listed below:

- Hashtag Recommendation for Multimodal Microblog Using Co-Attention Network [7]: They utilize multi-modal data present in Tweets to generate hashtags using a co-attention mechanism.
- Predicting Instagram tags with and without data [3]: They use zero-shot learning to predict the hashtags for a given image.
- A few other notable works that utilize multi-modal data are: [1], [4] and [2].
- Some of the datasets that could be possible used for training/evaluation are: YFCC100M [6] and HARRISON dataset [5]

The goal of literature review is to understand the prior work done in this field and utilize the method that would work best for our specific case and topics.

3 Implementation

Some of the basic aims/goals of this project are given below.

1. Collect a large dataset of Instagram posts and hashtags for a specific domain. Create a data collection pipeline that can be utilized for collecting data for other topics as well given different seed list of hashtags.
2. Implement few different models (Based on prior art) for hashtag generation using multimodal data. Evaluate them on our dataset and compare with simple baselines like image only / text only classifiers or non machine learning methods like synonym finder.
3. If time permits, evaluate the same models on a benchmark dataset like YFCC100M [6] and HARRISON dataset [5] and extend the same work for videos / short clips as well.

References

[1] Emily Denton, Jason Weston, Manohar Paluri, Lubomir Bourdev, and Rob Fergus. User conditional hashtag prediction for images. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1731–1740. ACM, 2015.

[2] Shivam Gaur. Generation of a short narrative caption for an image using the suggested hashtag. In *2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW)*, pages 331–337. IEEE, 2019.

[3] Shreyash Pandey and Abhijeet Phatak. Cs229 project report.

[4] Cesc Chunseong Park, Byeongchang Kim, and Gunhee Kim. Towards personalized image captioning via multimodal memory networks. *IEEE transactions on pattern analysis and machine intelligence*, 41(4):999–1012, 2018.

[5] Minseok Park, Hanxiang Li, and Junmo Kim. Harrison: A benchmark on hashtag recommendation for real-world images in social networks, 2016.

[6] Bart Thomee, David A Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. Yfcc100m: The new data in multimedia research. *arXiv preprint arXiv:1503.01817*, 2015.

[7] Qi Zhang, Jiawen Wang, Haoran Huang, Xuanjing Huang, and Yeyun Gong. Hashtag recommendation for multimodal microblog using co-attention network. In *IJCAI*, pages 3420–3426, 2017.