Automatic Hashtag Generation for Social Media Posts Using Neural Text Generation Models

Satwik Goyal, Shubham Shetty, and **Mahidhar Kommineni**
College of Information and Computer Sciences
University of Massachusetts Amherst

Abstract

In 2021, about 82% of adults in the United States used social media platforms. Social Media is used for various purposes including information sharing, marketing, and social interaction using posts. Each post may contain hashtags that help reach out to a wider audience by making it easier to search through the content. Hashtags create instant communities with people of similar interests. Analysing social media posts can help in several tasks such as targeted marketing, customer sentiment analysis, and market research. We propose a framework to generate hashtags for a given social media post using massive pre-trained language models (such as GPT-3 and BART) in order to augment the post, which can then be used for downstream analytical tasks.

1 Introduction

1.1 Motivation

Social media has increasingly become an essential mode of communication as a platform for self-expression and dissemination of information. In 2021, about 82% of adults in the United States used social media platforms (Statista, 2022). Social media posts cover a wide range of topics ranging from personal, political, social topics, to promotional content. With the explosion in the amount of posts being created and shared over various websites such as Twitter, Instagram, and Facebook, social media becomes a rich source of data which can be harnessed for multiple analytical use cases such as targeted marketing, sentiment analysis, and market research.

However, the inherent structure of a social media post provides several challenges. We define a social media post as a post shared on a social media platform (such as Twitter) which can consist of either text, images, and/or videos. Analyzing a social media post in isolation may not be able to provide meaningful results as single posts may have limited data. Additionally a post may depend on some contextual information such as comment chains, community posts, and replies.

To approach this issue of missing contextual information, several approaches have used different pooling methods. Community pooling (Albanese and Feuerstein, 2022) focuses on defining a community to aggregate tweets for topic modelling. The authors also explore other methods such as author pooling and conversation pooling. Hashtags can also provide some context in understanding the topics a post is dealing with. Hashtag pooling (Mehrotra et al., 2013) has been seen to outperform some baseline schemes for the downstream task of topic labelling. Therefore, our model may be used to label posts with hashtags which may be further used for hashtag pooling.

Hashtags are very often used to reach out to a wider targeted audience, however, some hashtags based on their popularity are more optimal than others. Therefore, we propose a model that will help generate optimal hashtags for social media posts to reach to a wider targeted audience.

1.2 Task Description

Users may use #hashtags to associate their posts with certain topics and drive engagement. Hashtags can further be used for downstream social media analytical tasks such as search, targeted marketing, and sentiment analysis. However, many social-media posts do not have any tagged hashtags (Wang et al., 2011). Hence, we propose a framework for automatic hashtag generation & augmentation for social media posts without any associated hashtags using massive pre-trained language models (such as GPT-3) in a few-shot or zero-shot setting. We hypothesize that the social media post augmented with generated hashtags would contain more context and hence perform better in downstream tasks.

2 Background & Related Work

Our work aims at generating hashtags from text, images, and video based social-media posts using neural text generation.

2.1 Automatic Hashtag Generation

Automatic hashtag annotation has been historically seen as a key-phrase extraction and classification task. Some earlier proposed methods for hashtag annotation include predicting hashtags from a pre-defined list of words (the candidate list) (Zhang et al., 2017), or using topic models to generate hashtags (Wu et al., 2016). The issue with these approaches is that they only generate single phrases as hashtags, whereas actual hashtags may be longer phrases which actually reflect the topic being referenced to in the post.

2.2 Neural Text Generation

Another issue with treating hashtag generation as a key-phrase extraction task is that hashtags might not appear in either the target posts or the given candidate list. Recent approaches to hashtag generation explore neural text generation models in order to generate phrase level hashtags beyond the provided post or candidate list. Current state-of-the-art approach (Wang et al., 2019) proposes a sequence-to-sequence text generation framework to generate hashtags based on social media conversations as a supplementary data source. However, it is unrealistic to expect conversation chains to contain required information, and has a high cost of annotation.

Further work builds on the sequence-to-sequence model proposed by (Wang et al., 2019). One such approach uses a semantic-fragmentation-based selection mechanism in transformer architecture to generate hashtags over large datasets (Mao et al., 2021). Another proposed framework utilises a retrieval-augmented sequence-to-sequence architecture (Zheng et al., 2021) in order to generate hashtags with current and emerging events reflected.

An issue with these approaches are that they require either external information or large amounts of data to generate meaningful results.

2.3 Massive Pre-Trained Language Models for Text Generation

Our proposed approach will try to address these issues by utilizing massive pre-trained language models (PLMs) with autoencoders for text generation, such as BART (Lewis et al., 2020) and GPT-3 (Brown et al., 2020), to generate hash-tags. We will also experiment with zero-shot approaches (Kumar et al., 2019) towards hashtag generation using these massive models.

BART has recently been used for topic modeling as BART-TL (Popa and Rebedea, 2021). BART-TL generates accurate representations of the most important topic terms and candidate labels by fine-tuning on a large number of potential labels generated by state-of-the-art models for topic labeling. A similar approach can be adopted for the task of hashtag generation.

GPT-3 (Brown et al., 2020) has shown near state-of-the-art performance in generative tasks such as summarization and news article generation in a few-shot or zero-shot setting (Tehranipour, 2020). We intend to perform similar experiments using zero-shot and few-shot approach, and compare hashtag generation performance against that of our other models.

2.4 Text Augmentation

Text augmentation involves augmenting the existing linguistic space to improve the performance of down-stream tasks without affecting the generalizability. There are two types of popular text augmentation techniques namely shallow and deep augmentation. In shallow augmentation technique, small noises in the form of words and phrases are added to the linguistic space to improve the performance of downstream tasks. EDA (Wei and Zou, 2019) employs shallow augmentation technique to improve the performance of text classification task. On the other hand, deep augmentation technique uses large scale pre-trained language models such as BERT (Devlin et al., 2018) and BART (Lewis et al., 2020) to augment the data in a more diverse manner. The deep augmentation technique using BART has been successful in augmenting dataset where the data labels are scarce (Kumar et al., 2020).

2.5 GPT-3 and Few Shot Learning

Large scale language models like GPT-3 and BERT can be easily controlled using natural text making them a good fit for few shot learning. There has been significant development and interest of using prompt-based mechanisms to help large scale pre-trained models to understand the task in hand. (Yoo et al., 2021) proposed a method of first selecting a subset of examples from the dataset, constructing a prompt using the selected set of examples including the meta-information about the dataset and finally extracting augmentation from the language model generation. In our proposed solution, we follow a similar approach of using prompt based NLP to help GPT-3 understand the hash tag generation task in a few shot setting by passing a list of train data points with the target hashtags and a list of testing data points without hashtags. The model is able to generate hashtags for the test data points by understanding the hash tag generation from train data points.

2.6 Hashtag Generation for Images & Videos

As far as we know, extensive work on hashtag generation from images and videos has not been done. HARRISON (Park et al., 2016) is a dataset provided as a baseline to test other models. hashtag generation from images. They consider the hashtag recommendation task as a multi-label classification problem and use CNN architectures to recommend hashtags.

The Graph Convolution Network based Personalized Hashtag Recommendation (GCN-PHR) model (Wei et al., 2019) was proposed for personalized hashtag recommendations on micro-videos. They use advances in GCN technology to graph complicated user interactions with micro-videos on social media, and generate hashtags for certain parts of the video.

Existing methods use CNN based approaches and may not provide best results. In our work, we will experiment with using image captioning methods first in order to generate captions, on which our text based approach can be extended.

3 Proposed Solution

We propose a framework to generate hashtags from a given social media post using a massive pre-trained language model and augmenting the post for downstream tasks. Figure 1 shows a high-level view of the framework. The framework will consist of two main modules -

1. Hashtag augmentation module, and
2. Downstream task training module.

The hashtag augmentation module will accept a social media post as an input and generate hashtags for that post using a massive pre-trained language model (GPT-3, BART, etc.) in a zero-shot

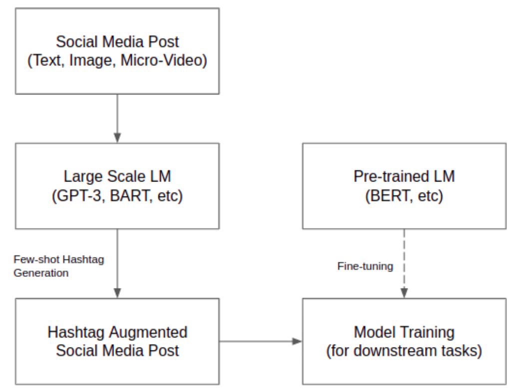

Figure 1: High level diagram for hashtag augmentation model

or few-shot setting. The generated hashtags will be used to augment the social media post in order to provide contextual information. Hashtags can be augmented either as text or as meta-data (in case of images or micro-videos).

Downstream task training module will accept augmented data and fine-tune models for the specific downstream task. Downstream task can be a classification task (such as controversy detection or stance detection), topic modelling, sentiment analysis etc.

Additionally, the hashtag augmentation module will include additional sub-modules to deal with different kinds of social media posts. As per our definition of a social media post, a post can contain either text, images, and/or micro-videos. Hashtags can be generated for text posts directly using GPT-3. For images we propose using a combination of image captioning and optical character recognition (OCR) to generate text for an image and then return hashtags via GPT-3. For videos, we propose to sample a few frames from the video and apply same method as used for images.

Evaluation of our model will combine evaluations of both modules separately. For the hashtag generation module, we will compare against existing hashtag generation and recommendation models using basic metrics such as precision, recall, f1-score, as well as hit-ratio. For the downstream training module we will compare against the baselines for the specific downstream task.

For downstream tasks, we will consider following models to evaluate our framework -

1. Sentiment analysis (Sun et al., 2019),
2. Stance detection (Glandt et al., 2021), and

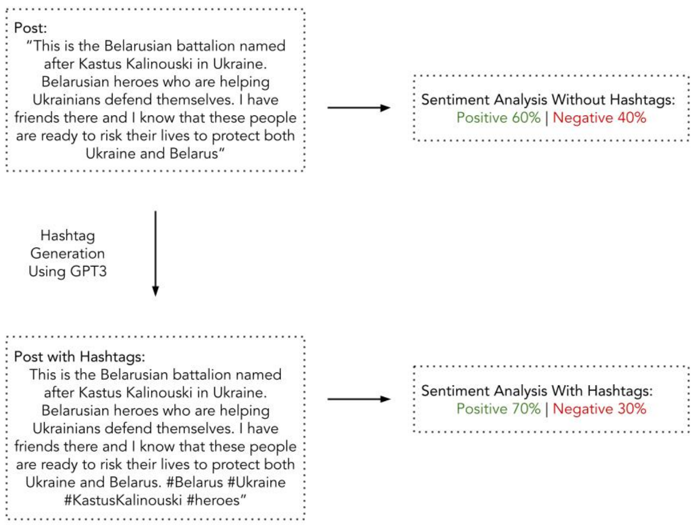

Figure 2: An illustration of the proposed framework. Addition of hashtags to original post should result in better results in downstream tasks.

3. Toxic comment classification (Caselli et al., 2021)

4 Experimental Results

4.1 Datasets

For the midpoint, we compiled a list of 15 popular hashtags on Twitter. Then for each hashtag we found 4 Tweets. For each of those Tweets we extracted the text and all hashtags in that Tweet. Further, we provided our industry mentors with 50 most popular hashtags on Twitter. They are currently creating larger dataset (consisting of only text documents). Currently, our dataset only includes text and hashtags, however, the new dataset will also consist of metadata such as date and location of the user. Additionally, we plan to create a small data set for images and micro-videos possibly from Instagram and/or Tiktok posts.

4.2 Evaluation: The Hit-Ratio Metric

To evaluate the performance of the models we are using the hit-ratio metric. For each tweet in the testing dataset there are a few hashtags. We then generate a few hashtags for each tweet. The length of all hashtags combined is limited to 60 characters. On an average the dataset contains 2.8 hashtags per tweet. On an average GPT3 generates 3.8 hashtags per tweet. If any hashtags generated is in the test set for the tweet it is given a Hit-Rate score (accuracy) of 1, 0 otherwise. i.e.

$$\text{Accuracy} = \begin{cases} 1 \text{ if } [\text{Test Hashtags} \cap \text{Generated Hashtags} \neq \phi] \\ 0, \text{ otherwise} \end{cases}$$

$$\text{Precision} = \frac{|\text{Test Hashtags} \cap \text{Generated Hashtags} \neq \phi|}{|\text{Generated Hashtags}|}$$

$$\text{Recall} = \frac{|\text{Test Hashtags} \cap \text{Generated Hashtags} \neq \phi|}{|\text{Test Hashtags}|}$$

$$\text{F-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4.3 Hashtag Generation for Text Documents

For text documents (i.e. tweets containing only text) we performed zero-shot and few-shot learning experiments using GPT3 to generate hashtags. The results are summarized in table 1.

As we see in table 2 GPT3 was able to generate #giveaway for Tweet 1 after few-shot learning, however, since #NFTGivaway was not seen during training it was not able to generate that. In table 2 for Tweet 2, the hashtag we aimed to generate was #TravelTuesday, however, since no metadata (i.e. date or day) was passed in, GPT3 was not able to generate that hashtag unless the text explicitly mentions it is a Tuesday. We believe this issue will be resolved with the new dataset that will consist of metadata as well.

4.4 Hashtag Generation for Image Documents

For image documents we first use the ViT + GPT2 image captioning model as well as EasyOCR to convert images to text. We then input the text into

	Accuracy	*Precision*	*Recall*	*F-Score*
Zero Shot	0.65	0.27	0.27	0.34
Few Shot	0.80	0.60	0.29	0.39

Table 1: Evaluation of GPT-3 Hashtag Generation

GPT3 to generate hashtags. We are yet to generate a dataset for images, so we have only been able to approach the problem with zero-shot learning. As seen in the table, the model was able to generate relevant hashtags. However, the model performs poorly on some images where there is no text and the image captioning model performs poorly as well. However, this issue can be overcome either by few-shot learning approach or by combining images with text data.

Figure 3: Input Image

5 Next Steps

Currently, we have experimented the hashtag generation module of our framework on a small dataset of 200 tweets manually created from twitter by searching through the list of most popular hastags(Chris Sabanty, 2019). Our next steps mostly revolve around expanding dataset and evaluating performance of hashtag augmented dataset on downstream tasks.

Broadly, our next steps include -

1. We are working with our industry mentor to create a larger dataset, which includes more data-points as well as metadata such as authors, date, replies, etc. We are also building datasets for posts containing images and micro-videos.

2. We experimented for few data-points by augmenting the textual content with image captions generated using image-captioning models like 'ViT+GPT2' and observed that the hashtags generated were more relevant with the augmented data. We will work on augmenting the textual content with image captions generated using image captioning model and video captions for micro-videos using video captioning model for each post before passing to the model for hashtag generation. For image captioning, along with 'Vit+GPT2' model we will also experiment with Visual BERT (Li et al., 2019) which performs state of the art results on Flickr30K dataset.

3. Performance on downstream tasks using hashtag-augmented data and comparison against baselines. We have chosen to compare our framework performance against baselines for sentiment analysis, hate speech detection, and stance detection.

4. Additional experimentation with BART and GPT-2, as well as different OCR and image captioning models.

We will evaluate the performance of downstream tasks like topic modelling/labelling using hashtag augmented data vs the data which do not contain hashtags.

6 Acknowledgements

We would like to thank our industry mentors Keen Sung and Andrew Curran from AuCoDe Inc., and our Ph.D. mentor Subendhu Rongali for helping us with this project.

References

Federico Albanese and Esteban Feuerstein. 2022. Improved topic modeling in twitter through community pooling.

Tweet	*GT Hashtags*	*Zero-Shot*	*Few-Shot*
Who wants a free NFT?	#NFTGivaway #NFT #Competition	#NFT #free	#NFT #giveaway
Where'd you go this summer? [emoticons]	#TravelTuesday	#summer #travel #fun	#travel #vacation #summer

Table 2: Hashtag Generation for Input Text

Model	*Generated Output*
ViT + GPT2 Image Captioning	an advertisement for a television show
EasyOCR Text	on this dah 14 years ago warriors nation steph currv itroduced hiiself to the iorld hnd scored 40 poiits aghiist hohzhca clutchpoints valdon:
Zero-Shot Hashtag Generation	#Warriors #StephCurry

Table 3: Hashtag Generation for Input Image

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners.

Tommaso Caselli, Valerio Basile, Jelena Mitrović, and Michael Granitzer. 2021. HateBERT: Retraining BERT for abusive language detection in English. In *Proceedings of the 5th Workshop on Online Abuse and Harms (WOAH 2021)*, pages 17–25, Online. Association for Computational Linguistics.

Chris Sabanty. 2019. Popular and most liked hashtags on twitter. [Online; accessed April 27, 2013].

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding.

Kyle Glandt, Sarthak Khanal, Yingjie Li, Doina Caragea, and Cornelia Caragea. 2021. Stance detection in COVID-19 tweets. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1596–1611, Online. Association for Computational Linguistics.

Abhay Kumar, Nishant Jain, Suraj Tripathi, and Chirag Singh. 2019. From fully supervised to zero shot settings for twitter hashtag recommendation.

Varun Kumar, Ashutosh Choudhary, and Eunah Cho. 2020. Data augmentation using pre-trained transformer models.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Liunian Harold Li, Mark Yatskar, Da Yin, Cho-Jui Hsieh, and Kai-Wei Chang. 2019. Visualbert: A simple and performant baseline for vision and language.

Qianren Mao, Xi Li, Hao Peng, Bang Liu, Shu Guo, Jianxin Li, Lihong Wang, and Philip S. Yu. 2021. Attend and select: A segment attention based selection mechanism for microblog hashtag generation.

Rishabh Mehrotra, Scott Sanner, Wray Buntine, and Lexing Xie. 2013. Improving lda topic models for microblogs via tweet pooling and automatic labeling. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '13, page 889–892, New York, NY, USA. Association for Computing Machinery.

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

Cristian Popa and Traian Rebedea. 2021. BART-TL: Weakly-supervised topic label generation. In *Proceedings of the 16th Conference of the European*

Chapter of the Association for Computational Linguistics: Main Volume, pages 1418–1425, Online. Association for Computational Linguistics.

Statista. 2022. Share of u.s. population who use social media 2008-2021.

Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. 2019. How to fine-tune bert for text classification? In *Chinese Computational Linguistics*, pages 194–206, Cham. Springer International Publishing.

Soheil Tehranipour. 2020. Openai gpt-3: Language models are few-shot learners.

Xiaolong Wang, Furu Wei, Xiaohua Liu, Ming Zhou, and Ming Zhang. 2011. Topic sentiment analysis in twitter: A graph-based hashtag sentiment classification approach. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, CIKM '11, page 1031–1040, New York, NY, USA. Association for Computing Machinery.

Yue Wang, Jing Li, Irwin King, Michael R. Lyu, and Shuming Shi. 2019. Microblog hashtag generation via encoding conversation contexts.

Jason Wei and Kai Zou. 2019. Eda: Easy data augmentation techniques for boosting performance on text classification tasks.

Yinwei Wei, Zhiyong Cheng, Xuzheng Yu, Zhou Zhao, Lei Zhu, and Liqiang Nie. 2019. Personalized hashtag recommendation for micro-videos.

Yong Wu, Yuan Yao, Feng Xu, Hanghang Tong, and Jian Lu. 2016. Tag2word: Using tags to generate words for content based tag recommendation. In *CIKM 2016 - Proceedings of the 2016 ACM Conference on Information and Knowledge Management*, International Conference on Information and Knowledge Management, Proceedings, pages 2287–2292. Association for Computing Machinery.

Kang Min Yoo, Dongju Park, Jaewook Kang, Sang-Woo Lee, and Woomyeong Park. 2021. Gpt3mix: Leveraging large-scale language models for text augmentation.

Qi Zhang, Jiawen Wang, Haoran Huang, Xuanjing Huang, and Yeyun Gong. 2017. Hashtag recommendation for multimodal microblog using co-attention network. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, pages 3420–3426.

Xiuwen Zheng, Dheeraj Mekala, Amarnath Gupta, and Jingbo Shang. 2021. News meets microblog: Hashtag annotation via retriever-generator.