Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022), pages 7191–7198
Marseille, 20-25 June 2022

Leveraging Hashtag Networks for Multimodal Popularity Prediction of Instagram Posts

Yu-Yun Liao
Graduate Institute of Linguistics
National Taiwan University
r09142002@ntu.edu.tw

Abstract

With the increasing commercial and social importance of Instagram in recent years, more researchers begin to take multimodal approaches to predict popular content on Instagram. However, existing popularity prediction approaches often reduce hashtags to simple features such as hashtag length or number of hashtags in a post, ignoring the structural and textual information that entangles between hashtags. In this paper, we propose a multimodal framework using post captions, image, hashtag network, and topic model to predict popular influencer posts in Taiwan. Specifically, the hashtag network is constructed as a homogenous graph using the co-occurrence relationship between hashtags, and we extract its structural information with GraphSAGE and semantic information with BERTopic. Finally, the prediction process is defined as a binary classification task (popular/unpopular) using neural networks. Our results show that the proposed framework incorporating hashtag network outperforms all baselines and unimodal models, while information captured from the hashtag network and topic model appears to be complementary.

Keywords: Hashtag Network, Instagram Popularity Prediction, Multimodal Deep Learning

1. Introduction

Over the past few decades, social media have become one of the most powerful and effective devices for spreading information. Companies and influencers around the world share millions of ideas and launch advertisement campaigns everyday through social media. In particular, Instagram, a photo and video sharing platform launched in 2010, has achieved widespread popularity among younger generations and Internet influencers these days. As a result, Instagram have gradually gained tremendous commercial and social importance as it unveils novel business strategies and marketing possibilities for not only companies and organizations, but also influencers, politicians, entertainers, and even country governments.

Due to the increasing significance of Instagram, multimodal prediction of popular content on Instagram has become of great interest to researchers in the field of digital marketing or social media analytics. On Instagram, users can utilize a digital arsenal of photos, videos, filters, captions, hashtags, or stories to create appealing content for their audience. Each of the modalities in an Instagram post often require different techniques to represent; for example, natural language processing for captions, computer vision for images, both of which involve knowledge in data analysis, neural networks, or deep learning. However, among different modalities, hashtags are often ignored or over-simplified as features such as hashtag length, number of hashtags in a post or hashtag frequency. Some even consider hashtags as part of the post caption, failing to represent hashtags as individual and informative features.

With the above being said, the present study proposes a multimodal framework considering post caption, image, and incorporating hashtag networks for predicting popular Taiwanese influencer posts on Instagram. Unlike existing Instagram multimodal frameworks, this study considers the co-occurrence relationship between all hashtags in the corpus and constructs hashtags as a homogeneous graph. The hashtag embeddings are then generated by combining structural embeddings extracted through GraphSAGE (Hamilton et al., 2017) and topic embeddings extracted using BERTopic. In addition to hashtag representations, contextualized representations of post captions are learned using sentence-transformers (Reimers et al., 2019), and image features are generated using the pretrained Inception-V3 model. Finally, following Carta et al. (2020)'s method, this study calculates the engagement moving average (EMA) of Instagram posts as the basis for determining whether a post is popular or unpopular. In sum, the multimodal framework encodes text, image, and hashtags as a set of modality features that are eventually used as inputs to a binary classifier.

It is the goal of this study to provide insights into not only social media marketing and influencer behaviors, but also leveraging a new form of hashtag feature – network graph as a part of the multimodal modeling of fragmented information on Instagram. Hopefully, the present study will illuminate new possibilities for representing hashtag connections as an indicator to the popularity of Instagram posts. In addition, as existing research in this field are mostly conducted on English posts, the present study not only provides precious insights into popularity prediction of Chinese Instagram posts, but it also explores Instagram content in different languages or the different posting patterns between Chinese and English Instagram users. Finally, the applications of the present study may extend to general social media analytics for companies, global brands, politicians, and influencers to form new marketing strategies.

2. Related Work

In this section, previous work that relate to the present study will be reviewed. To be more specific, this section presents studies under two subtopics: popularity prediction in social media and research in hashtags.

2.1 Popularity Prediction in Social Media

In recent years, popularity prediction of social media posts has gradually become a hot topic, and researchers in this field have proposed diverse frameworks concerning different datasets, model structures, feature representations, and defining the task as either a regression or classification problem. The mainstream method for social media popularity prediction is to incorporate deep learning models and combine features extracted from different modalities.

Zhang et al. (2018) train a dual-attention multimodal model using image caption pairs and topic features extracted from an LDA model on Instagram. De et al. (2017) implement stacked autoencoders connected to an MLP layer to predict popular posts from GQ India's Instagram account. On the other hand, Mazloom et al. (2016), gather posts from Flickr and trains SVR (support vector regression) models using textual information and CNNs to predict popularity scores, hence solving a regression task.

However, some research is able to achieve desirable results despite using features that are extracted without involving neural networks or deep learning. Huang et al. (2018) exploits post-related features such as text length, timestamp, geolocation and user-related features such as followers and following number as inputs to several regression models. Carta et al. (2020) collected about 100,000 posts from over 2,500 users on Instagram, utilizing several features such as post metadata, text length, hashtag counts, and average likes number to represent each post. It is worth noticing that they measure popularity using LMA (likes moving average). Unlike previous work considering raw likes or followers as the popularity metric, LMA compares the likes number of a post with the average likes number of its previous *k* posts to determine a popularity class (popular or unpopular), taking into consider the recent engagement performance of a post. The present study will also follow this method for popular post prediction, which will be explained in detail in section 4.2.2.

2.2 Hashtag Representation

As mentioned earlier, this study introduces hashtag networks as a novel method for including hashtags as a part of the multimodal post representation. Therefore, this section provides an overview of the different aspects of hashtags, including hashtags in social media linguistics to hashtags in networks and graph theory.

2.2.1 Linguistic Functions of Hashtags

Using a corpus-based approach, Zappavigna (2015) proposed that hashtags serve several different functions in social media, with hashtags as topic markers as the dominant function. To be more specific, hashtags encapsulates the main topic of a post, helping users to effectively highlight and emphasize central ideas. Scott (2015) also mentioned that hashtags are potent search tools. Following Messina (2007)'s proposal, hashtags are originally designed to track and tag content on Twitter, linking posts with similar topics and discourse environments together. In addition, Scott also claimed that hashtags are able to serve as highlighting devices of constituents in a post, enhancing readers' attention towards a specific chunk of text and increasing its saliency in readers' cognitive environments.

2.2.2 Network Representations and Hashtags

Unlike the linguistic aspects of hashtags, few research has touched upon learning hashtag representations in the form of graphs. A recent attempt is Hashtag2vec (Liu et al., 2018), a hashtag representation framework implementing a heterogeneous graph that connects hashtags with tweets and words in the corpus. They evaluate their framework using hashtag clustering tasks and compare the results to several different methods, including content-based, structure-based, and content-structure-based hashtag representation methods. An important discovery by Liu et al. is that content-based models perform better than structure-based models, but when the two are considered simultaneously, the model performs even better. Their claims align with that of Tu et al. (2017) in their CANE (context-aware network embedding) framework, pointing out that text and structure of a network are equally important for the calculation of network embeddings. TADW (text-associated deep walk) by Yang et al. (2015) is proposed based on similar concepts, they alter the DeepWalk algorithm (Perozzi et al.,2014) while taking into account the "sophisticated interactions between network structure and text information."

In terms of the structural aspect of hashtag networks, some research first constructs the network through co-occurring hashtags in a post, and then apply network embeddings frameworks such as node2vec (Grover et al., 2016), LINE (Tang et al., 2015), or DeepWalk (Perozzi et al.,2014) to extract node embeddings. Examples include Hashtag2vec (Liu et al., 2018) and Wang et al. (2016), who examined information virality through the co-occurrence relationship between hashtags. As for the textual aspect of hashtag networks or general networks, neural embedding techniques with attention mechanisms or topic modeling techniques such as LDA, PLSA seem to be common choices.

3. Problem Statement

This section defines the goal of this paper. For a set of n posts, each post is represented by 3 modalities and a set of metadata, denoted as $P = (X_t, X_i, X_h, X_m)$, where X_t, X_i, X_h, X_m denote text features, image features, hashtag features, and metadata respectively. Hashtag features X_h are generated by concatenating the average hashtag node embeddings $\overline{V}_h$ in a post with the post topic embedding X_T, represented as $X_h = \overline{V}_h \oplus X_T$. For the hashtag network $G = (V_h, E)$, the vertices consist of all hashtags in the corpus, and edges connect two hashtags co-occurring in the same post. For post metadata, 3 features including text length, hashtag counts, and time of post during a day are considered. Finally, each post is assigned a popularity label (explained in section 4.2.2), being either popular (1) or unpopular (0). These binary labels, along with the multimodal representation of each post are trained together as a supervised binary classification model. Given a set of future posts on Instagram, the model can then be used to predict whether the posts will be popular or unpopular.

4. Methodology

4.1 Data Collection

This section explains in detail the method of data collection and processing used in this study. The corpus is built from scratch using the python package "instagram-scraper" by arc298 on Github (https://github.com/arc298/instagram-scraper). The package is able to crawl a diverse array of information, including, post captions, images, followers and following count, hashtags used, post timestamp, etc. The present study gathers Instagram posts from top 100 social media influencers in Taiwan; the list of top 100 influencers is provided by Business Next, an online media focusing on global trends in technology, the Internet, and digital marketing. The ranking list is compiled based on an interdisciplinary metric calculated through engagement factors from Facebook, Instagram, Youtube, and excluding actors or singers whose popularity do not originate from Internet platforms. This study selects only influencer posts as corpus entries for a few reasons. First of all, the main goal of this study is to provide commercial insights into predicting popular posts on Instagram. Therefore, as Instagram influencers usually participate in business activities, their posting behaviors will be of prior consideration compared to those of the general public. Secondly, influencers have a larger base of online audience, so their likes or comments count can easily reflect whether their audience is fond of their content or not. Compared to influencers, an average Instagram user only has a limited number of followers, so the likes and comments count will remain pretty much the same regardless of what is posted, making popularity hard to track.

All posts are crawled and stored in the database on 2021/11/01. After the raw influencer posts are crawled, the following steps are taken to further process the posts:

1. Exclude influencer accounts with less than 50 posts, leaving posts from 83 influencers.
2. Collect only the most recent 550 posts for each influencer.
3. Exclude posts submitted between 2021/10/01 – 2021/11/01, as they are too recent to the crawling time, making their likes and comments count still unstable.
4. Exclude videos from the dataset. If a post contains multiple images, consider only the first image (thumbnail image) for image representation.
5. Extract post metadata and hashtags.
6. Construct hashtag network by linking all hashtags that co-occur in the same post. This results in a homogeneous graph with undirected edges.
7. Clean post captions by removing symbols, punctuations, URLs, and emojis. Each piece of text is then tokenized using ckip-transformers (level=3).
8. Perform topic modeling on the whole corpus using BERTopic. 250 topics are discovered and assigned to each post. Details of BERTopic will be explained in section 4.3.3.

The steps above leave the final dataset with a total of 19267 posts, and a hashtag network that contains 17906 unique nodes (hashtags) and 73300 edges. Every post contains 7 features ready for representation learning or model training, as shown in table 1 below.

Feature	Description
text	post caption of each post
image	thumbnail image of each post
hashtag	hashtags used in each post
topic	topic assigned for each post
timeofpost	time of post during a day
text_length	length of post caption
hashtag_num	hashtag count in a post

Table 1. Features derived from Instagram posts

Overall, the average length of post captions is about 50 words, and each post contains an average of 3 hashtags. In addition, top 100 influencers in Taiwan mostly submit their posts in the evening during 18:00-24:00.

4.2 Proposed framework

4.2.1 Multimodal Framework Overview

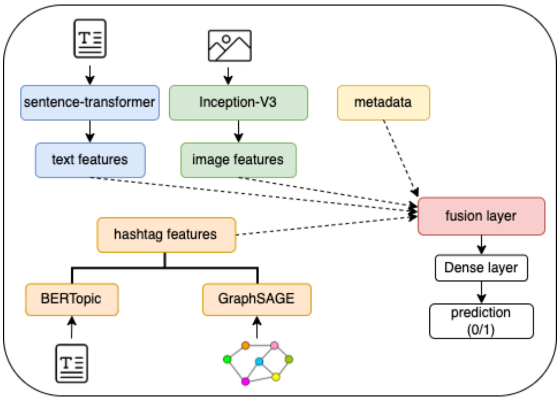


Figure 1. Illustration of the multimodal framework leveraging hashtag network

Figure 1 illustrates the multimodal post representation framework used in this study. For text modality, each post caption is encoded to a feature vector of length 512 using sentence-transformers. For image modality, each image is represented as a feature vector of length 1024 using the Inception-V3 featurizer. As for hashtag modality, two kinds of embeddings are involved: topic embeddings (of length 512) calculated using BERTopic and average node embeddings (of length 50) calculated from the hashtag network using GraphSAGE, which combines into a hashtag feature vector of length 562. Last but not least, 3 additional metadata are used in the framework, as shown in table 1. The four components – text, image, hashtags modalities and metadata, are then concatenated in the fusion layer, forming a final feature vector of length 2101. Ultimately, the concatenated features are connected to a dense layer with a sigmoid activation function outputting the final prediction probability. The popularity class is then derived from the probabilities accordingly.

4.2.2 Defining Popular Posts

In this section, how popular posts are determined will be explained in detail. The methods used in this study mostly follows that of Carta et al. (2020); however, instead of calculating the likes moving average (LMA) as the popularity metric, the present study calculates the engagement moving average (EMA). Engagement *E* in this study is defined as the number of likes count plus 2 times the number of comments in each post. Let P_i be the ith post in an influencer account, the formula can be shown as follows:

$$E(P_i) = \text{likes\_count}(P_i) + 2*\text{comments\_count}(P_i)$$

The EMA of a given post P_i can then be calculated by averaging the engagement of its previous *k* posts, as shown in the following formula:

$$EMA_k(P_i) = \frac{\sum_{i-k}^{i-1} E(P_i)}{k}, i > k$$

Finally, the popularity class (PC) of a given post P_i can be derived by comparing its engagement with its EMA times a parameter Δ, as shown in the following conditional formula:

$$PC_{k,\Delta}(P_i) = \begin{cases} 1\,(popular), & if\ \ E(P_i) > EMA\_k(P_i) * (1+\Delta) \\ 0(unpopular), & otherwise \end{cases}$$

According to the formula above, a post is labeled as a popular post (1) if its current engagement is greater than the average engagement of its previous k posts (EMA). Specifically, the parameter Δ controls the strictness of the threshold; for example, when Δ = 0.1, the engagement of a post needs to be greater than 1.1 times its EMA in order to be labeled as a popular post. In other words, as Δ becomes higher, it will also be harder for a post to become a popular post, decreasing the total number of popular posts as well. In Carta's work, they experimented with 3 levels of *k* (10,30,50), and 4 levels of Δ (0, 0.05, 0.1, 0.15). It is concluded that when *k* is larger, the accuracy of their model also rises accordingly. In light of this discovery, the present study simplifies Carta's experiments by directly assigning *k* as 50 while keeping 3 levels of Δ (0, 0.1, 0.15).

There are a few benefits to the current method used to determine popular posts. To begin with, this method negates the biases caused by difference in the number of account followers. Some influencer accounts have significantly more followers than others, therefore, their posts will naturally have more likes and comments as the follower base is larger. If raw likes and comments are used as the popularity metric, popular posts will mostly contain those from high-follower accounts. Since EMA is calculated independently for each account, each influencer can still have a fair number of popular posts regardless of how many followers they have. Secondly, this method also avoids issues caused by posts submitted too long ago. As time goes by, influencers gradually gain more followers; as a result, posts that are submitted more recently generally have more engagement than those submitted years ago. This way, popular posts will simply contain posts that are more recent, but not posts whose content are actually appealing to the audience. As EMA considers previous *k* posts of a given post, posts from a long time ago can still be evaluated on a scale that matches its follower base at that time, hence solving the issue caused by different posting times.

4.3 Feature Engineering

4.3.1 Text Representation

Using the segmented post captions, this study has first attempted and compared a few text representation methods before combining with other modalities:

- LSTM models incorporating self-trained word2vec embeddings as the embedding layer weights
- The sentence-transformers framework using "distiluse-base-multilingual-cased-v1" as the pretrained model.

Model	Accuracy (Δ=0.15)
Single-layer LSTM w/ word2vec	59.63%
Bidirectional LSTM w/ word2vec	56.93%
Sentence-transformers	62.53%

Table 2. Model accuracy comparison for text representation

Table 2 compares the testing set accuracy from the above models using only post captions (texts) as inputs when Δ=0.15. As shown in table 2, text embeddings produced by sentence-transformers yield the best results compared to other sequence models. Therefore, sentence-transformers is chosen to be the text encoder in this study.

4.3.2 Image Representation

Image is also a very important aspect for Instagram posts. Standard and state-of-the-art image representation methods often involve knowledge in convolutional neural networks (CNN). The present study implements pic2vec, a lightweight image featurizer by datarobot on Github (https://github.com/datarobot/pic2vec), and selects the Inception-V3 pretrained model to encode each image into a feature vector of length 1024.

4.3.3 Hashtag and Topic Representation

As mentioned in section 2.2.1, hashtags serve as effective topic markers and search tools in social media. In such context, the present study leverages the topicality and searchability of hashtags and attempt to utilize hashtags to reflect the topic and real-world information in a post.

Sometimes, certain posts on social media receive a high amount of engagement not because of attractive captions or images, but because they contain trending hashtags. For example, posts with hashtags like #metoo, #blacklivesmatter, #tokyo2020 easily bring in events or issues happening in the real world, making these posts more relatable to the public and thus earning more engagement. It is clear at this point that hashtags carry valuable information which should not be ignored in a social media post representation.

Since this study mainly focuses on influencer posts on Instagram, hashtags are widely presented throughout the corpus, making hashtags suitable targets for topic representation in the multimodal framework. However, as pointed out in section 2.2.2, representing textual or topical content is as important as representing the structural details of the hashtag network. Consequently, the present study combines two elements for hashtag representation: topic embeddings to represent hashtag content (content embedding) and node embeddings to represent the graphical structures between hashtags (structure embedding).

Content embeddings are topic embeddings, which are extracted using the python package BERTopic by MaartenGr on Github (https://github.com/MaartenGr/BERTopic). BERTopic is a topic modeling algorithm that uses the transformers framework and c-TFIDF to create topic representations. In short, BERTopic first creates a series of document embeddings using sentence-transformers, and then use UMAP (McInnes, 2018) for embeddings dimension reduction and HDBSCAN (Campello, 2013) for document clustering. Finally, they calculate the class-based TFIDF for each cluster (topic) and produce the importance scores of each word in every topic. The topic embedding for each topic is calculated by averaging all document embeddings within the same topic.

As for structure embeddings, 2 different frameworks are used and compared in this study: node2vec and GraphSAGE. Node2vec generates node sequences using random walk and feeds them to word2vec to create node embeddings; GraphSAGE implements an unsupervised framework that learns a binary classifier, predicting whether arbitrary node pairs are likely to co-occur in a random walk performed on the graph. Both frameworks can produce informative node embeddings, and can be easily implemented using the python package stellar-graph. These two specific frameworks are chosen since they each represent a node representation technique; node2vec represents random walk-based methods, and GraphSAGE represents GCN-based methods. The hashtag network constructed previously are then used as input to the 2 methods, which produces node embeddings for each hashtag. If multiple hashtags (>1) appear in a single post, the structure embedding of a post will be calculated by averaging the node embeddings of all hashtags in the post. On the other hand, if a post contains does not contain any hashtag, then a zero-vector will be assigned as the structure embedding.

Table 3 shows a comprehensive comparison between the performance (testing set accuracy) of different hashtag representation strategies. The ⊕ symbol represents the concatenation operation. The accuracy is calculated with Δ=0.15:

Model	Accuracy (Δ=0.15)
Topic embedding	61.49%
Node2vec	61.06%
GraphSAGE	61.75%
Topic embedding ⊕ node2vec	61.08%
Topic embedding ⊕ GraphSAGE	61.81%

Table 3. Model accuracy comparison for hashtag representation

It can be observed in table 3 that the topic ⊕ GraphSAGE combination outperforms all other models. Therefore, hashtags in this study are represented by concatenating the topic embedding of a post with its structure embedding derived from GraphSAGE.

4.3.4 Metadata

As shown in table 1, there are in total 3 kinds of metadata used in the present study. T-tests are performed respectively on *text_length* and *hashtag_num*, and both tests yield p-values smaller than 0.001. As for the time of post feature, a day is divided into 4 sections (as in figure 1 (c)) and represented as a categorical feature with 4 levels. A chi-square test performed on *timeofpost* also yields p-value smaller than 0.001. Therefore, *text_length*, *hashtag_num*, and *timeofpost* are all statistically significant features between popular and unpopular posts. Before the model training stage, *text_length* and *hashtag_num* are both normalized using the StandardScaler() function from scikit-learn.

4.4 Model Evaluation

4.4.1 Evaluation Metrics

In the results section, 2 different kinds of evaluation metrics will be used to present the performance of the models, including accuracy and balanced accuracy. Accuracy is defined as the ratio of the number of correct predictions to the total number of input samples, while balanced accuracy is defined as the average of recall obtained on each class. Balanced accuracy excels at evaluating the performance of a binary classifier since it avoids false high accuracy caused by an imbalanced dataset.

$$Balanced\ Accuracy = \frac{R_{c0} + R_{c1}}{2}$$

4.4.2 Baselines

To better evaluate the effectiveness of the features used in each model, 4 different baseline methods will be presented along the proposed framework. These baselines do not involve any machine learning techniques, but are instead based on naïve assumptions on the dataset. The first baseline simply predicts all posts in the dataset as unpopular posts:

$$Baseline\ 1\_PC_{k,\Delta}(P_i) = 0$$

For the other three baselines, the popularity class is derived based on a certain threshold. To be more specific, a post will be predicted as a popular post if its engagement exceeds an engagement threshold *T:*

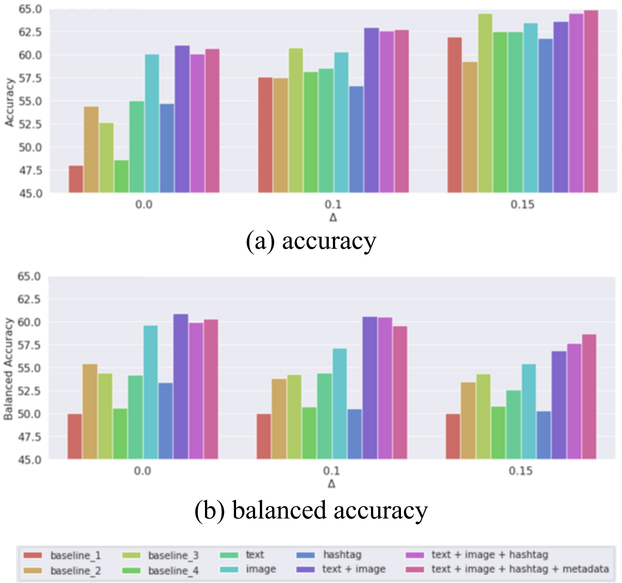
$$\text{Baseline n\_}PC_{k,\Delta}(\mathrm{P_i}) = \begin{cases} 1(\text{popular}), & \text{if} \quad \mathrm{E(P_i) > T} \\ 0(\text{unpopular}), & \text{otherwise} \end{cases}$$

$$with\ T = (30{,}000, 60{,}000, 100{,}000), \quad n = (2{,}3{,}4)$$

5. Results

5.1 Results Overview

This section compares the results from combining different modalities and popularity parameters. The proposed framework is evaluated together with the baselines established in section 4.2.2 using the metrics in section 4.4.1.



(a) accuracy

(b) balanced accuracy

Figure 2. Visualization of model accuracy and balanced accuracy

Figure 2 (a) plots the accuracy of each model along with the baselines on 3 different Δ levels. It can be observed that in general, model performance increases as delta goes up. A possible explanation for this trend is that as Δ rises, the threshold for a post to become a "popular post" rises as well, causing fewer popular posts to appear in the dataset. In this case, the accuracy of predicting a post as unpopular (0) will increase due to imbalanced dataset distribution. However, even when Δ = 0.15, predicting all instances as unpopular

	Δ =0		Δ =0.1		Δ =0.15	
Model	Accuracy	Balanced accuracy	Accuracy	Balanced accuracy	Accuracy	Balanced accuracy
baseline_1	48.01%	50.00%	57.58%	50.00%	61.96%	50.00%
baseline_2	54.47%	55.49%	57.56%	53.82%	59.31%	53.46%
baseline_3	52.68%	54.41%	60.79%	54.31%	64.49%	54.35%
baseline_4	48.65%	50.62%	58.19%	50.73%	62.56%	50.80%
text	55.06%	54.21%	58.59%	54.45%	62.53%	52.62%
image	60.09%	59.65%	60.35%	57.19%	63.47%	55.44%
hashtag	54.75%	53.43%	56.62%	50.53%	61.81%	50.35%
text + image	**61.08%**	**60.95%**	**62.95%**	**60.59%**	63.67%	56.90%
text + image + hashtag	60.09%	59.96%	62.58%	60.53%	64.50%	57.70%
text + image + hashtag + metadata	60.66%	60.32%	62.79%	59.62%	**64.87%**	**58.72%**

Table 4. Experiment results for *k*=50 (the best result for each Δ level is in bold)

(baseline 1), only yields an accuracy of 61.96%. At the same Δ level, all other multimodal models outperform the results of baseline 1 by 2~3%. Therefore, it can be inferred that even higher Δ levels lead to imbalanced dataset, the multimodal models are still able to capture additional information rather than just guessing posts as unpopular. Out of the three multimodal models, the *text + image* combination produces the best results when Δ = 0 and 0.1, but when Δ = 0.15, combining text, image, and other modalities produces even better results. As a result, it is possible that a stricter popularity threshold could bring out the effectiveness of combining multiple modalities. In addition, figure 2 and table 4 both demonstrate the fact that multimodal models outperform any unimodal models, proving that modality fusion is able to model the interaction between modalities and boost overall performance.

On the contrary to accuracy, overall balanced accuracy decreases as Δ rises, as shown in figure 2 (b). This is probably due to the recall obtained on popular posts decreases as popularity threshold goes up. Precisely, as Δ rises, the models start to predict posts as unpopular (0) more frequently, and rarely predicts a post as a popular post. Therefore, the recall for popular posts will drop significantly, which in turn lowers the balanced accuracy. In other words, the model's abilities to predict popular posts are hindered at higher Δ levels.

5.2 Effectiveness of the Hashtag Modality

An important goal of the present study is to harness the information provided from hashtag networks and improve the quality of post representations on social media. As mentioned in the introduction, we hope that hashtag networks can be used as a new form of hashtag representation method for other tasks as well. Consequently, this section describes in detail the effectiveness of hashtag representation framework used in this study.

To begin with, the hashtag modality gets increasingly effective as Δ rises. Compare the accuracies of *text + image* and *text + image + hashtag* models; when Δ = 0, *text + image* performs better by 0.99%, and when Δ = 0.1, *text + image* still performs better, but only by 0.37%. However, at Δ = 0.15, *text + image + hashtag* outperforms *text + image* by 0.83%. The same trend applies to balanced accuracy, with *text + image + hashtag* performing better than *text + image* when Δ = 0.15. In conclusion, at higher Δ levels, the addition of hashtag modality can further improve model performance.

Secondly, as mentioned previously, the hashtag modality consists of 2 components: structure (node) embeddings produced by GraphSAGE and content (topic) embeddings produced by BERTopic. From table 3 in section 4.3.3, it can be observed that when the two are combined together, they indeed perform better than GraphSAGE embeddings or topic embeddings alone. The results align with those mentioned in previous works, where they claim that semantic information has to be considered with structural information in a graph to produce better results. The hashtag network used in this study is presented below:

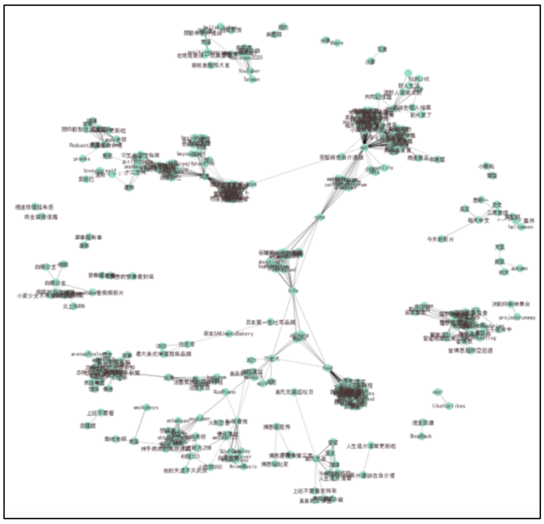


Figure 3. A simplified illustration of the hashtag network

Figure 3 shows the simplified hashtag network constructed from the co-occurrence relationship between hashtags in a post. Overall, the network shows an island-like distribution; some bigger node clusters are connected in the middle while smaller ones spread on the outer areas. Each hashtag cluster in the network represents and captures latent topics among the posts. However, in the current dataset, there is a considerable number of "lone hashtags", meaning that these hashtags only appear once in the entire dataset and are isolated from any node cluster. These hashtags usually contain a full sentence and are especially long in length; this is perhaps due to the fact that unlike English, words are connected without spaces in Chinese, so the hashtag can run on while maintaining readability. In this study, lone hashtags are still mapped into the hashtag network to preserve the overall structure of hashtags; however, they might become potential noise in the model training phase.

To evaluate semantic information captured by both the hashtag network and topic model, top 5 hashtag clusters in the network and top 5 topics from BERTopic are extracted and compared in the following table.

Top 5 hashtag clusters	Top 5 topics
#貓(cat)	解析_劇情_電影(movie)
#料理(cooking)	音樂_演唱會_歌曲(music)
#志祺七七 (Chih-Chyi77)[1]	衣服_西裝_穿搭(clothing)
#strnetwork[1]	料理_拌炒_醬油(cooking)
#占卜(fortune telling)	頭髮_髮型_短髮(hairstyle)

Table 5. List of top hashtag clusters and topics

From table 5, it can be inferred that the hashtag network and topic model seem to capture different aspects of Instagram influencer posts. Hashtag network captures granular textual information such as individual influencers, cats, or fortune telling, while topic model captures general information and topics such as movie, music, or clothing. Therefore, when combined together as the hashtag representation, hashtag network and topic model might mutually enhance their abilities to produce more informative embeddings.

6. Conclusion and Future Work

This paper proposes a multimodal framework incorporating hashtag network to predict popular Instagram posts of Taiwanese influencers. To be more precise, the hashtag modality of a post is constructed by fusing node embeddings from the hashtag network and topic embeddings from BERTopic. Instead of predicting raw likes or engagement, this study follows Carta et al. (2020) and builds a binary classifier that predicts posts as either popular or unpopular. Results indicate that the proposed multimodal framework surpasses all baselines and unimodal models. More importantly, the hashtag modality is shown to improve overall model performance at higher Δ levels, while its two components complement each other by capturing different post details.

There are some future directions for the present study. First of all, the quality of hashtag embeddings can be improved by trying out different node representation algorithms and topic models; in addition, the attention mechanism might also be of much help in order to design a more sophisticated hashtag representation framework. Secondly, lexical analyses should be performed on popular and unpopular posts, as the textual/multimodal information hidden in the posts can not only contribute to model improvement but also benefit influencers and the marketing industry. Finally, the ultimate goal of this study is to develop hashtag networks as a language resource for social media/text analytics. Besides popularity prediction, hashtag networks will also be applied to several other tasks to evaluate their full effectiveness or potential. We hope that hashtag networks can eventually become a convenient yet effective way of representing hashtags, and benefit future research in social media.

7. Acknowledgments

I would like to thank everyone from the LOPE lab for providing me with invaluable feedback and new research directions. I would also like to thank Tagnology.co for providing me with relavant data and advice on the model structure.

8. Bibliographical References

Carta, S., Podda, A. S., Recupero, D. R., Saia, R., & Usai, G. (2020). Popularity prediction of instagram posts. *Information*, 11(9), 453.

De, S., Maity, A., Goel, V., Shitole, S., & Bhattacharya, A. (2017, April). Predicting the popularity of instagram posts for a lifestyle magazine using deep learning. In *2017 2nd International Conference on Communication Systems, Computing and IT Applications (CSCITA)* (pp. 174-177). IEEE.

Grootendorst, M. (2020). Bertopic: Leveraging bert and c-tf-idf to create easily interpretable topics. *Version v0, 4*.

Grover, A., & Leskovec, J. (2016, August). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 855-864).

Hamilton, W. L., Ying, R., & Leskovec, J. (2017, December). Inductive representation learning on large graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp. 1025-1035).

Huang, F., Chen, J., Lin, Z., Kang, P., & Yang, Z. (2018, October). Random forest exploiting post-related and user-related features for social media popularity prediction. In *Proceedings of the 26th ACM international conference on Multimedia* (pp. 2013-2017).

Kim, S., Jiang, J. Y., & Wang, W. (2021, March). Discovering undisclosed paid partnership on social media via aspect-attentive sponsored post learning. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (pp. 319-327).

Liu, J., He, Z., & Huang, Y. (2018, July). Hashtag2Vec: Learning Hashtag Representation with Relational Hierarchical Embedding Model. In *IJCAI* (pp. 3456-3462).

Mazloom, M., Rietveld, R., Rudinac, S., Worring, M., & Van Dolen, W. (2016, October). Multimodal popularity prediction of brand-related social media posts. In *Proceedings of the 24th ACM international conference on Multimedia* (pp. 197-201).

Perozzi, B., Al-Rfou, R., & Skiena, S. (2014, August). Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 701-710).

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Scott, K. (2015). The pragmatics of hashtags: Inference and conversational style on Twitter. *Journal of Pragmatics*, *81*, 8-20.

[1] "Chih-Chyi77" and "strnetwork" are both Youtube creators in Taiwan

Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q. (2015, May). Line: Large-scale information network embedding. In *Proceedings of the 24th international conference on world wide web* (pp. 1067-1077).

Tu, C., Liu, H., Liu, Z., & Sun, M. (2017, July). Cane: Context-aware network embedding for relation modeling. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1722- 1731).

Wang, R., Liu, W., & Gao, S. (2016). Hashtags and information virality in networked social movement: Examining hashtag co-occurrence patterns. *Online Information Review*.

Yang, C., Liu, Z., Zhao, D., Sun, M., & Chang, E. (2015, June). Network representation learning with rich text information. In *Twenty-fourth international joint conference on artificial intelligence*.

Zappavigna, M. (2015). Searchable talk: The linguistic functions of hashtags. *Social Semiotics*, *25*(3), 274-291.

Zhang, Z., Chen, T., Zhou, Z., Li, J., & Luo, J. (2018, December). How to become Instagram famous: Post popularity prediction with dual-attention. In *2018 IEEE international conference on big data (big data)* (pp. 2383-2392). IEEE.