Proceedings of the Sixteenth International AAAI Conference on Web and Social Media (ICWSM 2022)

#Emoji: A Study on the Association between Emojis and Hashtags on Twitter

Yuhang Zhou, Wei Ai
College of Information Studies, University of Maryland, College Park, USA
{tonyzhou, aiwei}@umd.edu

Abstract

Prevalent on modern social media, both hashtags and emojis are text elements that function beyond plain text. While hashtags utilize free-formed strings and are highlighted by the platform, emojis are bonded by Unicode Standard and rendered by the platforms. Yet both are used to mark discussion topics, express sentiment, show identity, and highlight keywords. This paper analyzes and highlights the strong association between hashtags and emojis, not only in their usage frequency, but also in their semantics. We show that the association is strong enough for improving downstream tasks. To this end, we design a representation learning model that can learn emoji-based representations to improve hashtag prediction.

Introduction

The Making of #ChristmasTreeFarm 24 hours and a lot of xmas spirit🎄

On December 13, 2019, Taylor Swift, a world-wide singer posted a tweet to share the process of making the Christmas song "Christmas Tree Farm."[1] In this Tweet, Taylor used the hashtag *#ChristmasTreeFarm* and the Christmas tree emoji (🎄) to emphasize that the Tweet is related to Christmas.

This is one of many examples where emojis and hashtags jointly appear in users' microblogs and are semantically related. From a natural language perspective, both hashtags and emojis are the "unnatural" tokens that stand out in plain text, but the two differ in several ways. A hashtag is a string of characters preceded by the hash (#) character, which is usually a combination of words, numbers, or abbreviations. Although hashtags are still text strings, they are usually highlighted and hyperlinked by the platform. Unlike the free-form hashtags, emojis are defined by the Unicode standard, which limits the user's choice to an expanding list of supported emojis. These Unicode characters are rendered by the platform as vivid pictographs, representing faces, objects, flags, etc.

Despite the difference in appearance and flexibility, both hashtags and emojis encode rich semantics beyond a single word or phrase, which has attracted many researchers on their roles and functionalities. Although most existing research has separately focused on either hashtags or emojis, their conclusions reveal that the two have similar and overlapping roles in social media. For example, Yang et al. (2012) suggest that hashtags play a dual role on Twitter: a topic indicator of the content and a symbol of community around particular topics. As a result, hashtags have been widely used to mark discussion topics, express sentiment, show identity, and highlight keywords (Tsur and Rappoport 2012; Davidov, Tsur, and Rappoport 2010; Feng et al. 2015; Zhang et al. 2016). Similarly, research has found that emojis share similar functionalities of highlighting the topics (Lu et al. 2016), conveying sentiment (Ai et al. 2017), and indicating identities (Ge 2019). As a result of such functionalities, not only have both hashtags and emojis been popular among social media users, but they have also been widely used by researchers to study demographic and cultural phenomenons (Lu et al. 2016). Indeed, both hashtags and emojis have evolved into new-era languages (Ai et al. 2017; Zhang 2019).

Since both "languages" encode rich semantics within a short string, it is a common challenge to understand and analyze them at scale. Hashtags usually consist of abbreviated or concatenated concepts, making them hard to understand for outside observers. Emojis have also been associated with great ambiguity due to the rich information coded in the tiny pictures (Miller et al. 2016). On the other hand, although some previous work assumes a hidden correlation between emojis and hashtags (Park, Xu, and Fung 2018), little research has explicitly examined the relationship between hashtags and emojis. In this work, we take the initiative to study the relationship between emojis and hashtags and provide a new perspective to understand their usage. We demonstrate the strong association between emojis and hashtags, especially in their semantics. Specifically, we show how emojis can be applied to two downstream tasks on hashtag analysis, namely hashtag clustering and hashtag prediction, with satisfying performance.

We summarize our major contributions as follows:

- To the best of our knowledge, we make the first effort to construct a quantitative study that correlates hashtag usage to emoji usage.
- We present a comprehensive empirical analysis on the

[1] https://twitter.com/taylorswift13/status/1209113259737047040

semantic correlation between hashtags and emojis. We can rely on the semantic correlation to effectively cluster hashtags only dependent on their co-used emojis.

- We propose a novel machine learning model that incorporates emoji representation to predict the hashtags in Tweets. The derived model can achieve better performance than a model only using text information of Tweets.

Related Work

Our work mainly builds on two streams of previous literature – emoji and hashtag usage both in social media and related downstream tasks.

Emojis

The prevalence of emojis has gained researchers' focus from various areas. Researchers pay attention to the functionality of emojis in social media. Sentiment-related emojis play a complementary role of text and entity-related emojis can be used as alternatives to words (Ai et al. 2017). Except for the most common application for expressing sentiment, emoji usage also shows community attributes. Users apply emojis to make a message more engaging to the recipient, or for relationship maintenance (Hu et al. 2017). The differences in emoji usage are demonstrated across communities, such as, apps, platforms, languages, cultures and genders (Tauch and Kanjo 2016; Lu et al. 2016; Chen et al. 2018b; Barbieri et al. 2016; Miller et al. 2016). Besides the study of emoji usage in social media, researchers also develop models to incorporate emoji information in the downstream tasks. Sentiment classification is a primary application for including emoji features. Researchers learn emoji-powered sentence representations (Chen et al. 2019; Felbo et al. 2017) or directly use emoji embeddings for sentiment classification (Lou et al. 2020; Chen et al. 2018a; Eisner et al. 2016).

Hashtags

Hashtag as another label to convey the information in Tweets attracts much of researchers' attention. For the role of hashtag usage in social media, previous research has reflected that hashtags can serve as both a content indicator and community membership symbol (Yang et al. 2012). As content indicators, users commonly add hashtags to indicate the topics (Yan et al. 2013; Li et al. 2016; Hong et al. 2012) and key phrases in Tweets (Zhang et al. 2016). As community membership symbols, hashtags are able to define a virtual community of the same topic. For example, researchers learned about discussion trends of the Covid19 pandemic by discovering multiple communities of hashtag usage (Cruickshank and Carley 2020). Moreover, hashtag usage can be a measure of how active community members are (DeMasi, Mason, and Ma 2016). In reality, the differences for hashtag usage are also revealed in different demographic groups (An and Weber 2016; Olteanu, Weber, and Gatica-Perez 2016; Ye et al. 2018). The role of hashtag usage helps researchers to improve information retrieval downstream tasks, such as summarization (Zhang et al. 2012; Chang et al. 2013) and sentiment classification (Wang et al. 2011; Davidov, Tsur, and Rappoport 2010). However, not all microblogs are tagged with hashtags, so researchers also aim to automatically annotate the hashtags. They have designed various models for hashtag prediction, such as Latent Dirichlet allocation (LDA) models (Li et al. 2016), selecting hashtags from a predefined list (Gong, Zhang, and Huang 2015), or sequence-to-sequence (seq2seq) language generation models. Previous experiments' results show that seq2seq generation models largely outperform other types of prediction models (Wang et al. 2019).

Emoji-Hashtag Intersection

The two streams of literature on emojis and hashtags do not intersect frequently. Only a few research studies have hinted at the correlations between the two. Highfield (2018) conducted a case study of "emoji hashtags" in Instagram (using emojis in hashtags) and "hashflags" in Twitter (adding an emoji to the end of a hashtag), both being platform functionalities that support users in combining the two features. The data mining researchers have implicitly exploited the association between emojis and hashtags by incorporating both in building machine learning models for downstream tasks. For example, Sari et al. (2014) and Suttles and Ide (2013) use emojis and hashtags together to distantly label Tweets' sentiments and Park, Xu, and Fung (2018) apply the emoji-related feature and hashtag-related feature to improve the sentiment classification task. Compared with their work, our aim is to explicitly study the association between emojis and hashtags. Our experiment is to adopt emoji-related features to observe the effect of hashtag clustering and hashtag prediction, which quantitatively demonstrate the dependency between emojis and hashtags.

Tweets with Emojis and Hashtags

We begin our analysis by quantitatively measuring the correlation between emojis and hashtags in Tweets. We collect a large English Tweet dataset using the Twitter Gardenhose API. The collected Tweets span from January 2020 to January 2021, covering 4.82 billion Tweets from about 50 million active users with at least 5 posts. We aggregate the emoji and hashtag usage information of each user in the dataset for frequency analysis in Section *Frequency Association* and extract the frequency of co-used emojis for each hashtag in the semantic analysis in Section *Co-occurring Emojis in Trending Hashtags*.

Frequency Association

We hypothesize that users who use more emojis are more likely to use hashtags and vice versa, as earlier work (DeMasi, Mason, and Ma 2016; Hu et al. 2017) has confirmed that hashtags and emojis are both correlated to user engagement. Yet the alternative might also be true, that is, some users might have a strong preference of one type over another, rather than having similar passions for both. To answer this question, we calculate the percentage of Tweets containing emojis (%emoji-tweet) and the percentage of Tweets containing hashtags (%hashtag-tweet) for each user.

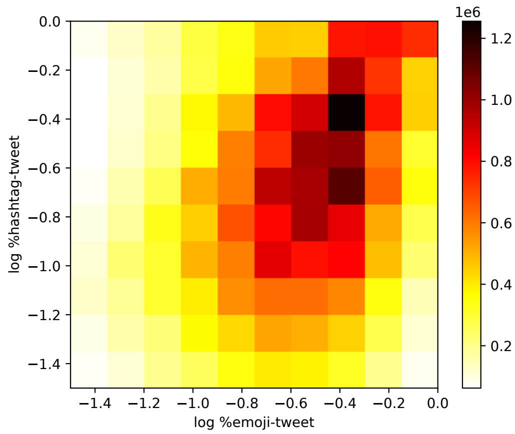

Figure 1: Heatmap of %emoji-tweet and %hashtag-tweet (percentage of Tweets containing emojis and hashtags)

Then, we measure the Pearson correlation between %emoji-tweet and %hashtag-tweet, which is **0.223** (p-value $\ll 0.01$, paired t-test). Given the large dataset, there are many potential factors that affect whether individual users use emojis and/or hashtags, which add a lot of noise in the aggregated measures of %emoji-tweet and %hashtag-tweet. Therefore, a magnitude of 0.223 indicates a strong correlation between the usage of emojis and hashtags, suggesting that users who use more emojis are more likely to use hashtags as well, and vice versa.

Further, we plot the heatmap of the %emoji-tweet and %hashtag-tweet in log scale. The color scale indicates the number of users in the range of each cell. As shown in Figure 1, most data points concentrate on the diagonal area of the heatmap, which indicates that for most users, more hashtag usage is correlated with more emoji usage.

Such frequency correlation between Tweets with emojis and hashtags, however, is studied at user level rather than Tweet level. To this end, we calculate the probability of Tweets containing hashtags, emojis, or both. 4.5% of Tweets contain both hashtags and emojis, while 14.43% and 21.64% Tweets contain only hashtags or emojis correspondingly. Such statistics yield a lift value of 1.44 (Ordonez and Omiecinski 1999). That is, a Tweet with hashtags is 0.44 times more likely to also use emojis compared with a random Tweet, which motivates us to investigate the co-occurrence of emojis and hashtags further.

So far, the frequency associations have been measured without considering the semantics of the emojis or hashtags. However, previous work suggests emojis' complementary role of text (Ai et al. 2017) and hashtags' content indicator role (Yang et al. 2012), which prompts us to further explore the associations in their semantics. In the next section, we analyze the correlation between emojis and individual hashtags co-used in the same Tweet to explore the semantic association between the hashtags and emojis.

Hashtag	Top 5 Co-used Emojis
#BlackOutBTS	✨ 😌 🥺 ❤️ 🖤
#BAEKHYUN	🔗 🍬 ❤️ ✨ 🍭
#MothersDay	❤️ 💕 💐 🤍 💫
#TheLastDance	🐐 😂 🍿 🐍 🔥
#SmackDown	👀 🔥 😂 😍 📺
#JusticeForGeorgeFloyd	✊ 🙏 🤬 ‼️ ❤️
#icantbreathe	😭 🙏 ✊ 🤬 😡
#BlackLivesMatter	‼️ ✊ ❤️ ❗ 👏

Table 1: Trending hashtags and their most frequently co-used emojis in May 2020

Semantic Association

Co-occurring Emojis in Trending Hashtags

We begin by hand-selecting a few popular hashtags and extract their most frequently co-used emojis. From the trending hashtags in the United States in May 2020, we select a combination of hashtags about fan cultures, holidays, sports, and social movements. The results are presented in Table 1.

We see that the hashtags are associated with different sets of frequently co-used emojis. Indeed, the co-used emojis reflect the general topics of the hashtags. *#BlackOutBTS* and *#BAEKHYUN* are about K-pop culture (the BTS boy band and the Korean idol Baekhyun), with which users tend to use ❤️ and ✨. For *#MothersDay*, people use ❤️ and particularly 💐 (bouquet) to express their emotion. *#TheLastDance* and *#SmackDown* are both sports TV shows, which is reflected in 🍿 (popcorn) and 📺 (TV). Also, 🔥 (fire) indicates the popularity and intensity of the sports. Lastly, *#JusticeForGeorgeFloyd*, *#icantbreathe*, and *#BLACK_LIVES_MATTER* all describe the "Black Lives Matter" movement in May 2020. Their most frequently co-used emojis, 🤬 (cursing face), ‼️ (bang), and 🙏 (pray) all express the anger and shock of users.

Among hashtags on similar topics, we can also see minor differences among their co-used emojis. A closer observation suggests that such minor differences do capture the nuances between the hashtags. For example, 🖤 (black heart) appears only in *#BlackOutBTS*, as the hashtag is to let black BTS fans shine and highlight their presence. It also provides evidence that both hashtags and emojis are used for the dual purpose of marking topics and showing identities. For another example, users particularly associate *#TheLastDancce* with 🐐 (goat), because the term G.O.A.T. (greatest of all time) is related to Michael Jordan, the protagonist of the TV show *The Last Dance*.

The observations from Table 1 suggest that the frequently co-used emojis can reveal the semantic meaning of hashtags, capturing the topic, identity, and sentiment. To this end, we question whether a hashtag can be well represented by its associated emojis, which could demonstrate the strength of their association. To answer this question, we conduct a clustering analysis to see if the emoji representation of hashtags can recover the clustering structure of the hashtags.

Clustering Hashtags with Emojis

For the clustering task, we construct a "bag-of-emoji" vector for each hashtag, where each dimension represents the co-occurrence frequency between the hashtag and the corresponding emoji. Since the popularity of both emojis and hashtags are power-law distributed, we normalize it with positive pointwise mutual information (PPMI) (Church and Hanks 1990). That is, in the vector for hashtag h, the value of the dimension regarding emoji e is

$$\text{PPMI}(h, e) = \max\left\{0, \log \frac{p(h, e)}{p(h)p(e)}\right\},$$

where $p(h), p(e), p(h, e)$ represent the marginal probabilities and joint probability of observing hashtag h and emoji e in a Tweet. PPMI measures how much the hashtag and emoji co-occur more frequently than we expect by chance. We calculate the distance between two hashtags as the cosine distance between their emoji-PPMI vectors and apply the hierarchical clustering with a single linkage (Bar-Joseph, Gifford, and Jaakkola 2001).

To select the hashtags for analysis, we obtain the Twitter Trending list of the United States in May 2020[2] and extract the 30 most frequent hashtags. In a span of one month, the selected top hashtags have a diverse distribution of topics, ranging from sports to entertainment and social movements. We also repeat the process for the entire year, but the top hashtags are all about entertainment or fan culture, and are less diverse in topics.

We visualize the clustering results in Figure 2. In the matrix heatmap, the color scale of each cell indicates the cosine distance between the two hashtags. The darker the cell is, the more similar the emoji-PPMI representations of the corresponding hashtags are. The dendrogram on the side of the heat map indicates the hierarchical clustering structure among the hashtags.

In Figure 2, we can identify five small but closely knitted clusters – WWE (World Wrestling Entertainment) TV programs (*#AEWDynamite*, *#WWENXT*, *#SmackDown*, and *#WWERaw*), webcast programs (*#InsecureHBO*, *#Verzuz*, and *#NS10v10*), social movements (*#icantbreathe*, *#Anonymous*, *#GeorgeFloyd*, and *#JusticeForGeorgeFloyd*), K-pop fan culture (*#BAKEHYUN*, *#ASTRO*, and *#NCTzenSelCaDay*), and three songs of Lady Gaga (*#SourCandy*, *#Chromatica*, and *#RainOnMe*).

The higher-level hierarchy between these clusters, however, is not well captured. For example, one may expect that the WWE cluster is closer to *#UFC249* rather than *#TrumpMeltdown*. Yet, this may be attributed to the single linkage used in the hierarchical clustering, which focuses more on local cluster structures. One may expect that the complete linkage could better handle the relationship between the small clusters.

[2]https://us.trend-calendar.com/, retrieved September, 2021

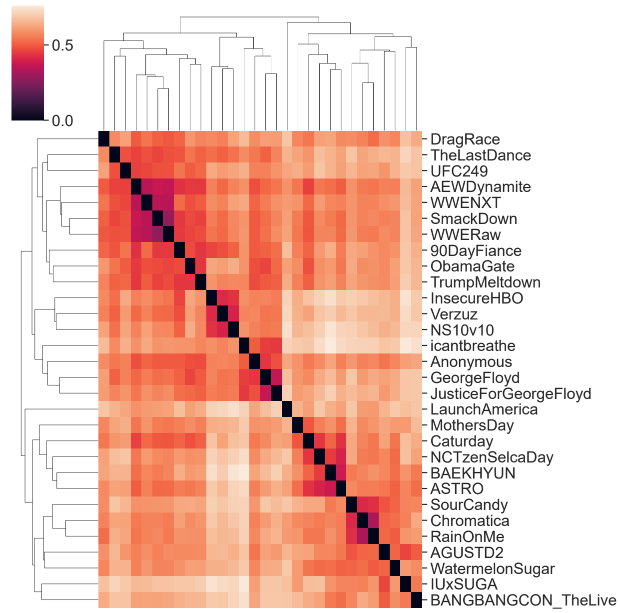

Figure 2: Clustering structure of top 30 frequent hashtags from USA Trending list of May 2020, based on their emoji-PPMI representations.

We repeat the clustering analysis on another two sets of hashtags. One is the 30 most frequent hashtags of an entire year and the other is a collection of popular hashtags in Japanese. Similar to Figure 2, we observe closely knitted local clusters. The results are presented in Appendix A.

So far, we have seen a strong semantic association between emojis and hashtags, particularly in their semantics. The frequently co-used emojis represent both the topic and the sentiment of a hashtag, and without any textual information, the bag-of-emoji representation itself can yield satisfying clustering results. Yet the unsupervised clustering task cannot quantify how strong the semantic association is. In the next section, we answer this question with a well-defined supervised learning task, namely hashtag prediction, which further quantitatively demonstrates the strength of the correlation.

Power of Emojis in Hashtag Generation

The hashtag prediction task has been studied ever since the hashtag functionality was introduced, and newer machine learning models keep being applied to increase the prediction accuracy. In recent years, seq2seq models have become more popular. These models treat each hashtag as a sequence of words or characters and transfer the hashtag prediction problem into a hashtag generation problem. By using a neural-network encoder to capture the textual information in the original Tweets and a neural-network decoder to generate word sequences as hashtags, seq2seq models can achieve state-of-the-art performance (Wang et al. 2019).

None of previous studies use emojis to enhance the predictions. By studying how emojis can be applied to improve

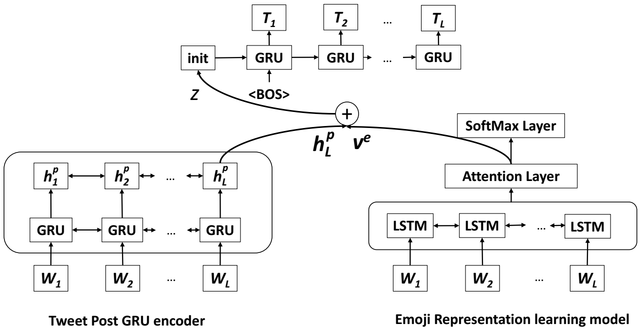

Figure 3: Our hashtag generation framework with a post encoder/decoder and representation learning model

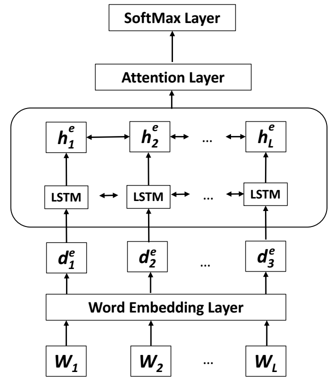

Figure 4: Network architecture for representation learning through emoji prediction

hashtag prediction, we not only quantify the semantic association between hashtags and emojis, but we may also examine their relationship with the Tweet text. Our hashtag generation model is also based on a seq2seq framework. However, we have designed a model that incorporates emoji information explicitly so as to easily evaluate how it can help improve the hashtag generation task.

A straightforward way to predict hashtag with emojis is to treat them as words and use the emoji information to predict the hashtag. Yet the approach may fail for two reasons. First, emojis carry richer semantics and often ambiguity compared with words (Ai et al. 2017), making it hard to represent individual emojis. Second, the approach may not help when there is no emoji in a Tweet.

To address these two caveats, we incorporate the emoji information through a representation learning model which learns an emoji representation for each Tweet. Specifically, the representation learning model predicts the emojis that occur in a Tweet through a neural network model, and a middle attention layer of the learning model is used as the emoji representation of the Tweet. The emoji representation is then fed into a seq2seq hashtag generation model. The whole generation model architecture is visualized in Figure 3, and we will introduce them each in detail.

Emoji Representation Learning

Before training the hashtag generation model, we learn the Tweet representations with emoji information and then incorporate representations into the generation model. Previous work applies representation learning of emojis for sentiment analysis (Chen et al. 2019), but the representation learning model is never used in hashtag prediction.

Representation learning model uses two bidirectional Long Short-Term Memory (Bi-LSTM) layers (Hochreiter and Schmidhuber 1997) and attention mechanism (Bahdanau, Cho, and Bengio 2014) to predict which emoji is used in the input sentence. We show the architecture of the representation learning model in Figure 4 (Chen et al. 2018b). The word embedding layer uses pre-trained Word2Vec embeddings (Mikolov et al. 2013) to transfer word sequence $[w_1, w_2, ..., w_L]$ to the embeddings $[d_1^e, d_2^e, ..., d_L^e]$, where L is the sequence length. We input the word embeddings into the LSTM layers. For the single word vector d_i^e, we concatenate the hidden layer output of both directions to construct the hidden state h_i^e, which is

$$h_i^e = [LSTM(d_i^e, \overrightarrow{h_{i-1}^e}), LSTM(d_i^e, \overleftarrow{h_{i+1}^e})].$$

In the attention layer, we use the simple approach (Bahdanau, Cho, and Bengio 2014) in common attention usage with a single weight matrix W_a.

$$a_i^e = \frac{exp(W_a u_i^e)}{\sum_{j=1}^{L} exp(W_a u_j^e)}$$

$$v^e = \sum_{i=1}^{L} a_i^e u_i^e$$

Here we concatenate the output of two Bi-LSTM layers and the word embedding layer to $u_i^e = [d_i^e, h_{i1}^e, h_{i2}^e]$, where h_{i1}^e and h_{i2}^e denote the first and second bi-directional LSTMs. Finally, the softmax layer takes the attention output v^e and then we minimize the loss of the softmax layer.

This representation learning process is conducted separately to avoid over-fitting. After learning the parameters of the network, we extract the output of the attention layer as emoji-powered Tweet representations.

Hashtag Generation Model

For the generation piece, we use a GRU post encoder to learn the text information in Tweets and combine hidden layer outputs with emoji-powered representations. The decoder contains both the text and emoji information of Tweets and generates hashtag words.

Post Encoder We use two bidirectional gated recurrent unit (Bi-GRU) (Cho et al. 2014) layers to encode the Tweet posts. We input the embeddings $[d_1^p, d_2^p, ..., d_L^p]$ after the word embedding layer into the Bi-GRU layers. Each embedding d_i^p is mapped into the hidden state h_i^p, which is:

$$h_i^p = [GRU(d_i^p, \overrightarrow{h_{i-1}^p}), GRU(d_i^p, \overleftarrow{h_{i+1}^p})].$$

We concatenate the output of the encoder hidden layer h^p_L and the emoji representation v^e to form a joint representation $[h^p_L; v^e]$. We add a linear layer with weight W_t to keep the dimension the same as the first decoder state and get representation z, which is:

$$z = W_t[h^p_L; v^e] + b_t.$$

We input the decoder with the vector z as the initial hidden state and a $<BOS>$ (sentence start) as the first token (as illustrated in Figure 3).

Decoder We apply attention-based GRU to generate word sequence T as a hashtag. The decoder shares the same word embedding layer with the post encoder. When predicting the t-th word in a hashtag, the GRU decoder generates a hidden state s_t and puts a global attention over the post encoder hidden state $h^p = [h^p_1, h^p_2, ..., h^p_L]$. We calculate the context vector c_t as:

$$\beta_i = \frac{exp(s_t W_{att} h^p_i)}{\sum_{j=1}^{L} exp(s_t W_{att} h^p_j)}$$

$$c_t = \sum_{i=1}^{L} \beta_i h^p_i$$

where W_{att} is the weight matrix of the attention layer and β_i is the attention score of the i-th word in the Tweet post. We concatenate the context vector c_t and decoder hidden state s_t and map the combined vector to the dimension of vocabulary. Finally, we add an additional softmax layer to generate the word probability distribution. We denote the proposed seq2seq model with emoji representations as **SEQ2SEQ-Emoji**.

Training and inference process In the training process, we use the cross entropy loss as our loss function and apply the teacher forcing mechanism to train our model. In the inference process, we apply the beam search algorithm to select the generated word for each time step.

Models for Comparison

To properly evaluate the effectiveness of the emoji representations in improving hashtag generation, we also implement several seq2seq baseline models and conduct an ablation study.

Base model and variations An intuitive baseline is to construct a base model by ablating the emoji representation module from the SEQ2SEQ-Emoji model. That is, we directly feed the hidden layer output of the post GRU encoder into the decoder without concatenating emoji-powered representations and we also delete the linear transformation between encoder and decoder. Depending on how emojis are preprocessed, we include three variations of Base models.

Base The first variation leaves all emojis "as is." That is, emojis are treated as the same type of tokens as words.

Base (Demoji) The second variation translates emojis into their description words, such as transferring 🔥 to "fire." We use the Demoji package to find the corresponding words.

Base (NoEmoji) In this third variation, we remove all the emojis from the Tweets. This variation is directly the post GRU encoder and decoder part of the SEQ2SEQ-Emoji model.

One may not be surprised to see that SEQ2SEQ-Emoji outperforms the Base model and variations. Yet a confounder to such improvement is the increased complexity of the SEQ2SEQ-Emoji model. The emoji representation module in SEQ2SEQ-Emoji increases the overall complexity, so it's not clear whether the improved performance over the Base models is attributed to the emoji information, or merely increased complexity. To rule out such a confounder, we design the SEQ2SEQ-Word model as follows.

SEQ2SEQ-Word The SEQ2SEQ-Word model shares the same structure as SEQ2SEQ-Emoji. However, for the emoji representation learning module, we change its predicting target from 64 emojis to 64 words. Since the 64 target emojis have different popularity, for a fair comparison, we control the popularity of the prediction target by replacing each target emoji with a word that has a similar document frequency (DF) in the Tweets corpus, even if they aren't semantically related. For example, instead of predicting the occurrence of the emoji 😂 (DF=531k in our dataset), we predict the occurrence of the word *one* (DF=489k). Instead of the emoji representation v^e, the learned "word" representation is concatenated to the encoder output for hashtag generation. In this way, SEQ2SEQ-Word shares the same model structure and complexity with SEQ2SEQ-Emoji without incorporating emoji information. It allows us to fairly evaluate how emojis help predict hashtags.

SEQ2SEQ-Conv This is the model implemented in Wang et al. (2019), which reports the state-of-the-art performance in hashtag generation. This model uses the co-attention mechanism to incorporate both Tweet post and conversation (replying messages) information in the encoder to enhance hashtag generation. We note that the conversation information is not available when the hashtag is added to the initial Tweet. Therefore, the direct comparison between SEQ2SEQ-Conv and SEQ2SEQ-Emoji is not fair. If we exclude the conversation module, SEQ2SEQ-Conv degenerates to the Base models.

Experiment Settings

We evaluate our proposed model on two real-world datasets.

Twitter 2011 We first adopt the dataset used in Wang et al. (2019), which allows us to compare with a more recent study. The dataset is a subset of the TREC 2011 microblog track,[3] and includes 44,793 Tweets with at least one of the 4,188 unique hashtags. However, since emojis were not

[3]https://trec.nist.gov/data/Tweets/

Model	Twitter 2011 F1@1	Twitter 2011 MAP@5	Twitter 2011 ROUGE F1	Twitter 2020 F1@1	Twitter 2020 MAP@5	Twitter 2020 ROUGE F1
Base (NoEmoji)	10.28	11.91	10.85	11.62	12.82	16.13
Base	-	-	-	11.82	12.84	16.16
Base (Demoji)	-	-	-	11.84	13.00	16.35
SEQ2SEQ-Word	10.53	12.25	10.75	11.92	13.18	16.91
SEQ2SEQ-Emoji*	**11.58**	**13.90**	**12.64**	**12.71**	**13.98**	**17.31**
SEQ2SEQ-Conv$^+$	12.29	15.49	13.73	-	-	-

Table 2: Comparison results between models on Twitter 2011 and Twitter 2020 dataset (in %). Max values in each column are bolded. $^+$: The results of the SEQ2SEQ-Conv model are cited from the original work. Note that SEQ2SEQ-Conv model uses additional conversation information, which is not fed to the first 5 models. *: The difference between SEQ2SEQ-Emoji and the baseline methods is statistically significant ($p < 0.05$) by McNemar's test.

	Twitter 2011	Twitter 2020
# of popular tags	52	41
# of unpopular tags	3,932	35,777
Tweets w/ popular tags	14,802	9,487
Tweets w/o popular tags	29,777	40,513

Table 3: Hashtag frequency distribution in Twitter 2011 and Twitter 2020.

yet supported during the data collection for TREC 2011, no emojis were observed in the Twitter 2011 dataset.

Twitter 2020 We construct a newer collection of Tweets where emojis are prevalent. Specifically, we sample 50,000 English Tweets from May 1st, 2020, to May 14th, 2020, which contain 35,818 unique hashtags. This dataset has a similar time span (half a month) and a similar size as the Twitter 2011 dataset. 8,053 Tweets (16%) in this set have emojis.

Besides the difference in emoji prevalence, we also note the difference in hashtag sparsity between the two datasets. There are far more unique hashtags (35,818 vs. 4,188 unique hashtags) in Twitter 2020, which prompts us to evaluate the model for hashtags of different popularity. We present the number of popular/unpopular hashtags and Tweets with/without popular hashtags in Table 3, where we define popular hashtags as those appearing more than 100 times in the dataset.

Evaluation Metrics We report standard information retrieval ranking metrics including the F-score for the Top 1 prediction (F1@1), and the mean average precision for Top 5 predictions (MAP@5). Besides, since we consider each hashtag as a sequence of words, we calculate ROUGE, a common evaluation metric in machine translation and summarization tasks (Lin 2004). We report ROUGE F1 for the top-ranked hashtag prediction.

Implementation Details We follow the procedure in the previous work to preprocess the dataset (Wang et al. 2019; Chen et al. 2019), and randomly split the labeled dataset into training, development, and test sets in the proportion of 8:1:1. We implement our SEQ2SEQ-Emoji and other baseline models based on the Pytorch framework. In the representation learning model, we choose 256-dimension Word2Vec embeddings and 512-dimension Bi-LSTM cells. The attention layer input $u_i = [d_i, h_{i1}, h_{i2}]$ is a 2304-dimension vector ($256 + 512 \times 2 + 512 \times 2$), and we extract the 2304-dimension output vector of the attention layer as the emoji representation vector. In the post GRU encoder, we set the vocabulary size to 30,000 and the dimension of word embeddings to 200. The encoders are comprised of two-payer 300 dimension Bi-GRU cells and for the decoder, we use one-layer GRU cells and also augment them with copy mechanism (Gu et al. 2016). The linear layer between post encoder and decoder has the output size of 300 dimension, the same dimension as the decoder GRU cell. For optimization, we use the Adam optimizer with learning rate 0.001 (Kingma and Ba 2014), with batch size 64 and 1000 iteration for epoch. In the training part, we adopt the early stopping strategy (Caruana, Lawrence, and Giles 2000) to avoid over-fitting based on the validation performance. SEQ2SEQ-Emoji achieves the highest validation accuracy at the second epoch. The dropout rate is 0.1 and the norm of gradient will be rescaled to 1 if $L2$-norm > 1. All the parameters in the model are uniformly initialized in [-0.1, 0.1] In the inference part, we adjust the beam size to 20 and the maximum hashtag length to 10.

We remove all the emoji tokens when training Base (No Emoji), SEQ2SEQ-Word and SEQ2SEQ-Emoji models, but keep them in Base and Base (Demoji). When training the SEQ2SEQ-Emoji model, we utilize the pretrained model from Felbo et al. (2017) as the representation model. They randomly extract 1.2 billion English Tweets containing the 64 most frequent emojis from January 1st 2013 to June 1st 2017 to learn the emoji representation model. When training the SEQ2SEQ-Word model, we follow the specifications in Felbo et al. (2017) using randomly sampled 4.5 million English Tweets in 2020 to retrain the word representation model. For training Tweets with multiple different emojis or word labels, we follow the preprocessing method in Felbo et al. (2017) and copy training Tweets to ensure each in-

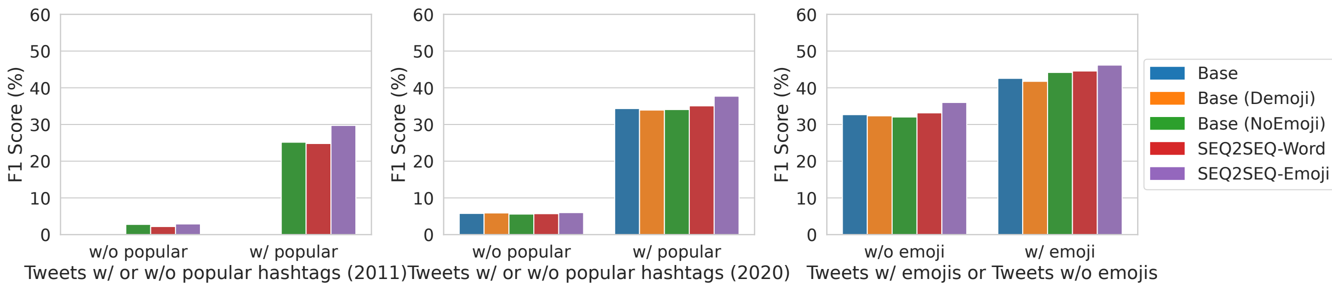

Figure 5: F1@1 on Twitter 2011 (the left subfigure), Twitter 2020 (the middle subfigure) in inferring hashtags with varying frequency and F1@1 on Twitter 2020 (the right subfigure) in inferring hashtags when Tweet post has emojis or not. SEQ2SEQ-Emoji model consistently performs better.

stance with a unique ground-truth emoji.

After obtaining the learned representation model, we train the GRU encoder and decoder of our model on Tweets and ground-truth hashtags of the Twitter dataset. For training instances with multiple ground-truth hashtags, we follow the method in Wang et al. (2019), copying them multiple times, each with one ground-truth instance.

During the inference stage, we input Tweets into both the encoder and the representation learning model. The representation learning model outputs emoji-powered representations and feeds them to the decoder, combined with encoder text representations.

Results and Discussion

Table 2 reports the results of our SEQ2SEQ-Emoji model and five other models for comparison. Since emojis were not yet supported in Twitter 2011, all Base models (that is, without emojis, with emojis, or with demojis) would yield the same result. The results of the SEQ2SEQ-Conv model are cited directly from Wang et al. (2019).

The SEQ2SEQ-Emoji model consistently outperforms all Base models and the SEQ2SEQ-Word model in all metrics (McNemar's test (Dietterich 1998) is performed on both the Top 1 prediction and Top 5 predictions and the differences are all statistically significant at the 5% level). Among the Base models, different preprocessing procedures regarding emojis do not make much difference in the performance - that is, removing emojis, treating emojis as a token, or replacing emojis with words using Demoji. Although the SEQ2SEQ-Word model has an extra representation module and increased complexity, its performance does not differ much from the Base models. These observations suggest that utilizing emoji through the representation model can improve hashtag generation, and the improvement cannot be attributed merely to increased complexity.

SEQ2SEQ-Conv still shows better results in the Twitter 2011 dataset. As noted earlier, the comparison is not fair because it incorporated additional information from other tweets in the same conversation. Still, the comparison evidences the scale of improvement of SEQ2SEQ. That is, compared with Base (NoEmoji), incorporating additional Tweets in the same conversation increases F1@1 from 10.3 to 12.3, while leveraging the emoji representation can already boost F1@1 to 11.6.

Beyond the overall performance metrics, we are more interested in understanding where the improvement comes from. To this end, we evaluate our model on different subsets of Tweets, based on hashtag popularity and on whether emojis are used in the Tweet.

Hashtag Popularity and Model Performance Different hashtags enjoy different levels of popularity, as shown in Table 3. Indeed, most tweets contain only rare hashtags, especially for Twitter 2020, which implies a sparsity issue that may affect the model performance. Therefore, we calculate the F1@1 scores separately for Tweets containing popular hashtags and those without. Unsurprisingly, for all models, the F1@1 scores are pretty low for Tweets with rare hashtags. The SEQ2SEQ-Emoji model outperforms other models both for Tweets with popular hashtags and for those without; however, we see more improvement when predicting popular hashtags. This is expected as there are more ample training examples for the popular hashtags.

Tweets with(out) Emoji and Model Performance One advantage of our SEQ2SEQ-Emoji model is that the learned emoji representation is able to boost hashtag prediction even for Tweets that do not have emojis. Is this advantage confirmed with data? On the other hand, if the original Tweets already contain emojis, does the model still outperform the baselines? To answer these questions, we calculate the F1@1 scores separately for Tweets containing emojis and those without in the Twitter 2020 dataset. Since previous discussion shows a huge performance gap between Tweets with popular hashtags and those without, we limit the experiment only to Tweets with popular hashtags (hashtag frequency > 100) to reduce noise. As shown in the right panel of Figure 5, the SEQ2SEQ-Emoji model achieves the highest score regardless of whether the Tweets have emojis or not. We also notice that among the variations of the Base models, the Demoji one under-performs on Tweets with emojis, which suggests that simply replacing emojis with words might indeed distort the semantics and hurt the prediction.

Tweet	Predicted Emoji	Model	Top two outputs
U gotta it just makes life more fun haha	😜🎶😉	Base (NoEmoji)	twitter after dark; a good girlfriend
		Emoji	**laugh at yourself**; laugh at
I do what I love	💁‍♀️😌😏	Base (NoEmoji)	yalitchat; scariest words ever
		Emoji	**i am proud to say**; random
Obama is my president	😎✋👏	Base (NoEmoji)	sotu; scariest words ever
		Emoji	**i am proud to say**; i am proud to
Enjoy your day mom 💕 You may be a mother or a stepmom, an aunt or have a further baby 💕💕	💁😘😉	Base	happy baekhyun day; mothers day
		Emoji	**happy mothers day**; mothers day
Here we go! two absolute savages about to throw down 😍	👊😈💪	Base	the last dance; ns10v10
		Emoji	**ufc 249**; the last dance

Table 4: Model outputs for Tweet examples. The ground-truth hashtags (used in the original Tweets) are highlighted in bold.

A Case Study Beyond the evaluation metrics, however, we are more interested in understanding how the emoji representation improves hashtag generation and why directly utilizing emoji tokens can not make the improvement. To this end, we conduct a case study on 5 selected posts, where the Base (NoEmoji) or Base models fail but the SEQ2SEQ-Emoji model predicts correctly. Table 4 lists the selected Tweets and the top two hashtags generated by models. Note that the first three Tweets do not use emojis and the last two Tweets use emojis. We also report the emojis predicted by the emoji representation model as the surrogate of the emoji representation.

By comparing the first three predictions, we see that the Base-model prediction is more relevant in topic, while the Emoji-model prediction is better at capturing the sentiment of the posts, and such sentiment is aligned with the predicted emojis. In the first Tweet, "*#TwitterAfterDark*" usually marks non-public and not-safe-for-work content, which seems relevant to the post, while "*#LaughAtYourself*" better reflects the joking sentiment of the Tweet, as are the predicted emojis of joking 😜 and winking faces 😉. For the second Tweet, "*#YALICHAT*" (Young African Leaders Initiative Chat) could indicate the topic of the Tweet (although we do not know the intention of the author for sure). The Emoji-model predicts "*#ImProudToSay*," which shares the same proud sentiment of the Tweet and is also aligned with the shy-face 😌 and tipping-hand emoji 💁‍♀️. In the last Tweet, the sentiment (of supporting Obama) is expressed implicitly, yet both the predicted emojis (sunglass-smiling 😎 and clapping 👏) and the Emoji-model prediction "*#ImProudToSay*" capture the supportive sentiment, while "*#SOTU*" (State of the Union) is more relevant to the Tweet topic-wise.

In the last two posts, the Base model (treating emojis as normal tokens) fails to predict the correct hashtags, while the SEQ2SEQ-Emoji model predicts the hashtags accurately even though it predicts the wrong emojis. It is likely that the Base model overfits the cooccurrence between emojis and trending hashtags, while the emoji representation model captures the semantic association in a smoother way. In the fourth Tweet, the two-heart emoji 💕 cooccurs frequently with the K-pop idol group Baekyun, which misleads the Base model to the wrong prediction. The SEQ2SEQ-Emoji model, on the other hand, predicts emojis with the aligned sentiment, which helps with generating the right hashtags for Mothers Day. Similarly, in the last Tweet, the Base model predicts "*#TheLastDance*" (documentary of Michael Jordan) and "*#NS10V10*" (musical artists competition), which is likely due to their frequent cooccurrence with the heart-eye emoji (😍). The predicted emojis (👊 and 😈) are more aligned with the phrase "absolute savages" and "throw down," which helps with ranking the boxing hashtag "*#UFC249*" first. The two examples suggest that the SEQ2SEQ-Emoji model indeed leverages the nuanced semantics beyond cooccurrence to predict the right hashtag.

Limitations and Implications

Our work is limited to Twitter and the hashtag prediction task is further limited to English Tweets, both of which may limit the generalizability of the conclusions. Future work may examine more platforms (such as Instagram) and other languages (such as Sina Weibo).

Although our proposed SEQ2SEQ-Emoji model achieves some promising results on improving hashtag generation, we have not yet applied more advanced machine learning models (such as BERT) which may further improve the hashtag generation performance. For this work, however, the contribution of such advanced models would be limited to improving metrics without revealing many additional insights about the relationship between emoji, hashtag, and the word context.

Results from our experiment provide new insights about the semantic relationship between texts, emojis, and hashtags. Below we discuss the broader implications of our work, which we hope can motivate future research.

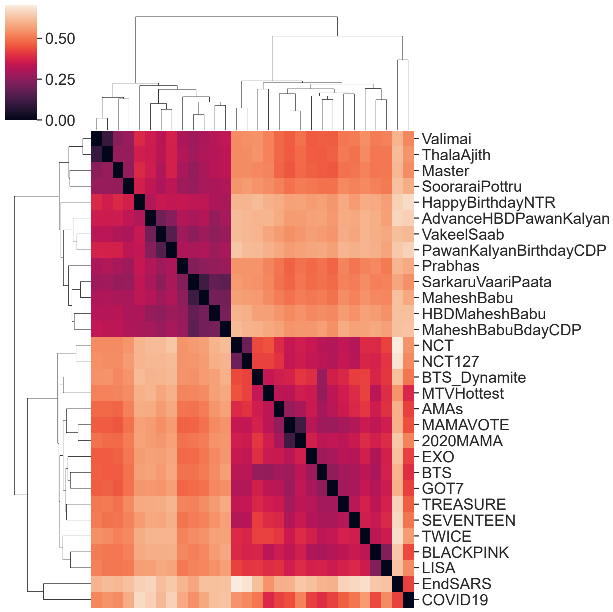

Figure 6: Clusters of top 30 frequent hashtags from 2020 full-year English Tweets.

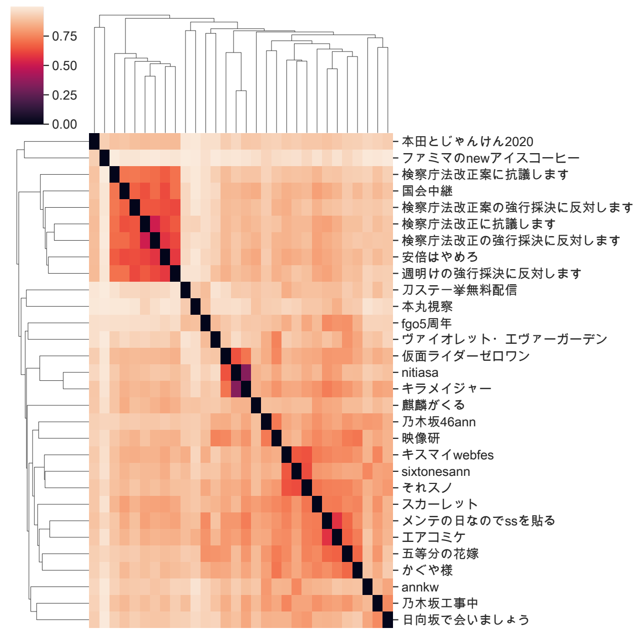

Figure 7: Clusters of top 30 frequent hashtags from Japan Trending list of May 2020.

First, despite cultural and demographic differences in their usage, emojis are widely considered a ubiquitous language across countries and languages, which allows us to understand hashtags and their semantics across different languages. For example, in our collected dataset, English and Korean hashtags for idols are all associated with ✨ (sparkles) and 🗳️ (ballot box), and previous research has pioneered in cross-lingual sentiment classification (Chen et al. 2019).

Second, the free-formed hashtags have infinite combinations, and the trending hashtags rise and fall rapidly. However, emojis have a limited "vocabulary" and may be utilized to understand trending hashtags from a semantic perspective. The clustering analysis in Section *Co-occurring Emojis in Trending Hashtags* provides initial evidence that an emoji-only representation can effectively cluster trending hashtags into topics. Beyond hashtags, the rich semantics embedded in emojis may also be used in further applications, such as meme and rumor detection (Zhao, Resnick, and Mei 2015).

Conclusion

In this work, we present the first empirical study of the association between the two new-era languages, emojis and hashtags. We show that there exist strong associations in their semantics, particularly topics and sentiment, and we show how the association can be used in downstream machine learning tasks. We propose a generation model that learns emoji representations for Tweets to improve hashtag prediction. The promising results suggest that emojis can be used in further study of hashtags and social media at large.

Acknowledgements

The authors would like to thank Jordan Boyd-Graber, Vanessa Frias-Martinez, Paiheng Xu, Xuan Lu and Qiaozhu Mei for helpful discussions. We thank the Foreseer research group at the University of Michigan for the help in preparing the Twitter 2020 dataset.

Appendix

A Additional Clustering Results

We extract the top 30 most frequent hashtags and their co-occurring emojis from the 2020 full-year English Tweets and the Twitter Japan Trending list of May 2020 respectively. The hierarchical clustering is applied on emoji-PPMI embeddings and the clustering results are visualized in the heatmap of Figure 6 and Figure 7.

Four large hashtag clusters appear in the heatmap of Figure 6. When looking into every hashtag cluster, the topic of the first and second cluster (the upper left corner) is about Indian celebrities, and the third cluster in the middle is about Korean idols. The final cluster in the bottom right corner contains *#COVID19* and *#ENDSARS*. The top 30 most frequent hashtags of the whole set 2020 English Tweets mostly concentrate on idols and celebrities, besides "COVID19." In the K-pop cluster, all hashtags are related to the Korean idols, but our emoji-PPMI embeddings can accurately discover hashtags describing the same idol and gather them into a small cluster. For example, hashtags about the *NCT* idol group, *#NCT* and *#NCT127* are embedded closely in the upper left corner.

In Figure 7, we can identify five tiny clusters, and the clusters of Japanese hashtags are accurate and follow our common sense. For example, except for hashtags about prosecutor change news (#検察庁法改正案に抗議します) on the upper left corner, hashtags about "Kamen Rider Series," a kind of TV series, (#仮面ライダーゼロワン) are also located in the middle. We can still observe that similarity between hashtags of the same event is higher than different but similar events. For example, compared with the "Kamen Rider Series" cluster in the middle, the "boy band" cluster (below the "Kamen Rider Series" cluster) containing three hashtags about three boy bands, #キスマイWEBFES, #sixtonesann and #それスノ has weaker similarity inside the cluster.

References

Ai, W.; Lu, X.; Liu, X.; Wang, N.; Huang, G.; and Mei, Q. 2017. Untangling emoji popularity through semantic embeddings. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11.

An, J.; and Weber, I. 2016. # greysanatomy vs.# yankees: Demographics and Hashtag Use on Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 10.

Bahdanau, D.; Cho, K.; and Bengio, Y. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* .

Bar-Joseph, Z.; Gifford, D. K.; and Jaakkola, T. S. 2001. Fast optimal leaf ordering for hierarchical clustering. *Bioinformatics* 17(suppl_1): S22–S29.

Barbieri, F.; Kruszewski, G.; Ronzano, F.; and Saggion, H. 2016. How Cosmopolitan Are Emojis? Exploring Emojis Usage and Meaning over Different Languages with Distributional Semantics. In *Proceedings of the 24th ACM International Conference on Multimedia*, MM '16, 531–535. New York, NY, USA: Association for Computing Machinery. ISBN 9781450336031.

Caruana, R.; Lawrence, S.; and Giles, C. 2000. Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping. volume 13, 402–408.

Chang, Y.; Wang, X.; Mei, Q.; and Liu, Y. 2013. Towards twitter context summarization with user influence models. In *Proceedings of the sixth ACM international conference on Web search and data mining*, 527–536.

Chen, Y.; Yuan, J.; You, Q.; and Luo, J. 2018a. Twitter Sentiment Analysis via Bi-sense Emoji Embedding and Attention-based LSTM. *Proceedings of the 26th ACM international conference on Multimedia* .

Chen, Z.; Lu, X.; Ai, W.; Li, H.; Mei, Q.; and Liu, X. 2018b. Through a Gender Lens. *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18* .

Chen, Z.; Shen, S.; Hu, Z.; Lu, X.; Mei, Q.; and Liu, X. 2019. Emoji-Powered Representation Learning for Cross-Lingual Sentiment Classification. In *The World Wide Web Conference*, WWW '19, 251–262. New York, NY, USA: Association for Computing Machinery.

Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* .

Church, K.; and Hanks, P. 1990. Word association norms, mutual information, and lexicography. *Computational linguistics* 16(1): 22–29.

Cruickshank, I. J.; and Carley, K. M. 2020. Characterizing communities of hashtag usage on twitter during the 2020 covid-19 pandemic by multi-view clustering. *Applied Network Science* 5(1): 1–40.

Davidov, D.; Tsur, O.; and Rappoport, A. 2010. Enhanced sentiment learning using twitter hashtags and smileys. In *Coling 2010: Posters*, 241–249.

DeMasi, O.; Mason, D.; and Ma, J. 2016. Understanding communities via hashtag engagement: A clustering based approach. In *Tenth International AAAI Conference on Web and Social Media*.

Dietterich, T. G. 1998. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural computation* 10(7): 1895–1923.

Eisner, B.; Rocktäschel, T.; Augenstein, I.; Bošnjak, M.; and Riedel, S. 2016. emoji2vec: Learning emoji representations from their description. *arXiv preprint arXiv:1609.08359* .

Felbo, B.; Mislove, A.; Søgaard, A.; Rahwan, I.; and Lehmann, S. 2017. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* .

Feng, W.; Zhang, C.; Zhang, W.; Han, J.; Wang, J.; Aggarwal, C.; and Huang, J. 2015. STREAMCUBE: Hierarchical spatio-temporal hashtag clustering for event exploration over the Twitter stream. In *2015 IEEE 31st International Conference on Data Engineering*, 1561–1572. doi: 10.1109/ICDE.2015.7113425.

Ge, J. 2019. Emoji Sequence Use in Enacting Personal Identity. In *Companion Proceedings of The 2019 World Wide Web Conference*, WWW '19, 426–438. New York, NY, USA: Association for Computing Machinery. ISBN 9781450366755.

Gong, Y.; Zhang, Q.; and Huang, X. 2015. Hashtag Recommendation Using Dirichlet Process Mixture Models Incorporating Types of Hashtags. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 401–410. Lisbon, Portugal: Association for Computational Linguistics.

Gu, J.; Lu, Z.; Li, H.; and Li, V. O. 2016. Incorporating copying mechanism in sequence-to-sequence learning. *arXiv preprint arXiv:1603.06393* .

Highfield, T. 2018. Emoji hashtags//hashtag emoji: Of platforms, visual affect, and discursive flexibility. *First Monday* .

Hochreiter, S.; and Schmidhuber, J. 1997. Long Short-Term Memory. *Neural Computation* 9(8): 1735–1780. ISSN 0899-7667.

Hong, L.; Ahmed, A.; Gurumurthy, S.; Smola, A. J.; and Tsioutsiouliklis, K. 2012. Discovering geographical topics in the twitter stream. In *Proceedings of the 21st international conference on World Wide Web*, 769–778.

Hu, T.; Guo, H.; Sun, H.; Nguyen, T.-v.; and Luo, J. 2017. Spice up your chat: the intentions and sentiment effects of using emojis. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* .

Li, J.; Liao, M.; Gao, W.; He, Y.; and Wong, K.-F. 2016. Topic Extraction from Microblog Posts Using Conversation Structures. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2114–2123. Berlin, Germany: Association for Computational Linguistics.

Lin, C.-Y. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*, 74–81. Barcelona, Spain: Association for Computational Linguistics.

Lou, Y.; Zhang, Y.; Li, F.; Qian, T.; and Ji, D. 2020. Emoji-Based Sentiment Analysis Using Attention Networks. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 19(5). ISSN 2375-4699.

Lu, X.; Ai, W.; Liu, X.; Li, Q.; Wang, N.; Huang, G.; and Mei, Q. 2016. Learning from the Ubiquitous Language: An Empirical Analysis of Emoji Usage of Smartphone Users. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, 770–780. New York, NY, USA: Association for Computing Machinery. ISBN 9781450344616.

Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* .

Miller, H.; Thebault-Spieker, J.; Chang, S.; Johnson, I.; Terveen, L.; and Hecht, B. 2016. "blissfully happy" or "ready to fight": Varying interpretations of emoji. Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016, 259–268. AAAI press.

Olteanu, A.; Weber, I.; and Gatica-Perez, D. 2016. Characterizing the demographics behind the# blacklivesmatter movement. In *2016 AAAI Spring Symposium Series*.

Ordonez, C.; and Omiecinski, E. 1999. Discovering association rules based on image content. In *Proceedings IEEE Forum on Research and Technology Advances in Digital Libraries*, 38–49. IEEE.

Park, J. H.; Xu, P.; and Fung, P. 2018. Plusemo2vec at semeval-2018 task 1: Exploiting emotion knowledge from emoji and #hashtags. *arXiv preprint arXiv:1804.08280* .

Sari, Y. A.; Ratnasari, E. K.; Mutrofin, S.; and Arifin, A. Z. 2014. User emotion identification in Twitter using specific features: hashtag, emoji, emoticon, and adjective term. *Jurnal Ilmu Komputer dan Informasi* 7(1): 18–23.

Suttles, J.; and Ide, N. 2013. Distant supervision for emotion classification with discrete binary values. In *International Conference on Intelligent Text Processing and Computational Linguistics*, 121–136. Springer.

Tauch, C.; and Kanjo, E. 2016. The Roles of Emojis in Mobile Phone Notifications. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, UbiComp '16, 1560–1565. New York, NY, USA: Association for Computing Machinery. ISBN 9781450344623.

Tsur, O.; and Rappoport, A. 2012. What's in a Hashtag? Content based Prediction of the Spread of Ideas in Microblogging Communities. 643–652. doi:10.1145/2124295.2124320.

Wang, X.; Wei, F.; Liu, X.; Zhou, M.; and Zhang, M. 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*, 1031–1040.

Wang, Y.; Li, J.; King, I.; Lyu, M. R.; and Shi, S. 2019. Microblog hashtag generation via encoding conversation contexts. *arXiv preprint arXiv:1905.07584* .

Yan, X.; Guo, J.; Lan, Y.; and Cheng, X. 2013. A Biterm Topic Model for Short Texts. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, 1445–1456. New York, NY, USA: Association for Computing Machinery. ISBN 9781450320351.

Yang, L.; Sun, T.; Zhang, M.; and Mei, Q. 2012. We Know What @you #tag: Does the Dual Role Affect Hashtag Adoption? In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, 261–270. New York, NY, USA: Association for Computing Machinery. ISBN 9781450312295.

Ye, Z.; Hashim, N. H.; Baghirov, F.; and Murphy, J. 2018. Gender differences in Instagram hashtag use. *Journal of Hospitality Marketing & Management* 27(4): 386–404.

Zhang, Q.; Wang, Y.; Gong, Y.; and Huang, X.-J. 2016. Keyphrase extraction using deep recurrent neural networks on twitter. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, 836–845.

Zhang, R.; Li, W.; Gao, D.; and Ouyang, Y. 2012. Automatic twitter topic summarization with speech acts. *IEEE transactions on audio, speech, and language processing* 21(3): 649–658.

Zhang, Y. 2019. Language in Our Time: An Empirical Analysis of Hashtags. In *The World Wide Web Conference*, WWW '19, 2378–2389. New York, NY, USA: Association for Computing Machinery. ISBN 9781450366748.

Zhao, Z.; Resnick, P.; and Mei, Q. 2015. Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts. In *Proceedings of the 24th International Conference on World Wide Web - WWW '15*, 1395–1405. New York, New York, USA: ACM Press.