Hashtag Recommendation System: HOOHLE

ÁNGEL GABRIEL RIVAS*, LUIS ANTONIO ORTEGA*, IGNACIO SALINAS*, and ANTONIO COÍN*, Autonomous University of Madrid, Spain

ABSTRACT

Nowadays, social networks and micro-blogging sites have undoubtedly become part of our everyday life. Among others, Twitter is one of the platforms with the highest engagement capacity, which has developed into an influential networking and knowledge gathering medium used by millions of people all over the world to share experiences in real time. One of the key identities of Twitter is the *hashtag*: a keyword preceded by the hash symbol (#) that helps users coordinate their tweets and search through a large amount of data. However, it can sometimes be difficult to choose an appropriate hashtag that accurately reflects the intended meaning of the message and at the same time makes the tweet have a meaningful impact, measured in the quantity of *retweets*, *likes* and interactions. In this work we present a novel hashtag recommendation system, *HOOHLE*, which combines standard recommendation techniques with automatic text generation and popularity prediction methods to produce potentially trending hashtags. Our system is aimed at relieving users of the burden of choosing a suitable hashtag to accompany their tweet, while simultaneously increasing its popularity and its chances of going viral. Differently from existing approaches, to accomplish this a new system is proposed, in which candidate hashtags are obtained both from current trends found in related tweets and a pool of newly generated hashtags via an embedding-based generative tool. All in all, our method presents an innovative approach that takes into account the context of the tweet to automatically generate new hashtags that will potentially be trending.

KEYWORDS

Hashtag recommendation, text generation, popularity prediction, topic modeling, social media.

1 INTRODUCTION

Twitter is one of the most common micro-blogging sites on the Internet, with over 187 million active users posting short messages or *tweets* every day [1]. The character limit of a tweet (280 characters) encourages users to casually upload new content, whereas conventional blogging requires more time investment to write a new message. Furthermore, since more and more people own mobile devices around the world, Twitter events are becoming increasingly big, resulting in over 500 million tweets being posted to the social network each day. The downside of Twitter's success is that users can become quickly overwhelmed by the large amount of data to which they are exposed on a daily basis [2]. As a mechanism to combat this issue, Twitter users have assimilated the usage of hashtags to filter which content they are exposed to. A *hashtag* is a word or expression that is preceded by the hash symbol # and can be used anywhere in the body of a tweet. Tagging tweets with hashtags allows trends to spread rapidly among millions of users, forming an immediate group of people with common interests. Since the introduction of Twitter's hashtag search tool, many individual users and company marketers have begun using tagging to group tweets into similar discussions in order to facilitate searching for posts using the associated hashtags.

One of the most attractive features of the microblogging sites is the fact that events of widespread interest to a group of users result in a surge in real-time mentions as they occur. For example, users live-tweet about sport events, discuss about breaking news, or celebrate certain events on a memorial day, among others. As a

*All authors contributed equally to the paper.

Authors' address: Ángel Gabriel Rivas, angelgabriel.rivas@estudiante.uam.es; Luis Antonio Ortega, luisa.ortega@estudiante.uam.es; Ignacio Salinas, ignacio.salinas@estudiante.uam.es; Antonio Coín, antonio.coin@estudiante.uam.es, Autonomous University of Madrid, Spain, 13 Einstein St. (Cantoblanco Campus), Madrid, 28049.

consequence, there is a spike in behavior correlated with the event in question, resulting in a social trend. While these social trends can reveal what is going on at the very moment and help discover certain events early on, a list of trends includes just the set of terms that are being mentioned outstandingly at the moment. One of the best tools to join current trends and achieve more popularity is the hashtag: similar tweets can be found by using it, and in this way more users are able to see the publications. In some social networks such as Twitter or Instagram, it is possible to follow hashtags or topics: although there is no follow-up relationship with other users, they will see the content tagged with the hashtags followed, so that if the appropriate terms are used, a high degree of popularity can be achieved. The importance of the hashtags in the tweets popularity has been analyzed and established in many research papers, e.g. [3–6].

In this new social atmosphere the trending phenomenon has arisen: some hashtags gain traction rapidly while other similar hashtags remain unseen and become irrelevant. Controlling and predicting this behaviour has become vital in many businesses [7–9], political campaigns [10] or discussions in general. There are even hundreds of jobs are being generated as a community manager: people who manage social networks accounts and try to get better opinions from users, promote products, influence voting decisions, etc. In this work, a new system for automatically recommending customized trending hashtags based on users' tweets is proposed, with the objective of helping those who intend to boost their popularity (regular users, community managers, politicians, etc.) to increase the reach of their tweets.

There is no denying that the development of hashtag recommender systems has attracted a great deal of research in recent years [2, 11–17]. In some of them (e.g. [2, 13]) the topic of the tweet is used to extract related and existing hashtags. But none of them can create and recommend new and innovative hashtags that can increase the tweet's impact and even anticipate new tendencies that have never appeared before. This is why, rather than limiting the collection of suggested hashtags and developing a recommendation system that learns to choose the best hashtag from that collection, we suggest a text generation approach in which topics, and hence hashtags, can be recommended in a domain-independent fashion. As a result, the recommendation system is more flexible and performs better suggestions compared to more classical approaches such as [18]. For example, it is very unlikely that a hashtag like #blacklivesmatter, which had never had any relevant impact before the year 2019, would have been suggested by the existing recommendation systems if no one had already written it. For the purposes of prediction and recommendation, our raw feature set is the collection of words and images for each tweet, which are preprocessed as explained in Section 4.

Differently from existing similar frameworks, our system allows the visualization of hashtags ordered according to the popularity prediction made. This paper's presented framework not only chooses and/or generates the hashtags and returns them to the user, but the text strings obtained are ordered with the help of a popularity prediction module and are returned in descending order of predicted popularity. There are some works which rank different tweets or users according to their popularity using different metrics [19–22], but all of them do it with already published tweets. In this way, the user will be able to choose hashtags not only based on the text displayed but also on the impact that the new tweet could potentially have. Another important aspect is that all this is achieved by simply entering the text in a text box and including the # symbol in the part where we want to get the hashtag, making it simple and straightforward. The importance of providing a clean and user-friendly interface to our system is critical, as analyzed for example in [23, 24]. In this way, any kind of stakeholders (i.e individual users, companies or governments) will be able to improve their publications without any effort or difficulty. The expected user workflow is shown schematically in Figure 1.

As a metric to measure a single tweet's popularity, will use a mix of likes (l), retweets (r) and interactions (i), similar to [25]. Specifically, the proposed formula for a tweet t is:

$$popularity(t) = r(t) + l(t) + i(t).$$

This is explained with more detail in Section 4.5.

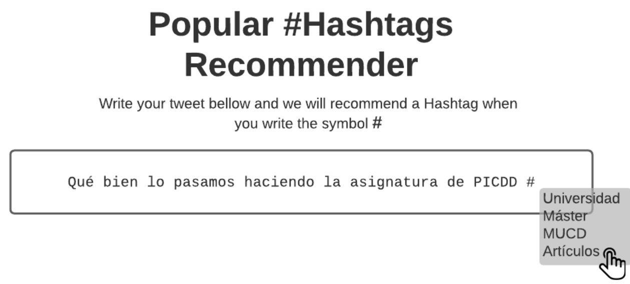

Fig. 1. System usage example.

In summary, this paper's main contributions are:

(1) Automatic text generation to produce new hashtags.
(2) Predict the tweet's impact using the different hashtags proposed *before* they are published.

The rest of this paper is organized as follows: Section 2 presents the related work on the field and an overview of the state of the art in Twitter trend prediction and hashtag recommendation. In Section 3 we study the main sources of data collected for experimentation. Then, Section 4 formulates the problem, objectives and methodology followed, along with the proposed framework details and the submodules developed. The outcome of the experimentation and the results are provided in Section 5, and finally Section 6 presents the conclusions on the work done and briefly addresses the next steps, further applications and possible extensions.

1.1 Background

We briefly define the terminology used to describe Twitter's services and functions.

- *Tweet*: A tweet is a 280-character-limit message (140 characters before October 2020) sent by a user on Twitter. To tweet is often used as a verb to describe the act of sending certain tweets. Unless an account is set to private, all tweets are publicly visible by default.
- *Follow*: Subscribing to other users' tweets on Twitter is referred to as a follow. Unless the user has set tweet protection or has blocked the user, users may become followers of others without their permission. Following, unlike friending in other social networks, is not reciprocal.
- *User*: A Twitter username, @user, identifies a user. Users may mention other users in their tweets by including that user's name. When a user is listed, the Twitter API notifies the user, and the mentioning tweet is shown on the user's feed.
- *Retweet*: If people wish to post someone else's message with their own fans, they will retweet it. Retweeting is a built-in feature on Twitter that spreads a tweet verbatim and marks it as a retweet. Users will retweet manually by copying the body of a message.
- *Hashtag*: A hashtag is a prefixed keyword that may be used anywhere in the body of a tweet to categorize or label words/phrases as keywords relevant to their tweets. Users can display all Tweets containing a hashtag by clicking on it in a Message. Extremely popular hashtags often become trends.
- *Trend*: To assist users in discovering emerging trends on Twitter, Twitter shows a collection of instantly popular keywords and hashtags on the user's homepage. Given this, from now on we will argue that any given hashtags is *trending* if twitter marks it as such.

2 RELATED WORK

Twitter is one of the most common micro-blogging sites on the Internet. Thus, there is a lot of work around it, most of it related to classification techniques for opinion mining and sentiment analysis [26]. A fairly big part of them use models based on LDA for topic modeling (around 27%, e.g. [27–30]), which is the first building block in our system. Topic modeling concerns itself with the extraction of common topics from a collection of documents, or in our case, tweets. On the other hand, a surprisingly but encouraging realization was the small number of articles that delved into text generation in hashtag recommendation, only about 2% of them [31]. This proves that a deep look into these topics is certainly not superfluous. Our work is focused in this gap in research found, i.e., hashtag recommendation via automatic text generation combined with popularity prediction in Twitter. The remaining of this section describes the current state of the art in different sub-areas inside our main area of research.

2.1 Predicting popular content

With the arrival and consolidation of social networks, the number and quality of research studies focused on them has risen accordingly, with several studies concentrating on the issue of predicting popular content. This problem can be tackled on different social networks: in [32] the authors focused on popularity description on Instagram Posts, in [33] on YouTube Videos and in [34] on Facebook Posts. In this paper the application domain considered is Twitter, where users and companies can greatly benefit from tools to predict whether their tweets are going to be popular, and adjust the content accordingly.

Among the reviewed existing work in popularity prediction (without recommendation), relatively little has been done to study the relationship between topics and hashtags to predict tweets popularity. For example, *Daga et al.* [35] rely only on the content of the tweet to make predictions, while *Silva et al.* [27] measure the relevance of influencers activity to predict popular tweets and suggest that the source of information has a greater impact than the content itself. They achieve an F1-score of 86.7% in a custom-made dataset of over 2 million tweets. In [36], the authors proposed a Recurrent Neural Network based hybrid model with Feature Concatenation (RNN-HMFC), an approach that allowed them to predict the future popularity of online posts with an F1-score of 91.60% in the SEISMIC dataset[1]. This is a dataset used in several popularity prediction research studies (e.g. [37, 38]), which consists on approximately 3 billion tweets and a little over 160.000 tweets labeled as being popular. Moving on, in [39] the main objective was to predict the popularity of hashtags on a 5 step scale. They used both content and contextual (hand-crafted) features for the model, achieving an F1-score of 86.10% in their custom-made dataset (collected from Twitter) with 230.000 tweets. Also, a recent study on predicting a tweets popularity is [40], which forecasts the popularity for the next day using the number of mentions in the past week of a hashtag. They repeated their experiment in two different datasets consisting on 200k and 700k extracted from Twitter, and also in a third article-based dataset consisting on 150k documents [41]. They achieve a *Root Mean Square Error* of 10.06, 83.299 and 3.463 respectively, outperforming other methods and techniques as shown in their comparison.

Various popularity metrics have been used in previous studies, such as the number of likes [42]. The number of likes is more accurate on a single user or specific datasets [43, 44] , whereas the engagement rate is more suitable for larger ones [45]. Among the studies reviewed there were discretized metrics, such as popular/unpopular [46] and viral/non-viral [47]. Other metrics were intrinsic popularity [48] and view count per day [49].

Regarding commonly used datasets for this task, the majority of studies reviewed use a custom dataset extracted from the twitter API over a specific period of time. Usually, tweets are gathered at the time of performing the experiment, thus reflecting the current trends and state of the community. This is explained with more detail in Section 3.

[1] http://snap.stanford.edu/seismic/

2.2 Topic modeling

In order to achieve a good hashtag recommendation system, one of the main tasks in our work is the development of a comprehensive topic modeling system. Topic modeling is an unsupervised machine learning technique for detecting words and phrase patterns within a set of documents (see Section 4 for details), with Latent Dirichlet Allocation (LDA, [50]) being one of the most common approaches. In this way the related topics for a given tweet can be retrieved, and thus finding related hashtags will be easier and more accurate.

Since the creation of the first social network the analysis of topics in Twitter and various micro-blogging messages is a research area with high and rapidly growing interest within the academic community. For example, in [51], the authors trained a specific topic model method with LDA in order to obtain a topical distribution for a collection of over 3 million tagged documents. Then, they chose the top word of the dominant topics in the tweet as their recommended tag, improving the F-score obtained with association rules in almost 20 percentage points. In [52] the authors trained a similar model with almost 2 million tweets extracted from the Twitter API, using LDA with 200 categories. The results on a smaller test set, manually evaluated by two different persons, were fairly accurate. Also, in *Zhao et al.* [13] the authors used an LDA-driven technique to identify users that were the most similar to a single user, and then recommended hashtags to the user based on the hashtags of the similar users. They use a custom dataset of over 1.5 million tweets an improve several methods such as EUCF and T-LDA (see [53]) in a hit-rate based metric.

In recent years, researchers have concentrated on issues such as Twitter message summarization and subject identification, as well as mass clustering of tweets. For example, TweetMotif by *O'Connor et al.* [54] takes an unsupervised approach to message clustering. One issue with this analysis is that O'Connor does not include success metrics for the method or comment on the approach's generalizability, probably due to a lack of applicable performance metrics. Another application of unsupervised methods on Twitter messages was a study by *Eisenstein et al.* [55], focused on predicting the geo-location of a tweet based on the text in the tweet, which made use of the geo-tagged information in the tweets.

Previous research has also explored the use of supervised methods for topic categorization of short social messages. For example, *Ranganath et al.* [56] presented a system that detects a speaker's intent to flirt using a spoken corpus of speed-dates; however, their dataset requires human transcription and heavy annotations. *Dela Rosa and Ellen* [57] also completed a set of experiments on classification of military chat posts, another form of short social messages, with algorithms such as support vector machines, k-nearest neighbors, Rocchio, and Naïve Bayes. However, like *Ranganath et al.*, their approach relies heavily on annotated data, which is usually not available for large scale micro-blogging messages, like those encountered in Twitter.

2.3 Hashtag recommendation system

Automatic hashtag recommendation is a novel trend that has attracted researchers' attention in recent years, as demonstrated in a survey by *Amiri et al.* [58]. To the authors' knowledge, *Amiri et al.* is the first survey article focused on hashtag recommendation, where a fairly comprehensive study is done to establish the current trends in Twitter-focused research. In their paper it is reflected that several techniques have been proposed over time to try to accurately recommend hashtags to users. For example, in [59] the authors proposed to establish a topic-term relationship, so that the user's perception was taken into account to recommend the hashtag. In their work, the dataset employed consisted of 14 million tweets collected over a period of 52 days, and the best model obtained reflected a mean F-Score of 81.73%. In [60], the authors approached the challenge of recommending appropriate hashtags for hyperlinked tweets as a learning-to-rank problem, and used the SVMRank model to rank and suggest suitable hashtags in each tweet. They conducted experiments with more than 350 million tweets crawled from the Twitter streaming API. The evaluations showed that this approach is capable of providing users with suitable recommendations for hashtags, reaching recall values of about 22% when presenting only five recommendations.

As an alternative approach, *Kywe et al.* [61] propose a collaborative filtering model to incorporate user preferences in hashtag recommendation, and *Gong et al.* [62] propose to model the type of hashtag as a hidden variable into their DPMM-based method (Dirichlet Process Mixture Models). Unlike other similar research works, the dataset in this last study was collected from Sina Weibo[2], which provides a Twitter-like service and is one of the most popular ones in China. The proposed method achieved a Precision of 50.2%, a Recall of 44.2% and a F1 score of 47%. On the other hand, in *Kywe et al.* the training and validation sets of tweets contained more than 2 million tweets from over 37,000 unique users. They used a TF-IDF and cosine similarity scheme as a basis for establishing users relationships, and the best score obtained on a custom hit-rate metric was 31.56%.

There are also various studies that have used a deep learning approach to solve the hashtag recommendation problem. For example, *Li et al.* [63] used a Long Short-Term Memory architecture (LSTM, see [64] for details on this architecture) to encode the intrinsic (syntactic or semantic) relations between sentences into micro-blog vectors, and these vectors were used as features to classify tweets into various hashtags. In their work, a dataset with more than 180 million tweets was collected, and among them there were around 16 million tweets that included hashtags annotated by users. They showed that the incorporation of a topical attention mechanism gives an improvement of more than 7% in the F1 score compared with the standard LSTM method. Alternatively, *Gong and Zhang* [65] proposed an attention-based convolutional neural network, which incorporated a local attention channel and a global channel for hashtag recommendation. Through experiments carried out using a dataset collected from real online services, they demonstrated the effectiveness of the proposed method by improving the performance of several state of the art recommenders by a value of 1.9% in the F1-Score metric.

It is worth pointing out that none of the studies consulted provided an explicit way of modeling the semantic disparity between a tweet without a hashtag and the corresponding version with the hashtag appended. This is why in Section 4.6 we propose a probabilistic metric to refine the hashtag recommendations, so that they are more faithful to the intended meaning of the tweet.

3 DATA COLLECTION

One of the pillars of these kinds of projects is the database to use, since all the modules developed are be trained on it, and therefore the results depend directly on it. In this section, a quick review of the different data sources used in the references as the basis for the development of their systems is done. Later, an explanation of the options chosen to carry out the implementation of the system is included.

3.1 Existing datasets

Most of the studies consulted provide their own dataset, built using the Twitter API. One of the most common methods in all of them is to make use of Tweepy [66], which is a Python library specially designed for accessing the Twitter API. Some of the references that make use of this facility provided by the social network are [27–31, 35, 39, 67–71].

A very common and reliable option consists on the use of sets of tweets provided by the web page Kaggle[3]. They are used in many works, but the two most representative and with which the best results are obtained are [72] and [73]. Other research works obtain data sets from independent pages as well. For example, Top Hashtags[4] is used in [32], Friend or Follow[5] in [74], and the already mentioned Twitter SEISMIC dataset is used in [36] and in many works related to cascade prediction. Another interesting option is the library Get Old Tweets[6], which

[2]https://weibo.com/login.php
[3]https://www.kaggle.com/
[4]https://top-hashtags.com/
[5]https://friendorfollow.com
[6]https://github.com/Jefferson-Henrique/GetOldTweets-python

allows to bypass some of the restrictions of the limited Twitter API in the free version, such as, for example, obtaining tweets with a publication date older than one week.

Finally, it is worth mentioning that other works develop popularity prediction systems for different social networks, and all of them provide different datasets to carry out this task. For example, [32] works on Instagram Posts, [33] on YouTube Videos and [34] on Facebook Posts.

3.2 Data collection

The proposed framework has two different stages: the initial training prior to the publication of the system and a daily update to continue improving the system. Notice that we only consider the content of the tweets for our recommendation purposes, that is, our feature set (prior to preprocessing) is the collection of words and images for each tweet. Also note that all the tweets we work with are in English. The initial training period is a vital phase for the final result of the project, since all the metrics provided in Section 5 are based on the program obtained after this initial training phase. We ultimately decided to reuse two of the datasets found in the work reviewed:

(1) As it was done in [72, 73] we take on of the datasets provided by Kaggle. This choice is due to the quality, reliability and ease of the data provided. A set of tweets from different prominent personalities[7] is chosen, to collect data on the messages that arguably have the most impact on a global level.
(2) Another of the data sets chosen for this task due to the phenomenon it represents is Twitter SEISMIC. It contains a large number of tweets demonstrating the cascade effect, which is closely related to the trend phenomenon in any social network.

In addition to these two datasets, the training data is complemented by a subset of the tweets available at Internet Archive[8] to obtain more samples, regardless of whether they are trendings or not, simply to add more elements to the initial collection so that we get a more complete and reliable training set. We use the most up-to-date data available: January 2021 and September-December 2020.

Next comes the labeling of the tweets in popular and unpopular. We found that it is difficult to select an appropriate threshold: many supposedly common tweets can cover the factually popular tweets if the threshold is set too low, and if the threshold is set too high, only a few tweets will be noticed. While the threshold could be selected subjectively, this could lead to disastrous results it ends up being unsuitable. To solve this problem, we implement the same algorithm as in [25] to acquire the threshold. This algorithm is based on clustering techniques, using the well-known methods DBSCAN and *k-medioids* to cluster the data in two groups, and selecting a value of popularity that separates them.

Once the system is published and after the initial training, we move on to the maintenance and improvement phase. As explained in Section 3.3, a dataset with trending and non-trending tweets is maintained throughout the process. This dataset is updated every day in order to have the latest trendings related to each topic, so that our algorithm recommendations are consistent and more faithful to current trends on Twitter. Every time the dataset is updated we retrain all the models in our framework (using data only from the last 30 days) to keep them up to date with current events. Like we said, the tweets are obtained using Tweepy, which is a Python library for accessing the Twitter API. A script is run every day at a fixed hour to retrieve tweets from the previous day.

4 METHODOLOGY

Popularity is an essential aspect that many users generally want to achieve in their tweets, and even more so when the issuer is a business account. Moreover, as mentioned in Section 1 social networks can be an essential tool for communicating urgent pieces of information to the general population, if done correctly. Our objective is

[7] https://www.kaggle.com/speckledpingu/RawTwitterFeeds
[8] https://archive.org/search.php?query=collection%3Atwitterstream&sort=-publicdate

to maximize the spread of the tweets and achieve a large viralization degree, which as we said before can help users, business and institutions get their message across effortlessly, by increasing the popularity of their tweets via an increase in retweets, likes and interactions count. With that in mind, we present a complete end-to-end hashtag recommendation system that builds on top of other already tested ideas for trend prediction and hashtag recommendation (e.g. [40, 75]). Specifically, we propose a system that given a text input (the tweet that the user wants to share) presents a list of recommended hashtags, ordered according to the predicted popularity of the tweet including those hashtags. The recommendation-prediction pipeline is outlined in Figure 2.

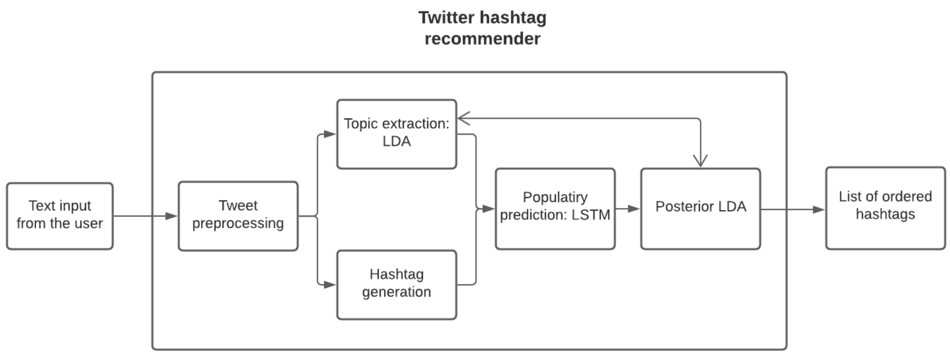

Fig. 2. HOOHLE complete pipeline.

The system consists on six major steps, which are summarized below.

(1) *Tweet preprocessing*. Once the user introduces the tweet, it undergoes a fairly comprehensive preprocessing step, with the aid of the NLTK library [76] and generic NLP techniques. This stage is essential, since it reduces the noise in the input, and has been shown to increase the recommendation accuracy and reduce the computational time required (see e.g. [77]).
(2) *Topic extraction*. The next step is to infer the principal topics of the tweet to produce a recommendation based on similar tweets. To this end, we use an LDA model to retrieve the latent topics and the most likely set of words linked with them, as has been done extensively in state-of-the-art recommender systems (see Section 2).
(3) *Existent hashtag retrieval*. The words obtained in the previous step are compared with a dataset of related tweets that is updated each day, and the most frequent words are used as a basis for the hashtag generation module.
(4) *New hashtag generation*. This is the novel generative part of our model. Using the text generation ideas of *Çetinkaya et al.* [78], for each tweet we retrieve an embedding to a certain latent space using the representation given by BERT [79]. From this embedding, a trained system is used to generate new points in the latent space, which are then projected to the original word space, giving new and not previously considered hashtags related to the tweet itself. We decided to use BERT instead of other well-known language representation models such as GloVe [80] or Word2Vec [81] because the evidence suggests that BERT is more reliable and performs better in practice, achieving state-of-the-art results in various NLP tasks [82–84].
(5) *Popularity prediction*. At this point, each tweet has been assigned different hashtags that may represent it. For each of these hashtags, we predict the impact it might have using a Long Short-Term Memory model,

trained using the guidelines given in [40]. For this task we use our up-to-date dataset, which is a way of incorporating temporal information into our model (see Section 3 for a detailed description of the datasets).

(6) *Hashtag and tweet similarity threshold.* While it would be reasonable to rank the hashtags by their predicted impact, we considered that the most popular hashtags might not be sufficiency related to the rest of the tweet to actually fulfill the intended purpose. In order to alleviate this potential flaw, we propose a metric to detect and solve this undesired behaviour. Considering the original tweet and the probabilities of it belonging to each of the latent topics from LDA on Step (2), the main idea is to check if appending each hashtag to the tweet changes those probabilities in a meaningful way. Using this, we can compute a tailored distance between the probabilities before and after adding the hashtag. Given a fixed threshold, we can discard the hashtags which change the latent topics of the tweet, meaning that they are not sufficiently related to it (even though their predicted popularity may be high).

4.1 Tweet preprocessing

Tweets contain a variety of elements such as extraneous symbols, URLS, emojis, alphanumeric characters, and so on. For this reason, natural language processing techniques must be used to preprocess the data and retain only the text actually needed for our model, reducing the noise and increasing the computational efficiency [77]. As it turns out, the NLTK libraries are specifically designed for this task. The followed procedure is simple and is carried out in several related studies with success, such as [11, 15, 17]. Firstly, each word of the tweet is converted to lowercase, and then special tokens like emojis or characters like '<3' are removed. Stopwords and connectors such as 'a', 'the', and so forth can be removed because they have little semantic importance and can disturb the rest of the content. Punctuation marks, numerical characters, URLs and duplicate words are also removed. Finally, a text lemmatization is done to convert every word to its lemma (e.g. 'learning' becomes 'learn'). An example of the preprocessing step is shown in Figure 3.

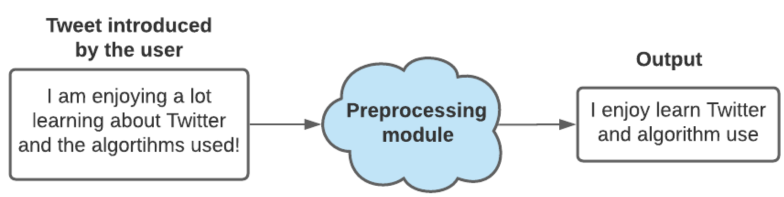

Fig. 3. Tweet preprocessing example.

Another important point is the use of pronouns to refer to nouns previously mentioned in the same tweet. *Kumar et al.* [85] suggested that by replacing pronouns with the associated nouns, the topic relevance detection will be more effective, and hence better hashtags will be retrieved. For this task we employ the system known as *BART* [86], which the authors describe as a "modular toolkit for coreference resolution". Lastly, any images attached to the tweet are processed using the Named Entity Recognition (NER) system proposed in [87], where text and images are jointly used to recognize named entities (such as celebrities, places, politicians, ...) which are then concatenated to the preprocessed text.[9] An open line of future research would be to extract some kind of descriptor from the images and use it to infer the topic of the tweet, which together with our current topic modeling strategy could potentially improve the accuracy of the final recommendation.

[9]Note that extraction of this kind of information from attached videos is not yet supported in our model, but we intend to implement this functionality in the near future.

4.2 Topic extraction

Topic analysis (also known as topic identification, topic modeling, or topic extraction) is a machine learning technique for organizing and comprehending large collections of text data by assigning tags or categories based on the topic or theme of each individual text. As mentioned before, there are several common techniques to fulfill this task, such as PLSA (Probabilistic Latent Semantic Analysis)[88], LDA (Latent Dirichlet Allocation)[50], and TNG (Topical N-Grams) [89]. Following the recommendations in [90], which compares these methods, in this paper we choose LDA because it is the one that more accurately fits the needs of this project. It gives us the posterior probabilities or topic proportion on each tweet, which will be useful in the next steps of the pipeline. Because of its importance in our model, we stop to describe it superficially.

Latent Dirichlet Allocation or *LDA* is a generative probabilistic model in natural language processing that allows a set of observations to be explained by unobserved groups that analyze why some parts of the data are similar. For example, observations may be words in a document (or tweet), regarded as a mixture of a small number of topics, and hence each word's presence is attributable to one of the document's topics. Learning corresponds to extracting information such as the set of topics, their associated word probabilities, the topic of each word, and the particular topic mixture of each document [27, 28, 70]. The main elements of the model are summarized below, and they can be seen schematically in Figure 4.

- K is the number of topics, V the number of words in the vocabulary, M the number of documents, N_d the number of words in document d, and N the total number of words.
- We have a probability distribution $\boldsymbol{\phi} = \{\phi_1, \dots, \phi_K\}$, where ϕ_k is the distribution of words in topic k. Each component $\phi_{k,n}$ is the probability of the n^{th} word in topic k.
- Each document d is associated with a vector of topic proportions θ_d, which is a $K-1$ simplex. Then each component $\theta_{d,k}$ is the probability of topic k in document d. We denote $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_K\}$.
- Each word in each document is assumed to be related with a single topic. The variable $Z_{d,n}$ indexes the topic of the n^{th} word in the d^{th} document.

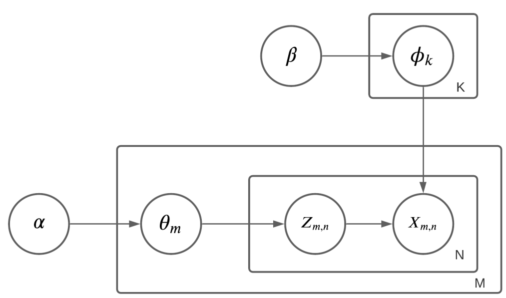

Fig. 4. LDA probabilistic model.

The model assumes that each document is generated with the following generative process: firstly, words proportion for each topic and topic proportions for each document are drawn using Dirichlet distributions, and then for each word, a topic and a word from that topic are drawn from a categorical distribution. The parameters of LDA can be estimated by the collapsed Gibbs sampling algorithm [91] or via variational inference [92].

An example of our topic extraction module can be seen in Figure 5, in which we observe that the number of latent topics in our tweets is fixed to 3. This number of topics is be a hyperparameter of our model, which needs to be optimized by experimentation, as it is done for example in [93]. The final implementation has been done following the steps described in [94][10].

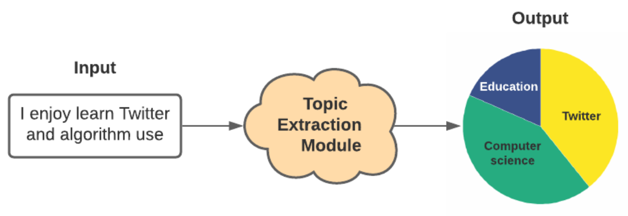

Fig. 5. Topic extraction example.

4.3 Existent hashtag retrieval

Once the topic of the tweet has been found, the next step in the recommendation pipeline consists on finding hashtags related to the tweet. In our case the objective is to find words related to the topics found as it is done in [52]. The steps followed to select the keywords for hashtag recommendation are:

(1) Sample tweets related to the topics found, using TF-IDF and cosine similarity.
(2) Obtain the distribution of terms in every topic.
(3) Select a set of top words for all the topics as the hashtags.

The result is a collection of keywords that are representative of the tweet's general subject, and they are concatenated to form a hashtag. Although we maintain a database with top trending tweets that is updated every day, as explained in Section 3, once the topic is located the latest tweets related to it are also obtained. In this way the system has updated data and the hashtags recommended are more faithful to current trends on Twitter and, consequently, to the reality of the population. An example to obtain some tweets related to the topic found is shown in Listing 1, and a schematic representation of the process is shown in Figure 6.

Listing 1. Tweepy library usage to obtain tweets related to a given topic.

```
new_search = "computer science-filter:retweets"
tweets = tw.Cursor(api.search,
                   q=new_search,
                   lang="en",
                   since='2021-03-23').items(1000)
all_tweets = [tweet.text for tweet in tweets]
```

4.4 New hashtag generation

One of this paper's main contribution is that, once the generation of hashtags is done, we create new candidate hashtags for each tweet using a text generation approach. As said before, the approach is to follow the procedure outlined in [78], where they developed a chat-bot capable of having conversations via retweets. However, since

[10] The code and datasets used to train the model are available in https://github.com/junyachen/BMM.

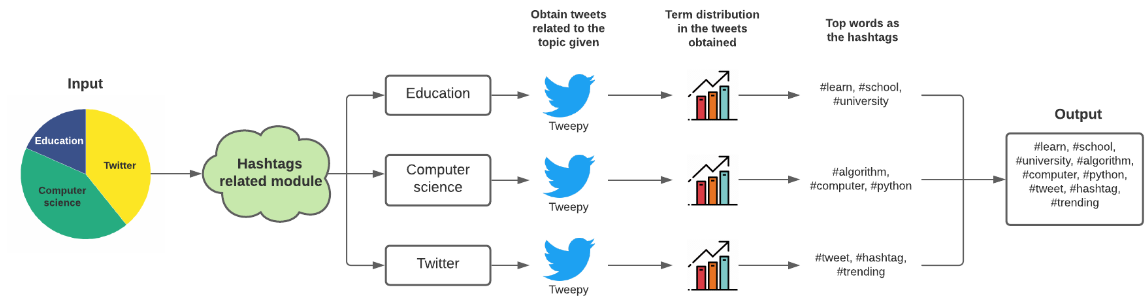

Fig. 6. Hashtags extraction example.

we do not aim to produce full texts but simple meaningful words that might be used as hashtags, we are not concerned by its performance at maintaining a full conversation, but only at its capability to retain the underlying topic of a message and generate new words in consequence. Roughly, the model is a composition of LSTM networks trained with the aid of a BERT embedding and a specifically designed metric for that work, called *Blue*. An example of the procedure can be seen in Figure 7.

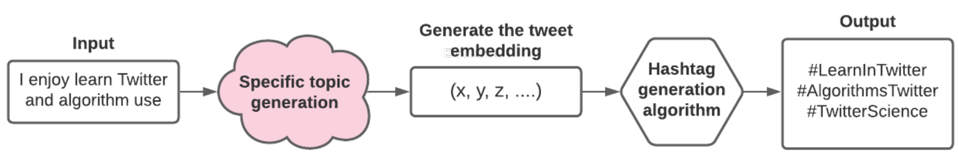

Fig. 7. New hashtag generation example.

4.5 Popularity prediction

Once we have all the possible hashtags the only remaining thing to do is to predict the potential popularity of each one and provide the user with a list of hashtags ordered from highest to lowest popularity. Before building the final hashtag ranking, a prediction based on a unified topic-popularity forecasting framework is constructed, reproducing the work in [40]. As explained in Section 2, we will follow the approach taken in [25] and use a combination of retweets, likes and replies to the tweet as a popularity measure.

The main idea here is that the popularity of a tweet is usually associated with, if not entirely dependent on, the popularity of other similar tweets. For example, before Donald Trump was president, almost no one cared about #MakeAmericaGreatAgain or the border wall with Mexico. The method followed to predict the popularity consists on three major phases:

(1) Frequent words extracted from tweets and hashtag are represented as high dimensional vectors using a BERT embedding, a standard technique to accomplish this task that works well in practice as stated before.
(2) The K-means clustering algorithm is used to group the data into a variety of meaningful topics. These topics are regarded as nodes in a network, with connections between them if the same user interacted with both topics, thus allowing the construction of a set of temporal topic networks.

(3) Finally, we train a Long Short Term Memory (LSTM) model that predicts the number of mentions of a given subject on the following day, using the network's structure and dynamic properties.

Note that the implicit measurement of popularity for a tweet t at day d is

$$popularity_d(t) = r_d(t) + l_d(t) + i_d(t), \tag{1}$$

where r_d, l_d and i_d represent the number of retweets, likes and replies to the tweet at day d. The full popularity prediction process is exemplified in Figure 8.

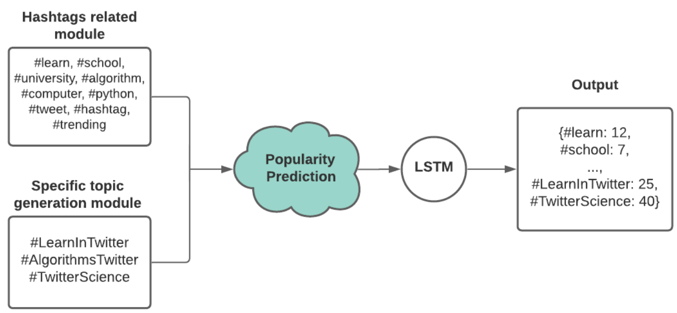

Fig. 8. Popularity prediction example.

The prediction system is trained separately in a supervised fashion using the tweets and the popularity labels available, as described in Section 3.

4.6 Hashtag ranking

As we said before, we have to consider that the hashtag recommendations could be biased to predict very popular hashtags that do not reflect the intended meaning of the rest of the tweet. To combat this issue, we propose a metric that tries to find a balance between the potential impact of a hashtag and how much it distorts the meaning of the original tweet.

The final ranking system is essentially done via a custom metric developed using the posterior probability of each topic computed with our pre-existing LDA model (see Section 4.2). We call this pseudo-recurrent model Hashtag-LDA, which in essence is incorporating the popularity prediction result into our LDA model. Specifically, given the initial topic probabilities $p = \{p_i\}$ from LDA and the updated topic probabilities $\tilde{p} = \{\tilde{p}_i\}$ computed after appending the hashtag in question to the tweet, we calculate

$$l(p, \tilde{p}) = \sum_i |p_i - \tilde{p}_i|.$$

We discard a hashtag if the value of the metric is higher than a given threshold T. This threshold parameter can be in the range $(0, K)$, where K is the total number of topics extracted from a single tweet. As we did with the parameter K, the threshold is tuned to a suitable value through experimentation.

Finally, it is worth pointing out that this system is trained independently from the popularity prediction module. Once the latter is trained and the pipeline is ready to function, we consider as positive labels (good recommendations) the hashtags that did not trigger the threshold, and negative labels those that do. In this way, we can quantify the performance of the method through standard binary metrics. The whole process can be seen schematically in Figure 9.

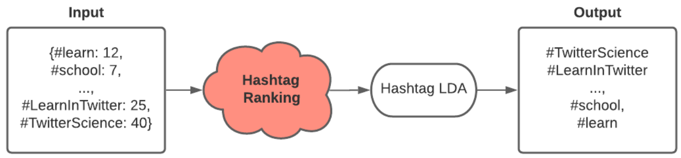

Fig. 9. Hashtag ranking example.

5 EXPERIMENTS

5.1 Evaluation metrics

As stated in Section 2, the dominating metric for testing the performance and effectiveness of the systems proposed is the F1-score, which is a combination of Precision (P) and Recall (R). They are two fairly important model evaluation metrics that are used extensively in this field [95]. In a classification context (popular vs non-popular), precision refers to the is the ratio of correctly predicted positive observations to the total predicted positive observations, while recall concerns itself with the the ratio of correctly predicted positive observations to all positive observations. The F1-score is then defined as:

$$F1 = \frac{2PR}{P + R}.$$

The F1-score is measured first for the popularity module using the popularity prediction metric defined in Eq. (1), which as was explained before is trained separately. For the hashtag module the custom metric explained in Section 4.6 is considered, which offers a trade-off between the predicted popularity and how well the hashtag fits in the original tweet. Then the F1-score is computed using as positive examples the hashtags that do not trigger the threshold.

Finally, apart from the on-line metrics outlined above, we consider an off-line metric to measure the impact of our tool through different time periods. Once the HOOHLE system is deployed, we will select a group of users to track how the popularity of their tweets evolve, and numerically assess how the system is performing (again, using our popularity metric). This *post hoc* study will be divided by stakeholders, to evaluate how the tool affects different target users.

5.2 Results

[This section will be completed when the framework is implemented in the near future, providing tables and graphics with the experimental results obtained.]

6 CONCLUSIONS AND FUTURE WORK

In recent years, the challenge of hashtag recommendation and the field of popularity prediction have been emerging tasks that have caught the interest of researchers across the world. Our work combines both topics and provides a complete end-to-end hashtag recommendation and popularity prediction system. Thus, our recommendation system exploits a gap in research to potentially obtain better results than its predecessors, while providing a convenient interface whose general functionality can be understood and tuned by non-experts for their particular purposes.

Differently from state of the art approaches, the final recommended hashtag is extracted from two different sources, i.e., a dataset of previous topic-related tweets and a hashtag generation algorithm based on novel approaches and state-of-the-art text generation techniques. But the "#" symbol is not the only relevant one on social networks, and more specifically on Twitter. The “@” symbol can also be used in any tweet to mention users whom the sender wants to personally alert. Since 2019 Twitter and other microblogging sites offer searching options such as “Related to a user/topic/hashtag”. We argue that if the appropriate mentions are made, it would be possible to reach a large number of users with which there is not even a followee relationship. In this regard, it has been shown that adding an automatic suggestion of a small list of candidate names can improve communication efficiency [96, 97].

On the other hand, apart from the mentions and hashtags, the rest of the tweet content remains a critical point in terms of popularity gain: the way it is written may make the impressions levels rise above unsuspected limits, or on the contrary, the tweet could go unnoticed. We think that providing users with a paraphrased version of their tweet text would improve the popularity of the post and even enhance the impact of our system [98], so we leave it as an open line of research.

To conclude, we list several approaches that could be followed to increase the performance of our system and be able to suggest more suitable hashtags in the future. One option would be to analyse hyperlinks and media content included in the tweets, so that we can obtain more information related to them. For example, in [99] a co-attention neural network model is proposed which learns to model both images and text. Another important line of research would be to repeat the experiments with other social networks, such as Facebook or Instagram, to see how well our findings translate to different scenarios. Finally, a promising approach that we would like to explore is to consider an extended set of features for each tweet, that provides not only content information but contextual data as well. We can find examples of this approach in [35, 39], where the authors leverage the retweets or user mentions to obtain more trustworthy results; and also the followers of each user to form a sort of social network graph in which relationship among users are highlighted. It is our hypothesis that by including this extended set of features we would obtain a representation of the tweet much more faithful to reality.

REFERENCES

[1] H. Tankovska, “Twitter: Most users by country,” 2021.

[2] E. Otsuka, S. A. Wallace, and D. Chiu, “A hashtag recommendation system for twitter data streams,” *Computational social networks*, vol. 3, no. 1, pp. 1–26, 2016.

[3] E. Lahuerta-Otero and R. Cordero-Gutiérrez, “Retweet or like? That is the question,” p. 18.

[4] E. Pancer, “The popularity and virality of political social media: hashtags, mentions, and links predict likes a,” p. 13.

[5] G. Kostygina, H. Tran, S. Binns, G. Szczypka, S. Emery, D. Vallone, and E. Hair, “Boosting Health Campaign Reach and Engagement Through Use of Social Media Influencers and Memes,” *Social Media + Society*, vol. 6, p. 205630512091247, Apr. 2020.

[6] N. Oliveira, “Retweet Predictive Model for Predicting the Popularity of Tweets,” p. 9.

[7] L. C. Schaupp and F. Bélanger, “The value of social media for small businesses,” *Journal of Information Systems*, vol. 28, no. 1, pp. 187–207, 2014.

[8] M.-J. Pérez-Serrano, J.-V. García-Santamaría, and M. Rodríguez-Pallares, “The social media presence of ibex 35 top executives and their role as influencers,” *Communication & Society*, vol. 33, no. 2, pp. 313–328, 2020.

[9] A. Dormanesh, A. Majmundar, and J.-P. Allem, “Follow-up investigation on the promotional practices of electric scooter companies: content analysis of posts on instagram and twitter,” *JMIR public health and surveillance*, vol. 6, no. 1, p. e16833, 2020.

[10] D. Owen, “New media and political campaigns,” in *The Oxford handbook of political communication*, 2017.

[11] C. Sharma and P. Bedi, “Community based Hashtag Recommender System (CHRS) for twitter,” *Journal of Intelligent & Fuzzy Systems*, vol. 34, pp. 1511–1519, Mar. 2018.

[12] N. Ben-Lhachemi and E. H. Nfaoui, “An Extended Spreading Activation Technique for Hashtag Recommendation in Microblogging Platforms,” p. 8, 2017.

[13] F. Zhao, Y. Zhu, H. Jin, and L. T. Yang, “A personalized hashtag recommendation approach using LDA-based topic model in microblog environment,” vol. 65, pp. 196–206.

[14] S. M. Kywe, T.-A. Hoang, E. P. Lim, and F. Zhu, “On recommending hashtags in Twitter networks,” p. 15.

[15] E. Otsuka, S. A. Wallace, and D. Chiu, “Design and evaluation of a Twitter hashtag recommendation system,” in *Proceedings of the 18th International Database Engineering & Applications Symposium on - IDEAS ’14*, (Porto, Portugal), pp. 330–333, ACM Press, 2014.

[16] B. Shi, G. Poghosyan, G. Ifrim, and N. Hurley, “Hashtagger+: Efficient High-Coverage Social Tagging of Streaming News,” *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, vol. 30, no. 1, p. 16, 2018.

[17] N. Ben-Lhachemi and E. H. Nfaoui, "Using Tweets Embeddings For Hashtag Recommendation in Twitter," *Procedia Computer Science*, vol. 127, pp. 7–15, 2018.
[18] B. Shi, G. Ifrim, and N. Hurley, "Learning-to-rank for real-time high-precision hashtag recommendation for streaming news," in *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, (Republic and Canton of Geneva, CHE), p. 1191–1202, International World Wide Web Conferences Steering Committee, 2016.
[19] S. Aleidi, D. Alsuhaibani, N. Alrajebah, and H. Kurdi, "A Tweet-Ranking System Using Sentiment Scores and Popularity Measures," in *Advances in Data Science, Cyber Security and IT Applications* (A. Alfaries, H. Mengash, A. Yasar, and E. Shakshuki, eds.), vol. 1097, pp. 162–169, Cham: Springer International Publishing, 2019. Series Title: Communications in Computer and Information Science.
[20] G. Yadav, M. Joshi, and R. Sasikala, "Twitter data analysis: temporal and term frequency analysis with real-time event," *IOP Conference Series: Materials Science and Engineering*, vol. 263, p. 042081, Nov. 2017.
[21] S. Iqbal, H. Ullah Khan, U. Ishfaq, M. Alghobiri, and S. Iqbal, "Finding influential users in social networks based on novel features & link-based analysis," *Journal of Intelligent & Fuzzy Systems*, vol. 40, pp. 1623–1637, Jan. 2021.
[22] R. E. Bacha and T. Thi Zin, "Ranking of Influential users based on User-Tweet bipartite graph," in *2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, (Singapore), pp. 97–101, IEEE, July 2018.
[23] A. Ali, M. Alrasheedi, A. Ouda, and L. F. Capretz, "A study of the interface usability issues of mobile learning applications for smart phones from the users perspective," *arXiv preprint arXiv:1501.01875*, 2015.
[24] B. Jitnupong and W. Jirachiefpattana, "Information system user interface design in software services organization: a small-clan case study," in *MATEC Web of Conferences*, vol. 164, p. 01006, EDP Sciences, 2018.
[25] X. Gao, Z. Zheng, Q. Chu, S. Tang, G. Chen, and Q. Deng, "Popularity prediction for single tweet based on heterogeneous bass model," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 5, pp. 2165–2178, 2021.
[26] F. Hemmatian and M. K. Sohrabi, "A survey on classification techniques for opinion mining and sentiment analysis," vol. 52, no. 3, pp. 1495–1545.
[27] M. G. Silva, M. A. Domínguez, and P. G. Celayes, "Analyzing the retweeting behavior of influencers to predict popular tweets, with and without considering their content," in *Information Management and Big Data* (J. A. Lossio-Ventura, D. Muñante, and H. Alatrista-Salas, eds.), vol. 898, pp. 75–90, Springer International Publishing. Series Title: Communications in Computer and Information Science.
[28] R. Geetha and S. Karthika, "Sensitive keyword extraction based on cyber keywords and lda in twitter to avoid regrets," in *Computational Intelligence in Data Science* (A. Chandrabose, U. Furbach, A. Ghosh, and A. Kumar M., eds.), vol. 578, pp. 59–70, Springer International Publishing. Series Title: IFIP Advances in Information and Communication Technology.
[29] V. V. Kaveri and V. Maheswari, "A framework for recommending health-related topics based on topic modeling in conversational data (twitter)," vol. 22, pp. 10963–10968.
[30] J. Osorio-Arjona, J. Horak, R. Svoboda, and Y. García-Ruíz, "Social media semantic perceptions on madrid metro system: Using twitter data to link complaints to space," vol. 64, p. 102530.
[31] Y. M. Çetinkaya, H. Toroslu, and H. Davulcu, "Developing a twitter bot that can join a discussion using state-of-the-art architectures," vol. 10, no. 1, p. 51.
[32] K. R. Purba, D. Asirvatham, and R. K. Murugesan, "An analysis and prediction model of outsiders percentage as a new popularity metric on instagram," vol. 6, no. 3, pp. 243–248.
[33] H. Pinto, J. M. Almeida, and M. A. Gonçalves, "Using early view patterns to predict the popularity of youtube videos," p. 365.
[34] S. Kumar, "Post Prediction On Facebook Page By Using Machine Learning Algorithm," vol. 9, no. 03, p. 4, 2020.
[35] I. Daga, A. Gupta, R. Vardhan, and P. Mukherjee, "Prediction of likes and retweets using text information retrieval," vol. 168, pp. 123–128.
[36] S. Roy, B. K. Suman, J. Chandra, and S. K. Dandapat, "Forecasting the future: Leveraging rnn based feature concatenation for tweet outbreak prediction," in *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*, pp. 219–223, ACM.
[37] C. Gou, H. Shen, P. Du, D. Wu, Y. Liu, and X. Cheng, "Learning sequential features for cascade outbreak prediction," *Knowledge and Information Systems*, vol. 57, no. 3, pp. 721–739, 2018.
[38] Z. T. Kefato, N. Sheikh, L. Bahri, A. Soliman, A. Montresor, and S. Girdzijauskas, "Cas2vec: Network-agnostic cascade prediction in online social networks," in *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, pp. 72–79, 2018.
[39] P. S. Devi, R. Geetha, and S. Karthika, "Trendingtags—classification & prediction of hashtag popularity using twitter features in machine learning approach proceedings," vol. 990, pp. 161–177. Series Title: Advances in Intelligent Systems and Computing.
[40] J. Sun and P. Gloor, "Towards re-inventing psychohistory: Predicting the popularity of tomorrow's news from yesterday's twitter and news feeds," vol. 30, no. 1, pp. 85–104.
[41] R. Misra, "News category dataset," 06 2018.
[42] K. R. Purba, D. Asirvatham, and R. K. Murugesan, "An analysis and prediction model of outsiders percentage as a new popularity metric on instagram," *ICT Express*, vol. 6, no. 3, pp. 243–248, 2020.
[43] S. De, A. Maity, V. Goel, S. Shitole, and A. Bhattacharya, "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, 2017.

[44] A. Zohourian, H. Sajedi, and A. Yavary, “Popularity prediction of images and videos,” in *2018 4th International Conference on Web Research (ICWR)*, pp. 111–117, IEEE, 2018.

[45] M. P. C. F. C.J. Qian, J.D. Tang, “Instagram popularity prediction via neural networks and regression analysis,„” IEEE, 2017.

[46] Z. Zhang, T. Chen, Z. Zhou, J. Li, and J. Luo, “How to become famous: Post popularity prediction with dual-attention,” in *2018 IEEE international conference on big data (big data)*, pp. 2383–2392, IEEE, 2018.

[47] A. Deza and D. Parikh, “Understanding image virality,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1818–1826, 2015.

[48] K. Ding, K. Ma, and S. Wang, “Intrinsic image popularity assessment,” in *Proceedings of the 27th ACM International Conference on Multimedia*, pp. 1979–1987, 2019.

[49] B. Wu, W.-H. Cheng, Y. Zhang, Q. Huang, J. Li, and T. Mei, “Sequential prediction of social media popularity with deep temporal context networks,” *arXiv preprint arXiv:1712.04443*, 2017.

[50] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.

[51] R. Krestel, P. Fankhauser, and W. Nejdl, “Latent dirichlet allocation for tag recommendation,” in *Proceedings of the third ACM conference on Recommender systems*, pp. 61–68, 2009.

[52] F. Godin, V. Slavkovikj, W. De Neve, B. Schrauwen, and R. Van de Walle, “Using topic models for Twitter hashtag recommendation,” pp. 593–596, 2013.

[53] W. X. Zhao, J. Jiang, J. Weng, J. He, E.-P. Lim, H. Yan, and X. Li, “Comparing twitter and traditional media using topic models,” in *Advances in Information Retrieval* (P. Clough, C. Foley, C. Gurrin, G. J. F. Jones, W. Kraaij, H. Lee, and V. Mudoch, eds.), (Berlin, Heidelberg), pp. 338–349, Springer Berlin Heidelberg, 2011.

[54] B. O’Connor, M. Krieger, and D. Ahn, “Tweetmotif: Exploratory search and topic summarization for twitter,” in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 4, 2010.

[55] J. Eisenstein, B. O’Connor, N. A. Smith, and E. Xing, “A latent variable model for geographic lexical variation,” in *Proceedings of the 2010 conference on empirical methods in natural language processing*, pp. 1277–1287, 2010.

[56] R. Ranganath, D. Jurafsky, and D. McFarland, “It’s not you, it’s me: Detecting flirting and its misperception in speed-dates,” in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pp. 334–342, 2009.

[57] K. D. Rosa and J. Ellen, “Text classification methodologies applied to micro-text in military chat,” in *2009 International Conference on Machine Learning and Applications*, pp. 710–714, IEEE, 2009.

[58] B. Amiri, R. Karimianghadim, N. Yazdanjue, and L. Hossain, “Research topics and trends of the hashtag recommendation domain,”

[59] A. Tariq, A. Karim, F. Gomez, and H. Foroosh, “Exploiting topical perceptions over multi-lingual text for hashtag suggestion on twitter,” in *The Twenty-Sixth International FLAIRS Conference*, Citeseer, 2013.

[60] E. Zangerle, W. Gassler, and G. Specht, “On the impact of text similarity functions on hashtag recommendations in microblogging environments,” vol. 3, no. 4, pp. 889–898.

[61] S. M. Kywe, T.-A. Hoang, E.-P. Lim, and F. Zhu, “On recommending hashtags in twitter networks,” in *Social Informatics* (K. Aberer, A. Flache, W. Jager, L. Liu, J. Tang, and C. Guéret, eds.), vol. 7710, pp. 337–350, Springer Berlin Heidelberg. Series Title: Lecture Notes in Computer Science.

[62] Y. Gong, Q. Zhang, and X. Huang, “Hashtag recommendation using dirichlet process mixture models incorporating types of hashtags,” in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 401–410, Association for Computational Linguistics.

[63] Y. Li, T. Liu, J. Jiang, and L. Zhang, “Hashtag recommendation with topical attention-based LSTM,” p. 12.

[64] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[65] Y. Gong and Q. Zhang, “Hashtag recommendation using attention-based convolutional neural network,” p. 7.

[66] J. Roesslein, “Tweepy,” *Python programming language module*, 2015.

[67] E. De Santis, A. Martino, and A. Rizzi, “An infoveillance system for detecting and tracking relevant topics from italian tweets during the covid-19 event,” vol. 8, pp. 132527–132538.

[68] K. Otake and T. Namatame, “Evaluation of the diffusion phenomenon using information from twitter,” vol. 11, no. 5.

[69] N. S. Murugan and G. U. Devi, “Feature extraction using lr-pca hybridization on twitter data and classification accuracy using machine learning algorithms,” vol. 22, pp. 13965–13974.

[70] V. Chakkarwar and S. Tamane, “Social media analytics during pandemic for covid19 using topic modeling,” in *2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)*, pp. 279–282, IEEE.

[71] R. Nagamanjula and A. Pethalakshmi, “A novel framework based on bi-objective optimization and lan2fis for twitter sentiment analysis,” vol. 10, no. 1, p. 34.

[72] P. G. Rao, M. Z. Khan, M. Rafeeq, N. Hitesh, P. D. Shenoy, and K. Venugopal, “An automated learning system for twitter trends,” in *2019 Fifteenth International Conference on Information Processing (ICINPRO)*, pp. 1–6, IEEE.

[73] F. H. A. Shibly, U. Sharma, and H. M. M. Naleer, “Classifying and measuring hate speech in twitter using topic classifier of sentiment analysis,” in *International Conference on Innovative Computing and Communications* (D. Gupta, A. Khanna, S. Bhattacharyya, A. E. Hassanien, S. Anand, and A. Jaiswal, eds.), vol. 1165, pp. 671–678, Singapore: Springer Singapore, 2021. Series Title: Advances in Intelligent Systems and Computing.

[74] L. Vega and A. Mendez-Vazquez, “Detecting of topic-specific leaders in social networks,” vol. 151, pp. 1188–1193.

[75] *A Survey on Hashtag Recommendations*, FRUCT Oy, Sept. 2020.

[76] E. Loper and S. Bird, “Nltk: The natural language toolkit,” *arXiv preprint cs/0205028*, 2002.

[77] D. Ramachandran and R. Parvathi, “Analysis of twitter specific preprocessing technique for tweets,” *Procedia Computer Science*, vol. 165, pp. 245–251, 2019. 2nd International Conference on Recent Trends in Advanced Computing ICRTAC -DISRUP - TIV INNOVATION , 2019 November 11-12, 2019.

[78] Y. M. Çetinkaya, I. H. Toroslu, and H. Davulcu, “Developing a twitter bot that can join a discussion using state-of-the-art architectures,” *Social Network Analysis and Mining*, vol. 10, no. 1, pp. 1–21, 2020.

[79] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” 2019.

[80] J. Pennington, R. Socher, and C. D. Manning, “Glove: Global vectors for word representation,” in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pp. 1532–1543, 2014.

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

[82] I. Tenney, D. Das, and E. Pavlick, “Bert rediscovers the classical nlp pipeline,” 2019.

[83] A. Khatri, P. P, and D. A. K. M, “Sarcasm detection in tweets with bert and glove embeddings,” 2020.

[84] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, “Roberta: A robustly optimized bert pretraining approach,” 2019.

[85] N. Kumar, E. Baskaran, A. Konjengbam, and M. Singh, “Hashtag recommendation for short social media texts using word-embeddings and external knowledge,” vol. 63, no. 1, pp. 175–198.

[86] Y. Versley, S. P. Ponzetto, M. Poesio, V. Eidelman, A. Jern, J. Smith, X. Yang, and A. Moschitti, “BART: a modular toolkit for coreference resolution,” pp. 9–12, 2008.

[87] D. Esteves, R. Peres, J. Lehmann, and G. Napolitano, “Named entity recognition in twitter using images and text,” in *International Conference on Web Engineering*, pp. 191–199, Springer, 2017.

[88] T. Hofmann, “Probabilistic latent semantic analysis,” *arXiv preprint arXiv:1301.6705*, 2013.

[89] X. Wang, A. McCallum, and X. Wei, “Topical n-grams: Phrase and topic discovery, with an application to information retrieval,” in *Seventh IEEE international conference on data mining (ICDM 2007)*, pp. 697–702, IEEE, 2007.

[90] D. Sharma, B. Kumar, and S. Chand, “A survey on journey of topic modeling techniques from svd to deep learning,” *International Journal of Modern Education and Computer Science*, vol. 9, no. 7, p. 50, 2017.

[91] Saura and Bennett, “A Three-Stage method for Data Text Mining: Using UGC in Business Intelligence Analysis,” *Symmetry*, vol. 11, p. 519, Apr. 2019.

[92] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, “Variational inference: A review for statisticians,” *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[93] M. Sokolova, K. Huang, S. Matwin, J. Ramisch, V. Sazonova, R. Black, C. Orwa, S. Ochieng, and N. Sambuli, “Topic modelling and event identification from twitter textual data,” *arXiv preprint arXiv:1608.02519*, 2016.

[94] J. Chen, Z. Gong, and W. Liu, “A Dirichlet process biterm-based mixture model for short text stream clustering,” *Applied Intelligence*, vol. 50, pp. 1609–1619, May 2020.

[95] S. S. Bama, M. Ahmed, and A. Saravanan, “A survey on performance evaluation measures for information retrieval system,” *International Research Journal of Engineering and Technology*, vol. 2, no. 2, pp. 1015–1020, 2015.

[96] T. Gui, P. Liu, Q. Zhang, L. Zhu, M. Peng, Y. Zhou, and X. Huang, “Mention recommendation in twitter with cooperative multi-agent reinforcement learning,” in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 535–544, 2019.

[97] B. Wang, C. Wang, J. Bu, C. Chen, W. V. Zhang, D. Cai, and X. He, “Whom to mention: expand the diffusion of tweets by@ recommendation on micro-blogging systems,” in *Proceedings of the 22nd international conference on World Wide Web*, pp. 1331–1340, 2013.

[98] J. Zhao and M. Lan, “Ecnu: Leveraging word embeddings to boost performance for paraphrase in twitter,” in *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*, pp. 34–39, 2015.

[99] S. Zhang, Y. Yao, F. Xu, H. Tong, X. Yan, and J. Lu, “Hashtag recommendation for photo sharing services,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 5805–5812, 07 2019.