On Predicting the Popularity of Newly Emerging Hashtags in Twitter

Zongyang Ma, Aixin Sun, and Gao Cong
School of Computer Engineering, Block N4, Nanyang Technological University, Nanyang Avenue, Singapore.
E-mail: zma4@e.ntu.edu.sg; {axsun, gaocong}@ntu.edu.sg

Because of Twitter's popularity and the viral nature of information dissemination on Twitter, predicting which Twitter topics will become popular in the near future becomes a task of considerable economic importance. Many Twitter topics are annotated by hashtags. In this article, we propose methods to predict the popularity of new hashtags on Twitter by formulating the problem as a classification task. We use five standard classification models (i.e., Naïve bayes, *k*-nearest neighbors, decision trees, support vector machines, and logistic regression) for prediction. The main challenge is the identification of effective features for describing new hashtags. We extract 7 content features from a hashtag string and the collection of tweets containing the hashtag and 11 contextual features from the social graph formed by users who have adopted the hashtag. We conducted experiments on a Twitter data set consisting of 31 million tweets from 2 million Singapore-based users. The experimental results show that the standard classifiers using the extracted features significantly outperform the baseline methods that do not use these features. Among the five classifiers, the logistic regression model performs the best in terms of the Micro-F_1 measure. We also observe that contextual features are more effective than content features.

1 Introduction

Twitter is a popular microblogging service that allows users to post short messages called "tweets." Twitter also provides social networking features that allow users to follow other users, to retweet (or repost) their received tweets, and to reply to other users' tweets. According to a Twitter blog post on March 21, 2012, more than 340 million tweets were posted daily by 140 million active Twitter users.[1] Because of Twitter's popularity and the viral nature of information dissemination on Twitter, trending topics become popular on Twitter in a very short time. In this article, we study the problem of effectively predicting the popularity of Twitter topics in the near future based on hashtags (keywords prefixed with # symbol in tweets). Our work has a number of practical applications to marketing and public relations. For example, it can be used by an advertising or public relations firm to assess the potential for success of their marketing campaigns (Kasiviswanathan, Melville, Banerjee, & Sindhwani, 2011; Schultz, Utz, & Gritz, 2011; Wei, Bu, & Liang, 2012).

Topic or event detection in Twitter remains a challenging research task because of the overwhelming information flow as well as the short and noisy content (Li, Sun, & Datta, 2012). However, hashtags are widely used in Twitter to define shared context for specific events, topics, or memes (Lehmann, Goncalves, Ramasco, & Cattuto, 2012). Newly created hashtags are frequently used to annotate emerging topics or events. In this article, we propose to predict the popularity of new hashtags in the near future (e.g., 1 day). The popularity of a hashtag is defined as the *number of users* who post at least one tweet containing the hashtag within the given time period. In our setting, a *new hashtag* can either be (a) a newly created hashtag that has not appeared before or (b) a hashtag created earlier, which was popular, then unpopular for a predefined time period (e.g., a week), and is now popular again. For instance, the hashtag #apple gains popularity from time to time when a new product from Apple is released.

We argue that predicting hashtag popularity on a daily basis is important in practice because of how fast information spreads on Twitter. However, the prediction task is also challenging because very limited information can be obtained for a newly created hashtag. There are at least two implications: (a) existing approaches on trend predication (Jeon, Croft, Lee, & Park, 2006; Liu et al., 2011; Liu, Huang, An, & Yu, 2007) cannot be applied to our problem because of the lack of historical data for a new hashtag, and (b) the identification of discriminative features from the limited information about a new hashtag becomes the key research issue. Our main focus is, therefore, to identify and

[1]http://blog.twitter.com/2012/03/twitter-turns-six.html. Accessed September 14, 2012.

Received August 2, 2012; revised September 10, 2012, and September 14, 2012; accepted September 17, 2012

evaluate the effectiveness of various features for the hashtag popularity prediction task. In particular, we have evaluated two types of features: *content* and *contextual* features. Content features are derived lexically from the hashtag string itself (e.g., if the hashtag contains digits) as well as from the content of the tweets containing the hashtag (e.g., the topic vector of the tweets). Contextual features are mainly derived from the social graphs formed by Twitter users. In a nutshell, the users who have adopted a hashtag form a virtual community, and we derive features from both the community graph and users who are not members of the community but have some relationships with the community members.

To the best of our knowledge, our work is the first to use both content and contextual features to predict the popularity of hashtags on a *daily* basis. This distinguishes our work from existing studies that merely consider one type of feature or predict hashtag popularity at much coarser time granularity (e.g., weekly popularity; Tsur & Rappoport, 2012). We formulate our problem as a classification task. In our experiments, we evaluated 7 content features and 11 contextual features extracted from more than 31 million tweets on their effectiveness for hashtag popularity prediction. We use five commonly used classification models (i.e., Naïve bayes [NB], *k*-nearest neighbors [KNN], decision trees, support vector machines [SVMs], and logistic regression [LR]) and three baseline methods. Our experimental results show that contextual features are more effective than content features for the prediction task, and that LR and KNN outperform the other three classification models. We also conducted experiments to evaluate the effectiveness of the features for popularity prediction for hashtags that have been popular for the past 2 days instead of 1 day.

2 Related Work

In this section, we first review existing research work on the dual role of Twitter as a microblog and social network. We then give a comprehensive survey on the characteristics of hashtags. Finally, we briefly introduce two research fields, graph evolution and statistical prediction, that are related to our work.

2.1 Twitter as a Microblog and Social Network

Twitter plays a dual role as both a microblog and a social network (Thelwall, Buckley, & Paltoglou, 2011). On one hand, posting tweets via the web or mobile interface is the central activity in Twitter and represents Twitter's function as a microblog. On the other hand, the following, retweeting, and mention behaviors in Twitter reflect its function as a social network.[2] However, unlike social networks such as Facebook and LinkedIn, the reciprocity in messages between Twitter users is low, and this skewed structure of the social network suits its function of information diffusion (Kwak, Lee, Park, & Moon, 2010). The dual role of Twitter and its unique characteristics have attracted the attention of a number of researchers. In the following, we briefly review some past research work on Twitter. The findings from previous work may help us in identifying features related to hashtag popularity prediction.

Twitter is now one of the most popular platforms for discussing current events for web users. Topic detection and tracking have become important issues in Twitter research. Naaman, Becker, and Gravano (2011, p. 908) defined a trend profile as a "collection of tweets with a topical keyword" (e.g., earthquake, Obama). The authors characterized the "exogenous trends" (trends originating from outside of Twitter, e.g., earthquake, hurricane) and "endogenous trends" (trends originating from within Twitter, e.g., popular tweets posted by Obama) using a set of features derived from a trend profile. Example features include the *retweet fraction*, *reply fraction*, *triangles*, and *components in graph* (see Table 1 for the definitions of these features). Their experiments show that exogenous and endogenous trends have different characteristics using Bonferroni correction based on the feature sets. Sakaki, Okazaki, and Matsuo (2010) proposed a framework to detect earthquakes in Japan and constructed a system for earthquake reporting. Treating each user as a sensor and each tweet as a sensor reading, the authors used SVMs to classify tweets as positive or negative (tweets referring to earthquakes are defined as positive). Then Kalman and particle filters using geographical information in tweets were applied to infer the location of an earthquake. Considering the aforementioned work on topic detection and tracking in Twitter, bursty hashtags (i.e., hashtags whose popularity rise, then fall very quickly) or keywords are usually treated as candidate topic indicators. Their goal is to identify whether tweets with a specific keyword or hashtag are event-related, whereas we aim to predict the *popularity* of a given hashtag.

Another interesting research task is sentiment analysis in Twitter (Calais Guerra, Veloso, Meira, & Almeida, 2011; Thelwall et al., 2011). Thelwall et al. (2011) proposed a novel lexicon-based classification model, namely, the SentiStrength model, to classify short texts like tweets by their sentiment. The SentiStrength model incorporates rules such as booster words (e.g., very, so) and emoticons to improve classification performance. They discovered that negative sentiment is dominant in tweets about emerging events. Intuitively, tweets with negative sentiment are more likely to be retweeted. Therefore, tweet sentiment might affect the adoption of event-related hashtags.

Because of the 140-character length constraint on tweets, users favor using short URLs to make their tweets more informative. Rowlands, Hawking, and Sankaranarayana (2010) used tweet content as the anchor text to facilitate Twitter URL searching. Dong et al. (2010) extracted different feature sets from Twitter URLs and web URLs, and their experimental results demonstrate that Twitter URLs linked

[2]In Twitter, users may repost a tweet to their followers through the retweet function; users may also reply or mention another user by using "@" followed by the username of the user.

DOI: 10.1002/asi

TABLE 1. The 7 content features (F_{c1}–F_{c7}) and the 11 contextual features (F_{x1}–F_{x11}).

Feature		Description
F_{c1}	*ContainingDigits*	Binary attribute checking whether a hashtag contains digits
F_{c2}	*SegWordNum*	Number of segment words from a hashtag
F_{c3}	*URLFrac*	Fraction of tweets containing URL in T_t^h
F_{c4}	*SentimentVector*	3-Dimension vector: ratio of neutral, positive, and negative tweets in T_t^h
F_{c5}	*TopicVector*	20-Dimension topic distribution vector derived from T_t^h using topic model
F_{c6}	*HashtagClarity*	Kullback–Leibler divergence of word distribution between T_t^h and tweets collection $\mathcal{T}$
F_{c7}	*SegWordClarity*	Kullback–Leibler divergence of word distribution between tweets containing any segment word in h and tweet collection $\mathcal{T}$
F_{x1}	*UserCount*	Number of users $\left\|U_t^h\right\|$
F_{x2}	*TweetsNum*	Number of tweets $\left\|T_t^h\right\|$
F_{x3}	*ReplyFrac*	Fraction of tweets containing mention @
F_{x4}	*RetweetFrac*	Fraction of tweets containing *RT*
F_{x5}	*AveAuthority*	Average authority of users in G_t^h
F_{x6}	*TriangleFrac*	Fraction of users forming triangles in G_t^h
F_{x7}	*GraphDensity*	Density of G_t^h
F_{x8}	*ComponentRatio*	Ratio between number of connected components and number of nodes in G_t^h
F_{x9}	*AveEdgeStrength*	Average edge weights in G_t^h
F_{x10}	*BorderUserCount*	Number of border users
F_{x11}	*ExposureVector*	15-Dimension vector of exposure probability $P(k)$

to new articles are read and shared by more users. We observe that, in our data set, 21.7% of hashtags co-occur with Twitter URLs in tweets. Given that a hashtag potentially is a topic indicator and Twitter URLs enrich tweet content with more information, the existence of Twitter URLs can be used as a feature for hashtag popularity prediction.

One of the key research tasks is identifying influential Twitter users. Weng, Lim, Jiang, and He (2010) proposed a framework to discover topically influential users in Twitter using the topic model (Blei, Ng, & Jordan, 2003) and PageRank. In Pal and Counts (2011), a set of raw features (e.g., *number of original tweets*, *number of retweets*, and *number of mentions*) from user interactions in Twitter was extracted; then user profiles consisting of fine-grained features were formed. After generating a feature vector for each user, users were clustered via the Gaussian mixture model and ranked via the Gaussian ranking algorithm. Welch, Schonfeld, He, and Cho (2011) presented an analysis of the retweeting graph as compared with the following graph in Twitter. Using PageRank on both graphs, they discovered that the retweeting graph can better preserve topical relevance than the following graph.

2.2 Characteristics of Hashtag

The complicated characteristics of hashtags in Twitter have sparked researchers' interest. Several aspects of hashtags have been studied in the literature, including hashtag sentiment analysis (Wang, Wei, Liu, Zhou, & Zhang, 2011), hashtag retrieval (Efron, 2010), and hashtag adoption (Yang, Sun, Zhang, & Mei, 2012). It is reported that users might have different purposes for adopting a hashtag: either to bookmark the content of tweets or to participate in a community graph concentrating on the same topic, or both (Yang et al., 2012). Therefore, both content and contextual features should be considered in studying hashtags. Most germane to this work are the studies in hashtag information diffusion (Romero, Meeder, & Kleinberg, 2011), hashtag popularity evolution (Lehmann et al., 2012), and content-based hashtag spread prediction on a weekly basis (Tsur & Rappoport, 2012).

Romero et al. (2011) analyzed differences in the mechanics of information diffusion of hashtags from eight predefined categories (e.g., politics, celebrity, and games) and discovered that given repeated exposures to a hashtag, politics hashtags are more likely to be adopted. Meanwhile, the initial graph for the politics hashtag is denser, illustrating that the graph on political topics contains more triangles. The user graph was constructed based on the mention relationship; namely, if user u_1 mentions user u_2, there exists a directed edge from u_1 to u_2. Note that the mention relationship is the key to studying interaction between users in Twitter (Huberman, Romero, & Wu, 2009). The authors also confirmed that there exists a strong relationship between contextual features and information diffusion. However, the authors did not use these features to predict the *popularity* of hashtags in the near feature. In Lehmann et al. (2012), the evolution of hashtag popularity over time (e.g., usage patterns before and after bursty peaks) was analyzed for hashtags with bursty peaks. Both studies are retrospective analysis of hashtags using historical data.

Using 25-week Twitter data, Tsur and Rappoport (2012) reported hashtag frequency prediction on a weekly basis using a regression model. The features are mainly derived from the hashtag itself (e.g., orthography, number of characters in a hashtag). In their work, a regression model was used to predict hashtag frequency (i.e., the number of tweets containing a hashtag). Their experiments demonstrated that content features from the hashtag itself improve the model's performance.

DOI: 10.1002/asi

2.3 Graph Evolution and Statistical Prediction

As more and more users adopt a hashtag, the community graph for the hashtag continuously grows. Thus, the popularity prediction problem is also related to graph evolution. Two models to simulate graph evolution, the community-guided attachment model and forest fire model, were proposed in Leskovec, Kleinberg, and Faloutsos (2005). Their models require that the degrees of all nodes are increasing and the distance between nodes is decreasing over time. Backstrom, Huttenlocher, Kleinberg, and Lan (2006) investigated the factors that affect users joining a new community, the growth of the community, and users' movement between communities. They concluded that the fraction of users having many friends is the most important factor in determining the growth of the community. Using a maximum-likelihood method, Leskovec, Backstrom, Kumar, and Tomkins (2008) modeled the evolution of the graph edge by edge and measured preferential attachment degree. All of these aforementioned studies focus on the evolution of the graph over a long period. However, most graph communities based on hashtags have an ephemeral lifetime. Hence previous graph methods cannot be directly applied to solve our problem. Nevertheless, we will evaluate some of the identified factors for their effectiveness in predicting hashtag popularity.

Hashtag popularity prediction can be considered as a trend/rank prediction task. The task is to predict future outcomes using historical data and it is often formulated as a classification task. Jeon et al. (2006) extracted nontext features from the data set of a question answering service and predicted the quality of answers using the maximum entropy method. Because of the strong relationship between viewer reviews of a movie and the revenue of the movie, Liu et al. (2007) analyzed sentiment information from blogs discussing movies and leveraged the autoregressive model to predict movie revenues in the near future. Liu et al. (2011) proposed using various regression algorithms to predict the satisfaction of web users who search with the community-based question-answering system, and they found that LR yields the best experimental outcomes.

3 Problem Setting and Prediction Methods

In our problem setting, all tweets received from a Twitter stream are partitioned into consecutive fixed-time intervals by their time stamps. The time interval could be an hour, a few hours, or a day, depending on the number of tweets received, as well as the time criticality of the prediction. We define the *popularity* of a hashtag h in time interval t to be the number of users who post at least one tweet annotated by h within the time interval t, and we denote this by Φ_t^h. Given a new hashtag at time t, our task is to predict its popularity at time $t+1$, or Φ_{t+1}^h. Note that predicting the *exact value* of Φ_{t+1}^h is extremely difficult and is often not necessary. Therefore, we relax the problem and predict the *range* of its popularity. We define five ranges of an exponentially increasing size: $[0, \phi]$, $[\phi, 2\phi]$, $[2\phi, 4\phi]$, $[4\phi, 8\phi]$, and $[8\phi, +\infty]$. We refer to these as being *not popular*, *marginally popular*, *popular*, *very popular*, and *extremely popular*, respectively. Note that, depending on the number of tweets received from the Twitter stream and the requirements of a specific prediction application, a different number of ranges may be defined. The value of ϕ controls the relative sizes of the ranges defined.

With the five ranges defined, our problem can be formulated as a classification problem. Given the features obtained for a hashtag h at time t, we predict its popularity range (i.e., one of the five categories) at time $t+1$. Because the key focus of this research is to identify and evaluate the effectiveness of features for the prediction, we apply five widely used classifiers in our evaluation: NB, KNN, decision trees (C4.5), SVM, and LR. We use *SVM*^{light}[3] to implement a multiclass SVM where linear kernels are used with all default parameter settings. For KNN, we use Euclidian distance and set $k = 3$. All the remaining classifiers are based on the Weka implementation using default parameter settings. In addition to the five standard classification methods, we have also evaluated three baseline methods, namely, Random, Lazy, and PriorDist. The three baseline methods do not use the extracted features.

- Random: Predict the popularity range of Φ_{t+1}^h randomly among the five ranges.
- Lazy: Predict the range of Φ_{t+1}^h to be the same as Φ_t^h.
- PriorDist: Predict the range of Φ_{t+1}^h randomly following a prior probability distribution on the five ranges.

4 Features for Hashtag Popularity Prediction

In this section, we detail the 7 content and 11 contextual features (see Table 1) that we have evaluated for hashtag popularity prediction.[4] We use T_t^h to denote the collection of tweets containing hashtag h published in time interval t, and U_t^h to denote the collection of Twitter users who have published at least one tweet containing hashtag h in time interval t. Thus, $\Phi_t^h = |U_t^h|$. Most of content features are derived from the hashtag itself lexically or T_t^h, and most contextual features are derived from U_t^h and the graph, denoted by G_t^h, constructed based on U_t^h considering the interactions between users. The interactions between users are quantified by the number of mentions in their published tweets, which will be detailed in the following discussion.

4.1 Content Features

In our study, content features are extracted not only lexically from the hashtag string but also using the collection of tweets annotated by the hashtag. As reported in Tsur and

[3] http://svmlight.joachims.org/svm_multiclass.html

[4] A preliminary study of a subset of the features is reported in Ma, Sun, and Cong (2012).

DOI: 10.1002/asi

Rappoport (2012), content features derived from a hashtag string (e.g., the number of words in the hashtag, digits usage in the hashtag) improve prediction performance. Therefore, we include these features in our evaluation. Moreover, we include the collection of tweets to enhance the representation of the hashtag.

4.1.1 Hashtag lexical features. The first two features, F_{c1} and F_{c2}, are lexical features derived from the hashtag string. The first is a binary feature to indicate whether the hashtag contains digits. Digits are widely used as temporal annotations in hashtags (e.g., #sgelection2011, #itshow2011) or for enumeration in Twitter game hashtags (e.g., #10thingsIlike, #5excuseforlate). The second is the number of segment words contained in a hashtag. Note that a hashtag usually consists of several words. The appropriate word compound can make the hashtag clearer, as well as encourage more users to adopt the hashtag. For example, the #bestthingyouneverheardof called Twitter game hashtag (Lehmann et al., 2012) can be parsed to "best thing you never heard of." Because of its clear meaning, many users put this hashtag in their tweets to share their experiences. We manually segment hashtags into separate words and count the number of separate words. Acronyms such as #sg, #ndp are considered as one word. Note that because our purpose is to evaluate the effectiveness of the features, manual segmentation would avoid potential errors introduced by automatic segmentation methods.

4.1.2 Content features from T_t^h. Four features are derived from T_t^h. The first feature, F_{c3}, is the fraction of tweets containing URLs. As discussed earlier in the Related Work section, Twitter URLs enrich tweet content by linking to a web page. Consequently, the higher the URL fraction, the more external information is introduced. Example hashtags with high URL fraction include #japanlife, #singapore, and #free . All these hashtags are widely adopted.

It is reported that topics or events expressing negative sentiments are more prevalent in Twitter (Thelwall et al., 2011); consequently, a hashtag's sentiment is a potentially useful feature for estimating its propagation. We consider each hashtag as having a neutral/positive/negative three-dimensional sentiment vector and implement a hierarchical sentiment classification model. For each tweet in T_t^h, we first use the subjective/objective model to classify the tweet as subjective or objective. If the tweet is classified as subjective, we further use the positive/negative model to classify the tweet as positive or negative. We calculate the fraction of neutral, positive, and negative sentiment tweets for T_t^h. The hierarchical sentiment classifier is implemented via LingPipe.[5] Feature F_{c4} is the three-dimensional vector with the fractions of the neutral, positive, and negative tweets in T_t^h.

Hashtags on similar topics (e.g., politics, music, sports) may follow similar popularity trends (Lehmann et al., 2012). For example, #sgelection and #sgpresident both refer to political events and demonstrate similar popularity trends in our data set. To identify the topic distribution of a hashtag, we use Latent Dirichlet Allocation (Blei et al., 2003). We consider the semantic meaning of a hashtag is to be defined by the collection of tweets containing the hashtag in a time interval. We therefore infer the topics of a hashtag from a virtual document formed by all tweets containing the hashtag in that time interval. More specifically, we consider each T_t^h as a virtual document, and 20 topics are inferred from all such documents. A 20-dimension topic vector (i.e., F_{c5}) is then assigned to each hashtag, with the entries quantifying the likelihood of the hashtag belonging to the corresponding topic. Note that the same hashtag at different time intervals may be assigned different topic distributions depending on the content of the tweets in T_t^h.

The hashtag clarity feature F_{c6} quantifies the topical cohesiveness of all tweets in T_t^h. A hashtag h_t is described by a set of words extracted from T_t^h to compute the clarity score. The word distribution is then compared with that of the entire tweet collection $\mathcal{T}$. If a hashtag refers to a specific topic, then the high probabilities of a few topic-relevant words distinguish its tweets from the background. For instance, #royalwedding refers to the wedding of Prince William and Catherine Middleton and has a clarity of 11.5, whereas a hashtag like #fb has a clarity slightly more than 2 because its meaning is unclear. Formally defined in Equation 1, a hashtag's clarity is the Kullback–Leibler divergence between the unigram language model inferred from T_t^h and the background language model from the entire tweet collection $\mathcal{T}$.

$$Clarity(h_t) = \sum_{w \in T_t^h} P(w|T_t^h) log_2 \frac{P(w|T_t^h)}{P(w|\mathcal{T})} \quad (1)$$

4.1.3 Segment words clarity. The last feature, F_{c7}, is an extension of hashtag clarity. The purpose is to evaluate whether the segment words in a hashtag are topically cohesive. Recall that we have manually segmented a hashtag into segment words $\{w_1, w_2 \ldots w_n\}$. Using the segment words as a keyword query, we search for the 2,000 most relevant tweets posted within the time interval $t-7$ and t. For example, for the hashtag #londonriot posted in time interval t, we first segment #londonriot into words *london* and *riot*, then use *london riot* as a query to search for the 2,000 most relevant tweets between $t-7$ and t. The clarity score is computed using these 2,000 tweets, in a similar way to the hashtag clarity.

4.2 Contextual Features

We use U_t^h to denote the collection of Twitter users who published at least one tweet containing hashtag h in time interval t. We consider these users form a virtual community for h_t, and the hashtag adoption could be largely affected by

[5]http://alias-i.com/lingpipe/

DOI: 10.1002/asi

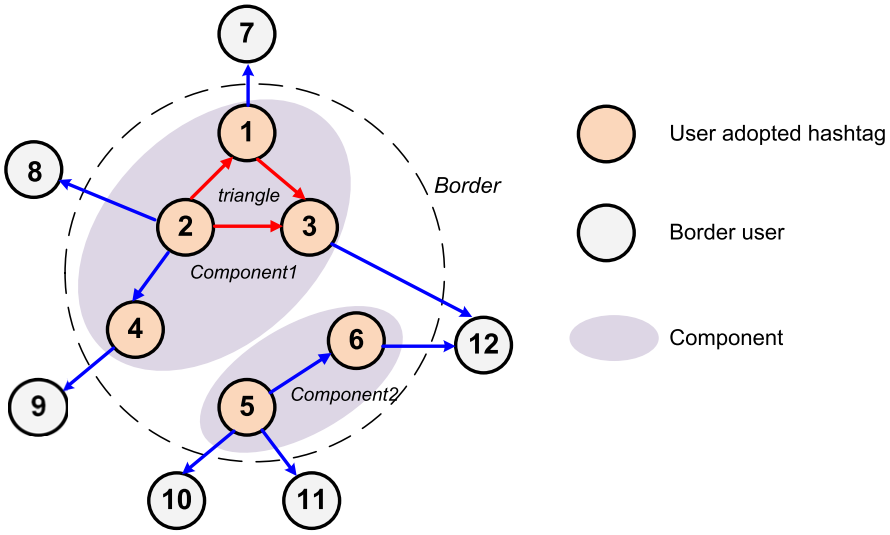

FIG. 1. Example virtual community of a hashtag. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the social relationships among these users, as well as their followers. Listed in Table 1, the first contextual feature F_{x1} is the number of users in U_t^h, that is, $|U_t^h|$. F_{x1} captures the *popularity* of the hashtag h in time interval t and is considered as a determined factor affecting the *popularity* in time interval $t+1$. The next feature, F_{x2}, number of tweets in T_t^h, can be considered to be either a content or a contextual feature. Nevertheless, in our data set, we observe that the number of tweets and the number of users are highly correlated (>.85 measured by the Pearson correlation) for all h_t's. We therefore simply put this as a contextual feature.

The next two contextual features, F_{x3} and F_{x4}, are derived from the social actions among users in U_t^h, that is, replying to a tweet or retweeting a tweet. Replying or mention is a key to study interaction behavior between users in Twitter (Huberman et al., 2009). If user u mentions user v in his or her tweet annotated by hashtag h, user v is unlikely to miss the tweet. The retweeting mechanism is another major force for promoting hashtags. The remaining seven contextual features are derived from the social graph formed by the users in the virtual community. To capture the relationships among users, we first construct a directed weighted graph $\mathcal{G} = \langle U, E \rangle$. In $\mathcal{G}$, a user $u \in U$ is a node and a directed edge $e(u_p, u_q) \in E$ from user u_p to u_q is weighted by the number of times u_p mentions u_q in his or her tweets, similar to that in Romero et al. (2011). The authority scores of users are computed in this global user graph using the PageRank algorithm. By extracting users' relationship from the global user graph, we form a community graph $G_t^h = \langle U_t^h, E_t^h \rangle$, from which we derive the remaining seven contextual features.

4.2.1 Contextual features derived from G_t^h. Feature F_{x5}, average authority, is adopted to measure the influential level of the community G_t^h. Intuitively, if a user is followed or mentioned by many users, he or she is likely to be influential. Feature F_{x6} is the fraction of users forming triangles in graph G_t^h, which reflects the strength of the ties among users in the graph. Three nodes form a triangle if any pair of nodes is connected by an edge. An example triangle is formed by nodes 1, 2, and 3, which is illustrated using three red lines in Figure 1. A higher triangle fraction indicates stronger ties among users. Different triangle fractions have been observed for community graphs of hashtags from different categories (e.g., the community graph for the political hashtag is denser and contains more triangles; Romero et al., 2011). Therefore, the triangle fraction feature distinguishes the category of the hashtag, which could benefit hashtag popularity prediction because hashtags of the same category share similar trends.

Graph density (feature F_{x7}) is a common feature in graph mining to measure sparsity of the graph. It is defined as $|E_t^h| / (|U_t^h| * (|U_t^h| - 1))$, which is the ratio of the number of edges and the number of possible edges in a graph. Average edge strength (feature F_{x8}) measures the overall degree of user interaction in G_t^h. The larger the edge weights, the more interactions among users. Let C_t^h denote the set of disconnected components in G_t^h. Feature F_{x9} is computed by $|C_t^h| / |U_t^h|$. Two components formed by nodes $\{1, 2, 3, 4\}$ and nodes $\{5, 6\}$, respectively, are illustrated in Figure 1 as examples. Different from triangle fraction, a higher component ratio indicates weaker ties among user nodes. It is considered as a complementary feature to triangle fraction.

Because our task is to predict the users who would adopt a hashtag, the users who have been "exposed" to the hashtag through the social relationships in Twitter could be potentially a very important feature. We derive two features for this purpose: F_{x10} and F_{x11}, known as *border user count* and *exposure vector*, respectively. With respect to the global graph, border users are those have at least one edge from users in G_t^h but have not adopted hashtag h, that is, $\{u_q \mid \exists e(u_p, u_q), u_p \in U_t^h, u_q \notin U_t^h\}$. Illustrated in Figure 1, nodes 1 to 6 are the users who have adopted hashtag h and nodes 7 to 12 are border users. Reported by Romero et al. (2011), a border user with more exposures to a hashtag is

DOI: 10.1002/asi

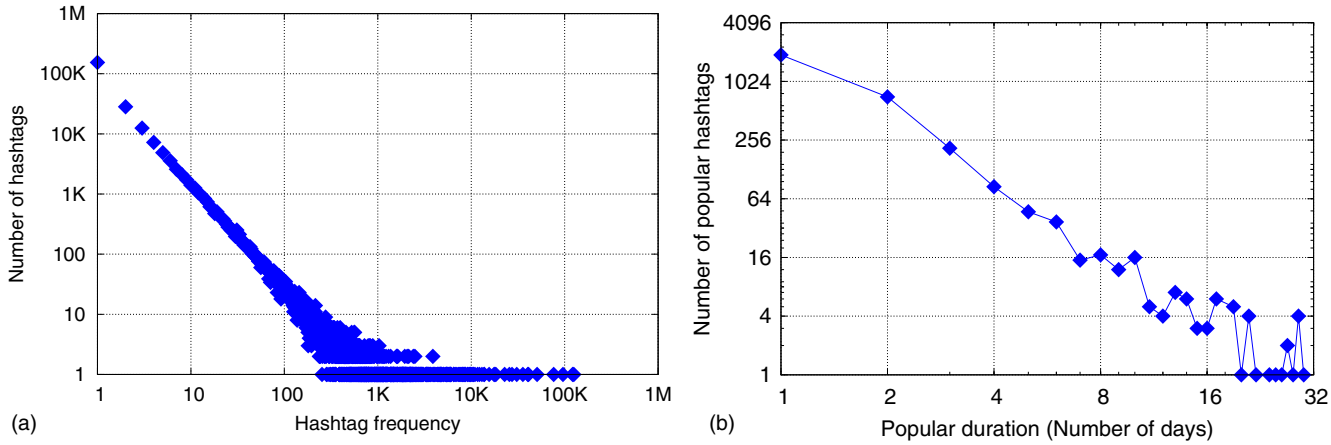

FIG. 2. Hashtag frequency distribution (a) and hashtag popularity duration distribution (b). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

expected to be more likely to adopt the hashtag. The number of exposures of a border user is the number of edges it has from the users in G_t^h. For example, in Figure 1, node 12 is exposed twice for having edges from nodes 3 and 6, respectively. The feature *exposure vector* is the exposure probability vector depicting border user distribution in a more detailed manner. It is a k-dimension vector, and the value for the k's dimension $P(k)$ is the ratio of the border users who have k edges from users in G_t^h. In our data set, we observe that a user can be exposed by a maximum number of 15 users; we therefore use a 15-dimension vector for the exposure vector and compute $P(1)$ to $P(15)$.

5 Experiments

5.1 Data Collection and Experimental Setting

Our data set consists of more than 31 million tweets from more than 2 million users. The tweets were published by Singapore-based users (based on the location specified in user profile) from January 1, 2011, to August 31, 2011. In our data set, more than 8.88% tweets contain hashtags, and the hashtag frequency distribution is plotted in Figure 2a. In Figure 2a, hashtag frequency refers to the number of tweets containing the hashtag. Observe from Figure 2a, the hashtag frequency distribution follows a power–law distribution similar to observations made in many other social data. Most hashtags were adopted by only a few users and were not popular. The global user graph contains 214,000 users and 680,000 edges. Note that only users who participated in mentions are included in this graph, and edge weight is proportional to mention times.

In our evaluation, we set the time interval to be *a day* and set $\phi = 25$; that is, a hashtag used by fewer than 25 users in a day is considered not popular. Note that $\phi = 25$ is a subjective setting based on our data set. A large ϕ may lead to insufficient instances in our evaluation, and a small ϕ may bring in too many noisy hashtags in the evaluation. A key issue in the experiments is that, among all new hashtags appearing at least once in a day, which hashtags should be selected for popularity prediction. Figure 2b plots the number of popular hashtags ($\Phi_t^h \geq \phi$) against their popularity duration in number of days. Observe that a large number of hashtags are popular for only a day. In fact, a much larger number of hashtags has never been popular (see Figure 2a). We therefore choose to predict the popularity of *newly appearing hashtags* at time $t+1$ that are *at least marginally popular* at time t, $\Phi_t^h \geq \phi$. A hashtag is considered *new* if it has not gained marginal popularity in the past 7 days. Using these two criteria, the number of hashtag instances (i.e., h_t's) falling into the five categories are listed in Table 2 under the first row "1-day prediction." The meaning of "2-day prediction" will be presented in our case study. In other words, given a hashtag that has not been popular in the past 7 days and has gained marginal popularity in the current day, we aim to predict its popularity for the next day.

TABLE 2. Hashtag distribution by categories.

Category	Not popular	Marginally popular	Popular	Very popular	Extremely popular
Φ_t^h range	$[0, \phi)$	$[\phi, 2\phi)$	$[2\phi, 4\phi)$	$[4\phi, 8\phi)$	$[8\phi, +\infty)$
1-Day prediction	1,609	614	353	211	111
2-Day prediction	685	355	186	131	49

5.2 Impact of Features

We conducted 10-fold cross validation and evaluated prediction accuracy using Micro-F_1, Macro-*Pr*ecision, Macro-*Re*call, and Macro-F_1. Because each instance has exactly one correct label, Micro-*Pr*ecision/*Re*call is the same as Micro-F_1.

Table 3 reports the prediction accuracies by the eight methods (see Problem Setting and Prediction Methods section) in ascending order of Micro-F_1. For each method,

DOI: 10.1002/asi

TABLE 3. Hashtag popularity prediction accuracy by eight methods.

Method	Features	Micro-F_1	Macro-*Pr*	Macro-*Re*	Macro-F_1
Random	–	.197	.198	.205	.163
Lazy	–	.254	.251	**.450**	.317
PriorDist	–	.385	.209	.210	.209
NB	F_c	.326	.263	.337	.235
	F_x	.345	.351	.367	.310
	F_{c+x}	.405	.348	.401	.348
KNN	F_c	.432	.317	.341	.326
	F_x	.501	.385	.383	.383
	F_{c+x}	.502	.398	.402	**.399**
C4.5	F_c	.488	.350	.332	.339
	F_{c+x}	.523	.382	.376	.378
	F_x	.534	.402	.373	.384
SVM	F_c	.555	.111	.200	.143
	F_x	.572	.316	.302	.261
	F_{c+x}	.585	.414	.330	.310
LR	F_c	.535	.235	.211	.174
	F_x	.592	.439	.346	.347
	F_{c+x}	**.598**	**.461**	.393	.396

Note. The highest accuracy achieved for each measure is shown in boldface.

LR = logistic regression; KNN = *k*-nearest neighbors; NB = Naïve bayes; SVM = support vector machines.

we conducted experiments with content features (F_c), contextual features (F_x), and all features (F_{c+x}). From the table, we make the following observations: First, three baseline methods, which do not use any feature, perform the worst. Second, generally, for a given classification method, contextual features lead to better prediction accuracy than content features, and the best accuracy is usually achieved using both content and contextual features. For instance, KNN using contextual features achieves 16% of increment over KNN using content features. The only exception is C4.5, where C4.5 with all features performs worse than that with contextual features. One possible reason is that high feature dimension complicates the decision tree model and hurts the prediction accuracy. Third, LR achieves the best Micro-F_1 of .598, which is triple of random and 55% of increment over the best baseline PriorDist. SVM is the second best performing method by Micro-F_1. Surprisingly, lazy prediction yields the best Macro-*Re*. The main reason is the skewed distribution of data (see Table 2). Because of the small number of instances in *very popular* and *extremely popular* categories, most classifiers fail to learn effective patterns for accurate prediction for these two categories. Lazy prediction enjoys high accuracy mainly in these two categories: by Macro-F_1, KNN is the best performing method followed by LR.

To better understand the effectiveness of content and contextual features, we rank all features based on their $\mathcal{X}^2$ scores. The 15 most effective features and 15 least effective features are listed in Table 4. Because some of the features listed in Table 1 are multidimensional (e.g., F_{x11} exposure vector is a 15-dimensional vector), the complete feature space used in our prediction is 53-dimensional.

TABLE 4. The 15 most effective features (rank 1–15) and 15 least effective features (rank 39–53).

Rank	Feature	Rank	Feature
1	F_{x1}: *UserCount*	39	F_{c5}: *TopicVector—T*(2)
2	F_{x10}: *BorderUserCount*	40	F_{c5}: *TopicVector—T*(14)
3	F_{x2}: *TweetsNum*	41	F_{x9}: *AveEdgeStrength*
4	F_{c6}: *HashtagClarity*	42	F_{c5}: *TopicVector—T*(17)
5	F_{x6}: *TriangleFrac*	43	F_{x8}: *ComponentRatio*
6	F_{x11}: *ExposureVector—P*(15)	44	F_{c5}: *TopicVector—T*(20)
7	F_{x11}: *ExposureVector—P*(14)	45	F_{c5}: *TopicVector—T*(9)
8	F_{x11}: *ExposureVector—P*(9)	46	F_{c5}: *TopicVector—T*(1)
9	F_{x11}: *ExposureVector—P*(10)	47	F_{c4}: *PosRatio*
10	F_{c5}: *TopicVector—T*(13)	48	F_{x5}: *AveAuthority*
11	F_{x11}: *ExposureVector—P*(11)	49	F_{c4}: *NegRatio*
12	F_{x11}: *ExposureVector—P*(5)	50	F_{c7}: *SegWordClarity*
13	F_{x11}: *ExposureVector—P*(8)	51	F_{c4}: *NeuRatio*
14	F_{x11}: *ExposureVector—P*(7)	52	F_{c2}: *SegWordNum*
15	F_{x11}: *ExposureVector—P*(12)	53	F_{c1}: *ContainingDigits*

The main observation made from Table 4 is consistent with that from Table 3; that is, contextual features are more effective than content features. Observe that *user count* is the most important feature, indicating that $\mathbf{\Phi}^h_{t+1}$ has strong relations with $\mathbf{\Phi}^h_t$. The next most effective feature is *border user count*, which is used to estimate the number of potential users who will adopt a hashtag. Because of the high correlation between *user count* and *number of tweets*, number of tweets plays a similar role as user count in hashtag popularity prediction. In addition to these three contextual features, *TriangleFrac* derived from the community graph describing the connectivity among users is also one of the most effective features. Among the top 15 most effective features, 9 are elements of the *exposure vector*. Recall that exposure vector is a 15-dimension vector describing the distribution of border users based on their exposure times. The results in Table 4 reveal a strong relationship between the times of exposure and the adoption of the hashtag. In general, larger exposure probability leads to more attention from users to a hashtag that increases the probability of hashtag adoption. In our analysis, *average authority*, however, is not as effective as expected. One possible reason is that most hashtags related to breaking news or ongoing events/topics are not created or retweeted by influential users. Such hashtags require no promotion from the influential users before their wide spread in Twitter.

Among all content features, *hashtag clarity* is the most effective one. Discussed earlier in the Features for Hashtag Popularity Prediction section, hashtag clarity quantifies topical cohesiveness of tweets in T^h_t. Hashtags with higher clarity scores usually have clear semantic meanings and refer to some specific topics or events (e.g., #royalwedding, #cancer). One dimension of topic vector (topic 13) has fairly good predicting ability, whereas most other dimensions are listed under the least effective features. To evaluate the effectiveness of topic vector, we conducted another set of experiments using the best performing classifier LR on the

DOI: 10.1002/asi

TABLE 5. Selected continuous and bursty hashtags.

Hashtag category	25 Selected hashtags
Continuous hashtags	#likeaboss, #cancer, #singapore, #sgpolitics, #sgedu, #pisces, #sg, #nowplaying, #fail, #justsaying, #fb, #ge2011, #fml, #badsgjokes, #mentionto, #greatsingaporesale, #apple, #disneywords, #ff, #happything, #mydreamjob, #hardtruths, #life, #love
Bursty hashtags	#arsenal, #sgfootball, #tsunami, #bbcwedding, #goldenglobe, #vma, #americanidol, #ndp2011, #supermoon, #londonriot, #earthquake, #sgelections, #happybirthdaychrisbrown, #sgheatwave, #billboardawards, #sgpresident, #gleefinale, #supportjapan, #f1, #oscar, #twittercrush, #ukriots, #ios5, #royalwedding, #2ne1lonely

feature set without *topic vector*. The prediction accuracy by Micro-F_1 declines 3% compared with the result of using *topic vector*, that is, the full feature set. In short, although *topic vector* is not as effective as most contextual features, it contributes to better prediction accuracy. Listed in Table A1, the topical keywords (by their generating probability under each topic) and their related hashtags indicate that the topic model does capture the general topics of the hashtags through their annotated tweets. These topics include bursty events (e.g., presidential election in Singapore, Japan earthquake, royal wedding), topics about celebrity (e.g., Lady Gaga, Taylor Swift, Justin Bieber), and daily life topics (e.g., birthday, weather, dinner, home). A closer look at the topics reveals that words from event-related topics are more cohesive. Listed in Table 4, the least effective features are sentiment vector and lexical features extracted from hashtag itself such as *ContainingDigits* and *SegWordNum*.

5.3 Case Study: Bursty Versus Continuous Hashtags

Reported by Lehmann et al. (2012), hashtags of the same category follow similar trends. Three categories (i.e., *bursty*, *continuous*, and *periodic* hashtags) are defined in their work. It is interesting to investigate whether there is any difference in prediction accuracy for hashtags of different categories. Because of the relatively small number of periodic hashtags (e.g., #gossipgirl, #bigbangtheory) in our data set, we conduct experiments to compare the prediction accuracy for bursty and continuous hashtags. Following the algorithm proposed by Lehmann et al. (2012), we chose the top 25 bursty hashtags in our evaluation. For continuous hashtags, we sorted the hashtags by their repetition frequency[6] and picked the top 25 as continuous hashtags. Table 5 lists the selected bursty and continuous hashtags. As expected, bursty hashtags mostly capture some important events during the 8 months covered by our data set. For example, #goldenglobe is about Golden Globe Awards, #sgpresident talks about the president election campaign in Singapore, and #supportjapan refers to the Japan earthquake. In contrast, the continuous hashtags selected likely gain long-period popularity (e.g., #cancer, #singapore, and #love).

Using all features and LR as the classifier, the Micro-F_1 achieved for bursty and continuous hashtags is .640 and .560, respectively. Compared with the overall prediction accuracy of .598, we conclude that the *popularity* of bursty hashtags can be predicted more effectively than continuous hashtags using the features evaluated.

5.4 Case Study: 2-Day Prediction

Our task of predicting the popularity of newly popular hashtag can be easily extended to be predicting popularity of hashtags that have been popular for 1 or more days. Nevertheless, as shown in Figure 2b, few hashtags are popular for a long time. Because of the limited number of train/test instances, we only evaluate the effectiveness of the features for popularity prediction of hashtags that have been popular for 2 days. The task then becomes that given hashtag h that has been at least marginally popular for time $t-1$ and t (i.e., $\Phi^h_{t-1} > \phi$ and $\Phi^h_t > \phi$), we predict the category of Φ^h_{t+1}. The number of instances by the five categories is reported in Table 2.

To extract features for h_{t-1} and h_t, there are two straightforward approaches: (a) *aggregation* to consider the tweets annotated by hashtag h at time $t-1$ and time t as one collection, and derive content and contextual features from these tweets; and (b) *concatenation* to extract content/contextual features for h_{t-1} and h_t, respectively, as we do in our earlier task and then concatenate the two feature vectors. In other words, the dimension of concatenated feature vector is doubled. Compared with the 1-day prediction, the aggregation approach for 2-day prediction extracts features from a larger sample of tweets (and their users); the concatenation approach, in contrast, could capture the feature changes from time $t-1$ to time t.

Figure 3 shows the Micro-/Macro-F_1 by the eight classification methods with aggregated features and concatenated features, respectively. Similar to our earlier observations, all five classification methods achieve better prediction accuracy than the three baseline methods. LR performs the best by Micro-F_1, and C4.5 is the best performing method by Macro-F_1. We also observe that methods with aggregation features slightly outperform methods with concatenation features by Micro-F_1 measure.

6 Discussion

In this article, we propose methods to predict the popularity of new hashtags on Twitter. We provide a comprehensive evaluation of both content and contextual features for short-term prediction. Some of the features evaluated in this

[6]Recall that in the problem setting, a *new* hashtag is defined as a hashtag not gaining popularity in the past 7 days at the time of evaluation. The same hashtag may be considered as "newly popular" hashtag at different time points.

DOI: 10.1002/asi

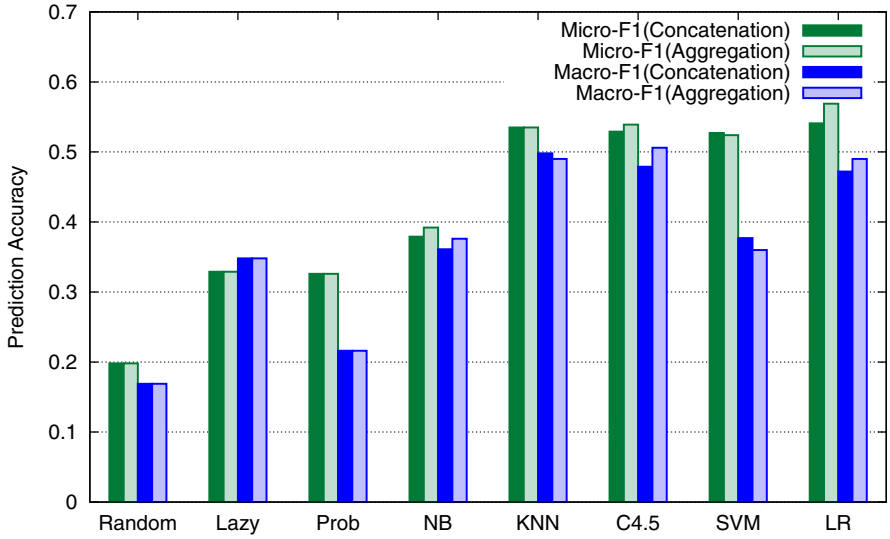

FIG. 3. Accuracy for 2-day prediction. LR = logistic regression; KNN = *k*-nearest neighbors; NB = Naïve bayes; SVM = support vector machines. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

article have been used in other studies (see Related Work section for a survey). For instance, a subset of the contextual features, including *triangles*, *density*, and *exposure*, was used in Romero et al. (2011) to study the differences in the mechanics of information diffusion of hashtags of different categories. Our evaluation complements these studies by showing the effectiveness of these features in short-term hashtag popularity prediction. On the other hand, to the best of our knowledge, our work is the first to conduct topic modeling for hashtag analysis in Twitter and to quantify hashtag topical cohesiveness using clarity. Our evaluation results showed that the two novel content features are effective in hashtag popularity prediction.

We note that our study is significantly different from the work on hashtag frequency prediction on a weekly basis (Tsur & Rappoport, 2012) in three aspects. First, we predict the number of users who will adopt a hashtag, not the number of tweets annotated by the hashtag. We therefore derive features from the community graph formed by users who have adopted a hashtag; these contextual features are not used in Tsur and Rappoport (2012). Second, our content features are derived from both a hashtag itself and the content of the tweets annotated by the hashtag. Almost all content features in Tsur and Rappoport (2012), however, are derived from the hashtag itself. Third, we target time-critical applications by predicting hashtag popularity on a daily basis rather than a weekly basis. This is important because most bursty hashtags are popular only for a few days (Lehmann et al., 2012).

Our technique can benefit advertising and public relations companies by providing predictions on the popularity of hashtags related to the organization in a timely manner. However, as new hashtags are constantly introduced by Twitter users, determining whether a new hashtag is related to an organization remains an open problem. Last, we believe that our techniques can be easily extended to predict the popularity of any predefined string (e.g., a company name, brand, or product name) by considering each such string as a hashtag and monitoring all tweets containing the strings. For instance, if a tweet contains a predefined product name, we can consider the tweet contains the product's "hashtag." Together with content and sentiment analysis of all tweets containing these product hashtags, popularity prediction can lead to more successful marketing and PR campaigns.

7 Conclusion and Future Work

In this article, we propose methods to predict the hashtag popularity of new topics on Twitter by formulating the problem as a classification task and evaluating three baseline methods and five classification methods. The main focus of our work was to identify and evaluate the effectiveness of content and contextual features derived from tweets annotated with candidate hashtags. Our experiments demonstrated that contextual features are more effective than content features. This is consistent with the finding that the property of the community graph plays a dominant role in information diffusion. We also show that our prediction technique is more effective for bursty hashtags than continuous hashtags.

In our future work, we will analyze other potentially useful features, and more importantly, propose more effective models for hashtag popularity prediction. As reported in the Related Work section, 21.7% of hashtags co-occur with Twitter URLs in tweets in our data set. Each URL links to a web document. This type of information can be used to enrich hashtag representation (e.g., by a transfer learning approach). Content features from web documents will be presented and evaluated in our future work. In addition, the

DOI: 10.1002/asi

model for 2-day prediction does not outperform that for 1-day prediction as expected. One possible reason is that the hashtag's change in popularity over the 2 days is not fully utilized. Hence we plan to investigate new techniques to fully use all available information for hashtag popularity prediction.

References

Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). Group formation in large social networks: Membership, growth, and evolution. In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 44–54). New York, NY: ACM.

Blei, D.M., Ng, A.Y., & Jordan, M.I. (2003, March). Latent dirichlet allocation. Journal of Machine Learning Research, 3, 993–1022.

Calais Guerra, P.H., Veloso, A., Meira, W. Jr., & Almeida, V. (2011). From bias to opinion: A transfer-learning approach to real-time sentiment analysis. In Proceedings of the 17th International Conference on Knowledge Discovery and Data Mining (pp. 150–158). New York, NY: ACM.

Dong, A., Zhang, R., Kolari, P., Bai, J., Diaz, F., Chang, Y., . . . Zha, H. (2010). Time is of the essence: Improving recency ranking using Twitter data. In Proceedings of the 19th International Conference on World Wide Web (WWW) (pp. 331–340). New York, NY: ACM.

Efron, M. (2010). Hashtag retrieval in a microblogging environment. In Proceedings of the 33rd ACM SIGIR International Conference on Research and Development in Information Retrieval (pp. 787–788). New York, NY: ACM.

Huberman, B.A., Romero, D.M., & Wu, F. (2009, January 5). Social networks that matter: Twitter under the microscope. First Monday, 14. Retrieved from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2317/2063

Jeon, J., Croft, W.B., Lee, J.H., & Park, S. (2006). A framework to predict the quality of answers with nontextual features. In Proceedings of the 29th ACM SIGIR Annual International Conference on Research and Development in Information Retrieval (pp. 228–235). New York, NY: ACM.

Kasiviswanathan, S.P., Melville, P., Banerjee, A., & Sindhwani, V. (2011). Emerging topic detection using dictionary learning. In Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM) (pp. 745–754). New York, NY: ACM.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? In Proceedings of the 19th International Conference on World Wide Web (WWW) (pp. 591–600). New York, NY: ACM.

Lehmann, J., Goncalves, B., Ramasco, J.J., & Cattuto, C. (2012). Dynamical classes of collective attention in Twitter. In Proceedings of the 19th International Conference on World Wide Web (WWW) (pp. 251–260). New York, NY: ACM.

Leskovec, J., Backstrom, L., Kumar, R., & Tomkins, A. (2008). Microscopic evolution of social networks. In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 462–470). New York, NY: ACM.

Leskovec, J., Kleinberg, J., & Faloutsos, C. (2005). Graphs over time: Densification laws, shrinking diameters and possible explanations. In Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (pp. 177–187). New York, NY: ACM.

Li, C., Sun, A., & Datta, A. (2012, October). Twevent: Segment-based event detection from tweets. In Proceedings of ACM Conference on Information and Knowledge Management (CIKM). New York, NY: ACM.

Liu, Q., Agichtein, E., Dror, G., Gabrilovich, E., Maarek, Y., Pelleg, D., & Szpektor, I. (2011). Predicting web searcher satisfaction with existing community-based answers. In Proceedings of the 34th ACM SIGKDD International Conference on Research and Development in Information Retrieval (pp. 415–424). New York, NY: ACM.

Liu, Y., Huang, X., An, A., & Yu, X. (2007). Arsa: A sentiment-aware model for predicting sales performance using blogs. In Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 607–614). New York, NY: ACM.

Ma, Z., Sun, A., & Cong, G. (2012). Will this # hashtag be popular tomorrow? In Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 1173–1174). New York, NY: ACM.

Naaman, M., Becker, H., & Gravano, L. (2011, May). Hip and trendy: Characterizing emerging trends on Twitter. Journal of the American Society for Information Science and Technology, 62(5), 902–918.

Pal, A., & Counts, S. (2011). Identifying topical authorities in microblogs. In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (WSDM) (pp. 45–54). New York, NY: ACM.

Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter. In Proceedings of the 20th International Conference on World Wide Web (WWW) (pp. 695–704). New York, NY: ACM.

Rowlands, T., Hawking, D., & Sankaranarayana, R. (2010). New-web search with microblog annotations. In Proceedings of the 19th International Conference on World Wide Web (WWW) (pp. 1293–1296). New York, NY: ACM.

Sakaki, T., Okazaki, M., & Matsuo, Y. (2010). Earthquake shakes Twitter users: Real-time event detection by social sensors. In Proceedings of the 19th International Conference on World Wide Web (WWW) (pp. 851–860). New York, NY: ACM.

Schultz, F., Utz, S., & Gritz, A. (2011). Is the medium the message? Perceptions of and reactions to crisis communication via Twitter, blogs and traditional media. Public Relations Review, 37(1), 20–27. doi:10.1016/j.pubrev.2010.12.001

Thelwall, M., Buckley, K., & Paltoglou, G. (2011, February). Sentiment in Twitter events. Journal of the American Society for Information Science and Technology, 62(2), 406–418.

Tsur, O., & Rappoport, A. (2012). What's in a hashtag? Content based prediction of the spread of ideas in microblogging communities. In Proceedings of the Fifth ACM International Conference on Web Search and Data Mining (WSDM) (pp. 643–652). New York, NY: ACM.

Wang, X., Wei, F., Liu, X., Zhou, M., & 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 (CIKM) (pp. 1031–1040). New York, NY: ACM.

Wei, J., Bu, B., & Liang, L. (2012). Estimating the diffusion models of crisis information in micro blog. Journal of Informetrics, 6(4), 600–610. doi:10.1016/j.joi.2012.06.005

Welch, M.J., Schonfeld, U., He, D., & Cho, J. (2011). Topical semantics of Twitter links. In Proceedings of the Fourth ACM International Conference on Web Search and Data Mining (WSDM) (pp. 327–336). New York, NY: ACM.

Weng, J., Lim, E.-P., Jiang, J., & He, Q. (2010). Twitterrank: Finding topic-sensitive influential twitterers. In Proceedings of the Third ACM International Conference on Web Search and Data Mining (WSDM) (pp. 261–270). New York, NY: ACM.

Yang, L., Sun, T., Zhang, M., & 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) (pp. 261–270). New York, NY: ACM.

DOI: 10.1002/asi

Appendix

List of 20 Topics

TABLE A1. Topics with keywords and example hashtags.

Topic ID	Topical keywords and example hashtags
Topic 01	school time life day money homework sleep study late home class hard friends test fun #firstdayofsummer, #thismorning
Topic 02	pap pm lee rally vote party singapore opposition grc singaporeans people george govt #sg, #sgelection, #sgpolitics
Topic 03	mum dad mom teacher bieber justin fat steven lim lesbians loves called #happyfathersday, #happymothersday
Topic 04	singapore day national ndp chinese home happy song proud red free country fun #ndp, #sosingaporean, #happybirthdaysingapore
Topic 05	ipad tak nak kita saf ni aku dah app la ah angry lagi kan kl yang makan macam army #ipadsforsale, #lessstrict
Topic 06	day time morning rain bus sunday night home dinner weekend week weather train car #fb, #fail, #presidentswouldsay
Topic 07	song lady listening taylor time swift gaga ne mars perry justin bruno katy baby #nowplaying, #bornthisday, #vma
Topic 08	sg ah la lol eh omg damn lah leh liao sia time haha call watch die top watching #aprilfool, #badsgjokes
Topic 09	curry bangla love bangala bad lady gaga lol rock run trending rolling day black katy #curryday, #replacesongnameswithbangala
Topic 10	trending happy birthday trend omg love shine tt awesome worldwide singapore day lol #happybaeday, #happywooday
Topic 11	love world heart day life wanna night feel miss girl beautiful perfect cry smile #everygirl, #gladyoucame, #iloveyoubecause
Topic 12	time feel start watch gonna bad damn lol real wtf guess bring hear till reading #ff, #justsaying, #sad
Topic 13	tan tony president cheng vote ah results tt yam votes bock mee cna election win #sgpresident, #sgpresidentialelection
Topic 14	people talk person stop talking bitch twitter ur stupid wrong ppl fucking #beforetwitter, #bigmistake, #mentionto
Topic 15	watching omg watch kate prince wedding game william united goal match live dress #royalwedding, #bbcwedding
Topic 16	pants bacon harry potter voldemort grind days transformers movie lol fast king #moviesilove, #nowwatching
Topic 17	pancakes super generation snsd girls junior lol simple pancake sooyoung time boys #snsdtour2011, #ilovesnsd, #snsdvisualdreams
Topic 18	love hate im friends lol dont eat people food family music lot sleep play school #meatschool, #wheneverimbored
Topic 19	japan hope god people world news safe earthquake pray tsunami stay dead hit stop strong #japan, #helpjapan, #earthquake
Topic 20	duck school students twitter class trending facebook social teachers tv playing top sec school #sgedu, #iftwitterwashighschool

DOI: 10.1002/asi