Learning Topical Social Sensor on Twitter: A Longitudinal Study

First Author[1] **and Second Author**[2]

[1]**Address of first author**
[2]**Address of second author**

ABSTRACT

Social media sources such as Twitter represent a massively distributed social sensor over a kaleidoscope of topics ranging from social and political events to entertainment and sports news. However, due to the overwhelming volume of content, it can be difficult to identify novel and significant content within a broad theme in a timely fashion. To this end, this paper proposes a scalable and practical method to automatically construct social sensors for generic topics. Specifically, given minimal supervised training content from a user, we learn to identify topical tweets from millions of features capturing content, user and social interactions on Twitter. On a corpus of over 800 million English Tweets collected from the Twitter streaming API during 2013 and 2014 and learning for 10 diverse themes ranging from social issues to celebrity deaths to the "Iran nuclear deal", we empirically show that our learned social sensor automatically generalizes to unseen future content with high ranking and precision scores. Furthermore, we provide an extensive analysis of features and feature types across different topics that reveals, for example, that (1) largely independent of topic, simple terms are the most informative feature followed by location features and that (2) the number of unique hashtags and tweets by a user correlates more with their informativeness than their follower or friend count. In summary, this work provides an effective, and efficient way to learn topical social sensors requiring minimal user curation effort and offering strong generalization performance for identifying future topical content. removed novel (this work provides a novel, effective, ...)

Keywords: Social Media Sensor, Machine Learning, Social Information Retrieval

INTRODUCTION

Social media sites such as Twitter present a double-edged sword for users. On one hand these sources contain a vast amount of novel and topical content that challenge traditional news media sources in terms of their timeliness and diversity. Yet on the other hand they also contain a vast amount of spam and otherwise low-value content for most users' information needs where filtering out irrelevant content is extremely time-consuming. Hence, while it is widely acknowledged that social media sources can be used as topical content sensors (indeed, an entire European Union project was focused on related "Social Sensor" research[1].), automatically learning high-precision sensors (i.e., ranking and retrieval methods) for arbitrary topics that generalize to future unseen content have been addressed recently only by a handful of researchers (Lin et al., 2011; Yang et al., 2014; Magdy and Elsayed, 2014).

In this work, we coalesce recent ideas on learning social sensors for general topic detection. We expand these works to learn a generalizable supervised method with minimal user curation for detecting and ranking topical content over a variety of topics and on a long-term dataset. We believe that none of the earlier works covers all the aspect of the work that is presented here. Recently, (Lin et al., 2011; Yang et al., 2014; Magdy and Elsayed, 2014) have explored use of social media sensors for detection and/or tracking of general topics from Twitter. One of the key challenges on dealing with general topics and large number of tweets is automatic labeled data aquisition. Lin et al. (2011) discusses automatic labeling of tweets by using one hashtag as topic proxy. Magdy and Elsayed (2014) uses a user-defined query to label tweets and Yang et al. (2014) takes a co-training approach based on embedded URLs in the tweet and tweet text to label tweets. We build and extend on (Lin et al., 2011)'s idea of automatic labeling of tweets, however we choose a set of hashtags for each topic instead of a single hashtag which we will show to be imperative for evaluating generalization. To learn social sensors for general topic detection, (Lin et al., 2011) uses information retrieval method (language models), Yang et al. (2014) take advantage of topic modeling techniques and Magdy and Elsayed (2014) applies SVM classifier.

[1]http://www.socialsensor.eu/

Here, we leverage various supervised learning methods for the purpose of detection and ranking of topical content. However, we present a unique method for splitting hashtags and Twitter data that encourages generalization to new unseen future content. Then we proceed to train supervised classification and ranking methods to learn topical content from a large feature space of source users and their locations, terms, hashtags, and mentions. On a corpus of over 800 million English Tweets collected from the Twitter streaming API during 2013 and 2014 and covering 10 diverse topics ranging from ”social issues” to ”celebrity deaths” to the ”Iran nuclear deal”, we empirically show that two simple and efficiently trainable methods — logistic regression and naive Bayes — generalize well to unseen future topical content (including content with no hashtags) in terms of their mean average precision (MAP) and Precision@n for a range of n. Our results suggest that these sensors generalize well to unseen future topical content and provide a novel paradigm for the extraction of high-value content from social media. Furthermore, we show that terms and locations are among the most useful features — surprisingly more so than hashtags, even though hashtags were used to label the data. And perhaps even more surprisingly, the number of unique hashtags and tweets by a user correlates more with their informativeness than their follower or friend count.

In summary, we build on the recent existing works on tracking general topics and we expand these works by (1) providing a long-term study on performance of the detectors, (2) testing future generalization to novel topical content, (3) providing a novel and comprehensive longitudinal feature analysis to investigate the importance of features and their attributes in regards to detection of topical content.

RELATED WORK

The concept of social media as a sensor is prevalent in the literature and in this section we survey five related areas of active research: (1) trending topic detection, (2) tweet recommendation, (3) tracking general topics, (4) friend sensors, (5) and specific event detection such as earthquake or influenza sensors.

Trending Topic Detection represents one of the most popular types of social sensor and can be subdivided into many categories. The first general category of methods define trends as topically coherent content and focus on clustering across lexical, linguistic, temporal and/or spatial dimensions (Petrović et al., 2010; Ishikawa et al., 2012; Phuvipadawat and Murata, 2010; Becker et al., 2011; O’Connor et al., 2010; Weng and Lee, 2011). The second general category of methods define trends as temporally coherent patterns of terms or keywords and focus largely on detecting bursts of terms or phrases (Mathioudakis and Koudas, 2010; Cui et al., 2012; Zhao et al., 2011; Nichols et al., 2012; Aiello et al., 2013). The third category of methods extends the previous categories by additionally exploiting network structure properties (Budak et al., 2011). Despite this important and very active area of work that can be considered a type of social sensor, trending topic detection is intrinsically unsupervised and not intended to detect targeted topics. In contrast, the work in this paper is based on supervised learning of a specific topical social sensor derived from the topical set of hashtags provided by the user.

Tweet Recommendation represents an alternate use of social sensors and falls into two broad categories: personalized or content-oriented recommendation and retweet recommendation. For the first category, the objective of personalized recommendation is to observe a user’s interests and behavior from their user profile, sharing or retweet preferences, and social relations to generate tweets the user may like (Yan et al., 2012; Chen et al., 2012). The objective of content-oriented recommendation is to use source content (e.g., a news article) to identify and recommend relevant tweets (e.g., to allow someone to track discussion of a news article) (Krestel et al., 2015). For the second category, there has been a variety of work on retweet prediction that leverages retweet history in combination with tweet-based, author-based, and social network features to predict whether a user will retweet a given tweet (Can et al., 2013; Xu and Yang, 2012; Petrovic et al., 2011). Despite the fact that all of these methods recommend tweets, they — and recommendation methods in general — are not focused on a specific topic but rather on predicting tweets that correlate with the preferences of a specific user or that are directly related to specific content. Rather the focus with learning topical social sensors is to learn to predict for a broad theme (independent of a user’s profile) in a way that generalizes beyond existing labeled topical content to novel future topical content.

Tracking General Topics represents use of social media sensors for detecting and tracking general topics such as ”Baseball” and ”Fashion”. Researchers collect labeled data either by using a single hashtag for each topic(Lin et al., 2011), a user-defined query for each topic (Magdy and Elsayed, 2014), or co-training based on the URLs and text of the tweet (Yang et al., 2014). Lin et al. (2011) leverages language models to train models using unigrams and bigrams, Magdy and Elsayed (2014) applies SVM classifier on extracted hashtags, unigrams, users and mentions as features, and Yang et al. (2014) defines the problem as topic modeling of tweets. We expand on (Lin et al., 2011)’s work and use a set of hashtags instead of a single hashtag. We extract hashtags, mentions, unigrams, users as features inline with these works. However, we add location as another feature which we will show later that location is the second most important feature for detection of topical content. We take advantage of various supervised learning methods and provide a novel framework for learning in terms of splitting the data and hashtags as topical proxies that would ensure generalization to future unseen content. While these works provide a good basis for this work, there are many fine-grain but important differences between previous works and this work with the most important ones being: (1) we analyzed long-term sensor performance on detecting topical content over two years of Twitter data and across a variety of topics, (2) we provide a novel and clear framework for splitting hashtags to train, validation and test in a way ensuring generalization to future unseen content, (3) we present ranking in addition to correct classification while none of the other works provide ranking, (4) we deliver a comprehensive longitudinal study on features and their attributes over two years of tweets that supports our insights for learning and relevance of features to topicality while these works had little or none analysis over their features, (5) we extract *Location* as one of the features which none of these works do and as we show in our feature analysis, *Location* is the second most important feature beating even hashtags in terms of correlation with topicality.

Specific Event Detection builds social sensors as we do in this work but focuses on highly specific events such as disasters or epidemics. For the use case of earthquake detection, an SVM can be trained to detect earthquake events and coupled with a Kalman filter for localization (Sakaki et al., 2013). In another example use case to detect health epidemics such as influenza, researchers build purpose-specific classifiers targeted to this specific epidemic (Culotta, 2010; Aramaki et al., 2011), e.g, by exploiting knowledge of users' proximity and friendship along with the contageous nature of influenza (Sadilek et al., 2012). While these targeted event detectors have the potential of providing high precision event detection, they are highly specific to the target event and do not easily generalize to learn arbitrary event-based or topic-based social sensors as provided in this work.

Friend Sensors are a fourth and final class of social sensors intended for early event detection (Kryvasheyeu et al., 2014; García-Herranz et al., 2012) by leveraging the concept of the "friendship paradox"(Feld, 1991), to build user-centric social sensors. We note that our topical social sensors represent a *superset* of friend sensors since our work includes author features that the predictor may learn to use if this proves effective for prediction. However, as shown in our feature analysis, user-based features are among the least informative feature types for our topical social sensors suggesting that general social sensors benefit from a wide variety of features well beyond those of author features alone.

LEARNING TOPICAL SOCIAL SENSORS

In this section, we reduce the problem of learning topical content from large space of features and small set of examples provided by the user to the following setting that will match standard supervised learning paradigm.

Here, the problem statement is that the user has an information need for high-precision topical content from Twitter. The first step for the user is that he/she must provide labeled data to represent this information need for use in a targeted supervised learning setting. We assume that for each topic, the user will provide us with a set of hashtags. For example for the topic of *Natural Disaster*, the user will give us {#earthquake,#flood,#prayforthephilippines, ...}. The goal is, given this topic and related hashtags, return a ranked list of tweets to the user that are highly relevant to the topic and match users information needs; meaning instead of returning a set of tweets that only match "disaster" or "natural", we realize the actual information need behind the searched topic and present all tweets matching this need. To this end, we need a methodology that learns from the set of hashtags provided by users on how to pick up sensors (i.e., useful terms, mentions, hashtags, locations, and users) and weight them to ensure picking new, unseen topical hashtags in future tweets. The following discussion intends to answer how to develop such methodology.

Our objective in learning social sensors is to train an automatic system for ranking documents by their topical relevance. Formally, given an arbitrary document d and a set of topic classes $C = \{c_1, \dots, c_K\}$, we wish to train a scoring function $f{:}d \to \mathbb{R}$ over a set of training documents $D = \{d_1, \dots, d_N\}$. Considering M number of features extracted from dataset, each document $d_i \in D$ has a boolean feature vector $(d_i^1, \dots, d_i^M) \in \{0,1\}^M$ and boolean label $d_i^c \in \{0,1\}$ indicating whether the document d_i is topical (1) or not (0). We define the set of positively occurring features for a document d_i as $D_i^+ = \{d_i^j | d_i^j = 1\}_{j=1 \dots M}$ and note that D_i^+ may include features for the content of d_i (e.g., terms, hashtags) as well as its meta-data (e.g., author, location).

There are two catches that make our training setting somewhat non-standard and which underlie subtle but critical contributions in this work:

1. Manually labeling documents is time-consuming so we need a way to label a large number of tweets with minimal user curation effort; *We achieve this by using hashtags as topical proxies.*

2. We need to train our social sensor on known topical content, but tune it on novel topical validation content that ensures the tuning achieves optimal generalization; *We achieve this by excising training content from our validation data so that our scoring function hyperparameter tuning ensures generalization.*

We next explain these key innovations in detail.

A critical bottleneck for learning targeted topical social sensors is to achieve sufficient supervised content labeling. With data requirements often in the thousands of labels to ensure effective learning and generalization over a large candidate feature space (as found in social media), manual labeling is simply too time-consuming for many users and crowdsourced labels are both costly and prone to misinterpretation of users' information needs. Fortuitously, hashtags have emerged in recent years as a pervasive topical proxy on social media sites — hashtags originated on IRC chat, were adopted later (and perhaps most famously) on Twitter, and now appear on other social media platforms such as Instagram, Tumblr, and Facebook. Hence as a simple enabling insight that serves as a catalyst for effective topical social sensor learning, for each topic class $c \in C$, we leverage a (small) set of user-curated topical hashtags H^c to efficiently provide a large number of supervised topic labels for social media content. Next we will provide the formal procedure for labeling data with H^c and training.

With the data labeling bottleneck resolved, we proceed to train supervised classification and ranking methods to learn topical content from a large feature space (e.g., for Twitter, this feature space includes terms, hashtags, mentions, authors and their locations). The training process includes the following two steps:

1. **Temporally split train and validation using H^c:** As usual for machine learning methods, we divide our training data into train and validation sets — the latter for hyperparameter tuning to control overfitting and ensure generalization to unseen data. As a critical insight for topical generalization where we view identification of previously unseen hashtags as a proxy for topical generalization, we do not simply split our data temporally into train and test sets as usually done. Instead, we split H^c into two disjoint sets H^c_{train} and H^c_{val} according to a time stamp t_{split} and the first usage time stamp t^*_h of hashtags $h \in H^c$. This procedure is shown visually in Fig **??**. Formally, we define the following:

$$\begin{aligned} H^c_{\text{train}} &= \{h \mid h \in H^c \wedge t^*_h < t_{\text{split}}\}, \\ H^c_{\text{val}} &= \{h \mid h \in H^c \wedge t^*_h \geq t_{\text{split}}\}. \end{aligned}$$

Once we have split our hashtags into training and validation sets according to t_{split}, we next proceed to temporally split our training documents D into a training set D^c_{train} and a validation set D^c_{val} for topic c based on the posting time stamp t_{d_i} of each document d_i as follows:

$$\begin{aligned} D^c_{\text{train}} &= \{d_i \mid d_i \in D \wedge t_{d_i} < t_{\text{split}}\}, \\ D^c_{\text{val}} &= \{d_i \mid d_i \in D \wedge t_{d_i} \geq t_{\text{split}}\}. \end{aligned}$$

Then for each set of *train* and *val* tweets, we use the respective hashtag sets H^c_{train} and H^c_{val} for labeling each $d^c_i \in D^c_{\text{train}}$:

$$d^c_i = \begin{cases} 1 : \exists_{h \in H^c_{\text{train}}} \, h \in D^+_i \\ 0 : \text{otherwise} \end{cases} .$$

and similarly for each $d^c_i \in D^c_{\text{val}}$:

$$d^c_i = \begin{cases} 1 : \exists_{h \in H^c_{\text{val}}} \, h \in D^+_i \\ 0 : \text{otherwise} \end{cases} .$$

The critical insight here is that we not only divide the train and validation temporally, but we divide the hashtag labels temporally and label the validation data with an entirely disjoint set of topical labels from the training data. The purpose behind this training and validation data split and labeling is to ensure that learning hyperparameters are tuned so as to prevent overfitting and maximize generalization to unseen topical content (i.e., new hashtags).

2. **Training and hyper-parameter tuning:** Once D^c_{train} and D^c_{val} have been constructed, we proceed to train our scoring function f on D^c_{train} and select hyperparameters to optimize Average Precision (AP) on D^c_{val}. Once the optimal f is found for D_{val}, we return it as our final learned topical scoring function for topic t.

Having defined our topical social sensor learning paradigm, it now remains to empirically evaluate this methodology in a social media setting, which we describe next.

Feature Usage in #Tweets

Feature	Max	Avg	Median	Max entity
From	10,196	8.67	2	running_status
Hashtag	1,653,159	13.91	1	#retweet
Mention	6,291	1.26	1	tweet_all_time→null
Location	10,848,224	9,562.34	130	london
Term	241,896,559	492.37	1	rt

Feature Usage by #Users

Hashtag	592,363	10.08	1	#retweet
Mention	26,293	5.44	1	dimensionist
Location	739,120	641.5	2	london
Term	1,799,385	6,616.65	1	rt

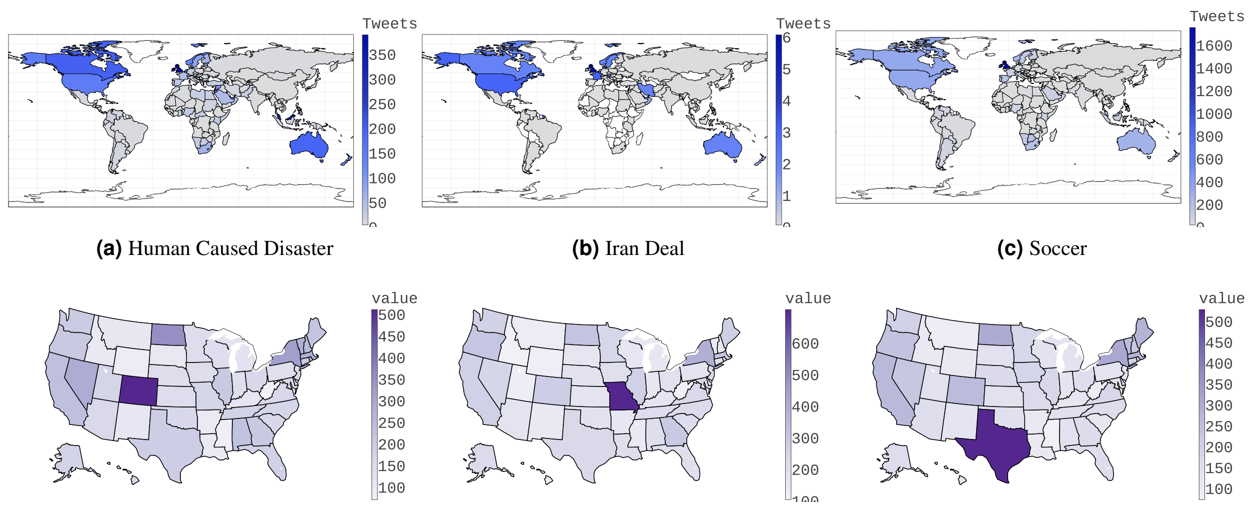
Feature Using #Hashtags

From	18,167	2	0	daily_astrodata
Location	2,440,969	1,837.79	21	uk

#Unique Features

From	Hashtag	Mention	Location	Term
95,547,198	11,183,410	411,341,569	58,601	20,234,728

Table 1. Feature Statistics of our $829,026,458$ tweet corpus.



(a) Human Caused Disaster **(b)** Iran Deal **(c)** Soccer

(d) Health Epidemics **(e)** Social Issues **(f)** Space

Figure 1. Per capita tweet frequency across different international and U.S. locations for different topics. The legend provides the number of tweets per 1 Million capita.

DATA DESCRIPTION

Now we provide details of the Twitter testbed for topical social sensor learning that we evaluate in this paper. We crawled Twitter data using Twitter Streaming API for two years spanning 2013 and 2014 years. We collected more than 2.5 TB of compressed data, which contains a total number of $829,026,458$ english tweets. In the context of Twitter, we consider five feature types for each tweet. Each tweet has a *From* feature (i.e., the person who tweeted it), a possible *Location* (i.e., a string provided as meta-data), and a time stamp when it was posted. A tweet can also contain one or more of the following:

- *Hashtag*: a topical keyword specified using the # sign.
- *Mention*: a Twitter username reference using the @ sign.

	Tennis	Space	Soccer	IranDeal	HumanDisaster	CelebrityDeath	SocialIssues	NaturalDisaster	Epidemics	LGBT
#TrainHashtags	58	98	126	12	49	28	31	31	52	29
#TestHashtags	36	63	81	5	29	16	19	19	33	17
#TopicalTweets	55,053	239,719	860,389	8,762	408,304	163,890	230,058	230,058	210,217	282,527
Sample Hashtags	#usopenchampion	#asteroids	#worldcup	#irandeal	#gazaunderattack	#robinwilliams	#policebrutality	#earthquake	#ebola	#loveislove
	#novakdjokovic	#astronauts	#lovesoccer	#iranfreedom	#childrenofsyria	#ripmandela	#michaelbrown	#storm	#virus	#gaypride
	#wimbledon	#satellite	#fifa	#irantalk	#iraqwar	#ripjoanrivers	#justice4all	#tsunami	#vaccine	#uniteblue
	#womenstennis	#spacecraft	#realmadrid	#rouhani	#bombthreat	#mandela	#freetheweed	#abfloods	#chickenpox	#homo
	#tennisnews	#telescope	#beckham	#nuclearpower	#isis	#paulwalker	#newnjgunlaw	#hurricanekatrina	#theplague	#gaymarriage

Table 2. Test/Train Hashtag samples and statistics.

	Threshold	#Unique Values
From	159	361,789
Hashtag	159	184,702
Mention	159	244,478
Location	50	57,767
Term	50	317,846
Features (CF)	-	1,166,582

Table 3. Cutoff threshold and corresponding number of unique values of candidate features *CF* for learning.

		Tennis	Space	Soccer	IranDeal	HumanDisaster	CelebrityDeath	SocialIssues	NaturalDisaster	Epidemics	LGBT	Mean
LR	**AP**	**0.918**	0.870	0.827	0.811	0.761	0.719	0.498	**0.338**	**0.329**	**0.165**	**0.623±0.19**
NB	**AP**	0.908	**0.897**	0.731	**0.824**	**0.785**	**0.748**	**0.623**	0.267	0.178	0.092	0.605±0.22
Rocchio	**AP**	0.690	0.221	**0.899**	0.584	0.481	0.253	0.393	0.210	0.255	0.089	0.407±0.18
RankSVM	**AP**	0.702	0.840	0.674	0.586	0.603	0.469	0.370	0.248	0.136	0.082	0.471±0.18
LR	**P@10**	**1.000**	0.000	0.200	0.700	**0.600**	0.000	0.100	0.200	0.300	**0.500**	0.360±0.24
NB	**P@10**	**1.000**	**0.900**	0.700	0.600	**0.600**	**0.700**	**1.000**	0.100	0.400	0.100	**0.610±0.23**
Rocchio	**P@10**	0.800	0.000	**1.000**	**0.900**	0.000	0.000	0.000	**0.500**	**0.500**	0.100	0.380±0.29
RankSVM	**P@10**	**1.000**	0.800	0.600	0.800	0.400	0.300	0.000	0.100	0.000	0.200	0.420±0.26
LR	**P@100**	0.950	0.580	0.650	0.870	0.620	0.490	0.640	**0.690**	**0.790**	**0.210**	**0.649±0.15**
NB	**P@100**	**0.980**	**0.850**	0.600	**0.880**	0.750	**0.860**	**0.730**	0.230	0.090	0.190	0.616±0.23
Rocchio	**P@100**	**0.980**	0.000	**1.000**	0.690	0.170	0.000	0.280	0.170	0.680	0.120	0.409±0.28
RankSVM	**P@100**	0.730	0.720	0.310	0.700	**0.880**	0.440	0.480	0.340	0.020	0.100	0.472±0.20
LR	**P@1000**	**0.963**	**0.954**	0.816	**0.218**	0.899	0.833	**0.215**	0.192	**0.343**	**0.071**	**0.550±0.26**
NB	**P@1000**	0.954	**0.954**	0.716	**0.218**	**0.904**	**0.881**	**0.215**	**0.195**	0.141	0.060	0.524±0.28
Rocchio	**P@1000**	0.604	0.000	**0.925**	**0.218**	0.359	0.000	**0.215**	0.167	0.144	0.065	0.270±0.21
RankSVM	**P@1000**	0.799	0.922	0.764	**0.218**	0.525	0.547	**0.215**	0.173	0.154	0.064	0.438±0.22

Table 4. Performance of topical social sensor learning algorithms across metrics and topics with the mean performance over all topics shown in the right column. The best performance per metric is shown in bold.

- *Term*: any non-hashtag and non-mention unigrams.

We provide more detailed statistics about each feature in Table 1. For example, there are over 11 million unique hashtags, the most frequent unique hashtag occurred in over 1.6 million tweets, a hashtag has been used on average by 10.08 unique users, and authors (*From* users) have used a median value of 2 tweets.

Fig. 1 shows per capita tweet frequency across different international and U.S. locations for different topics. While English speaking countries dominate English tweets, we see that the Middle East and Malaysia additionally stand out for the topic of Human Caused Disaster (MH370 incident), Iran, U.S., and Europe for nuclear negotiations the "Iran deal", and soccer for some (English-speaking) countries where it is popular. For U.S. states, we see that Colorado stands out for health epidemics (both whooping cough and pneumonic plague), Missouri stands out for social issues (#blacklivesmatter in St. Louis), and Texas stands out for space due to NASA's presence there.

EMPIRICAL EVALUATION

With the formal definition of learning topical social sensors provided in Sec. and the overview of our data in Sec. , we proceed to outline our experimental methodology on our Twitter corpus. We manually curated a broad thematic range of 10 topics shown in the top row of Table 2 by annotating hashtag sets H^t for each topic $t \in T$. We used 4 independent annotators to query the Twitter search API to identify candidate hashtags for each topic, requiring an inner-annotator agreement of 3 annotators to permit a hashtag to be assigned to a topic set. Per topic, hashtags were split into train and test sets according to their first usage time stamp roughly according to a 3/5 to 2/5 proportion. The train set was further temporally subdivided into train and validation hashtag sets according to a 5/6 to 1/6 proportion. We show a variety of statistics and five sample hashtags per topic in Table 2. Here we can see that different topics had varying prevalence in the data with *Soccer* being the most tweeted topic and *IranDeal* being the least tweeted according to our curated hashtags.

As noted in Sec. , positively occurring features D_i^+ in our d_i may include *From*, *Mention*, *Location*, *Term*, and *Hashtag* features. Because we have a total of $538,365,507$ unique features in our Twitter corpus, it is critical to pare this down to a size amenable for efficient learning and robust to overfitting. To this end, we thresholded all features according to the frequencies listed in Table 3. The rationale in our thresholding was initially that all features should have the same frequency cutoff in order to achieve roughly 1 million features. However, in initial experimentation, we found that a high threshold pruned a large number of informative terms and locations. To this end, we lowered the threshold for terms and locations noting that even at these adjusted thresholds, we still have more authors than terms. We also removed common English stopwords which further reduced the unique term count. Overall, we end up with $1,166,582$ candidate features (*CF*) for learning social sensors.

0.1 Supervised Learning Algorithms

With our labeled training and validation datasets defined in Sec. and our candidate feature set *CF* defined previously, we proceed to apply different probabilistic classification and ranking algorithms to generate a score function f for learning social sensors as defined in Sec. . In this paper, we experiment with the following four state-of-the-art classification and ranking methods:

1. **Logistic Regression** using LibLinear Fan et al. (2008)
2. **Bernoulli Naïve Bayes** McCallum and Nigam (1998)
3. **Rocchio** Manning et al. (2008)
   (a centroid-based classifier)
4. **RankSVM** Lee and Lin (2014)

As outlined in Sec , tuning of hyperparameters on a validation dataset is critical. In our experiments, we tune the following hyperparameters:

- *Logistic Regression*: L_2 regularization constant C is tuned for $C \in \{1E-12, 1E-11, ..., 1E+11, 1E+12\}$.
- *Naïve Bayes*: Dirichlet prior α is tuned for $\alpha \in \{1E-20, 1E-15, 1E-8, 1E-3, 1E-1, 1\}$.
- *All Classfiers*: The number of top features M selected based on their Mutual Information is tuned for $M \in \{1E2, 1E3, 1E4, 1E5, 1166582 \text{ (all features) }\}$.

We remark that many algorithms such as Naive Bayes and Rocchio performed better with feature selection and hence we used feature selection for all algorithms (where it is possible to select all features). Hyperparameter tuning is done via exhaustive grid search and using the Average Precision (AP) to select the best scoring function f on the validation data. Once found, f can be applied to any tweet d_i to provide a score $f(d_i)$ used to rank tweets in the test data.

0.2 Performance Analysis

We now proceed to evaluate the performance of each of the four aforementioned supervised learning algorithms for the task of learning social sensors. Once each a scoring function is trained via each method, we use it to rank tweets and then compute the following ranking metrics on the resulting ranked list:

- **AP:** Average precision over the ranked list; the mean over all topics provides mean AP (MAP).
- **P@k:** Precision at k for $k \in \{10, 100, 1000\}$.

While P@10 may be a more standard retrieval metric for tasks such as ad-hoc web search, we remark that the short length of tweets relative to web documents makes it more plausible to look at a much larger number of tweets, hence the reason for also evaluating P@100 and P@1000.

Tennis	**Space**
✓rt @espntennis: shock city. darcis drops rafa in straight sets. first time nadal loses in first rd of a. major...	✗rt @jaredleto: rt @30secondstomars: icymi: mars performing a cover of @rihanna's #stay on australia's @trip...
✓@ESPNTennis: Shock city. Darcis drops Rafa in straight sets. First time Nadal loses in first rd of a...	✗voting mars @30secondstomars @jaredleto @shannonleto @tomofromearth xobest group http://t.co/dls...
✓@ESPNTennis: Djokovic ousts the last American man standing @Wimbledon, beating Reynolds 7-6...	✗rt @jaredleto_com: show everyone how much you are proud of @30secondstomars !#mtvhottest 30 seconds to...
✓Nadal's a legend. After 3 years; Definitely He's gonna be the best of all the time. Unbelievable perf...	✗rt @30secondstomars: missed the big news? mars touring with @linkinpark + special guests @afi this summer...
✓@calvy70 @ESPNTennis @Wimbledon I see, thanks for the info and enjoy #Wimbledon2014	✗rt @30secondstomars: to the right,to the left,we will fightto the death.go #intothewildonvyrt with mars, starting...
Soccer	**IranDeal**
✗rt @tomm_dogg: #thingstodobeforeearthends spend all my money.	✓rt @iran_policy: @vidalquadras:@isjcommittee has investigated 10 major subjects of iranŐs controversial #nuc...
★@mancityonlineco nice performance	✓rt @iran_policy: @vidalquadras:@isjcommittee has investigated 10 major subjects of iranŐs controversial #nuc...
★rt @indykaila: podolski: "let's see what happens in the winter. the fact is that i'm not happy with it, th...	✗rt @negarmortazavi: thank you @hassanrouhani for retweeting. let's hope for a day when no iranian fears retur...
★rt @indykaila: wenger: "i don't believe match-fixing is a problem in england." #afc	✗rt @iran_policy: iran: details of savage attack on political prisoners in evin prison http://t.co/xdzuakqdiv #iran...
✗@indykaila you never got back to me about tennis this week	✓rt @iran_policy: chairman ros-lehtinen speaking on us commitment 2 protect camp liberty residents. #iranhr...
HumanDisaster	**CelebrityDeath**
✓rt @baselsyrian: there've been peaceful people in #homs not terrorists! #assad,enemy of #humanity...	★rt @sawubona_chris: today is my birthday & also the day my hero @nelsonmandela has died. lets never...
✓what a helpless father, he can do nothing under #assad's siege!#speakup4syrianchildren http://t.co/vg...	★rt @nelsonmandela: Ňdeath is something inevitable.when a man has done what he considers to be his duty to...
★exclusive: us formally requested #un investigation; russia pressured #assad to no avail;chain of evidence...	★rt @nelsonmandela: la muerte es algo inevitable.cuando un hombre ha hecho lo que considera que es su...
★#save_aleppo from #assadwarcrimes#save_aleppo from #civilians -targeted shelling of #assad regime...	✗#jacques #kallis: a phenomenal cricketing giant of all time - #cricket #history #southafrica http://t.co/ms5p...
✓rt @canine_rights: why does the #un allow this to continue? rt@tintin1957 help raise awareness of the...	✗@sudesh1304 south africa has the most beautiful babies....so diverse,so unique...so god!! lol #durban #southa...
SocialIssues	**NaturalDisaster**
★the us doesn't actually borrow is the thing. i believe in a creationist theory of the us dollar @usanationdebt...	✗us execution in #oklahoma : not cruel and unusual? maybe just barbaric, inhumane and reminiscent of the...
★rt @2anow: according to @njsenatepres women's rights do not include this poor nj mother's right to defend...	✗#haiti #politics - the haiti-dominican crisis - i agree with how martelly is handling the situation: i totally... http...
★rt @2anow: confiscation ? how many carry permits are in the senate and assembly? give us ours or turn ...	★rt @soilhaiti: a new reforestation effort in #haiti. local compost, anyone? http://t.co/xpad0rqbjk @richardbran...
★rt @2anow: vote with your wallet against #guncontrolforest city enterprises does not support the #2a http...	✗mes cousins jamais ns hantent les nuits de duvalier #haiti #duvalier
★@2anow @momsdemand @jstines3 they dont have a plan for that,which is why they should never be allow...	✓tony burgener of @swisssolidarity says you can't compare the disaster response in #haiti with the response to...
Epidemics	**LGBT**
✓rt @who: fourteen of the susp. & conf. ebola cases in #conakry, #guinea, are health care workers, of...	★rt @jackmcoldcuts: @lunaticrex @fingersmalloy @toddkincannon @theanonliberal anthony kennedy just wro...
✗@who who can afford also been cover in government health insurance [with universal health coverage]	✗@toddkincannon your personal account, your interest. separate from your business.
✓#ebolaoutbreak this health crisis..unparalleled in modern times,Ó @who dir. aylward - requires $1 billion ...	✗why would you report someone as spam if he is not spam? @illygirlbrea @toddkincannon
✗rt @medsin: @who are conducting a survey on the social determinants of health in medical teaching. fill...	✗rt @t3h_arch3r: @toddkincannon thanks for your tl having the female realbrother. between them is 600 lbs....
✗augmentation vertigineuse de 57,4% en 1 an des actes islamophobes en france, dit le collectif contre l'is...	✗@toddkincannon who us dick trickle.

Table 5. Top tweets for each topic from *Logistic Regression* method results, marked with ✗as irrelevant, ✓as relevant and labeled as topical, and ★as relevant but labeled as non-topical

Table 4 evaluates these metrics for each topic. *Logistic Regression* is the best performing method on average except for $P@10$. We conjecture the reason for this is that *Naïve Bayes* tends to select fewer features for training, which allows it to achieve higher precision over the top of the ranked list but which causes it suffer slightly more lower down the list due to having fewer features and lower recall. These results suggest that in general both *Logistic Regression* and *Naïve Bayes* make for effective topical social sensor learners with *Naïve Bayes* being a good choice in terms of its efficiency compared to it's overall performance.

To provide more insight into the general performance of our learning topical social sensor framework, we provide the top five tweets for each topic returned by *Logistic Regression* in Table 5. We've annotated all tweets in this table with the following symbols:

- ✓: the tweet was topical according to our curated test hashtag set.
- ★: the tweet was determined to be topical through manual evaluation even though it did not contain a hashtag in our curated hashtag set (this corresponds to a false negative due to non-exhaustive labeling of the data).
- ✗: the tweet was not topical.

In general, we remark that our learning social sensor based on logistic regression performs even better than the quantitative results in Table 4 would indicate: many of the highly ranked tweets are false negatives — they are actually relevant. Furthermore, we remark that even though we use hashtags to label our training, validation, and testing data, our learning social sensor has highly (and correctly) ranked topical tweets that do not contain hashtags indicating encouraging generalization properties from a relatively small set of curated topical hashtags.

FEATURE ANALYSIS

In this section, we analyze the informativeness of our defined features in Sec and the effect of their attributes on learning targeted topical social sensors. To this end, our goal in this section is to answer the following questions:

- What are the best features for learning social sensors and do they differ by topic?
- For each feature type, do any attributes correlate with importance?

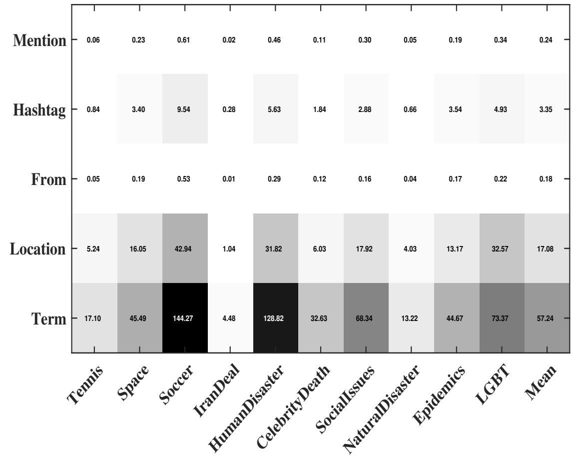


Figure 2. Matrix of mean Mutual Information values for different feature types vs. topics. The last column as average of mean values across all topics. All values should be multiplied by $1E+10$.)

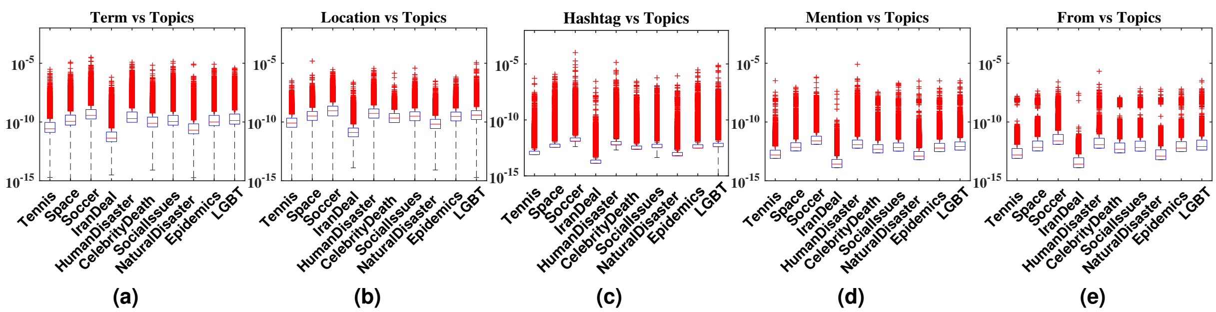


Figure 3. Box plots of Mutual Information values per feature type across topics.

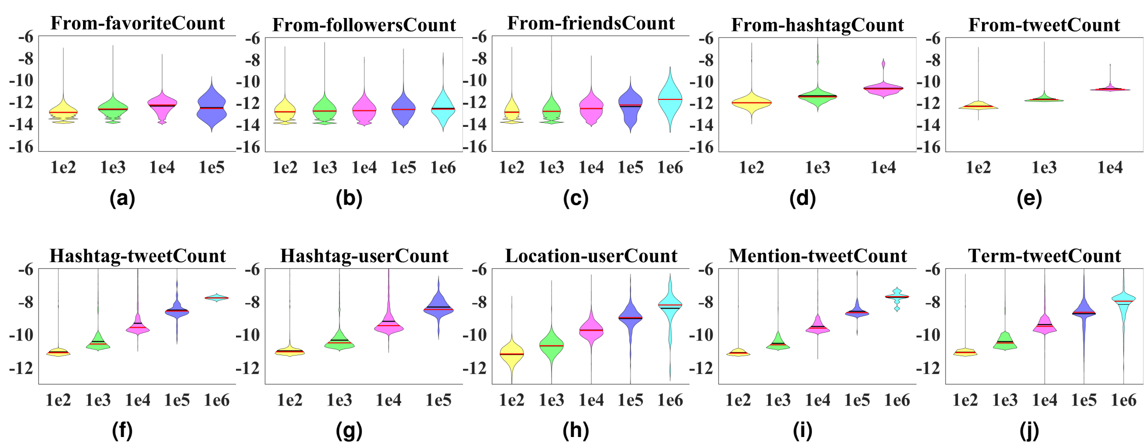


Figure 4. Violin plots for the distribution of Mutual Information values of different features as a function of their attributes. Plots (a-e) respectively show attributes {favoriteCount, followerCount, friendCount, hashtagCount, tweetCount} for *From* feature. Plots (f-j) respectively show attributes tweetCount and userCount for *Hashtag*, userCount for *Location* feature, tweetCount for *Mention* and *Term* features.

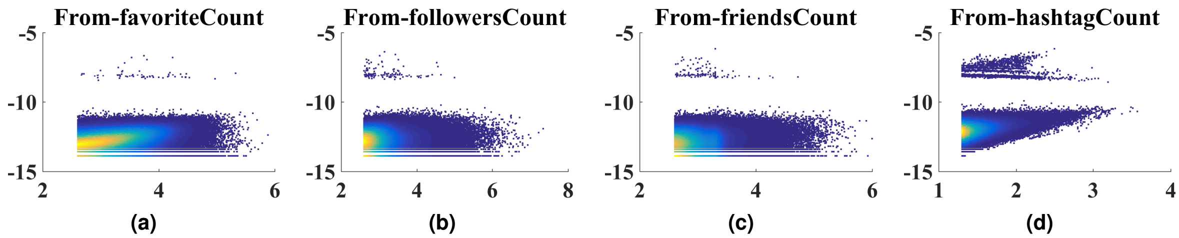


Figure 5. Density plots for the frequency values of feature attributes vs. Mutual Information. Plots (a-d) respectively show attributes {favoriteCount, followerCount, friendCount, hashtagCount} for the *From* feature.

Topics/Top10	NaturalDisaster	Epidemics	IranDeal	SocialIssues	LBGT	HumanDisaster	CelebrityDeath	Space	Tennis	Soccer
From	earthquake_wo	changedecopine	mazandara	nsingerdebtpaid	eph4_15	ydumozyf	nmandelaquotes	daily_astrodata	tracktennisnews	losangelessrh
From	earthalerts	drdaveanddee	hhadi119	debtadvisoruk	mgdauber	syriatweeten	boiknox	freesolarleads	tennis_result	shoetale
From	seelites	joinmentornetwk	140iran	debt_protect	stevendickinson	tintin1957	jacanews	houston__jobs	i_roger_federer	sport__agent
From	globalfloodnews	followebola	setarehgan	negativeequityf	lileensvf1	sirajsol	ewnreporter	star_wars_gifts	tennislessonnow	books_you_want
From	gcmcdrought	localnursejobs	akhgarshabaneh	dolphin_ls	truckerbooman	rt3syria	paulretweet	lenautilus	kamranisbest	makeupbella
Hashtag	earthquake	health	iran	ferguson	tcot	syria	rip	science	wimbledon	lfc
Hashtag	haiyan	uniteblue	irantalks	mikebrown	p2	gaza	riprobinwilliams	starwars	usopen	worldcup
Hashtag	storm	ebola	rouhani	ericgarner	pjnet	isis	ripcorymonteith	houston	tennis	arsenal
Hashtag	tornado	healthcare	iranian	blacklivesmatter	uniteblue	israel	mandela	sun	nadal	worldcup2014
Hashtag	prayforthephilippines	depression	no2rouhani	fergusondecision	teaparty	mh370	nelsonmandela	sxsw	wimbledon2014	halamadrid
Location	philippines	usa	tehran	st.louis	usa	malaysia	southafrica	germany	london	liverpool
Location	ca	ncusa	u.s.a	mo	bordentown	palestine	johannesburg	roodepoort	uk	manchester
Location	india	garlandtx	nederland	usa	newjersey	syria	capetown	houston	india	london
Location	newdelhi	oh-sandiego	iran	dc	sweethomealabama!	israel	pretoria	austin	pakistan	nigeria
Location	newzealand	washington	globalcitizen	washington	aurora	london	durban	tx	islamabad	india
Mention	oxfamgb	foxtramedia	4freedominiran	deray	jjauthor	ifalasteen	nelsonmandela	bizarro_chile	wimbledon	lfc
Mention	weatherchannel	obi_obadike	iran_policy	natedrug	2anow	revolutionsyria	realpaulwalker	nasa	usopen	arsenal
Mention	redcross	who	hassanrouhani	antoniofrench	govchristie	drbasselabuward	robinwilliams	j_ksen	andy_murray	realmadriden
Mention	twcbreaking	obadike1	un	bipartisanism	a5h0ka	mogaza	rememberrobin	jaredleto	serenawilliams	ussoccer
Mention	abc7	c25kfree	statedept	theanonmessage	barackobama	palestinianism	tweetlikegiris	30secondstomars	espntennis	mcfc
Term	philippines	health	iran	police	obama	israel	robin	cnblue	murray	madrid
Term	donate	ebola	regime	protesters	gun	gaza	williams	movistar	tennis	goal
Term	typhoon	acrx	nuclear	officer	rights	israeli	nelson	enero	federer	cup
Term	affected	medical	iranian	protest	america	killed	mandela	cimperdible	djokovic	manchester
Term	relief	virus	resistance	cops	gop	children	cory	greet	nadal	match

Table 6. The top 5 features for each feature type and topic based on Mutual Information.

To answer these questions, we use Mutual Information (MI) as our primary metric for feature evaluation. Mutual Information is a general method for measuring the amount of information one random variable contains about another random variable. In order to calculate the amount of information that each feature $j \in \{From, Hashtag, Mention, Term, Location\}$ provides w.r.t. each topic label $t \in \{NaturalDisaster, Epidemics, \ldots\}$, Mutual Information is formally defined as

$$I(j,t) = \sum_{t \in \{\text{true,false}\}} \sum_{j \in \{\text{true,false}\}} p(j,t) \log \left(\frac{p(j,t)}{p(j)p(t)} \right),$$

where higher values for this metric indicate more informative features for the specified topic.

In order to answer the first question regarding the best features for learning social sensors, we provide the mean Mutual Information values for each feature across different topics in Fig. 2. The last column in Fig. 2 shows the average of the mean Mutual Information for each feature type. From analysis of Table 2, we can make a set of observations:

- The *Term* and *Location* features are the most informative features on average.
- The Location feature provides the most information regarding the topics of *HumanDisaster*, *LBGT*, and *Soccer* indicating that a lot of content in these topics is heavily localized.
- Looking at the overall average values, the order of informativeness of feature types appears to be the following: *Term*, *Location*, *Hashtag*, *Mention*, *From*.

To further analyze the relationship between the informativeness of feature types and topics, we refer to the box plots of Fig. 3. Here we see the quartiles and outliers of the distribution rather than just the average of the MI values in order to ensure the mean MI values were not misleading our interpretations. Overall, however, the story is the same: term and location features dominate in terms of Mutual Information followed by the other less informative features. Furthermore, two observations are apparent: (1) terms have more outliers indicating that the most useful individual features may be terms, and (2) the topic has little impact on which feature is most important indicating stability of feature type informativeness over topics.

As anecdotal evidence to inspect which features are most informative, we refer to Table 6, which displays the top five feature instances for each feature type and topic. Among many remarkable insights in this table, one thing we note are that the terms appear to be the most generic (and hence most generalizable) features, providing strong intuition as to why these features figure so prominently in terms of their informativeness. The top locations are also highly relevant to most topics indicating the overall importance of these tweet features for identifying topical tweets.

In order to answer the second question on whether any attributes correlate with importance for each feature, we provide two types of analysis. The first analysis shown in Fig. 4 analyzes the distributions of Mutual Information values for features when binned by the magnitude of various attributes of those features, outlined as follows:

- **From** vs.
  - *Favorite count:* # of tweets user has favorited.
  - *Followers count:* # of users who follow user.
  - *Friends count:* # of users followed by user.
  - *Hashtag count:* # of hashtags used by user.
  - *Tweet count:* # of tweets from user.
- **Hashtag** vs.
  - *Tweet count:* # of tweets using hashtag.
  - *User count:* # of users using hashtag.
- **Location** vs. *User count:* # of users using location.
- **Mention** vs. *Tweet count:* # of tweets using mention.
- **Term** vs. *Tweet count:* # of tweets using term.

As we can see in the Violin plots of Fig. 4, the general pattern is that the greater the number of tweets, users, or hashtag count a feature has, the more informative the feature is in general. This pattern also exists to some extent on the attributes of the *From* feature, although the pattern is less visible in general and not clear (or very weak) for the follower or friend count. In general, the informativeness of a user appears to have little correlation with their follower or friend count.

Fig. 5 provides a further analysis by showing density plots of favorite count, follower count, friends count, and hashtag count attributes of the *From* feature. Here we see an interesting phenomenon that was not clear in the Violin plots: there is a very clear bimodality of the density. On further investigation it turns out that the top mode feature occurs in at least one topical tweet whereas the bottom mode occurs in no topical tweets. While the bottom mode features may serve as good indicators of non-topicality, the top mode are inherently more indicative of topicality, which justifies feature selection by mutual information.

CONCLUSIONS AND FUTURE WORK

This work fills a major gap in event detection and tracking from social media on identifying emerging topics from long-running themes with minimal user supervision. We contribute a novel supervised method for training social sensors with minimal user curation by using a small seed set of hashtags as topical proxies for automatic supervised data labeling. Our results suggest that these learned social sensors generalize well to unseen future topical content and provide a novel paradigm for the extraction of high-value content from social media. Furthermore, an extensive analysis of features and feature attributes across different topics has revealed two key insights: (1) largely independent of topic, simple terms are the most informative feature followed by location features and that (2) the number of unique hashtags and tweets by a user correlates more with their informativeness than their follower or friend count.

Among many interesting directions, future work should explore the following enhanced topical social sensor learning tasks: (1) optimizing rankings not only for topicality but also to minimize the lag-time of novel content identification, (2) optimizing queries for boolean retrieval oriented APIs such as Twitter, and (3) utilizing more social network structure to exploit a more expressive graph-based features. Altogether, we believe this and future work will pave the way for a new class of social sensors that learn to identify broad themes of topical information with minimal user interaction and enhance the overall social media user experience.

ACKNOWLEDGMENTS

So long and thanks for all the fish.

REFERENCES

Aiello, L. M., Petkos, G., Martín, C. J., Corney, D., Papadopoulos, S., Skraba, R., Göker, A., Kompatsiaris, I., and Jaimes, A. (2013). Sensing trending topics in twitter. *IEEE Transactions on Multimedia*, 15(6):1268–1282.

Aramaki, E., Maskawa, S., and Morita, M. (2011). Twitter catches the flu: Detecting influenza epidemics using Twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11.

Becker, H., Naaman, M., and Gravano, L. (2011). Beyond trending topics: Real-world event identification on twitter. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*.

Budak, C., Agrawal, D., and El Abbadi, A. (2011). Structural trend analysis for online social networks. *PVLDB*, 4(10):646–656.

Can, E. F., Oktay, H., and Manmatha, R. (2013). Predicting retweet count using visual cues. In *22nd ACM International Conference on Information and Knowledge Management, CIKM'13, San Francisco, CA, USA, October 27 - November 1, 2013*, pages 1481–1484.

Chen, K., Chen, T., Zheng, G., Jin, O., Yao, E., and Yu, Y. (2012). Collaborative personalized tweet recommendation. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 661–670, New York, NY, USA. ACM.

Cui, A., Zhang, M., Liu, Y., Ma, S., and Zhang, K. (2012). Discover breaking events with popular hashtags in Twitter. In *21st ACM International Conference on Information and Knowledge Management, CIKM'12, Maui, HI, USA, October 29 - November 02, 2012*, pages 1794–1798.

Culotta, A. (2010). Towards detecting influenza epidemics by analyzing Twitter messages. In *Proceedings of the First Workshop on Social Media Analytics*, SOMA '10.

Fan, R., Chang, K., Hsieh, C., Wang, X., and Lin, C. (2008). LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 9:1871–1874.

Feld, S. L. (1991). Why your friends have more friends than you do. *American Journal of Sociology*, pages 1464–1477.

García-Herranz, M., Egido, E. M., Cebrián, M., Christakis, N. A., and Fowler, J. H. (2012). Using friends as sensors to detect global-scale contagious outbreaks. *PloS one*, abs/1211.6512.

Ishikawa, S., Arakawa, Y., Tagashira, S., and Fukuda, A. (2012). Hot topic detection in local areas using Twitter and wikipedia. In *ARCS Workshops (ARCS), 2012*, pages 1–5.

Krestel, R., Werkmeister, T., Wiradarma, T. P., and Kasneci, G. (2015). Tweet-recommender: Finding relevant tweets for news articles. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 53–54, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Kryvasheyeu, Y., Chen, H., Moro, E., Hentenryck, P. V., and Cebrián, M. (2014). Performance of social network sensors during hurricane sandy. *PLoS one*, abs/1402.2482.

Lee, C.-P. and Lin, C.-J. (2014). Large-scale linear RankSVM. *Neural Computing*, 26(4):781–817.

Lin, J., Snow, R., and Morgan, W. (2011). Smoothing techniques for adaptive online language models: topic tracking in tweet streams. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 422–429. ACM.

Magdy, W. and Elsayed, T. (2014). Adaptive method for following dynamic topics on twitter. In *ICWSM*.

Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA.

Mathioudakis, M. and Koudas, N. (2010). Twittermonitor: trend detection over the Twitter stream. In *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2010, Indianapolis, Indiana, USA*, pages 1155–1158.

McCallum, A. and Nigam, K. (1998). A comparison of event models for naive bayes text classification. In *In AAAI-98 Workshop On Learning For Text Categorization*, pages 41–48. AAAI Press.

Nichols, J., Mahmud, J., and Drews, C. (2012). Summarizing sporting events using Twitter. In *17th International Conference on Intelligent User Interfaces, IUI '12, Lisbon, Portugal, February 14-17, 2012*, pages 189–198.

O'Connor, B., Krieger, M., and Ahn, D. (2010). Tweetmotif: Exploratory search and topic summarization for Twitter. In *Proceedings of the Fourth International Conference on Weblogs and Social Media, ICWSM 2010, Washington, DC, USA, May 23-26, 2010*.

Petrović, S., Osborne, M., and Lavrenko, V. (2010). Streaming first story detection with application to twitter. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT '10, pages 181–189, Stroudsburg, PA, USA. Association for Computational Linguistics.

Petrovic, S., Osborne, M., and Lavrenko, V. (2011). Rt to win! predicting message propagation in Twitter. In *ICWSM*.

Phuvipadawat, S. and Murata, T. (2010). Breaking news detection and tracking in Twitter. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and International Conference on Intelligent Agent Technology - Workshops, Toronto, Canada, August 31 - September 3, 2010*, pages 120–123.

Sadilek, A., Kautz, H. A., and Silenzio, V. (2012). Modeling spread of disease from social interactions. In *Proceedings of the Sixth International Conference on Weblogs and Social Media, Dublin, Ireland, June 4-7, 2012*.

Sakaki, T., Okazaki, M., and Matsuo, Y. (2013). Tweet analysis for real-time event detection and earthquake reporting system development. *Knowledge and Data Engineering, IEEE Transactions on*, 25(4):919–931.

Weng, J. and Lee, B. (2011). Event detection in Twitter. In *Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, July 17-21, 2011*.

Xu, Z. and Yang, Q. (2012). Analyzing user retweet behavior on Twitter. In *International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2012, Istanbul, Turkey, 26-29 August 2012*, pages 46–50.

Yan, R., Lapata, M., and Li, X. (2012). Tweet recommendation with graph co-ranking. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers - Volume 1*, ACL '12, pages 516–525, Stroudsburg, PA, USA. Association for Computational Linguistics.

Yang, S.-H., Kolcz, A., Schlaikjer, A., and Gupta, P. (2014). Large-scale high-precision topic modeling on twitter. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1907–1916. ACM.

Zhao, S., Zhong, L., Wickramasuriya, J., and Vasudevan, V. (2011). Human as real-time sensors of social and physical events: A case study of Twitter and sports games. *Technical Report TR0620-2011, Rice University and Motorola Mobility*, abs/1106.4300.