Evaluating GeoContext: A System for Creating Geographical Topics from a Social Media Stream

Elizabeth Williams[1], Jeff Gray[2], Brandon Dixon[3]
Department of Computer Science
University of Alabama
Tuscaloosa, AL
[1]eawilliams2@crimson.ua.edu, [2]gray@cs.ua.edu, [3]dixon@cs.ua.edu

***Abstract*—Social media can provide an insight into different contextual and temporal events that are not easily discovered with traditional search engines. In this paper, we evaluate different configurations for creating a social media analysis engine. We briefly describe our system, GeoContext, for analyzing a Twitter stream. GeoContext creates clusters of tweets that have the same topic, and then analyzes the topic clusters to determine if each cluster is centered in a geographical location. In this paper, we present our evaluation of four threshold values present in GeoContext: the threshold value of GeoContext's similarity score calculation by which two tweets are considered to have similar topics, the time between pruning sessions, at which old irrelevant clusters of tweets are removed, the time at which a cluster is considered to be irrelevant, or "stale," and the threshold value of GeoContext's geographical analysis algorithm, by which a cluster of tweets is considered to be centered at a location. We posit that these types of threshold values are likely to be present in any system for clustering and analyzing a social media stream.**

Keywords—social media, social media analysis, topic modeling

I. INTRODUCTION

Social media can provide more temporal and contextual information for certain scenarios than traditional search engines such as Google. For example, weather information is often disseminated via social media because the nature of social media is real-time and far-reaching. Social media can also be a useful tool for gathering real-time information about the spread and reach of earthquakes and other natural disasters [1]. Critical information can be spread more rapidly via social media than other traditional forms of media. Twitter and other social media platforms were used by the FBI in the aftermath of the 2013 Boston Marathon explosions to disseminate data regarding potential suspects [2]. Social media can allow the discovery of the opinions and information of the over 3 billion worldwide Internet users, rather than relying on a limited number of traditional media sources.

Approximately 6,000 tweets are published every second [3]. In order to discover information within a social media stream, the stream must be consolidated into clear and well-defined topics that represent what Twitter users are talking about. Our system, GeoContext, provides topical analysis of a stream as well as geographical analysis to determine where topics within the stream are located and centered.

GeoContext can be configured via four main threshold values, described in detail in Section II. We posit that many social media analysis with the same aim may contain similar threshold values. In this paper, we provide an evaluation of the various configurations possible for GeoContext.

GeoContext utilizes Twitter as a social media platform for several reasons. First, because tweets are limited in length, they often contain only one topic, which makes topical clustering more feasible. Also, Twitter provides methods for tweets to be geotagged, or have explicit geographical coordinates attached, making geographical analysis possible. Lastly, Twitter allows users to form both bidirectional and unidirectional relationships. However, GeoContext is able to utilize any social media platform with similar characteristics with minimal configuration.

In this paper, we analyze our assessment and evaluation of the configuration of GeoContext. In Section II, we give a brief overview of the implementation of GeoContext. In Section III, we outline the experimental process and describe the results of our evaluation. Section IV describes work related to social media analysis and evaluation, and Section V concludes the paper.

II. OVERVIEW OF GEOCONTEXT

This section provides an overview of GeoContext, our system for discovering geographical topics within a social media stream. For a more detailed description of GeoContext, please refer to [4] and [5].

A. Stream Initialization

The pipeline followed by GeoContext is shown in Fig. 1. GeoContext can accept two optional parameters prior to the Twitter stream starting: geographical coordinates and keywords. Both parameters filter the stream of tweets that are located within the given coordinates or contain the given keywords, respectively. GeoContext utilizes JoBimText [6], a distributional semantics framework, in order to expand the specified keywords into a set of related keywords. This way, the stream can contain tweets that are conceptually related to the given keywords, but may not contain the exact term. For example, if the given keyword is "weather," GeoContext will track tweets containing the word "weather," as well as tweets containing the terms "rain" and "thunderstorm."

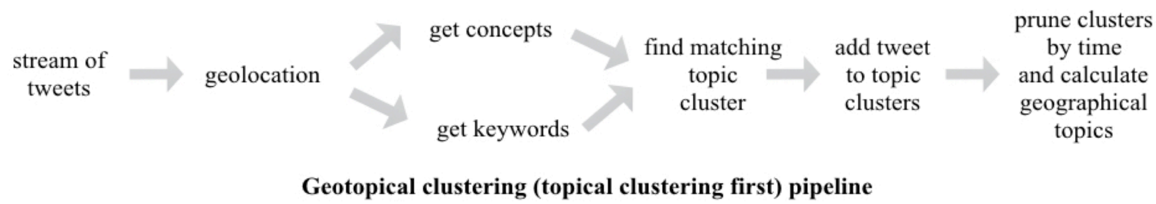

Fig. 1. GeoContext Pipeline

After parameters are received, the stream of tweets is started via the twit framework [1]. GeoContext uses the "Gardenhose" stream from Twitter, which provides an estimated 15% of the public Twitter stream [7].

B. Geolocation

Less than 3% of all tweets in the stream contain explicit coordinates [4]. However, in order to perform geographical clustering, GeoContext needs all tweets to be associated with a location. To accomplish this, the tweets are first passed through the geolocation step, which predicts the location using the user account location, content of the tweet, locations of the user's friends and followers, and the topic of the tweet to predict the tweet's location.

GeoContext gathers any potential locations from each of the four sources within the tweet, and then predicts a final set of coordinates for the tweet from the information found. GeoContext uses AlchemyAPI's Entity Extraction API [2] as well as Dbpedia, a database that includes the data found in Wikipedia infoboxes, to recognize potential locations found within the user account location and the content of the tweet. GeoContext also obtains the user account locations of the user's friends and followers and clusters them using DBSCAN [8], a density-based clustering algorithm. GeoContext then calculates the midpoint of the largest cluster, which represents the location where the majority of the user's friends and followers are located. Lastly, GeoContext calculates the topic of the tweet using AlchemyAPI's Concept Tagging API [3] and Keyword Extraction API [4]. GeoContext then determines whether tweets with the same topic are centered in a specific location. If users that are tweeting about a certain topic are located predominately in some location, then GeoContext infers that the tweet with the same topic also comes from that location.

After extracting location information from all four sources, GeoContext clusters any extracted coordinates using DBSCAN. If any clusters appear, then at least two of the sources produced similar coordinates. In that case, GeoContext chooses the midpoint of the cluster as the final predicted location for the tweet. If no cluster appears, then GeoContext chooses coordinates from one of the sources based on a preliminary study described in [4].

C. Topical Clustering

After the geolocation step, GeoContext clusters the tweet into topics. GeoContext uses the concepts and keywords extracted from the AlchemyAPI Concept Tagging and Keyword Extraction APIs, which also return a relevance score with every concept and keyword to indicate how relevant the concept or keyword is to the tweet. GeoContext calculates a similarity score by taking the average of the relevance scores of any matching concepts or keywords between two tweets. Tweets that have a score over a threshold value will be clustered together into the same topic. This threshold value is the first that we evaluate in Section III.

D. Geographical Clustering

GeoContext performs geographical analysis on the topics produced from the topical clustering step in order to determine whether each topic is centered at a location or spread across a larger region. GeoContext uses TF-IDF, a statistic used in natural language processing that shows how important or meaningful a term is to a document [9]. GeoContext calculates an adapted version of the TF-IDF statistic in order to determine how important a location is to a topic cluster. If the TF-IDF value for a location and a topic cluster is above a threshold, then the topic cluster is considered to be centered at that location. This threshold value is the second that we evaluate in Section III.

E. Pruning

After a period of time, the collection of topic clusters is pruned to remove any clusters that have not had tweets added recently, or "stale" clusters. If "stale" clusters are not removed, the storage and analysis of so many tweets can greatly affect performance. The length of time between pruning sessions is the third threshold value evaluated in Section III. The length of time between the last tweet added to a cluster and the cluster becoming "stale" is the fourth threshold value evaluated in Section III.

III. Evaluation of GeoContext Threshold Values

We performed an empirical evaluation on GeoContext in order to determine the accuracy of the threshold values used in the clustering process. We collected a dataset of 14,817 tweets throughout April 2016. We clustered the tweets using

[1] https://www.npmjs.com/package/twit
[2] http://www.alchemyapi.com/api/entity-extraction
[3] http://www.alchemyapi.com/api/concept-tagging
[4] http://www.alchemyapi.com/api/keyword-extraction

GeoContext with 48 different configurations and then tested the TF-IDF threshold with 4 different values separately.

A. Experimental Setup

In this evaluation, we tested the accuracy of the resulting clusters from GeoContext with various configurations. Specifically, we tested four different threshold values present with GeoContext: the similarity score threshold, the time value between prunings of the topic clusters, the time threshold at which a topic cluster is considered "stale," and the adapted TF-IDF threshold value.

Due to rate limits imposed by AlchemyAPI, we obtained the concepts and keywords for all tweets, as well as geolocated the tweets, prior to running the experiment. All tweets were then clustered using GeoContext. All data gathering as well as the clustering experiment was performed on a MacBook Pro 2.5 GHz with 16 GB of RAM.

TABLE I. VARIED EXPERIMENTAL VALUES

Threshold Value	***Possible Values***
Similarity Score	0.2, 0.4, 0.6, 0.8
Pruning Time	15 min, 30 min, 12 hrs, 24 hrs
Stale Cluster Time	24 hrs, 48 hrs
TF-IDF Threshold	0.2, 0.4, 0.6, 0.8

B. Evaluation

We first present the evaluation of the topical clusters produced by GeoContext. We evaluated 48 different configurations of the three threshold values that affect the topic clusters: the similarity score value, the pruning time value, and the stale cluster time value. Table 1 shows the possible values for each of these threshold values. Table 2 shows the resulting five largest topic clusters for each configuration. The concepts and keywords that matched within the similarity score calculations are shown for each topic cluster. Because the matching concepts and keywords are the factor that makes tweets within the cluster similar, we believe that they give an accurate representation of the overall topic of the cluster.

a) Similarity Score: Because relevance scores from the Alchemy API Concept Tagging and Keyword Extraction APIs range from (exclusive) 0 to 1, the similarity score value also ranges from (exclusive) 0 to 1. We decided to choose sample values of 0.2, 0.4, 0.6, and 0.8 so that the range is covered in evaluation.

As seen in Table 2, it is clear that the lower the similarity score, the broader the concepts within the topic cluster. This is not surprising, due to the fact that more keywords and concepts contribute to the similarity score of the tweets if there is a lower similarity score threshold. With a lower threshold value, several of the topic clusters contain more than one topic that is not related. For example, the (0.2, 15 min., 24 hr.) topic cluster contains tweets about both weather and people. These tweets are separated into two distinct clusters with the higher similarity score threshold values.

Interestingly, many of the five largest topic clusters for each similarity score value are the same or very similar, even as the pruning time value varies. This correlation suggests that the similarity score value and the stale cluster value are the strongest in influencing the topic clusters.

b) Pruning Time: We chose the values 15 minutes, 30 minutes, 12 hours, and 24 hours for the time between pruning sessions. We believe that waiting longer than 24 hours will keep too many old topics in the system, since many trends in Twitter are fairly short-lived. Also, there are often so many topic clusters after 24 hours that if old ones are not removed, so many tweets are analyzed within the topical clustering step that performance is affected.

As displayed in Table 2, there exists basically no discernable difference between the clusters produced by the various pruning time values, while holding the similarity score threshold value and the stale cluster threshold value constant. This indicates that the pruning time value does not have a discernible effect on the topic clusters. Therefore, the configuration of the pruning time threshold can be determined by any other means desired.

c) Stale Cluster Time: We chose the values 24 hours and 48 hours for the time threshold at which a topic cluster becomes "stale." This value represents the time between the addition of the last tweet to the cluster and the time at which the cluster becomes "stale" and should be removed. We believe that a value shorter than 24 hours would result in topic clusters being removed while they are still relevant, because trends on Twitter tend to occur over at least one day.

There is only a slight difference between the topic clusters produced by the 24 hour and 48 hour values. Unexpectedly, there are a few topics that appeared with the 24 hour value that did not appear in the 48 hour value clusters. For example, "Internet slang," "photography," and "guys" were all matching concepts or keywords that appeared in topic clusters with the 24 hour value. Prior to the experiment, we expected the 48 hour value clusters to have more range in topics because more clusters are kept, because the clusters are allowed to be older. However, the additional concepts found in the 24 hour value clusters may be overshadowed by the larger group of clusters with the 48 hour value.

Overall, the evaluation shows that the similarity score between tweets should be higher in order to produce topic clusters that consist of one topic each. Also, the time at which topics are pruned does not have any discernible effect on the topic clusters. Lastly, the time at which a topic cluster becomes stale produces more concepts within topic clusters with a lower value.

d) Adapted TF-IDF Threshold: We also evaluated the adapted TF-IDF statistic threshold value. This is the value at which a topic cluster is considered to be centered at a geographical location.

TABLE 2. TOPIC CLUSTERS WITH VARIOUS CONFIGURATIONS

Configuration (Sim. Score, Pruning time, Stale time)	*1*	*2*	*3*	*4*	*5*
0.2, 15 min, 24 hrs.	Weather, climate change, people	English-language films, American films, 1990s music groups	Apple Inc., bid	Friends, time	Thanks
0.4, 15 min, 24 hrs.	People	English-language films, American films	Friends, time	Retweets	Thanks
0.6, 15 min, 24 hrs.	People	English-language films	Time	Retweets	Thanks
0.8, 15 min, 24 hrs.	English-language films	People	Lol	Retweets	Thanks
0.2, 30 min, 24 hrs.	Weather, people	Bid, dress	Guys, thanks, mom	English-language films, American films	Thermodynamics, time
0.4, 30 min, 24 hrs.	Things, people	English-language films	Friends, time	I'm, retweets	Bid
0.6, 30 min, 24 hrs.	People	English-language films	Time	Retweets	Thanks
0.8, 30 min, 24 hrs.	English-language films	Retweets	People	Lol	Thanks
0.2, 12 hrs, 24 hrs.	Weather, people	Bid	English-language films, American films	Thermodynamics, time	Internet slang, I'm, guys, retweets
0.4, 12 hrs, 24 hrs.	People	English-language films	Time	Life, retweets	Retweets
0.6, 12 hrs, 24 hrs.	People	English-language films	Time	Life, retweets	Retweets
0.8, 12 hrs, 24 hrs.	English-language films	Retweets	People	Lol	Thanks
0.2, 12 hrs, 24 hrs.	Weather, people	Bid, dress	English-language films	Thermodynamics, time	Internet slang, retweets
0.4, 12 hrs, 24 hrs.	Things, people	English-language films	I'm, retweets	Bid	Life, photography
0.6, 12 hrs, 24 hrs.	People	English-language films	Time	Retweets	Thanks
0.8, 12 hrs, 24 hrs.	English-language films	Retweets	People	Lol	Thanks

TABLE 2 (CONT). TOPIC CLUSTERS WITH VARIOUS CONFIGURATIONS

0.2, 15 min, 48 hrs.	Bed, thermodynamics, people	Cause, heart, English-language films	Bid	Life, retweets	Time
0.4, 15 min, 48 hrs.	People	English-language films	Retweets	Time	Thanks
0.6, 15 min, 48 hrs.	People	English-language films	Retweets	Time	Thanks
0.8, 15 min, 48 hrs.	English-language films	Retweets	People	Lol	Thanks
0.2, 30 min, 48 hrs.	Bed, thermodynamics, people	Cause, English-language films	Bid	Life	Time
0.4, 30 min, 48 hrs.	People	English-language films	Time	Life, retweets	Retweets
0.6, 30 min, 48 hrs.	People	English-language films	Retweets	Time	Thanks
0.8, 30 min, 48 hrs.	English-language films	Retweets	People	Lol	Thanks
0.2, 12 hrs, 48 hrs.	Bed, thermodynamics, people	Cause, heart, English-language films	Bid	Life	Time
0.4, 12 hrs, 48 hrs.	People	English-language films	Time	Life, retweets	Retweets
0.6, 12 hrs, 48 hrs.	People	English-language films	Retweets	Time	Thanks
0.8, 12 hrs, 48 hrs.	English-language films	Retweets	People	Lol	Thanks
0.2, 24 hrs, 48 hrs.	Bed, thermodynamics, people	Cause, heart, English-language films	Bid	Life	Time
0.4, 24 hrs, 48 hrs.	People	English-language films	Time	Life, retweets	Retweets
0.6, 24 hrs, 48 hrs.	People	English-language films	Retweets	Time	Thanks
0.8, 24 hrs, 48 hrs.	English-language films	Retweets	People	Lol	Thanks

TABLE 3. TOPIC CLUSTERS WITH RECOMMENDED LOCATIONS

Topic	*TFIDF Value*	*Topic Cluster*	*Location*
Aftermath of Brussels attack	0.2	RT @WPXI: Local prayer vigil held for victims of terrorist attacks in Brussels, Pakistan : s: t.co , Modi leads attack on Nuke terror at global summit, warns of state actors working with terrorists : , RT @USATODAY: Brussels Airport partially opens 12 days after terror attack : via @usatoday , Brussels Airport Partially Reopens 12 Days After Terror Attack: The Brussels airport is expected to restart fl...	38.88333333333333,-77.01666666666667
Spam tweets	0.4	RT @jedydynysem: Selfies you weren't meant to see Shhh!!, RT @jedydynysem: Selfies you weren't meant to see Shhh!!RT @jedydynysem: Selfies you weren't meant to see Shhh!!, RT @jedydynysem: Selfies you weren't meant to see Shhh!!, RT @jedydynysem: Selfies you weren't meant to see Shhh!!, RT @jedydynysem: Selfies you weren't meant to see Shhh!!	38.88333333333333,-77.01666666666667
None	0.6	None	None
None	0.8	None	None

The geographical analysis is a unique aspect of GeoContext, in that it can reveal topics that are specific and important to a geographical location. In this dataset of tweets taken from April 2016, there were tweets containing information about events occurring after the terrorist attack in Brussels, Belgium, in March 2016. We were specifically interested in whether GeoContext could discover these tweets as a topic. This type of information can reveal the opinions of people in different locations about a large worldwide event. Discovering these tweets as a topic can also show that GeoContext is able to consolidate tweets about a certain topic into one cluster. Grouping the tweets can assist anyone performing social media analysis about the event, from news agencies to individuals.

Table 3 displays the largest resulting topic cluster that has a recommended location for each adapted TF-IDF threshold value. This means that GeoContext considers the topic cluster to be centered at that geographical location. As shown, with an adapted TF-IDF value of 0.2, a topic cluster consisting of tweets about the aftermath of the Brussels attack was revealed. The recommended location for this topic cluster was Washington, D.C., which is not surprising as the event is related to national security and therefore the topic contains many tweets from news agencies and government programs located in Washington, D.C.

Also shown in Table 3, the adapted TF-IDF value of 0.4 was not able to reveal the Brussels attack topic. Rather, this value resulted in topics that contained spam tweets. We believe that the prevalence of spam topics with this value was due to the fact that spam tweets generally come from a similar location, and the spam accounts simply post retweets from each other. Due to the high volume of tweets being retweeted by the spam account, a larger topic cluster was created, and because the tweets all come from the same location, GeoContext considered the topic to be centered at that location.

Lastly, as displayed in Table 3, the threshold values of 0.6 and 0.8 do not result in any topic clusters being centered at any location. These threshold values are simply too high for any geographical locations to be discovered as meaningful to a topic cluster.

Overall, it is clear from this evaluation that the adapted TF-IDF threshold value that is able to produce topic clusters such as the Brussels attack aftermath that are geographically centered is 0.2. It is clear that a higher value adapted TF-IDF value requires almost all tweets within the topic cluster to be at one specific location, rather than more slightly spread out.

C. *Runtime Evaluation*

Because GeoContext is intended as a way to provide contextual information about temporal events, it runs in real time. As mentioned previously, the Gardenhose variety of the Twitter stream is estimated to provide about 15% of the public Twitter stream, which equates to approximately 18 tweets per second. In our evaluation, since the tweets were pre-geolocated and concepts and keywords were pre-extracted, we were able to determine the fastest possible time that GeoContext is able to calculate similarity scores and cluster tweets. The clustering process occurs at an average rate of 350 tweets per second. Because the clustering process occurs at a faster rate than the rate at which GeoContext can receive tweet objects from Twitter, it is clear that GeoContext is able to work in real-time and process tweets as they come in.

IV. RELATED WORK

In this section, we outline work related to the evaluation process of GeoContext described in this paper. We present the state of the art of existing research in the area of discovering geographical topics from a social media stream and the evaluation process for each approach.

Kim et al. [10] discovered trending topics on Twitter by normalizing high frequency words within tweets over time. They were able to discover words that had a dramatic increase in usage in a period of time, therefore discovering some terms that related to bursty topics. They also used a Louvain community detection algorithm [11] to analyze the geographical locations of the topics by U.S. state. The evaluation consisted of matching the discovered topics to current events and holidays during the time period of the evaluation. For example, the authors matched a sports topic to sporting events and games that were occuring, and the rise of terms like "joke" to April Fools Day.

Yin et al. [12] implemented Latent Geographical Topic Analysis, or LGTA, which extends Latent Dirichlet Allocation, or LDA [13], to discover topics that are grouped within a geographical region. The authors performed their evaluation of LGTA on a dataset of keywords from Flickr[5], a photo-sharing service. They analyzed the resulting groups of keywords and manually determined whether groups consisted of keywords mixed from different topics. They compared LGTA against a location-driven model, text-driven model, and a topic model that incorporates both text and spatial information.

Zhang et al. [14] produced six different topical and geographical social media clustering algorithms from combinations of LDA and DBSCAN. They utilized a dataset that consisted of different already-known topics and analyzed which of the six algorithms was able to discover each topic. They manually determined whether a resulting cluster contained mixed topics.

Vosecky et al. [15] created their Multi-Faceted Topic Model, which is a topic model incorporating multiple pieces of information present in tweets, including people and location data. They evaluated their model against LDA using the perplexity measure, which measures likelihood on a held-out test set. However, it has been shown that perplexity is not strongly correlated with human judgement on topic models [16].

Sakaki et al. [1] implemented an earthquake detection system that analyzes a stream of tweets by filtering the stream for keywords related to earthquake detection. Their approach uses a classifier to determine which tweets are relevant and which are irrelevant. They evaluate the system against real-world earthquake data and compare their system's estimation of the location of the earthquake and subsequent aftershocks to the real-world data.

Because evaluation metrics for topic modeling such as the perplexity statistic are not strongly correlated with human perception of topics, these types of metrics are not used as frequently for social media analysis systems whose intended purpose is for human readability. In much of the existing research, a more manual approach is used for evaluation. Like Kim et al. [10], we aimed to discover a major event with the adapted TF-IDF threshold evaluation. However, unlike existing approaches, our evaluation consists of comparing the various configurations of GeoContext. We propose that other social media analysis systems similar to GeoContext may contain the same types of threshold values, yet no evaluation has been conducted of these configuration values.

[5] https://www.flickr.com/

V. Future Work and Conclusion

In future work, we plan to perform more analysis on the spam tweet clusters that appear within the location recommendations. These types of topics are not typically relevant for users, yet there does not exist a completely effective detection algorithm for these types of tweets. We also plan to improve the geolocation module within GeoContext. Currrently, the geolocation module is very effective at recognizing locations such as cities or large venues that appear within Wikipedia, but it is not as effective as discovering smaller locations such as stores or restaurants. We plan to utilize the Google Maps API[6] in order to improve the location extraction.

In this paper, we described the evaluation of our system, GeoContext for performing topical and geographical analysis on a social media stream. We performed an evaluation of the four values that are customizable for various configurations of GeoContext: the similarity score threshold value, the time between pruning sessions, the time at which a topic is considered stale, and the adapted TF-IDF value.

In summary, our evaluation showed that the similarity score threshold value produces the largest effect on the resulting topic clusters, and a larger value produces more defined topics. The time between pruning sessions has no effect on the clusters, and a shorter time at which a topic is considered stale produces more concepts within some clusters. Using a smaller adapted TF-IDF threshold value, GeoContext was able to discover a topic cluster consisting of a major worldwide event, the aftermath of the Brussels attack.

References

[1] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo, "Tweet Analysis for Real-Time Event Detection and Earthquake Reporting System Development, " *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 4, pp. 919-931, 2013.

[2] Fang Jin, Edward Dougherty, Parang Saraf, Yang Cao, and Naren Ramakrishnan, "Epidemiological modeling of news and rumors on Twitter, " in *Proceedings of the 7th Workshop on Social Network Mining and Analysis*, Chicago, IL, 2013, pp. 8:1-8:9.

[3] Raffi Krikorian. (2013) Twitter Blog. [Online]. https://blog.twitter.com/2013/new-tweets-per-second-record-and-how

[4] Elizabeth Williams, Jeff Gray, and Brandon Dixon, "Improving Geolocation of Social Media Posts, " University of Alabama, SERG-042216, http://eawilliams2.students.cs.ua.edu/techreport2.pdf.

[5] Elizabeth Williams, Jeff Gray, and Brandon Dixon, "Mobile Context Recommendations from Social Media through Geotopical Clustering, " University of Alabama, SERG-042116, http://eawilliams2.students.cs.ua.edu/techreport.pdf.

[6] https://developers.google.com/maps/

[6] Chris Biemann and Martin Riedl, “Text: Now in 2D! A Framework for Lexical Expansion with Contextual Similarity, ” *Journal of Language Modelling*, vol. 1, no. 1, pp. 55-95, 2013.

[7] J. Eisenstein, B. O'Connor, N. A. Smith, and E. P. Xing, “A Latent Variable Model for Geographic Lexical Variation, ” in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, 2010, pp. 1277-1287.

[8] Martin Ester, Hans-Peter Kriegel, Jorg Sander, and Xiaowei Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise, ” in *Second International Conference on Knowledge Discovery and Data Mining*, Portland, OR, 1996, pp. 226-231.

[9] Karen Sparck Jones, “A Statistical Interpretation of Term Specificity and Its Application in Retrieval, ” *Journal of Documentation*, vol. 28, pp. 11-21, 1972.

[10] Hwi-Gang Kim, Seongjoo Lee, and Sunghyon Kyeong, “Discovering Hot Topics using Twitter Streaming Data, ” in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, Niagara, Ontario, 2013, pp. 1215-1220.

[11] Vincent Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre, “Fast Unfolding of Communities in Large Networks, ” *Journal of Statistical Mechanics*, vol. 2008, no. 10, p. P10008, 2008.

[12] Zhijun Yin et al., “Geographical Topic Discovery and Comparison, ” in *Proceedings of the 20th International Conference on World Wide Web (WWW)*, Hyderabad, India, 2011, pp. 247-256.

[13] David Blei, Andrew Ng, and Michael Jordan, “Latent Dirichlet allocation, ” *Journal of Machine Learning Research*, vol. 3, pp. 993-1022, 2003.

[14] Li Zhang, Xiaoping Sun, and Hai Zhuge, “Topic Discovery of Clusters from Documents with Geographical Location, ” *Concurrency and Computation: Practice and Experience*, vol. 27, no. 15, pp. 4015-4038, 2015.

[15] Jan Vosecky, Kenneth Wai-Ting Leung, and Wilfred Ng, “Dynamic Multi-Faceted Topic Discovery in Twitter, ” in *Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM)*, San Francisco, CA, 2013, pp. 879-884.

[16] Jonathan Chang, Jordan Boyd-Graber, Chong Wang, Sean Gerrish, and David Blei, “Reading Tea Leaves: How Humans Interpret Topic Models, ” in *Neural Information Processing Systems*, Vancouver, BC, 2009.