Discovering Location Information in Social Media

Fred Morstatter, Huiji Gao, Huan Liu
Arizona State University
{Fred.Morstatter, Huiji.Gao, Huan.Liu}@asu.edu

Abstract

Social media is immensely popular, with billions of users across various platform. The study of social media has allowed for deeper inquiries into questions posed by computer scientists, social scientists, and others. Social media posts tagged with location have provided means for researchers to perform even deeper analysis into their data. While location information allows for rich insight into social media data, very few posts are explicitly tagged with geographic information. In this work, we begin by introducing some state-of-the-art analysis techniques that can be performed using the location of a social media post. Next, we introduce some systems that help first responders provide relief with the help of the location of social media posts. Finally, we discuss how machine learning techniques can be applied to infer the location of a social media post, bringing this analysis to any message posted on social media.

1 Introduction

Social media sites provide ways for their users to conveniently share their lives in real-time, from their current mood to the music they are listening to, and even information pertaining to their physical activities. Among the myriad ways to share information, the ability for users to share their location has come to the forefront of many sites. Sites such as Twitter and Facebook allow users to tag their posts with their current location, either with the venue or "place", or the exact GPS coordinates. Sites such as Foursquare have built their entire platform around users sharing their geographic information.

Increasingly, researchers and practitioners have found ways to make use of this new source of information for novel applications, such as recommending new venues to users, and predicting the number of people who will check in at a certain location. It has also been used to increase the effectiveness of existing problems such as helping deliver the right information to first responders in humanitarian assistance and disaster recovery.

In this work we discuss state-of-the-art challenges in leveraging geographic information in social media research. We begin by discussing how researchers use this information to make recommendations and to predict the next location a user will visit. Next, we discuss systems that have been created to help make sense of users mobility patterns in online social networks and to use these patterns to understand the greater picture of an event of disaster as it unfolds on social media. Finally, we introduce techniques that can predict a user's location in absence of explicit information on their post. These techniques have the potential to bring this analysis to the entirety of social media posts, and not just those explicitly tagged with geographic information.

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

2 Using Location Data for Novel Applications

2.1 Personalized Point of Interest Recommendation on Location-Based Social Networks

The rapid growth of cities has developed an increasing number of points of interest (POIs), e.g., restaurants, theaters, stores, and hotels, providing us with more choices of life experience than before. People are willing to explore the city and neighborhood in their daily life and decide "where to go" according to their personal interest and the various choices of POIs. At the same time, making a satisfying decision efficiently among the large number of POI choices becomes a touch problem for a user. To facilitate a user's exploration and decision making, POI recommendation has been introduced by location-based services such as Yelp and Foursquare. However, such recommendation models are commonly based on majority users' preference on POIs, which ignore a user's personal preference. Comparing to visiting places that best fit a user's interest, visiting places against a user's taste may give him very terrible experience, especially in a situation when the user travels to a new place. Therefore, personalized POI recommendation is proposed to help users filter out uninteresting venues according to their own taste and save their time in decision making.

2.1.1 Background

Before the Web 2.0 era, analyzing user's mobility for personalized POI recommendation is based on cellphone-based GPS data. Due to the lack of mapping information between geographical coordinates and real-world POIs on GPS data, a POI is usually determined by the stay points (geographical points at which a user spent sufficient long time) extracted from hundreds of users' GPS trajectory (a sequence of time-stamped latitude/longtitude pairs collected repeatedly at intervals of a short period) logs [25, 26]. With the rapid development of location-based social networking services, e.g., Foursquare, Yelp, and Facebook Places[1], users are able to check in at real-world locations with specific POI information and share such check-ins with their friends through mobile devices, resulting in more abundant information to improve personalized location recommendation.

This abundance of information has led to a new class of social network, called a "location-based social network". Location-based social networks not only refer to the social connections among users, but also consist of the "location-based" context including geographical check-in POIs, check-in time stamps, and check-in related content (e.g., tips, comments, POI descriptions, etc.), as shown in Figure 1. Compared with other online social networks that consist of user activities interacting with the virtual world, LBSNs reflect a user's geographical action in the real world, residing where the online world and real world intersect, therefore bridging the gap between the real world and the virtual world, providing both opportunities and challenges for researchers to investigate users' check-in behavior for personalized POI recommendation in spatial ("where"), temporal ("when"), social ("who") and content ("what") aspects.

In this work, we use POI, venue, and location as interchangeable terms. Let $\mathbf{u} = \{u_1, u_2, ..., u_m\}$ be the set of users and $\mathbf{l} = \{l_1, l_2, ..., l_n\}$ be the set of POIs where m and n are the numbers of users and POIs, respectively. The problem of personalized POI recommendation on LBSNs is defined as:

Given a user $u \in \boldsymbol{u}$, a set of POIs (locations) $\boldsymbol{l}_u \in \boldsymbol{l}$ that u has checked-in, recommend him some POIs for his future visits based on the LBSN context (e.g., social connections, content information of check-ins, time stamps of check-ins) related to him.

In the last decade, recommender systems have been widely studied among various categories, e.g., movie recommendation on Netflix, dating recommendation on Zoosk, item recommendation on Amazon. However, it is not sufficient to directly apply these technologies as personalized POI recommendation on LBSNs presents unique challenges due to the heterogeneous information layout and the specificity of human mobility. Designing efficient POI recommendation approaches on LBSNs inevitably needs to consider the following properties.

[1] http://www.facebook.com/about/location

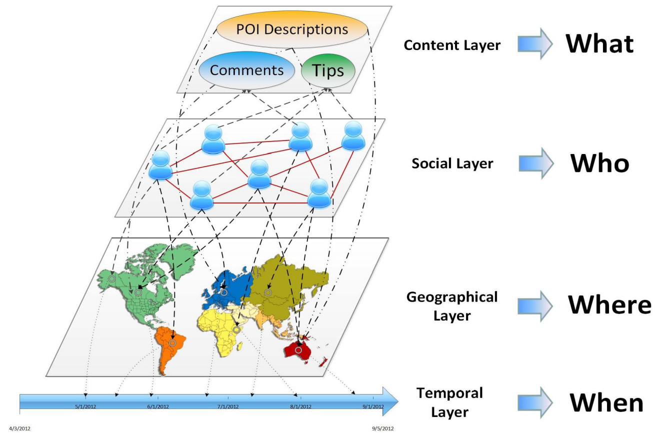


Figure 1: The information layout of location-based social networks. The geographical layer contains the historical check-ins of users, while the social layer contains social friendship information, and the content layer consists of user feedbacks or tips about different places. All these three layers share one timeline, indicating the temporal information of the user "check-in" behavior.

Geographical Property of Social Connections

Geographical property and social connections are coherent and affect each other in human behavior. For example, a user is more likely to be friends with other users who are geographically close to him, e.g, co-workers, colleagues. Likewise, a user may check-in at a location due to the influence from his friends, such as following friends' suggestions to visit a restaurant. Such coherence results in a new property, commonly referred to as socio-spatial properties [21]. Thus, considering the social information together with the geographical property enables us to capture the user preferences more precisely in personalized POI recommendation on LBSNs.

Temporal Patterns of Geographical Check-ins

Human geographical movement exhibits strong temporal patterns [3, 15, 24] and is highly relevant to the location property. For example, a user regularly goes to a restaurant for lunch around 12:00 pm, watches movie on Friday night, and shops during weekends. This is generally referred to as temporal cyclic patterns. Such temporal patterns are not widely observed in other recommender systems. For instance, it is not common to observe a user regularly watching a specific movie (e.g., Batman, Avatar) or purchasing a specific item (e.g., camera, cellphone) at specific hour of the day, or day of the week. (Although birthdays or holidays like Thanksgiving may affect human behavior a bit, they are not commonly considered).

Semantic Indications of Check-in Content

Content information on LBSNs could be related to a user's check-ins, providing a unique opportunity for location recommendation from a conceptual perspective. For example, By observing a user's comment on a Mexican restaurant discussing its spicy food, we observe if the user is interested in spicy food or not. This is an example of *user interests*. By observing a location's description as "vegetarian restaurant", we may infer that the restaurant serves "vegetarian food" and users who check-in at this location might be interested in the vegetarian diet. This is an example of *location properties*. These two types of information are representatives of user-generated content and location-associated content on LBSNs. The former refers to comments that left by users towards specific locations when they check-in; the latter can be descriptive tags associated with specific locations.

The above three properties indicate the three unique relationships between geographical information and social, temporal, and content information, commonly referred to as *geo-social correlations*, *geo-temporal patterns*, and *geo-content indications*. For more information, please refer to Gao and Liu 2015 [9].

2.2 Geolocated Information for Crisis Response Applications

In this section we discuss how geolocated social media data can be used to help with disaster relief. We introduce two classes of systems: crisis maps that help first responders match need with resources, and tools which help first responders get a deeper understanding of the situation.

2.2.1 Crisis Mapping

Crisis mapping consists of tools that help first responders to coordinate resources in times of disaster. Usually, the requests for assistance are obtained through SMS, as well as through social media. Twitter is often used in such applications [11].

Ushahidi [8] is one of the first crisis mapping systems. It has helped to coordinate relief in Kenya [16], Afghanistan, and Haiti. The system features a request engine that allows for those affected to seek out the resources they need for their specific situation. Volunteers and disaster relief organizations can then use this map to allocate aid and to see where their services can be of the most use.

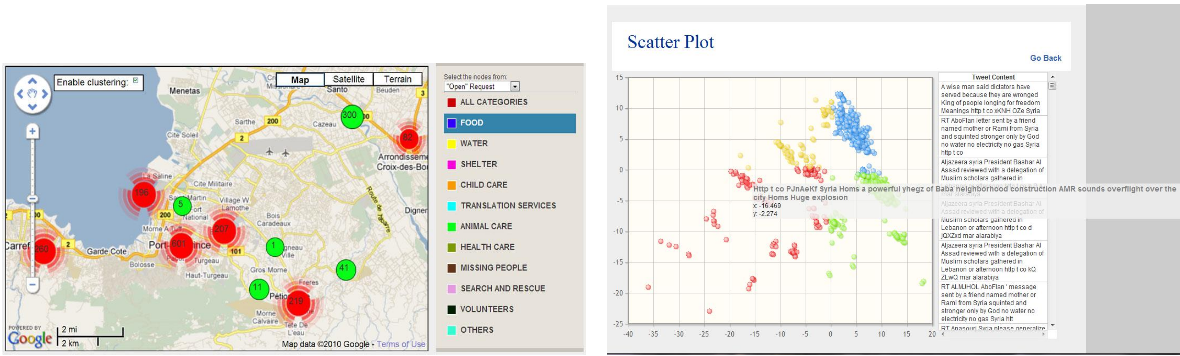
TweetTracker [10] enables a first responder to collect Twitter data pertaining to a crisis by specifying parameters about the crisis. These parameters can come in one of three forms: 1) keywords which describe words that pertain to the crisis, 2) geographical bounding boxes which specify the region or regions affected by the crisis, and 3) user names which can be users that tweet about the crisis. TweetTracker collects tweets that match any of these parameters and shows them to the first responder.

ASU Coordination Tracker (ACT) [7], is designed to collect crowdsourced requests, keep first responders aware of the current situation, and help them coordinate for disaster relief. The main goal of ACT is to analyze crowdsourced requests and promote inter-agency coordination to prevent duplication of effort during crisis. It comprises of five functional modules: request collection, request analysis and visualization, response, coordination, and situational awareness. Figure 2(a) shows an overview of classified requests and the quantity of requests on ACT's crisis map. ACT collects two types of requests: requests from crowds (crowdsourcing) and requests from groups (groupsourcing). Crowdsourcing refers to requests submitted by people (e.g. victims, volunteers) who are not from certified organizations. The groupsourcing [1,6] requests originate from responding organizations such as United Nation, Red Cross, etc. Specially, crowdsourcing data are collected in forms of web, SMS and tweets collected through TweetTracker. The data analysis takes advantage of both data mining technology and expert knowledge to iteratively capture the essential content of raw requests and classify them into several categories (food, shelter, missing persons, etc.). Figure 2(b) shows a clustering visualization for expert labeling and decision making based on active learning techniques.

2.2.2 Visualizing Crisis Data

Another way that geolocated social media data has helped first responders is by giving them a picture of what is unfolding on the ground. While looking at the raw data is very difficult, several systems have emerged in the past few years to make sense of the massive volume of data that is generated during a crisis.

TweetXplorer [17] is a system that is designed to help first responders get situational awareness. An overview of the system is shown in Figure 3(a). Queries are issued to the system using the query pane on the bottom right. The network in the top right shows the most retweeted users in the dataset. The top left shows a map, which shows the locations that gave the most geotagged tweets about the user's query. This map can also be used to help the user's explore the data. The map can be combined with the network to show the geo-tagged retweets of a particular user, as shown in Figure 3(b). This can help the analysts understand which locations care the most



(a) ACT Request Window

(b) Request Classification and Visualization

Figure 2: ASU Coordination Tracker overview. This shows two important views of the system. On the left we see a request window which shows requests made by specific organizations classified in terms of quantity. On the right we see the active learning module which helps the human expert to classify requests.

about a particular tweet. The map can also be "brushed", causing it to show a tag cloud of the most important words in the selected region.

Another visual analytic system designed to help first responders understand events on the ground is SensePlace2 [13]. The system helps analysts find important tweets by allowing them to query in two ways: through keywords, and through a spatial filtering interface. In this way the users can find both the content they are interested in and where it originates. More details about the system can be found on the project web page[2].

3 Inferring Location Information in Social Media

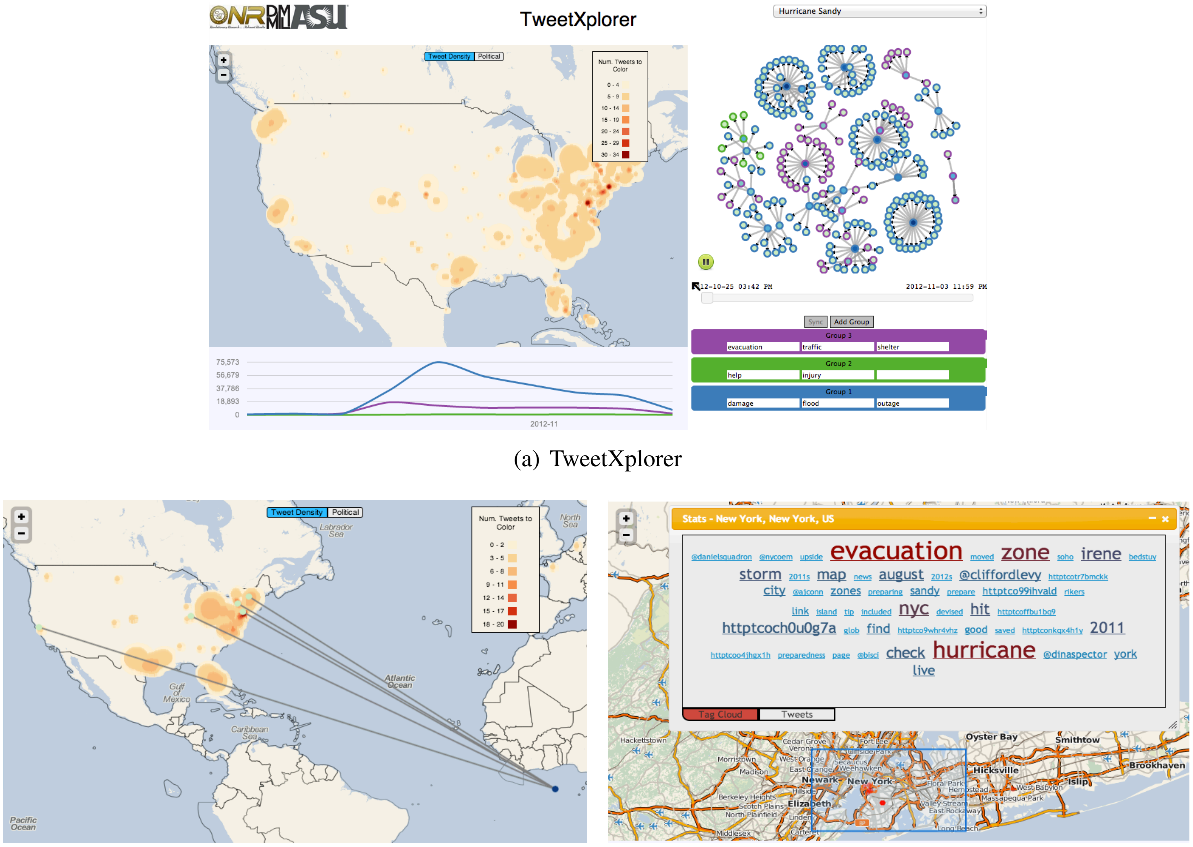
While location-based user analysis has taken off in recent years, the number of users providing their location has not kept pace. Only about 1% of all of the tweets posted on Twitter are geotagged [19]. This is largely due to Twitter's "opt-in" policy for providing user location. The disparity prevents existing techniques being used to assist the vast majority of users of a service who do not use their geographical information.

To bridge the gap analysis and users who lack location information, researchers have focused on uncovering the locations of users who do not share their location on social media. This location can be uncovered from three perspectives: the user's *profile location*, where he lives; the *tweet's location*, where the message was published; and the *event location*, where the message is talking about. In this section we discuss attempts that researchers have made to address these problems.

3.1 Inferring a User's Location

A "user's location" refers to the location where the user lives, or the location that he would give in his profile. The problem of user location prediction of Twitter users was first investigated in [4], where the authors used the language of the user's tweets to estimate his home location. The authors manually asses the statistical distribution of words in the tweets to find words that contained a "strong geo-scope". One example of such a word is "Red Sox", which occurs much more in the Boston, Massachusetts area than anywhere else in the USA. This work was extended in [2], where the authors propose an automated approach to finding words with a strong geo-scope. Linguists have also made an attempt at this problem looking for statistical variation in different geographic regions. In [5], the authors propose geographical topic models to model the language across the entire continental United States. [23] and [20] use a grid-based approach to define regions using the data.

[2] http://www.geovista.psu.edu/SensePlace2/



(a) TweetXplorer

(b) Social Network Overlaid on Map

(c) Geo-Aware Tag Cloud

Figure 3: Overview of TweetXplorer. The top pane shows an overview of the system. The bottom left shows how the network can be combined with the map to analyze who is retweeting the user. The bottom right shows a tag cloud that shows the most important words from a particular geographic area.

Researchers have approached this problem not just from one perspective (e.g. text, or network) but instead take a holistic approach trying to incorporate as much signal as possible into their models. In [22], the authors combine signals such as the tweeter's "location field" in his profile, any URLs in his biography (top-level domains such as .uk may provide a country indicator), and the time zone the user has set. This heuristic-based approach is furthered by [14], who adds more information including patterns of users posting time, and points of interest. This work also studies how different types of spatial aggregation and ensemble approaches can lead to better classification results.

In the context of disaster response, user location has been used to help first responders to find "eyewitness accounts": accounts that are both geographically near the disaster and discussing the topics of those affected. Kumar et. al 2013 [12] proposed a method for finding eyewitness accounts by measuring all of the users who tweet about a crisis along two dimensions: the location of their geotagged tweets, and their affinity for a set of topics. An example of these dimensions is shown in Figure 4. The users who score above-average on both dimensions, putting them in the upper-right quadrant, are considered "Q1" users. The authors find that the properties of Q1 users reflect the properties of eyewitness users: they tweet first on pressing topics and often relay relevant information that is not found in the other dimensions.

One important requirement of all of the above work is the number of tweets required in order for the approaches to make accurate predictions. The geo-scope approaches described in [2,4] requires at least 700 tweets for an accurate prediction, while the linguistic [5, 20, 23] and heuristic-based [14, 22] methods require approxi-

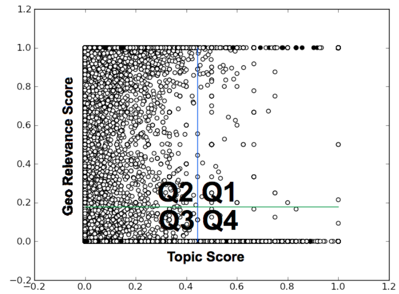
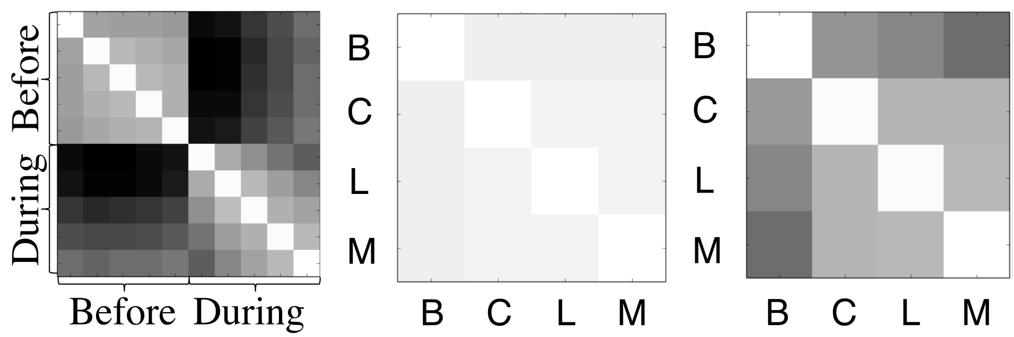


Figure 4: Quadrants used to find eyewitness users. Here, the authors focus on users with an above-average Geo Relevance Score (vertical axis), and an above-average Topic Score (horizontal axis).



(a) Temporal language differences. (b) Geographic language differences: tranquil time. (c) Geographic language differences: crisis time.

Figure 5: Temporal and geographic differences of language (calculated using Jensen-Shannon divergence); darker shades represent greater difference. To illustrate geographic differences, we compare Boston (B) with three other major U.S. cities: Chicago (C), Los Angeles (L), and Miami (M).

mately 200 tweets.

3.2 Inferring a Tweet's Location

While current approaches to user location prediction have shown promising results, one limitation is that they need a substantial history of a user's tweets in order to make accurate predictions. This much data is often unavailable for the vast majority of users on Twitter. Moreover, even for users who have posted this much information, it can be very difficult to collect this history under duress. Here we discuss alternatives to this problem, that allow users to geotag a single tweet. Often this is necessary in times of crisis when it is not feasible to collect a user's entire history to estimate his location.

Disaster response agencies often look to Twitter to understand what is unfolding on the ground in real time. To get a sense of the area most effected by the disaster, these agencies look at geo-tagged tweets. Since geo-tagged tweets only account for 1% of all activity on Twitter, these first responders are left looking for other methods to find a tweet's location. However, with the requirement of hundreds of past tweets for a particular user, existing methods to finding a user's location become infeasible during crisis situations.

In the absence of explicit geographic information, it is unlikely that a single tweet contains enough information to locate its exact position. Instead, to accommodate the lack of information, in [18] we change the problem to reflect what first responders are actually looking for during times of crisis: whether or not a tweet actually originates from within the crisis region. By simplifying the problem from predicting two continuous values to predicting one boolean value, we make the problem more tractable with such sparse data.

To differentiate the users within a crisis region from those outside by the text of their tweet, we must first verify that the text that is generated from within a crisis region is actually different from the text outside of it. We perform this analysis along two dimensions: within the area of the crisis before and during the crisis, and during the time of the crisis across different locations. The results of this analysis are shown in Figure 5. Figure 5(a) shows the temporal difference by hour over the course of April 15, 2013, the day of the Boston Marathon Bombing. We see that the hours leading up to the bombing are much more similar than the hours after the bombing. Furthermore, in the location comparisons, we see that the cities are similar before the disaster (Figure 5(b)), and exhibit different behavior after the beginning of the crisis (Figure 5(c)). Thus, a linguistic difference exists between the linguistic patterns during the crisis within the crisis location.

Now that we have established that a difference exists between the locations during the crisis, we can continue to build a machine learning model that can capture these differences and aid first responders in finding tweets coming from within the crisis region. To do this, we hypothesize some linguistic features within the tweet that may be useful in identifying whether it originates from within the crisis region: *Word Unigrams and Bigrams*, *Part-of-Speech Tags*, *Shallow Parsing*, and *Crisis-Sensitive Features*. Crisis-sensitive features are some features identified by inspecting the text produced in the tweet. These consist of some part-of-speech patterns that are commonly observed in crises.

To test the effectiveness of our linguistic features, we build basic classifiers to test our features. The model then outputs its prediction of whether the tweet is inside region or outside region. We compare all possible combinations of individual feature classes and find that a combination of Unigram + Bigram + Crisis Sensitive features perform best for both crises.

We see that in both crises all of the top performing feature combinations still contain both the Bigram and Unigram feature classes. These classifiers massively outperform traditional approaches in the geolocation problem. This shows that inferring the tweet's location is possible, and that by modifying the problem to focus on the binary question of "within location" and "outside location" we are able achieve superior performance on this problem.

4 Conclusion

Social media is immensely popular, allowing users to share their lives in new ways. By allowing users to share their location, social media sites have enabled their users with richer means to express themselves. Location information is an important part of social media analysis, allowing researchers to obtain new insights into the behavior of users online and practitioners to develop new applications. In this work, we have shown how location can be leveraged to find users' interests and to predict what location a user will visit next. Furthermore, location can be leveraged to help those affected by disaster, both by helping them to find the right information and by making sure their requests for help are sourced to the correct agencies.

One of the main difficulties with studying location in social media is the lack of explicit information. This comes, in part, from the low number of users who share their information on social media sites. We have presented work which seeks to address this problem. We have also presented an algorithm that helps find users in the region affected by a crisis. By focusing only on whether the user is inside or outside the region, we are able to achieve higher performance than traditional approaches.

The problem of discovering location information in social media is a challenging one with a long way to go. Future work consists of finding the "event location", the location that the user is talking about. This can differ from both the user's tweet location and his user location. Another area for future work is location privacy. In discovering location, we may uncover the location of users who do not want to be discovered, such as users participating in protests. While existing approaches illuminate the potential for privacy concerns, future work will be to address them in a way that does not bring users into harms way. Additionally, data reliability is a problem within the context of location discovery. Users providing fake and incorrect values for their location add noise to the data. Future work seeks to identify these fake and incorrect locations and remove them to increase the performance of the location discovery task.

Acknowledgments

This work is sponsored in part by Office of Naval Research (ONR) grant N000141410095 and by the Department of Defense under the Minerva Initiative through the ONR grant N000141310835.

References

[1] Geoffrey Barbier, Reza Zafarani, Huiji Gao, Gabriel Fung, and Huan Liu. Maximizing benefits from crowdsourced data. *Computational and Mathematical Organization Theory*, 18(3):257–279, 2012.

[2] Hau-wen Chang, Dongwon Lee, Mohammed Eltaher, and Jeongkyu Lee. @Phillies Tweeting from Philly? Predicting Twitter User Locations with Spatial Word Usage. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, ASONAM '12, pages 111–118, Washington, DC, USA, 2012. IEEE Computer Society.

[3] Z. Cheng, J. Caverlee, K. Lee, and D.Z. Sui. Exploring millions of footprints in location sharing services. In *Proceedings of the Fifth International Conference on Weblogs and Social Media*, 2011.

[4] Zhiyuan Cheng, James Caverlee, and Kyumin Lee. You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 759–768. ACM, 2010.

[5] Jacob Eisenstein, Brendan O'Connor, Noah A Smith, and Eric P Xing. A Latent Variable Model for Geographic Lexical Variation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 1277–1287. Association for Computational Linguistics, 2010.

[6] H. Gao, G. Barbier, and R. Goolsby. Harnessing the crowdsourcing power of social media for disaster relief. *Intelligent Systems, IEEE*, 26(3):10–14, 2011.

[7] H. Gao, X. Wang, G. Barbier, and H. Liu. Promoting coordination for disaster relief–from crowdsourcing to coordination. *Social Computing, Behavioral-Cultural Modeling and Prediction*, pages 197–204, 2011.

[8] Huiji Gao, Geoffrey Barbier, and Rebecca Goolsby. Harnessing the crowdsourcing power of social media for disaster relief, 2011.

[9] Huiji Gao and Huan Liu. *Mining Human Mobility in Location-Based Social Networks*. Data Mining and Knowledge Discovery. Morgan & Claypool, 2015.

[10] Shamanth Kumar, Geoffrey Barbier, Mohammad Ali Abbasi, and Huan Liu. TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief. *ICWSM*, 2011.

[11] Shamanth Kumar, Fred Morstatter, and Huan Liu. *Twitter data analytics*. Springer, 2014.

[12] Shamanth Kumar, Fred Morstatter, Reza Zafarani, and Huan Liu. Whom Should I Follow?: Identifying Relevant Users During Crises. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, HT '13, pages 139–147, New York, NY, USA, 2013. ACM.

[13] Alan M. MacEachren, Anuj Jaiswal, Anthony C. Robinson, Scott Pezanowski, Alexander Savelyev, Prasenjit Mitra, Xiao Zhang, and Justine Blanford. Senseplace2: Geotwitter analytics support for situational awareness. In *Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on*, pages 181–190. IEEE, 2011.

[14] Jalal Mahmud, Jeffrey Nichols, and Clemens Drews. Home Location Identification of Twitter Users. *Transactions on Intelligent Systems and Technology*, 5(3), 2014.

[15] E. Malmi, T.M.T. Do, and D. Gatica-Perez. Checking in or checked in: Comparing large-scale manual and automatic location disclosure patterns. *The 11th International Conference on Mobile and Ubiquitous Multimedia (MUM 2012)*, 2012.

[16] Patrick Meier and Kate Brodock. Crisis mapping kenyas election violence: Comparing mainstream news, citizen journalism and ushahidi. *iRevolution Blog, October*, 23, 2008.

[17] Fred Morstatter, Shamanth Kumar, Huan Liu, and Ross Maciejewski. Understanding Twitter Data with TweetXplorer. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pages 1482–1485, New York, NY, USA, 2013. ACM.

[18] Fred Morstatter, Nichola Lubold, Heather Pon-Barry, J"urgen Pfeffer, and Huan Liu. Finding eyewitness tweets during crises. *ACL Workshop on Language Technologies and Computational Social Science*, 2014.

[19] Fred Morstatter, Jürgen Pfeffer, Huan Liu, and Kathleen M. Carley. Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. *Proceedings of ICWSM*, 2013.

[20] Stephen Roller, Michael Speriosu, Sarat Rallapalli, Benjamin Wing, and Jason Baldridge. Supervised Text-Based Geolocation using Language Models on an Adaptive Grid. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 1500–1510. Association for Computational Linguistics, 2012.

[21] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo. Socio-spatial properties of online location-based social networks. *Proc. 5th International AAAI Conference on Weblogs and Social Media*, 11, 2011.

[22] Axel Schulz, Aristotelis Hadjakos, Heiko Paulheim, Johannes Nachtwey, and Max Mühlhäuser. A multi-indicator approach for geolocalization of tweets. In *Proceedings of ICWSM*, 2013.

[23] Benjamin Wing and Jason Baldridge. Simple Supervised Document Geolocation with Geodesic Grids. In *ACL*, volume 11, pages 955–964, 2011.

[24] M. Ye, K. Janowicz, C. Mülligann, and W.C. Lee. What you are is when you are: the temporal dimension of feature types in location-based social networks. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 102–111. ACM, 2011.

[25] Vincent W Zheng, Yu Zheng, Xing Xie, and Qiang Yang. Collaborative location and activity recommendations with gps history data. In *Proceedings of the 19th international conference on World wide web*, pages 1029–1038. ACM, 2010.

[26] Y. Zheng, L. Zhang, X. Xie, and W.Y. Ma. Mining interesting locations and travel sequences from gps trajectories. In *WWW*, pages 791–800. ACM, 2009.