International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

A SURVEY OF LOCATION PREDICTION ON TWITTER

[1]P.Bharath,[2]ChennamsettyBhargavi Lakshmi,[3] Kolisetty Jyothi,Venkatesh, [4]AmaraDivyabhargavi, [5]Cherukupalli Prameela Jyothika
[1,2,3,4,,5]Assistant professors, Department of CSE in Narasaraopet Institute Of Technology

ABSTRACT

Locations, e.g., countries, states, cities, and point-of-interests, are central to news, emergency events, and people's daily lives. Automatic identification of locations associated with or mentioned in documents has been explored for decades. As one of the most popular online social network platforms, Twitter has attracted a large number of users who send millions of tweets on daily basis. Due to the world-wide coverage of its users and real-time freshness of tweets, location prediction on Twitter has gained significant attention in recent years. Research efforts are spent on dealing with new challenges and opportunities brought by the noisy, short, and context-rich nature of tweets. In this survey, we aim at offering an overall picture of location prediction on Twitter. Specifically, we concentrate on the prediction of user home locations, tweet locations, and mentioned locations. We first define the three tasks and review the evaluation metrics. By summarizing Twitter network, tweet content, and tweet context as potential inputs, we then structurally highlight how the problems depend on these inputs. Each dependency is illustrated by a comprehensive review of the corresponding strategies adopted in state-of-the-art approaches. In addition, we also briefly review two related problems, i.e., semantic location prediction and point-of-interest recommendation. Finally, we make a conclusion of the survey and list some future research directions.

1. INTRODUCTION:

He last decade has witnessed an unprecedented proliferation of online social networks. Those include general-purpose platforms like Twitter and Facebook, location-based ones like Foursquare and Gowalla, photo sharing sites like Flicker and Interest, as well as other domain-specific platforms such as Yelp and LinkedIn. On these platforms, users may establish online friendship with others sharing similar interests. Users may also share with online friends their daily lives in forms of texts, photos, videos, or check-ins. Among all online social networks, Twitter is characterized by its unique way of following friends and sending posts. On the one hand, Twitter friendships are not necessarily mutual. For example, users may "follow" celebrities without requiring them to follow back. On the other hand, textual posts on Twitter, a.k.a. tweets or micro blogs, are limited to 140 characters. Users are encouraged to post frequently but casually about anything, such as moods, activities, opinions, local news, etc. Users, online friendships, and tweets make twitter a virtual online world. This virtual world intersects with the real world, where locations acting as intermediate connections. Twitter users have long- term residential addresses. Their home locations cause them to notice, get interested, and tweet news or events around their daily activity regions. With increasing popularity of GPS-enabled devices such as smart phones and tablets, users may casually attach real-time locations when sending out tweets.

In this survey, we concentrate on the above three types of Twitter-related locations, namely user home location, tweet location, and mentioned location. Knowing physical

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

locations involved in Twitter helps us to understand what is happening in real life, to bridge the online and offline worlds, and to develop applications to support real-life demands, among many applications. For example, we can monitor public health of residents recommend local events or attractive places to tourists, summarize regional topics and identify locations of emergency or even disasters. Although Twitter users may casually reveal locations either manually or with the help of GPS, location information on Twitter are far from complete and accurate. Cheng et al. find that only 21% of users in a U.S. Twitter dataset provide residential cities in their profiles, while 5% give coordinates of their home addresses. Despite the low availability, Hecht et al. Report that self-declared home information in many user profiles are inaccurate or even invalid. Hecht et al. and Ryoo et al. Observe that only 0.77% and 0.4% of tweets have location information attached in their datasets, respectively. Similar percentages are also reported by Bartosz et al. And Priedhorsky et al. Therefore, completing Twitter-related locations acts as the prerequisite for many other studies and applications, and is worth careful investigation.

2. Literature Survey

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy n company strength. Once these things r satisfied, ten next steps are to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

1.)Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: a content-based approach to geo-locating twitter users," in Proc. 19th ACM Conf. on Information and Knowledge Management, 2010, pp. 759–768.

We propose and evaluate a probabilistic framework for estimating a Twitter user's city-level location based purely on the content of the user's tweets, even in the absence of any other geospatial cues. By augmenting the massive human-powered sensing capabilities of Twitter and related micro blogging services with content-derived location information, this framework can overcome the scarcity of geo enabled features in these services and enable new location based personalized information services, the targeting of regional advertisements, and so on. Three of the key features of the proposed approach are: (i) its reliance purely on tweet content, meaning no need for user IP information, private login information, or external knowledge bases; (ii) a classification component for automatically identifying words in tweets with a strong local geo-scope; and (iii) a lattice-based neighborhood smoothing model for refining a user's location estimate. The system estimates k possible locations for each user in descending order of confidence. On average we find that the location estimates converge quickly (needing just 100s of tweets), placing 51% of Twitter users within 100 miles of their actual location.

2.)Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "Who, where, when and what: discover spatio-temporal topics for twitter users," in Proc. 19th ACM Int. Conf. on

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

Knowledge Discovery and Data Mining, 2013, pp. 605–613.

Micro-blogging services, such as Twitter, and location-based social network applications have generated short text messages associated with geographic information, posting time, and user ids. The availability of such data received from users offers a good opportunity to study the user's spatial-temporal behavior and preference. In this paper, we propose a probabilistic model W4 (short for Who+Where+When+What) to exploit such data to discover individual users' mobility behaviors from spatial, temporal and activity aspects. To the best of our knowledge, our work offers the first solution to jointly model individual user's mobility behavior from the three aspects. Our model has a variety of applications, such as user profiling and location prediction; it can be employed to answer questions such as "Can we infer the location of a user given a tweet posted by the user and the posting time?" Experimental results on two real-world datasets show that the proposed model is effective in discovering users' spatial-temporal topics, and outperforms state-of-the-art baselines significantly for the task of location prediction for tweets.

3. Systemanalysis: `

3.1 Existingsystem:

The characteristics of Twitter pose emerging challenges for these existing research problems in new problem settings. On the one hand, users often write tweets in a very casual manner. Acronyms, misspellings, and special tokens make tweets noisy, and techniques developed for formal documents are error-prone on tweets.

Disadvantages:
• Most cost effective
• Performance is less

3.2 Proposed system:

The prediction models proposed based on Twitter can also be adapted to other social media sites, while might require some changes. But before considering model adaptations, we need to be clear on whether the three geo location problems on Twitter, i.e., prediction of home location, tweet location and mentioned location, are applicable to the target platform or not. For example, tweet and mentioned location prediction on some image and video sharing platforms like Instagram and Pinterest may not be applicable.

Advantages:
• Less cost effective
• Performance is high

4. SYSTEM DESIGN

4.1 UML DIAGRAMS:

UML represents Unified Modeling Language. UML is an institutionalized universally useful showing dialect in the subject of article situated programming designing. The fashionable is overseen, and become made by way of, the Object Management Group.

The goal is for UML to become a regular dialect for making fashions of item arranged PC programming. In its gift frame UML is contained two noteworthy components: a Meta-show and documentation. Later on, a few type of method or system can also likewise be brought to; or related with, UML.
The Unified Modeling Language is a popular dialect for indicating, Visualization, Constructing and archiving the curios of programming framework, and for business demonstrating and different non-programming frameworks.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

The UML speaks to an accumulation of first-rate building practices which have verified fruitful in the showing of full-size and complicated frameworks.

The UML is a essential piece of creating gadgets located programming and the product development method. The UML makes use of commonly graphical documentations to specific the plan of programming ventures.

GOALS:

The Primary goals inside the plan of the UML are as in step with the subsequent:

1. Provide clients a prepared to-utilize, expressive visual showing Language on the way to create and change massive models.
2. Provide extendibility and specialization units to make bigger the middle ideas.
3. Be free of specific programming dialects and advancement manner.
4. Provide a proper cause for understanding the displaying dialect.
5. Encourage the improvement of OO gadgets exhibit.
6. Support large amount advancement thoughts, for example, joint efforts, systems, examples and components.
7. Integrate widespread procedures.

4,2 USE CASE DIAGRAM:

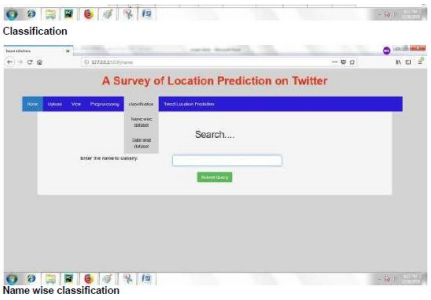
A use case diagram in the Unified Modeling Language (UML) is a type of behavioral diagram defined by and created from a Use-case analysis. Its purpose is to present a graphical overview of the functionality provided by a system in terms of actors, their goals (represented as use cases), and any dependencies between those use cases. The main purpose of a use case diagram is to show what system functions are performed for which actor. Roles of the actors in the system can be depicted.

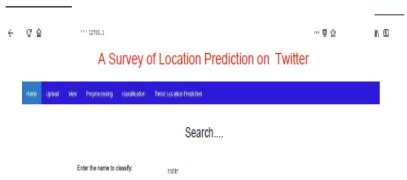
5. Test cases and test resultsTest cases

S.No	TEST CASE	EXEPTED RESULTS	TEST RESULTS
1	Upload the data set	Upload successful	Successful
2	Doing the pre-processing	Pre-processing successful	Successful
3	Convert the consolidate data set	Convert consolidate data set viewed successful	Successful
4	View the Top spam and Ham words.	Viewed top spam and ham words successfully	Successful
5	Calculate the accuracy score in different algorithms	Successfully evaluate the accuracy score in different algorithms	Successful
6	Predict the spam or not	Successfully predict the spam and ham words	Successful
7	Generate the graph	Successfully generate the graph values	Successful

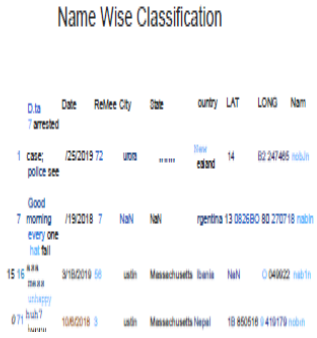
TEST RESULT: All the test cases mentioned above passed successfully. Nodefects encountered.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362





LIST OF TABLES



6. Conclusion:

We review and summarize techniques of three geolocation problems on Twitter: home location, tweet location, and mentioned location. Compared with similar problems on formal documents, i.e., document geolocationand named entity recognition & disambiguation, geolocation problems on Twitter face unique challenges and opportunities. The challenges generally arise from the noisy and short nature of tweet content. The opportunities, on the other hand, are enabled by the massive Twitter network and rich tweet context. All the three prediction problems rely heavily on tweet content.As a significant feature of the platform, Twitter network plays a key role in home location prediction. Various hypotheses have been made on the connections between friendship and home proximity. Inspired by Backstrom et al. Many works try to formulate the relationship between the probability of friendship and home location distance. However, the indication is not very strong on Twitter. To fix this issue, social- closeness-based methods are proposed to differentiate noisy friendship. Explicit factors like friends with interactions are employed as useful information to predict home proximity. Implicit factors like influence scope are captured by sophisticated models. Finally, we note that Twitter network causes the predictions for different users to depend on each other. Therefore, it is necessary to involve global inference approaches

REFERENCE

[1] Z. Cheng, J. Caverlee, and K. Lee, "You are where you tweet: a content-based approach to geo-locating twitter users," in Proc. 19th ACM Conf. on Information and Knowledge Management, 2010, pp. 759–768.

[2] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "Who, where, when and what: discover spatio-temporal topics for twitter users," in Proc. 19th ACM Int. Conf. on Knowledge Discovery and Data Mining, 2013, pp. 605–613.

[3] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo, "Mining user mobility features for next place prediction in location-based services," in Proc. 12th IEEE Int. Conf. on Data Mining, 2012, pp. 1038–1043.

[4] V. Rakesh, C. K. Reddy, D. Singh, and M. Ramachandran, "Location-specific tweet detection and topic summarization in twitter," in Proc. Advances in Social Networks Analysis and Mining, 2013, pp. 1441–1444.

[5] J. Ao, P. Zhang, and Y. Cao, "Estimating the locations of emergency events from twitter streams," in Proc. 2nd Int. Conf. on Information Technology and Quantitative Management, 2014, pp. 731–739.

[6] J. Lingad, S. Karimi, and J. Yin, "Location extraction from disaster-related microblogs," in Proc. 22nd Int. World Wide Web Conf. Companion Volume, 2013, pp. 1017–1020.

[7] Z. Cheng, J. Caverlee, and K. Lee, "A content-driven framework for geolocating microblog users," ACM Transactions on Intelligent Systems and Technology, vol. 4, no. 1, p. 2, 2013.

[8] B. Hecht, L. Hong, B. Suh, and E. H. Chi, "Tweets from justinbieber's heart: the dynamics of the location field in user profiles," in Proc. Int. Conf. on Human Factors in Computing Systems, 2011, pp. 237–246.

[9] K. Ryoo and S. Moon, "Inferring twitter user locations with 10 km accuracy," in Proc. 23rd Int. World Wide Web Conf. Companion Volume, 2014, pp. 643–648.

[10] B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti, "Geo- located twitter as proxy for global mobility patterns," Cartography and Geographic Information Science, vol. 41, no. 3, pp. 260–271, 2014.

[11] R. Priedhorsky, A. Culotta, and S. Y. Del Valle, "Inferring the origin locations of tweets with quantitative confidence," in Proc. 17th ACM Conf. on Computer Supported Cooperative Work and Social Computing, 2014, pp. 1523–1536.

[12] B. P. Wing and J. Baldridge, "Simple supervised document geolocation with geodesic grids," in Proc. 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1, 2011, pp. 955–964.