ENGINEERING IN ADVANCED RESEARCH SCIENCE AND TECHNOLOGY

ISSN 2278-2566
Vol.02, Issue.03
October -2018
Pages: -131-136

A SYSTEM FOR REAL-TIME COUNTRY-LEVEL LOCATION CLASSIFICATION OF ONLINE SOCIAL NETWORK

Shaik Ameerunnisa[1], Prasanthi.M [2]
[1] *M. Tech., Dept of CSE, Sri Sunflower College of Engineering and Technology,*ameern84@gmail.com
[2] *M. Tech. Asst Prof, Dept of CSE. Sri Sunflower College of Engineering and Technology,* **,**
santhi512@gmail.com

ABSTRACT
The expansion of enthusiasm for utilizing online life as a hotspot for research has spurred handling the test of consequently geo locating tweets, given the absence of express area data in the lion's share of tweets. As opposed to much past work that has concentrated on area arrangement of tweets confined to a particular nation, here we attempt the assignment in a more extensive setting by arranging worldwide tweets at the nation level, which is so far unexplored in a constant situation. we propose a framework enabling Online Social Network clients to have an immediate control on the messages posted on their dividers. This is accomplished through an adaptable standard based framework, that enables clients to alter the separating criteria to be connected to their dividers, and a Machine Learning based delicate classifier consequently naming messages in help of substance based sifting.
Key-Words :-Twitter,geoLoccation,Online Social Network(OSN),Filtering.

I.INTRODUCTION:-

Social media are increasingly being used in the scientific community as a key source of data to help understand diverse natural and social phenomena, and this has prompted the development of a wide range of computational data mining tools that can extract knowledge from social media for both post-hoc and real time analysis. Thanks to the availability of a public API that enables the cost-free collection of a significant amount of data, Twitter has become a leading data source for such studies [53]. Having Twitteras a new kind of data source, researchers have lookedinto the development of tools for real-time trend analytics [32], [56] or early detection of newsworthy events [51], as well as into analytical approaches for understanding the sentiment expressed by users towards a target [24], [26], [52], or public opinion on a specific topic [5]. However, Twitterdata lacks reliable demographic details that would enable a representative sample of users to be collected and/or afocus on a specific user subgroup [36], or other specific applications such as helping establish the trustworthiness of information posted [34]. Automated inference of social media demographics would be useful, among others, to broaden demographically aware social media analyses that are conducted through surveys [16]. One of the missing demographic details is a user's country of origin, which we study here. The only option then for the researcher is to try to infer such demographic characteristics before attempting the intended analysis. This has motivated a growing body of research in recent years looking at different ways of determining automatically the user's country of origin and/or – as a proxy for the former – the location from which tweets have been posted [1]. Most of the previous research in inferring tweet Geo location has classified tweets by location within a limited geographical area or country; these cannot be applieddirectly to an unfiltered stream where tweets from any location or country will be observed. The few cases that have dealt with a global collection of tweets have used an extensive set of features that cannot realistically be extracted in a real-time, streaming context (e.g., user tweeting history or social networks) [14], and have been limited to a selected set of global cities as well as to English tweets. This means they use ground truth labels to pre-filter tweets originating from other regions and/or written in languages other than English. The classifier built on this pre-filtered dataset may not be applicable to a Twitter stream where every tweet needs to be geo located. An ability to classify tweets by location in real-time is crucial for applications exploiting social media updates as social sensors that enable tracking topics and learning about location-specific trending topics, emerging events and breaking news. Specific applications of a real-time, country-level tweet geolocation system include country-specific trending topic detection or tracking senti- ment towards a topic broken down by country. To the best of our knowledge, our work is the first to deal

with global tweets in any language, using only those features present within the content of a tweet and its associated metadata. We also complement previous work by investigating the extent to which a classifier trained on historical tweets can be used effectively on newly harvested tweets. Motivated by the need to develop an application to identify the trending topics within a specific country1, here we document the development of a classifier that can geo locate tweets by country of origin in real-time. Given that within this scenario it is not feasible to collect additional data to that readily available from the Twitter stream [14], we explore the usefulness of eight tweet-inherent features, all of which are readily available from a tweet object as retrieved from the Twitter API, for determining its geolocation.We perform classification using each of the features alone, but also in feature combinations. We explore the ability to perform the classification on as many as 217 countries, or in a reduced subset of the top 25 countries, as judged by tweet volume. The use of two datasets, collected in October 2014 and October 2015, gives additional insight into whether historical Twitter data can be used to classify new instances of tweets. These two datasets with over 5 million country coded tweets are publicly available. Our methodology enables us to perform a thorough analysis of tweet geo location, revealing insights into the best approaches for an accurate country-level location classifier for tweets. We find that the use of a single feature like content, which is the most commonly used feature in previous work, does not suffice for an accurate classification of users by country and that the combination of multiple features leads to substantial improvement, outperforming the state-of-the-art real-time tweet geo location classifier; this improvement is particularly manifest when using metadata like the user's self-reported location as well as the user's real name. We also perform a per-country analysis for the top 25 countries in terms of tweet volume, exploring how different features lead to optimal classification for different countries, as well as discussing limitations dealing with some of the most challenging countries. We show that country-level classification of an unfiltered Twitter stream is challenging. It requires careful design of a classifier tha uses an appropriate combination of features. Our results at the country level are promising enough in the case of numerous countries, encouraging further research into finer grained geo location of global tweets. Cases where country level geo location is more challenging include English and Spanish speaking countries, which are harder to distinguish due to their numerous commonalities. Still, our experiment show that we can achieve F1 scores above 80% in many of these cases given the choice of an appropriate combination of features, as well as an overall performance above 80% in terms of both micro-accuracy and macro-accuracy for the top 25 countries.

II. Literature Survey:-

1) Securing personal health records in OSN Patient-centric and fine-grained data access control in multi-owner settings

AUTHORS: M. Li, S. Yu, K. Ren, and W. Lou

Online personal health record (PHR) enables patients to manage their own medical records in a centralized way, which greatly facilitates the storage, access and sharing of personal health data. With the emergence of cloud computing, it is attractive for the PHR service providers to shift their PHR applications and storage into the cloud, in order to enjoy the elastic resources and reduce the operational cost. However, by storing PHRs in the cloud, the patients lose physical control to their personal health data, which makes it necessary for each patient to encrypt her PHR data before uploading to the cloud servers. Under encryption, it is challenging to achieve fine-grained access control to PHR data in a scalable and efficient way. For each patient, the PHR data should be encrypted so that it is scalable with the number of users having access. Also, since there are multiple owners (patients) in a PHR system and every owner would encrypt her PHR files using a different set of cryptographic keys, it is important to reduce the key distribution complexity in such multi-owner settings. Existing cryptographic enforced access control schemes are mostly designed for the single-owner scenarios.

In this paper, we propose a novel framework for access control to PHRs within cloud computing environment. To enable fine-grained and scalable access control for PHRs, we leverage attribute based encryption (ABE) techniques to encrypt each patient's PHR data. To reduce the key distribution complexity, we divide the system into multiple security domains, where each domain manages only a subset of the users. In this way, each patient has full control over her own privacy, and the key management complexity is reduced dramatically. Our proposed scheme is also flexible, in that it supports efficient and on-demand revocation of user access rights, and break-glass access under emergency scenarios.

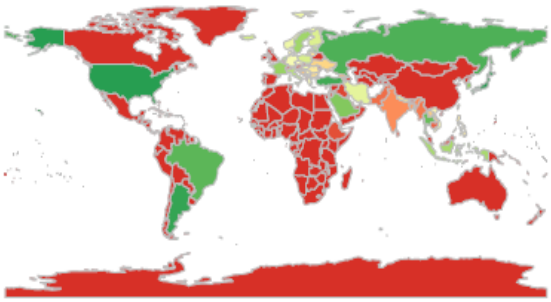
2) Securing the Health Records in OSN

AUTHORS: H. L¨ohr, A.-R. Sadeghi, and M. Winandy

Modern information technology is increasingly used in healthcare with the goal to improve and enhance medical services and to reduce costs. In this context, the outsourcing of computation and storage resources to general IT providers (cloud computing) has become very appealing. E-health clouds offer new possibilities, such as easy and ubiquitous access to medical data, and opportunities for new business models. However, they also bear new risks and raise challenges with respect to security and privacy aspects.

In this paper, we point out several shortcomings of current e-health solutions and standards, particularly they do not address the client platform

security, which is a crucial aspect for the overall security of e-health systems. To fill this gap, we present a security architecture for establishing privacy domains in e-health infrastructures. Our solution provides client platform security and appropriately combines this with network security concepts. Moreover, we discuss further open problems and research challenges on security, privacy and usability of e-health cloud systems.



User language

3) Authorized private keyword search over encrypted personal health records in OSN

AUTHORS: M. Li, S. Yu, N. Cao, and W. Lou

In Online Social Network clients usually outsource their data to the cloud storage servers to reduce the management costs. While those data may contain sensitive personal information, the cloud servers cannot be fully trusted in protecting them. Encryption is a promising way to protect the confidentiality of the outsourced data, but it also introduces much difficulty to performing effective searches over encrypted information. Most existing works do not support efficient searches with complex query conditions, and care needs to be taken when using them because of the potential privacy leakages about the data owners to the data users or the cloud server. In this paper, using on line Personal Health Record (PHR) as a case study, we first show the necessity of search capability authorization that reduces the privacy exposure resulting from the search results, and establish a scalable framework for Authorized Private Keyword Search (APKS) over encrypted cloud data. We then propose two novel solutions for APKS based on a recent cryptographic primitive, Hierarchical Predicate Encryption (HPE). Our solutions enable efficient multi-dimensional keyword searches with range query, allow delegation and revocation of search capabilities. Moreover, we enhance the query privacy which hides users' query keywords against the server. We implement our scheme on a modern workstation, and experimental results demonstrate its suitability for practical usage.

4) Public standards and patients' control: how to keep electronic medical records accessible but private

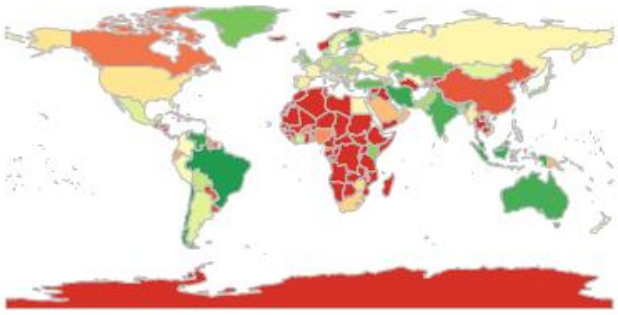
AUTHORS: K. D. Mandl, P. Szolovits, and I. S. Kohane

A patient's medical records are generally fragmented across multiple treatment sites, posing an obstacle to clinical care, research, and public health efforts.1 Electronic medical records and the internet provide a technical infrastructure on which to build longitudinal medical records that can be integrated across sites of care. Choices about the structure and ownership of these records will have profound impact on the accessibility and privacy of patient information. Already, alarming trends are apparent as proprietary online medical record systems are developed and deployed. The technology promising to unify the currently disparate pieces of a patient's medical record may actually threaten the accessibility of the information and compromise patients' privacy. In this article we propose two doctrines and six desirable characteristics to guide the development of online medical record systems. We describe how such systems could be developed and used clinically.

5) Patient controlled encryption: ensuring privacy of electronic medical records

AUTHORS: J. Benaloh, M. Chase, E. Horvitz, and K. Lauter

We explore the challenge of preserving patients ' privacy in electronic health record systems. We argue that security in such systems should be enforced via encryption as well as access control. Furthermore, we argue for approaches that enable patients to generate and store encryption keys, so that the patients ' privacy is protected should the host data center be compromised. The standard argument against such an approach is that encryption would interfere with the functionality of the system. However, we show that we can build an efficient system that allows patients both to share partial access rights with others, and to perform searches over their records. We formalize the requirements of a Patient Controlled Encryption scheme, and give several instantiations, based on existing cryptographic primitives and protocols, each achieving a different set of properties.



User location

III.ENHANCEMENT:

One central issue in today On-line Social Networks (OSNs) is to enable clients to control the messages posted alone private space to stay away from that undesirable substance is shown. Up to now OSNs give little help to this necessity. Users are sharing some location from there user walls that will help to find the locations of different user and also helps this project to advice the friends and neighbours about the different places to visit.

IV. Future Enhancement and Conclusion

Later on we intend to test elective cost-touchy learning ways to deal with the one utilized here, centering particularly on gathering of more information for under-spoke to nations, so the classifier can be additionally enhanced for every one of the nations. Moreover, we intend to investigate more complex methodologies for substance examination, e.g. discovery of points in substance (e.g. do a few nations speak more about football than others?), and also semantic treatment of the content. We likewise intend to create better grained classifiers that take the yield of the nation level classifier as info.

V.REFERENCES

[1] O. Ajao, J. Hong, and W. Liu. A survey of location inference techniques on twitter. Journal of Information Science, 1:1–10, 2015.

[2] E. Amig´ o, J. C. De Albornoz, I. Chugur, A. Corujo, J. Gonzalo,T. Mart´ın, E. Meij, M. De Rijke, and D. Spina. Overview of replab 2013: Evaluating online reputation monitoring systems. In Proceedings of CLEF, pages 333–352. Springer, 2013.

[3] F. Atefeh and W. Khreich. A survey of techniques for event detection in twitter. Computational Intelligence, 31(1):132–164, 2015.

[4] H. Bo, P. Cook, and T. Baldwin. Geolocation prediction in social media data by finding location indicative words. In Proceedings of COLING, pages 1045–1062, 2012.

[5] J. Bollen, H. Mao, and A. Pepe. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In Proceedings of ICWSM, pages 450–453, 2011.

[6] J. D. Burger, J. Henderson, G. Kim, and G. Zarrella. Discriminating gender on twitter. In Proceedings of MNLP, pages 1301–1309, 2011.

[7] H.-w. Chang, D. Lee, M. Eltaher, and J. Lee. @ phillies tweeting from philly? predicting twitter user locations with spatial word
usage. In Proceedings of ASONAM, pages 111–118, 2012.

[8] Y. Chen, J. Zhao, X. Hu, X. Zhang, Z. Li, and T.-S. Chua. From interest to function: Location estimation in social media. In Proceedings of AAAI, pages 180–186, 2013.

[9] Z. Cheng, J. Caverlee, and K. Lee. You are where you tweet: a content-based approach to geo-locating twitter users. In Proceedings of CIKM, pages 759–768, 2010.

[10] R. Compton, D. Jurgens, and D. Allen. Geotagging one hundred million twitter accounts with total variation minimization. In IEEE Big Data, pages 393–401, 2014.

[11] M. Conover, J. Ratkiewicz, M. R. Francisco, B. Gonc¸alves, F. Menczer, and A. Flammini. Political polarization on twitter. In Proceedings of ICWSM, pages 89–96, 2011.

[12] M. D. Conover, B. Gonc¸alves, J. Ratkiewicz, A. Flammini, and F. Menczer. Predicting the political alignment of twitter users. In IEEE PASSAT/SocialCom, pages 192–199, 2011.

[13] D. Doran, S. Gokhale, and A. Dagnino. Accurate local estimation of geo-coordinates for social media posts. arXiv preprint arXiv:1410.4616, 2014.

[14] M. Dredze, M. Osborne, and P. Kambadur. Geolocation for twitter: Timing matters. In Proceedings of NAACL-HLT, pages 1064–1069, San Diego, California, 2016.

[15] M. Dredze, M. J. Paul, S. Bergsma, and H. Tran. Carmen: A twitter geolocation system with applications to public health. In HIAI Workshop, pages 20–24, 2013.

[16] M. Duggan. The demographics of social media users. Pew Research Center, 2015.

[17] J. Eisenstein, B. O'Connor, N. A. Smith, and E. P. Xing. A latent variable model for geographic lexical variation. In Proceedings of EMNLP, pages 1277–1287, 2010.

[18] M. Graham, S. A. Hale, and D. Gaffney. Where in the world are you? geolocation and language identification in twitter. The Professional Geographer, 66(4):568–578, 2014.

[19] B. Han, P. Cook, and T. Baldwin. Text-based twitter user geolocation prediction. Journal of Artificial Intelligence Research, pages 451–500, 2014.

[20] B. Hecht, L. Hong, B. Suh, and E. H. Chi. Tweets from justin bieber's heart: the dynamics of the location field in user profiles . In Proceedings of CHI, pages 237–246, 2011.

[21] B. R. Heravi and I. Salawdeh. Tweet location detection. In Computation + Journalism Symposium, 2015.

[22] W. Huang, I. Weber, and S. Vieweg. Inferring nationalities of twitter users and studying inter-national linking. In Proceedings of Hypertext, pages 237–242, 2014.

[23] D. J. Hughes, M. Rowe, M. Batey, and A. Lee. A tale of two sites: Twitter vs. facebook and the personality predictors of social media usage. Computers in Human Behavior, 28(2):561–569, 2012.

[24] L. Jiang, M. Yu, M. Zhou, X. Liu, and T. Zhao. Target-dependent twitter sentiment classification. In Proceedings of ACL, pages 151–160, 2011.

[25] D. Jurgens. That's what friends are for: Inferring location in online social media platforms based on social relationships. In Proceedings of ICWSM, pages 273–282, 2013.

[26] E. Kouloumpis, T. Wilson, and J. D. Moore. Twitter sentiment analysis: The good the bad and the omg! In Proceedings of ICWSM, pages 538–541, 2011.

[27] R. Krishnamurthy, P. Kapanipathi, A. P. Sheth, and
K. Thirunarayan. Knowledge enabled approach to predict
the location of twitter users. In The Semantic Web. Latest Advances
and New Domains, pages 187–201. Springer, 2015.

[28] V. Lampos, N. Aletras, J. K. Geyti, B. Zou, and I. J. Cox. Inferring the socioeconomic status of

social media users based on behavior and language. In Proceedings of ECIR, pages 689–695, 2016.
[29] K. Lee, R. K. Ganti, M. Srivatsa, and L. Liu. When twitter meets foursquare: tweet location prediction using foursquare. In Proceedings of MobiQuitous, pages 198–207, 2014.
[30] C. Li and A. Sun. Fine-grained location extraction from tweets with temporal awareness. In Proceedings of SIGIR, pages 43–52, 2014.
[31] W. Liu and D. Ruths. What's in a name? using first names as features for gender inference in twitter. In AAAI Spring Symposium: Analyzing Microtext, volume 13, page 01, 2013.
[32] A. Madani, O. Boussaid, and D. E. Zegour. Real-time trending topics detection and description from twitter content. Social Network Analysis and Mining, 5(1):1–13, 2015.
[33] J. Mahmud, J. Nichols, and C. Drews. Home location identification of twitter users. ACM TIST, 5(3):47:1–47:21, 2014.
[34] S. E. Middleton and V. Krivcovs. Geoparsing and geosemantics for social media: spatio-temporal grounding of content propagating rumours to support trust and veracity analysis during breaking news. ACM Transactions on Information Systems, 34(3):1–27, 2016.
[35] Z. Miller, B. Dickinson, and W. Hu. Gender prediction on twitter using stream algorithms with n-gram character features. International Journal of Intelligence Science, 2(04):143, 2012.
[36] A. Mislove, S. Lehmann, Y.-Y. Ahn, J.-P. Onnela, and J. N. Rosenquist.Understanding the demographics of twitter users. In Proceedings of ICWSM, volume 11, page 5th, 2011.
[37] D. Murthy, A. Gross, and A. Pensavalle. Urban social media demographics: An exploration of twitter use in major American cities. J. of Computer-Mediated Communication, 21(1):33–49, 2016.
[38] M. Naaman, A. X. Zhang, S. Brody, and G. Lotan. On the study of diurnal urban routines on twitter. In Proceedings of ICWSM, pages 258–265, 2012.
[39] T. Palpanas and P. Paraskevopoulos. Fine-grained geolocalisation of non-geotagged tweets. In Proceedings of ASONAM, pages 105– 112. IEEE, 2015.
[40] M. Pennacchiotti and A.-M. Popescu. Democrats, republicans and starbucks afficionados: user classification in twitter. In Proceedings of KDD, pages 430–438, 2011.
[41] M. Pennacchiotti and A.-M. Popescu. A machine learning approach to twitter user classification. In Proceedings of ICWSM, pages 281–288, 2011.
[42] D. Preot¸iuc-Pietro, V. Lampos, and N. Aletras. An analysis of the user occupational class through twitter content. In Proceedings of ACL, pages 1754–1764, 2015.
[43] D. Preot¸iuc-Pietro, S. Volkova, V. Lampos, Y. Bachrach, and N. Aletras. Studying user income through language, behaviour and affect in social media. PloS one, 10(9):e0138717, 2015.
[44] A. Priante, D. Hiemstra, T. van den Broek, A. Saeed, M. Ehrenhard, and A. Need. #whoami in 160 characters? classifying social identities based on twitter profile descriptions. In NLP+CSS, pages 55–65, 2016.
[45] V. Rakesh, C. K. Reddy, D. Singh, and M. Ramachandran. Location-specific tweet detection and topic summarization in twitter. In Proceedings of ASONAM, pages 1441–1444, 2013.
[46] D. Rao, D. Yarowsky, A. Shreevats, and M. Gupta. Classifying latent user attributes in twitter. In Proceedings of the workshop on Search and mining user-generated contents, pages 37–44, 2010.
[47] C. Robusto. The cosine-haversine formula. The American Mathematical Monthly, 64(1):38–40, 1957.
[48] E. Rodrigues, R. Assunc¸ ˜ao, G. L. Pappa, R. Miranda, andW. Meira. Uncovering the location of twitter users. In Proceedings of BRACIS, pages 237–241, 2013.
[49] S. Roller, M. Speriosu, S. Rallapalli, B. Wing, and J. Baldridge. Supervised text-based geolocation using language models on an adaptive grid. In Proceedings of EMNLP, pages 1500–1510, 2012.
[50] D. Rout, K. Bontcheva, D. Preot¸iuc-Pietro, and T. Cohn. Where's@ wally?: a classification approach to geolocating users based on their social ties. In Proceedings of Hypertext, pages 11–20, 2013.
[51] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In Proceedings of WWW, pages 851–860, 2010.
[52] R. Townsend, A. Tsakalidis, Y. Zhou, B. Wang, M. Liakata, A. Zubiaga, A. Cristea, and R. Procter. Warwickdcs: from phrase-based to target-specific sentiment recognition. SemEval, page 657, 2015.
[53] K. Weller, A. Bruns, J. Burgess, M. Mahrt, and C. Puschmann. Twitter and society. Peter Lang New York, 2013.
[54] B. P. Wing and J. Baldridge. Simple supervised document geolocation with geodesic grids. In Proceedings of ACL, pages 955–964 , 2011.
[55] A. Zubiaga, I. San Vicente, P. Gamallo, J. R. Pichel, I. Alegria, N. Aranberri, A. Ezeiza, and V. Fresno. Tweetlid: a benchmark for tweet language identification. Language Resources and Evaluation, 50(4):729–766, 2016.
[56] A. Zubiaga, D. Spina, R. Mart´ınez, and V. Fresno. Real-time classification of twitter trends. JASIST, 66(3):462–473, 2015.

Shaik Ameerunnisa is a student of sri sunflower college of Engineering and Techonology, Lankapalli Present she is Pursuing her M.tech[Computer Science & Engineering] from this college and she received B.tech Degree(Bachelor of techonology) from the University of JNTUK,Kakinada.

Title : A System for Real-Time, Country-Level Location Classification of Online Social Network.

Prasanthi.M Associate Professor in Sri Sunflower College of Engineering and Technology. Lankapalli and Also Received M.tech from JNTUK University.Having 10 years of Experience in Teaching Faculty

INTERNATIONAL JOURNAL OF ENGINEERING IN ADVANCED RESEARCH SCIENCE AND TECHNOLOGY
Volume.02, **IssueNo.**03**,** October -2018**, Pages:** 131-136