A SURVEY ON IDENTIFYING IDENTICAL USERS   AMONG MULTIPLE SOCIAL MEDIA SITES

Anju Viswam[1], Gopu Darsan[2]

[1] PG scholar, Department of Computer Science and Engineering, Sree Buddha College of Engineering, Alappuzha, India.
[2] Assistant Professor, Department of Computer Science and Engineering, Sree Buddha College of Engineering, Alappuzha, India.

---***---

Abstract: *In social media networks, profile details of one user can be used by others to create account with original user identity or the original user may have multiple accounts in multiple social media sites. Discovery of multiple accounts that belong to the same person is an interesting and challenging work in social media analysis. Profiles, contents and network structures can be used for user identification in social media sites. We are conducting a survey in this paper to identify the identical users among multiple social media sites based on the profile details of the users.*

Keywords:-**social networks, user identification, cross-communities, social-tagging systems.**

1. INTRODUCTION

There are many types of social networking sites which immensely contributed large number of real world data on social behaviors. People tend to use multiple social media sites for different purposes and all the existent social media sites will satisfy some user needs. Before using any social network, the user has to be registered to the network by submitting their profile information. It is possible to gather the user's information from different social networks. For instance, Facebook offers a view to the profiles of the users', LinkedIn reveals professional profiles and Twitter broadcasts interests. This knowledge can be exploited and it is a serious threat to the user's privacy.

By using the profile information an anonymous user can create an account in the social media site with the identity of the original user. As the available social media sites provides different purposes, a user can have multiple accounts in multiple or same social network. Two profiles pointing to the same person can have different profile attribute such as name, address. It may be modeled on different schemas depending on the social networks. These profile information may consist of structured or semi-structured data. So identifying unique identities from multiple profiles is a challenging task.

Aggregating profiles among multiple social networks reveals more information about a person and it is beneficial for personalization. However, for privacy reasons people do not want their accounts to be connectable in different social networks. Indeed, connecting profiles is risky. So it is difficult to identify anonymous users in different social media networks. No solution can fully identify the identical anonymous users, but with some social media networks elements a portion of identical users can be identified. In this paper we are conducting a survey to identify the users across social media sites based on the profile attributes.

3. LITERATURE SURVEY

The profile attributes can be used to identifying anonymous yet identical users in multiple social media sites.

Perito et al. [1] presented a scheme to linking multiple online profiles using usernames. For that they devise an analytical model to identify the uniqueness of a username which can be used to assign a probability to a single username from two different online services. The same model was used for the user identification when usernames are different across many online services. Their method identifies the uniqueness of usernames and shows that users will choose their usernames from a small set and reuse them across multiple services.

Liu et al. [2] deals with a way to differentiate users with the same usernames. This approach was used for automatic acquisition of training data that is based on the following two observations. Rare usernames are owned by a single natural person and common usernames are owned by different natural person. Using n-gram probability of a username they estimated the rareness or commonness of a username. The training instance shows the positive or negative nature of the usernames. From that they verified the effectiveness of the classifiers trained with the

automatically generated training data and differentiated users with the same usernames.

R. Zafarani and H. Liu [3] analyzed the cross-communities. Here cross -communities refers to multiple websites. They performed mapping among identities across multiple websites and provided a method for connecting these websites. This approach is based on the concept that usernames appear in the URLs of the profile web pages. They use this principle to extract username sets for each username for user identification. Their observation shows that usernames can be used to identify corresponding usernames in various communities. But the main challenge is that the same usernames does not necessarily guarantee the same identity.

R. Zafarani and H. Liu [4] further developed a user mapping method on cross-communities by modeling user behavior on screen names. They demonstrated a methodology for connecting individuals across social media sites (MOBIUS). The system categorizes the behavioral patterns as patterns due to human limitations, exogenous factors, and endogenous factors. Limited time and memory, limited knowledge are the patterns due to human limitations. Exogenous factors are behaviors observed due to the environment that the user is living in. Endogenous factors refer to the profile attributes and characteristics of a person. From these various features are constructed to capture redundancy of information. MOBIUS employs a supervised learning to connect users and this can be used for user identification to link users in cross-communities.

Acquisti et al. [5] addressed the user identification task using profile photos. They conducted the experiment on Facebook using face recognition algorithm. Facebook profile photos are visible to all by default. Most of the members use photos of them as primary profile image and use real first and last names on their profiles. Face recognition of everyone or everywhere or all the time is not yet feasible. Both screen name and profile image can identify users but they cannot be applied to large social media networks. This is because some users may have the same screen name and profile images.

Iofciu et al. [6] investigated whether the users can be identified on the basis of tags across social tagging systems. They combine the information such as user ids and their tags. They analyze the profiles of users from Flickr, Delicious and StumbleUpon. They exploit tagging behavior and usernames to construct user profiles for identifying the users. Their study shows that it is possible to identify users across systems based on their tagging behavior even though the tagging behavior varies across different platforms. They introduced a symmetric variant of BM25 using site specific statistics and when compare it against other measures like TF, TFIDF and conventional BM25, it produces better performance.

M. Motoyama and G. Varghese [7] used the profile attributes such as name, age, geographical location, etc. for searching and matching users across social networks. They randomly selected profiles from Facebook and MySpace. They trained a classifier using boosting to identify whether a match exists in the considered social networks. Initially they considered email address for the searching purpose. They used tools from learning theory and natural language processing that can assist in the analysis of social networks. Their results present an analysis of the overlapping profiles in the social networks.

Goga [8] linked accounts by exploiting usernames, real names, locations, and photos. They performed the experiment on five social networks: Twitter, Facebook, Google+, MySpace, and Flickr. Similarity matching was performed on the profile attributes. Naive Bayes algorithm was used to decide if two accounts match based on the probability that each feature's similarity belongs to the matching class. A decision tree was used to decide if two accounts match by traversing a tree of questions until they reach a leaf node. They identified about 80% of the matching pairs of user accounts between any combination of two social networks among Twitter, Facebook and Google+.

Cortis [9] presented a weighted ontology-based user profile resolution technique which discovers the multiple online profiles that refer to the same person. They proposed a profile matching algorithm with Text Analytics that can be applied on both semi structured and structured profile information. The use of Linked Open Data (LOD) improved the Named Entity Recognition (NER) process. They used a Semantic-based matching extension to find any possible semantic relations between incomplete and unstructured profile attributes. The proposed work can be used for various real life applications such as detection of anonymous profiles and contact recommendations based on common locations, activities, friends and interests.

Elie Raad et al [10] proposed a framework for identifying social profiles that refer to the same person between two social media sites. They investigate the areas such as social network profile heterogeneity, similarity measuring between attribute values, and decision making about whether two profiles refer to the same person or not. This framework was able to discover the possible number of

profiles that refer to the same user that existing approaches are unable to detect. Aggregation functions developed were used for data fusion and for decision making.

4. CONCLUSIONS

Identifying anonymous identical users among multiple social media sites is a challenging task. We conducted a survey on identifying multiple user identities across multiple social media networks. The above studies address the user identification problem by considering the profile attributes such as screen name, gender, profile photo etc. User identification can also be done using the social media network attributes such as content and network. These attributes can also used for user identification process in social networks. The choice may rely on the human user's involvement.

REFERENCES

[1]   D. Perito, C. Castelluccia, M. A. Kaafar, and P. Manils, "How unique and traceable are usernames?" in Proc. 11th Int. Conf. Privacy Enhancing Technol., 2011, pp. 1–17.

[2]   J. Liu, F. Zhang, X. Song, Y. I. Song, C. Y. Lin, and H. W. Hon, "What's in a name?: An unsupervised approach to link users across communities," in Proc. 6th ACM Int. Conf. Web Search Data Mining, 2013, pp. 495–504.

[3]   R. Zafarani and H. Liu, "Connecting corresponding identities across communities," in Proc. 3rd Int. ICWSM Conf., 2009, pp. 354–357.

[4]   R. Zafarani and H. Liu, "Connecting users across social media sites: a behavioral-modeling approach," in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2013, pp. 41–49.

[5]   A. Acquisti, R. Gross, and F. Stutzman, "Privacy in the age of augmented reality," in Proc. Nat. Acad. Sci.,2011,pp.3653.Available:https://www.unisex.org/legacy/events/sec11/tech/slides/acquisti.pdf.

[6]   T. Iofciu, P. Fankhauser, F. Abel, and K. Bischoff, "Identifying users across social tagging systems," in Proc. 5th Int. AAAI Conf. Weblogs Social Media, 2011, pp. 522–525.

[7]   M. Motoyama and G. Varghese, "I seek you: searching and matching individuals in social networks," in Proc. 11th Int. Workshop Web Inf. Data Manage., 2009, pp. 67–75.

[8]   O. Goga, D. Perito, H. Lei, R. Teixeira, and R. Sommer, "Large-scale correlation of accounts across social networks," University of California at Berkeley, Berkeley, California, Tech. Rep. TR-13-002, 2013.

[9]   K. Cortis, S. Scerri, I. Rivera, and S. Handschuh, "An ontology based technique for online profile resolution," in Proc. 5th Int. Conf. Social Informat., 2013, pp. 284–298.

[10]   E. Raad, R. Chbeir, and A. Dipanda, "User profile matching in social networks," in Proc. 13th Int. Conf. Netw.-Based Inf. Syst., 2010, pp. 297–304.