



International Journal of Advance Engineering and Research Development

Special Issue on Recent Trends in Data Engineering

Volume 4, Special Issue 5, Dec.-2017

Detecting Unknown or Fake User Accounts on Different Micro blogging and Social Media Networks

Meenakshi A.Thalor¹, Ashish Mehrotra², Chaitali Kelkar², Priyanka Shirsat², Shantaling Kapse²

¹Assistant Professor, Computer Engineering, AISSMS's IOIT, Pune

²Student, Computer Engineering, AISSMS's IOIT, Savitribai Phule Pune University, Pune

Abstract — Nowadays different kinds of social media networking sites have become apparent and each and every person is associated with such type of social media networks with different or similar identity. Each and every social media networking site has a different network structure and detecting anonymity, however matching users among multiple social media network is a challenge. Most of the existing algorithms work on profile based or network structure to find an anonymous user which is not effective. This paper highlights Friend Relationship Based User Identification algorithm which identifies duplicate users among heterogeneous network structure.

Keywords-Anonymous Users, Identical Users, Social Media Network

I. INTRODUCTION

Nowadays most widely used social media networks are Titter, Facebook, Linkedin, Youtube where the identity of a user is composed of three dimensions of attributes as given below:

1. Profile: Profile information of person includes his/her name, gender, address, age, profile image etc.
2. Content: Content information includes the text/images/video creates or shared by his/her.
3. Network: Connection information which describes his/her network of friend, number of friends etc.

Detecting duplicate users among cross social media networking site using profile based approach is simplest and easiest approach in literature where the profile name, age, gender and even profile image is used in order to find out matching score. Content based user identification is moderated approach where the geo location along with the text posted by a person is also used for user matching. Network structure based user identification is complex approach where some sort of graph mining came into picture and network of friends and friends of friend is also considered to find out user similarity on heterogeneous social networking sites.

II. LITERATURE SURVEY

This paper presents surveyed across the web regarding the research, implementations and methodologies used to detect anonymous users in Social Media Networks which is summarized below in table 1.

Table1: Literature Review of Existing work

Research Paper	Approach	Limitation
"How Unique and Traceable are Usernames?"[1]	It is clearly possible to tie digital identities together and, most likely, to real identities in many cases only using ubiquitous usernames. Spammers could gather information across the web to send extremely targeted spam, which we dub E-mail spam.	Authors are not able to test their tool in linking profiles of certain types of web services in which users are more privacy aware, like dating and medical websites (e.g., WebMD).
"What's in a name?: an unsupervised approach to link users across communities"[2]	1. Social Relationship based Features. 2. Post Content based Features	Because of using usernames only 60 % results are accurate.
"Connecting Corresponding	An Approach to Cross-Community	It provides correct result up to 60%

Identities across Communities"[3]	Corresponding Username Elicitation	only because it is using username.
"Identifying Users Across Social Tagging Systems"[4]	1. Matching Users based on their Tags 2. Matching Users based on Explicit Usernames 3. Matching Users based on Combined Profiles. 4. Combining Username and Tags	Using only username and tagged based attribute.
"I Seek You: Searching and Matching Individuals In Social Networks"[5]	The data collection system we used to perform the seeding and searching is comprised of three components: supervisors, vaults, and workers.	It works on Facebook pages which must be online. Dataset work on different attribute which found online only so time consumption is less.
"An Ontology-based Technique for Online Profile Resolution"[6]	1. Ontology-driven Profile Resolution Technique 2. Text Analytics 3. Semantic-based matching Extension 4. Ontology-based Attribute Weighting	1. The real 5-person dataset was used to determine the threshold 2. Eliminate matches found between profiles stating a different gender.
"User profile matching in social networks"[7]	1. Intersection between SN1 and SN2 2. Union between SN1 and SN2 3. Difference between SN1 and SN2	The user profiles are matched accordingly 1. Data Isolated Islands 2. Site/User Objectives
"Cross-system User Modeling and Personalization on the Social Web"[8]	1. Form-based Profiles 2. Tag-based Profiles 3. Mypes	Completeness of user profiles is dependent on service to service.
"Finding Nemo: Searching and Resolving Identities of Users Across Online Social Networks"[9]	1. Profile Search. 2. Content Search 3. Self-mention Search 4. Network Search	Results are dependent on identity exposure since some users can hide their information thereby making it difficult to user information.
"Connecting Users across Social Media Sites: A Behavioral-Modeling Approach"[10]	Mobius: Behavioral Patterns And Feature Construction	Limitations in Time and Memory: 1. Selecting the Same Username. 2. Username Length Likelihood.

III. PROPOSED SYSTEM

Identifying a user is represented by different keywords like user matching, user identity resolution, user recognition and anchor linking. No doubt, Profile attributes provides effective information for user matching but easy to impersonate. Geo location is one of the widely used attribute comes under content based user identification but limitation of this approach is that very few people share their geo location while post their message. Under Network structure various algorithms have been proposed like Network- Structure (NS), Joint-Link attribute (JLA) and Friend Relationship Based User Identification (FRUI). We are using here FRUI algorithm along with profile and content based user identification to find out anonymous users on SMN. For Profile and content based user identification methods are used together to find out initial user matched pairs. FRUI uses user matched pair generated by profile and network structure based user identification method and computes more candidate user matched pairs and user matched pair with highest score will be considered as identical users. FRUI works in iteration mode and obtain or identify more UMPS using prior UMPS.

Propose system (refer fig. 1) contains basic three modules as listed below:

1. Profile based user Identification: From limited number of profiles, first we will compute user matching using profile attributes and these user matches will act as Prior UMPS.
2. Identifier: The purpose of Identifier is to detect a new set of UMPS using network structure information of prior UMPS in the iteration process. In next iteration resultant UMPS are union with Prior UMPS and produce new UMPS and that Process keeps up until no more resultant UMPS.
3. When there is no more resultant UMPS from previous module then retrieve user matching based on highest rank of UMPS.

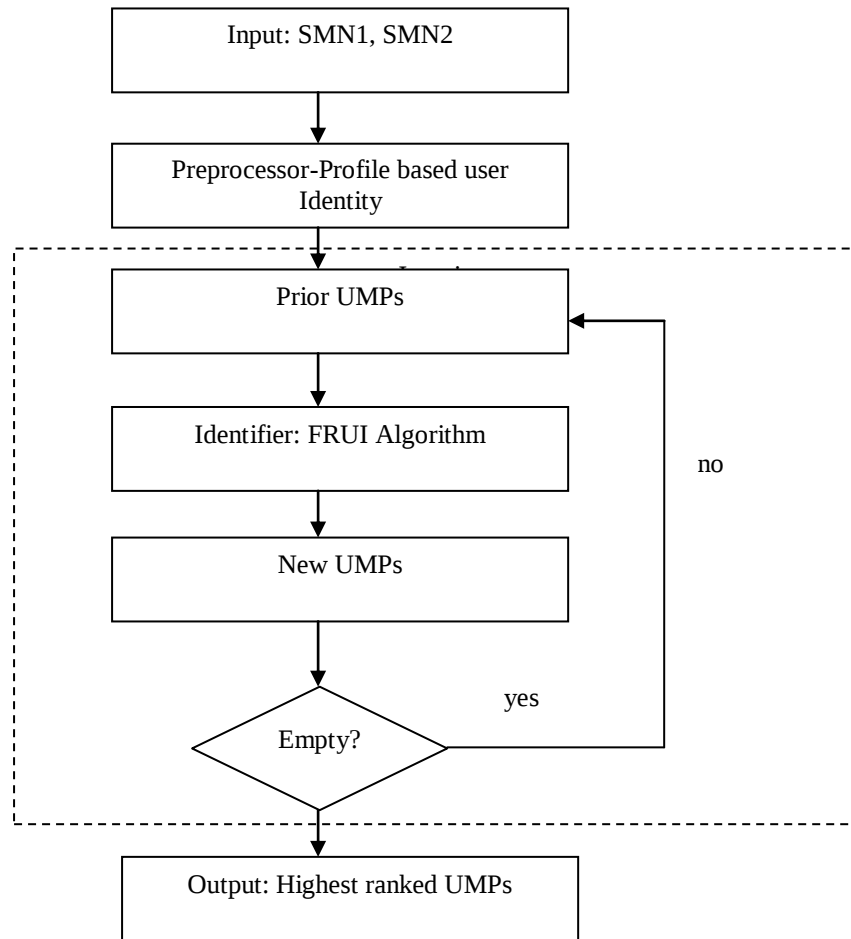


Fig.1: Proposed System Architecture

In literature different algorithms are used which treated ad identifier. In our approach here we will use FRUI algorithm which uses the number of the shared known friends among cross platform social media network sites.

Conclusion

To conclude, we have presented a FRUI algorithm which is an attempt to highlight the identity resolution problem in social media networks. Rather than rely on one type of identity resolution approaches here we are emphasizing on combination of identity resolution approaches where the outcome of one of identity resolution approach will be act as input for FRUI algorithm. The performance of system will improve with inclusion of different identity resolution techniques. At initial stage, we will apply FRUI to two social media networking sites which will be created by us. But at later stage we will test our approach with other social media services so that usefulness of proposed algorithm can be proved.

REFERENCES

- [1] Perito D., Castelluccia C., Kaafar M.A., Manils P., "How Unique and Traceable Are Usernames?.", *Lecture Notes in Computer Science*, vol 6794. Springer, Berlin, Heidelberg, 2011
- [2] J Liu, F Zhang, X Song, YI Song, CY Lin, HW Hon , "What's in a name?: an unsupervised approach to link users across communities" *Proceedings of the sixth ACM international conference on Web search and data mining*, 2013
- [3] Reza Zafarani and Huan Liu, "Connecting Corresponding Identities across Communities", *Proceedings of the Third International ICWSM Conference*, 2009
- [4] Tereza Iofciu, Peter Fankhauser, Fabian Abel, Kerstin Bischoff , "Identifying Users Across Social Tagging Systems", *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011
- [5] Marti Motoyama and George Varghese, "I Seek You: Searching and Matching Individuals in Social Networks", *Proceedings of the eleventh international workshop on Web information and data management*, pp. 67-75, 2009

- [6] Keith Cortis, Simon Scerri, Ismael Rivera, and Siegfried Handschuh , “An Ontology-based Technique for Online Profile Resolution”, *International Conference on Social Informatics*, pp. 284-298,2013.
- [7] Elie Raad, Richard Chbeir, Albert Dipanda. “User profile matching in social networks”, *Network Based Information Systems (NBIS)*, Sep 2010, Japan. pp.297-304, 2014.
- [8] Fabian Abel, Eelco Herder, Geert-Jan Houben, Nicola Henze, Daniel Krause, “Cross-system User Modeling and Personalization on the Social Web”, *User Modeling and User-Adapted Interaction*, Volume 23, Issue 2–3, pp 169–209,2013
- [9] Paridhi Jain, Ponnuram Kumaraguru,“Finding Nemo: Searching and Resolving Identities of Users Across Online Social Networks”, *arXiv preprint arXiv:1212.6147*.
- [10] Reza Zafarani and Huan Liu ,“Connecting Users across Social Media Sites: A Behavioral- Modeling Approach”, *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, PP. 41-49,2013