ABHIYANTRIKI
An International Journal of Engineering & Technology
(A Peer Reviewed & Indexed Journal)
Vol. 3, No. 5 (May, 2016) http://www.aijet.in/ eISSN: 2394-627X

Social Identity Linkage using Heterogeneous Behavior Model

Prasanna K Bhat*
PG Student
Deptt. of Computer Science Engineering
RNS Institute of Technology, Bengaluru,
Karnataka, India

G T Raju
Professor
Deptt. of Computer Science Engineering
RNS Institute of Technology, Bengaluru,
Karnataka, India

Abstract

In today's world people make use of different social media platforms for different purposes. Often the information provided on the individual site is not complete and fragmented. Linkage of identity across social media helps to gain deeper understanding of user profiles. Deeper understanding of user data mainly helps in business intelligence. This paper contains a framework called hydra which consists of 3 steps: (I) model the heterogeneous behavior of user (II) build the structure consistency model (III) final step is optimization. Experiments have been conducted on databases where the framework does the linkage correctly and a profile will be built.

Keywords*: Identity linkage, Social platform, Heterogeneous model.*

****Author for correspondence*** pkbhat123@gmail.com

1. Introduction

The recent development in the social network in all services has made revolution in our social life by allowing us to share all kinds of data. We can share text, video, images, blogs, tweets and many things. Due to this large amount of data is created and it is very difficult handle this big social data. People are struggling get the information of a use because many of the times data of a user is incomplete and fragmented. To overcome this problem, a better idea would be linking a particular user across many platforms so that it helps to gain more information about a user and better business intelligence. It helps us to gain following benefits:

Completeness: One social network may give partial information about a user using one particular perspective. It will be better if we connect user across various social media to get detailed information.

Consistency: User may give false and inconsistent information in a social network. We get consistent information if we cross check the information in other social networks.

Continuity: Due to some reasons some social networks may stop their service but the users who are using it remains the same. And they shift to other social network. So we get better information if we gather information from multiple social networks.

In this paper, the problem of linking users across social platforms is studied. This paper makes use of linkage algorithm for linking users across sites.

2. Related Work

Various researches have been done on the area of social identity linkage. Previous research is divided in to three categories: linking based on user profile, linking based content generated by user and last one is linking based on user behavior. In user profile based methods we use the tagging information provided by user [2]. The methods used in this type lack in user tagging, personal identifiable information and user profile privacy. Second method is linking based on content generated by user [3], it makes the assumption of consistent usernames. So it lacks much information. Linking based on user behavior considers the user behavior in social media [4]. The previous methods have not handled the missing information properly and also the reason behind that one. Authorship identification is the process where that identifies the authors by checking language and writing style. Previous studies contained two methods: based of content and behavior model. First method checks for content features across vast number of documents [5, 6]. To check content ownership, second method absorbs writing style features [8] or builds language models. Most authorship identification methods compromised due to complex network structure and high degree of missing information.

3. Methodology

Problem formulation: Consider p as set of all natural persons in real world. For one social network platform s, let Ts be set of all user names belonging to a distinct user and Θs: Ts ->p, the injective function mapping every online user of s to a natural person.

Definition of social identity linkage: Given two social networks platforms a and b, the problem of Social Identity Linkage is to find out a function f to check if any two users from a and b respectively correspond to the same natural person, i.e. Ts × Ts'-> {0, 1} such that for any pair of users (ui,ui') € Ts × Ts', we have

$$f(ui, ui') = \begin{cases} 1, \text{ if } \Theta s(ui)=\Theta s'(ui'), \\ 0, \text{ otherwise} \end{cases}$$

In this paper, a framework called HYDRA [1] is proposed which combines user's heterogeneous behavior and core social structure. This framework consists of three main modules.

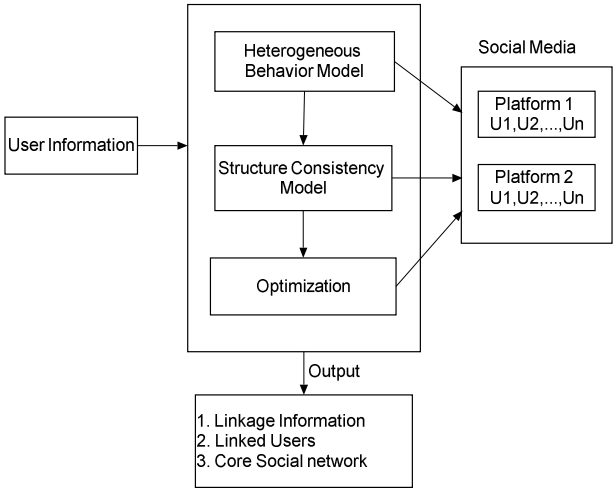

Fig. 1: System Architecture

Step 1 – Behavior Similarity Modeling: In this paper, the measurement of behavior similarity between two users is done.

Step 2 – Structure Consistency Modeling: Here the structure consistency model is built by considering core social network structure of user and similarity in behavior.

Step 3 – Optimization: In this step dealing with missing information is done. In many approaches they have not considered the missing information.

Figure 1 shows the system architecture of project. It consists of three steps. Those are heterogeneous behavior model, structure consistency model and optimization. In heterogeneous behavior model, consideration of all structured and unstructured data generated by user is done.
(a) Attributes: Usually it contains the structured data generated by a user. It contains basic information of user like name, age, gender, location and contact details etc.
(b) Generated content: It contains the unstructured data generated by a user. It can be text, reviews, comments, images and tweets etc.
(c) User behavior trajectory. It means behavior of a user along the time line.

Modeling of User Attributes
Textual Attributes: These attributes are nothing but the structured information like name, age, gender, education, work, location and email etc. This information is used to distinguish user profiles. Other attributes than email are not so effective in distinguishing because other attributes can be same for many users.
Visual Attributes: This attribute is the images in profile. It also helps in distinguishing the profile. Here we make use of face recognition tools to detect and crop the face in the images.

Modeling of User Topics
The important feature of social network is over a period of time behaviors of the user can change. And he can generate large number of interests. We can use the interests and favorites to compare between two users of different platforms like facebook and twitter. If the users have same interests we can have to take them for pairing the users.

Modeling of User Style
User writing style is factor in distinguishing users. User style is seen in tweets, comment and re-tweets. First the removing stop words from the collected tweets, comments and re-tweets of the specific user, then select the unique keywords and compare with the user across the platform [7].
In structure consistency model, consider the core social network structure and behavior similarity of users for linkage between user pairs. Core social network structure is the frequently interacting friends. For matched users usually friends will be the same. Sentiment analysis is also done here. Generally users bring their friends to social network. Finally perform the optimization where dealing with the missing information for the user behavior model and structure consistency model. Here consider the null values instead of ignoring it.

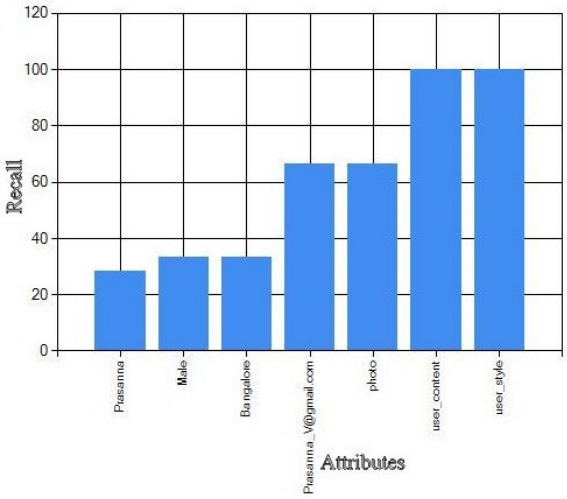

Fig. 2: Recall

4. Result and Conclusion
For experimentation purpose, we made use of databases similar to facebook and twitter. Then we have performed the necessary operations on the database to retrieve the linkage information of a user. In figure 2, recall value is calculated against attributes and precision remains constant.

Hereby concludes that the usage of HYDRA framework helps in social identity linkage to get overall information of a user which in turn helps in business intelligence.

References

[1] Siyuan Liu, Shuhui Wang, & Feida Zhu. (2015). Structured Learning from Heterogeneous Bheavior for Social Identity Linkage. *IEEE Transactions on Knowledge and Data Engineering*. Vol. 27, no. 7.

[2] T Iofciu, P Fankhauser, F Abel, & K Bischoff. Identifying users across social tagging systems. *In:* ICWSM'11.

[3] 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. *In:* WSDM'13.

[4] R Zafarani, & H Liu. Connecting users across social media sites: A behavioral-modeling approach. *In:* KDD'13.

[5] O de Vel, A Anderson, M Corney, & G Mohay. (2001). Mining e-mail content for author identification forensics. *SIGMOD Record*. 30(4), pp. 55–64.

[6] R Cilibrasi, & PMB Vitanyi. (2005). Clustering by compression. *IEEE Transactions on Information Theory*. pp. 1523–1545.

[7] J Weston, C Leslie, E Ie, D Zhou, A Elisseeff, & W Noble. (2005). Semi-supervised protein classification using cluster kernels. *Bioinformatics*. pp. 55–64.

[8] R Zheng, J Li, H Chen, & Z Huang. (2006). A framework for authorship identification of online messages: Writing-style features and classification techniques. *Journal of the Association for Information Science and Technology*. 57(3).