

# IDENTIFYING THE REAL USER IN THE SOCIAL NETWORK USING CLASSIFIERS

Neha Doggalli<sup>1</sup>, R Senthil Kumar<sup>2</sup>

<sup>1</sup>Computer Science and Engineering, The Oxford college of Engineering, and Bangalore  
*neha.doggalli@gmail.com*

<sup>2</sup>Computer Science and Engineering, The Oxford college of Engineering, and Bangalore  
*senkr.raj@gmail.com*

## Abstract

Social identity linkage is essential to connect different social media platforms. It is critically needed for business intelligence to investigate about the user profiler of the social media. To address this issue HYDRA framework is used. It has three steps mainly heterogeneous behavior, structure consistency and normalized margin based linkage. In heterogeneous behavior model, the behavior of certain attributes are matching against the information missing and the behavior similarities are described to know the person is real or not. In structure consistency model, the purpose is to construct the structure and behavior regularity of users across different social networks thus the identity link can be performed on set of users. In normalized margin based linkage, it propose a function and learn the linkage function by considering multiple optimization where both overseen pair wise linkage function and structure consistency are conducted to get optimal solution. This model can find the missing information and helps to find out the linkage between the two platforms. This work summarizes about the identification of user profiles or the information using HYDRA.

**Keywords:** Data Mining, Structure Consistency, Normalized Margin.

\*\*\*

## I. INTRODUCTION

“Many people working from years for detecting multiple identities in the social media sites but in the cyber virtual world this has been achieved in recent years. The current social sites have a provision of sharing video, favorite food, likes, dislikes, memorable events to share with their loved ones on the other-hand due to large scale of social media database the business man are searching for a method to handle this big data for improving their profit margin. The common people or companies like Amazon are unable to find the right information of the user due to non-indexed social media data. In most of the cases the information of the user is obtained but it won't complete some part of it will be omitted, disruptive and inconsistent.

The solution to this problem to link all the social sites and believe that common information of user in different platform is true. Social network like Orkut will provide information of the user for one particular aspect, in the same way the information can be obtained from LinkedIn, Facebook etc by performing this task the person's behavior can be judged technically this is titled as completeness.

The user can have multiple accounts in social network therefore some information mentioned will be obviously false so we need to check in various platforms to match the true information of the user in other words it is called as consistency.

In order to link the information of the user which is present in the different social media without human intervention a

machine learning method is essential by observing the data huge records of the user's profile which are available online. The challenges facing by research scholars are as follows

- **Attributes variation:** the user may provide different names to the social network during the time of registration. For example some people may add the family name before or after their name and this varies from region to region. In worst case for misleading the opposite sex the person may hide his marital status, age etc. Some male users are going to create their profile with female names mainly to mislead. Therefore some of the classifiers like SVM [4], Neural networks [5] are user to determine the genuine information of the user in social network. SVM is going to generate the hyper plane which accurately differentiate between two classes it make use of support vectors to decide to which class does the test image belongs when they fall near to hyper plane. Neural Network uses the back propagation algorithm to find the accurate operating point using the weights to the respective nodes.
- **Misplaced Data:** behavior of the user on social networks various from one platform to other For example the person personal recreational activities can shared in Facebook but about his professional life is shared on twitter i.e., the user information which are going to be gathered are platform dependent. The social network having N number of media categories like technical blogs, forms, video, image etc. due to these reasons the linkage for identities will become tough.
- **Similar upload of data in different platform:** based on the interest of the user, they may upload the same content at

different time slots this leads to collection of false information due to variation in the day, time and platform. The system should be able to read the data compare with the existing (similarity models are used for this purpose) data and it should avoid the storage of repeated information.

## II. BACKGROUND

### A. Discovering the linkage using web data

The method to determine the linkage points of the users in the data sources i.e., determine the attributes which are shared between stored records. The matching between the stored databases to find the similarities is the basis idea to determine the linkage in the [1] social network to determine the accurate user. The match operator is incorporated to match the databases of two social networks. The very first step in collecting the data is to find the link points which are used in the later stage to equivalent the attributes. The Goal of the work is to align the un-structured data and separate the overlapped information collected from different social networks. The objective of the work is to align the overlapped data provided from the data providers. They proposed a new method called as Search by set similarities, it measures the value of the each node and provides the score to for node to find the exact attribute, it align the schema structure and improve the efficiency in linkage discovery of the web data. The results have shown that the proposed methods efficiently worked on the real time data.

### B. Unsupervised approach to link the user over social networks

Worked on the linkage of users in different online data sources. More specifically they are able to distinguish the users who had given the similar user-name to achieve this they have incorporated an alias-disambiguation [2] test before beginning the node linkage. They have collected the names of about seventy thousand similar names. Due to pre-testing the success rate was around fifty percent raise, in accurately distinguish the users with similar user Ids'. By the survey of these names they have came up with the novel method called as unsupervised learning which can able to label the training set automatically. The positive labels are provided to rare name i.e. unfamiliar names. Common names are included under the negative label. Proposed a method by name n-gram to differentiate between the familiar and non-familiar names. They have tested the proposed work on the yahoo datasets.

## III. EXISTING SYSTEM

These system uses the single class classifiers model, they provide efficient results when the user are providing only two labels (Positive and negative label) which determines the information of the overlapped username.

The single class classifiers fails while classifying the users data in other language like Hindi, Marathi, Chinese and this results in the low throughput in linking the users profile to gather the missing data.

## IV. PROPOSED SYSTEM

The key features the work includes are explained in this section

- **Heterogeneous Behavior Model:** have designed a novel model which is able to detect the similarities present in the user's behavior which includes IP number of device, location, image sharing etc. By collecting these information the model is able to align the un-structured data and find the missing attribute of the user.
- **Structure Consistency:** a new method is proposed which observes the user's core structure instead of behavioral similarity. We are going to check the different social platform and common information obtained are considered as true. This model is able to find the user profile even though the ground truth label is absent.
- **Multi-objective Model Learning:** We make use of multi-objective optimization (MOO) framework [8] to determine the social linkage, which includes both ground truth label and supervised learning models are joined together to meet Pareto optimality. The structure of the kernel has been modified and developed a normalized margin based approach which finds the missing information where similarity modeling fails to do it.

The proposed system makes use of multi class classifiers and multi-level objective level therefore the efficiency of the system has increased by four times than the existing system.

## V. IMPLEMENTATION

This segment explains how the theoretical part turns into working system. In this part it is implemented using different modules such user, admin and the linkage shown between the single user in different platforms. To achieve this linkage between two social network we make use of different steps:

### 1. Heterogeneous Behavior

This model describes the heterogeneous behaviour Of a particular user by considering many attributes across different platform so that they can get the missing information and consider the behavior similarity which are described by dimensional similarity vector for each user pair.

### 2. Structure Consistency

This module construct structure consistency models to maximize the structure and behavior regularity of users' across different social platforms, thus the task of identifying the linkage can be performed on set of users through this consistency model.

### 3. Social Structure Linkage

The social structure linkage to link the overall structure for people using the social networks .structure consistency maximization by modelling the core social networks behavior consistency. They are complementary to each other by jointly measuring the behavior similarity of both individual and group levels. There are multiple social network available for this paper. social networks identifying

the user(profile and content and overall structure) data with Structure.

### C. Modules

Proposed system uses different modules such as user registration, user login and Admin.

In the user registration module, the user uploads their basic profile data. For identifying the particular user the username

and password could have duplicate values and then it stores in the database. User logins using username and password and user can update the basic profile information with profile photo and his interests, user can upload the post and blogs and can sends the friend request to known person in the network and can accept the request who sent to him by doing this he will be having list of friends in his list as shown in the Fig 1

Fig 1. user profile update

User upload the various post using the tag name, description about the particular post and user and friends circle can comment on that particular post as shown in fig 2

Fig 2. upload post.

## VI. EVALUATION RESULTS

The fake accounts are increasing day by day earlier based on the email, location the real people were identified but this methodology is less accurate. We are proposing the framework which is able to identify the real user based on the blogs written, user attributes. For the experimental phase

we are creating the two database consider one as face-book and another as twitter. The database includes blogs, profile photo, location and some basic personal information. After creating the database based on the behavior of the user the system is able to identify the real user as shown in the Fig 3.

**Social Identity Linkage**

**USERS**

Name	DOB	Mobile	Email	Gender	Address	State	District	Nationality	FavGame	FavMusic	FavPlace	Profile Pic
NEHA	1991-02-15	8553723868	neha@gmail.com	FEMALE	#212/ZNr RoadDavangere	karnataka	DAVANGERE	INDIAN	volley ball	pop	MYSORE	
swaroop	1986-03-12	9844791292	swaroop@gmail.com	MALE	hno 234chamraj nagamysore	karnataka	mysore	indian	Not Specified	Not Specified	Not Specified	

**Fig 3.** social identity Linkage

The experimental result is shown in the above diagram which consider several different attributes about the basic profile information, user interest and the user profile picture the identity of the same user will be greater than 70% than we can say the user is real across the different platform.

## VII. VII CONCLUSION

We proposed a Framework which is able to identify the real user based on the user attributes, the blogs written. Based on the behaviour of the user the system is able to identify the real user if this condition is failed a structure consistency and multi-objective model learning methods are incorporated to identify the real people the latter method make use of the multiple attributes which includes collection of information from the mutual friends and other attributes through machine phase.

The results are simulated in the eclipse and the algorithm is working efficiently to identify the real user.

## ACKNOWLEDGMENT

Proposed work is a job of great enormity and it can't be accomplished by an individual all by them. Eventually, I am grateful to a number of individuals whose professional guidance, assistance and encouragement have made it a pleasant endeavor to undertake this project. I am very much grateful to the Vice Principal and Head of the Department Dr. R.J. Anandhi for her unfailing encouragement and suggestion given to me in the course of my project work. I convey my sincere gratitude to my guide Mr R Senthil Kumar, Associate professor, Department of CSE, for having constantly guided and monitored the development of the project work.

A note of thanks to the Department of Computer Science Engineering, both teaching and non-teaching staff for their

co-operation extended to us. I thank my parents for their constant support and encouragement. Last, but not the least, I would like to thank my peers and friends.

## REFERENCES

- [1] O. Hassanzadeh, K. Q. Pu, S. H. Yeganeh, R. J. Miller, M. Hernandez, L. Popa, and H. Ho, "Discovering linkage points over web data," *PVLDB*, vol. 6, no. 6, pp. 444–456, 2013.
- [2] R. Zafarani and H. Liu, "Connecting corresponding identities across communities," in *ICWSM'09*, 2009, pp. –1–1.
- [3] 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 *WSDM'13*, 2013.
- [4] R. Zafarani and H. Liu, "Connecting users across social media sites: A behavioral-modeling approach," in *KDD'13*, 2013.
- [5] P. Jain and P. Kumaraguru, "@i to @me: Identifying Users across multiple online social networks," *CoRR*, 2013.
- [6] J. Vosecky, D. Hong, and V. Shen, "User identification across multiple social networks," in *NDT'09*, 2009, pp. 360–365.
- [7] P. Jain and P. Kumaraguru, "@i to @me: An anatomy of username changing behavior on twitter," *CoRR*, 2014.
- [8] T. Iofciu, P. Fankhauser, F. Abel, and K. Bischoff, "Identifying users across social tagging systems," in *ICWSM'11*, 2011, pp. –1–1.
- [9] I. Bhattacharya and L. Getoor, "Collective entity resolution in relational data," *ACM Transactions on Knowledge Discovery from Data*, pp. –1–1,