International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-2, Issue-5, May 2016
ISSN: 2395-3470
www.ijseas.com

Smart Connect-Cross Platform User Linkage Model Based on Heterogeneous behaviour

Pavan Bharadwaj H
Dept. of Computer Science & Engg.,
K.S Institute of Technology,
Bengaluru, Karnataka.

Ashiq M.E
Dept. of Computer Science & Engg.,
K.S Institute of Technology,
Bengaluru, Karnataka.

B S Pavan
Dept. of Computer Science & Engg.,
K.S Institute of Technology,
Bengaluru, Karnataka.

Vijayalaxmi Mekali
Dept. of Computer Science & Engg.,
K.S Institute of Technology,
Bengaluru, Karnataka.

***Abstract* - With the concept of interacting with people the emergence of social networks achieved a greater success in the current decade. This has led to the existence of many social media platforms where in which people tend to have their identity in most of the popular social networks. With this development the challenge was how to gain a business insight with huge collection of social data as the users had multiple identities in different platforms. In this paper we propose a solution in which the linkage is performed among same users with multiple identities so that a better profile can be built to achieve business intelligence. We build our model using three key steps where first, we analyse users' basic profile and topical discussion over a period of time. Secondly, analysis on the content is made across different platform and linkage is established via the contents shared. Finally, we analyse on users' core social structure and try linking among groups so that level of accuracy is increased. The model will be able to efficiently handle with the sparse values and aggregating among key profile values.**

Keywords—heterogeneous behavior, business intelligence, topical analysis.

I. INTRODUCTION

The recent advent of social medias has made the users connect among each other easily irrespective of the location or identity. This trend in the past few years has led in the drastic increase of the number of people interacting via social networking platforms. This has resulted in the evolution of different social medias with its own features. With every new platforms coming into existence the corresponding number of users migrating has also been increasing.

The major challenge with the users migrating from one platform to another platform is that the user identity may not be the same in all the different platforms. Thus it becomes difficult to understand the users' core identical characteristics for any purpose which can be for example, business intelligence. The difficulty here is how to gain a thorough understanding about the user with multiple identities. Therefore we require a model which efficiently links the users among multiple platforms and give a better profiling of users. The major advantage with this are briefed as follows.

Completeness: User may provide some additional information in one platform which he might have not provided in other platform. Linkage across the platform may reduce the incompleteness level as a whole.

Continuity: As the social networks come and go the underlying person will remain the same. Therefore continuity can be achieved if cross-platform linkage is performed.

Handle missing values: User may not provide complete information about him in a single platform. Meanwhile, user may give some information in one platform and he may not give the same in other. Therefore this model can efficiently handle with missing values.

In this paper what we propose is, we analyse user's heterogeneous behaviour across the platform. Here by behaviour we mean user actions over a social network. Moreover user activities can be represented by various types of media such as tweets, status updates, videos, images, location which we call as heterogeneous behaviour . We consider much more challenging aspect where we examine multiple features over a period of time with the sparse and misaligned information and cross-platform linkage is done.

Some of the key contributions are discussed as follows.

- Heterogeneous behaviour model: It is used to measure the similarity in user behaviour using user's core social data. This would be able to efficiently handle with the sparse representations.
- Structure consistency: We try to maximise users' behaviour consistency by analysing behaviour over different intervals of time.
- Multi-objective learning: Identities can be linked by both ground truth and cross-platform linkage which increases consistency.

In Social data two important features are considered to be unique (I) Behaviour of the user over a period of time (II) Users core structure which is formed by those who are close to the user. Users core structures across the different platforms do have some common similarity and offers higher characterization.

II. Related Work

User linkage was considered as connecting among similar identities over multiple networks and a web based approach was proposed for it in [12]. Social structure based linkage perform linkage based on the features in social circles. Korula et al. [1] solve the reconciliation of users social network by starting from nodes with higher degrees. Koutra et al. [5] propose a solution by learning an optimal permutation function between two graph affinity matrices. Based on users social, geographical, temporal and textual information, Kong et al.[7] propose a Multi-Network anchoring to find the links between users from different platform. Zhang et al.[3] propose to predict social and location links inside the target social network given a set of anchor links among users from target network and source network.

We can identify the authors of documents by their writing styles and the language from their corresponding documents. This is achieved by two key methods: content based and behaviour-model based method. Content based methods identify features across a large number of documents while behaviour model based capture the writing style.

III. Problem Overview

Given two social networks platform P and P', the problem of linking is to find a function to decide if any two users from P and P' correspond to the same user.

Direct approach to solve the problem by examining each pair of user would result in high computational cost. We calculate similarity among pair of users via heterogeneous behaviour modelling. Later we construct the structure consistency by analysing user behaviour over a period of time. As we know social relationship can be better understood using social graph rather than multi-dimensional database. When linkage across the platform has to be done it becomes necessary to consider every individual aspect associated with the user. In multi-dimensional databases every dimension refers to a attribute and it becomes to gain a deeper insight to the data. Thus social graph help us to build a better linkage function which can help us achieve our objective efficiently as the analysis has to be made on groups to which the user may belong to.

IV. Heterogeneous Behaviour Model

As this paper is mainly concentrated on user behaviour analysis across the platform several challenges comes into picture while we mean by heterogeneous behaviour. Therefore very high heterogeneity of user social data can be achieved by the following categorization.

A. *Based on attributes*

This attribute provides the distinguishing properties for different users. The common attributes are basic information like user's name, DOB, gender, email id, company name, nationality, education. In this model there is a possibility of having same properties between the users like having same name, gender, nationality etc. so proper training of model is required. By the above credentials a simple matching strategy can be built. The relative attributes of credential can be estimated by data counting. In this model we also provide the user to upload profile photo so that it can be used to guess who is the user and it also help in linking with other users. However the user may not provide his true image or provide image with poor illumination. For that false constraint we design a matching module to compare every bit of that pixel and provide the proper information to user about profile images.

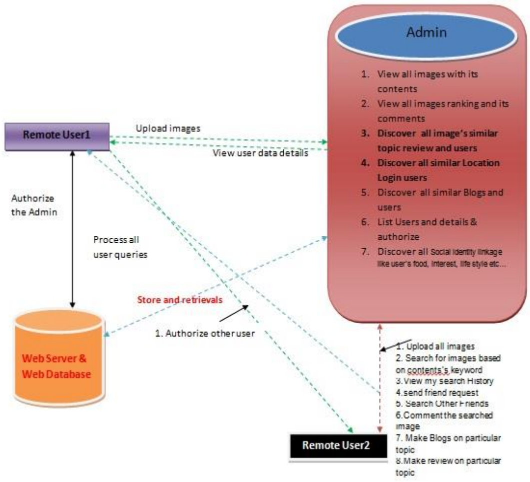

Fig.1: System Architecture

Interaction between the users and admin of the platform involves authorization. Users can also interact with each other. However, the admin can view all the users and also their details and use these details for designing the linkage function.

B. *Topical Analysis*

This is one of the interesting model as the user is provided with an option to search the contents by his wish, however faking of information by the other end user provides no proper data to the proper users. Meanwhile the framework would be overcoming this hurdle by monitoring the activities of the user over a platform in order to make sure that the information provided would be fair as per the activities. This is achieved using latent Dirichlet allocation along with the collected textual pre-processing procedure. If two persons view the single person profile then their inclination tends to be similar in the temporal range and the temporal range is also calculated with respect to the multi scale temporal topic distribution.

C. *Based on users expressing style*

This model is helpful in judging the user behaviour based on the language style he makes use of while interacting with other users. The language style of user is reflected in different areas like comments section where he has to express his views and also express his emotion so that model could be able to judge his expressing styles. By the reviews we can distinguish particular user from the other users. To judge the characteristic style this model analyses the statements and scans each word and add them to the database. The most unique words are added to the database and frequencies of those words are checked. The similarity among the words is also found for the further analysis.

D. Modelling on Multi-resolution behaviour

The behaviour of an individual person is unique and it is closely examined by the system. The social behaviour of the user exhibited on the timeline of this system like commenting the post, friend request etc. is closely analysed by the model and the process is termed as user behaviour trajectory. This paper mainly focuses on the following patterns.

Social media sites having the feature of location based service provide a strong path to share the location of user using recording of location information and this pattern of sharing is called as Location and Mobile Trajectory Information. Users post or share the unstructured data in the system like video, audio, images, text and this data might be unique or duplicate and this pattern of sharing is called Multimedia Content Generation and Sharing. The solution to the above problems of duplicate data sharing and wrong location sharing can be solved is by constructing pattern matching sensors so that each information like location, Image, video, audio is recorded by the sensors analyses the information based on the user behaviour.

V. CORE SOCIAL NETWORK FEATURES

It is understood that user bring other users with similar behavior and friends to multiple social network platforms. Thus the information about the users friends and their activities are also very informative in designing the linkage function. Given two users S and S' over different platforms, the behavior data of most frequently interacting friends are collected. This is used as measure for behavior consistency of the user and their friend's group.

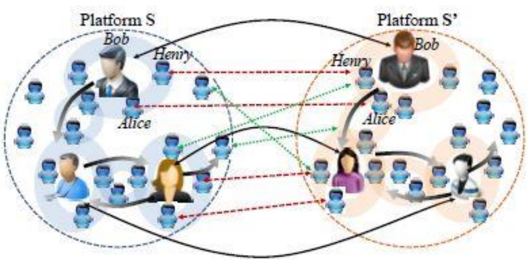

Fig.2: Structure consistency model
Given two platforms, we measure the similarity in behavior among the frequently communicating friends, especially among users with linkage information. The arrows within the platform indicate how the linkage is done along the social structure of each user.

A. Multi objective Structure learning

Based on the heterogeneous behaviour model from user attributes we proposed a linkage function via multi objective function. Different social media allow the user to login to different system with one account .For example the user can login to the twitter system with the help of providing the credential to facebook system .Similarly in the same way our system provides the opportunity for the user to login to the system, with the other social media platform. We collect the user data from the authorized social media, and the data collected will be off high security such that no other person can utilize the data. And with those data the user is registered with our system.

B. Structure Consistency Modelling:

As said above in the Structure Learning, the ground truth information about the user and his credentials is every important to login to the system and to judge the behaviour of the user. If the user has provided false information, then that data would affect other end users. The proper analysis of this system is explained with the help of an example of 3 users Alice, Bob and Henry. If the 3 users are friends in real life and if they are friend in our system then the frequency of interaction among them would be probably high. In the figure 1, it explains clearly about the nature of user. The dashed red arrow indicates that those peoples are strongly connected and sharing same set of properties. The process of separating same object to one group and different object to another group is called cluster. The dashed green arrows indicate the incorrect user and they are separated with the help of clusters, and they are not connected to the strongly connected users. When the strong ground truth information is not available between the users still they are connected on the basis of ground truth provided in the other social platforms. Our System uses 3 modules in order to execute the complete operation of the system. They are,

- Basic Information Linkage
- Content Oriented Linkage
- Social Structure Linkage

C. Basic Information Linkage

User profile-based methods collect tagging information provided by users or user profiles from several social networks and then represent user profiles in vectors, of which each dimension corresponds to a profile field. Methods in this category suffer from huge effort of user tagging, different identifiable personal information types from site to site, and privacy profile**.**

D. Content Oriented Linkage

We build structure consistency models to maximize consistency on user's core social structure across different platforms. Thus the task of identity linkage can be performed on groups of users. It collects personal identifiable information from public pages of user-generated content. Yet these methods still make the assumption of consistent usernames across social platforms, which is not the case in large-scale social networks platforms. User-behavior-model based methods analyze behavior patterns and build feature models from usernames, language and writing styles.

E. Social Structure Linkage

The main responsibility of this module is to link the overall structure of the people using the social networks. Jointly measuring the behaviour similarity of individual user and groups. Here analysis is made on groups as a whole in order to achieve optimization. In the proposed system, the system proposes HYDRA, a framework for cross platform user identity linkage via heterogeneous behavior modeling Compared with the long studied record linkage problem.

VI. MODEL ANALYSIS

The system analyses the model based on the two objective functions. First it uses supervised learning method to obtain the ground truth provided by the user and then structure consistency maximization by the core social networks behaviour. Both are linked together and they are depended on each other i.e., If the user provide some information, and the ground level truth is off less percentage of about 30% then it searches social network platform for the information about the user and this is done with the authentication of user. Even when the ground truth information is sufficient, the model uses generalization power in order to achieve good group behaviour, so that the system can provide better efficiency and consistency.

VII. CONCLUSION

In this paper, our approach of aggregating the user profile across multiple social media platform helps in gaining a better insight for business intelligence where in accuracy of user profile is of critical importance. This paper overcomes the complexity of achieving better identity of users with multiple identities in several platforms by behaviour analysis. The major principle behind this is the user profile can be faked but not the behaviour. Thus, achieving better result than profile analysis technique.

ACKNOWLEDGMENT

This study was supported by Mrs.Vijayalaxmi Mekali, Asst. Professor, Department of Computer Science and engineering, K.S Institute of Technology, Bengaluru with her valuable suggestions and guidance.

REFERENCES

[1] N. Korula and S. Lattanzi, "An efficient reconciliation algorithm for social networks," PVLDB, pp. 377-388, 2014.

[2] P. Jain and P. Kumaraguru, "@i to @me: An anatoni of user name changing behaviour on twitter," CoRR, 2014

[3] J. Zhang, X. Kong, and P. S. Yu, "Transferring heterogeneous links across location-based social networks," in WSDM'14, 2014, pp. 303– 312.

[4] R. Zafarani and H. Liu, "Connecting users across social media sites: A behavioural modelling approach," in KDD'13, 2013.

[5] D. Koutra, H. Tong, and D. Lubensky, "Big-align: Fast bipartite graph alignment," in ICDM'13, 2013, pp. 389–398

[6] J. Liu, F. Zhang, X. Song , C.-Y. Lin, and H.-W. Hon, "What's in a name?: an unsupervised approach to link user across Communities," in WSDM'13, 2013.

[7] X. Kong, J. Zhang, and P. S. Yu, "Inferring anchor links across multiple heterogeneous social networks," in CIKM'13, 2013, pp. 179–188.

[8] L. Chu, S. Jiang, S. Wang, Y. Zhang, and Q. Huang, "Robust spatial consistency graph model for partial duplicate image retrieval," IEEE Transactions on Multimedia, 2013.

[9] A. Malhotra, L. C. Totti, W. M. Jr., P. Kumaraguru, and V. Almeida, "Studying user footprints in different online social networks," in ASONAM'12, pp. 1065–1070, 2012.

[10] J. Wang, G. Li, J. X. Yu, and J. Feng, "Entity matching: How similar is similar," PVLDB, pp. 622–633, 2011.

[11] G. Pickard, W. Pan, I. Rahwan, M. Cebrian, R. Crane, A. Madan, and A. Pentland,"Time-critical social mobilization," Science, vol. 334, no. 6055, pp. 509–512, 2011.

[12] R. Zafarani and H. Liu, "Connecting corresponding identities across communities," in ICWSM'09, 2009, pp. –1–1.

[13] G. Chechik, G. Heitz, G. Elidan, P. Abbeel, and D. Koller, "Max-margin classification of data with absent features," Journal of Machine Learning Research, pp.1–21, 2008.

[14] R. Zheng, J. Li, H. Chen, and Z. Huang, "A framework for authorship identification of online messages: Writing-style features and classification techniques," Journal of the Association for Information Science and Technology, vol. 57, no. 3, 2006.