

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.

B S Pavan
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.

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.

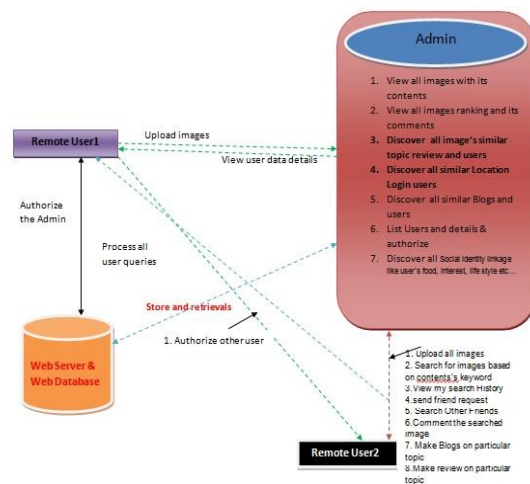


Fig.1: System Architecture

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

