EasyChair Preprint

№ 2018

Privacy preserving in online social networks using recommendation system

R Indumathy

November 24, 2019

PRIVACY PRESERVING IN ONLINE SOCIAL NETWORKS USING RECOMMENDATION SYSTEMS

R.Indumathy 1
1 Department of Computer Technology
Anna University
Chennai 600 044, India

Abstract—**Social media platforms allow users to communicate, share information and innovating the web content. In social media, personalized recommendation is essential to assist users realize pertinent data. However delivering such user activity data makes users liable to inference attacks, as private data typically conclude from user activity data.To protect the user privacy when releasing the pubic data from any other third party services and to avoid the inference attack data masking technique is used. For data masking Differential privacy is used to ensure the privacy regardless of the adversary's prior knowledge. To save the utility of the data, data distortion using a pairwise ranking loss metric, i.e., Jensen-Shanon distance is utilized.**

***Keywords*: Recommendation sytem, Privacy protection, Data masking.**

I. INTRODUCTION

Social organizing service (also social organizing site or social media) is an online stage which individuals use to assemble interpersonal organizations or social relationships with other individuals who offer comparable individual or profession interests, exercises, foundations or genuine associations. The social network is classified into varied pc networks. The social networks are inherited by pc networks, linking folks, organization, and information Social media platforms allow users to communicate, share information and innovating the web content. In social media, personalized recommendation is essential to assist users realize pertinent data. However delivering such user activity data makes users liable to inference attacks, as private data typically conclude from user activity data. To apply privacy preserving data publishing techniques in social media, one quick strategy is to change the user details before sent to social media. To protect the user privacy when releasing the pubic data from any other third party services and to avoid the inference attack data masking technique is used. For data masking Differential privacy is used to ensure the privacy regardless of the adversary's prior knowledge. To save the utility of the data, data distortion using a pairwise ranking loss metric, i.e., Jensen-Shanon distance is utilized.

The purpose of this work is to to provide user with customizable privacy protection. To provide the continuous protection of user specified private data against inference attack. To guarantee the utility of the data using pair wise ranking loss.

The first part of the Chapter describes the Introduction about the topic Privacy preserving Online social networks. The second part of the chapter provides the various techniques used in Existing System and their limitations. Chapter 3 provides the proposed system of architecture and their modules. Chapter 4 provides the Implementation. Chapter 5 discusses the results and discussion.

II. LITERATURE SURVEY

To preserve privacy in social media for data publishing, anonymization techniques for hiding the sensitive attributes, privacy preserving in Recommendation system, privacy preserving in Machine learning techniques, and Clustering methods are used .

Hatem AbdulKader et al has proposed a framework for masking sensitive data in OSN user's profiles. The proposed framework based on two main steps. First, reconstruct the profile data by setting privacy level to each attribute. After that , construct an association rule hiding algorithm based on the utility of privacy setting. A mining analysis attack can be conducted by other users or third party social network application on user's profiles data to discover the relevant pattern of users. The proposed framework will protect the user's profiles sensitive frequent attribute-sets against attacks from

OSN profiles data against attacks from user even OSN application .

Davis et al has proposed an unsupervised method—SociRank—which recognizes news topics common in both social media and the news media, and then ranks them by their MF, UA, and UI as relevance factors. The secular prevalence of a particular topic in the news media is considered the MF of a topic, which gives insight into its mass media popularity.

Xu Yuan et al (2017) has proposed a two-sided Cross Domain Collaborative Filtering model. Assume that there exist two auxiliary domains, i.e., user-side domain and item-side domain, where the user-side auxiliary domain shares the same aligned users with the target domain, and the item-side shares the same aligned items. Also both the two auxiliary domains contain heavy rating data.

III. PROPOSED APPROACH

A. Pre-processing

.Data Pre processing is a data mining technique that involves transforming raw data into an understandable format. technique that involves transforming raw data into an understandable format. Data Preprocessing is required because of Incomplete, Noisy, Inconsistent. Data cleaning, also called data cleansing or scrubbing. Fill in missing values, smooth noisy data, identify or remove the outliers, and resolve inconsistencies. Data cleaning is required because source systems contain “dirty data” that must be cleaned .

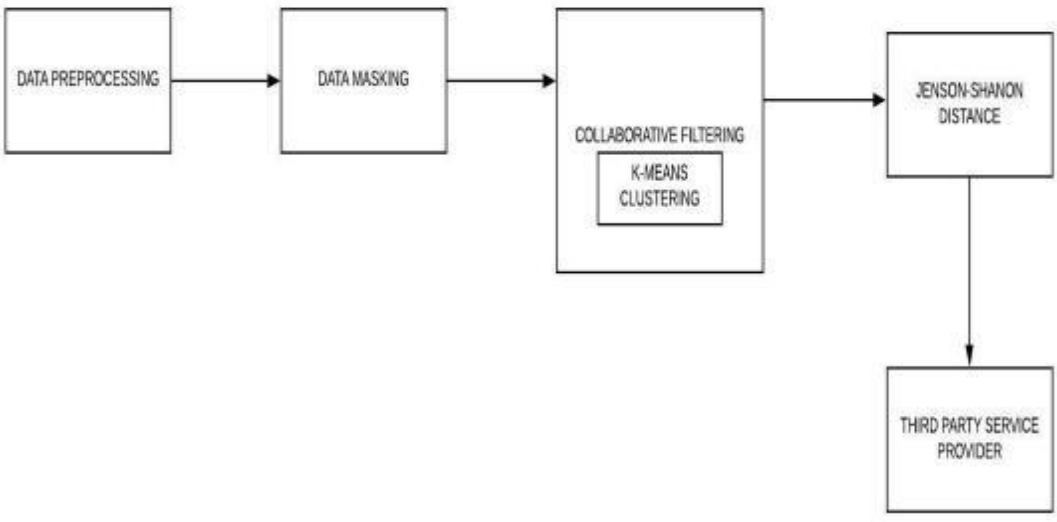
B. Data masking

Data masking is the process of hiding original data with modified content. Data masking is important because it protects the private information but it gives the freedom to conduct robust testing with realistic data and outsource testing and development.For Data masking, Differential privacy method is used. Mask individuals within the data by creating a sanitization point between user interface and data.

Differential privacy

Differential privacy is a system for publicly sharing information about a dataset by describing the patterns of groups within the dataset while withholding information about individuals.

Fig 1 Proposed system



IV. IMPLEMENTATION DETAILS

DATASET CREATION

The dataset is collected from Foursquare check-in. It contains global-scale Foursquare check-ins collected via Twitter Public Streams for about 18 months (from Apr. 2012 to Sep.2013). A user is uniquely identified by her ID. An activity is represented by the category corresponding to the POI such as gym or restaurant. By using Anaconda the dataset is preprocessing for removing the missing values.

V. EXPERIMENTAL RESULTS

In python, the data preprocessing is used to filling the missing values, smooth noisy data, identify or remove the outliers, and resolve inconsistencies.

userId	venueId	venueCate	venueCate	latitude	longitude	timezoneO	utcTimestamp
1541	4f0fd5a8e4	4bf58dd8d4	Cosmetics Shop		139.61959	540	Tue Apr 03 18:17:18 +0000 2012
868	4b7b884ff9	4bf58dd8d4	Ramen / N	35.715581	139.80032	540	Tue Apr 03 18:22:04 +0000 2012
114	4c16fdda96040f477cc47	Convenien		35.714542	139.48007	540	Tue Apr 03 19:12:07 +0000 2012
868	4c178638c2	4bf58dd8d4	Food & Drir	35.725592	139.77663	540	Tue Apr 03 19:12:13 +0000 2012
1458	4f568309e4	4f2a210c4b	Housing De	35.656083	139.73405	540	Tue Apr 03 19:18:23 +0000 2012
1541	4b83b207f9	4bf58dd8d4	Furniture /	35.705074	139.6195	540	Tue Apr 03 19:20:09 +0000 2012
NAN	4ea281c302	4d954b0ea	Convenience Store		139.61778	540	Tue Apr 03 19:21:00 +0000 2012
114	4b3eae5cf9	4bf58dd8d4	Train Statio	35.700253	139.48025	540	Tue Apr 03 19:35:36 +0000 2012
1635	4cca7bd679	4bf58dd8d4	Other Grea	35.755759	139.73357	540	Tue Apr 03 19:51:50 +0000 2012
2033	4b5c7671f9	4bf58dd8d4	Ramen / N	35.693121	139.69945	540	Tue Apr 03 19:51:59 +0000 2012
589	4b5ed39cf9	4bf58dd8d4	Airport	35.548963	139.78461	NAN	Tue Apr 03 19:59:06 +0000 2012
589	4e014c11c6	4bf58dd8d4	Bus Station	35.575028	139.75879	540	Tue Apr 03 20:01:54 +0000 2012
2290	4e538e914	4d954b0ea	Convenien	35.750751	139.58206	540	Tue Apr 03 20:08:42 +0000 2012
NAN	4ba88a28f9	4bf58dd8d4	Ramen / N	35.70919	139.77382	540	Tue Apr 03 20:09:35 +0000 2012
589	4d69a46cde	4bf58dd8d4	Bridge	35.609929	139.82566	540	NAN
2290	4b53b05ef5	4bf58dd8d4	Train Statio	35.749538	139.58654	540	Tue Apr 03 20:14:18 +0000 2012
886	4b5a91e5f9	4d954b0ea	Convenien	35.613889	139.7248	540	Tue Apr 03 20:19:58 +0000 2012
NAN	4b6e3e46f5	4bf58dd8d4	Train Statio	35.729025	139.7111	540	Tue Apr 03 20:28:32 +0000 2012
265	4b3b5742f9	4bf58dd8d4	Train Statio	35.5532	139.64681	540	Tue Apr 03 20:30:32 +0000 2012
159	4b243a7df964a5203564	Train Statio		35.729865	139.71096	540	Tue Apr 03 20:33:55 +0000 2012
589	4b4162b1f9	4bf58dd8d4	Train Statio	35.645924	139.82692	540	Tue Apr 03 20:34:39 +0000 2012

Fig 2.Check-in dataset

The dataset contains user id, venue id, venue category id, venue category, latitude, longitude, time zone offset, utc time stamp. In this activity is represented by the category to the POI such as gym or restaurant. In this Tokyo Check-in dataset it contains several missing values of user id, venue id, category id, venue category, latitude, longitude, time zone offset, utc time stamp. The dataset contains large missing values, so it can be replaced by NAN (Not a Number) values. By using numpy and pandas library package the dataset is extracted. Numpy is fundamental package for scientific computing with python. Pandas is used for data manipulation and analysis. Matplotlib is a python 2D plotting library that produces publication quality figures in a variety of hard copy formats and interactive environments across platforms.

CONCLUSION

To apply privacy preserving data publishing techniques within the case of social media , one immediate strategy is to modify user aspect before sent to social media . However such an approach is fantastic because it hinders key edges for users . In real world use case , Social media provides users with a social sharing platform , Wherever they will move with their friends by sharing their comments/ratings on things , blogs , photos , videos or perhaps their locations . As it is inappropriate to alter user public data before sent to social media , an alternative answer is to protect user privacy when releasing private data from social media to other third party service. To protect the user privacy when releasing the pubic data from any other third party services and to avoid the inference attack data masking technique is used.

REFERENCES

1. Hatem AbdulKader, Emad EIAbd, Waleed Ead , " Protecting online social networks profiles by hiding sensitive data attributes", Procedia Computer Science, vol.82, no.3, pp.20-27, 2016.

2. G. Zhao, X. Qian and X. Xie, "User-Service Rating Prediction by Exploring Social Users' Rating Behaviors," in IEEE Transactions on Multimedia, vol. 18, no. 3, pp. 496-506, 2016.

3. Hongdi Lina, Mengmeng Zhanga, Jinquan Zhanga, "Hybrid Filtrations Recommendation System based on Privacy Preserving in Edge Computing ", Procedia Computer Science, vol.129, no.2, pp.407-409, 2017.

4. Sánchez, David and Batet, Montserrat, "Toward sensitive document release with privacy guarantees ",

in Engineering Applications of Artificial Intelligence, vol.59, no.1, pp. 23–34, 2017.

5. S.Khater, D. Gračanin and H. G. Elmongui, "Personalized Recommendation for Online Social Networks Information: Personal Preferences and Location-Based Community Trends," in IEEE Transactions on Computational Social Systems, vol. 4, no. 3, pp. 104-120, 2017.