International Engineering Journal For Research & Development

Vol.4
Issue 5

E-ISSN NO:-2349-0721

Impact factor : 6.03

A NEW APPROACH FOR SOCIAL MEDIA NETWORK PRIVACY PROTECTION USING DATA MINING

Ms. P.S.Khorgade
PG Department of Computer Science
SGBAU,Amravati, India
lanushruti.raut@gmail.com

Dr . S. S. Sherekar
PG Department of Computer Science
SGBAU,Amravati , India
swatisherekar@gmail.com

Dr. V. M. Thakare
PG Department of Computer Science
SGBAU,Amravati, India
vilthakare@gmail.com

Abstract-

Online Social Networks (OSNs) have become an important part of daily digital interactions for more than half a billion users around the world. Privacy protection model uses type, frequency, and initiation factor of social interactions to calculate relationship strength. This model minimum real-life interaction patterns and makes online social networks more privacy friendly. Social media provided user with a social sharing platform, they can interact with their friends by intentionally sharing their comments /rating on items, block, photos, videos or real time locations. Users in the same social circle (group) have similar behavior, such as similar education background, hobbies, and similar privacy references.

The proposed approach to recommend privacy policies for newly uploaded data items or newly added contacts. And facilitate online social network users to group their contacts into social circles with common interests. The proposed social interaction-based audience segregation model for online social networks.

Keywords— ***Social media, Social network, Security, Privacy protection.***

I. INTRODUCTION

The Personalized recommendation is crucial to help users find pertinent information. This method proposes PrivRank, a customizable and continuous privacy-preserving Data publishing framework protect users against inference Attacks while enabling personalized ranking-based recommendation. As ranking-based recommendation algorithms mainly leverage the ranking of items for preference prediction, they are sensitive to the ranking loss Incurred from the data obfuscation process. PrivRank achieves both a better privacy protection and a higher utility in all the ranking-based recommendation use cases tested. [1]. general-purpose scheme is also proposed which is broadly applicable. Many advanced vector-based retrieval systems can be seamlessly used with the proposed approach. a "general purpose technique" which can be leveraged to improve the effectiveness of a wide variety of techniques. Popular techniques include Latent Semantic Indexing (LSI), Probabilistic Latent Semantic Indexing (PLSI), and Latent Dirichlet Allocation (LDA) [2]. In interest-based online social media networks, users can easily create and share personal content of interest, such as tweets, photos, music tracks, and videos. The topic-level influence by utilizing textual information and link information [3]. In many such applications, standard community detection methods, such as the Louvain method, info map, label propagation, or Newman's leading eigenvector, focus on the relationship connections between users, usually depending on links and friendship links. These techniques represent the network as a static structure that has stable links [4]. The socioscope model for analyzing the properties of social structures and human behavior; is also proposed, the socioscope consists of several components, including zoom, scale, and analysis tools, which are used for analyzing network structures, for discovering social groups and events, for quantifying relationships, and so on. The method is spreading activation-based technique to predict potential churners by examining a current set of churners and its underlying social network. [5].

Proposed method of this paper, social media protection network PrivRank, a customizable and continuous privacy-preserving Data publishing framework protect users against inference Attacks. The PrivRank achieves both a better privacy protection and a higher utility.

II. BACKGROUND

As per the studied analysis of a new approach for social media network privacy protection using data mining in recent past year. Such a approaches are:

Privacy-preserving publishing of user social media data is proposed by considering both the specific requirements of user privacy on social media and the data utility for enabling high-quality personalized recommendation. Online services such as e-commerce applications typically rely on a large collection of user data, particularly user activity data on social media, such as tagging/ rating records, comments, check-ins, or other types of user activity data. [1].

The proposed, geometric structure of MOs by their content links to capture their distribution in the underlying latent space. It is NP-hard to directly solve the optimization problem of determining the lowest rank approximation [2].

The hybrid threshold method combines the universal and SURE threshold methods. This method uses the universal threshold if the signal-to-noise ratio is low with a sparse wavelet coefficient [3].

In interest-based online social media networks, users can easily create and share personal content of interest, such as tweets, photos, music tracks, and videos [4].

The method also proposed an influence propagation model, in which the method has identified users with high frequency of interaction and determine their influence on neighboring users [5].

This paper is organized as follows:

Section I Introduction. **Section II** discusses Background. **Section III** discusses previous work. **Section IV** discusses existing methodologies. **Section V** discusses attributes and parameters and how these are affected on analysis of user behaviour **Section VI** proposed method and outcome result possible. Finally, **section VII** Conclude this review paper.

PREVIOUS WORK DONE

In research literature, many methods for analysis of for social media network protection using data mining a proposed:

Dingqi Yang et al. (2019) [1], has proposed Supervised learning that can be used to detect review spam by looking at it as the classification problem of separating reviews into two classes: spam and non-spam reviews.

Guo-Jun Qi et al. (2012) [2], has proposed semi-supervised learning to explore both labeled and unlabeled images in photo sharing websites while exploring the associated keywords in the text modality.

Huiqi Zhang et al (2011) [3], has proposed, a novelty detection approach based on the identification of sentence level patterns. Also proposed a probabilistic model to incorporate both content and time information in a unified framework. This approach gives new representations of both news articles and news events.

Quan Fang et al. (2014) [4], has proposed, the previous hypergraph learning methods incorporating different relations into one regularize for learning task, the proposed method utilizes different relations for different learning purposes. The social link relations to construct heterogeneous hyperedges for influence ranking in the network.

Noha Alduaiji et al. (2018) [5], has proposed the PageRank methods to solve the problem of loading balanced parts to processors. They define an active edge based on the edge plus its neighbours and call this a biased active edge. In the context of detecting temporal interaction-biased communities, they argue that weakly connected edges are important.

III. EXISTING METHODOLOGIES

There are many techniques that are implemented for providing a new approach for social media network privacy protection in data mining:

A. Probabilistic Historical Data Obfuscation Function Learning:

The activity-wise obfuscation function is learned based on individual activity. The optimal obfuscation function is learned based on user clusters still need to bridge the gap between clusters and users to obfuscate individual public data vectors.

Algorithm 2. Probabilistic Historical Data Obfuscation

Require: Obfuscation functions $p_{\hat{G}|G}$ for all possible Y

1: **for** $u \in \mathcal{U}$ **do**
2: Get u-specified private data Y
3: Get obfuscation function $p_{\hat{G}|G,Y}$ for Y
4: Get u's cluster G, where $u \in G$
5: Obfuscate the user's cluster G to $\hat{G}$ based on $p_{\hat{G}|G,Y}(\hat{G}|G)$
6: Randomly select a user $\hat{u}$ in cluster $\hat{G}$
7: Obfuscate $V^{\hat{u}}$ to V^u
8: **end for**

Algorithm: Probabilistic Historical Data Obfuscation Function Learning:

B. Context-and-Content Multimedia Retrieval:
This approach formulates multimedia retrieval by unifying the content and context-based approaches. Context and content links are explored in a unifying framework. the learned latent space ought to be more optimal than the other methods which separately mine these two kinds of links in MINs.

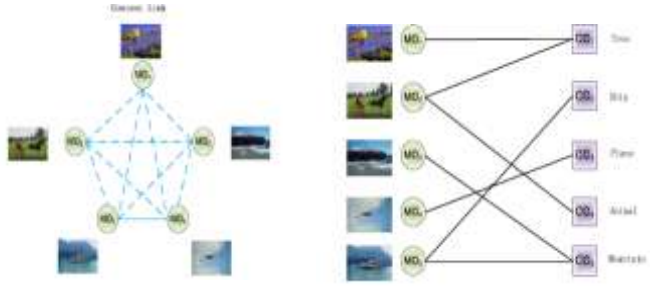


Figure 1: Context and Content Multimedia

C. Wavelet Denoising method:
The general wavelet-denoising procedure follows three steps: wavelet selection, threshold selection, and inverse wavelet transform (IWT)The wave Shrink method is widely used to estimate signal *x*. The two commonly used shrinkage functions are *hard* and *soft thresholding* functions defined as:

$$S_\delta^{\text{hard}}(x) = \begin{cases} x & |x| > \delta \\ 0 & |x| \le \delta \end{cases}$$

$$S_\delta^{\text{soft}}(x) = \begin{cases} x - \delta & x > \delta \\ \delta - x & x < -\delta \\ 0 & |x| < \delta \end{cases}$$

where $\delta \geq 0$ is the threshold.

D. Algorithm Topic-sensitive Influence Ranking:
The algorithm learned topic distribution and the constructed hypergraph, perform a topical affinity propagation on the hypergraph with the erogenous hyperedges. the influence of users and images is recursively updated until it achieves the optimal condition. Users share topical similarity which can be computed through the connected images and social links.

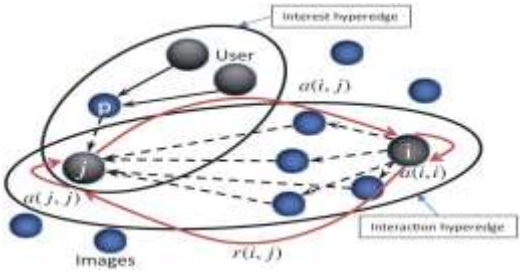


Figure 2: Topic-sensitive Influence Ranking

E. Algorithm 1 Pseudocode of the Partition Phase Require:
The algorithms for solving the problem of TIB communities in a massive graph. MapReduce is a well-known approach used to distribute function over processors to boost efficiency. Low-degree distribution and PageRank methods are also proposed to solve the problem of loading balanced parts to processors.

$$J(r) = \left[\sum (P^e - \bar{E})\right] + \left[\sum (P^v - \bar{V})\right]$$

Where, *Pe* is the number of edges in the processor *P*, *Pv* is the number of vertices.

ANALYSIS AND DISCUSSION

It is observed, that PrivRank still outperforms all baselines. Similar to the impact of the number of private attributes because a larger domain size implies less obfuscation budget on each value for the private attribute [1]. Author develop an analytical utility model that allows the SCserver to quantify the probability that a task request disseminated in a certain GR is accepted by a worker. The proposed dynamic WorkerPSD algorithm is to spend a fraction of the privacy budget to build the structure of the adaptive grid in the first-time instance. [2]. Author analyzed Apple company product – iPhone, and used SM-PIF to gather the opinionated resource to make a decision that to improve the performance and customer satisfaction of the year [3]. A novel idea implemented to develop a smart real time fear appeal system. Once the system is built, it can promote users to more risk mitigating behaviors on the social media sites [4]. Topic clustering methods to find the attention of users giving to various topics concerning Golden Week in China, which can obtain growth tendencies and geographic distributions of travelers [5].

Proposed Methods and Techniques	**Characteristics**	
	Advantages	**Disadvantages**
Privacy-Preserving Social Media Data Publishing of Personalized Ranking-Based Recommendation	-The PrivRank can continuously provide customized protection of user specified private data. -High-quality personalize ranking base recommendation to the user.	-Forced ranking can limit creativity, risk taking; it can weaken ethics and negatively impact the corporate culture.
Exploring Context and Content Links in Social Media: A Latent Space Method	-Easy implement, understand and use. There are many practical and scalable implementations available.	-The model is not humanly readable. Debug/evaluation is possible through finding similar words for each word in the latent space through. But otherwise not easy to interpret like, say LDA
Socioscope: Human Relationship and Behaviour Analysis in Social Networks	-To useful for homeland security and for detecting unwanted calls.	-The proposed scheme should also investigate new, non-numeric encoding domains, that is, categorical and alphanumeric attributes.
Topic-Sensitive Influencer Mining in Interest-Based Social Media Networks via Hypergraph Learning	-In interest-based online social media networks, users can easily create and share personal content of interest, such as tweets, photos, music tracks, and videos.	-The setup of a network security system can be a bit expensive.
Influence Propagation Model for Clique-Based Community Detection in Social Networks.	-Freely available to researchers and Twitter application developers. -good scalability on large graphs.	-The little memory consumption and showed a significant decrease in computation time, because the model runs on partitions and only computes the influence probability for the inactive edges in these partitions.

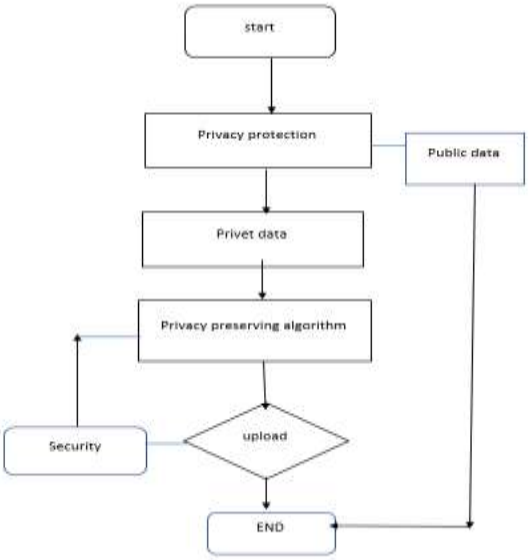
Table 1: Comparison Between Different Schemes

IV. PROPOSED METHODOLOGY

A new proposed method is used for privacy protection. The Online Social Networks (OSNs) have become an important part of daily digital interactions. In privacy preserving algorithm, personal data has made privacy preserving and data mining issue is an important one. General privacy preservation methods are committed to data protection at a lower privacy level. such as association rules mined in horizontally partitioned and vertically partitioned data, clustering mining, classification mining, and decision tree mining etc.
There are many methods of data mining for privacy protection.

Data distribution: Currently, some algorithms execute privacy protection data mining on a centralized data attribute values in different sites.

Data distortion: This method is to modify original data-base record before release, so as to achieve privacy protection purpose.



Flowchart: Privacy Preserving Algorithm using Data Mining

The proposed method interaction ranking base recommendation model. These interaction attributes are interaction quality, seriousness in interaction, a long period and common interest. In interest-based online social media networks, users can easily create and share personal content of interest, such as tweets, photos, music tracks, and videos. Historical data publishing a data paper about the datasets.

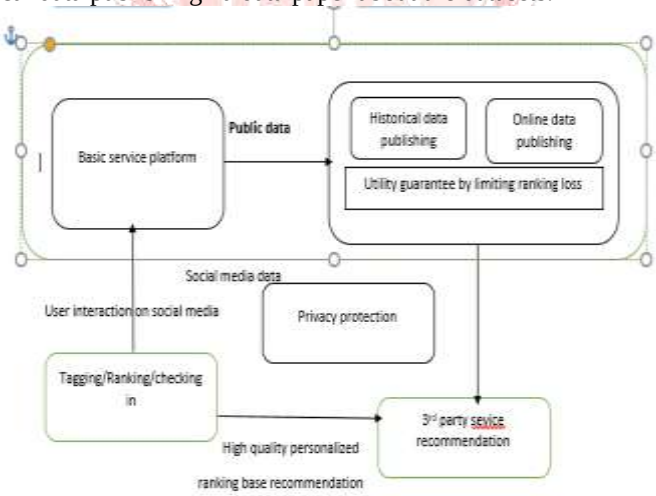


Figure 3: Block Diagram privacy protection Using Social Network

V.ACKNOWLEDMENT AND POSSIBLE RESULT

In this way the proposed method is performed for new approach for social media network privacy protection using data mining, the parameters that control the distortion budget for our method and baselines, and observe the resulting trade-off between privacy (AUC) and utility. The proposed algorithm is quite effective to integrate the content and context links for semantic retrieval over all 81 from Flickr data set.

CONCLUSION

This paper focused on the proposed scheme implement a PrivRank, a continuous privacy-preserving social media data publishing framework. the problem of topic-sensitive influencer mining (TSIM) in Flickr. TSIM aims to find the influential nodes in the networks and socioscope model for social network and human-behavior analysis based on mobile-phone call-detail records.

FUTURE SCOPE

From observations of the proposed method the future work will include investigate more information such as semantics, users' relation from other sources, other events, social-network dynamics, and evolution.

REFERENCES

Dingqi Yang, Binging Qu, and Philippe Cudr_e-Mauroux, "Privacy-Preserving Social Media Data Publishing of Personalized Ranking-Based Recommendation", *IEEE transactions on knowledge and data engineering*.Vol. 31, No. 3, Pg. No., 507-520 March 2019.

Guo-Jun Qi, Charu Aggarwal, Fellow, IEEE, Qi Tian, Senior Member, IEEE Heng Ji, Member, IEEE, and Thomas S. Huang, Life Fellow, IEEE," Exploring Context and Content Links in Social Media: A Latent Space Method*", IEEE transactions on pattern analysis and mechanic intelligence*, Issue 34 No. 05, Page No.: 850-862 Year: May 2012.

Huiqi Zhang, Ram Dantu, Member, IEEE, and João W. Cangussu, Member, IEEE," Socioscope: Human Relationship and Behavior Analysis in Social Networks*", IEEE transactions on systems man, and cybernetics—part a: systems and humans*, Vol, Issue 41 No. 06, Page No.: 1122-1143 Year: November 2011.

Quan Fang, Jitao Sang, Changsheng Xu, Fellow, IEEE, and Yong Rui, Fellow, IEEE.," Topic-Sensitive Influencer Mining in Interest-Based Social Media Networks via Hypergraph Learning", *IEEE transactions on multimedia* Vo, Issue: 16 No: 3 Page No.: 796-812 Year: April 2014.

Noha Alduaiji, Amitava Datta, and Jianxin Li, "Influence Propagation Model for Clique-Based Community Detection in Social Networks", *IEEE transactions a computational social system*. Vol. 05, No. 02, Pg. No., 563-575, June 2018.