Apr – Jun © 2019 Transactions; 6(22): 1209-1214

International Journal of Research and Applications

ISSN (online): 2349-0020 ISSN (print): 2394-4544 http://www.ijraonline.com/

Review Report

Privacy-preserving of social network data and application scenario with data mining

Shoban babu Sriramoju

Corresponding Author:
babuack@yahoo.com

DOI:
http://dx.doi.org/
10.17812/IJRA.6.22(3)2019

Manuscript:
Received: 26th Apr, 2019
Accepted: 15th May, 2019
Published: 15th Jun, 2019

Publisher:
Global Science Publishing Group, USA
http://www.globalsciencepg.org/

ABSTRACT

To handle and also evaluate side data check out organization chances deriving from the analytics of side information. Team up with the business to recognize existing side system as well as the potential use for information. It wrapped up from the findings that Business are still trying to find the best framework tools that will allow them to properly manage their big-data with their organization demands. These significant info can be obtained utilizing some data mining jobs. In short we can call big data as a ‟ property ‟ and also data mining is a „ handler ‟ that is made use of to give beneficial results. To carry out these evaluation data mining formulas can be utilized and likewise the big data techniques. This paper briefly explains about the privacy-preserving of social network data and application scenario with data mining.

Keywords: privacy preserving, data mining, big data.

Project Manager, Kenexcel Software Pvt. Ltd., India
Research Scholar, Kalinga University, India.

IJRA - Year of 2019 Transactions:
Month: April - June
Volume – 6, Issue – 22, Page No's:1209-1214
Subject Stream: Computers

Paper Communication: Author Direct

Paper Reference Id: IJRA-2019: 6(22)1209-1214

DOI: 10.17812/IJRA.6.22(3)2019

International Journal of Research and Applications
Apr - Jun © 2019 Transactions 6(22): 1209-1214

eISSN : 2349 – 0020
pISSN : 2394 – 4544
www.ijraonline.com

COMPUTERS REVIEW REPORT

Privacy-preserving of social network data and application scenario with data mining

Shoban babu Sriramoju

Project Manager, Kenexcel Software Pvt. Ltd., India
Research Scholar, Kalinga University, India.

ABSTRACT

To handle and also evaluate side data check out organization chances deriving from the analytics of side information. Team up with the business to recognize existing side system as well as the potential use for information. It wrapped up from the findings that Business are still trying to find the best framework tools that will allow them to properly manage their big-data with their organization demands. These significant info can be obtained utilizing some data mining jobs. In short we can call big data as a " property " and also data mining is a „ handler " that is made use of to give beneficial results. To carry out these evaluation data mining formulas can be utilized and likewise the big data techniques. This paper briefly explains about the privacy-preserving of social network data and application scenario with data mining.

Keywords: privacy preserving, data mining, big data.

1. INTRODUCTION

In data mining various Authors defined data accumulating from numerous data sources and additionally review uncompromising data explanation with numerous applications in real time. currently a day's traditional approaches making use of Data Mining And also Big Data, Artificial Intelligence Techniques. There are lots of future vital obstacles in Big Data monitoring and also analytics that develop from the nature of information: huge, diverse, and progressing.

These are several of the obstacles that researchers as well as practitioners will certainly have to deal throughout the next years:

Analytics Architecture: It is not clear yet exactly how an optimal design of analytics systems ought to be to manage historic information as well as with real-time information at the same time. An interesting proposal is the Lambda design of [4] The Lambda Design fixes the issue of calculating arbitrary features on approximate data in real-time by breaking down the issue right into three layers: the batch layer, the offering layer, and also the rate layer. It integrates in the exact same system Hadoop for the batch layer, and Storm for the speed layer. The residential or commercial properties of the system are: durable as well as mistake forgiving, scalable, basic, and also extensible, permits ad hoc questions, minimal upkeep, as well as debug gable.

Analytical Importance: It is essential to accomplish considerable analytical outcomes, and not be deceived by randomness.

Distributed Mining: Lots of data mining methods are not trivial to incapacitate. To have dispersed variations of some methods, a lot of research study is needed with useful as well as theoretical analysis to offer new techniques.

Time Evolving Information: Information might be evolving with time, so it is important that the Big Data mining strategies need to be able to adjust as well as in many cases to discover adjustment initially.

Compression: Managing Big Data, the amount of space needed to keep it is very relevant. There are 2 major methods: compression where we do not

lose anything or tasting where we select what is the data that is a lot more representative. Using compression, we might take even more time and much less room, so we can consider it as a makeover from time to space. Making use of sampling, we are losing info, but the gains precede might remain in orders of magnitude.

Visualization: A major job of Big Data evaluation is exactly how to picture the outcomes. As the information is so big, it is extremely tough to discover easy to use visualizations.

Hidden Big Data: Huge quantities of useful information are getting lost because new data is largely untagged file based and also disorganized information The 2012 IDC research study on Big Data [3] describes that in 2012, 23% (643 Exabyte) of the digital universe would certainly work for Big Data if marked as well as evaluated. Nevertheless, currently just 3% of the possibly beneficial data is marked, and even less is assessed.

DM is the process of exploration and evaluation, with normal or semi-automatic methods, of vast of amounts of details [1] it is made use of for checking out and also analyzing huge amount of data to locate patterns for Big Data. Recently all business as well as Marketing professionals gather all the info of transaction saved in a Large Information Base. it requires the performance of new strategies, new technology and also administration that are collectively being referred to as Huge Data [2] The main aim of Data Mining is either forecast or category, clustering. In forecast, to anticipate a worth Example. Service individuals or Atmosphere forecasting. In category is to specifically estimate the unbiased course made use of for each instance in the info. For e.g. leading student, Typical Student, Below Pupil. In clustering organizing the information..

2. USER ROLE-BASED METHODOLOGY

Present models and formulas suggested for PPDM mainly concentrate on exactly how to conceal those delicate info from specific mining procedures. Nonetheless, as portrayed in Fig, the whole KDD process entail multi-phase procedures.

Besides the mining phase, personal privacy problems may likewise emerge in the phase of information collecting or data preprocessing, even in the distribution procedure of the mining results. In this paper, we examine the personal privacy elements of data mining by thinking about the whole knowledge-discovery process.

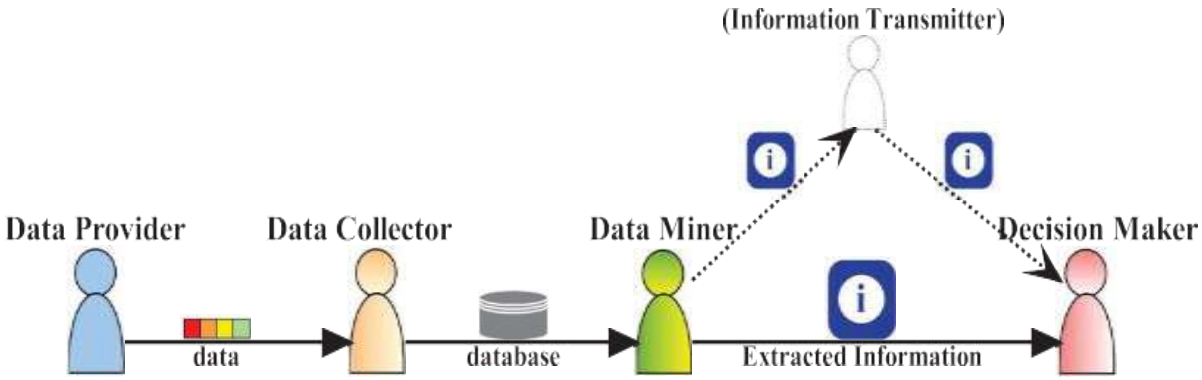
We offer a review of the many methods which can help to make proper use sensitive information as well as shield the safety and security of delicate details found by data mining. We use the term" sensitive information" to describe blessed or proprietary details that only specific individuals are permitted to see which is therefore not available to everyone. If delicate info is lost or made use of by any means apart from planned, the outcome can be severe damages to the person or company to which that details belongs. The term" sensitive information" refers to data from which sensitive info can be drawn out. Throughout the paper, we consider both terms" privacy" as well as" delicate details" are compatible.

In this paper, we develop a user-role based technique to carry out the review of relevant researches. Based upon the stage department in KDD process, we can recognize four various types of users, namely 4 customer duties, in a normal data mining circumstance (see Fig. 1).

Information Carrier: the individual who has some information that are preferred by the data mining task.
Information Enthusiast: the individual who collects data from information companies and then release the data to the data miner.
Data Miner: the customer that executes data mining tasks on the data.
Choice Manufacturer: the individual that makes decisions based on the data mining results in order to accomplish certain goals.

In the data mining scenario illustrated in Fig. 1, a customer stands for either an individual or an organization. Likewise, one customer can play multiple roles simultaneously. For example, in the Target tale we stated above, the client plays the role of data provider, as well as the store plays the duties of information collector, information miner and also decision maker.

Figure 1: A simple illustration of the application scenario with data mining at the core



3. CHALLENGING ISSUES WITH BIG DATA

Big data has actually been just one of the current and also future study problem. In the year 2014, Gartner noted „ Top ten Strategic Technologies patterns for 2013 ‟ as well as „ Top 10 crucial Technology Fads for the next 5 years ‟ and also big data is noted in both 2. [5]

Obstacles in big data are large. On one hand big data had lots of chances as well as on the other hand it is encountering lot of obstacles also.

When managing big data difficulties occurs in the following locations.
Information Capture as well as Storage Space.
Data Transmission.
Data Curation.
Data Evaluation.
Data Visualization.

According to [1] challenges of big data mining is normally separated into three rates.

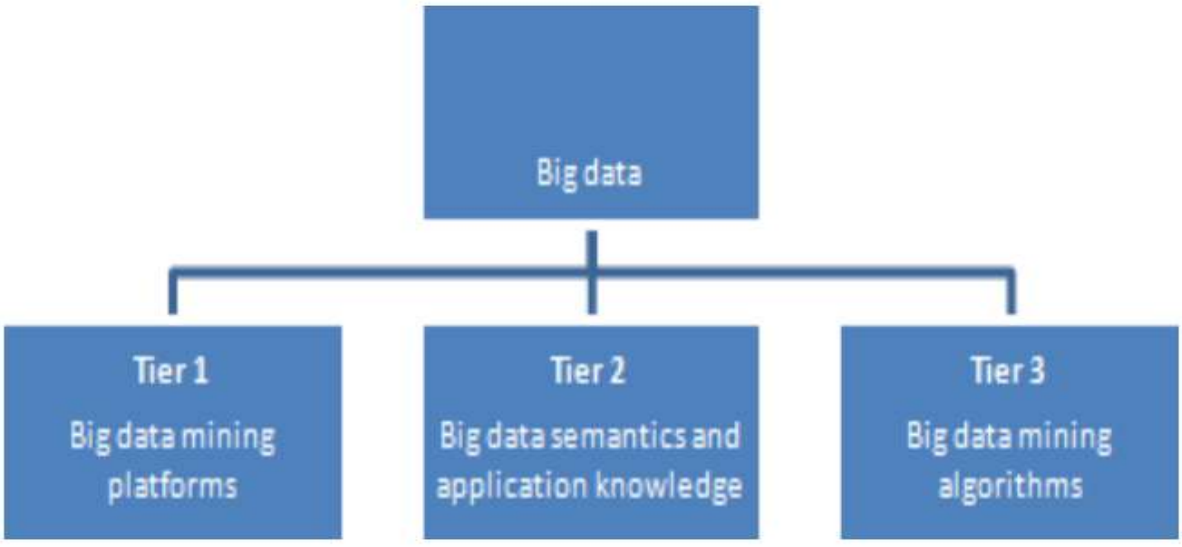


Figure 2: Phases of Big data challenges

The first tier includes the setup of data mining platforms.
The second one includes Information sharing and Data privacy, Domain and Application Knowledge.

The third one includes Local Learning and model fusion for multiple information sources. Mining from sparse, uncertain and incomplete data, Mining complex and dynamic data.

Generally mining of data from different data sources is tedious one as the data size is larger.

4. RESERVING PUBLISHING OF SOCIAL NETWORK DATA

Social networks have gained great development in recent years. Aiming at discovering interesting social patterns, social network analysis becomes more and more important. To support the analysis, the company who runs a social net- work application sometimes needs to publish its data to a third party. However, even if the truthful identifiers of individuals are removed from the published data, which is referred to as naïve anonymized, publication of the network data may lead to exposures of sensitive information about individuals, such as one's intimate relationships with others. Therefore, the network data need to be properly anonymized before they are published.

A social network is usually modeled as a graph, where the vertex represents an entity and the edge represents the relationship between two entities. Thus, PPDP in the context of social networks mainly deals with anonymizing graph data, which is much more challenging than anonymizing relational table data. [4] Have identified the following three challenges in social network data anonymization:

First, modeling adversary's background knowledge about the network is much harder. For relational data tables, a small set of quasi-identifiers are used to define the attack models. While given the network data, various information, such as attributes of an entity and relationships between different entities, may be utilized by the adversary.

Second, measuring the information loss in anonymizing social network data is harder than that in anonymizing relational data. It is difficult to determine whether the original network and the anonymized network are different in certain properties of the network.

Third, devising anonymization methods for social network data is much harder than that for relational data. Anonymizing a group of tuples in a relational table does not affect other tuples. However, when modifying a network, changing one vertex or edge may affect the rest of the network. Therefore, ''divide-and-conquer'' methods, which are widely applied to relational data, cannot be applied to network data.

To deal with above challenges, many approaches have been proposed. According to [5], anonymization methods on simple graphs, where vertices are not associated with attributes and edges have no labels, can be classified into three categories, namely edge modification, edge randomization, and clustering-based generalization. Comprehensive surveys of approaches to on social network data anonymization can be found.

5. ATTACK MODEL

Given the anonymized network data, adversaries usually rely on background knowledge to de-anonymize individuals and learn relationships between de-anonymized individuals. [4] Identify six types of the back- ground knowledge, i.e. attributes of vertices, vertex degrees, link relationship, neighborhoods, embedded subgraphs and graph metrics. [2] Propose an algorithm called *Seed-and-Grow* to identify users from an anonymized social graph, based solely on graph structure. The algorithm first identifies a seed sub-graph which is either planted by an attacker or divulged by collusion of a small group of users, and then grows the seed larger based on the adversary's existing knowledge of users' social relations. [6] Design a *structural attack* to de-anonymize social graph data.

The attack uses the cumulative degree of *N-hop* neighbors of a vertex as the regional feature, and combines it with the simulated annealing-based graph matching method to re-identify vertices in anonymous social graphs. [3] Introduce a relationship attack model called *mutual friend attack*, which is based on the number of mutual friends of two connected individuals. Fig. 3 shows an example of the mutual friend attack. The original social network *G* with vertex identities is shown in Fig. 3(a), and Fig. 3(b) shows the corresponding anonymized network where all individuals' names are removed. In this network, only Alice and Bob have 4 mutual friends. If an adversary knows this information, then he can uniquely re-identify the edge *(D, E)* in Fig. 3(b) is *(Alice, Bob)*. The *friendship attack* where an adversary utilizes the degrees of two vertices connected by an edge to re-identify related victims

in a published social network data set. Fig. 4 shows an example of friendship attack. Suppose that each user's friend count (i.e. the degree of the vertex) is publicly available. If the adversary knows that Bob has 2 friends and Carl has 4 friends, and he also knows that Bob and Carl are friends, then he can uniquely identify that the edge *(2, 3)* in Fig. 4(b) corresponds to *(Bob, Carl)*. In [3], another type of attack, namely *degree attack*, is explored. The motivation is that each individual in a social network is inclined to associate with not only a vertex identity but also a community identity, and the community identity reflects some sensitive information about the individual. It has been shown that, based on some background knowledge about vertex degree, even if the adversary cannot precisely identify the vertex corresponding to an individual, community information and neighborhood information can still be inferred.

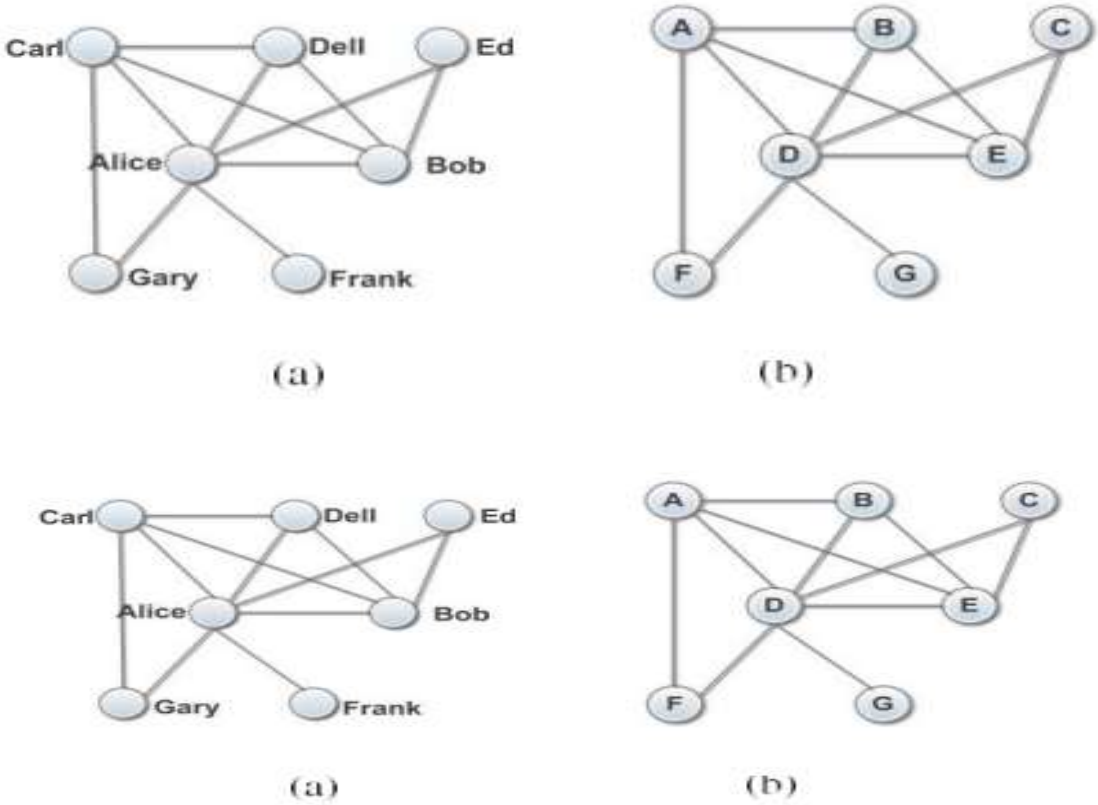


Figure 3. Example of mutual friend attack: (a) original network; (b) Naïve anonymized network.

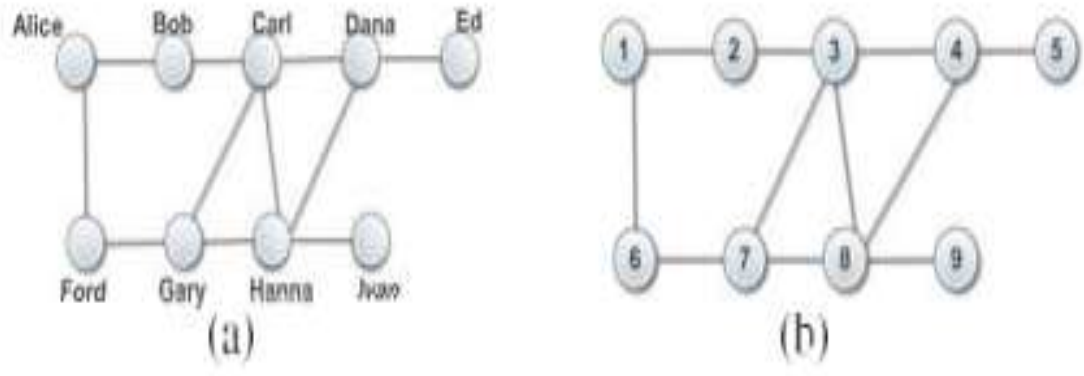


Figure 4. Example of friend attack: (a) original network; (b) naïve anonymized network.

CONCLUSION

For data provider, his privacy-preserving objective is to effectively control the amount of sensitive data revealed to others. To achieve this goal, he can utilize security tools to limit others access to his data, sell his data at auction to get enough compensations for privacy loss, or falsify his data to hide his true identity. This paper explained about the privacy-preserving of social network data and application scenario with data mining.

REFERENCES

1) Wei Fan, Albert Bifet, "Mining Big Data: Current Status and Forecast to Future," SIGKDD Exploration, vol. 14, Issue 2, 2013.
2) J. Gantz and D. Reinsel. IDC: The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East. Dec. 2012.
3) N. Marz, J. Warren. Big Data: Principles and best practices of scalable realtime data systems. Manning Publications, 2013.
4) Defending Networks with Incomplete Information: A Machine Learning Approach, BlackHat Briefings USA 2013 A Trend Micro White Paper | September 2012.
5) Mounika Reddy, Avula Deepak, Ekkati Kalyani Dharavath, Kranthi Gande, Shoban Sriramoju, "Risk-Aware Response Answer for Mitigating Painter Routing Attacks" in "International Journal of Information Technology and Management", Volume VI, Issue I, Feb 2014 [ISSN : 2249-4510].
6) Mounica Doosetty, Keerthi Kodakandla, Ashok R, Shoban Babu Sriramoju, "Extensive Secure Cloud Storage System Supporting Privacy-Preserving Public Auditing" in "International Journal of Information Technology and Management", Volume VI, Issue I, Feb 2012 [ISSN : 2249-4510].
7) Shoban Babu Sriramoju, "An Application for Annotating Web Search Results" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 2,Issue 3,March 2014 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798].
8) Shoban Babu Sriramoju, "Multi View Point Measure for Achieving Highest Intra-Cluster Similarity" in "International Journal of Innovative Research in Computer and Communication Engineering" Vol 2,Issue 3,March 2014 [ISSN(online) : 2320-9801, ISSN(print) : 2320-9798].
9) Shoban Babu Sriramoju, Madan Kumar Chandran, "UP-Growth Algorithms for Knowledge Discovery from Transactional Databases" in "International Journal of Advanced Research in Computer Science and Software Engineering", Vol 4, Issue 2, February 2014 [ISSN : 2277 128X]..
10) Shoban Babu Sriramoju, Azmera Chandu Naik, N.Samba Siva Rao, "Predicting The Misusability Of Data From Malicious Insiders" in "International Journal of Computer Engineering and Applications" Vol V, Issue II,Febrauary 2014 [ISSN : 2321-3469].
11) Ajay Babu Sriramoju, Dr. S. Shoban Babu, "Analysis on Image Compression Using Bit-Plane Separation Method" in "International Journal of Information Technology and Management", Vol VII, Issue X, November 2014 [ISSN: 2249-4510].