RESEARCH ARTICLE OPEN ACCESS

The Role of Mobile Network-Apps Security Using Network Techniques

S.Amutha
Lecturer in PG Dept.of.Computer Science & Application, Bharathidasan university constituent college for Women Orathanadu-614625

Abstract

Now-a-days, Apps based on Mobile Network Technology that are known as Mobile network Apps are very popular among all kinds of users. But, sufficient knowledge of users about the Apps is questionable. Many users don't know which application is useful for their smart phone and which is not and which apps exploit their ignorance etc. Exploiting the unawareness / ignorance of users, some fraudulent activities happened through mobile applications. In this situation, App designers use shady means, such as raise the App values or App ratings to commit fraud. For instance, Apple took action against App designers, who submitted misrepresentation in the Apple's App Store. App developers resolve to some fake ranks and comments to intentionally help their Apps for ranking on an App store. In this thesis, we propose that these fraudulent activities can be identified, controlled and conquered by an innovative method with the investigation of three types of evidences, i.e., ranking based evidences, rating based evidences, review based evidences and their historical records. These three types of evidences are used to create Evidence Aggregation Certificate. We evaluate the proposed system with real world App data collected from the iOS App store over a span of time. In this experiment, we validate the effectiveness of the proposed system and show the scalability of the detection algorithm as well as some regularity of ranking fraud activities.

Keywords **— Mobile apps, customer website.**

I. Introduction

The boom of Internet has given rise to an ever increasing amount of text data associated with multiple dimensions (attributes), for example, customer reviews in shopping websites (e.g., Amazon) are always associated with attributes like price, model, and rate. A traditional OnLine Analytical Processing (OLAP) data cube can be naturally extended to summarize and navigate structured data together with unstructured text data. Such a cube model is called text cube. A cell in the text cube aggregates a set of documents/tuples with matching attribute values in a subset of

dimensions. Keyword query, one of the most popular and easy-to-use ways to retrieve useful information from a collection of plain documents, is being extended to RDBMSs to retrieve information from text-rich attributes. Given a set of keywords, existing methods aim to find relevant tuples or joins of tuples (e.g., linked by foreign keys) that contain all or some of the keywords. Internet search engines have popularized keyword based search. Users submit keywords to the search engine and get a ranked list of documents as response. An alternative to keyword search is structured search where users direct their search by browsing classification hierarchies. Both models are tremendously valuable today.

1.1. Database support:

We are in the Information Age, which leads to power and success. We collect tremendous amounts of information by using sophisticated technologies such as computers, satellites etc. and storing all sorts of data in mass digital storage. Unfortunately, these massive collections of data stored on disparate structures very rapidly became overwhelming. This initial chaos has led to the creation of structured databases and Database Management Systems (DBMS). The efficient DBMS have been very important assets for management of large groups of data and especially for effective and efficient retrieval of required/particular information from a large collection. The proliferation of DBMS has also contributed to recent massive gathering of all sorts of information. Today, we have far more information than we can handle: from business transactions and scientific data, to satellite pictures, text reports and military intelligence. Information Retrieval (IR) is simply not enough anymore for decision-making. Confronted with huge collections of data, we have now created new needs to help us make better managerial choices. These needs are automatic summarization of data, extraction of the "essence" of information stored, and the discovery of patterns in raw data.

Find the application Applications in network security process:

Data Mining is the procedure of mining knowledge from data. The information or knowledge extracted so can be used for any of the following applications:

- Fraud Detection
- Customer Retention
- Production Control
- Science Exploration

- Market Analysis and Management
- Corporate Analysis & Risk Management

Apart from these, data mining can also be used in the areas of production control, customer retention, science exploration, sports, astrology, and Internet Web Surf-Aid.

Classification and Prediction in cyber security process:

Classification is the process of finding a model that describes the data classes or concepts. The purpose is to be able to use this model to predict the class of objects whose class label is unknown. This derived model is based on the analysis of sets of training data. The derived model can be presented in the following forms –

- Classification (IF-THEN) Rules
- Decision Trees
- Mathematical Formulae
- Neural Networks

The list of functions involved in these processes is as follows.

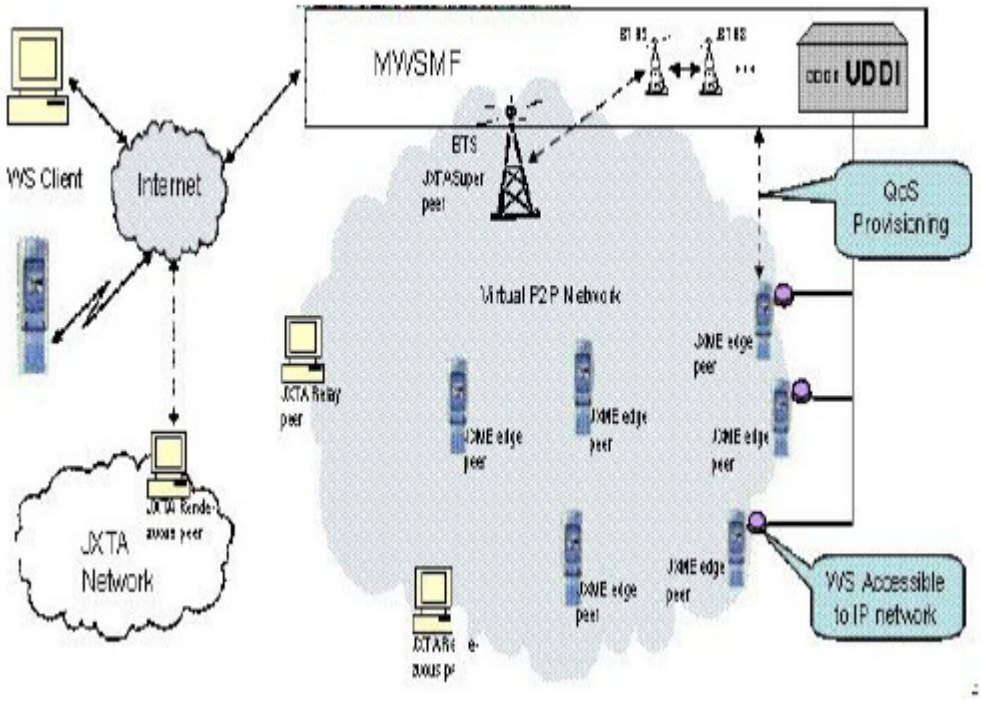
What are the network Classification Based on the Applications Adapted:

We can classify a data mining system according to the applications adapted. These applications are as follows –

- Finance
- Telecommunications
- DNA
- Stock Markets
- E-mail

Network analysis for problem description:

The fraudulent activities involved in promoting Apps can be identified and controlled by an empirical method which needs the evidences of three types: Ranking, Rating and Review. In order to understand the problem, let us discuss these methods of evidences briefly.



Ranking

In recent years, improvement in ubiquitous technologies and sensor networks has motivated the application of data mining techniques to network organized data. Network data describe entities represented by nodes, which may be connected with (related to) each other by edges. Many network datasets are characterized by a form of autocorrelation where the value of a variable at a given node depends on the values of variables at the nodes it is connected with. This phenomenon is a direct violation of the assumption that data are independently and identically distributed (i.i.d.). At the same time, it offers the unique opportunity to improve the performance of predictive models on network data, as inferences about one entity can be used to improve inferences about related entities. The learned models can be used to predict the ranking of nodes in the network for new time periods. SVM Rank algorithm emphasizes the importance of models learned in time periods during which data follow a data distribution that is similar to that observed in the new time period.

Rating

Internet Movie Database (IMDb), a free, user-maintained, online resource of production details for over 390,000 movies, television series and video games, which contains information such as title, genre, box-office taking, cast credits and user's ratings. In this case, a series of interesting facts and relationships using a variety of data mining techniques are gathered. In particular, the following are gathered: attributes relevant to the user ratings of Online products, such as discovering if big-budget laptops are more popular than their low budget counterparts, if any relationship between movies produced during the "golden age" (i.e. Citizen Kane, It's A Wonderful Life, etc.) can be proved, and whether any particular actors or actresses are likely to help a movie to succeed. The paper also reports on the techniques used, giving their implementation and usefulness. It is found that the IMDb is difficult to perform data mining upon, due to the format of the source data. It is also found some interesting facts, such as the budget of a film is no indication of how well-rated it will be, there is a downward trend in the quality of films over time, and the director and actors/actresses involved in a film are the most important factors to its success or lack thereof. The data used in this paper is not freely distributable, but remains copyright to the Internet Movie Database Inc. It is used here within the terms of their copying policy.

Review

In the Information Technology era information plays vital role in every sphere of the human life. It is very important to gather data from different data sources, store and maintain the data, generate information, generate knowledge and disseminate data, information and knowledge to every stakeholder. Due to vast use of computers and electronics devices and tremendous growth in computing power and storage capacity, there is explosive growth in data collection. The storing of the data in data warehouse enables entire enterprise to access a reliable current database. To analyse this vast amount of data and drawing fruitful conclusions and inferences it needs the special tools called data mining tools. This paper gives overview of the data mining systems and some of its applications.

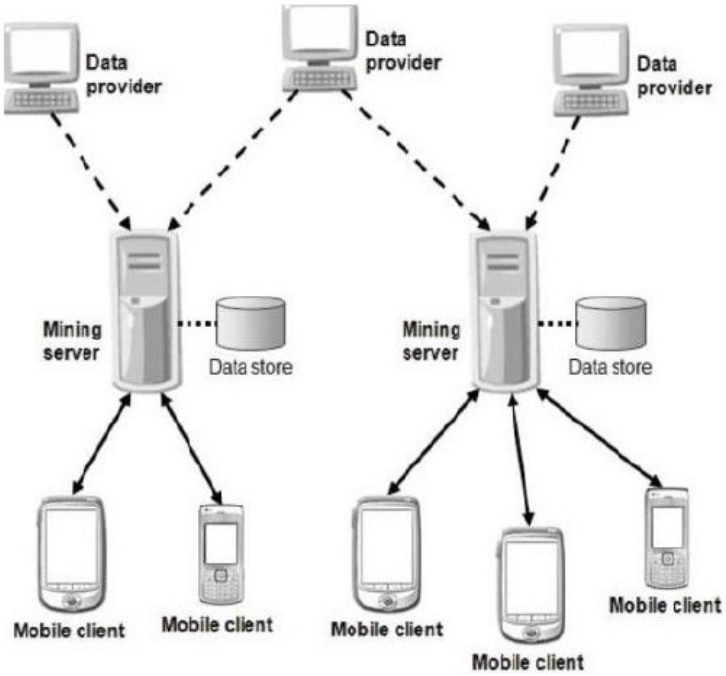
A report on Implementation process in paper

Identifying leading sessions

Ranking fraud usually happens in leading sessions. Therefore detecting ranking fraud mobile apps is actually to detect ranking fraud within leading sessions of mobile apps. Specifically, we first propose a simple yet effective algorithm identify the leading sessions of each app based on its historical ranking records. Then with the analysis of Apps ranking behaviours, we find that the fraudulent apps often have different ranking patterns in each leading session compared with normal Apps.

Mining leading sessions:

There are two main steps for mining leading sessions. First we need to discover leading events from the Apps historical ranking records. Second, we need to merge adjacent leading events for constructing leading sessions.



Ranking based evidences

A leading session is composed not several leading events. Therefore, we should first analyse the basis characteristics of leading events for extracting fraud evidences. By analysing the apps historical ranking records, we observe that apps ranking behaviours in a leading event always satisfy a specific ranking pattern, which consists of three different ranking phase, namely raising phase maintenance phase and recession phase. Specifically in each leading event an App's ranking firs increase to a peak position in the leader board(i.e. rising phase), then keeps such peak position

for a period (i.e. maintenance phase) and finally decrease till the end of the event (i.e. recession phase).

Rating based evidences

The ranking based evidences are useful for ranking fraud detection. However, sometimes it is not sufficient to only use ranking based evidences. Specifically after an app has been published it can be rated by any user who downloaded it. Indeed, user rating is one of the most important features of App development. An app which has higher rating may attract more users to download and can also be ranked higher in the leader board. Thus the rating manipulation is also an important perspective of ranking fraud. Intuitively, if an app has ranking fraud in a leading session the ratings during the time period of a may have anomaly patterns compared with its historical ratings, which can be used for constructing rating based evidences.

Review based evidences

Besides ratings most of the apps stores also allow users to write some textual comments as App reviews. Such reviews can reflect the personal perceptions and usage experiments of existing users for particular mobile apps. Indeed, review manipulation is one of the most important perspectives of App raking fraud. Specifically, before downloading or purchasing a new mobile app, users often firstly read its historical reviews to ease their decision making, and a mobile App contains more positive reviews may attract more users to download. Therefore, imposter's often fake reviews in the leading sessions of a specific App in order to inflate the App download and thus propel the apps ranking position in the leader board. Although some previous works on review spam detection have been reported in recent years, the problem of detecting the local anomaly of reviews in the leading sessions and capturing them as evidences for ranking fraud detection are still under- explored.

Evidence aggregation

After extracting three types of fraud evidences, the next challenge is how to combine them for ranking fraud detection. Indeed there are many ranking and evidences aggregation methods in the literature, such as permutation based models score based models and Dempster-shafer rules. However, some of these methods focus on learning a global ranking for all candidates. This is not proper for detecting ranking fraud for new Apps. Other methods are

based on supervised learning techniques which depend on the labeled training data and are hard to be explored. Instead, we propose an unsupervised approach based on fraud similarity to combine these evidences.

CONCLUSION AND FUTURE WORKS

We have proposed and studied the problem of keyword based top-k search in text cube (i.e., multidimensional text data). Flexible query language and relevance scoring formula are developed. Four efficient algorithms are designed for this problem. Among them, the most efficient one that network being process and get carry out the , search-space ordering approach, uses different search space pruning techniques for finding the top-k relevant cells by exploring only a small portion of the whole text cube (when k is small).

For future work, we would improve the support more types of queries, and different queries.

(i) it is interesting to compare the effectiveness of different scoring formulas experimentally in different real datasets;

(ii) it is also interesting to design efficient algorithms for different forms of relevance scoring formula

(iii) while our approaches work in a bottom-up manner (starting from rows or base cells), it is interesting to study whether there is any efficient top-down approach (e.g., starting from the 0-dim cell).

Reference:

1. Bolin Ding, Bo Zhao, Cindy Xide Lin, Jiawei Han, Fellow, “Efficient Keyword-Based Search for Top-K Cells in Text Cube”, *IEEE,* Chengxiang Zhai, Ashok Srivastava, *Senior Member, IEEE,* Nikunj C. Oza.

2. C. X. Lin, B. Ding, J. Han, F. Zhu, and B. Zhao, “Text cube: Computing ir measures for multidimensional text database analysis,” in Proc. IEEE Int’l Conf. Data Mining (ICDM), 2008, pp. 905–910.

3. S. Agrawal, S. Chaudhuri, and G. Das, "Dbxplorer: A system for keyword-based search over relational databases," in *Proc. IEEE Int'lConf. Data Eng. (ICDE)*, 2002, pp. 5–16.

4. G. Bhalotia, A. Hulgeri, C. Nakhe, S. Chakrabarti, and S. Sudarshan, "Keyword searching and browsing in databases using banks," in Proc. IEEE Int'l Conf. Data Eng. (ICDE), 2002, pp. 431–440.

5. B. Ding, J. X. Yu, S. Wang, L. Qin, X. Zhang, and X. Lin, "Finding top-k min-cost connected trees in databases," in Proc. IEEE Int' Conf. Data Eng. (ICDE), 2007, pp. 836–845.