



MOVIE TICKET BOOKING FRAUD DETECTION AND PREVENTION USING GOTCHA MODEL

¹Usha Nandini.D, ²Dr. R. Maruthi

¹M.Phil Research Scholar, ²Professor

Department of Computer Science

PRIST University, India

¹Sumithra6612@gmail.com, ²rmaruthi2014@gmail.com

ABSTRACT

In the last decade, the benefit of on-line payment has opened several new opportunities for e-commerce, lowering the geographical boundaries for retail. While e-commerce continues to be gaining quality, it's additionally the playground of fraudsters UN agency try to misuse the transparency of on-line purchases and also the transfer of master-card records. . We introduce GOTCHA!, a new approach on how to define and extract features from a time-weighted network, and how to exploit and integrate network-based and intrinsic features in fraud detection. The combination of all features (i.e., intrinsic and network features) is fed to the machine learning algorithms. This is the Gotcha! Model. As the creation of network features drastically increases the number of features to learn from, ensemble methods like Random Forest are used to train the models.

Keywords: OTB, FD, Gotcha, DOT NET.

1. INTRODUCTION

The incredible growth of the internet use for all sort of applications such as data production and storage, business transactions, professional, cultural and personal information management, etc. are pushing back the frontiers of traditional computer and digital data management. This overwhelming activity allows all kinds of players to propose new services and offers.

Unfortunately, some did not hesitate to take advantage of this space to be engaged in fraudulent activities, such as Identity Theft Fraud. The objective of this study is to work on a new way to address large scale ticket booking fraud detection by combining real-time processing and batch processing in data warehouse and Dot-Net Distributed File System (DDFS).

Fraud is often characterized by irregular concentration of activities on subsets of nodes in sub networks of the internet, particularly on online ticket booking (OTB).

This calls for linking data, which were not likely to be linked, because they do not belong to the same networks. Linking ticket booking data, spread upon different heterogeneous data repositories, calls for addressing several challenging problems such as algorithms optimization and parallelization, new knowledge representation paradigms for heterogeneous, redundant, noncertified or false information, association mechanisms, graph analysis for clustering and partitioning.

To address this multi-dimensional problem, we will adopt the following approach:

- 1) Identify community sub networks by using community detection algorithms running in a parallel environment,
- 2) Represent data and knowledge stored in these networks in a common knowledge scheme,

3) Apply iterative algorithms for clustering and partitioning.

The paper is organized as follows. We present in the second section, some of the main characteristics of OTB data, especially in the case of fraudulent activity. Then, we describe some recent works in different areas such as community detection in social networks, the analysis of large graphs, the clustering and partitioning of bi-partite graph and fraud detection.

Then we introduce the basis of our approach. In the third section, we present how we intend to develop our study, and how we are going to test the proposed solutions through experiments. In the last part, we will give some preliminary conclusions.

2. SOCIAL NETWORK ANALYSIS FOR FRAUD DETECTION

In the last decade, the use of social media websites in everybody's daily life is booming. People can continue their conversations on online social network sites like Facebook, Twitter, LinkedIn, Google+, Instagram, and so on and share their experiences with their acquaintances, friends, family, and others. It only takes one click to update your whereabouts to the rest of the world.

Plenty of options exist to broadcast your current activities: by picture, video, geo-location, links, or just plain text. You are on the top of the world—and everybody's watching. And this is where it becomes interesting. Users of online social network sites explicitly reveal their relationships with other people. As a consequence, social network sites are a (almost) perfect mapping of the relationships that exist in the real world. We know who you are, what your hobbies and interests are, to whom you are married, how many children you have, your buddies with whom you run every week, your friends at the wine club, etc.

This whole interconnected network of people knowing each other, somehow, is an extremely interesting source of information and knowledge. Marketing managers no longer have to guess who might influence whom to create the appropriate campaign. It is all there—and that is exactly the problem. Social network sites acknowledge the richness of the data sources they have, and are not willing to share them as such and free of cost. Moreover, those data are often privatized and regulated, and well-hidden from commercial use.

In this paper, we will briefly introduce the reader to networks and their applications in a fraud detection setting. One of the main questions answered in this chapter is how unstructured network information can be translated into useful and meaningful characteristics of a subject. We will analyze and extract features from the direct neighbourhood (i.e., the direct associates of a certain person or subject) as well as the network as a whole (i.e., collective inference). Those network-based features can serve as an enrichment of traditional data analysis techniques.

3. FRAUD PREVENTION

Since fraud is thus arduous to prove in courts, most organizations and people attempt to forestall fraud from happening by blanket measures. This includes limiting the quantity of harm the fraudster will impact on the organization moreover as early detection of fraud patterns. For instance, MasterCard corporations will cut the MasterCard limit across the board in anticipation of many negative fraud cases. Advertisers will forestall advertising campaigns with low variety of qualifying events.

These actions area unit typically in distinction with the corporate efforts to draw in additional customers and end in general discontentment. To the rescue area unit new technologies like Dot-Net, Influence Diagrams and theorem Networks that area unit computationally pricy (these area unit NP-hard in applied science terminology) however area unit additional correct and prophetic.

4. WHY DOT-NET?

Apache Dot-Net may be a distributed system for process giant amounts of knowledge. In an exceedingly recent Dot-Net Summit 2010 Yahoo, Facebook, and different corporations declared that they presently method many TBs of knowledge per day and also the volumes area unit growing at exponential rates. Dot-Net is very important for finding the fraud detection drawback because:

- Sampling doesn't work for rare events since the possibility of missing a fraud in fact case ends up in important deterioration of model quality.
- Dot-Net will solve abundant tougher issues by leverage multiple cores across thousands of machines and search through abundant larger drawback domains.
- Dot-Net is combined with different tools to manage moderate to low response latency needs.

Let's bear these reasons one by one. Sampling may be a common technique for modeling rare events. One amongst the issues with sampling is that we tend to cannot afford to throw away rare positive cases. Even in an exceedingly stratified or sampling theme one needs to retain all positive cases since the model accuracy heavily depends on them (one will typically discard some negative cases though). Given the on top of, the system still needs to bear the total dataset to sieve through the positive and negative cases.

Dot-Net is understood for its gnawing power. Nothing will compare with the output power of thousands of machines every of that has multiple cores. As was according recently at the Dot-Net Summit 2010, the most important installations of Dot-Net have two,000 to 4,000 computers with eight to twelve cores every, amounting to up to forty eight,000 active threads yearning for a pattern at constant time. this enables either (a) searching through larger periods of your time to include events across a bigger timeframe or (b) taking additional sources of data under consideration. it's quite common among social network corporations to comb through twitter blogs in search of relevant knowledge.

Finally, one amongst the fraud interference issues is latency. The agencies wish to react to an occurrence as shortly as attainable, typically inside many minutes of the event. Yahoo recently according that it will alter its activity model in an exceedingly response to a user click event inside 5-7 minutes across many hundred of innumerable customers and billions of events per day. Cloudera has developed a tool, Flume, which will load billions of events into HDFS inside many seconds and analyze them victimization Map Reduce.

Often fraud detection is like "finding a needle in an exceedingly haystack". One needs to bear mountains of relevant and on the face of it unsuitable info, build dependency models, appraise the impact and thwart the fraudster actions. Dot-Net helps with finding patterns by Process Mountains of data on thousands of cores in an exceedingly comparatively short quantity of your time.

5. GOTCHA

We introduce GOTCHA!, a new, generic, scalable, and integrated approach on however network analytics will improve the performance of ancient fraud detection tools during a ticket booking. We identify 5 challenges that concur with fraud; that is, fraud is associate degree uncommon, well-considered, time-evolving, carefully organized, and unnoticeably hid crime that appears in many alternative varieties and forms. Whereas current analysis fails to integrate of these dimensions into one encompassing approach, GOTCHA! Is that the 1st to address every of those challenges along in one high-performance, time-dependent detection technique.

In short, GOTCHA! Contributes to the fraud detection domain by proposing a completely unique approach on the way to spread fraud through a (i) time-weighted network and options extracted from a (ii) bipartite graph. We have a tendency to exploit dynamic network-based options area unit hidden (dashed line). Derived from the direct neighbourhood and develop a new propagation rule that infers associate degree initial exposure score for every node victimisation the complete network.

The exposure score measures the extent to that a node is influenced by dishonest nodes. We have a tendency to integrate each intrinsic and network-based option into one scalable algorithm. We have a tendency to argue that fraud may be a time-dependent phenomenon, and as a consequence, GOTCHA! Is designed specified a subject's characteristics and fraud probability will modification the over time.

6. RESULT AND ANALYSIS

By using the above logic, we studied the execution of fraud detection by theatres ticket booking. Here depending upon the GOTCHA method the unwanted processing charges can be eliminated. So that the user can be save the amount form online booking charges & other things. The results obtained are below

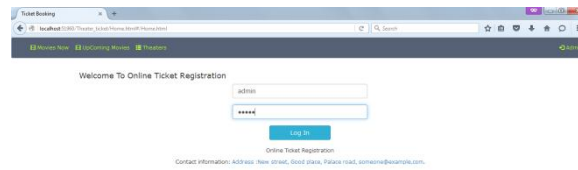


Fig 1: Admin Login

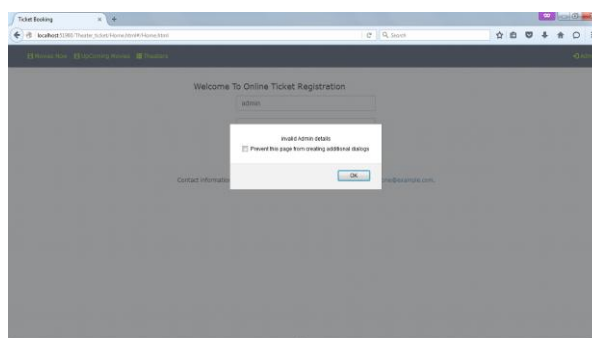


Fig 2: login ok

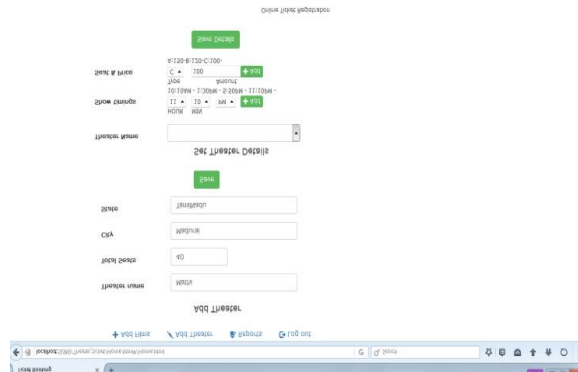


Fig 3: Admin movie entry details

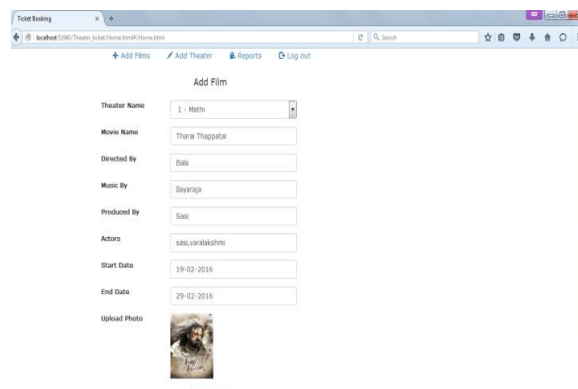


Fig 4: Admin movie entry final page

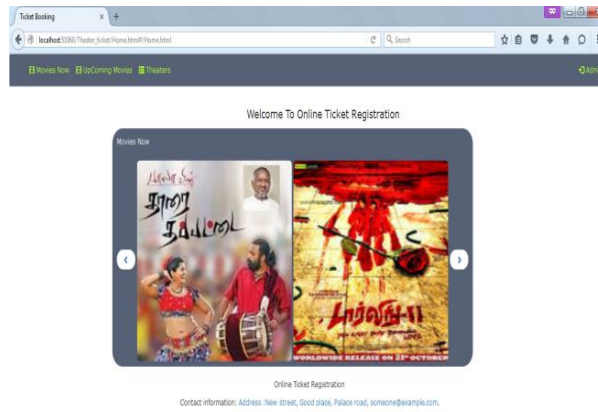


Fig 5: online movie display page

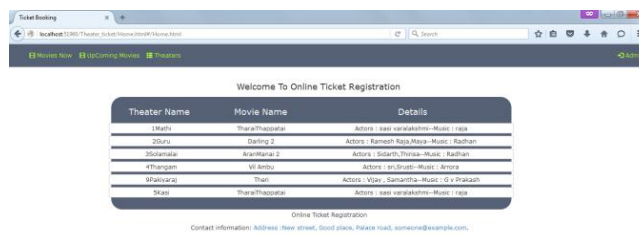


Fig 6: User need to select their option of theatre it will redirect to ticket booking page



Fig 7: Ticket booking page

6. CONCLUSION

In this paper, we tend to improve the performance of ancient classification techniques for ticket booking fraud detection by together with domain-driven network information mistreatment GOTCHA!, a brand new fraud detection approach. We tend to begin by distinguishing the challenges that concur with fraud and style GOTCHA! Specified it addresses every of those challenges to find future fraud. In this paper we have presented our motivations to study large scale social networks for characterizing communities. Our study will address the problems of linking information spread over several heterogeneous networks, algorithms parallelization and optimization for network analysis, and graph partitioning and clustering for structure extraction. We expect that this work will provide an answer to fraud detection.

7. FUTURE SCOPE

Presently, the work caters only to identify and classify the events into normal and attack. It can be extended to detect and classify the attacks into multiple attacks. Dynamic updating of the Anomaly Model using CNM (Clusset, Newman, and Moore) can also be considered for future enhancement.

8. REFERENCES

- [1] Armstrong, J. S. (2001). Selecting Forecasting Methods. In J.S. Armstrong, ed. Principles of Forecasting: A Handbook for Researchers and Practitioners. New York: Springer Science + Business Media, pp. 365–386.
- [2] Baesens, B. (2014). Analytics in a DOT NETWorld: The Essential Guide to Data Science and Its Applications. Hoboken, NJ: John Wiley & Sons.
- [3] Bolton, R. J., & Hand, D. J. (2002). Statistical Fraud Detection: A Review. *Statistical Science*, 17 (3): 235–249.
- [4] Caron, F., VandenBroucke, S., Vanthienen, J., & Baesens, B. (2013). Advanced Rule-Based Process Analytics: Applications for Risk Response Decisions and Management Control Activities. *Expert Systems with Applications*, Submitted.
- [5] Chakraborty, G., Murali, P., & Satish, G. (2013). Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS. Cary, NC: SAS Institute.
- [6] Cressey, D. R. (1953). *Other People's Money; A Study of the Social Psychology of Embezzlement*. New York: Free Press.
- [7] Duffield, G., & Grabosky, P. (2001). The Psychology of Fraud. In *Trends and Issues in Crime and Criminal Justice*, Australian Institute of Criminology (199).
- [8] Elder IV, J., & Thomas, H. (2012). *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*. New York: Academic Press.
- [9] Fawcett, T., & Provost, F. (1997). Adaptive Fraud Detection. *Data Mining and Knowledge Discovery* 1–3 (3): 291–316.
- [10] Grabosky, P., & Duffield, G. (2001). Red Flags of Fraud. *Trends and Issues in Crime and Criminal Justice*, Australian Institute of Criminology (200).
- [11] Han, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques*, Third Edition: Morgan Kaufmann.