International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)

A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.

Research Paper
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105

Volume 6, Issue 9
May 2018.
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Received on: 09.09.2018 Published on: 27.09.2018 Pages : 1 - 6

FRAUD DETECTION AND PREVENTION USING GOTCHA MODEL

USHA NANDINI.D
M.Phil Scholar, PRIST UNIVERSITY
Email: Sumithra6612@gmail.com.

Dr. R. MARUTHI
Professor of Computer science,PRIST UNIVERSITY
E-Mail: rmaruthi2014@gmail.com.

ABSTRACT

In the last decade, the benefit of online 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 online purchases and also the transfer of master-card records. We establish GOTCHA! A new method on how to describe and take out features from a time-weighted network, and how to develop and integrate network-based and intrinsic quality in fraud detection (FD). The grouping of all features (i.e., intrinsic and network features) is fed to the machine learning methods. This is the Gotcha! Model. As the formation of network features drastically increases the number of features to learn from, all together methods like Random Forest are used to train the models.

Keywords: OTB, FD, Gotcha, Dot-NetB(DN).

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 push to back the frontiers of conventional computer and digital data management. This tempting activity allows all kinds of players to suggest new services and offers.

Unluckily, some did not be uncertain to take benefit of this space to be busy in fraudulent actions, such as Identity Fraud. The aim of this study is to work on a new way to address large-scale ticket booking FD by combining real-time processing and batch processing in the data warehouse and Dot-Net Distributed File System (DDFS). Fraud is often characterized by an irregular concentration of activities on subsets of nodes in subnetworks of the internet, particularly on online ticket booking (OTB). This calls for connecting data, which were not probable to be correlated because they do not fit in to the same networks. Linking ticket booking data, spread upon different heterogeneous data repositories, calls for addressing several challenging problems such as techniques 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 technique:

1) Recognize neighbourhood subnetworks by using neighbourhood detection technique running in a similar environment,

2) Stands for data and information stored in these networks in a common information scheme,

3) Apply iterative techniques for cluster-ing and partition-ing.

The paper is prepared as follows. We present in the 2 section, some of the main description of OTB data, particularly in the case of fraudulent activity. Then, we explain some recent works in dissimilar areas such as neighbourhood discovery in social networks, the analysis of big graphs, the cluster and partition of the bi-partite graph and fraud detection. Then we introduce the basics of our technique. In the 3 section, we present how we intend to expand our study, and how we are going to test the projected solutions through experiment. In the final part, we will give conclusions.

2. SOCIAL NETWORK (SN) ANALYSIS FOR FRAUD DETECTION

In the last decade, the use of social media websites in everybody's day by day life is booming. People can continue their conversations on online SN 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 revise your position to the respite of the globe. Plenty of options exist to broadcast your current activities: by the 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 SN sites explicitly reveal their relationships with other people.

As a consequence, metallic element sites are associate (almost) excellent mapping of the relationships that exist within the universe. We all know UN agency you're, what your hobbies and interests are, to whom you're married, what number youngsters you've got, your buddies with whom you run each week, your friends at the wine club, etc.

This whole interconnected network of individuals knowing one another, somehow, is an especially fascinating supply and knowledge. Promoting managers now not need to guess UN agency would possibly influence whom to form the acceptable campaign. it's all there - and that's precisely the drawback

SN 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 FD 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 neighborhood (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 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 the 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 DN, Influence Diagrams and theorem Networks that area unit computationally pricy (these area units NP-hard in applied science terminology) however area unit additional correct and prophetic.

4. WHY DOT-NET?

Apache DN may be a distributed system for process giant amounts of knowledge. In an exceedingly recent DN 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. DN is very important for finding the FD 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.

- DN will solve abundant tougher issues by leverage multiple cores across thousands of machines and search through abundant larger drawback domains.
- DN 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.

DN 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 DN Summit 2010, the most important installations of DN 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 SN 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.

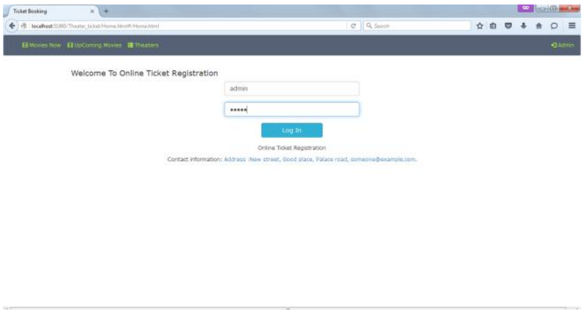
Often FDis like “finding a needle in an exceedingly haystack”. One needs to bear mountains of relevant and on the face of it unsuitable info, build dependency models, appraise the impact and thwart the fraudster actions. DN helps with finding patterns by Process Mountains of data on thousands of cores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

5. GOTCHA

We introduce GOTCHA!, a new, generic, scalable, and integrated technique on however network analytics will improve the performance of ancient FD 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 technique, 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 FD domain by proposing a completely unique technique 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 technique. 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 FD 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



Welcome To Online Ticket Registration

admin

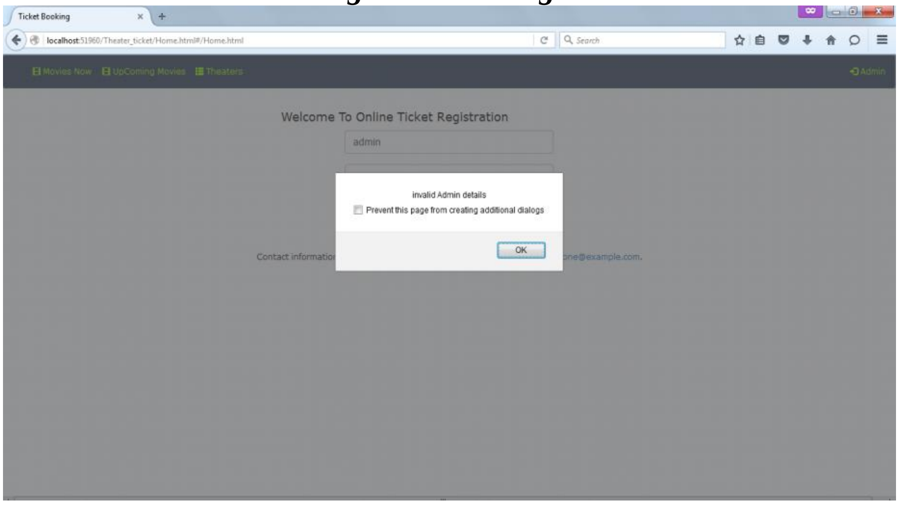
•••••

Log In

Online Ticket Registration

Contact information: Address :New street, Good state, Palace road, someone@example.com.

Fig 1: Admin Login



Welcome To Online Ticket Registration

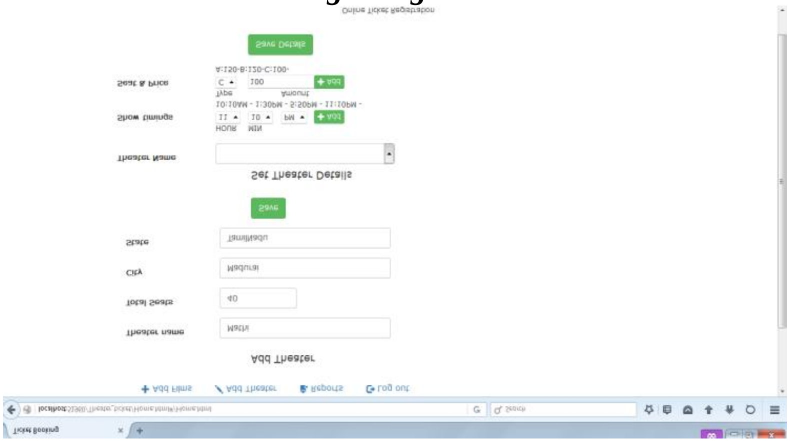
admin

invalid Admin details

□ Prevent this page from creating additional dialogs

OK

Fig 2: login ok



Add Theater

Theater name: Mathi

Total Seats: 40

City: Madurai

State: Tamilnadu

Save

Set Theater Details

Theater Name

Show timings: Hour 11 Min 10 PM Add

10:10AM - 1:30PM - 5:20PM - 11:10PM -

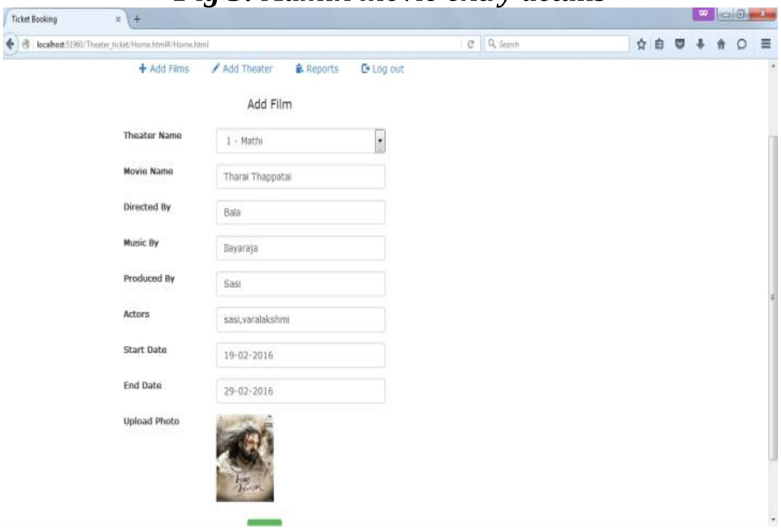
Seat & Price: Type C Amount 100 Add

A:150-B:120-C:100-

Save Details

Online Ticket Registration

Fig 3: Admin movie entry details



Add Film

Theater Name: 1 - Mathi

Movie Name: Tharai Thappatai

Directed By: Bala

Music By: Ilayaraja

Produced By: Sasi

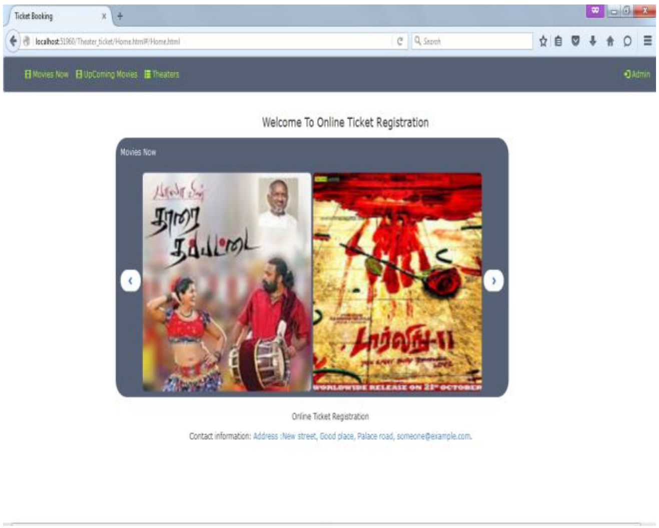
Actors: sasi,varalakshmi

Start Date: 19-02-2016

End Date: 29-02-2016

Upload Photo

Fig 4: Admin movie entry final page



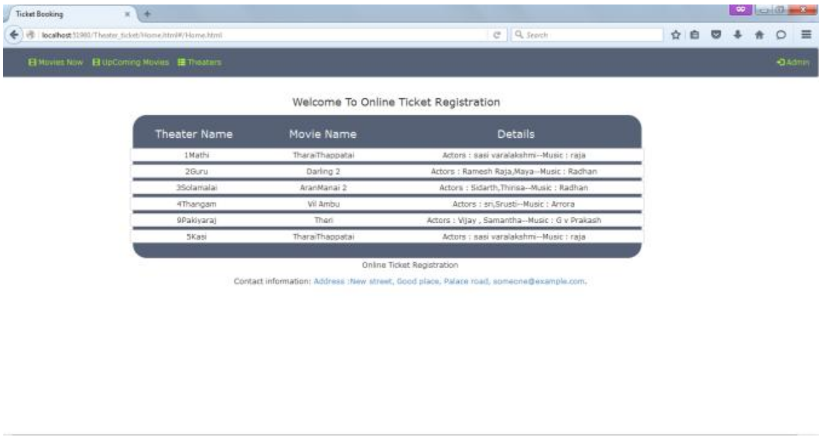
Welcome To Online Ticket Registration

Movies Now

Online Ticket Registration

Contact information: Address :New street, Good place, Palace road, someone@example.com.

Fig 5: online movie display page



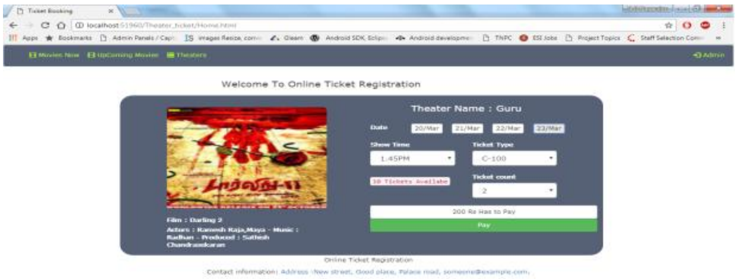
Welcome To Online Ticket Registration

Theater Name	Movie Name	Details
1Mathi	TharaiThappatai	Actors : sasi varalakshmi--Music : raja
2Guru	Darling 2	Actors : Ramesh Raja,Maya--Music : Radhan
3Solamalai	AranManai 2	Actors : Sidarth,Thirisa--Music : Radhan
4Thangam	Vil Ambu	Actors : sri,Srusti--Music : Arrora
9Pakiyaraj	Theri	Actors : Vijay , Samantha--Music : G v Prakash
5Kasi	TharaiThappatai	Actors : sasi varalakshmi--Music : raja

Online Ticket Registration

Contact information: Address :New street, Good place, Palace road, someone@example.com.

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



Welcome To Online Ticket Registration

Theater Name : Guru

Date 20/Mar 21/Mar 22/Mar 23/Mar

Show Time 1:45PM

Ticket Type C-100

10 Tickets Availabe

Ticket count 2

200 Rs Has to Pay

Pay

Film : Darling 2

Actors : Ramesh Raja,Maya - Music : Radhan - Produced : Sathish Chandrasekaran

Online Ticket Registration

Contact information: Address :New street, Good place, Palace road, someone@example.com.

Fig 7: Ticket booking page

6. CONCLUSION

In this paper, we tend to improve the performance of ancient classification methods for ticket booking FD by together with domain-driven network information maltreatment GOTCHA!, a trademark of new FD technique. We tend to begin by distinctive the challenges that correspond with fraud and style GOTCHA! Specific it addresses each of those challenges to find future fraud.

In this paper we have obtainable our motivation to learning big scale SN for characterizing communities. Our study will address the trouble of linking information spread over several various networks, techniques parallelization and optimization for network analysis, and graph partition and cluster for constitution extraction. We expect that this work will meet the expense of an answer to fraud detection.

7. FUTURE SCOPE

Currently, the work caters only to recognize and class the actions 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 (Clauset, Newman, and Moore) can also be considered for future enhancement.

8. REFERENCES

[1]. Armstrong, J. S. (2001). choosing prognostication strategies. In J.S. Armstrong, ed. Principles of Forecasting: A reference for Researchers and Practitioners.

[2]. Baesens, B. (2014). Analytics in an exceedingly DOT NETWorld: The Essential Guide to knowledge Science and Its Applications. Hoboken, NJ: John Wiley & Sons.

[3]. Bolton, R. J., & Hand, D. J. (2002). applied mathematics Fraud Detection: A Review.Statistical Science, seventeen (3): 235–249.

[4]. Caron, F., VandenBroucke, S., Vanthienen, J., &Baesens, B. (2013). Advanced Rule-Based method Analytics: Applications for Risk Response choices and internal control Activities. professional Systems with Applications, Submitted.

[5]. Chakraborty, G., Murali, P., & Satish, G. (2013). Text Mining and Analysis: sensible strategies, Examples, and Case Studies victimization SAS. Cary, NC: SAS Institute.

[6]. Cressey, D. R. (1953). different People’s Money; A Study of the psychology of misappropriation. New York: public press.

[7]. Duffield, G., &Grabosky, P. (2001). The psychological science of Fraud. In Trends and problems in Crime and Criminal Justice, Australian Institute of sociology (199).

[8]. Elder IV,, J., & Thomas, H. (2012). sensible Text Mining and applied mathematics Analysis for Non-Structured Text knowledge Applications. New York: tutorial Press.

[9]. Fawcett, T., & Provost, F. (1997). adaptational Fraud Detection. data processing and information Discovery 1–3 (3): 291–316.