DETECTION OF OPINION FRAUD USING TEMPORAL APPROACH

A
Dissertation
submitted in partial fulfilment of the
requirements for the award of degree of

Master of Technology
in
Computer Science and Engineering
by

Rachna Rani
16535030
M.Tech (CSE)

under the guidance of

Dr. Durga Toshniwal

Department of Computer Science and Engineering
Indian Institute of Technology Roorkee
Roorkee- 247667, India

MAY, 2018

DETECTION OF OPINION FRAUD USING TEMPORAL APPROACH

CANDIDATE'S DECLARATION

I hereby declare that the dissertation entitled “**Detection of Opinion Fraud Using Temporal Approach**” submitted by me in partial fulfilment of the requirements for the award of the Degree of Master of Technology in Computer Science and Engineering to the Department of Computer Science and Engineering, Indian Institute of Technology Roorkee is my original work carried during May 2017 to April 2018 under the guidance of Dr. Durga Toshniwal, Associate Professor, Department of Computer Science and Engineering, Indian Institute of Technology, Roorkee.

The content presented in this dissertation has not been submitted by me for award of any other degree of this or any other institute.

Date:

Place: Roorkee   Rachna Rani

DETECTION OF OPINION FRAUD USING TEMPORAL APPROACH

CERTIFICATE

This is to certify that the statement made by the candidate in the declaration is correct to the best of my knowledge and belief.

Date: Dr. Durga Toshniwal

Place: Roorkee (Associate Professor)

(Department of Computer Science & Engineering)

(Indian Institute of Technology Roorkee)

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude towards supervisor Dr. Durga Toshniwal for the continuous support, motivation, and guidance. I would also like to thank my lab-mates and friends for their support and help.

I am also grateful to the Department of Computer Science & Engineering of IIT Roorkee for providing valuable resources to aid my research.

Rachna Rani

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.1 Opinion Fraud

The social opinion for products and services advertised on online platforms such as Amazon, TripAdvisor or YouTube plays a vital role in shaping the decision of users. The products and services most liked by the users appear on top of the recommendation list of e-commerce site and garner maximum attention. Fraudsters manipulate the ratings, reviews or followers count to give undue promotion to some products and dishonest negative feedback to the others. Fraudsters adopt camouflaging and deceptive style of writing to evade Fraud Detection Systems.

Opinion Fraud is on the rise more than ever. A Daily Mail report [1] makes it evident how Trip Advisor has been crippled with fake reviews problem. Reviews are being sold for as cheap as $3. These solicitors are ordinary people who are willing to write honest looking reviews against minimal charges. Many hotel owners have taken help from these fraudsters to promote their hotels on Trip Advisor and malign their opponents' reputation. Thus, our recommendation systems, e commerce platforms and social networking sites are not equipped to catch fraudsters. The social opinion for products on e commerce sites influences the prospective buyers. Social platforms have now become an integral part of marketing strategies to attract buyers. Another case reported recently by CNBC [2] shows how e commerce is susceptible to online fraud. The sales of a company named Pure Daily Care selling skin care products on Amazon were demolished in a month by swarming their top 2 selling product with negative feedbacks. The fraudsters thus cheat the system by manipulating the ratings, inflating the popularity or demolishing that of their competitors. Such type of online fraud not only cause monetary loss to the user but also diminishes their trust in the social platforms. In order to protect the interests of people, the social sites must be equipped with techniques to eradicate any form of fraud and present interactions in their unadulterated form.

1.2 Problem Description

The problem of opinion fraud is widespread in online social networks. Fake profiles and followers on Twitter, rating manipulation on Google Play & Amazon, Fake subscribers on YouTube, fake postings on Yelp & Trip Advisor are a few examples. In simple terms, we undertake the problem to identify fraudulent reviewers and in turn fake reviews in an online recommendation system. The online recommendation system consists of a set of actors (customers), a set of targets (products) and the reviews. Customer is either spammer or benign and review is either genuine or fake. Each actor writes a review for a particular target which is often quantified as an integer between 1 and 5.

The subset of the customers that write reviews directed towards promoting or harming reputation of a product are opinion fraudsters. However, fraudsters could also write some genuine reviews to merge into category of benign customers and camouflage themselves.

Problem. Given data of all review tuples R, the model needs to detect opinion fraudsters, i.e. the task is to find a prediction function f that labels customers:

$$f : C \rightarrow \{\text{fraudster, genuine}\}.$$

1.3 Organization of the Report

The thesis report aims at finding the feature for characterizing spam and depicting how the feature plays role in distinguishing fraud apart. The rest of the report is organized as follows: literature review, proposed framework, experiments & discussion, results, conclusion and future work.

CHAPTER 2. LITERATURE REVIEW

2.1 Opinion Spam Detection Algorithm

The first work on Opinion Fraud was carried out by Jindal and Liu [3]. They (a) used a linear classifier on statistical logistic regression model which was built on behavioural and textual features, (b) thoroughly examined the vast review set obtained from Amazon to obtain 12 features and categorized reviews into Untruthful, Brand-targeted and Non-reviews, and (c) trained their model using synthetic training set and modified their classifier to fit for each type of review. Their work was remarkable but had shortcomings like targeting only a subset of opinion spams- Duplicate Opinions whereas fraudsters have evolved to adulterate review sites with Deceptive Opinions. Moreover, supervised algorithms for fraud detection are not adaptive to detect new frauds. The practice followed in [4] by Li et al. of manually annotating crawled review dataset from Epinions to identify spam features is not efficient as humans are prone to misclassify reviews. Ott et al. in [5] used n-grams and psycholinguistic features to create a classifier with crowd sourced gold standard dataset produced by Amazon Mechanical Turk but detection can be evaded by avoiding using the clue words that the classifier look for. In another work [6] Jindal focused on studying patterns suggestive of fraudsters in reviewing behaviour of certain reviewers. Mukherjee et al. [7] targets on detecting fraudster groups and proposed several new group oriented features. Fei et al. [10] explores how the burstiness of reviews could be related to spam activity. Their findings show better results in detecting fraudster groups but fails to consider that review time is dictated by product type, its usage pattern(more in one season) , etc. and hence the bins created for reviews according to time could give more accuracy if these factors were considered. Lim et al. in [11] showed that fraudster reviews are majorly inclined towards certain product groups. Their rating pattern towards that group can be suggestive of fraud, i.e. if a reviewer gave multiple positive reviews or multiple negative reviews to a particular product category and has no or less reviews written for other categories then the reviewer is in an attempt to steer the ratings. Another important feature used is Early Deviation, the fraudsters tend to review a product early in their operational time to garner attention and drive sales.

2.2 Time Series Anomaly Detection Algorithm

There are many applications where temporal data is generated like stock markets, astronomical data, census data, medical data, etc. The enormous size of the data poses great challenges. The first class of time-series outlier detection algorithm uses modeling and forecasting models (e.g. ARIMA [8], Exponential Smoothing [9], State Space Models [12], etc.) to detect deviation from normal behavior and report outliers. Another common strategy for detecting change points in the literature is to move two side-by-side windows on the time-series and compute the difference between the behavior of the time-series in the two windows as a measure of the deviation metric [15, 16, 13, 14]. The behavior of the time-series in each window is typically modeled by the distribution of the values, motifs, frequencies, etc. that are present in the time-series.

2.3 Research Gaps

1. Several different approaches have been employed to identify fraud using Supervised and Unsupervised learning. There is a scope to attempt detection of fraudster using temporal approach.
2. There has been a great deal of research work in the domain of fraud detection with each methodology separately. There is scope of adopting a hybrid approach in the domain of Opinion Fraud.

CHAPTER 3. PROPPOSED FRAMEWORK

3.1 Motivation

There are two strategies to curb fraud- detecting presence of fraud and predicting occurrence of fraud. The existing work in the domain of Opinion Fraud does not adopt Pre Detection method, though it has been widely explored in other types of fraud like credit card fraud. This thesis takes up to explore the early detection in the context of fake opinions posted online. The features indicative of fraud must be explored to categorize users as normal and abnormal. However features that distinguish fraud are dependent on the domain. Primarily, features can be categorized as graph based, behavioural and text based. Often review text is a good measure of trustworthiness of the customer. However, fraudsters are expert in writing deceptive custom made for the product at hand. We determine several features that characterize fraud.

3.2 Fraud Detection

A time series is defined for every user. The time series of fraudsters contains timestamps with abnormal feature values. Since the properties of anomalous timestamp stand out of the rest, these anomalous timestamps can be determined by exploiting their relationship with other timestamps. So, the timestamp with least similarity to all timestamps gets the minimum score. If the score is lower than a defined threshold, the time series is categorized as anomalous and customer as fraudulent.

Once label of each customer has been determined, model must be learned for early detection of fraud. In order to detect the fraudsters early in time their operational dynamics must be studied. The fraudsters work in congruence to steer the ratings of the target product group. The operational period of fraudsters called Active period is determined.

Since fraudsters accomplish their tasks and then vanish from the face of social setting, the detection time of fraud can be greatly reduced if the feature value is monitored for intervals equal to active period. Any deviation in those values is then reported. These feature values are indicative of spamming activity and are stronger in active period. Thus, fraudsters can be spotted in the earliest time possible in this way.

CHAPTER 4. EXPERIMENTS AND DISCUSSION

4.1 Dataset Used

The data spanning between May 1996 and July 2014 is segregated into different categories with large number of reviews for products. The enormous dataset used allowed better determination of the features and better credibility of experiment results. Each line of data consists of the following 7 parts:<reviewer id> <product id> <date> <number of helpful votes> <rating> <review text><review summary>.

Review pre-processing: The data also contains users which have less than 3 reviews, such users were removed. Some repeating products were removed.

Category	**Products Count**	**Reviews Count**
Clothing	1,503,384	5,748,920
Cell Phones	346,793	3,447,249

Table 1- Dataset Used

4.2 Results and Analysis

The proposed work was run on dataset. The features were computed and fraudsters detected. The result of the algorithm is summarized in the table given below.

Category	Reviewers Count	Fraudsters Count
Clothing	3117268	56
Cell Phones	2261045	173

Table 2- Fraudsters detected in each category

The fraudsters detected are clustered using the proposed method. The following Time series diagram shows the clustering of similar time series (i.e. coordinating fraudsters).

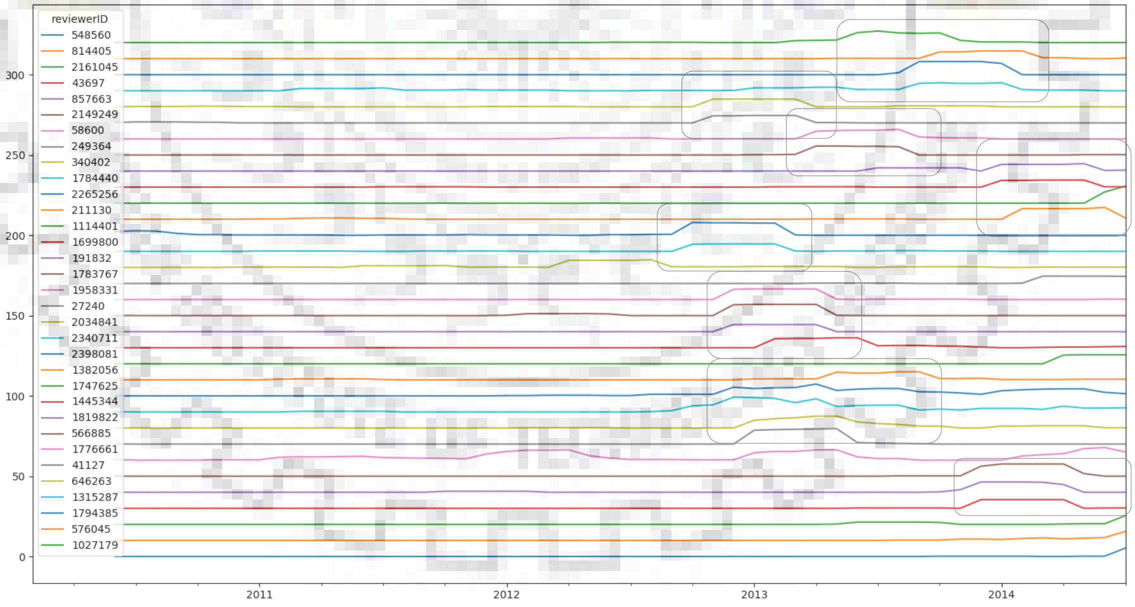

Figure 1 – Coordinating Behavior of fraudster group

With the knowledge of fraudster groups, the active period can be computed. Following graphs show the distribution of active period of different groups in different categories.

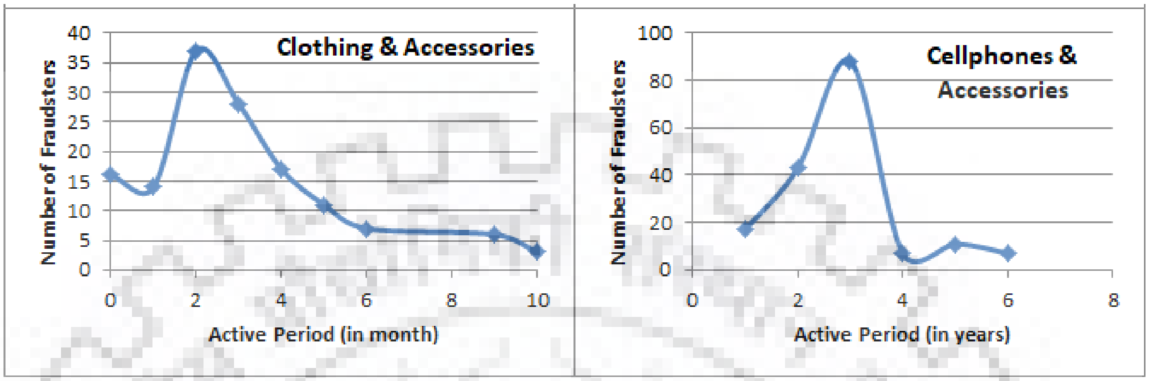

Figure 2- Distribution of Active Period in different Product Category

In every graph there are large number of fraudsters with short active period and small number of fraudsters with long active period.

Product Category	**Active period of Opinion Fraudster**	**Products Reviewed**	**Positive Rating**	**Negative Rating**
Clothing	2 months	217	99.05 %	0.95%
Cell Phones	3 years	4109	96.41 %	3.59 %

Table 3 - Average Active Period of Fraudster

The next part of thesis work deals with developing a model for distinguishing between normal and abnormal behavior. The plot below detects the change point in feature. It shows the distributions for normal customers in blue and fraudulent customers in orange.

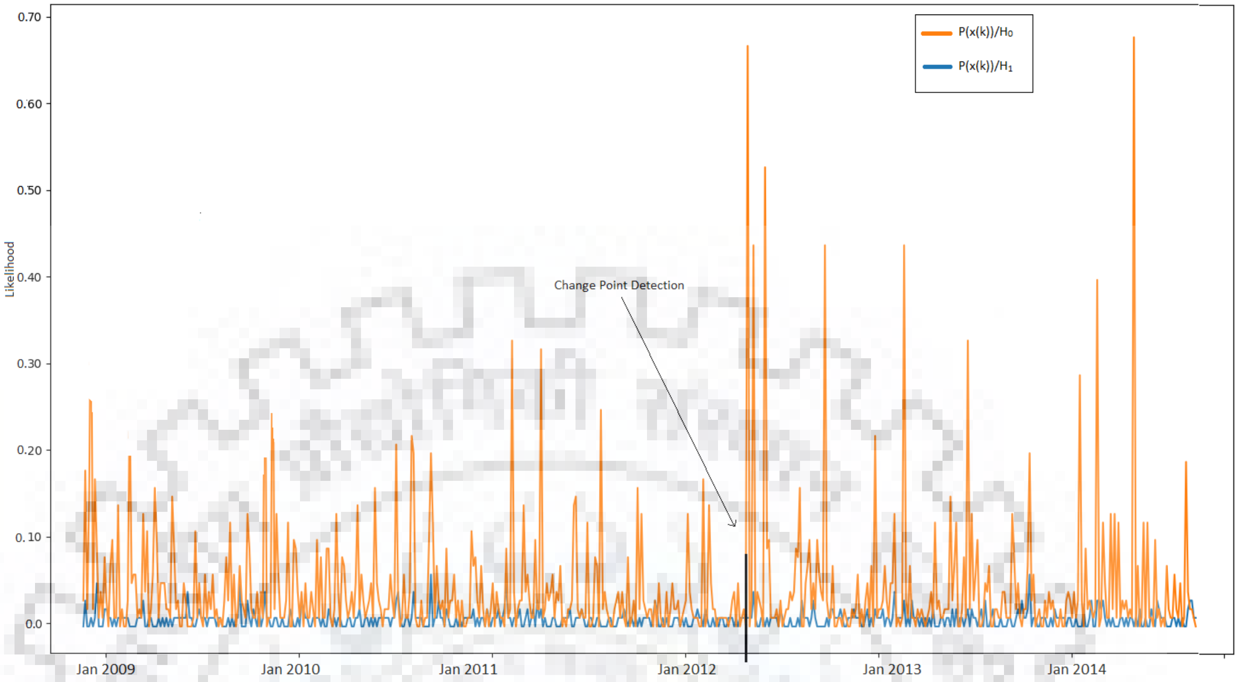

Figure 3 – Normal and Abnormal customer Model

CHAPTER 5. CONCLUSION & FUTURE WORK

Opinion Fraud is nuisance to social platforms. It undermines the recommendation system of e commerce sites where sales are driven by peer recommendations. It is difficult to prevent opinion fraud and so focus must be towards detecting and eradicating the malignant entities. In this work, we demonstrated that fraudsters are not lasting member of social platform; they often disappear after exploiting the misled buyers.

We developed a fraud detection system. The model utilizes features to define a score. This score denotes the dishonesty of the user. Once the fraudulent users are captured, the other coordinating members playing along are also identified. The knowledge of different group of fraudster presents the opportunity to learn the operational dynamics of the fraudster. With this framework, the fraudsters are caught successfully in earliest possible time.

REFERENCES

[1] T. Rawstorne, 'Disturbing proof the online review that made you book your holiday may be FAKE: Investigation reveals an entire industry is dedicated to generating bogus appraisals for cash', DailyMail, 2015.

[2] A. Levy, ‘This Amazon seller lost $400,000 in sales after being attacked by self-proclaimed 'virus of Amazon',CNBC,2017

[3] Nitin Jindal and Bing Liu. 2008. Opinion spam and analysis. In Proceedings of the 2008 International Conference on Web Search and Data Mining (WSDM '08). ACM, New York, NY, USA, 219-230.

[4] Fangtao Li, Minlie Huang, Yi Yang, and Xiaoyan Zhu. 2011. Learning to identify review spam. In Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume Three (IJCAI'11), Toby Walsh (Ed.), Vol. Volume Three. AAAI Press 2488-2493.

[5] Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T. Hancock. 2011. Finding deceptive opinion spam by any stretch of the imagination. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1 (HLT '11), Vol. 1. Association for Computational Linguistics, Stroudsburg, PA, USA, 309-319.

[6] Nitin Jindal, Bing Liu, and Ee-Peng Lim. 2010. Finding unusual review patterns using unexpected rules. In Proceedings of the 19th ACM international conference

on Information and knowledge management (CIKM '10). ACM, New York, NY, USA, 1549-1552.

[7] Arjun Mukherjee, Bing Liu, and Natalie Glance. 2012. Spotting fake reviewer groups in consumer reviews. In Proceedings of the 21st international conference on World Wide Web (WWW '12). ACM, New York, NY, USA, 191-200.

[8] W. W.-S. Wei. Time series analysis. Addison-Wesley publ, 1994

[9] R. H. Jones. Exponential smoothing for multivariate time series. Journal of the Royal Statistical Society. Series B (Methodological), pages 241–251, 1966.

[10] Fei, G & Mukherjee, A & Liu, B & Hsu, M & Castellanos, M & Ghosh, R. (2013). Exploiting burstiness in reviews for review spammer detection. Proceedings of the 7th International Conference on Weblogs and Social Media, ICWSM 2013. 175-184.

[11] Ee-Peng Lim, Viet-An Nguyen, Nitin Jindal, Bing Liu, and Hady Wirawan Lauw. 2010. Detecting product review spammers using rating behaviors. In Proceedings of the 19th ACM international conference on Information and knowledge management (CIKM '10). ACM, New York, NY, USA, 939-948.

[12] J. Durbin and S. J. Koopman. Time series analysis by state space methods. Number 38. Oxford University Press, 2012.

[13] S. Liu, M. Yamada, N. Collier, and M. Sugiyama. Change-point detection in time-series data by relative density-ratio estimation. Neural Networks, 43:72–83, 2013.

[14] B. K. Ray and R. S. Tsay. Bayesian methods for change-point detection in long-range dependent processes. Journal of Time Series Analysis, 23(6):687–705, 2002.

[15] Y. Kawahara, T. Yairi, and K. Machida. Change-point detection in time-series data based on subspace identification. In ICDM, pages 559–564. IEEE, 2007

[16] Y. Xie, J. Huang, and R. Willett. Change-point detection for high-dimensional time series with missing data. Selected Topics in Signal Processing, IEEE Journal of, 7(1):12–27, 2013.