International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 7 Issue 11 Ver II || Nov 2018 || PP 68-71

Anomaly DetectionOn Images Using Machine Learning Algorithm

Urmi Priyadarshani Das,Ms.N.Deepika
M.Tech CSE, New Horizon College Engg, Bangalore
Asst Prof, CSE Dept., New Horizon College Engg, Bangalore
Corresponding Author:Urmi Priya Darshani Das

Abstract: *Now a days OSN's (Online Social Network) addiction of users towards the networks which has many users, group of people. News making event causes rise in user activity in OSN.Facebook is the widely used social network which is the interest of the malicious attackers. Huge financial and reputation loss caused by massive malicious activities in the OSN by sending spam messages, fake applications. Malicious activities caused by larger part of the fake Facebook accounts. In Facebook other user's posts/reaction influence users.Anomalous behaviour of the corresponding account can be observed by the change in the behaviour and reaction. Anomalous activities in the image further classified by applying ensemble machine learning algorithm using decision tree .Data set on the features of the images used to detect the anomalous activities of the images of the Facebook. Change in the features of the images determine malicious activity in the image.*
Keywords*: Anamoly Detection, Facebook, Online Social Networks, Ensemble algorithm.*

Date of Submission: 19-11-2018 Date of acceptance: 04-12-2018

I. Introduction

Modern society is use OSN (Online Social Network) such as Facebook, twitter in all sphere of life. Among all the OSN's Facebook is the widely used by the users. Number of active user s escalation is more as compared to other OSN.The number of Facebook users in 2008 was 100 million and 2017 it has reached to 1968million and half of the Facebook users have 200 friends[2][21].Data theft, spreading fake news, spreading spam messages are the malicious activities occurs in the OSN's like facebook.Anonymization, deanonymization,Sybilattack, social engineering attack ,spam, malware and botnets are the major issues of the Facebook. The various threats are explored by many researchers[3][6][21].serious economic and reputation loss caused by the use of malicious network usuage.Its a challenging task to figuring out the Sybil attack and compromised accounts. Facebook by themselves are doing

Research on new techniques to detect fake accounts and according to delete 30,000 fake accounts out of 1968 million accounts[7],that shows the detection rate ,they obtained in every negligible compared to total number of fake accounts[21].

The Facebook pages are widely used by the politicians, celebreties.They have large amount of friends and followers. The post and the comment on other post helps to determine malicious users. The malicious users post effect the community to which the user is belong to. News making events also causes malicious activities on the account. Anamoly is the set of activities which are different from the normal activity. Personal information can be accessed using spam messages of a particularaccount. Account details of users can be theft by the cyber criminals by sending messages to the friend s, followers. Behaviour analysis and modelling of the user's account various usersdetermine the malicious activities of the user account. The different information available in the Facebook is leading information, post information and lagging information according to Meire[21].The posts of the user is commented by the text emoji's ,analysis of which determines the user behaviour in the OSN's such as Facebook, twitter. Social influence analysis and malicious activities are inter related. Malicious users post effects on the community to which the user is belong to and in turn it effects on the social influence analysis.

A large amount of the dataset used by the analyser to detect anomalous activities. The data set can be collected using the Facebook graph API.To detect the anomalous activities in the images posted in the Facebook a large number of dataset is required. Taking features of the images into account dataset can be formed. The dataset is being applied to the images using different machine learning algorithms to determine the anomalous activities in the images. The accuracy of the anomalous detection depends on the data contain in the data set. If the data set is a large the accuracy percentage is more compared to the small dataset.

This paper discuss about the algorithm and the methods used to determine the anomalous activity of the images posted in the Facebook.

II. Basic Concept and Model

A.Basics of Machine Learning

Machine learning is the multidisciplinary research field that spans multiple disciplines including computer science,statistic,psycology and brain science[28]..Imitating human learning concept is the main objective of the machine learning algorithms. Depending on the learning concept machine learning is divided into three types supervised, Agriculture,chemistry,cognitiveapplications, naturallanguage processing uses supervised learning algorithm to train applications, Unsupervised, rein enforcementlearning. Supervised learning in the training of ladled data to the algorithm to makecomputer to be intelligent. Decisiontree, support vector machine algorithm, neural network are the algorithms used in the supervise model.Unlabled data used in the unsupervised miodel.Though data is unlabelled learning happens using clustering and auto encoder method. Learning in reinenforcement happen with the rewards and punishment of the algorithm technically it is the try-and error method.

B.Dataset

Dataset for the classier algorithms cane be formed using different techniques, truth data for the spam, fake profiles other malicious content. In this paper data set formation for the image processing to detect anomalous activities is required. Crowd sourcing method such human annotation, through service like Amazon mechanical Turk, annotation/coding by topic experts etc. are the third partyURL blacklist for phishing,malware,spam and other kind of malicious URLs[24].Large data sets containing human annotationcause scalability issue cannot be used tom identify the ground truth. Machine learning application on a small data set from a large sample of data set will determine the true positive affect which cannot be determined by the large set of the model. Small datasets may causes over fitting issue. Which will cause not produce the accurate result. That issue can also be addressed separately to get the accurate result from the algorithm.

For the detection of the anomalous activities in the images posts in the Facebook data set can be formed by taking features of the images into account. The combination of the image features used for the formation of the dataset. Combination of the features used to form a large number of data set from the limited features of the dataset. The datasets applied continuously to the classifier algorithms to determine the anomalous activities in the images. The data set formed has to undergo through null value treatment.

Base dataset is converted to the analytic dataset to being used by the model. Dataset has to undergo number of processes to find out analytic dataset. The processes are I. Null value treatment: Null values present in the column of the data set is being replaced by some other values. ii.outliertreatment: Presence of outlier in the dataset can bias the output of the dataset.so outlier treatment is required to get the proper result from the dataset. ,iii.Garbage value treatment: garbage value can also bias the result., to get accurate result from the dataset garbage value treatment is required. iv.univariateanalysis: Analysis of the single variable. Bivariate analysis: Analysis of two variable to get the output..

C.Models for anamoly detection

To determine the anomalous images there are two models can be followed

I. Baseline model
II. Benchmarking model.

Baseline model is the model which can be determined by using algorithm to the data set to get the required result for the algorithm. In case to detect anomalous activities in the image .Ensemble mechanism used on classifier algorithm such as decision tree. Ensemble mechanism is that in which multiple algorithms run at a time to get detect accurate anomalous activities. In this scenario ensemble model with decision tree is being used which is also called as Rain Forest algorithm. In this algorithm continuously decision tree applied to the classifier dataset formed by the features of the images. With each iteration result can be determined. The accuracy of the result of the always being verified. If the result is more accurate the machine learning algorithm then detectanomalousactives of the images .If the accuracy is not acceptable then Benchmarking model isused to determine the accurate anomalous activities.

Benchmarking model is the model which applied to the system when Baseline model marking model could not give accurate output. Bench marking model follow the algorithms such as SVM classifier algorithms, Naïve Bayes, Logisticregression, KNN classifier algorithm. The bench marking model will give nearly accurate result for the anamoly detection.

In bench marking model each time algorithm is used linearly distance of the element from the plane will remains same. The iteration takes place with each elements applying the classifier algorithms. If benchmarking model could not give the accurate result.

Boosting model is applied to the data set to get the accurate result .the boosting model applied algorithms in all prespective.It is applied to all combination of the dataset from dataset formed by taking characteristic of the images. Into account. Boosting model work on minus infinity to plus infinity.If the result is accurate during Baseline model then Boosting model is not applied to the data set, only baseline model is applied to the data set and benchmark model will verify the accuracy of the result obtained from the baseline model.

The above three models used to find the accuracy of the anamoly detection on images. In general these three models are used for dataset to determine malicious activities.

D.ANN learning for anomaly detection of images

ANNs have the ability to learn and model non-linear and complex relationships, which is really important because in real-life, many of the relationships between inputs and outputs are non-linear as well as complex. After learning from the initial inputs and their relationships, it can infer unseen relationships on unseen data as well, thus making the model generalize and predict on unseen data. Unlike many other prediction techniques, ANN does not impose any restrictions on the input variables (like how they should be distributed). Additionally, many studies have shown that ANNs can better model heteroskedasticity i.e. data with high volatility and non-constant variance, given its ability to learn hidden relationships in the data without imposing any fixed relationships in the data.ANNs based models have some advantages on K-mean model which make them most suitable for certain problem like anomaly detection of images in OSN like Facebook. ANN algorithm can result more percentage of accuracy compared to other algorithms in detection of anomalous activities.

III. Conclusion

Machine learning not only used in image processing, Natural language processing but it is included in each and every applications.OSN such as Facebook, twitter has affected by the malicious user. Detection of the anomalous activities in OSN is required.IN this paper determination of anomalous activities of images is being discussed ,How to prevent account form the Users who causes malicious activities. The type of data set need to use and the type of the model need to use to detect anamoly in images posted in Facebook account.

Acknowledgement

I would like to thank Ms N.Deepika assistant professor, of the Department Computer science and Engineering in New Horizon college of Engineering for all her supportand guidance. The Facility provided helped to complete good quality projector. Your guidance on technology helps me to develop a strong technical system.Aslo I would like to thank facility provided by the institution to complete my project.

Reference

[1]. I. Statista., "Leading social networks worldwide as of September 2016,ranked by number of active users (in millions)," accessed 2017-0123.[Online].Available: https://www.statista.com/statistics/272014/global-social-networksranked-by-number-of-users/
[2]. A. Sedghi, "Facebook: 10 years of social networking, in numbers," accessed 2017-10-25. [Online]. Available: https://www.theguardian.com/news/datablog/2014/feb/04/facebook-innumbers-statistics
[3]. A. Madain, A.-Z. AlaM, R. Al-Sayyed et al., "Online social networks security: Threats, attacks, and future directions," in Social Media Shaping e-Publishing and Academia. Springer, 2017, pp. 121–132.
[4]. M. Fire, R. Goldschmidt, and Y. Elovici, "Online social networks: threats and solutions," IEEE Communications Surveys & Tutorials, vol. 16, no. 4, pp. 2019–2036, 2014.
[5]. A. Sadeghian, M. Zamani, and B. Shanmugam, "Security threats in online social networks," in Informatics and Creative Multimedia (ICICM), 2013 International Conference on. IEEE, 2013, pp. 254–258.
[6]. A. Al Hasib, "Threats of online social networks," IJCSNS International Journal of Computer Science and Network Security, vol. 9, no. 11, pp. 288–93, 2009.
[7]. W. Kurt, "Facebook is deleting a bunch of spam accounts in its effort to fight fake news," accessed 2017-10-25. [Online].Available:https://www.recode.net/2017/4/14/15306220/facebook-fake-newsdeleting-spam-accounts
[8]. N. Ismail, "1 in 5 SMEs have fallen victim to social media hackers," accessed 2017-07-19. [Online]. Available: http://www.informationage.com/1-5-smes-fallen-victim-social-media-hackers-123466013/
[9]. M. Mccann, "Potential legal fallout from LaremyTunsils hacked Twitter account," accessed 2017-10-12. [Online]. Available: https://www.si.com/nfl/2016/04/29/nfl-draft-round-one-laremy-tunsiltwitter-hack-miami-dolphins
[10]. M. Meire, M. Ballings, and D. Van den Poel, "The added value of auxiliary data in sentiment analysis of facebook posts," Decision Support Systems, vol. 89, pp. 98–112, 2016.
[11]. H. Grigonis, "Facebook celebrates World Emoji Day by showing how many are shared each day," accessed 2017-10-22. [Online]. Available:https://www.digitaltrends.com/social-media/world-emoji-day2017/
[12]. D. Krackhardt, N. Nohria, and B. Eccles, "The strength of strong ties," Networks in the knowledge economy, p. 82, 2003.
[13]. M. S. Granovetter, "The strength of weak ties," American journal of sociology, vol. 78, no. 6, pp. 1360–1380, 1973.
[14]. C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao, "User interactions in social networks and their implications," in Proceedings of the 4th ACM European conference on Computer systems. Acm, 2009, pp. 205–218.
[15]. D. Savage, X. Zhang, X. Yu, P. Chou, and Q. Wang, "Anomaly detection in online social networks," Social Networks, vol. 39, pp. 62– 70, 2014.
[16]. J. Cao, Q. Fu, Q. Li, and D. Guo, "Discovering hidden suspicious accounts in online social networks," Information Sciences, vol. 394, pp. 123–140, 2017.
[17]. G. Srivastav and A. Gupta, "Going private in public: A study on shift in behavioral trend using facebook," Computers in Human Behavior, vol. 73, pp. 55–63, 2017.

[18]. F. Amato, A. Castiglione, A. De Santo, V. Moscato, A. Picariello, F. Persia, and G. Sperlí, "Recognizing human behaviours in online social networks," Computers & Security, 2017.
[19]. P. Bindu and P. S. Thilagam, "Mining social networks for anomalies: Methods and challenges," Journal of Network and Computer Applications, vol. 68, pp. 213–229, 2016.
[20]. X. Ruan, Z. Wu, H. Wang, and S. Jajodia, "Profiling online social behaviors for compromised account detection," IEEE transactions on information forensics and security, vol. 11, no. 1, pp. 176–187, 2016.
[21]. SavyanP. V. and S. Mary SairaBhanu, "Behaviour Profiling of Reactions in Facebook Posts for Anomaly Detection",2017 Ninth International Conference on Advanced Computing (ICoAC),978-1-5386-4349-5/17/$31.00 ©2017 IEEE.
[22]. Sneha C. Vishwakarma, Pooja R. Shejwalkar, Aishwarya R. Sadigale, Shyamal G. Palkhede, Prof. D. S. Thosar, "Detection of Malicious Applications on Facebook using Machine Learning Algorithm",IJARIIE-ISSN(O)-2395-4396Vol-3 Issue-3 2017
[23]. DominicSeyler,LunanLi,ChengXiangZhai,"IdentifyingCompromisedAccountsonSocialMediaUsingStatisticalTextAnalysis",arXiv:1804.07247v1 [cs.SI] 19 Apr 2018
[24]. PrateekDewan, PonnurangamKumaraguru,"Facebook Inspector (FbI): Towards Automatic Real Time Detection of Malicious Content on Facebook",3http://multiosn.iiitd.edu.in/fbapi/endpoint/?version=2.0&fid= <post id>
[25]. Jagath Sri LalSenanayaka, Surya TejaKandukuri, Huynh Van Khang, Kjell G. Robbersmyr,"Early Detection and Classification of Bearing Faults using Support Vector Machine Algorithm",978-1-5090-5853-2/17/$31.00 ©2017 IEEE
[26]. [26]QIANG LIU1, (Member, IEEE), PAN LI1, WENTAO ZHAO1, WEI CAI2, (Member, IEEE) , SHUI YU3, (Senior Member, IEEE), VICTOR C. M. LEUNG2, (Fellow, IEEE),"A Survey on Security Threats and Defensive Techniques of Machine Learning: A Data Driven View",2169-3536 (c) 2018 IEEE. Translations and content mining are permitted for academic research only. Personaluseisalsopermitted,butrepublication/redistributionrequiresIEEEpermission.Seehttp://www.ieee.org/publications_standards/publications/rights/index.html for more information.
[27]. M.SwarnaSudhaK.ArunPriya, A.Kanaka Lakshmi, A.Kruthika, D.LakshmiPriya,"Data Mining Approach for Anomaly Detection in Social Network Analysis",978-1-5386-1974-2/18/$31.00 ©2018 IEEE
[28]. https://books.google.co.in/books?hl=en&lr=&id=-eqpCAAAQBAJ&oi=fnd&pg=PA2&dq=concepts+of+machine+learning&ots=Yy4nkGhG5ptBv4xK06xi2Xbv1vg#v=onepage&q=concepts%20of%20machine%20learning&f=false

Urmi Priyadarshani Das "Anomaly Detection on Images Using Machine Learning Algorithm "International Journal of Engineering Science Invention (IJESI), vol. 07, no. 11, 2018, pp 68-71