International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 01 | Jan 2021 www.irjet.net p-ISSN: 2395-0072

A Survey on Approaches for Behaviour Analysis of Users in Online Social Networks

Midhun Shaji

Department of Computer Science and Engineering, College of Engineering Trivandrum, Trivandrum, India

---

Abstract - *Advancements in Online Social Networks (OSNs) is having a greater impact on the lives of individuals. Social-economic influence of social networking site is also increasing at an alarming rate. This has lead to the importance of analysing the behaviour of a user in those networks. Behaviour analysis in online social networks is an active field since the social media revolution. Existing social media services use different techniques for behaviour analysis which help them to tackle some of the cyber issues in social networking. But still, solutions for tackling attacks in social media is an open research area. Analysis of the behaviour of a user in social networking sites become important when decentralised social networks like diaspora become more popular. This survey discusses different approaches used in analysing user behaviour and their effectiveness in solving the problems.*

Key Words*: *Online Social Networks (OSN), Behaviour analysis, Cyberbullying, Cybercrimes, Clustering, community detection.

1. INTRODUCTION

Online Social networks (OSNs) are increasing their influence on the lives of people. Recent studies show that it is becoming common for an active user to face cyberbullying in the social network. Spreading of fake news, spamming and targeted attacks are also becoming an issue. This eventually leads to a point where people move away from these networks. Envisaging such a situation social networking services are trying to stop such attacks. Recent profile lock feature introduced by Facebook is a good example. Twitter and Youtube have detailed their approach to tackle these attacks in their security policy. In the current scenario, any user who is under an attack can report it to the social networking authority or the policing systems. This process of reporting an issue is tiresome for some. This increases the importance of analysing the behaviour of a user and methods to find trouble makers in social networks. This has lead to studies on the behaviour of a user in these networks.

Social networks are a platform for a person to express his views and ideas. Such a platform has to provide some amount of freedom to the user. Restrictions on user activity can lead to the failure of the networks. This was the case during an initial period of social networking sites. The advancements in social networking platforms along with the increase in usage of smartphone applications lead to an increase in the number of people using these platforms. From this point on, freedom was not the only concern. People became more conscious of the privacy factor. Data leakage scams increased the demands for privacy. Privacy vs freedom tradeoff was a new concern. Some of the social networking platforms had to close down their operations because of this. Cambridge Analytica scandal [5] in 2018 was a concern for users of Facebook. People started demanding for more privacy. It was not easy to gain privacy from a centralised social networking platform. This is the reason for a shift towards decentralized social networks. Such networks store user data in a decentralised manner. Decentralised networks such as the Diaspora gives more freedom to its users. In such networks, a user is set free to go to any extent in stealing others freedom, which becomes a security concern. Behaviour analysis of a user in an online social network becomes more relevant in such a scenario.

Three different entities in an online networking site are the User, the service provider [7] and the law enforcement system in the country. Behaviour analysis helps these three entities in tackling the cyber issues related to online social network Reason for increasing usage of social networking services is the ease with which a person can present his ideas and opinions to a large audience. Attacks targeted at a user is one of the problems that get solved by behaviour analysis. A system that analyses the behaviour of the user will help another user to accept or reject any request or messages from that user. Social network service providers also find it difficult to stop some attacks such as spamming and Sybil attacks. Law enforcement system in different countries finds it difficult to stop fake news and cyberbullying in social media services. Any analysis model needs to consider these three entities. Only then the system will become effective. Approaches used for analysis should be having a clear output and its implementation in a social networking service should be easy.

Some of the approaches used for behaviour analysis of a user in online social networks are community detection [3], two phases risk assessment [1], text and post-feature-based analysis [2], techniques based on principal component analysis [4] and centrality analysis [6]. Section 2 discusses these methods.

2. APPROACHES FOR USER BEHAVIOUR ANALYSIS

Each of the methods in analysing the behaviour of a user in an online social network is unique in their way of identifying a user who deviates from what is said to be normal. A set of normal traits is identified for this analysis. In community detection methods, a user is divided into communities to identify the deviation from the expected behaviour of the community members. The posts, comments, tweets or replies that a user has in a social network is also a source to detect

misbehaviour. The centrality of a node in social media can also be used to analyse a user. Such a centrality study may yield the most influential user in a network who can be a potentially harmful user. Approaches to analysing the behaviour of a user differ in the result they produce. Most of the approach discussed ends up in finding a parameter to analyse the user.

A community can be considered as a subset of a set of users who interact frequently. In other words, they are closely related to users or users who share a similar area of interest. Community detection and user behaviour analysis using this method can help in preventing some of the common security issues such as worm propagation and node influence.

A General Stochastic Block (GSB) model can be employed to detect different communities for a set of users [3]. Stochastic block model generates a structure for large scale networks. MapReduce based general stochastic algorithm is a parallel community detection algorithm which is used to calculate parameters of the general stochastic block model. Analysis of the effectiveness of the stochastic block model-based approach and other approach shows that the GSB model-based approach is having a lesser execution time with larger data size. Updating data to Hdfs which is a Hadoop database for big data analysis can be considered as a limitation as user information changes over time and any data become irrelevant in a very short period. Apart from the general stochastic block model, there are many community detection algorithms, that are specific to some networks including the hierarchical clustering model [14] and modularity based algorithms [15]

Risk assessment of a user in a social network is analysed by considering different behavioural features, group identification features [1], textual content features and posting features. Group identification features include the number of friends or followers a user has.. A user who is having a more or moderate number of friends are different from isolated users. The activity level of a user includes the number of messages, comments, posts or replies a user have. This may also include the responses the user received from others in the form of likes. Passive users have less number of such information. Public details made available can be a parameter. A user who has no public information made available can be either a person with privacy concerns or a user who is fake.

Some of the behaviour features are, Friendship Rate, Mutual Friendship Rate, Comment Rate, Post Rate, Post Propagation Speed, Like Propagation speed, Comment feedback rate and Post feedback rate. This and many other features can be used as a parameter to the clustering algorithms used.

Two-phases of clustering is used to analyse the risk factor[1]. Probabilistic clustering methods find the probability of a user being part of a cluster. User behaviour and group identification features are given as a user data vector to the clustering algorithm. Expectation-Maximization algorithm is used to determine the clusters. A risk score is given based on the probability of the user being in that cluster. High membership probability indicates a low risk and a low membership probability indicate a high deviation from the normal behaviour of the cluster members.

A user risk score is given based on the probability of a user falling in a cluster. A user will have different probability values for different cluster. Identifying abnormal behaviour by this method can cause erroneous results, as there may be users who have an equal probability of falling in two clusters. If one cluster is a cluster with good users and another with malicious users then the chance of a user getting a wrong risk score is high.

Principal component analysis can be used to identify behavioural features [4]. A high dimension data is projected to a low dimensional data. The highest value in the result indicates high variability in data. Data and noise need to be distinguished for an analysis.

PCA analysis considers a subspace of training data. Data vector indicates this subset of data. Eigen coefficients are found out for each of the eigenvectors. A subset of principal components attributes to most of the variability. A residual subspace is identified from the normal subspace. This residual subspace can be used to identify anomalous behaviour. In social platforms such as Facebook, there exist black market services that aim at providing a user with the desired likes and other favouring reactions. Such services can be identified and stopped using this method. Compromised attacks too can be identified for a larger deviation from normal behaviour. Click spam in Facebook is studied and used for test-ing the approach based on the principal component analysis. This unsupervised learning approach can be considered as a great tool in the process of analysing user behaviour.

The behaviour of a user can be studied based on the activities obtained from the user profiles [2]. User-generated data such as tweets in Twitter are analysed based on the following.

Text specific features: This includes the textual content of user activity. Natural language processing tools are used in this case to study user tweets. Lexical and syntactic features are taken into consideration. Lexical character and word features are analysed in detail. Syntactic features such as the use of punctuations and the structural features such as the number of sentences are used for analysis.

Post specific features: Posting behaviour of the user is being analysed. Posting pattern is identified for a user post and a reaction pattern are found. Amount of likes and shares for a post can be considered as a post-specific feature. The algorithm assumes that the oldest 100 tweets are done by a legitimate user. The number of training samples is increased

periodically. The algorithm analyses the tweets and returns the full set of suspicious tweet

Centrality studies can be of great importance in finding the most influential user in a social networking platform. Centrality studies are mainly used in predicting the categories of product a user is likely to buy. This can be extended to analysing user behaviour in OSNs [6]. Three types of centrality studies are carried out,

Degree centrality Considers the number of edges a node has in a network. Nodes in these case are the users and the edges are the relations they have. Higher the number of nodes linked to a node. higher is the probability of the user influencing others into doing something.

Closeness centrality: Ability of the user to reach other nodes. It evaluates how close the users are. This can be used to find the speed at which the data can spread through a network.

Eigenvector centrality: The importance of a node in the network is found using Eigenvector centrality. This can be found by finding nodes that are connected to another important node in the network

Most central nodes are identified using the above approaches. A distributed framework for network centrality measures can be developed through this approach.

Some other types of anomalous behaviour in social media include spamming. User data is being compromised to effectively spam a user. One of the methods uses the Dirichlet mixture model for identifying spammer in social media [8]. Url type anomalies occur in popular social media platform such as Twitter. URL posted by the user could lead to some serious security issues. This URL can be analysed and can be ranked for the amount of spam they can generate [9].

User trustworthiness can be assessed based on user attributes and interactions [10]. Social Ties Strength (STS) and Friends Of A Friend (FOAF) can be used for generating a trust model. Exact trust is not found but only determinants of trust. STS gives the intimacy of two users. This can be used as a factor of influence of a user on others. Factors can be users' attitudes, experiences, behaviour, and users' attributes similarities, such as demographic similarity, spatial proximity, and interests similarity. Trust is computed with these parameters. This trust model can be used to gain an overall view of the user profile. Trustworthy users can be rightly found. This approach can be used as a factor for assigning user score and identifies and ranks the trusted users more accurately than the other algorithms.

User identification is another perspective of user behaviour analysis. This can be used to prevent Sybil attack. User profiles are matched to find the similarity of two users [11]. User profile information or user-generated data could be used for identification of two users for any similarity.

3. CONCLUSION

User behaviour analysis in online social networks is becoming more relevant with the increase in cybercrimes through these social networking platforms. Behaviour analysis of users in online social networks aims to give more power to a user. Each of the approaches discussed in this survey can be utilized for an effective analysis of user behaviour in online social networks. Social networks are evolving from textual contents to images and videos, this makes the analysis of user data more complex. Approaches considering such contents also need to be studied and developed. Privacy of a user in online social networks needs to be preserved in any of the behaviour analysis approaches. Also, the number of false-positive results need to be minimized. The Features used to analyse a user can be further increased to get more precise risk assessment. Most of the approaches envisage implementation in social networking platforms. With all such limitations in mind, behaviour analysis can be further developed in an effective manner, which may be useful for users of any online social networking site.

REFERENCES

[1] Naeimeh Laleh, Barbara Carminati and Elena Ferrari,Risk Assessment in Social Networks based on User Anomalous Behaviours.IEEE Trans-actions on Dependable and Secure Computing 2016.

[2] Raad Bin Tareaf, Philipp Berger, Patrick Hennig, and Christoph Meinel. Malicious Behaviour Identification in Online Social Networks. Interna-tional Federation for Information Processing 2018.

[3] Cong Wan, Sancheng Peng , Cong Wang, Ying Yuan Communities De-tection Algorithm Based on General Stochastic Block Model in Mobile Social Networks. ATC'08 USENIX Annual Technical Conference, 2008.

[4] Bimal Viswanath, M. Ahmad Bashir Towards Detecting Anomalous User Behavior in Online Social Networks, 23rd USENIX Security Symposium 2014.

[5] Jim Isaak, Mina J. Hanna User Data Privacy: Facebook, Cambridge An-alytica, and Privacy Protection, COMPUTER, vol. 51, issue 2,2018,pp. 56-59.

[6] Ranjan Kumar Behera, Santanu Kumar Rath, Sanjay Misra, Robertas Damasevi and Rytis Maskeliunas,Distributed Centrality Analysis of Social Network Data Using MapReduce, Algorithms 2019.

[7] Long Jin, Yang Chen, Tianyi Wang, Pan Hui, Athanasios V. Vasilakos, Kuwait University Understanding User Behavior in Online Social Net-works: A Survey. IEEE Communications Magazine, September 2013.

[8] Farnoosh Fathaliani, Mohamed Bouguessa A Model-Based Approach for Identifying Spammers in Social

Networks, IEEE International Confer-ence on Data Science and Advanced Analytics (DSAA) 2015.

[9] Vishal Chauhan, Ajay Pilaniya, Anomalous behaviour detection in social networking, Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT), Jul. 2017.

[10] Jebran Khan, Sunchang Lee, Implicit User Trust Modeling Based on User Attributes and Behavior in Online Social Networks, IEEE Access 2019.

[11] Kaikai Deng, Ling Xing, A User Identification Algorithm Based on User Behavior Analysis in Social Networks.IEEE Access 2019.

[12] Andra, Z.: 10 alarming cybersecurity facts that threaten your data. Heimdalsecurity (2015)

[13] Bin Tareaf, R., Berger, P., Hennig, P., Meinel, C.: Identifying audience attributes: predicting age, gender and personality for enhanced article writing. In: International Conference on Cloud and Big Data Computing, pp. 79–88. ACM (2017).

[14] M. Girvan and M. E. J. Newman, Community structure in social and biological networks, Proceedings of the National Academy of Sciences of the United States of America, vol. 99, no. 12, pp. 7821–7826, 2002.

[15] M. E. Newman, Fast algorithm for detecting community structure in networks, Physical Review E Statistical Nonlinear and Soft Matter Physics, vol. 69, no. 6, pp. 066 133–066 133, 2004.