© AUG 2020 | IRE Journals | Volume 4 Issue 2 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV4I2-1723102

A Review of Hybrid Machine Learning Approaches for Fraud Detection in Online Social Networks

DR. USHA RANI
Associate Professor in Computer Science, Smt Sushma Swaraj Govt. College for Girls Ballabgarh

Abstract- Online social networks (OSNs) have become an essential platform for human communication, business and public expression, however the popularity of these OSNs has given rise to large-scale abuse by way of fraud, identity deception, spamming and digital manipulation. It is difficult to detect this misuse since the fraudulent accounts are designed as legitimate users. Hybrid machine learning methods that integrate several analysis methodologies (versus using a single method) have been proposed and demonstrated as a potentially effective solution.In this paper, we summarize hybrid Machine Learning techniques for OSN fraud detection in terms of their architectures, effectiveness and limitations. The survey also focuses on important frontiers in feature engineering, graph-based learning, and ensemble learning, and ends with some research gaps that are still open till today. It further underlines the increasing demand for scalable models to cope with real-time detection over large-scale social media.

Keywords: OSN, Machine Learning, Ensemble Learning, Scalable models, Spamming.

I. INTRODUCTION

Social media/networks such as Facebook, Twitter, Instagram, LinkedIn have billions of users and interactions taking place every day. Although they are a convenient and global medium, the platforms can also serve as vehicles for fraudulent agents who use false identities, spam accounts, bots or coordinated propaganda systems. Studies published between 2010 and 2018 has repeatedly stressed that a large proportion of social media users are not real, with Twitter itself reported to harbor millions of robots or scam accounts [1][2].

Early methods using rules, blacklists or manual moderation for fraud detection. Those approaches became outdated too as attackers modified their tactics and humanized what they were doing. Machine learning has been applied to the problem of automatically detecting fraud, but "mono-models," pure support vector machines (SVM) or a clustering algorithm alone were inadequate for capturing the range and dynamics of online fraud types and volumes [3].In the meanwhile, some researchers have introduced hybrid approaches by combing supervised learning, unsupervised clustering, graph mining, natural language processing (NLP) and statistical methods to promote more reliable results.

II. TYPES OF FRAUD IN ONLINE SOCIAL NETWORKS

Fraud in social networks takes various forms, including:

- Fake and impersonated profiles
- Automated bot accounts
- Mass spam campaigns
- Social engineering or phishing attempts
- Political or ideological manipulation
- Financial scams and fake promotions
- Malicious social bots and click-farms

In nature, while conventional cyber-fraud is not related mainly on social trust, the OSN-based fraud relies itself on it[9]. If a bot is acting like a human, using another person's identity and interacting in the same style as any other user on that platform — well, how should it be treated differently? This complicates detection and calls for data-driven approaches [4].

III. HYBRID MACHINE LEARNING METHODOLOGIES

3.1. Combining Supervised and Unsupervised Learning

Hybrid detection by combining classification models with clustering algorithm is a commonly used method. For instance, [5] fused K-means clustering and support vector machines to detect spammers on

Facebook. When applied to supervised learning, it can expose typical patterns of fraudulent activity, and utilise clustering to find those patterns amongst unknown (or unlabelled) accounts that never before were marked as fraud.

3.2. Graph-Based Analysis + Classification Models

As OSNs are networks, a large number of approaches include such graph features as the ratio of followers to followees, interaction density or edge betweenness among communities. Social graph analysis and supervised detection models were applied [6] to distinguish normal from fraudulent users. This hybrid method leverages social relationships, something purely text-based models are unable to encode.

3.3. Linguistic + Behavioral Feature Integration

Text contents, punctuations patterns, the interval between submission and reply time, profile completeness and interaction behavior are usually combined in a hybrid way. The author [7] also leveraged lexical features in combination with timing and activity based signals for separating human users from bots. When user metadata is combined with content-based analysis, detection performance significantly improves.

3.4. Ensemble and Multi-Model Learning

Many studies attempted ensemble learning by integrating multiple weak classifiers to have higher overall reliability. Random Forest, boosting methods, and voting-based decision systems belong to such kind. It was observed that ensemble based spam detection achieved better performance than single-level classifiers, since they were less dependent on a particular feature set or algorithm [8].

IV. ADVANTAGES OF HYBRID SYSTEMS

Hybrid ML methods have several competitive advantages over designing a single algorithm system for the purposes of OSN fraud detection. They are more robust against adversarial attacks as they combine different analytical viewpoints and are thus more difficult for the attacker to evade. These systems can also recognize known and emergent fraud behaviors, because they can combine supervised and unsupervised learning methodology. They are modular so that they can adapt; individual parts of the system can be upgraded to defend against new threats. Moreover, the hybrid model can achieve better performance with less false positives because of diverse decision-making manners. They utilize the above-mentioned various feature types like behavioral, social and textual cues to build more complete representations of user activity [10]. In addition, they show greater scalability than rule-based systems only.

V. CHALLENGES AND LIMITATIONS

Despite their progress, hybrid approaches face several technical and practical barriers:

- Lack of high-quality, labeled datasets
- Class imbalance (fraud cases are rare compared to real users)
- Computational overhead from multi-step pipelines
- Difficulty evaluating cross-platform performance
- Rapid changes in attacker behavior
- Limited access to private platform data due to API restrictions

While hybrid systems are often more accurate, they can also be harder to understand and explain, making it difficult for moderators or law enforcement to deploy them in real-world settings.

VI. CONCLUSION

Hybrid machine learning systems have a clear edge in the identification of abuse in online social media. These hybrid models use various analysis methods in comparison to single algorithm, achieving better understanding of user behaviors and enhancing detection accuracy dramatically. These frameworks are more effective at modeling varied types of fraud, mitigating false positive cases and remaining robust even from attackers attempting to mimic legitimate users. Although issues like limited availability of data, changing attacker behavior and system scalability are still challenges to be solved, the improvement obtained by hybrid models suggest their promise as a practical model. In conclusion,

hybrid models are a significant step towards more reliable and secure social networks one can trust.

REFERENCES

[1] Varol, O., Ferrara, E., Davis, C., Menczer, F., & Flammini, F. (2017). Online human-bot interactions: Detection, estimation, and characterization. *Proceedings of the International AAAI Conference on Web and Social Media, 11*(1), 280–289. Crossref

[2] Benevenuto, F., Rodrigues, T., Almeida, V., Almeida, J., & Gonçalves, M. (2010). Detecting spammers on Twitter. *Proceedings of the Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS'10).*

[3] Yang, Z., Wilson, C., Wang, X., Gao, T., Zhao, B. Y., & Dai, Y. (2014). Uncovering social network sybils in the wild. *ACM Transactions on Knowledge Discovery from Data, 8*(1). Crossref

[4] Cresci, S., Di Pietro, R., Petrocchi, M., Spognardi, A., & Tesconi, M. (2017). The paradigm shift in social bot research. *Communications of the ACM, 61*(3), 72–80.

[5] Wang, G., Mohanlal, M., Wilson, C., Wang, X., Metzger, M., Zheng, H., & Zhao, B. (2013). Social Turing tests: Crowdsourcing sybil detection. *Proceedings of the 20th Network and Distributed System Security Symposium (NDSS).*

[6] Stringhini, G., Wang, G., Egele, M., Kermarrec, A., Kruegel, C., Vigna, G., & Zhang, H. (2015). Follow the money: Distinguishing regular and fraudulent accounts in online social networks. *NDSS.*

[7] Lee, K., Eoff, B., & Caverlee, J. (2011). Seven months with the devils: A long-term study of content polluters on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media, 5*(1), 185–192. Crossref

[8] Ahmed, F., & Abulaish, M. (2013). A generic statistical approach for spam detection in Online Social Networks. *Computer Communications, 36*(10–11), 1120–1129. Crossref

[9] Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, F. (2016). The rise of social bots. *Communications of the ACM, 59*(7), 96–104. Crossref

[10] Almaatouq, A., Alsaleh, M., Alarifi, A., Al-Salman, A., & Al-Khalifa, H. (2016). Semantic-graph based approach for efficient spam detection in online social networks. *IEEE Access, 4*, 9400–9412. Crossref