Fraud Analytics: A Research Agenda

Baesens B.

Abstract

Fraud is as old as humankind and appears in many types and forms. Popular examples are credit card fraud, tax evasion, identity theft, insurance fraud, counterfeit, click fraud, anti-money laundering, and payment transaction fraud. In earlier research we defined fraud as an uncommon, well-considered, imperceptibly concealed, time-evolving, and carefully organized crime (Baesens, Van Vlasselaer, and Verbeke 2015). Nowadays, fraud is typically tackled using state-of-the-art analytical techniques with many accompanying challenges. It is the purpose of this article to highlight twelve research topics (**RTs**) that we believe prioritize high on the agenda of contemporary fraud analytics models. We do this by reviewing fraud analytics from a data, model, performance, and deployment perspective.

The data perspective

Based upon many benchmarking studies and research we conducted, we firmly believe that the best way to boost the performance of analytical fraud models is by carefully considering and enriching the data rather than developing yet another highly-parametrized, overly complex machine learning technique. This can be achieved in three possible ways.

Both structured as well as unstructured data can be used for fraud analytics. Structured data can easily be represented in a tabular format with rows typically depicting transactions, customers, companies, etc., and columns representing variables (also called predictors or covariates) that represent socio-demographic status, transactional information, contextual information, etc. Popular examples of unstructured data are text, images, audio, video and graph data. Many new types of unstructured data have recently emerged, such as data obtained from drones, wearables, or sensors. Studying the impact of these new data sources for fraud is an interesting research topic that may, in certain settings, also come with important privacy and ethics challenges (**RT1**).

Another way to leverage the data to boost analytical fraud models is by carefully engineering it taking into account the characteristics of the analytical technique that will be used to analyze it (Baesens, Höppner, and Verdonck 2021). Structured data engineering can be decomposed into instance (also called observation) and feature engineering. Instance engineering relates to fraud being uncommon and the fact that real-life fraud data sets are typically very skewed, necessitating the adoption of special preprocessing routines, which are basically all variants of under- or oversampling. Throughout our extensive empirical research and industry experience, we found that undersampling and SMOTE are usually the preferred methods to deal with skewed data sets. Hence, we do not expect much performance gains from yet another new instance engineering method. Structured feature engineering concerns the careful crafting of features from variables in the data to improve either the statistical performance or interpretability of the analytical fraud model (Baesens et al 2021). Popular examples are creating smart time features using, e.g., a von Mises distribution, creatively defining Recency, Frequency and Monetary (RFM) features, features based on Benford's law (an interesting law describing the distribution of the first digit of a set of numbers), and including temporal trend features. Somewhat more ambitious is the definition of social networks (e.g. between credit

cards and merchants in credit card fraud; between policy holder, claim, broker, expert and garage in car insurance; between bank accounts in anti-money laundering) and the ingenious featurization thereof which requires both analytical knowledge as well as business creativity (Óskarsdóttir et al 2021; Van Belle, De Weerdt, and Baesens 2022; Van Vlasselaer et al 2017). Also, variable transformations can be considered a type of feature engineering, such as the logarithmic, exponential, Box-Cox, and powerful Yeo-Johnson transformations (Baesens et al 2021). We believe much more research is needed in structured feature engineering to understand which type of features work well for what type of fraud, either in terms of statistical performance, profit, interpretability, or operational efficiency (**RT2**).

Unstructured data engineering aims at generating numerical representations, also called embeddings, for text, images, audio, video, sensor, geospatial, graph data, etc. Often this is done using powerful deep learning neural networks trained to generate condensed representations of the data by squeezing it into a lower dimensional space without significant loss of semantics. Unfortunately, many of these embeddings generated are merely black box numbers, which poses a serious research challenge in terms of interpretation in case understanding the fraud mechanics (e.g., to develop fraud prevention strategies) is important (**RT3**).

The analytical technique perspective

Fraud analytics can be tackled in either a supervised (fraud indicator available) or unsupervised (no fraud indicator available) way. The former is also referred to as predictive analytics, whereas the latter is called descriptive analytics or anomaly detection. A plethora of analytical techniques have been developed for both analytical tasks in the literature.

For supervised learning on structured data, various benchmarking studies indicated that when interpretability is unimportant, ensemble methods such as random forests or XGBoost are the method of choice (Gunnarsson et al. 2021). However, when interpretability is needed, one can either resort to standard techniques such as logistic regression and decision trees, or a complex one such as XGBoost and complement this with explanation facilities such as partial dependence plots, LIME and especially Shapley values which are commonly adopted in the industry (Molnar 2022). However, a key research challenge is that many of these explanatory techniques yield explanations that are sometimes hard to interpret unambiguously, and typically have difficulties dealing with multicollinearity and interaction effects between variables (vanden Broucke and Baesens 2021). Both challenges require a unique blend of quantitative and qualitative research using, e.g., surveys and interviews conducted in close collaboration with fraud investigators (**RT4**).

For unsupervised learning on structured data, many methods have been suggested as well, such as Local Outlier Factor (LOF), isolation forest (iForest), one-class SVMs, DBSCAN, and deep-learning autoencoders (Stripling et al. 2018). Recent research has shown that iForest is the method of choice for anomaly detection performance (Tiukhova et al. 2022). However, further research is needed to develop interpretation facilities to properly understand why an observation is considered an anomaly or, in our setting, suspicious of fraud. (**RT5**)

For supervised or unsupervised learning on unstructured data (e.g., images, audio, video, etc.), deep learning techniques are the method of choice. However, with nearly a continuous avalanche of new deep learning techniques being introduced at an unseen pace, it needs to be

clarified when what deep learning technique should be used on what type of unstructured data necessitating additional benchmarking studies (**RT6**).

The performance perspective

Analytical fraud models are typically evaluated in terms of precision, recall, F1 score, or the area under the precision-recall curve, which is tailored toward skewed class distributions. However, for many businesses, what matters is the bottom line impact or in other words increased profit or reduced cost. Even modest improvements in fraud detection might result in significant dollar savings. However, more research is needed to develop fraud application-specific cost metrics to measure the economic performance of either supervised or unsupervised fraud models (**RT7**).

Another interesting research challenge concerns the estimation of the analytical models themselves. Right now, this is typically done by optimizing a statistical objective function such as a likelihood error in logistic regression, entropy in decision trees, or Euclidean distance in an unsupervised setting. However, it would be much more interesting if the objective function considered is not statistical but economically meaningful, as this directly optimizes the target of interest. Though some successful preliminary attempts have been made at this in, e.g., payment transaction fraud (Höppner et al 2022), much more research is needed in other fraud settings (**RT8**).

Recently another performance measure emerged related to the carbon footprint of analytical models (Lasse, Kanding and Selvan 2020). This especially pertains to deep learning methods, which often require tons of parameters to be estimated on energy-consuming hardware accelerators. New research studying and minimizing the ecological impact of analytical fraud models is another interesting research topic to consider (**RT9**).

The deployment perspective

Whenever analytical fraud models are being put into production, they become immediately outdated since the data and fraud tactics are subject to continuous change, also referred to as the cat-and-mouse game between fraudsters and fraud preventors. Hence, analytical fraud models should be complemented with backtesting facilities to permanently monitor their performance and decide when corrective actions are needed. This offers some interesting avenues for future research, such as what performance metrics to monitor, how to tweak the model based on new incoming data incrementally, or when to decide on a complete model rebuild (**RT10**).

In many settings, analytical models for marketing, risk, or other business decisions, are being deployed through a public (company-wide, customers, or the general public) query interface, i.e., as an application programming interface (API). In 2016, a group of researchers (Tramèr et al. 2016) worked out how a fraudster with access to such a query interface could duplicate the model's functionality without having any prior knowledge of the model's parameters or training data. This is a malicious practice that is often referred to as model exfiltration or model theft (vanden Broucke and Baesens 2021). Obviously, this introduces a new type of fraud with new accompanying challenges and requires highly sophisticated analytical detection mechanisms, which is another interesting research topic since we are now using analytics to fight analytics (**RT11**).

Conclusion

This article highlighted eleven research topics that we consider highly needed in fraud analytics. We want to add one more overarching topic, which essentially applies to all the other ones above and relates to the reproducibility of new research approaches and findings (**RT12**). Though more and more scientific outlets and authors try to address this already, we still feel there is further room for improvement on this important topic.

Disclosure statement

No potential conflict of interest was reported by the author.

References

[1] Baesens B., T. Verdonck, M. Óskarsdóttir, and S. vanden Broucke. 2021. "Special issue on feature engineering editorial". *Machine Learning.* doi: doi.org/10.1007/s10994-021-06042-2.

[2] Baesens, B., S. Höppner, and T. Verdonck, T. 2021. "Data Engineering for Fraud Detection." *Decision Support Systems* 150. doi: doi.org/10.1016/j.dss.2021.113492.

[3] Baesens, B., V. Van Vlasselaer, and W. Verbeke. 2015. *Fraud Analytics Using Descriptive, Predictive, and Social Network Techniques: A Guide to Data Science for Fraud Detection.* Wiley, ISBN 1119133122.

[4] Gunnarsson B. R., S. van den Broucke, B. Baesens, M. Óskarsdóttir, and W. Lemahieu W. 2021. "Deep learning For Credit Scoring: Do or Don't." *European Journal of Operational Research* 295(1):292-305. doi: doi.org/10.1016/j.ejor.2021.03.006.

[5] Höppner, S., B. Baesens, W. Verbeke and T. Verdonck. 2022. "Instance dependent cost-sensitive learning for detecting transfer fraud." *European Journal of Operational Research* 297(1): 291-300. doi: doi.org/10.1016/j.ejor.2021.05.028.

[6] Lasse F. W. A., B. Kanding, and R. Selvan. 2020. "Carbontracker: Tracking and Predicting the Carbon Footprint of Training Deep Learning Models." *Proceedings of the ICML Workshop on Challenges in Deploying and monitoring Machine Learning Systems*. doi: doi.org/10.48550/arXiv.2007.03051

[7] Molnar C. 2022. "Interpretable Machine Learning." ISBN: 979-8411463330.

[8] Óskarsdóttir M., A. Waqas, K. Antonio, B. Baesens, R. Dendievel, T. Donas and T. Reynkes T. 2021. "Social network analytics for supervised fraud detection in insurance." *Risk Analysis* 42. doi:10.1111/risa.13693.

[9] Stripling E., B. Baesens, B. Chizi, and S. vanden Broucke. 2018. "Isolation-Based Conditional Anomaly Detection on Mixed-Attribute Data to Uncover Workers' Compensation Fraud." *Decision Support Systems* 111. doi: doi.org/10.1016/j.dss.2018.04.001.

[10] Tiukhova E., M. Reusens, B. Baesens, and M. Snoeck. 2022. "Benchmarking Conventional Outlier Detection Methods." *Proceedings of the International Conference on Research Challenges in Information Science*. Springer, Cham.

[11] Tramèr, F., F. Zhang, A. Juels, M.K. Reiter, and T. Ristenpart. 2016. “Stealing Machine Learning Models via prediction APIs.” Proceedings of the 25th USENIX Security Symposium. doi: doi.org/10.48550/arXiv.1609.02943

[12] Van Belle R., J. De Weerdt, and B. Baesens. 2022. “CATCHM: A novel network-based credit card fraud detection method using node representation learning.” *Decision Support Systems*. doi: doi.org/10.1016/j.dss.2022.113866.

[13] Van Vlasselaer V., C. Bravo, O. Caelen, T. Eliassi-Rad, L. Akoglu, M. Snoeck, and B. Baesens. 2015. “APATE: A Novel Approach for Automated Credit Card Transaction Fraud Detection using Network-Based Extensions.” *Decision Support Systems* 75:38-48. doi: doi.org/10.1016/j.dss.2015.04.013.

[14] Van Vlasselaer V., T. Eliassi-Rad, L. Akoglu, M. Snoeck, and B. Baesens. 2017. “GOTCHA! Network-based Fraud Detection for Security Fraud.” *Management Science* 63(9):3090-3110. doi: doi.org/10.1287/mnsc.2016.2489.

[15] vanden Broucke S., and B. Baesens. 2021. *Managing Model Risk*, ISBN: 979-8521686988.