Visual Analytics for Fraud Detection and Monitoring

Roger A. Leite, Theresia Gschwandtner, Silvia Miksch, Erich Gstrein, and Johannes Kuntner *†

ABSTRACT

One of the primary concerns of financial institutions is to guarantee security and legitimacy in their services. Being able to detect and avoid fraudulent schemes also enhances the credibility of these institutions. Currently, fraud detection approaches still lack Visual Analytics techniques. We propose a Visual Analytics process that tackles the main challenges in the area of fraud detection.

Keywords: Visual Knowledge Discovery, Time Series Data, Business and Finance Visualization.

1 INTRODUCTION

Financial operations management systems need to be in constant evaluation to avoid frauds and to provide risk management. Operations, such as a bank transaction and credit control, usually involve data with time-oriented and multivariate aspects. Both aspects are targets of interest to Visual Analytics (VA) community. Due to its complex nature [1], time-oriented data as well as multivariate data require detailed exploration and analysis.

Fraudulent schemes such as 'money laundry', or 'straw persons' should be identified and fought as fast as possible by financial systems. Government, banks, and others financial institutions that provide credit, and money transaction services are always interested in improving operations monitoring, and fraud detection systems.

Artificial Intelligence (AI) algorithms, and fraud detection metrics are currently used to identify suspicious patterns and outliers in the financial field [3]. However, fraud techniques are constantly changing with the aim to generate new action flows that do not fit into known patterns. In other words, novel fraudulent schemes are hard to detect. Search for known fraud patterns is not enough. To this end, VA methods can be used to improve new pattern discovery.

VA methods allow data understanding, monitoring, and analysis from a visual cognitive perspective, and thus, can help financial institutions in suspicious behavior detection. There are approaches that already combine both fields (AI and VA). However, each of those has its own limitations. We identified important tasks in collaboration with a financial company. After a literature study regarding fraud identification, common shortcomings were identified and commented in Section 2.

Our contributions are: (1) listing the challenges that are associated with the fraud detection problem, (2) the conceptual design of a VA process tackling fraud detection by combining the use of detection metrics with interactive visual exploration, and (3) a first prototype based on the proposed process.

2 RELATED WORK

Fraud detection and risk evaluation are not exclusive tasks from the financial field. Malware detection and network monitoring also involve the same type of data as financial transactions. Attacks from those fields are of similar nature. They also cover suspicious behavior identification in a wide range of sequential multivariate data. In the following section, we illustrate VA approaches used in the financial market and for malware analysis.

A survey of visualization systems for malware analysis is presented by Wagner, et al. [7]. From 25 papers, only seven combine time-oriented and interaction approaches. It also outlines open challenges such as the improvement of attack classification, enhancement of expert involvement during the tool development, and the merge of known analytical methods with visualizations.

One of the pioneers of fraud detection visualization analysis was Kirkland, et al. [5]. They developed the NASD Regulation Advanced-Detection System (ADS) which uses five different visualizations. ADS combines detection and discovery components, supports multiple regulatory domains, and shares the same data among all visualizations. ADS also uses AI, visualization, pattern recognition, and data mining to support regulatory analysis, alerts (pattern detection), and knowledge discovery.

WireVis [2] presents multiple coordinated view visualizations based on identifying specific keywords within the wire transactions. These different views aim to depict relationships among keywords and accounts over time. This approach is directly related to our proposed step of *interactive visual filtering* (see Figure 2).

Huang, et al. [4] presents a new framework of VA for stock market security. They present an approach with two stages (with different visualizations): (1) visual inspection of market performance (using 3D treemap display), and (2) behavior-driven visual analysis of trading networks (using a node-link diagram), for the identification of attackers.

According to our literature study the following challenges can be derived: (a) development of a comprehensive VA design; (b) enhancement of the scalability for frauds analysis; (c) modifying analytical methods and visualizes it's consequence; (d) knowledge base construction in order to aid in forensics; (e) event analysis and prediction (monitoring).

3 CONCEPTUAL DESIGN

Following the design triangle [6], to generate interactive VA methods we first tackle the three main aspects: data, users, and tasks.

Data Financial transaction events constitute multivariate and time-oriented data which may include details regarding the person or company involved in the transaction.

Users We consider two different types of users. (1) Expert users from financial institutions that are interested in the security and credibility of their company. (2) Credit managers, who are not experts in fraud detection but at the same time want to evaluate or monitor the some given credit flow and occasionally match it with a suspicious behavior.

Tasks We identified the following tasks: fraud detection, fraud classification, credibility analysis, and customer behavior monitoring.

Most existing VA approaches in the field of fraud detection are node-link diagrams which represent sequences of suspicious events identified by AI and metric techniques. However, these approaches do not exploit the full potential of VA in terms of interactive computation and visualization. To fill the current gaps, we propose a visualization process (see Figure 1) that utilizes VA methods. From

*Roger A. Leite, Theresia Gschwandtner, and Silvia Miksch are with the Vienna University of Technology. Erich Gstrein, and Johannes Kuntner are with the Erste Group IT.

†E-mail: {roger.leite, gschwandtner, miksch}@ifs.tuwien.ac.at, {erich.gstrein, johannes.kuntner}@s-itsolutions.at

a literature review we derived several tasks of which we focus on fraud detection and classification as well as customer monitoring.

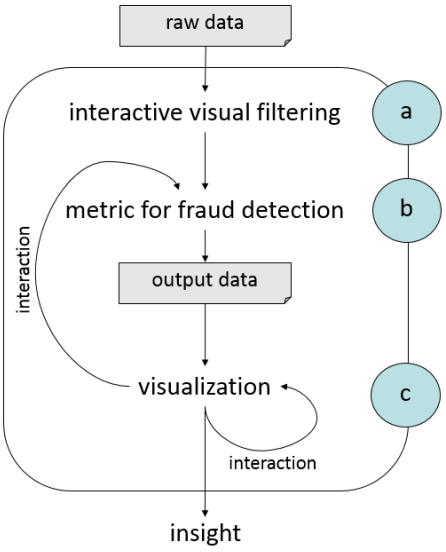

Figure 1: Financial VA generation process. Interactive visual filtering (a) allows for initial insights and filtering of the data set. Metric for fraud detection (b) is the step where different risk management formulas can be applied and edited, generating "output data", which is used in (c). Visualization (c) is the phase where VA methods are employed to make the results from fraud detection metrics visually explorable. Interaction techniques and immediate visual feedback aid the user to fine-tune these metrics.

Fraud Detection The first task is the classification of known frauds. This process generates visual signatures for each group of similar fraud cases. Those signatures are later used, during monitoring, to identify potential hybrid attacks trough signature comparisons. This can be done automatically to some extend. However, fraud attacks are changing constantly. The visual comparison allows for appraising suspicious cases where automatic methods fail. This information is then used to modify the detection metrics accordingly.

Customer monitoring Costumers monitoring focuses on real-time monitoring of the data for an early identification or even prediction of suspicious behavior. To this end, we use the fraud detection mechanisms described above.

Our proposed process include the following steps: Figure 1 (a) we generate different interactive visualizations from the raw multivariate data which are used to filter the most interesting attributes as well as their value ranges. Besides a focus on features desired by the experts, this step also allows for generating first hypotheses about attribute relationships. We have prototypically implemented this first phase (see Figure 2).

In the next step (see Figure 1 (b)), metrics or AI techniques are applied. Fraud detection analysts need to be able to set and edit available financial metric formulas, and also create their own metrics. Previously generated signatures of different types of frauds are used to automatically identify suspicious cases. Subsequently, we present these results in an interactive visualization (see Figure 1 (c)). This visualization allows the analyst to dynamically modify the visual representation of data features to meet his/her interests and questions in mind. The visual comparison of fraud signatures helps to reason about possible frauds. Insights from this exploration are then used to adjust and fine-tune the fraud detection metrics accordingly, closing the feedback loop (see arrows in Figure 1).

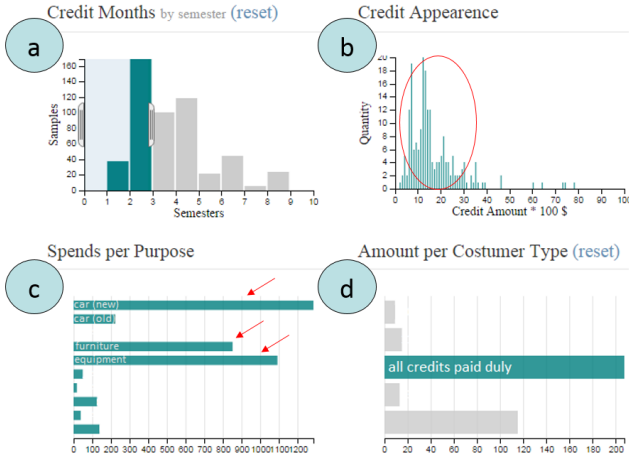

Figure 2: Multivariate interactive filtering. In this scenario, we filter for credit solicitation samples for two semesters, or less (see (a)) that were ordered from customers that have "all credits paid duly"(middle bar in (d)). Using this double filter, we can see that those credits usually involve amounts among 500$ and 3.500$ (red circle in (b)), and are used mainly to buy cars, furniture, and equipment (red arrows in (c)). Besides allowing for an analysis of group trends, this filtered data is used in all subsequent steps (see Figure 1).

4 Conclusion and Further Work

We summarized the main challenges of fraud detection in the financial field. While existing approaches employ static data displays, we propose a VA approach, focusing on fraud detection, and customer monitoring. The proposed pipeline combines efficient fraud detection techniques (i.e., AI techniques and fraud detection metrics) with VA methods. Our approach may also be applicable in similar domains, such as malware detection or tax usage analysis.

Acknowledgements

This work was supported by the Austrian Federal Ministry of Science, Research, and Economy via CVAST, a Laura Bassi Centre of Excellence (No. 822746).

References

[1] W. Aigner, S. Miksch, H. Schumann, and C. Tominski. *Visualization of time-oriented data*. Springer Science & Business Media, 2011.

[2] R. Chang, M. Ghoniem, R. Kosara, W. Ribarsky, J. Yang, E. Suma, C. Ziemkiewicz, D. Kern, and A. Sudjianto. Wirevis: Visualization of categorical, time-varying data from financial transactions. In *Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on*, pages 155–162. IEEE, 2007.

[3] T. Fawcett, I. Haimowitz, F. Provost, and S. Stolfo. AI approaches to fraud detection and risk management. *AI Magazine*, 19(2):107, 1998.

[4] M. L. Huang, J. Liang, and Q. V. Nguyen. A visualization approach for frauds detection in financial market. In *Information Visualisation, 2009 13th International Conference*, pages 197–202. IEEE, 2009.

[5] J. D. Kirkland, T. E. Senator, J. J. Hayden, T. Dybala, H. G. Goldberg, and P. Shyr. The nasd regulation advanced-detection system (ads). *AI Magazine*, 20(1):55–69, 1999.

[6] S. Miksch and W. Aigner. A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data. *Computers & Graphics*, 38:286–290, 2014.

[7] M. Wagner, F. Fischer, R. Luh, A. Haberson, A. Rind, D. A. Keim, and W. Aigner. A Survey of Visualization Systems for Malware Analysis. In R. Borgo, F. Ganovelli, and I. Viola, editors, *EG Conference on Visualization (EuroVis) - STARs*, pages 105–125. The EGA, 2015.