EVA: Visual Analytics to Identify Fraudulent Events

Roger A. Leite, Theresia Gschwandtner, Silvia Miksch, Simone Kriglstein,
Margit Pohl, Erich Gstrein, and Johannes Kuntner

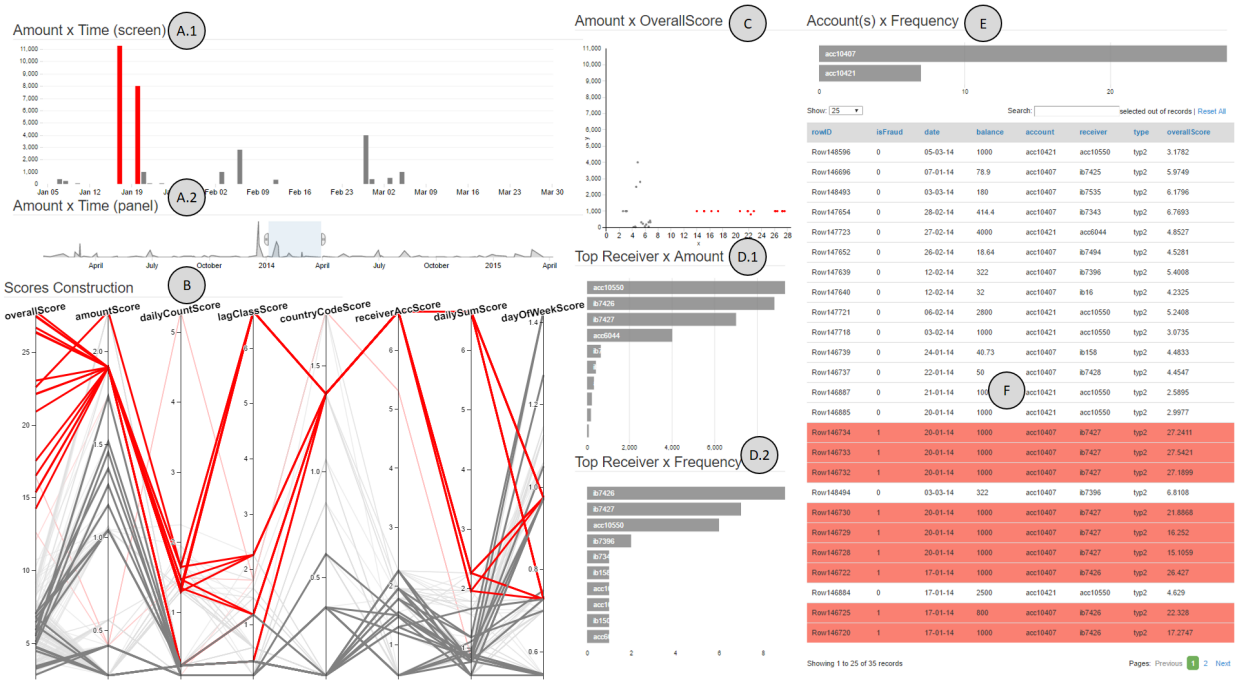

Fig. 1. Screenshot of EVA (Event detection with Visual Analytics). (A.1, A.2) Temporal Views: a filter was applied in (A.2) to the period from January 2014 until April 2014. (B) Score Construction View: each line represents a transaction and its scores. (C) Amount vs Overall Score Scatterplot. (D.1, D.2) Ranks of accounts that received the highest amounts of money from the selected account and accounts that received the highest number of transactions from the selected account. (E) Accounts Selector: bars shows amount of transactions from each account. (F) Dynamic Table of raw transaction data. In all views, elements that represents suspicious data are highlighted in red.

Abstract— Financial institutions are interested in ensuring security and quality for their customers. Banks, for instance, need to identify and stop harmful transactions in a timely manner. In order to detect fraudulent operations, data mining techniques and customer profile analysis are commonly used. However, these approaches are not supported by Visual Analytics techniques yet. Visual Analytics techniques have potential to considerably enhance the knowledge discovery process and increase the detection and prediction accuracy of financial fraud detection systems. Thus, we propose EVA, a Visual Analytics approach for supporting fraud investigation, fine-tuning fraud detection algorithms, and thus, reducing false positive alarms.

Index Terms—Visual Knowledge Discovery, Time Series Data, Business and Finance Visualization, Financial Fraud Detection

1 Introduction

Event detection is an important task in many domains such as finding interesting changes in stock markets, spotting problems in health parameters, or detecting financial fraud. Analyzing these events in a temporal context allows the identification of insights such as frequency, trends, and changes. Moreover, the investigation of outliers allows the analyst to identify risks, drastic changes, or rare occurrences. In this work we focus on the identification of anomalous events in the financial sector.

Financial institutions handle millions of transactions from clients per year. Although the majority part of these transactions being legitimate, a small number of them are criminal attempts, which may cause serious harm to customers or to the financial institutions themselves. Thus, the trustability of each transaction has to be assessed by the institution. However, due to the complex and multidimensional data at hand, financial fraud detection (FFD) is a difficult task.

The well renowned Oxford Dictionary defines fraud as "wrongful or criminal deception intended to result in financial or personal gain" [1].

- *Roger A. Leite, Theresia Gschwandtner, Silvia Miksch, and Margit Pohl are with Vienna University of Technology. E-mail: {firstname.lastname}@tuwien.ac.at*
- *Simone Kriglstein is with University of Vienna Faculty of Computer Science, Austria. E-mail: simone.kriglstein@univie.ac.at*
- *Erich Gstrein and Johannes Kuntner are with Erste Group IT International, Austria. E-mail:{firstname.lastname}@erstegroup.com*

Manuscript received 31 Mar. 2017; accepted 1 Aug. 2017.
Date of publication 28 Aug. 2017; date of current version 1 Oct. 2017.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TVCG.2017.2744758

[1] http://www.oxforddictionaries.com/definition/english/fraud (accessed December 10, 2016)

Software environments handling sensitive data such as financial operation management systems, systems for insurance evaluation, or companies' internal control systems, need to be in constant evaluation to prevent fraud, to provide risk management, and, thus, to avoid serious consequences. All these scenarios deal with similar data with the aim to detect suspicious events and, thus, to identify frauds. For instance, the two tasks of monitoring bank transactions and credit control usually involve data with time-oriented and multivariate aspects. Due to its complex nature [1], time-oriented and multivariate data require sophisticated means for detailed analysis and exploration. By consequence, both are subjects of interest to the Visual Analytics (VA) community.

Besides its challenging nature, FFD has also a strong social and financial importance. For instance, fraudulent schemes such as 'money laundering', 'unauthorized transaction', or 'straw person' should be detected and fought as fast as possible by financial systems, since the negative economical and social impact increases with time. Thus, governments, banks, and other financial institutions that provide credit and money transaction services have a strong interest in improving operation monitoring and fraud detection.

Kielman et al. [14] describe fraud detection as an open VA problem that requires visual exploration, discovery, and analysis. However, many of the current solutions involve mainly data mining techniques, while neglecting the potential of VA techniques to integrate human analysis into the process [13]. In this paper, we aim at closing this gap by presenting a VA approach for the investigation of suspicious financial transactions and fine-tuning of an existing automatic alert system. VA approaches may be utilized to identify different types of frauds. In this work, we focus on detecting "unauthorized transactions" within a financial institution. We designed our VA approach for FFD according to the nested model [29] paying attention that our solution is flexible and extensible enough to be applied in similar domains with similar multivariate and time-oriented aspects. The main contributions are:

- In tight collaboration with domain experts we analyzed the real world problem of FFD and iteratively designed EVA, a VA approach to improve their current work flow;
- EVA interweaves well-known visualization techniques, which our domain experts are mostly familiar with, and automatic methods;
- To the best of our knowledge, we present the first VA approach based on a scoring system for FFD;
- We present our findings from an evaluation with three target users (not involved in the design process) and categorize the types of insights that could be gained with our prototype;
- We derived open challenges and possible future research directions in the field.

2 Related Work

There is a number of surveys that focus on fraud detection. In 2002, Bolton and Hand [32] published a review about fraud detection approaches. They described the available tools for statistical fraud detection and identified the most used technologies in four areas: credit card fraud, money laundering, telecommunication fraud, and computer intrusion. Kou et al. [20] presented a survey of techniques for identifying the same types of fraud as described in [32]. The different approaches are broadly classified into two categories: misuse and anomaly detection. Both categories present techniques such as: outlier detection, neural networks, expert systems, model-based reasoning, data mining, state transition analysis, and information visualization. These works helped us to understand diverse fraud domains and how they are normally tackled. When looking on surveys of visual approaches for financial data, we identified FinanceVis [7] which is a browser tool including over 85 papers related to financial data visualization. FinanceVis was instrumental in analyzing how data that is similar to our data is usually visualized. Motivated by a lack of information, Ko et al. [19] presented a survey of approaches for exploring financial data. In this work, financial data experts were interviewed concerning their preferences of data sources, automated techniques, visualizations, and interaction methods.

When it comes to visual solutions to support FFD, Kirkland et al. [15] published one of the first works in fraud detection using visual techniques. In their work they combined Artificial Intelligence (AI), visualization, pattern recognition, and data mining to support regulatory analysis, alerts (fraud detection), and knowledge discovery. In our approach, we use a similar combination of techniques, but we also provide means for an interactive exploration of the visualized data.

WireVis's [4] main idea is to explore big amounts of transaction data using multiple coordinated views. In order to aid fraud detection, they highlight similarities between accounts based on keywords over time. Yet, WireVis does not support the detailed analysis of single accounts without clustering a set of accounts by their similar keywords usage. This is the most similar approach to EVA. However, instead of focusing on hierarchical analysis of keywords patterns within the transactions, EVA enables a broader and more flexible analysis. A deeper comparison with our approach is provided in Section 5.1.1. A first financial data flow is presented by [34]. In this approach, data are aggregated in order to allow users to draw analytical conclusions and make transaction decisions. EventFlow [28] was designed to facilitate analysis, query, and data transformation of temporal event datasets. The goal of this work is to create aggregated data representations to track entities and the events related to them. When looking at approaches for event monitoring in general, Huang et al. [10] presented a VA framework for stock market security. In order to reduce the number of false alarms produced by traditional AI techniques, this work presents a visualization approach combining a 3D tree map for market performance analysis and a node-link diagram for network analysis. Dilla et al. [6], presented the current needs in FFD. The authors presented a theoretical framework to predict when and how the investigators should apply VA techniques. They evaluated various visualization techniques and derived which visualizations support different cognitive processes. In addition, the authors also suggest future challenges in this research area and discuss the efficacy of interactive data visualization for fraud detection, which we used as a starting point for our approach.

Carminati et al. [3] presented a semi-supervised online banking fraud analysis and decision support based on profile generation and analysis. While this approach provides no visual support for fraud analysis, it is directly related to our approach since we are also focusing on profile analysis. However, we believe that VA methods have great potential to foster the investigation of the data and enable the analyst to better fine-tune the scoring system.

In the health domain, Rind et al. [33] conducted a survey study focusing on information visualization systems for exploring and querying electronic health records. Moreover, Wagner et al. [36] presented a systematic overview and categorization of malware visualization systems from a VA perspective. Both domains of these studies are similar to FFD, since they both involve multivariate and temporal aspects. However, the FFD domain demands for special consideration due to the complexity of the involved tasks (see Section 3.2).

3 Financial Fraud Detection

We developed our prototype called EVA (Event detection with Visual Analytics) in tight collaboration with a national bank institution [8] with the aim to improve and support their current FFD techniques.

In this section, we (1) describe the characteristics of transaction data, (2) discuss the complexity of the problem at hand, (3) present the currently used methodology for FFD at the bank, and (4) sketch EVA's scoring approach.

3.1 Transaction Data

We use an anonymized data set of real money transactions from our collaborating bank. This data set contains all transactions (e.g., payments, money transfers) executed or received by one of its customers within a given time period. Each transaction event is composed by several categorical, numerical, geospatial, and temporal dimensions. Some

examples are: sender/receiver, amount of money, location, and time of execution. The combination of these different aspects of data results in complex analysis scenarios that require the combination of different techniques in order to be tackled. More details concerning the data set used during development and evaluation are given in Section 5.

3.2 Problem Complexity

Automated FFD techniques are suited for well-defined problems and scenarios where the investigator knows exactly which patterns he/she is examining. However, the majority of fraudulent cases are not easily predictable by common rules and require some human investigation. Consequently, new methods such as VA are needed for these ill-defined problems. Besides the complexity that comes with the multivariate data set, there are several additional aspects that add up to the complexity of FFD.

Scalability. Financial institutions execute hundreds of thousands of transactions per day. To validate the veracity of all these transactions requires visually and analytically scalable solutions [22].

Context complexity. To better understand frauds, we need to consider the motivations that guide this criminal act. It is known that geopolitical, social, and economical contexts influence this criminal behavior [12]. Considering the ever changing local and global scenarios, FFD techniques need to be adapted frequently.

Frequent Changes. Not only there are many different types of frauds, but new ones are constantly being created and old ones are constantly being adapted in order to hide from current detection mechanisms.

False Positives. For each transaction that is flagged as suspicious by the automatic system, an investigator has to decide if the accusation of fraudulent behavior is correct, or not. Depending on the fraudulent classification (i.e., in case of money laundering suspicion), the owner account is then sued. To bring the accusation to court, involves professionals and costs a lot of money. This means that as the levels of positive alarms increase, the bank wastes money and, also, looses customers. Besides, even if identified during the process, false positive alarms overloads the investigators and waste their time of analysis.

False Negatives. Frauds that are neither detected by the automatic scoring system nor by investigators produce a financial damage to the bank and impact its clients' safety. They also impact the trustworthiness of the institution. Moreover, false negatives overlook actual recurrent frauds and, by consequence, result in fraudulent harm [23]. In other words, in order to be more helpful than harmful, the solutions need to be precise in estimating possible threats.

Fraud Classification. Fraudulent techniques are constantly being updated and reinvented. The definition of a set of features that classify fraud techniques is a difficult task which increases with the amount and complexity of data dimensions.

Time-Oriented Analysis. FFD not only requires the identification of temporal outliers, but also of periodic behavior (e.g., disguising fraudulent transactions as monthly bill payments). If well planned, frauds can avoid automatic algorithms detection. Thus, synchronous and asynchronous temporal aspects should be observed during analysis. However, due to its complexity [1], there are many aspects of temporal data that need to be analyzed efficiently.

3.3 Methodology for FFD

In this subsection we give an overview of the methodology that is used for FFD by our collaborating bank. Since we are using real data that is quite sensitive, we need to respect privacy and security regulations. Thus, we are not allowed to get into details about the actual algorithms. However, we roughly sketch the four phases of the FFD methodology applied: Profile Generation System, Scoring System, Results Interpretation, and Fraud Validation.

Profile Generation System. For each customer account the automatic system for FFD generates profiles based on this account's transaction history (see Figure 2 A.1, A.2, and A.3). A single account can have several profiles which reflect different aspects, for instance, separate sender and receiver profiles for one account. The result of this profile construction is then used for further classification. Profile generation is not a phase that is sequentially linked with the other phases. It has its own rules of execution. The bank can define a period of time for when it has to be executed (every week), or after a certain amount of events (after 100 transactions).

Scoring System. The system compares each of the incoming transactions (see Figure 2 B.1) with the corresponding customer's profile. To this end, it uses metrics to compute several different scores that are summarized in one overall score (see Figure 2 B.2); a single float number that represents how suspicious a transaction is. For example, each time a customer makes a new transaction, the FFD system automatically compares this transaction with the customer's profile of past transactions in order to compute a score that flags this transaction as either suspicious (possible fraud) or not suspicious. The higher the score, the more suspicious is the transaction (i.e., different aspects of the transaction that influence the score).

Results Interpretation. After calculating the scores, the non-automatic part of the investigation takes place. In this phase, investigators analyze multiple transactions simultaneously, due to time constraints. Transactions whose scores exceed a given threshold are further filtered by predefined rules. For example, all transactions below 20 euros are excluded from the list of suspicious transactions, since the amount is too low. The remaining transactions are then manually explored with the help of spreadsheet tools. During this exploration, investigators use their personal experience to decide whether a transaction should be considered fraudulent or not (see Figure 2 D).

Fraud Validation. Once investigators have decided that a suspicious transaction is possibly fraudulent, they call the account owner to ask him/her about the transaction's veracity. The bank stops the transaction in case the account owner did not authorize it.

We incorporated a VA component into the described work flow to tackle the complexity of fraud analyses (compare Section 3.2). The new process is illustrated in Figure 2.

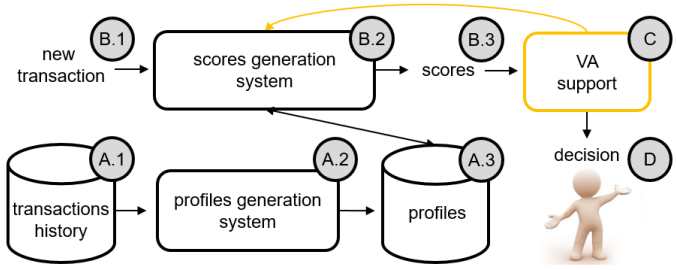

Fig. 2. The transaction evaluation system. The newly added interactive VA approach for investigating suspicious behavior and for evaluating the scoring system is highlighted in orange.

3.3.1 Types of Fraud

Different scores can be constructed for representing different types of frauds. There are more definitions of fraud types and their subtypes than the ones described in this section. However, we opt for describing the ones that mattered most to our collaborators.

Money Laundering. The main goal of money laundering is to transform profit gained from crime and corruption into 'legal' money. Usually, this type of fraud is composed of a number of events involving a network of accounts. Thus, the computation and analysis of scores needs to include network analysis for detecting this type of fraud.

Unauthorized Transaction. This type of fraud involves transactions from the account of a customer of a financial institution made by a non-authorized user with the aim of financial profit. This fraud is usually detected by profile analysis, i.e., comparing this transactions to other transactions usually done by this customer. Uncommon transactions receive high scores and need to be further investigated.

Embezzlement. In this type of fraud a criminal person misappropriates the money entrusted to him/her. This fraud may happen in the public or private sector and it is usually considered an internal fraud. To detect this fraudulent behaviour, scores need to consider transaction flows. This is usually done based on records from management software (e.g., log-files).

Straw Person. This type of fraud is sometimes related to money laundering. It describes a person A who receives money instead of person B, because B is not legally allowed to receive this money. In order to detect this type of fraud, scores need to consider customer profiles and detect outliers.

During the design and development of EVA, we focused on detecting "unauthorized transactions". Since EVA uses scores for decision support, an extension of these scores to detect other types of frauds would be an appropriate way to perform other types of fraud analyses.

3.4 EVA's Scoring Approach

Our profile-based algorithm is a self-adaptive, histogram-based approach according to the (mandatory) guidelines from the European Banking Authority (EBA) and monitors the behavior of internet originated payments. The proposed algorithm computes individual customer profiles, which are created on basis of historical transactions. These profiles are used to score new transactions in real-time. Thus, depending on the relative deviation of the score from the profile's standard range, the payment might be classified as suspicious.

Due to privacy and security regulations of our collaboration partners of the bank, we are not allowed to describe our profile-based algorithm in detail. However, our approach is comparable with Carminati et al. [3], who generate customer profiles in a semi-supervised way and provide different kinds of statistical analysis. As a result, their approach correctly ranks complex transactions as suspicious.

We evaluated EVA's profile-based algorithm on a representative sample of internet-based transactions consisting of 13 million payments ranging over a period of 15 months (1.1.2015 - 31.3.2016). To create the customer profiles the transactions of the first 12 months were taken (year 2015, 11.9 mill). For scoring - and consequently for evaluating - the remaining transactions of 2016 (about 1.1 million single transactions) were used. 24 transactions - out of this 1.1 mill - were flagged as confirmed frauds. Furthermore, as a reference system, an amount-threshold-based strategy was implemented, thus simulating the detection rules previously used.

Our profile-based approach documented a good performance and outperformed the threshold-based previously used strategy by far. For example: Taking the current number of confirmed fraudulent transactions identified by the threshold-based strategy as the constraint to be met, our approach found 500% more confirmed fraudulent transactions, and thus, preventing 86% of all potential amount losses. In a statistical analysis, applying Receiver Operating Characteristic (ROC) curve and calculating the Area Under the Curve (AUC) on both approaches, the profile-based approach produced an AUC of 0.944 while the threshold-based approach demonstrated less efficiency with an AUC of 0.78.

4 EVA's Design and Implementation

In the design phase of EVA we collaborated with two domain experts from the bank institution (referred as "collaborators"). Following the design triangle [27], to generate interactive VA methods we designed EVA with respect to the data, users, and tasks at hand.

Data. Financial transaction events constitute multivariate and time-oriented data which include details about the transactions such as amount, time, receiver, etc.

Users. Investigators from financial institutions that investigate and validate transactions alerts.

Tasks. The overall tasks are fraud detection by means of profile analysis. This task includes the reduction of false negative and false positive alarms, history comparison, as well as the manual investigation of suspicious transactions.

4.1 Requirements

When looking at currently used FFD solutions, there are still many opportunities for improvements. Instead of running queries in spreadsheets and judging alarms by a single overall score value, we propose EVA to support investigators during their decision-making process. From the study of related work and in collaboration with our project partners, we derived the follow requirements:

R1: Visual Support for Scoring System. Considering the constantly changing fraudulent behaviour, the scoring system and the profiling systems should be in constant evaluation. They should be frequently updated in order to stay effective. In the current system, investigators are not able to explore which transaction features and which sub-scores influenced the overall score to what extent. This information would be beneficial for understanding the construction of scores and deciding if the algorithm needs to be adapted. Moreover, investigators should be able to compare single transactions and their scores with the client's history of transactions.

R2: Account Comparison. Another important task in order to better understand suspicious events is to analyze the relationship between two accounts (i.e., their money exchanging behaviour). However, currently, this task is not supported besides the manual analysis of the two separate accounts by means of spreadsheets. Our solution needs to support the analysis of money exchange relationships of different accounts to enable the user to analyze and detect fraudulent collaborators.

R3: Reasoning About Potential Frauds. During the fraud validation phase (see Section 3.3), the investigator has to decide if a transaction flagged as suspicious is going to be confirmed as being fraud or not. To aid this task, our system needs to provide visual means to support the investigation of the automatically computed results. The system needs to allow for visually analyzing flagged transactions in contrast with non-flagged transactions, and thus, support the identification of false-positively flagged transactions.

R4: Identification of Hidden Frauds. Due to the data complexity (see Section 3.2), automatic methods such as the one used in our approach are not fully accurate. This can lead investigators to overlook fraudulent transactions that were not detected by the automatic system. In order to better support this task, our solution needs to make similarities between flagged and non-flagged transactions visible during the validation phase (see Section 3.3). Thus, the system needs to facilitate the identification of false-negative frauds.

4.2 Event Detection with Visual Analytics (EVA)

Following a user-centric iterative design process [29] we had regular meetings with our collaborators (about two hours each other month for one and a half years). We discussed the data, users, and tasks at hand in order to gain a thorough understanding of the problem and we designed a number of prototypes, ranging from low-fidelity mock-ups to interactive prototypes. Some design ideas we had to discard while others were iteractively refined and integrated into the final prototype. EVA is composed of six views displaying different aspects of the data (see Figure 1). All views are connected via brushing and linking (i.e., multiple coordinated views). EVA was developed as a web application by using Angular and D3.js technologies.

After some discussions with our collaborators, we opted for simple and well-known visualizations they are mostly familiar with, such as bar charts, line charts, and scatterplots [5, 6]. The goal was to

keep the learning curve for investigators as low as possible, and thus, foster acceptance of the system. However, since investigators are used to exploring data by using spreadsheets, we also provide a table representation in Figure 4, **F** which allows for assessing all details of the underlying data.

The visualization techniques were chosen with respect to the suitability of their visual attributes (e.g., element position, length, angle, color) to effectively and accurately encode the data types at hand [24]. In particular, we chose different visual encodings to achieve the best possible balance between distinguishability, separability, and pop-out of important information.

A.1, A.2: Temporal Views. Both views represent the temporal dimension of transactions. In both views, time is laid out on the x-axis, while the y-axis represents the total amount of money transacted per day. Thus, in **A.1** (see Figure 4), each bar represents a day. Days that contain at least one suspicious transaction are highlighted in red. **A.2** (see Figure 4) serves as an overview visualization of the inspected time period and as an interactive temporal filter. View **A.1** is tightly linked to view **A.2**. When brushing an interval in **A.2**, view **A.1** zooms in or zooms out to this interval. Selected periods of time are propagated to the other views for the analysis of other data characteristics. In particular, **A.1** shows a more detailed temporal representation of the selected period, while A.2 preserves an overview of the whole time interval.

We opted for a line chart in **A.2** because this representation of time series data is well-known to the investigators. Moreover, our data at hand contains daily sums of money transacted, which are usually quite stable. However, there are days with unusual high amounts of money transacted and these are the interesting days to identify and investigate, since high amounts of money may indicate frauds. Line charts allow for efficiently identifying such peaks even in very long time series. Alternative representations of time series data, such as calendar heatmaps would have been feasible too, but they bring some weaknesses about. These weaknesses include requiring more space, a more complex selection of temporal intervals, and a less intuitive visual encoding of the daily sum of money by color. For a more fine grained selection of days, we provide selectable bars in **A.1**. View **A.1** also represents the daily amount of money transacted on a temporal x-axis. However, daily amounts are represented by bars. In contrast to View **A.2** we decided for using bars in View **A.1** since investigators need to select suspicious days to investigate them in detail. For such a selection bars offer self-contained bodies each representing a single day, which can be more easily selected than regions on a line chart. Moreover, bar charts foster the accurate perception of the data by using bar length to encode quantitative information, which is accurately perceived [24] and thus, presents a suitable visual encoding for this type of data. Discussions with our collaborators also showed that selecting one or multiple bars in a bar chart presented a simple and intuitive way of filtering the data.

B: Score Construction View. We use parallel coordinates to present a visual history of transactions where each line represents a transaction of the selected account (see Figure 4, view **B**). Transactions whose overall score exceeds a given threshold are considered suspicious and are highlighted in red. Besides the "overallScore" axis at the very left (which was an explicit request of our collaborators), all axes represent sub-scores computed by the automatic system that are used for constructing the overall score. Thus, this view supports investigators to understand how overall scores were constructed by the automatic system and how the score of each transaction fits into the overall scoring scope. This view supports filtering by brushing any set of axis and these filters are reflected in all the other views. Moreover, selections in any other view also filter the transactions displayed in this view. The Score Construction view **B** highlights the selected data while graying out other transactions. This feature allows investigators to keep the context of filtered transactions.

Parallel coordinates are well suited to represent multiple dimensions side by side, which makes them a rational choice for representing the different sub-scores that contribute to the overall suspiciousness score of transactions. Although the investigators were not familiar with parallel coordinates we still decided to use them for various reasons. We needed to provide a visualization that enables investigators to identify sub-scores with strong influences on the overall score. They also need to identify groups of transactions with similar sub-scores in order to better understand fraudulent patterns. In a previous version of the prototype, we used a scatterplot matrix to show these relations. However, this visualization technique confused our collaborators, while the parallel coordinates were well perceived (see Section 5). The sactterplot matrix provided too many scatterplots that failed to give an effective overview and only allowed for analysis of pairwise relations between sub-scores. All sub-scores of one transaction could only be related by brushing and linking the dots in the different scatterplots, while parallel coordinates represent transactions by lines and the connection of all scores of such a transaction can easily be spotted. Thus, parallel coordinates facilitate comparing and relating these scores when reasoning why some transactions scored high or low. In addition, representing transactions by lines instead of dots in separate scatterplots also facilitates the identification of groups of transactions with similar patterns.

C: Amount vs Overall Score Scatterplot. In this view each dot represents a transaction. The overall score is encoded on the x-axis, while the amount of money exchanged is encoded on the y-axis. Thus, clusters of dots represent transactions with similar characteristics in these two dimensions, while outliers indicate uncommon transactions. Since investigators are interested in identifying outliers in contrast to clusters of normal transactions, overplotting in regions of normal transactions does not present a problem.

We decided for a scatterplot since it most efficiently encodes the relations between two variables. Using a scatterplot allows investigators to select a group of transactions (dots) according to the amount of money transferred and their overall suspiciousness score by area brushing. This supports the analysis of the relation of two of the most important dimensions for fraud detection: investigators emphasized that the amount of money is always a good place to start the investigation since small amounts of money are not of interest to them; combining this information with the overall score of a transaction facilitates the identification of cases that require further investigation. On the other hand, also transactions with high amounts of money that did not score very high are easily identified as outliers in this scatterplot and may hint at false negative cases.

D.1, D.2: Ranks. For analyzing money exchange relationships among clients, we provide two bar charts. These visualizations are utilized to represent who received money from transactions of the selected account. View **D.1** shows the rank of the top 10 receivers that received the biggest amount of money from the currently selected account, while the bar length encodes the sum of money received (see Figure 4). In **D.2**, we display the top 10 accounts which received money most frequently from the selected account, and thus, the bar length encodes the number of transactions received. Both types of information are important since frequently transferring money to the same receiver can hint at a fraudulent pattern, as well as transferring high amounts of money to one receiver. Investigators can select different bars in these two views to filter the data in all other views to show only transactions to the selected receivers. This way investigators can detect temporal patterns (e.g., frequent transactions to a specific receiver), analyze the history of transactions to this receiver, how they were ranked by the automatic scoring system, and drill-down into money exchange details by means of the Dynamic Table **F**. We decided for using bar charts to represent this information since they give a good overview of the ranking relationship of different receivers (i.e., very frequent receivers are emphasized by both, position at the top of the chart and bar length). Moreover, bars again allow for easy selection of interesting receivers for filtering and further investigation.

E: Accounts Selector. When investigating more than one account, this view facilitates comparison and switching between accounts. The bar length represents the amount of transactions that each account executed, which already facilitates the selection of accounts of interest. By selecting a bar, investigators filter the other views to show only data of the selected account. This functionality can be used for comparing a small group of accounts in more detail.

F: Dynamic Table. Currently, investigators are used to apply queries within spreadsheets in order to find insights. Besides providing a good amount of details, tables hinder pattern recognition and scale badly. However, tables are known and appreciated by investigators and thus, we provide an interactive table view in addition to the other views. Each row represents one transaction and each column one of its dimensions. Filters and selections in other views are automatically reflected by the table view and the other way around. Moreover, it is possible to sort rows by column values and to execute manual search queries.

Multiple Connected Data Perspectives. Since transaction logs are composed of multiple heterogeneous dimensions that need to be analyzed in relation to each other, EVA provides multiple perspectives on the data in multiple connected views. This set of views presents a variety of abstraction levels of the same subset of the data. In all views that represent transactions, we chose a colorblind-safe color encoding [9] to indicate transactions flagged as suspicious. Using the color red makes these suspicious transactions stand out immediately.

5 Evaluation

To assess the usefulness of EVA, we conducted a qualitative evaluation which aimed to address the following research questions (RQ):

RQ1: Comparison. What are the advantages and disadvantages of EVA compared to the tools which users usually use?

RQ2: Insights. What kind of insights can be generated with EVA?

RQ3: Improvements. Do users miss any features or have suggestions for improvement?

We decided for a qualitative study because it allowed us to get users' feedback and to understand insights they gained while using EVA.

Sample. We recruited three target users of EVA, i.e. FFD investigators, from our collaborating bank, who were not involved in the design process and have never seen our prototype before. Although the number of participants was low, qualitative evaluation studies are still useful to understand if the approach is useful for domain experts and if it fits their workflow [11, 21]. All three male participants had basic knowledge of working with visualizations. They usually use visualizations for presentation purposes to show the key message and the structure of the data. However, one participant also noted that he uses visualizations for exploration tasks (e.g., to analyze algorithms via heatmaps). For fraud detection tasks, they primarily use rule-based management systems which provide mainly spreadsheet representations including the automatic generated scores for each transaction.

Dataset and Tasks. We used an anonymized real world dataset from our collaborators covering an interval from January 2013 to April 2015. The dataset consists of 413 different accounts with a total of 1,128,147 transactions of different types (e.g., netbanking transactions). These tasks are structured according to the analytical task taxonomy by Andrienko and Andrienko [2], distinguishing elementary and synoptic tasks. In order to evaluate our solutions with respect to our requirements, we have defined a list of typical tasks together with two collaborating domain experts. Each task was designed with a specific focus on one or more requirements (see Figure 3). Requirements such as interactivity **(R2)**, data conservancy **(R3)**, and visual scalability **(R5)** were considered in all tasks.

Task 1: Explore the top three frequent receivers from the account acc10407 during the period of January 2014 to April 2014.

Task 2: Explore the transactions of account acc10421 and find the reason about the scoring of all transactions that happened on day(s) where fraud(s) were detected.

Task 3: Analyze two fraudulent accounts (acc10407 and acc10421) with respect to their similarities and differences in their fraudulent behavior.

	R1	R2	R3	R4
Task 1		•	•	
Task 2	•		•	•
Task 3		•	•	•

Fig. 3. This table shows the relation between tasks and requirements in our evaluation.

Procedure. The study took place in a quiet meeting room at the bank's head office. In addition to the respective participant, one test moderator, one observer for taking notes, and one developer as contact person for technical questions were present in the room. Furthermore, audio recording and screen capturing software was used. The test session began with a short introduction of the goal and the schedule of the study. Next, EVA was presented and participants had the possibility to ask questions to clarify any issues. We then conducted a semi-structured interview with the participants in order to learn about their experience regarding visualizations and which tools they usually use to solve their fraud detection tasks. After the interview, the participants were asked to interact with EVA in order to fulfill the three tasks outlined above. While the participants interacted with the prototype, they were encouraged to think aloud. After they finished the tasks, again a semi-structured interview was conducted. They were asked about their impressions of EVA, if they missed anything in particular, to compare the prototype with the tools they would typically use for fraud detection, and if there were any further tasks which they would like to solve with this kind of VA tool.

Data Analysis. The collected qualitative data (observation and interview notes as well as the audio and video recordings) were analyzed in order to find out what works well, what needs further improvement, and what are possible missing features (cf. research questions RQ1 and RQ3). However, we were also interested in which kinds of insights they gained with EVA while they solved the tasks (cf. RQ2). EVA supports processes of exploration and sensemaking. There are two well-known approaches explaining sensemaking with visualization - the model by Pirolli and Card [30] and Klein's sensemaking model (see also [17,18]). The model by Pirolli and Card has been criticized because it applies only to a very narrow range of activities of intelligence analysis [31], while Klein's model is much broader. Therefore, Klein's categories were chosen for this analysis. Thus, we adapted the five categories from Klein [16] for gaining insights:

Connection. These insights result from a connection between two or more events which provides new information. For example, two visualizations present the same data set from different viewpoints. The combination of these visualizations allows the viewer to get additional detail information about the data.

Coincidence. Coincidence insights result from events which seem related but do not have an obvious connection. In contrast to the connection insights a coincidence insight results from repetition and not from detail information. For example, if data points have the same value in the visualization then this can be a result from a specific event.

Curiosity. These insights differ in one way from the coincidence insights: it results from a single event. For example, a data point with a specific value in a visualization arouses the interest of the viewer.

Contradiction. Such insights often occur if there is discrepancy between events which causes doubts. For example, a data point in the visualization has an unrealistic value.

Creative Desperation. These insights result from events which tend to be a dead-end and require finding new ways. For example, if it is not

possible to get relevant information with a specific type of visualization then another visualization type might be helpful.

Based on these categories, the observation notes as well as audio and video recordings were coded and categorized.

5.1 Results

All participants solved all tasks. The average duration needed for the tasks was about 18 minutes. The interview sessions (before and after the participants solved the tasks) took about 40 minutes in total. Next, we will present and discuss the results according to our research questions.

5.1.1 RQ1: Comparison

The investigators stated that they typically use visualizations for presentation tasks which they typically generate with Microsoft Office tools (e.g., Excel and PowerPoint) [25, 26]. Therefore, they argued that it is difficult to compare EVA with these tools. The challenge in using these tools is to find the interesting hot spots. All three investigators agreed that a powerful visual tool for exploration tasks would be helpful for browsing the data, and for gaining insights which they were not even looking for, or for giving them hints to look at specific parts in the data set more closely. They highlighted that EVA was very intuitive (e.g., color-coding), easy to use, more dynamic than the tools which they usually use and that it provides a good overview and quick access to detail information. For example, one expert mentioned: *"I liked it because it is very interactive and you can browse the data, even if you don't know what you are looking for, and find new insights"*. Furthermore, the usage of EVA led to a positive attitude towards using VA approaches in the future. For example, one expert noted: *"Here we could see what is possible [...] So we can rethink what we can offer to the bank."*.

Since, the investigators did not know any visual approaches to support FFD, they could not compare EVA to actual VA approaches. However, if we compare EVA, for instance, to WireVis [4], a state of the art VA approach focusing on FFD, EVA takes more aspects into account in order to identify fraudulent behaviour. While WireVis is used to analyze keywords used in transaction descriptions with a focus on detecting money laundering, EVA is aimed at detecting unauthorized transactions by analyzing a variety of objective aspects about transactions (e.g., amount, date and time, frequency, etc.). Thus, EVA presents a broader approach that is in line with existing FFD mining techniques. Moreover, EVA allows for a deeper exploration of multiple aspects of transactions as well as reasoning about how they influenced the automatically generated scores for fraud detection. Instead of analyzing keywords, our FFD approach constructs individual profiles for each account and computes suspiciousness scores that indicate how unusual the transaction is, given the history of this account. This supports the detection of different types of fraudulent transactions which would not be possible from keyword analysis alone.

5.1.2 RQ2: Insights

In total we found 77 insights. Most of these insights were connection insights (53.2%). Coincidence insights contributed 26%. Curiosity insights (6.5%) and contradiction insights (9.1%) played a marginal role. Creative desperation insights also only showed up in 5.2% of the cases. Most insights (35 insights) were found for the **Task 3**, followed by **Task 2** with 24 insights and **Task 1** with 18 insights. The number of found insights correlated with increasing task complexity. For example, **Task 1** focused on the identification of specific values in order to offer the investigators an easy start with EVA whereas **Task 3** allowed for more data exploration since it required to analyze and compare two accounts in order to find their similarities and differences. Next, we will discuss each insight category in more detail.

Connection. In total, 41 connection insights were found and two types of connection insights were identified. Connection insights from the first type resulted from a connection between the different views of EVA. For example, one expert compared the Dynamic Table view (see Figure 1, F) with the Score Construction view (see Figure 1, B) to detect connections between the amount and the scores of the suspicious transactions. Connection insights from the second type, resulted from a connection between the different variables. For example, one expert noted: *"country code and daily count are suspicious since unusual many transactions were made from this foreign country"*. In total, they derived slightly more connection insights from the views (53.7%) than from the variables (46.3%). These results show that the different views helped investigators to analyze the data from different viewpoints. However, also the comparison of the different variables played a role in finding insights.

Coincidence. From the 77 found insights 20 were coincidence insights. These insights resulted from comparing values of the same variables. For example, one expert noted: *"the chance is high that these both transactions are also fraudulent since the receiver has already a confirmed fraudulent transaction"*. Or another investigator mentioned: *"when you detect one fraud in this case, you can detect all frauds because they were all made by only one person"*. Although we found more connection insights than coincidence insights during **Task 1** (15 versus 2 insights) and during **Task 2** (13 versus 4 insights), slightly more coincidence insights than connection insights were gained during **Task 3** (13 versus 14 insights). It seems that the differences between **Task 3** and the other two tasks arose from the comparison of the two accounts in **Task 3**. For example, we observed that the investigators compared the values of the variables separately for each account and next compared these between the accounts to detect similar behaviour.

Curiosity. We found 5 curiosity insights. These insights resulted from investigators' observations which stimulated their interest to explore the data further. For example, although one expert had already detected fraud cases with the bar chart visualization in the Time Panel (see Figure 1, A.1), he interacted further with the time slider from the area visualization (see Figure 1, A.2) in order to detect further possible cases. Curiosity insights only occurred during the last two tasks which were more exploratory in nature than the first one (**Task 2**: 3 insights and **Task 3**: 2 insights).

Contradiction. In total, 7 insights were contradiction insights. The contradiction insights arose from conflicts and doubts in their own observations but also in EVA. For example, one expert explained his decision not only to select the suspicious cases automatically marked from EVA: *"I would also select the surrounding in the scatterplot - because maybe the system did not detect all fraudulent cases"*. Most contradiction insights were found during **Task 3** (5 insights). It seems that comparing two accounts added to the complexity of Task 3. This showed that the participants had less confidence in their own observations. Thus, we plan to find solutions to minimize users' doubts in the future, for example, by directly highlighting the differences between accounts.

Creative Desperation. Only 4 creative desperation insights were found. These insights resulted from revising their own, previously phrased interpretations or from finding alternative ways when the desired interaction or view was not available. For example, one expert assumed that the interaction technique linking and brushing is not possible between the Dynamic Table view (see Figure 1, F) and the Score Construction view (see Figure 1, B) since he saw no changes between the two entries in the parallel coordinates after he selected them in the table. After he tried a third entry, he realized that the first two entries had the same values and hence there were no visual differences in view B.

5.1.3 RQ3: Improvements

The investigators made useful suggestions for possible new features and improvements. One suggestion was to expand the filter and selection functionality. For example, all investigators noted that filter options especially for the table representation (e.g., only to show transactions with a certain amount) would be helpful. One expert highlighted that he would also like to directly select suspicious cases in the Time Panel by selecting bars instead of using the temporal filter (see Figure 1, A.1). Another suggestion was to provide the possibility to put suspicious receivers on a black list to stop and investigate all transactions to these receivers. Furthermore, one expert suggested to have a simulation feature to being able to play around with the composition, weights, and thresholds of sub-scores and see how it affects the overall score in order

to optimize the scoring algorithm.

However, all three investigators propose to include: support of network analysis. Network visualization is highly interesting in the area of fraud detection in order to analyze the relations between accounts, receivers, and dimensions. Such a network visualization could help them to see the connection between suspicious transactions and other transactions as well as involved accounts.

In addition to the mentioned opportunities for improvement we observed several minor usability issues (e.g., labels were sometimes too small or they overlapped) which will be resolved in the next iteration of EVA.

6 Discussion

In this section, first, we discuss which features presented in EVA fulfill each of the defined requirements. Next, we illustrate the challenges and opportunities that we identified during the work process.

6.1 Requirements

The set of views presented in EVA supports a better understanding of the data by presenting it in different abstraction levels to the investigator. Identifying time-oriented aspects, analysing the score construction, as well as drill-down inspection are tasks that can be executed simultaneously in different views. EVA also presents fully responsive interaction techniques that allow a natural understanding of the relationships between the multiple coordinated views. Aiming to perform a more concise decision about an alarmed transaction, the investigator can explore data features, identify patterns, and evaluate scores by selecting and filtering the views. All interaction that excludes or include data into a view ensures the consistency of the data by also excluding, including, or rearranging data representations on the other views as well.

The time-oriented analysis is presented through the views A.1 and A.2 and through their link with the other views. Investigators can observe sending, receiving, scores, amount and others feature patterns over the time. Different periods of time can be specified during the analysis. The multivariate feature of transactions makes a complete visual encoding of all important characteristics within a single view impossible. By presenting multiple-coordinated views we present an overview of different features that our collaborators declared essential for their tasks. However, by presenting a dynamic table in addition to these multiple views, EVA allows detailed analysis of raw data features. Although EVA provides features to compare multiple customer accounts, it was primarily developed for the analysis of individual accounts. The simultaneous investigation of multiple accounts usually includes transactions of a time span from one to three years. However, due to national law restrictions, the maximum period of time that a bank is allowed to keep this transaction data is seven years. Thus, even for the most extreme outlier case present in our real world data (2,000 transactions per year), EVA scales fine.

In view B we can analyse how each overall score is constructed by others specific scores (**R1**). By observing how each line passes through the axis it is possible to identify which sub-scores influence most. In addition, once all transactions are visible at once in this view, EVA allows an analysis of the usual construction pattern for an account. Thus, when analyzing a single transaction, the investigator can compare the construction of it with the normal construction behaviour. Another feature is the range filters for any of the axis. By doing that, the investigator can judge if high/low sub-scores are being ignored, overestimated, or present correlation and, thus, help on evaluating the scoring system.

In order to aid multiple account comparison and analysis, view E present an interactive row chart that allows the investigators to filter each account's transactions (**R2**). By doing that, it is possible to compare patterns among the different views presented by the prototype.

Similarities and differences between suspicious and non-suspicious transactions become more evident during visual analysis. This can be done for single or multiple accounts. Thus, transactions that are wrongly flagged as suspicious (false positive cases) can be more easily identified when comparing them with other transactions from the same account (**R3**). On the other hand, false-negative cases are also more easily detected by the exploration of the suspicious patterns through visual means (**R4**). A false-negative case is illustrated and described in Figure 4. An efficiency way to perform both tasks is by using the Dynamic Table (F) in combination with the Score Construction View (B).

6.2 Solving Real-World Tasks with EVA

We chose the tasks for our evaluation session in order to reflect the investigators' real world tasks (we elaborated them together with two collaborating domain experts). In this section, we outline how an investigator solved his real-world tasks with EVA and the insights he derived during the evaluation session. While one investigator used EVA to solve Task 2, he examined account acc10421 (see Figure 4). Interactively investigating the different views of EVA, he used the Scores Construction View to filter out transactions with a low suspiciousness of the country the money was transferred to (see arrow in Figure 4, view B). For this filter selection only one transaction was not automatically flagged as suspicious (i.e., the gray element in Figure 4) and EVA shows several clues that indicate that the only non-flagged transaction might be fraudulent too and should at least be considered as suspicious. All transactions (including the non-flagged transaction) are going out to the same receiver (see arrow in Figure 4, view D.1), on the same date (see arrow in Figure 4, view A.2). In addition, the non-flagged transaction involves a high amount of money (see arrow in Figure 4, view C) and scored quite high in several sub-domains (see Figure 4, view B). However, the overall score was not high enough to flag this transaction as suspicious. After some more exploration the investigator confirmed this transaction as a false-negative case. Without the VA support of EVA, it would have been (nearly) impossible to spot this mistake of the automatic scoring system, which illustrates the benefits of a VA approach compared to a pure FFD approach. This insight led the investigators to actually fine-tune the automatic alert system (see Section 3.4).

6.3 Limitations & Further Work

Demands to detect, analyze, and monitor suspicious behavior are constantly increasing not only in FFD. Based on some of EVA's limitations, we present possible further work and open research challenges.

Network Analysis. Although our prototype shows promising results for investigating long time intervals of transactions and relating a small number of accounts, we do not support the investigation of networks of accounts yet. An interactive network visualization would allow investigators to better reason about suspicious money transfer relationships and patterns within their contexts [4].

New Customer Classification. When a new customer is added to a profile system, he/she does not have enough transactions to derive a reliable profile by EVA's profile generation algorithm (see Section 3.4). This makes it impossible for the scoring algorithm to detect fraudulent attempts from new accounts. There is a need for an exploratory VA environment which allows for the analysis of suspicious behavior of new customer accounts.

Knowledge Base Construction. One aspect that adds up to the complexity of fraud detection is that finding suitable solutions for detecting and deciding about suspicious cases is not enough [6]. Currently, investigators are using their experience to judge if a transaction is fraudulent or not. This can be tricky and results vary with investigators. We suggest the construction of a knowledge base based on former fraud detection that supports investigators to choose suitable scoring thresholds during analysis. For instance, by logging investigators' interaction data, we could collect filter set ups or filter combinations that obtained most success on detecting fraudulent behavior. This would not only support fine-tuning of the automatic scoring system but also ease the knowledge transfer to inexperienced investigators. Furthermore, it could also keep all investigators updated on new fraudulent discoveries.

Multiple Customers Monitoring. In our work, we can handle a small group of customers. Usually, these are accounts that were flagged as suspicious by the automatic scoring system. However, it should be possible to monitor all customers (or at least a big parts of them), which could lead to new insights. During evaluation (see

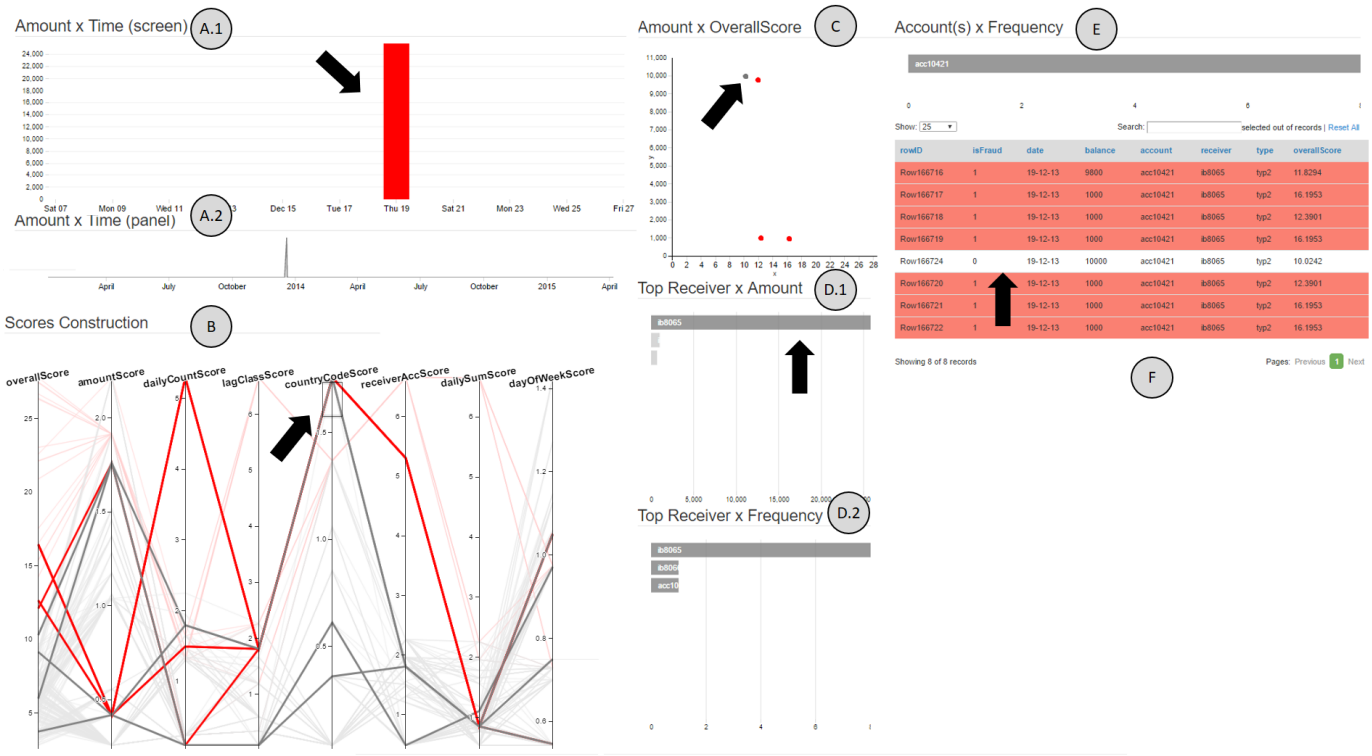

Fig. 4. False-negative case spotted during the evaluation of EVA. We composed this figure to show different steps in the analysis process. Black arrows indicate interesting insights that are discussed in subsection 6.2.

Section 5.1.3) investigators also suggested to keep a "black list" of fraudulent customers to block them as well as to use them as blueprints for the identification of new fraudulent attempts.

Fine-Tuning Fraud Detection Algorithms. EVA's Score Construction view already facilitates reasoning about how suspiciousness scores were constructed by sub-scores and which combinations of sub-scores are well suited as indicators. This feature also supports investigators in evaluating their algorithm (see Section 6.2). However, directly manipulating the weights of sub-scores for overall score construction is not supported by EVA at its current state. A visual interactive support for fine-tuning the algorithm would be a valuable addition to this work.

Different Types of Fraud. EVA was designed and implemented to meet the specific needs of the investigators at our collaborating financial institution, and thus, to tackle a specific kind of financial fraud, i.e., identifying and analyzing unauthorized transactions. VA support for detecting other types of fraud (see Section 3.3.1) is still needed, which prompts important research challenges for further work.

7 Conclusion

During the development of EVA we followed an iterative design over a period of 1.5 years in close collaboration with domain experts from a national bank. EVA follows the VA principles of interweaving intuitive interactive visualizations and analytical techniques in a seamless way. We selected our visualization techniques as well as the interaction techniques with special consideration of our design requirements, derived from discussions with our collaborating domain experts, who had limited experience with visual exploration tools. Therefore, we decided to use well-known interactive visualization techniques our experts are mostly familiar with. In the background we use the automatic computation of profile-based suspiciousness scores, which helps to monitor the behavior of costumers (in particular, payment transactions, which are characterized by multivariate and time-oriented aspects). We pursued an interactive multi-coordinated view approach, which took Tufte's [35] principles of graphical integrity into account. In particular, we obeyed Tufte's principle of *"show data variation, not design variation"* [35] (page 61). EVA eased the overall FFD process, which was enjoyable verbalized by one investigator during the evaluation sessions as *"...it is very interactive and you can browse the data, even if you don't know what you are looking for, and find new insights"*.

We evaluated EVA with real world data and could demonstrate that EVA was able to scale well even for extreme cases and to perform the required tasks in a suitable and appropriate way. The analysis of the insights, that were discovered the study participants, supported a more comprehensive understanding of the professional usage of EVA. Participants predominantly looked for connections in the data. They primarily got these connections from comparing different views, but also from comparing different variables. It was fascinating to observe that the number of connection insights decreased with the complexity of the task while other types of insights (coincidence, contradiction) emerged. Moreover, this real case of a overlooked fraudulent transaction was discovered during this evaluation study due to the indisputable benefits of visual exploration. We are aware that transferability from three study participants is limited, but nevertheless, this analysis indicates that future research in this area might lead to striking results concerning insight generation with VA approaches in relation to the complexity of tasks.

Based on our study, we also propose possible future research directions in the field. Since the tasks involved in FFD are similar in different event detection domains, our approach may be transferable to other domains too, such as malware risk analysis, health parameter monitoring, terrorist detection, and governmental fraud.

8 Acknowledgements

This work was supported by Centre for Visual Analytics Science and Technology CVAST (funded by Austrian Federal Ministry of Science, Research, and Economy in the exceptional Laura Bassi Centres of Excellence initiative, project number: 822746).

REFERENCES

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

[2] N. Andrienko and G. Andrienko. *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. doi: 10.1007/3-540-31190-4_3

[3] M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero. Banksealer: an online banking fraud analysis and decision support system. In *IFIP International Information Security Conference*, pp. 380–394. Springer, 2014.

[4] 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. VAST. IEEE Symposium on*, pp. 155–162. IEEE, 2007.

[5] W. S. Cleveland. Graphical methods for data presentation: Full scale breaks, dot charts, and multibased logging. *The American Statistician*, 38(4):270–280, 1984.

[6] W. N. Dilla and R. L. Raschke. Data visualization for fraud detection: Practice implications and a call for future research. *International Journal of Accounting Information Systems*, 16:1–22, 2015.

[7] M. Dumas, M. J. McGuffin, and V. L. Lemieux. Financevis.net - a visual survey of financial data visualizations. In *Poster Abstracts of IEEE VIS 2014*, November 2014. Poster and Extended Abstract.

[8] Erste Bank, Austria. Erste Group IT International. https://www.erstegroupit.com/en/home.

[9] M. Harrower and C. A. Brewer. Colorbrewer. org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37, 2003.

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

[11] T. Isenberg, P. Isenberg, J. Chen, M. Sedlmair, and T. Möller. A systematic review on the practice of evaluating visualization. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2818–2827, 2013.

[12] K. Jaishankar. *Cyber criminology: exploring internet crimes and criminal behavior*. CRC Press, 2011.

[13] D. A. Keim, F. Mansmann, J. Schneidewind, J. Thomas, and H. Ziegler. *Visual analytics: Scope and challenges*. Springer, 2008.

[14] J. Kielman, J. Thomas, and R. May. Foundations and frontiers in visual analytics. *Information Visualization*, 8(4):239, 2009.

[15] 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, 1999.

[16] G. Klein. *Seeing what others don't: The remarkable ways we gain insights*. PublicAffairs, 2013.

[17] G. Klein, B. Moon, and R. R. Hoffman. Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems*, 21(4):70–73, July 2006.

[18] G. Klein, B. Moon, and R. R. Hoffman. Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent Systems*, 21(5):88–92, Sept. 2006.

[19] S. Ko, I. Cho, S. Afzal, C. Yau, J. Chae, A. Malik, K. Beck, Y. Jang, W. Ribarsky, and D. S. Ebert. A survey on visual analysis approaches for financial data. In *Computer Graphics Forum*, vol. 35, pp. 599–617. Wiley Online Library, 2016.

[20] Y. Kou, C.-T. Lu, S. Sirwongwattana, and Y.-P. Huang. Survey of fraud detection techniques. In *Networking, sensing and control, 2004 IEEE international conference on*, vol. 2, pp. 749–754, 2004.

[21] S. Kriglstein and M. Pohl. Choosing the Right Sample? Experiences of Selecting Participants for Visualization Evaluation. In W. Aigner, P. Rosenthal, and C. Scheidegger, eds., *EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (EuroRV3)*. The Eurographics Association, 2015. doi: 10.2312/eurorv3.20151146

[22] R. A. Leite, T. Gschwandtner, S. Miksch, E. Gstrein, and J. Kuntner. Visual analytics for fraud detection and monitoring. In *Visual Analytics Science and Technology (VAST), 2015 IEEE Conference on*, pp. 201–202. IEEE, 2015.

[23] J. Luell. *Employee fraud detection under real world conditions*. PhD thesis, UNIVERSITY OF ZURICH, 2010.

[24] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions On Graphics (Tog)*, 5(2):110–141, 1986.

[25] Microsoft. Excel. office.microsoft.com/en-us/excel/ (accessed: 2016-12-09).

[26] Microsoft. Powerpoint. office.microsoft.com/en-us/powerpoint/ (accessed: 2016-12-09).

[27] 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.

[28] M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman. Temporal event sequence simplification. *IEEE transactions on visualization and computer graphics*, 19(12):2227–2236, 2013.

[29] T. Munzner. A nested model for visualization design and validation. *IEEE transactions on visualization and computer graphics*, 15(6):921–928, 2009.

[30] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*, pp. 2–4, 2005.

[31] M. Pohl, M. Smuc, and E. Mayr. The user puzzle—explaining the interaction with visual analytics systems. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2908–2916, Dec 2012.

[32] D. J. H. Richard J. Bolton. Statistical fraud detection: A review. *Statistical Science*, 17(3):235–249, 2002.

[33] A. Rind, T. D. Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant, B. Shneiderman, et al. Interactive information visualization to explore and query electronic health records. *Foundations and Trends® in Human–Computer Interaction*, 5(3):207–298, 2013.

[34] J. P. Steidlmayer and G. Kummel. Financial data event flow analysis system with study conductor display, Sept. 26 1995. US Patent 5,454,104.

[35] E. R. Tufte. *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CT, 2011.

[36] 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, eds., *EG Conference on Visualization (EuroVis) - STARs*, pp. 105–125. The Eurographics Association, 2015. doi: 10.2312/eurovisstar.20151114