MtVis: A visual analytics Framework for mobile money transactions analysis and exploration

Bocoum Ousmane[1],Yadong Wu[2]

[1,2] *School of computer science and Technology,Southwest University of science and Technology,China*

Abstract

Mobile money transfer systems (MMTS) are progressively becoming the standard banking system in countries with limited banking.The analysis of transactions performed on these systems helps to detect fraudulent and criminal activities.In this paper we introduce MtVis a novel visual analytics framework for mobile money transactions analysis and exploration.Our tool enables the exploratory analysis of mobile money transactions data using multiple views to reveal the temporal,geospacial and categorical aspects of the transactions. Through the process of implementing a fraud detection model for our system, we identified several challenges related to the given MMT datasets,we offer the principles used to overcome these challenges as a step towards understanding the difficult task of designing flexible and scalable fraud detection models.

Keywords: Visual analytics,Fraud detection,Financial data visualization,mobile money

1 Introduction

Mobile money transactions systems are being intensively deployed in countries with limited banking.Like any monetary system it is necessary to analyze transactions to detect criminal and fraudulent transactions such as money laundering and terrorism financing.Traditional methods has shown their limits with multi feature datasets like mobile transactions datasets,hence the need for more sophisticated methods.In an attempt to solve this problem In this paper we present MtVis a visual analytics system for mobile transactions analysis and exploration.We used a binary neural network classifier to perform the fraud detection.During the implementation of our model we identified several challenges specific to the given MMT datasets.We offer the principles used to overcome these challenges as a step towards understanding the difficult task of designing flexible and scalable fraud detection models.This paper covers the description of MtVis,the evaluation of the system by two domain experts and a brief discussion on how to design a suitable fraud detection model for datasets generated by Mobile money transactions services.

The rest of the paper is organized as follows. In Section 2, we briefly explain the related work.Section 3 discusses Fraud detection models;Section 4 presents the Visual Components of Mtvis.In the section 5 we present our case studies,section 6 describes the implementation of our system we finally conclude in the section 7.

2 Related Works

Numerous research works on visual analytics have been done for financial data. A large number of systems have been designed to help analysts perform efficient risk management and enterprise decision making. Our tool for mobile money transactions analysis is informed from the fields of financial data visualization and financial fraud detection.

2.1 Financial data analysis and visualization

With the growing concern of organizations to transform their data into valuable assets,a large number of visualization systems has been developed. A survey of these systems is presented by Sungahn et al(1) ,Chang et al.(2) presented wirevis a financial time varying data visual analytics,Didimo et al (3) introduced VisFan,a network visualization system for financial data.Netsuite(4) is a financial planning visual analytics for decision making .However Mobile money services are subsystems of the traditional banking system with their own characteristics and challenges.The two major operator of mobile money services are Mpesa(5) and Orange Money(6)deployed in Africa and Middle east regions.Works specific to MMTS are few due to their relative newness,Novikova et al. (7) worked on anomalous activity visualization in MMTS, and Gaber et al. (8) studied the behavior of users in MTTS environment.

2.2 Financial fraud detection

Fraud is the cause of major losses in financial systems,to overcome this problem various fraud detection models have been proposed. Abbassi et al.(9) designed an excellent framework for detecting financial Fraud using meta-learning.(Adedoyin et al.(10) used case based reasoning to predict fraud in mobile money transfer.Kappelin and jimmie(11) used mathematical and statistics to tackle the problem of detecting fraud in Mobile Money systems. Novikova et al. (7) used the radviz(16) visualization technique to detect anomalous activities in MMT environment,Didimo et al. (3) used network visualization to detect financial crime.Albashrawi et al.(12) did a review on fraud detection techniques using data mining.To provide researcher with experimental data, Lopez-Rojas et al.(13) designed Paysim,a financial mobile money simulator for fraud detection,this simulator can be used to simulate mobile transactions and generate data similar to the original dataset.

Differently from our system, all the works cited above focus either on the customer or the financial statement as entity,our system on the other hand focus on transactions as entity. The analysis of transactions as entities helps to understand the overall state of MMT systems and reveals valuable trends useful in the fight against financial crime and terrorism financing.

3 Proposed Model for Fraud Detection

3.1 Dataset description

Our dataset was collected from 12 months of activity from a mobile network operator,it contains more than 400000 transactions,for security and privacy issues the original dataset was processed to remove the sensible informations.The basic entity of our data is a transaction.Each transaction record is characterized by the following features

- Type: The type of the transaction performed,taken as value:Cash-in,Cash-out,Transfer,payment,debit
- Location: Localization of the transaction.
- Amount: Amount of money of the transaction
- NameOrigin: Name of the sender
- NameDest: Name of the receveir
- BeforeBalanceSender: balance of the sender before the transaction
- AfterBalanceSender: balance of the sender after the transaction
- BeforeBalanceReceiver: balance of the receiver before the transaction
- AfterBalanceReceiver: balance of the receiver after the transaction
- Time: Time of the transaction

Our dataset also contains benchmarked transactions reported to be fraudulent.Flaged by a fraud or non fraud feature.

However due to the scarcity of benchmarked mobile transactions datasets and their privacy nature we used paysim (12) to generate mobile transactions datasets based on our original dataset without changing our original features.

type	amount	nameOrig	oldbalanceOrig	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud
CASHOUT	134991.97	C576685194	0	0	C36322011	391909.49	526901.46	0
TRANSFER	37991.39	C26423367	13136	0	C1764084529	0	37991.39	0
TRANSFER	3845765.36	C196788126	3845765.36	0	C1000407130	0	0	1
PAYMENT	16871.11	C652001878	0	0	M849216230	0	0	0
CASHIN	252508.29	C736570021	1828637.42	2081145.71	C928094130	27284930.5	27032422.21	0

Table 1: Sample of data generated using paysim

3.2 Challenges in designing the proposed Fraud detection model

Fraud detection is basically a classification problem,given $X_1...n$ features we are trying to output a $y_1...n$ classes,but due to its features and the fact that fraud features can vary over time some algorithms are more suitable than others to perform this task.
And it requires various data preparation to obtain good results. In this section we briefly describe the challenges in designing a good model for fraud detection in MMTS and how to overcome them.

- Data Scarcity
  The lack of datasets on mobile money(15) transactions to perform research on in the domain of fraud detection is a big problem,to overcome this a simulator such as paysim(12) can be used to generate transactions data from a given input of original dataset.

- Imbalanced Data
  The big challenge(17) in modeling a fraud detection model as a classification problem is that the majority of real world transactions are not fraudulent.
  There are different ways of dealing with imbalanced data.among others:
  -SMOTE(14): Synthetic Minority Over-sampling Technique
  It consist of creating new synthetic samples by interpolating new points between marginal outliers and inliers.
  -Oversampling
  A common technique to deal with imbalanced data is oversampling(18).It consist of creating new observations in our data belonging to the under-represented class.
  -Undersampling
  It is the opposite of oversampling,it consist of reducing the observations of the dominant class.

- Choosing the right model
  Fraud detection being a classification problem(19),various classifiers can be used to solve it.
  However due to the fact that fraud features can change overtime,we found that some classifier are more suitable than others.
  For example using a decision trees classifier(20) and a naive bayesian classifier(21) we obtained good results on our dataset after performing feature engineering.
  We noticed that these models are not good compared to other models such as a binary neural classifier and Gated Recurrent Unit Neural Network due to their dependency to the features engineered.
  The neural classifiers are feature engineering free and more flexible to adapt to new fraud features.We therefore chose the Gated Recurrent Unit Neural Network to perform the fraud detection.

3.3 The proposed Gated Recurrent Unit Neural Network model

As explained in section 3.2 we used a GRU neural classifier(22) to implement the fraud detection module of our system.The model is characterized as follows:

- Model overview

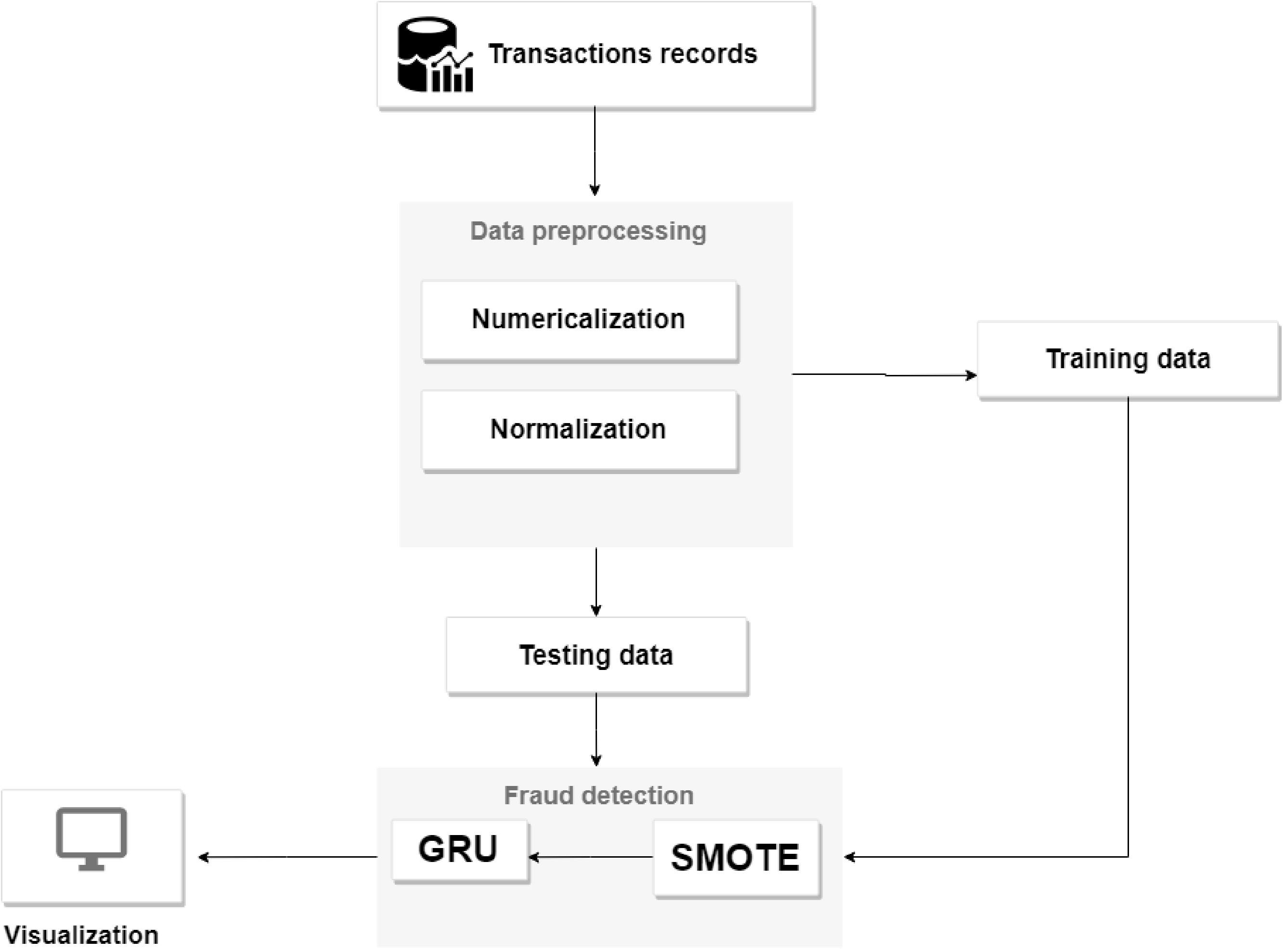


Figure 1: Fraud detection module

Figure 1 shows the flowchart of the proposed fraud detection model.First of all preprocessing operations such as numericalization and normalization are performed on the original dataset.The SMOTE algorithm is then used to process the low-frequency samples in the training dataset. The processed data is finally used to train the GRU network.

- Model parameters

  The parameters of the model used include the following:

  (a) The number of neurons in the input layer, output layer, and hidden layer of the GRU network and batch size during network training.

  (b) The parameters of Adam algorithm, including step size and first-order and second-order exponential damping decrement and nonzero constant.

  (c) The number of nearest neighbors and oversampling rate of the LA-SMOTE algorithm.

  The specific parameter values used in the experiments are shown in Table 2.

- Metrics
  To mesure the effectiveness of the model,evaluation indicators including ACC (accuracy), DR (detection rate), and FDR (false detection rate) were adopted to evaluate the detection performance of the proposed model and

Algorithm	Parameter	Value
SMOTE	Number of nearest neighbors	65
	Over-sampling rate	600%
GRU	Number of nodes in input layer	122
	Number of neurons in hidden layer	75
	Number of neurons in output layer	5
	Batch size	500
Adam	Step size	0.001
	First-order exponential damping decrement	0.9
	Second-order exponential damping decrement	0.999
	Non-zero constant	10 8

Table 2: Models parameters

compare it with other state-of-the-art intrusion detection models and methods.These evaluation indicators are obtained as follows:

$$ACC = \frac{TN + TP}{TN + TP + FN + FP}$$

(1)

$$DR = \frac{Tp}{TN + FN}$$

(2)

$$FDR = \frac{FP}{FP + TN}$$

(3)

where TP represents the true positive, which means normal samples predicted as correctly, TN represents the true negative, which means fraudulent samples correctly detected, FP represents the false positive, which means fraudulent samples classified as normal mistakenly, and FN represents false negative, which means normal samples predicted as fraudulent wrongly.

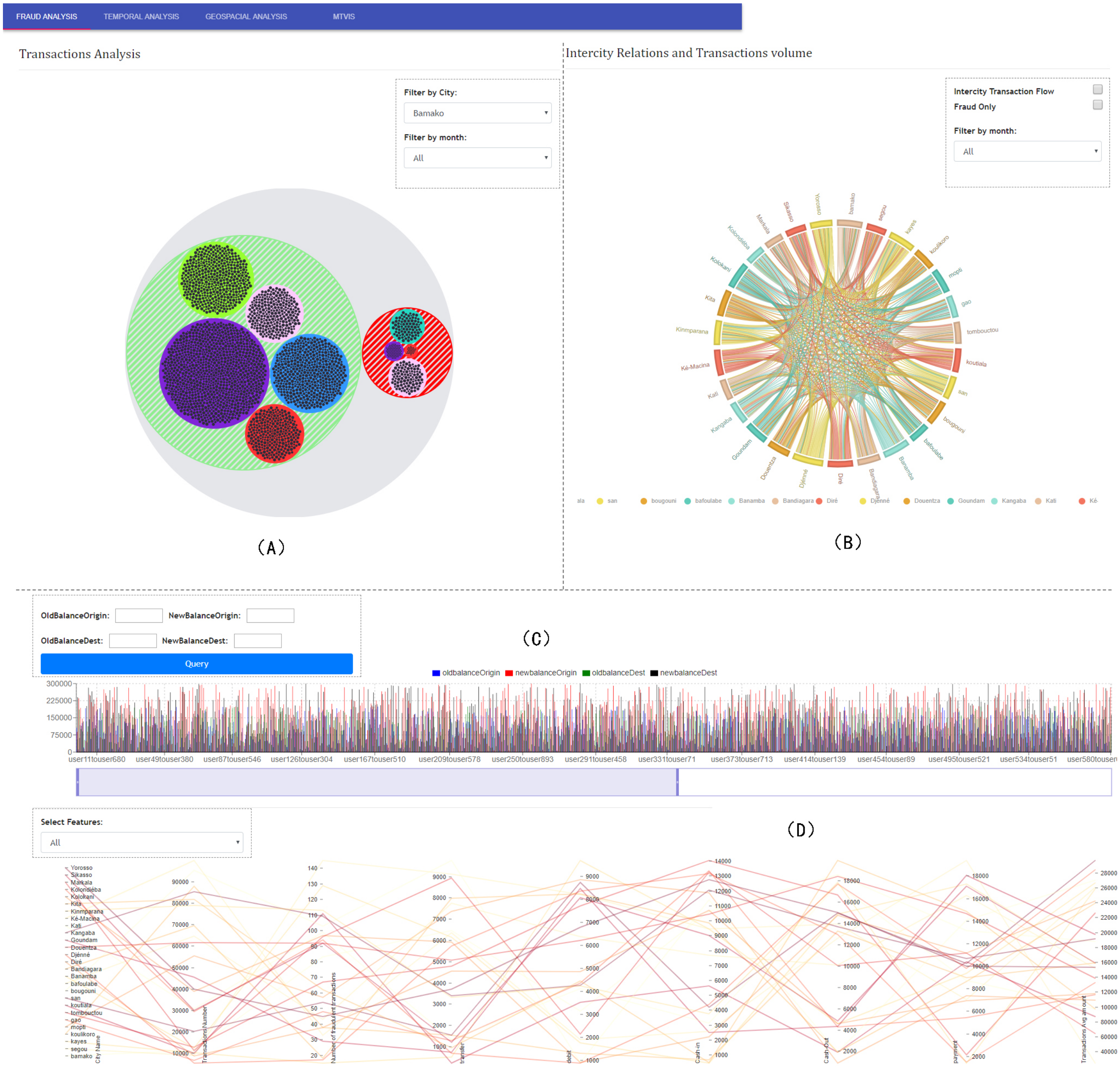
- Experiment and results

To get optimal results for the fraud detection module multiple experiments were made and parameter tuning was performed on parameters such as the number of neurons in the hidden layers for GRU neural network and Oversampling rate and number of nearest neighbors for the SMOTE algorithm. Optimal values were choosed empirically for each parameters. The optimal accuraccy was (99.33% ACC,99.25 DR ,0.093% FDR) obtained with a system architecture of [122,200,1].

4 Visual Components

After performing the classification on input data using the GRU Neural Network, we feed the visualization components with the classified data for visualization and analysis. The system is composed of three main views each one performing a specific task: Fraud analysis view,Geospacial Analysis view and temporal analysis view. In this section we describe each view and explain its functionalities.

4.1 Fraud analysis View



Transactions Analysis

Filter by City: Bamako

Filter by month: All

Intercity Relations and Transactions volume

Intercity Transaction Flow □

Fraud Only □

Filter by month: All

OldBalanceOrigin: NewBalanceOrigin:

OldBalanceDest: NewBalanceDest:

Query

Select Features: All

Figure 2: Fraud analysis view displaying the bubble chart with clusters of transactions,a chord to show intercity relations, a bar chart displaying the state of the customers balance before and after performing the transaction and a parallel coordinates graph plotting multiple features for each city

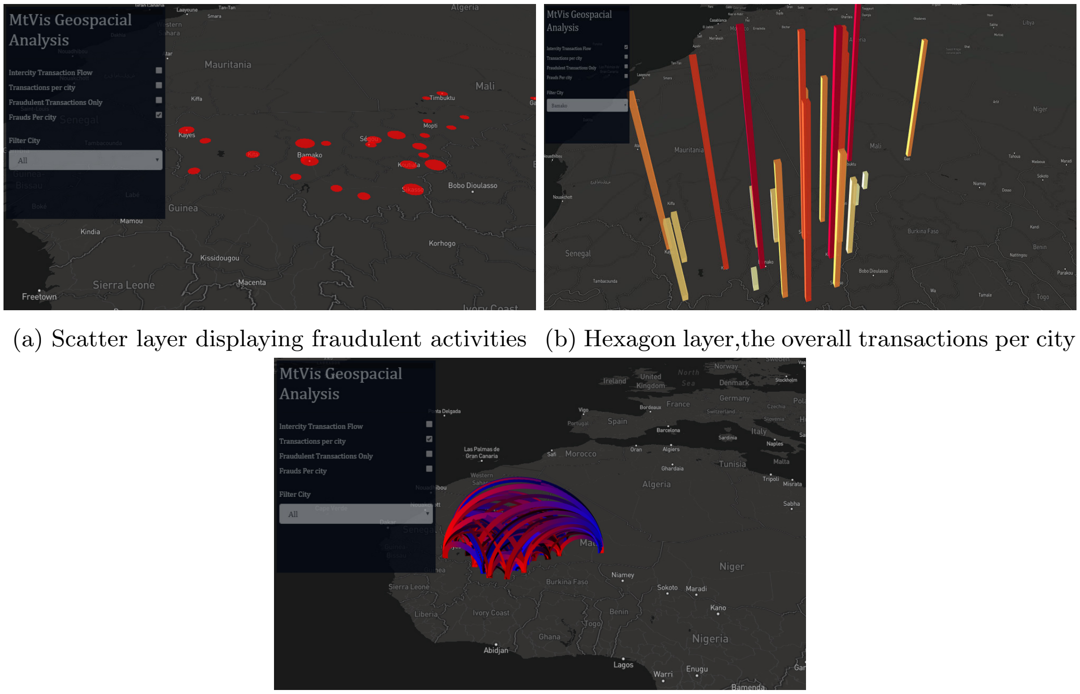
The fraud analysis view is the main view of our system and allow the user to grasp the principal trends in the transactions.It is composed of the following components:

- Bubble chart(A)
  The bubble chart displays clusters of transactions according to their type and their fraudulent or non fraudulent nature.The first level is a green cluster containing the normal transactions and a red one containing the

fraudulent transactions,in the second level the transactions are organized by category using colors encoding,and in the third level each transaction is displayed as a circle,the size of the circle is proportional to the amount of money involved in the transaction.Each level of the bubble chart is zoomable and interactive giving the user a better analysis experience.

- Chord chart(B)
  The chord chart is used to show the relation between the transactions location.This is useful to know how users from different cities interact and assess the volume of the transactions between cities.
- Barchart(C)
  The filterable bar chart displays the situation of the customers balance before and after performing the transaction.The variation of the balance is an important fraud feature.A query panel allows the user to query different transactions.The query panel is used to make hypothesis and check them against the dataset.
- Parallel coordinates graph(D)
  The parallel coordinates graph is an abstraction of the transactions for each city across many features. The purpose of the parallel coordinates graph is to allow the user to quickly analyze different dimensions of the transactions per city.

4.2 Geospacial Analysis



(a) Scatter layer displaying fraudulent activities (b) Hexagon layer,the overall transactions per city

(c) Arc layer displaying the transaction activity between cities

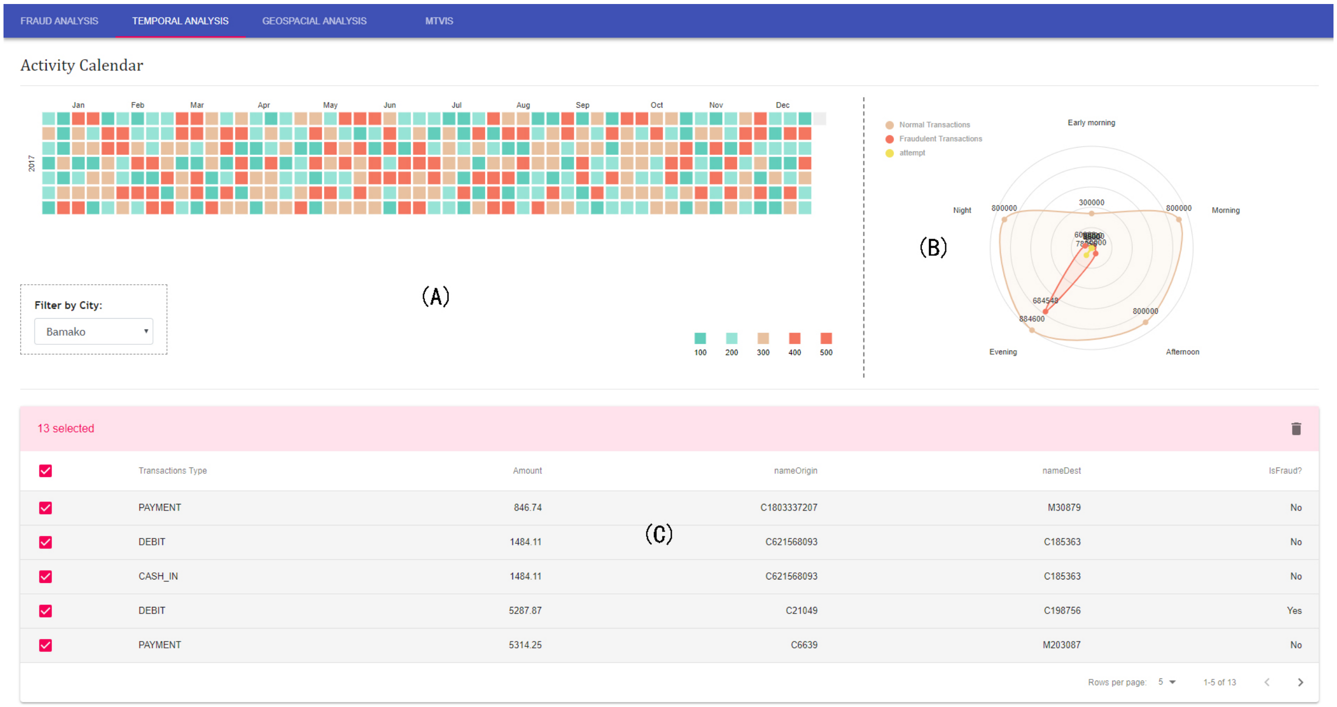
Figure 3: Layers of the geospacial analysis view

The geospacial analysis view allow the user to grasp the geographical state of the transactions.It is composed of three layers:

- A scatter (a) shows the geographical distribution of fraudulent transactions. The radius of the scatter is proportional to the number of fraudulent transactions.
- A hexagon layer(b) is used to display the volume of transactions performed per city.

- Arc layer(c) displays the transaction flow among cities.Using the thickness of the arcs and colors we abstract the volume of transactions In and Out between two cities.

4.3 Temporal Analysis View



13 selected

☑	Transactions Type	Amount	nameOrigin	nameDest	IsFraud?
☑	PAYMENT	846.74	C1803337207	M30879	No
☑	DEBIT	1484.11	C621568093	C185363	No
☑	CASH_IN	1484.11	C621568093	C185363	No
☑	DEBIT	5287.87	C21049	C198756	Yes
☑	PAYMENT	5314.25	C6639	M203087	No

Rows per page: 5 ▾ 1-5 of 13 ‹ ›

Figure 4: Temporal analysis view,with a calendar activity view(A) displaying intensity of transactions per day, a radar chart displaying the time distribution of transactions per category(B) and a table view displaying the raw data(C).

The temporal analysis view is used to display the time aspect of the transactions and composed of three charts:

- A calendar activity view shows the distribution of transactions per day.
- A radar view is used to display the variations of activity at different time of the day.This is useful to identify the exact time when fraudulent transactions occur the most.
- At the bottom is an interactive and filterable table displaying the classified data, allowing the user to analyze it deeper.

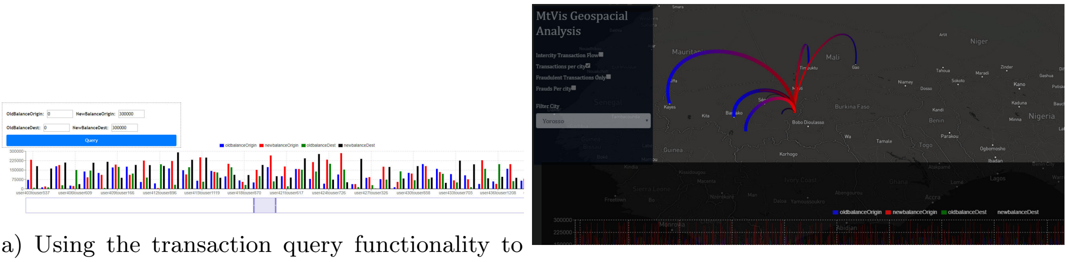
To enhance the analysis and exploration experience,we implemented a wide range of interactions and filters.The user can hover on components for more detail and filter the information.

5 Case studies and Expert Evaluation

To validate the effectiveness of our system we conducted a case study with two experts (in the following referred to as Expert A and Expert B).Each expert used our system to analyze the transactions and gave us feedback.

5.1 Case study 1: Terrorism financing and money laundering

Expert A is an anti-terrorism agent,he used the system to target transactions between big cities and sensitive areas where terrorists are located by using the filtering option available in the fraud analysis view.The expert then analyzed the transactions occurred between sensitive areas by using the geospacial analysis feature. The fraudulent transactions found will be the subject of future investigations.He then commented about the system "The geospacial view is really powerful when tracking terrorism financing transactions, and analyzing transactions in sensible regions". Money laundering happens when illegal money is introduced in the financial system through placements.To track such placements our transaction query can be used to filter transactions in which big amounts of money are placed(cash in).



OldBalanceOrigin: 0 NewBalanceOrigin: 300000

OldBalanceDest: 0 NewBalanceDest: 300000

MtVis Geospacial Analysis

Transactions per city ☑

Filter City: Yorosso

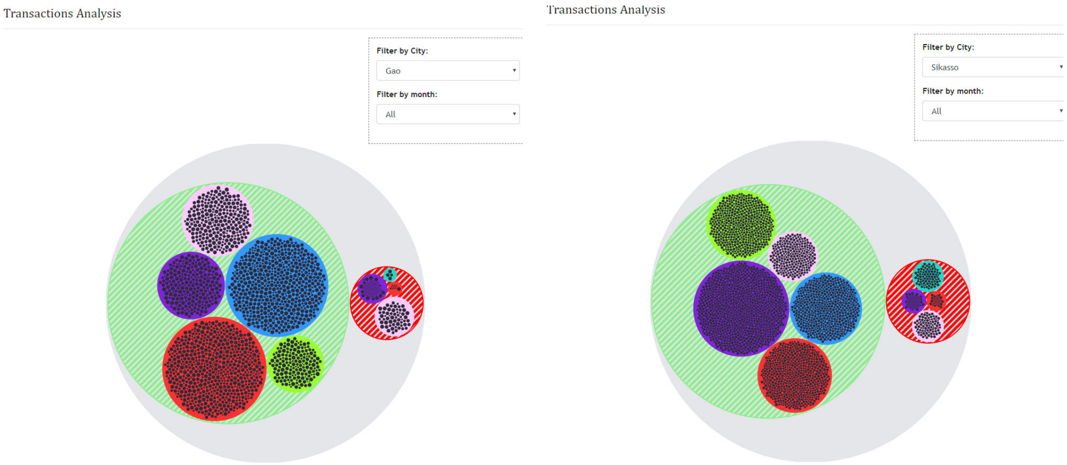
(a) Using the transaction query functionality to make hypothesis and verify them,useful for money laundering tracking

(b) Targeting transactions between big cities and terrorism areas using the arc layer

Figure 5: Case study 1

5.2 Case study 2: Users behavior

Expert B is a senior financial advisor for mobile money he used the system to analyze the behaviour of users in different cities,he found that users in biggest cities tend to perform transactions involving cash in,cash out and payment,the users in small cities on the other hand tend to receive money from bigger,they perform more cash out,transfer,payment activities.Using the temporal analysis view the expert analyzed the temporal behaviors of their customers and found interesting trend in the timing of the transactions.The expert stated "This is very useful to get relevant information when deploying our new services and planning the coverage of new regions ".



Transactions Analysis

Filter by City: Gao

Filter by month: All

Transactions Analysis

Filter by City: Sikasso

Filter by month: All

(a) Users behavior in city 1

(b) Users behavior in city 2

Figure 6: Case study 2

6 System Implementation

Our system is a client-server application with a React js(23) front and a Flask(25) back end. The Fraud detection module was developed using the open source deep learning framework pytorch(**?**). The Detection module is deployed on the server where the input data is prepossessed and sent to the front end for visualization.

7 Conclusion

We presented MtVis a visual analytics framework for mobile money transactions analysis and exploration, and showed the relevance of our tool in understanding the trends and all aspects in a mobile transaction system, our system has been proven to be effective for fraud detection and planning tasks.We also proved the importance of the transaction entity in a mobile money service system in detecting anomalous activities. However, in this work we only focused on the transaction entity which is only a small component of a mobile money transfer service, in future works we will dive deeper in other components such as users behavior and users categorization.

References

[1] Ko, Sungahn, et al. *"A survey on visual analysis approaches for financial data."*. Computer Graphics Forum. Vol. 35. No. 3. 2016.

[2] Chang, Remco, et al. *"Wirevis: Visualization of categorical, time-varying data from financial transactions."* Visual Analytics Science and Technology, 2007. VAST 2007. IEEE Symposium on. IEEE, 2007.

[3] Didimo, Walter, et al.
"An advanced network visualization system for financial crime detection." Visualization Symposium (PacificVis), 2011 IEEE Pacific. IEEE, 2011.

[4] *NETSUITE(2018) https://http://www.netsuite.com.*

[5] *Mpesa(2018) https://www.mpesa.in/.*

[6] *https://orangemoney.orange.fr.*

[7] Novikova, Evgenia, and Igor Kotenko. *"Visual Analytics for Detecting Anomalous Activity in Mobile Money Transfer Services."*. International Conference on Availability, Reliability, and Security. Springer, Cham, 2014.

[8] Gaber, Chrystel, et al. *Analyse des comportements dans un système de transfert d'argent sur mobile.*. 8ème Conférence sur la Sécurité des Architectures Réseaux et Systèmes d'Information (SAR SSI). 2013.

[9] Abbasi, Ahmed, et al. *Metafraud: a meta-learning framework for detecting financial fraud.*. Mis Quarterly (2012): 1293-1327.

[10] Adedoyin, Adeyinka, et al. *"Predicting fraud in mobile money transfer using case-based reasoning."*. International Conference on Innovative Techniques and Applications of Artificial Intelligence. Springer, Cham, 2017.

[11] Kappelin, Frida, and Jimmie Rudvall. *"Fraud Detection within Mobile Money: A mathematical statistics approach."*. (2015)

[12] Albashrawi, Mousa, and M. Lowell.. *"Detecting financial fraud using data mining techniques: A decade review from 2004 to 2015."*. Journal of Data Science 14.3 (2016): 553-569.

[13] Lopez-Rojas, Edgar, Ahmad Elmir, and Stefan Axelsson. *"PaySim: A financial mobile money simulator for fraud detection."*. 28th European Modeling and Simulation Symposium, EMSS, Larnaca. Dime University of Genoa, 2016.

[14] Chawla, Nitesh V., et al. *"SMOTE: synthetic minority over-sampling technique."*. Journal of artificial intelligence research 16,321-357.

[15] Jack, William, Adam Ray, and Tavneet Suri. *"Transaction networks: Evidence from mobile money in Kenya."*. American Economic Review 103.3 (2013): 356-61.

[16] Sharko, John, Georges Grinstein, and Kenneth A. Marx. *"Vectorized radviz and its application to multiple cluster datasets."*.IEEE transactions on Visualization and Computer Graphics 14.6 (2008): 1444-1427.

[17] Sun, Yanmin, et al. *"Cost-sensitive boosting for classification of imbalanced data."*.Pattern Recognition 40.12 (2007): 3358-3378.

[18] Chu, J., et al. *"Flow based oversampling technique for multiscale finite element methods."*Advances in Water Resources 31.4 (2008): 599-608.

[19] Major, John A., and Dan R. Riedinger. *"A hybrid knowledge/statistical-based system for the detection of fraud."*Journal of Risk and Insurance 69.3 (2002): 309-324.

[20] Breiman et al. *Breiman, L., 2017. Classification and regression trees. Routledge.*

[21] Corani, Giorgio, et al. *"Statistical comparison of classifiers through Bayesian hierarchical modelling."*Machine Learning 106.11 (2017): 1817-1837.

[22] Jozefowicz, Rafal, Wojciech Zaremba, and Ilya Sutskever. *"An empirical exploration of recurrent network architectures."*International Conference on Machine Learning. 2015.

[23] Aggarwal, Sanchit. *"Modern Web-Development using ReactJS."*International Journal of Recent Research Aspects 5 (2018): 133-137.

[24] Grinberg, Miguel. *Flask web development: developing web applications with python. "*O'Reilly Media, Inc.", 2018.

[25] Ju, Shengtai, et al. *A PyTorch Framework for Automatic Modulation Classification using Deep Neural Networks.*(2018).