Proceedings of the 10th International Conference on Applied Informatics
Eger, Hungary, January 30–February 1, 2017. pp. 121–128
doi: 10.14794/ICAI.10.2017.121

Event Detection on Call Detail Records*

Katalin Hajdú-Szücs[a], Attila Kiss[b]

[a] Eötvös Loránd University
szucsk@caesar.elte.hu

[b] Eötvös Loránd University
kiss@inf.elte.hu
J. Selye University
kissae@ujs.sk

Abstract

The previously unseen amount of data from mobile phones creates new possibilities to model human behaviour. Several data mining methods have been applied and tested on these datasets, providing interesting results. The goal of the paper is to present an overview of the applicability of anomaly detection in mobile phone data analysis. After describing the wide range of possibilities for anomaly detection in this context and presenting the state of the art, we detail the most important methods. Finally, we present our own findings in this field on CDR data covering one-month of traffic of a Hungarian mobile communication service provider.

Keywords: Mobile phone metadata, CDR, anomaly detection

MSC: 94A13

1. Introduction

Millions of people communicate via calls, SMS or MMS every day. Even without knowing the exact context of these calls and messages, the increasing traffic of mobile phone devices provides us with the unforeseen opportunity to learn more about our society.

Mobile communication network providers store all the data related to the phone traffic in the shape of CDR (Call Detail Record) files. A call detail record contains metadata that describe a specific instance of a telecommunication transaction

*This publication is the partial result of the Research & Development Operational Programme for the project “Modernisation and Improvement of Technical Infrastructure for Research and Development of J. Selye University in the Fields of Nanotechnology and Intelligent Space”, ITMS 26210120042, co-funded by the European Regional Development Fund.

(telephone calls, text messages). In general it contains the following fields: ID of subscriber originating the call, starting time of the call, call duration, call type (voice, SMS, etc.), ID of the equipment writing the record, Cell ID.
There are several ways one can extract important information from these datasets. Using the calls between the costumers one can build a graph in which the vertices are people and edges are drawn between two vertices who call each other. These graphs can be analysed by complex network research methods and social networks can be constructed.

The location of a given person of interest can be specified easily through the identification of geographic coordinates of the transmitting tower the user connected to first. This provides us the opportunity to learn more about human movement, as well as about distribution of population or simply about the whereabouts of individuals, that can be of paramount importance for instance in case of emergencies.

Our work is focused on the analysis of cell phone call activity by means of data mining and anomaly detection techniques. These methods make it possible to discover underlying human behaviour and use the results to detect potentially anomalous events. The wide applicability of this concept leads to multiple research directions. One of these areas is crowd detection. Over the last decade much effort has been devoted to introduce a methodology of detecting social events in massive mobile phone data to help resource allocation of service providers and traffic management. Another area concerns the ability to follow the movements of unusual gatherings of people. This information is beneficial to raise the situational awareness of emergency and disaster managers to perform effectively during an event. A third direction involves the utilisation of cell phone data in case of incidents as part of an early warning systems. Furthermore, forensics analysis and fraud detection of phone networks can help to identify criminal organisations that are structurally different from common social network.

In the next section we are going to detail some of the most important methods from the literature, then in Section 3 we present the results of our own research in the field of emergency detection.

2. Event Detection Methods

Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behaviour [4]. Detecting outliers or anomalies in data has been studied in the statistics community as early as the 19th century [5]. Over time, a variety of anomaly detection techniques have been developed in several research communities. The next paragraphs summarise those techniques that have been applied on CDR data successfully.

Crowd Detection The ability to immediately detect crowd formations can provide many advantages. First of all, it can improve public safety. For example, in case of an emergency situation the exact knowledge of the affected people's position facilitates easier assistance. It also helps traffic management, but can serve marketing purposes as well. The work of Traag et al. [14] focuses on unusually large gatherings of people and introduce a methodology of detecting social events in massive mobile phone data, based on a Bayesian location inference frameworks. They also develop a scheme for deciding who is attending an event and they demonstrate the method on some examples. The reason for taking a probabilistic approach is the phenomena when a user switches between antennas while making calls, although it is unrealistic that the user is actually moving so fast. Furthermore when events take place it is more likely that multiple antennas serve the costumers at the event area because of load balancing.

Detection of error-prone network configurations The current service structure faces several challenges. First of all, the possibility of error prone manual configurations and management to deliver new networking services presents a constant challenge for mobile operators as well as service providers. Second, the planning of additional services can be imperfect and it can cause the deployment life-cycle of new changes to take too long. One solution to these network management problems was presented in 2014 [10], in this work the goal of the authors is to analyse the users calling activities and detect the abnormal behaviour of user movements in a real cellular network of AVEA, a mobile service provider in Turkey. A rule based anomaly detection technique have been proposed to help mobile service providers improve their system consistency and reduce the time to detect location based anomalies.

Fraud Detection A fraudulent phone call is one in which there is no intention to pay. In the context of telecommunication a fraudulent event can have many forms. There are some basic fraud types that can be found in the literature [4]. Subscription fraud happens when one signs up for service with no intent to pay. Intrusion fraud is when a legitimate account is compromised. In this case the anomalous traffic is mixed with normal behaviour. Another type is fraud based on loopholes in technology. For example a voice mail systems can be used to make outgoing calls. If inadequate passwords are used to secure the mailboxes, it creates a vulnerability. Sometimes, instead of exploiting technological loopholes, human interaction with the system is exploited. These type of attack are called social engineering.

Although there are several fraud detection techniques in the literature, most evasion methods use CDR data to create behaviour profiles for the customer, and detect deviations from these profiles. Early fraud detection used thresholds on the number of messages and the number of minutes of calling within a specified time period. This method is called threshold-based altering. But these early solutions were not always reliable enough to detect all fraudulent events and more sophisticated methods were needed. Following that, the signature-based altering solutions emerged [1]. In these techniques each incoming call is compared to the current signature for that customer and also to a generic fraud signature. Whenever the current activity is more similar to the fraudulent behaviour, an alert is raised. Later stochastic models [12, 3] and solutions based on neural networks [6, 8, 11, 7] outperformed all these earliest methods. The accuracy of some of these models exceeds 98%.

Emergency Detection Quality communication capabilities and high level situational awareness are the greatest needs for disaster and emergency response managers. Reports from on-scene coordinators, first responders, public safety officials, the news media, and the affected population are often inaccurate, conflicting and incomplete with gaps in geographical and temporal coverage.

Population movements following a disaster can cause important increases in morbidity and mortality. Without knowledge of the locations of affected people, relief assistance is compromised. In the work of Linus Bengtsson et al. [2] geographic positions of SIM cards have been utilized to estimate the magnitude and trends of population movements following the Haiti 2010 earthquake and cholera outbreak. Their findings corresponded well with the results from a large retrospective, population-based UN survey, but was very different from the official government estimates which were widely used during the relief operations.

In the paper of Steenbruggen et al. [13] mobile phone usage data is utilised as a reliable predictor of motorway incidents for an early warning system. Greater Amsterdam is used here as a case study. In the article a three-stage approach is followed. Firstly, the effect of motorway traffic flow and traffic incidents on mobile phone usage is analyzed. Then they investigate whether increase in motorway traffic flows effect the probability of accident. And finally, the marginal effects of different types of mobile phone usage on the probability of having a motorway incident is estimated.

3. Emergency Detection: A Case Study

Our work concentrated on the utilisation of cell phone data as part of an early warning system. In particular, at 10:36 pm on the 24th of September 2016 a bomb went off in central Budapest. We aimed to analyse the effect of the explosion in the CDR data and we propose an algorithm to detect such events as soon as possible.

The dataset composed of anonymized CDR data covering one-month of traffic of a Hungarian mobile communication service provider. The following attributes were available: call ID, subscriber ID (anonymized), Network ID, device type (anonymized), timestamp, cell ID, event type (originating call, terminating call, originating SMS, terminating SMS).

In case of an emergency situation, an elevation in the phone traffic is expected. Since human behaviour is periodic, the CDR data shows periodicity also. Every cell tower's traffic has its own pattern that depends on the day of the week and on the time of the day as well. Due to this phenomenon it is not possible to detect an unexpected elevation in traffic by simply setting a threshold on the number of CDR events. A more complex solution is needed.

As a first step, feature vectors were generated by aggregating phone calls and sms together in every five minutes, independently for each antenna. Based on the first two weeks in the dataset, the average of the normal daily pattern of each antenna was defined for each day of the week. This information is updated in every five minutes if there is no emergency detected. Whenever a new aggregate of an antenna arrives, the corresponding average is subtracted and the difference is further analysed whether the deviation from the average is significant or not. For this purpose we applied Bayesian surprise detection [9].

For each antenna a, at each time step t, let $A(t, a)$ be the five-minute aggregate. $A(t, a)$ is used to update the prior distribution and obtain the posterior distribution, using the Gaussian model.

$$P^a_{prior} = P(.|A(t-1,a),\dots,A(t-N,a))$$
$$P^a_{post} = P(.|A(t,a),A(t-1,a),\dots,A(t-N,a))$$

The surprise value is calculated as

$$SA(t,a) = DKL(P^a_{post}||P^w_{prior})$$

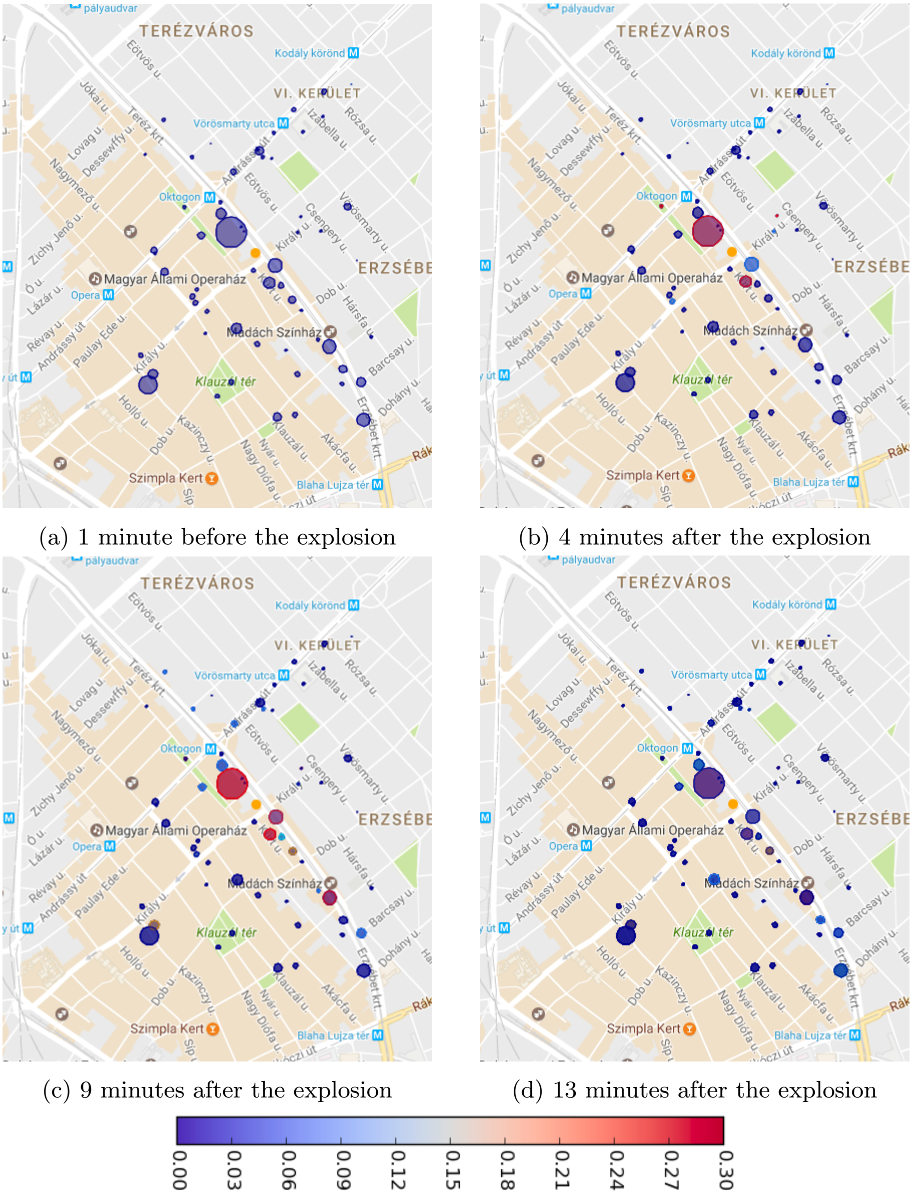
where DKL is the Kullbach Leibler Divergence (a.k.a. information gain). Whenever the surprise values exceeds a predefined threshold (specific for each antenna), the traffic is considered to be higher than usual. In case there are multiple antennas with elevated traffic in the same region at the same time, an alarm is set indicating a possible emergency situation.

On Figure 1 the surprise values of the antennas near to the explosion is presented one minute before (Figure 1a), then 4, 9 and 13 minutes after the explosion (Figures 1b, 1c and 1d). It can be seen on Figure 1a that one minute before the explosion all the neighbouring antennas had a surprise value close to zero. However, some minutes later, after the detonation, the surprise values of the closest antennas suddenly rose on Figures 1b and 1c. Finally, after 13 minutes, although the traffic was still increased, the surprise value started to decrease back to normal, because the model got used to the elevated traffic.

Figure 2 presents the data corresponding to one of the near antennas on the day of the explosion. The explosion took place at 22:36. By having a closer look at this part of the graph on Figure 2a, some elevation in the traffic can be observed. Although, the type of the increase is not obvious by looking only at this picture. It could be part of the average daily pattern of the antenna, like the usual elevation around 1 am in the morning. However, after the application of Bayesian surprise detection, the surprise values on Figure 2b clearly show that the increased activity at 22:36 is extremely suspicious.

4. Conclusions

In our work we investigated the applicability of anomaly detection techniques on CDR data. After describing the different research directions and the state of the art, we presented our own solution for emergency detection on Call Detail records.



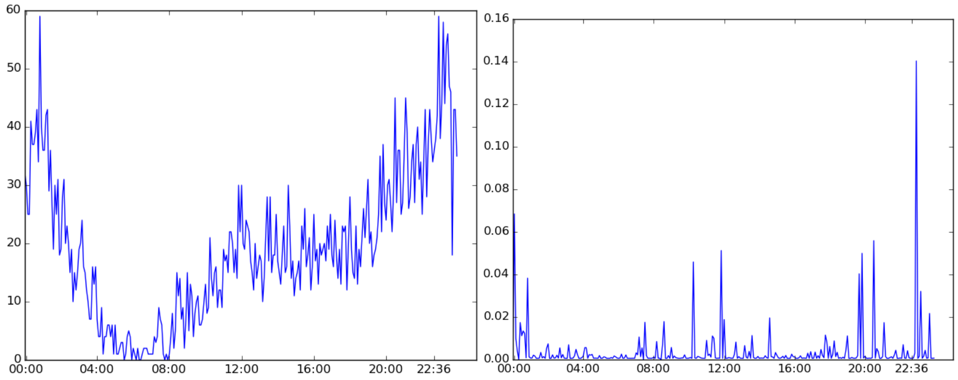
(a) 1 minute before the explosion

(b) 4 minutes after the explosion

(c) 9 minutes after the explosion

(d) 13 minutes after the explosion

Figure 1: The maps show the surprise values of the antennas close to the detonation at different time steps. Each circle indicates an antenna. The size of the circles represent the average volume of transacted traffic, the colour indicates the surprise value. The location of the explosion is shown by a yellow circle.



(a) The number of phone calls and sms aggregated in every five minutes.

(b) The corresponding surprise values.

Figure 2: The figures show the amount of traffic and the surprise values of an antenna that was close to the scene on the day of the detonation.

The future work in this topic involves further tests on data with more emergency situations in order to precisely assess the performance of the solution. It could be also interesting to investigate the types of the events generated after an emergency situation instead of treating calls and SMS together.

References

[1] Richard A Becker, Chris Volinsky, and Allan R Wilks. Fraud detection in telecommunications: History and lessons learned. *Technometrics*, 2012.

[2] Linus Bengtsson, Xin Lu, Anna Thorson, Richard Garfield, and Johan Von Schreeb. Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: a post-earthquake geospatial study in haiti. *PLoS Med*, 8(8):e1001083, 2011.

[3] Michael H Cahill, Diane Lambert, José C Pinheiro, and Don X Sun. Detecting fraud in the real world. In *Handbook of massive data sets*, pages 911–929. Springer, 2002.

[4] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.

[5] FY Edgeworth. Xli. on discordant observations. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 23(143):364–375, 1887.

[6] Abdikarim Hussein Elmi, Subariah Ibrahim, and Roselina Sallehuddin. Detecting sim box fraud using neural network. In *IT Convergence and Security 2012*, pages 575–582. Springer, 2013.

[7] Hernan Grosser, Paola Britos, and Ramón García-Martínez. Detecting fraud in mobile telephony using neural networks. In *International Conference on Industrial,*

Engineering and Other Applications of Applied Intelligent Systems, pages 613–615. Springer, 2005.

[8] AJ Hussain and E Chew. Data mining and telecommunication fraud detection using artificial neural networks. In *IWSSIP'02: international workshop on systems, signals and image processing*, pages 474–483, 2002.

[9] Laurent Itti and Pierre F Baldi. Bayesian surprise attracts human attention. In *Advances in neural information processing systems*, pages 547–554, 2006.

[10] Ilyas Alper Karatepe and Engin Zeydan. Anomaly detection in cellular network data using big data analytics. In *European Wireless 2014; 20th European Wireless Conference; Proceedings of*, pages 1–5. VDE, 2014.

[11] Sameer Qayyum, Shaheer Mansoor, Adeel Khalid, Zahid Halim, A Rauf Baig, et al. Fraudulent call detection for mobile networks. In *Information and Emerging Technologies (ICIET), 2010 International Conference on*, pages 1–5. IEEE, 2010.

[12] Steven L Scott. A bayesian paradigm for designing intrusion detection systems. *Computational statistics & data analysis*, 45(1):69–83, 2004.

[13] J. G. M. Steenbruggen, Emmanouil Tranos, and Piet Rietveld. Can Motorway Traffic Incidents be detected by Mobile Phone Usage Data? Technical report, Faculty of Economics and Business Administration, 06 2015.

[14] Vincent A Traag, Arnaud Browet, Francesco Calabrese, and Frédéric Morlot. Social event detection in massive mobile phone data using probabilistic location inference. In *Privacy, security, risk and trust (PASSAT) and 2011 IEEE Third inernational conference on social computing (SocialCom), 2011 IEEE Third International Conference on*, pages 625–628. IEEE, 2011.