cost
EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

UCA Universidad de Cádiz

DIGFORASP

COST ACTION CA17124

BOOK OF ABSTRACTS

3rd Working Group Meeting

October 23rd – 24th, 2019 · La Valletta, Malta

Editors:
Jesús Medina

Associate Editors:
Roberto G. Aragón, M. Eugenia Cornejo, David Lobo,
Eloísa Ramírez-Poussa

Table of Contents

DigforASP Workshop. Session 1

DigforASP Workshop. Session 2

DigforASP Workshop. Session 3

DigforASP Workshop. Session 4

Typical Prosecutor's requests collection to create and share a requests collection for NLP purposes

Raffaele Olivieri, Ph.D.
Arma dei Carabinieri - University of L'Aquila, Italy

Digital forensic science is the science that deals with identifying and prosecuting digital criminals that involve digital technologies such as computer devices, network devices, mobile devices and so on. The digital forensic investigation is carried out to identify and collect the evidence that will subsequently be accepted by the court. The activities of Digital Forensics and Digital Investigations in most of the jurisprudence begin with an act of delegation by the investigating magistracy (Prosecutor).

For this reason, the fundamental act that gives rise to the aforementioned activities is the Prosecutor's request, in which the main lines of investigation to be followed by the Law Enforcement or by the Consultants are outlined.

Based on these considerations, new application scenarios are opening up for AI, in particular in the Natural Processing Language sector, for the automation of investigation schemes, which require a thorough analysis of the investigation methodologies adopted in the various countries to understand the common problems in dealing with the contrast of digital crime.

Tools supporting obtaining digital evidence in the fight against cybercrime used in the Cybercrime Department

Wojciech Lis
CyberCrime Department in Cracow, Poland

Presentation of the structures and tasks of the Department of Fighting Cybercrime in Krakow Police. Information about the equipment and software used for Department's forensics and operational cases. Directions of cybercrime development in Poland and related problems with digital evidence. Paying attention to hardware and software that may be available in the daily work of the Cybercrime Police in specialties in the field of digital data analysis.

Improving digital forensic model explainability through model-agnostic methods

***Marko Krstic*[1]*, Milan Cabarkapa*[2] *and Aleksandar Jevremovic*[3]

[1] Regulatory Agency for Electronic Communications and Postal Services
[2] School of electrical engineering, Belgrade, Serbia
[3] Sinergija University, Bosnia and Herzegovina

In order to find the best solution for the observed digital forensics problem different mathematical, AI, AR tools should be combined. By separating the explanation methods from digital forensic models, it is possible to choose and combine wide variety of algorithms and still have interpretable models. Thus, in this study we examined Local Surrogate (LIME), Shapley Values and SHAP (SHapley Additive exPlanations) as representatives of model-agnostic methods and discuss their applicability to digital forensics domain. Special attention was dedicated to law requirements and Big Data scalability issue.

Neural Networks and other Machine Learning Techniques for Data Analysis

***Stefania Costantini*[1]*, Lorenzo De Lauretis*[2] *and Aleksandar Jevremovic*[3]**
[1] Dipartimento di Ingegneria e Scienze dell'Informazione e Matematica, Univ. dell'Aquila, Italy
[2] Università degli Studi dell'Aquila, Italy
[3] Singidinum University, Belgrade, Serbia

Machine learning is an application of Artificial Intelligence, that provides systems with the ability to automatically learn and improve from experience without being explicitly programmed. During the past few years, machine learning became ever more popular. It can be used to simplify everyday life and is applicable to a lot of scenarios. In our work, we have examined a dataset concerning London crime data, from January 2008 to December 2016, represented using a CSV file. Elaborating those data through the software RapidMiner, we have been able to create a neural net that provides us with useful statistical data about the crimes committed in the districts of London. Applying machine learning techniques via our neural net, we will be able to make classification on the crime data, in particular, we can discover whenever there is an increase of a particular kind of crime in certain areas. This can help the police districts for the assignment task of policemen, cars and other resources, increasing the attention concerning districts with an increasing crime rate. By using a neural network, we are able to make previsions on the crime rate, but we are not able to know the reason underlying the increase or decrease of crime rate and the real happenings of crime events. The dataset that we have analyzed unfortunately lacks more specific information, such as the precise date of each crime and its geolocalization, features that would have made our results more interesting. Moreover, as emphasized by Judea Pearl in the recent book

"The book of why", we are unable to understand the reasons underlying the increase/decrease of crime rates in various areas. To tackle this problem, we propose for future work the adoption of Computational Logic. In particular, we will experiment the adoption of Inductive Logic Programming which is a form of Machine Learning that however learns rules, that should represent causal connections extracted from data to understand "why" they increase/decrease in each specific area. In complement, forms of reasoning such as Answer Set Programming might elicit future plausible scenarios of crime distribution. To do this however, the datasets to be analyzed should be richer of significant features.

Applications of AI in Cyber Security - opportunities and challenges

Goran Shimic
Military Academy, Belgrade, Serbia

Due to the necessity to process and analyze a huge amount of data represented in different formats, artificial intelligence (AI) technology and science have become a mandatory part for contemporary systems designed for evidence analysis. Intelligently extracting data and metadata, their clustering, classification and the advanced search performed on them represent the improvements introduced by AI. Pattern recognition in the behavior of malicious users in their correspondence, in their attacks on the systems exposed on the Web, or in the content that they send to their victims, shows the complexity of evidence analytics in today's digital forensic.

Comparing Models for Time Series Analysis and Forecasting of London Crime Data

Aleksandra Dedinec, Sonja Filiposka and Anastas Mishev
Ss. Cyril and Methodius University - Skopje, North Macedonia

Crime forecasting has become a major trend over the past years based on the availability of new technologies and methods that can be used to improve prevention efforts by supporting decisions related to efficient resource allocation. The ability to accurately forecast the future crime trends heavily depends on the data quality and quantity, characteristics of the time series analyzed, and model and its tuned parameters used to make the forecast. In depth investigation is needed to uncover the best model that provides the most accurate forecast while the careful tuning of its parameters has a tremendous impact on the confidence interval size.

This paper aims to provide an initial analysis made by applying time series forecasting methods and models on the open data London crime dataset. The obtained results uncover different patterns in the dataset related to seasonal activities. The forecasting techniques presented are used to discuss the accuracy and expectations that can be made from the future crime forecasting based on the methodology used.

Mathematical tools applied to digital forensic. London Crime Data, 2008-2016

Roberto G. Aragón, M. Eugenia Cornejo, David Lobo, Jesús Medina and Eloísa Ramírez-Poussa
Department of Mathematics, University of Cádiz, Spain

Formal concept analysis (FCA) was introduced in the eighties by Ganter and Wille and since then, it has become an appealing research topic both from theoretical and applied perspective. FCA is a tool for extracting pieces of information from databases containing a set of attributes and a set of objects together with a relation between them. These pieces of information are called concepts, which can be hierarchized to obtain concept lattices. FCA allows us to handle uncertainty, imprecise data or incomplete information, which are important characteristics nowadays. In this work, we focus on applying fuzzy formal concept analysis techniques to London crime dataset, in order to obtain interesting information from this dataset which allows us to answer the following questions: What category crimes are related? If a Robbery is done, what other crimes will be done, by borough? What are the boroughs and months more dangerous?, among others.

CASPER - Children Agents for Secure and Privacy Enhanced Reaction

Aleksandar Jevremovic[1], Milan Cabarkapa[2] and Marko Krstic[3]
[1] Sinergija University, Bosnia and Herzegovina
[2] School of electrical engineering, Belgrade, Serbia
[3] Regulatory agency for Electronic Communications and Postal Services

We present our recently approved project within the Horizon 2020 framework, specifically, grant No 825618 within the NGI TRUST call. The project goal is to apply A.I. technologies to protect children on the Internet. Different types of threats are addressed - cyberbullying, predators, challenges, etc. The consortium consists of organizations from Serbia, Portugal, and North Macedonia, but is very likely to be increased in future.

On fuzzy measure's role in ranking objects based on the metadata

***Ivana Štajner-Papuga*[1] *and Andreja Tepavčević*[1,2]**
[1] Department of Mathematics and Informatics, Faculty of Sciences, University of Novi Sad, Novi Sad, Serbia
[2] Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, Serbia

The aim of this paper is to show a potential applicability of some well-known fuzzy integrals, i.e., of the Choquet integral and the Sugeno integral, in construction of extraction tools in the forensics metadata analysis. Possible benefit of the use of fuzzy integrals in general is in the flexibility of a fuzzy measure which is in the core of the aggregation process and which is being used for modeling predefined classification requirements. The proposed method is sensitive not only to the change of the search parameters, but to the change within interaction of the search parameters as well.

Analysis of Chaotic Properties in Synthetic Databases

Fatih Özkaynak and Yılmaz Aydın
Firat University, Elazig 23119, Turkey

The main purpose of science and engineering studies is to understand real world systems and to use these results for the benefit of mankind. During these studies, chaos theory became increasingly important. Because this phenomenon is needed to understand the logic of real world events. Therefore chaos theory has started to find its place in many applications. In the simplest expression, chaos theory is defined as the randomness of a deterministic system. In other words, despite the fact that real world events are mathematical models, they contain an unpredictable randomness. Since chaotic behavior is an important characteristic, many researchers want to examine the existence of chaos in their systems. In this process, methods such as phase space portrait, power spectrum, Poincare mapping bifurcation diagram have been some of the most common methods used to determine chaotic behavior. However, the common point of these methods is that they are qualitative approaches. In other words, there is a need for an expert to interpret and evaluate the results. The chaos analysis

method known as Lyapunov exponents has become more popular than others because it is a quantitative approach.

The idea that fixed (invariant) exponents could be used to determine the stability states of the sets of differential equations of nonlinear dynamic systems was first shown by Sonya Kovalevskaya in 1889. Following the introduction of this hypothesis, it was based on theoretical foundations by Alexandr Mikhailovich Lyapunov. In the Lyapunov study, he explained only the basics of his thoughts about the change of trajectories of a dynamic system (as a function of time) with Lyapunov exponents. The starting point of the chaos analysis using the Lyapunov exponents is the dependence on the initial conditions and control parameters of chaotic systems. When a chaotic system is initiated from two very close neighboring initial conditions, chaos analysis can be carried out by moving the orbits away from each other or by convergence. Lyapunov exponents are a mathematical method that measures this distance between neighboring orbits. Lyapunov exponentials are likened to eigenvalues used in linear systems.

Lyapunov exponents can be calculated for continuous time system, discrete time systems and time series obtained from experimental or simulation results. Sensitivity to the initial conditions of a dynamic system is measured by Lyapunov exponents. Firstly, two trajectories have been determined with very close initial conditions on an attractor. If the attractor is showing chaotic behavior, the orbits are divided on an exponential rate, characterized by the largest Lyapunov exponent. The detection of a positive Lyapunov exponential is sufficient for the existence of chaos and indicates instability in a particular direction.

In this study, it is aimed to determine chaotic and periodic behaviors of information in synthetic datasets. In order to realize this aim, the properties of the databases have been converted to a time series data and the presence of positive Lyapunov exponents have investigated. It is planned to contribute to the process of evidence analysis with intelligent approaches in the future thanks to the chaotic and periodic behaviors to be discovered from the datasets.

AI and malware - future of malware and anti-malware algorithms

Ivan Zelinka

Department of Computer Science, Faculty of Electrical Engineering and Computer Science, VSB-Technical University of Ostrava, Czech Republic

In this talk, we outline a possible dynamics, structure, and behaviour of a hypothetical (up to now) swarm malware based on swarm intelligence and classical malware as an extension of the botnet. Our findings can serve also a background for a future antimalware system. We suggest how to capture and visualize the behaviour of such malware when it walks through the file system of an operating system. The swarm virus prototype, designed here, mimics a swarm system behaviour and thus follows the main idea underlying the swarm intelligence algorithms. The information of the prototype's behaviour is stored and visualized in the form of a complex network, reflecting virus communication and swarm behaviour. The network nodes are then individual virus instances. The network has certain properties associated with its structure that can be used by the virus instances in its activities like locating the target and executing a payload on the right object. As the paper shows, the swarm behaviour pattern can be incorporated also to an antimalware system and can be analyzed for a future computer system protection. In the end, we sketch how such a system can be hybridized with an artificial neural network in order to get a very robust swarm intelligence system, that can represent near future threats as well as its use in antimalware systems.

Implementation of artificial intelligence algorithms in digital forensics using Python programming language and its libraries

***Aleksandar Miljkovic*[1]*, Slobodan Nedeljkovic*[1]*, Milan Cabarkapa*[2] *and Aleksandar Jevremovic*[3]**

[1] Ministry of Interior, Sector for analytics, telecommunication and information technologies, Serbia

[2] School of electrical engineering, Belgrade, Serbia

[3] Sinergija University, Bosnia and Herzegovina

AI is based on the fact that there are generic algorithms that are able to extract useful and interesting information from a data set without having to write specific code to solve the problem, but it is sufficient to pass the data to the appropriate algorithm and algorithm will be able to learn from it. Eventually the user will get information from that data. For digital forensics AI is another tool in the toolbox that is helping law enforcement agencies (and corporate in-house investigators) comb through the available data for insights–digital needles in the proverbial haystack. AI functions can help with spotting and identifying elements in photos and videos, observing commonalities in communication, location, and times, and based on history, make educated guesses about where and when the next incident or crime might occur. Python is uniquely positioned as a programming language to perform cyber investigations and perform forensics analysis. While complex algorithms and versatile workflows stand behind machine learning and AI, Python's simplicity allows developers to write reliable systems. Therefore, AI researchers/developers could put all their effort into solving ML problem instead of focusing on the technical nuances of the language. All these advantages Python achieves through support of community, libraries and frameworks which are freely available on the Internet. In this research we are going to analyze and demonstrated the use of these libraries and frameworks in real-world examples (NLP based internet portal search and WhatsApp message list search).

Scene Analytics - Improving Situational Awareness using Artificial intuition

Florin Ciocan
Nokia, Romania

Authorities and companies are increasing their use of video and thermal cameras to monitor their cities and sites for safety and security threats. To get full value from these cameras, they need solutions that can help operations staff process massive amounts of video data, identify relevant footage and react to it in a timely way. Scene Analytics addresses these needs with computer vision technology and machine learning techniques that analyze raw video in real time and generate insights that help authorities and companies identify suspicious or unsafe activities. It turns cameras into IoT devices that enable energy companies to manage physical safety and critical infrastructure in a more proactive, automated and cost-effective way.

Building an anonymized dataset for AI algorithms

David Billard
University of Applied Sciences in Genova, Switzerland

The purpose of this talk is to introduce a dataset derived from a real case. For obvious privacy reasons, the identities of the people involved, their exchanged messages, as well as the name of the case itself have to be anonymized. This talk present the dataset and how the anonymization procedure is intended to be done. The audience will be prompted to interact and to provide leads to anonymization.

Can machine learning outperform standard biometrics?

Saša Adamović[1], Vladislav Miškovic[1], Nemanja Maček[2], Milan Milosavljević[1], Marko Šarac[1] and Milan Gnjatović[2,3]

[1] Faculty of Informatics and Computing, Singidunum University, Belgrade, Serbia
[2] Faculty of Computer Science, Megatrend University, Belgrade, Serbia
[3] Faculty of Technical Sciences, University of Novi Sad, Serbia

This paper deals with an iris recognition system based on machine learning methods. Unlike the pioneering work of Daugman, we managed to remove the needs for wavelet transformations. Yet, we transformed a classic normalized iris image into plain text and sub-sequentially extracted features that appear to be stylometric ones. Several databases containing iris images have been experimentally evaluated in order to bullet-proof the superiority of our algorithm. According to the experimental evaluation on these databases, the system performs a virtually perfect classification, with zero false acceptance rate, extremely low false rejection rates, extremely low template sizes, and computational costs.

Cyber Security Challenges in Connected Autonomous Vehicles

Jamal Raiyn
Computer Science Depatment
Al Qasemi Academic College, Baqa Al Gharbiah, Israel

An autonomous vehicle (AV) is a vehicle that operates and performs tasks under its own power. Some features of autonomous vehicle are sensing the environment, collecting information, and managing communication with other vehicles. Many autonomous vehicles in development use a combination of cameras, sensors, GPS, radar, LiDAR, and an on-board computer. These technologies work together to map the vehicle's position and its proximity to everything around it. AVs support various types of communication, such as, vehicle-to-vehicle (V2V), vehicle-to-infrastructure (V2I), and vehicle-to-mobile device (V2D) communication. However, they are susceptible to cyber- attacks, which take advantage of the situation when systems are easily accessible to an attacker. If a cyber attacker discovers a weakness in a certain vehicle type or a company's electronic system, a lack of information security could lead to criminal and terrorist acts that could eventually cost lives. The mostly used data communication security schemes are based on cryptographic approaches. The main goal of any encryption system is to disguise the classified message and render it unreadable to all unauthorized persons. Cryptographic systems are most often used in communication technologies. Encryption cannot prevent sent files from getting to the unauthorized person, but they can prevent the recipient from understanding their contents. The information to be secured is usually known in specialized terminology as plain text. The security process is called encryption. The rules for the encryption of plain text are called an encryption algorithm. The operations of this algorithm are derived from the encryption key, which, together with the text of the message, are the input information for the algorithm. If the recipient of the encrypted message wants to read the original, he or she has to use a decryption algorithm, which uses a decryption key to convert the encrypted text to the original.

This proposal gives an overview of scenarios of attacks on an AV. In particular, it focuses on cyber attack detection in AVs that use an augmented GNSS based on vehicle localization. The augmented GNSS has been designed as a cyber immune system for analyzing and detecting anomalous activities and their relation to malicious activities. The concept is to build in self-immunization through anomaly-based statistical measurements (see Figure 1). Furthermore, the augmented GNSS includes speed adaptation, estimation of vehicle location and cyber security. The cyber security in the augmented GNSS aims to secure stored and newly collected data. When the AV receives wrong data, the control unit in AV takes immediate measures, such as decreasing the speed. The proposed system model can detect the error that is caused by environmental factors or cyber attackers.

The novelty of the proposed approach lies in building in a self-immunization system and adding new components based on biometric data for message authentication and message communication. The proposed method provides enhanced security by combining biometrics features with randomly generated keys. The proposed method is based on eye human verification. A set of biometrics features is first extracted from the user's eye images. The feature extraction module provides discriminant and low dimension biometrics representation. The extracted features are then combined with user specific input, which is associated with a secret key, and the generated templates are stored for authentication. Furthermore, some information that is based on eye movement, such as nervousness and fatigue, can be obtained about divers.

Verified Computational Logic and European Transport Regulations

***Ana de Almeida Borges*[1], *Juan José Conejero Rodríguez*[1], *David Fernández-Duque*[2], *Mireia González Bedmar*[1], *Bjørn Jespersen*[3] *and Joost J. Joosten*[1]**

[1] University of Barcelona, Spain

[2] Ghent University, Belgium

[3] University of Utrecht, Netherlands

Our aim is to collaborate with industry, lawyers and legislators to develop verified legal software. The industrial and social need is evident: various legal decisions are made on the basis of algorithmic processing of data, which may lead to fines or imprisonments. Software tends to contain errors, but in the legal context no errors should be accepted.

In this presentation we will elaborate on the challenges involved and the tools we are employing in order to overcome them. For the sake of illustration, we will focus on a few problematic passages from the main European transport regulation.

Caveats for using synthetic data sets in financial fraud detection algorithms training

***Aleksandar Jevremovic*[1]*, Justin Bowling*[2] *and Zona Kostic*[2]**

[1] Sinergija University, Bosnia and Herzegovina

[2] Harvard University, USA

In this paper, we consider the problem of the feature extraction process in case of using synthetic data sets. Particularly, we analyze synthetic data sets for financial fraud detection. We analyze data sets from the perspective if the Benford's law is respected or not. However, the scope of this approach is much wider, since the Benford's distribution is present in many naturally generated data sets.

Deepfake and facemorphing detection

Zeno Geradts

Netherlands Forensic Institute The Hague

University of Amsterdam, Science Park 904, Amsterdam

In this presentation an overview will be given of different methods for deepfake-detection and face morphs that are used at the Netherlands Forensic Institute in current research. Altering faces by deepfakes as well as generating new faces based on Artificial Intelligence is a challenge for manipulation detection. Several methods exist on detecting these manipulations. One can detect if the camera fingerprint PRNU has been changed. For this it is preferable to have the camera that made the images. Also, methods with deep learning to detect the morphing with SVMs has been researched. Manual detection of image manipulation can also help with the guidelines of the Scientific Working Group of Digital Evidence. For forensic purposes one should keep in mind the time needed to make a deepfake as well as the chain of evidence. Once a method for detection is published, it is possible to improve the algorithms to make a better deepfake that is not easy to detect anymore.