Proceedings of the International Workshop on Applications of AI to Forensics 2020 (AI2Forensics 2020)

in association with
the 17th International Conference
on Principles of Knowledge Representation and Reasoning
(KR-2020)

AI2Forensics Organizers

Floris Bex, Utrecht University & Tilburg University, The Netherlands
Stefania Costantini, Universita' degli Studi dell'Aquila, Italy
Antonis Kakas, University of Cyprus, Cyprus
Raffaele Olivieri, Arma dei Carabinieri, Italy
Ken Satoh, National Institute of Informatics, Japan

September 14, 2020

Preface

International Workshop on Applications of AI to Forensics 2020 (AI2Forensics 2020) was held associated with 17th International Conference onPrinciples of Knowledge Representation and Reasoning (KR-2020) on September 14, 2020.

Digital forensics is a part of the criminalistics sciences which deals with digital evidence recovery and exploitation in the solution of criminal cases through the application of scientific principles. There are several and increasingly sophisticated methods for collecting digital evidence. As a matter of fact, the evolution of technology continuously pushes such kind of methods. Rough evidence must however be used to elicit hypotheses concerning events, actions and facts (or sequences of them) with the goal to obtain evidence to present in court. Evidence analysis involves examining fragmented incomplete knowledge, and reconstructing and aggregating complex scenarios involving time, uncertainty, causality, and alternative possibilities. No established methodology exists today for digital evidence analysis.

We had three submissions for technical papers and two demonstration papers and all papers were presentend. Furthermore, as invited speakers, we had Francesca A. Lisi from Università degli Studi di Bari "Aldo Moro", Italy. She gave a talk titled "Combining Knowledge Representation and Machine Learning in Forensics".

And last but not least, we would like to thank all the authors for the papers and the members of PC for reviewing the papers.

September 2020

AI2Forensics Co-organizers
Floris Bex, Utrecht University & Tilburg University, The Netherlands
Stefania Costantini, Universita' degli Studi dell'Aquila, Italy
Antonis Kakas, University of Cyprus, Cyprus
Raffaele Olivieri, Arma dei Carabinieri, Italy
Ken Satoh, National Institute of Informatics, Japan

PC members

David BILLARD, University of Applied Sciences in Geneva, Switzerland
Rafael BORDINI, FACIN-PUCRS, Brasil
Pedro CABALAR, Universidade da Coruña, Spain
Zeno GERADTS, Netherlands Forensic Institute, The Netherlands
Aleksandra DEDINEC, The Ss. Cyril and Methodius University in Skopje, North Macedonia
Martin DIEGUEZ, ENIB – Lab-STICC, France
Hans HENSELER, Magnet Forensics, The Netherlands
Aleksandar JEVREMOVIC, Sinergija University, Bosnia and Herzegovina
Viviana MASCARDI, University of Genova, Italy
Jesus MEDINA, Universidad de Cadiz, Spain
Jelte MENSE, Netherlands National Police, The Netherlands
Alessandra MILEO, Dublin City University, Ireland
Yoshiaki NISHIGAI, Chiba University, Japan
Manuel OJEDA ACIEGO, Universidad de Málaga, Spain
Alessandro PROVETTI, Birbeck University of London, UK
Bas TESTERINK, Netherlands National Police, The Netherlands
Wataru ZAITSU, Mejiro University, Japan

Table of Contents

AI2Forensics

Combining Knowledge Representation and Machine Learning in Forensics

Francesca A. LISI

Dipartimento di Informatica & Centro Interdipartimentale di Logica e Applicazioni (CILA),
Università degli Studi di Bari "Aldo Moro", Italy
FrancescaAlessandra.Lisi@uniba.it

Abstract

This invited talk overviews 20 years of work at the intersection between the two AI areas of Knowledge Representation (KR) and Machine Learning (ML). The distinguishing feature of this research is the extension of the methodological apparatus of Inductive Logic Programming (ILP) along a couple of directions towards the realm of Description Logics (DLs). One aims at learning hybrid rules that tightly integrate DATALOG and DLs, whereas the other aims at learning axioms in fuzzy DLs. Both have turned out to be alternative suitable ways to treat spatial knowledge in several applications and could be successfully applied also in the field of Forensics.

1 Introduction

Inductive Logic Programming (ILP) (Muggleton 1990) was proposed in the early 90s as a powerful setting for Concept Learning (Mitchell 1982) within the framework of Logic Programming (LP) (Lloyd 1987), often limited to the fragment of DATALOG (Ceri, Gottlob, and Tanca 1990) for computational reasons. ILP has been historically focused on learning rules from examples and background knowledge with the aim of prediction. However, it has also been applied to tasks other than classification - such as association rule mining - where the scope of induction is description rather than prediction. Notable examples of ILP systems supporting these two kinds of tasks are FOIL (Quinlan 1990) and WARMR (Dehaspe and Toivonen 1999).

With the advent of ontologies, mostly expressed with languages based on the family of Description Logics (DLs) (Baader et al. 2003), new challenges and opportunities have been presented to ILP. This invited talk summarizes the work done in ILP over the past 20 years, first on learning so-called onto-relational rules (Section 2) and later on learning fuzzy ontology axioms (Section 3). Both extensions of ILP beyond the original setting provide means for representing and reasoning over spatial knowledge, thus showing a potential for application also in Forensics, especially during the phase of Evidence Analysis where the spatial dimension plays a key role in the analysis of evidences related to crime scenarios (Costantini, Lisi, and Olivieri 2019).

2 Learning onto-relational rules with ILP

LP and DLs are both based on fragments of First Order Logic (FOL). However, they are characterized by different semantic assumptions (Motik and Rosati 2010). Though a partial overlap exists between LP and DLs, even more interesting is a combination of the two via several integration schemes that are aimed at designing very expressive FOL languages and ultimately overcoming the aforementioned semantic mismatch (see, *e.g.*, (Drabent et al. 2009) for a survey). A popular example of this class of hybrid KR formalisms is $\mathcal{AL}$-LOG (Donini et al. 1998) which tightly integrates DATALOG and $\mathcal{ALC}$. Several works in ILP testify the great potential of these formalisms also from the perspective of machine learning and inductive reasoning (Rouveirol and Ventos 2000; Kietz 2003; Lisi 2008; Lisi 2010; Lisi 2014). Originally motivated by a spatial data mining application (Appice et al. 2003; Lisi and Malerba 2004) and inspired by WARMR, $\mathcal{AL}$-QUIN (Lisi 2011) is an ILP system for mining association rules at multiple levels of granularity. It can perform taxonomic reasoning, *e.g.*, over hierarchies of spatial objects such as regions, by relying on the KR framework of $\mathcal{AL}$-LOG.

3 Learning fuzzy ontology axioms with ILP

Spatial notions such as the distance between two sites can be naturally represented with fuzzy sets if one is interested in their human perception rather than in precise measurements. In order to deal with imprecision in Ontology Reasoning several fuzzy extensions of DLs have been proposed (see, *e.g.*, (Straccia 2015) for an overview). However, the problem of automatically managing the evolution of fuzzy DL ontologies still remains relatively unaddressed (Konstantopoulos and Charalambidis 2010; Iglesias and Lehmann 2011). Lisi and Straccia (2013) propose *Soft*FOIL, a FOIL-like method for learning fuzzy $\mathcal{EL}$ GCI axioms from fuzzy DL assertions. In (Lisi and Straccia 2014), the same authors present FOIL-$\mathcal{DL}$, another FOIL-like method which, conversely, is designed for learning fuzzy $\mathcal{EL}(\mathbf{D})$ GCI axioms from crisp DL assertions. As opposite to *Soft*FOIL, FOIL-$\mathcal{DL}$ has been implemented and tested (Lisi and Straccia 2015), notably in a real-world tourism application. More recently, a granular computing method for OWL 2 ontologies has been proposed in (Lisi and Mencar 2018) with the ultimate goal of optimizing the learning process when dealing with a huge number of relations, *e.g.*, those concerning the distance between places.

Acknowledgments

This talk is partly based upon work from COST Action 17124 "Digital forensics: evidence analysis via intelligent systems and practices (DigForASP)", supported by COST (European Cooperation in Science and Technology).

References

Appice, A.; Ceci, M.; Lanza, A.; Lisi, F. A.; and Malerba, D. 2003. Discovery of spatial association rules in georeferenced census data: A relational mining approach. *Intelligent Data Analysis* 7(6):541–566.

Baader, F.; Calvanese, D.; McGuinness, D.; Nardi, D.; and Patel-Schneider, P., eds. 2003. *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press.

Ceri, S.; Gottlob, G.; and Tanca, L. 1990. *Logic Programming and Databases*. Springer.

Costantini, S.; Lisi, F. A.; and Olivieri, R. 2019. DigForASP: A European cooperation network for logic-based AI in digital forensics. In Casagrande, A., and Omodeo, E. G., eds., *Proceedings of the 34th Italian Conference on Computational Logic, Trieste, Italy, June 19-21, 2019*, volume 2396 of *CEUR Workshop Proceedings*, 138–146. CEUR-WS.org.

Dehaspe, L., and Toivonen, H. 1999. Discovery of frequent DATALOG patterns. *Data Mining and Knowledge Discovery* 3:7–36.

Donini, F. M.; Lenzerini, M.; Nardi, D.; and Schaerf, A. 1998. $\mathcal{AL}$-log: Integrating Datalog and Description Logics. *Journal of Intelligent Information Systems* 10(3):227–252.

Drabent, W.; Eiter, T.; Ianni, G.; Krennwallner, T.; Lukasiewicz, T.; and Maluszynski, J. 2009. Hybrid Reasoning with Rules and Ontologies. In Bry, F., and Maluszynski, J., eds., *Semantic Techniques for the Web, The REWERSE Perspective*, volume 5500 of *Lecture Notes in Computer Science*. Springer. 1–49.

Iglesias, J., and Lehmann, J. 2011. Towards integrating fuzzy logic capabilities into an ontology-based inductive logic programming framework. In *Proc. of the 11th Int. Conf. on Intelligent Systems Design and Applications*. IEEE Press.

Kietz, J.-U. 2003. Learnability of description logic programs. In Matwin, S., and Sammut, C., eds., *Inductive Logic Programming, 12th International Conference, ILP 2002, Sydney, Australia, July 9-11, 2002. Revised Papers*, volume 2583 of *Lecture Notes in Computer Science*, 117–132. Springer.

Konstantopoulos, S., and Charalambidis, A. 2010. Formulating description logic learning as an inductive logic programming task. In *Proc. of the 19th IEEE Int. Conf. on Fuzzy Systems*, 1–7. IEEE Press.

Lisi, F. A., and Malerba, D. 2004. Inducing Multi-Level Association Rules from Multiple Relations. *Machine Learning* 55:175–210.

Lisi, F. A., and Mencar, C. 2018. A granular computing method for OWL ontologies. *Fundamenta Informaticae* 159(1–2):147–174.

Lisi, F. A., and Straccia, U. 2013. A logic-based computational method for the automated induction of fuzzy ontology axioms. *Fundamenta Informaticae* 124(4):503–519.

Lisi, F. A., and Straccia, U. 2014. A FOIL-like Method for Learning under Incompleteness and Vagueness. In Zaverucha, G.; Santos Costa, V.; and Paes, A., eds., *Inductive Logic Programming - 23rd International Conference, ILP 2013, Rio de Janeiro, Brazil, August 28-30, 2013, Revised Selected Papers*, volume 8812 of *Lecture Notes in Computer Science*, 123–139. Springer.

Lisi, F. A., and Straccia, U. 2015. Learning in description logics with fuzzy concrete domains. *Fundamenta Informaticae* 140(3-4):373–391.

Lisi, F. A. 2008. Building Rules on Top of Ontologies for the Semantic Web with Inductive Logic Programming. *Theory and Practice of Logic Programming* 8(03):271–300.

Lisi, F. A. 2010. Inductive Logic Programming in Databases: From Datalog to $\mathcal{DL}$+log. *Theory and Practice of Logic Programming* 10(3):331–359.

Lisi, F. A. 2011. $\mathcal{AL}$-QUIN: An Onto-Relational Learning System for Semantic Web Mining. *International Journal on Semantic Web and Information Systems* 7(3):1–22.

Lisi, F. A. 2014. Learning onto-relational rules with inductive logic programming. In Lehmann, J., and Völker, J., eds., *Perspectives on Ontology Learning*, volume 18 of *Studies on the Semantic Web*. IOS Press/AKA. 93–111.

Lloyd, J. W. 1987. *Foundations of Logic Programming*. Springer, 2nd edition.

Mitchell, T. M. 1982. Generalization as search. *Artificial Intelligence* 18:203–226.

Motik, B., and Rosati, R. 2010. Reconciling description logics and rules. *J. ACM* 57(5).

Muggleton, S. H. 1990. Inductive logic programming. In Arikawa, S.; Goto, S.; Ohsuga, S.; and Yokomori, T., eds., *Proceedings of the 1st Conference on Algorithmic Learning Theory*. Springer/Ohmsma.

Quinlan, J. R. 1990. Learning logical definitions from relations. *Machine Learning* 5:239–266.

Rouveirol, C., and Ventos, V. 2000. Towards Learning in CARIN-$\mathcal{ALN}$. In Cussens, J., and Frisch, A. M., eds., *Inductive Logic Programming, 10th International Conference, ILP 2000, London, UK, July 24-27, 2000, Proceedings*, volume 1866 of *Lecture Notes in Artificial Intelligence*, 191–208. Springer.

Straccia, U. 2015. All about fuzzy description logics and applications. In Faber, W., and Paschke, A., eds., *Reasoning Web. Web Logic Rules - 11th International Summer School 2015, Berlin, Germany, July 31 - August 4, 2015, Tutorial Lectures*, volume 9203 of *Lecture Notes in Computer Science*, 1–31. Springer.

Machine Learning to predict London crime rates

Lorenzo De Lauretis[1] , **Stefania Costantini**[1]
[1]Università Degli studi dell'Aquila, Italy
lorenzo.delauretis@graduate.univaq.it, stefania.costantini@univaq.it

Abstract

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, and in particular to various problems related to Evidence Analysis in Digital Forensics. 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, 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 were 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 be used to help the police districts for the assignment task of policemen, cars and other resources, increasing the attention concerning districts with an increasing crime rate.

1 Introduction

Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems can use in order to perform a specific task effectively without using explicit instructions, relying on patterns and inference instead. Machine learning algorithms build a mathematical model based on sample data, known as training data, in order to make predictions or decisions without being explicitly programmed to perform the task (Bishop 2006; Koza et al. 1996).

Deep learning is part of a broader family of machine learning methods based on neural networks. Learning can be supervised, semi-supervised or unsupervised (Bengio, Courville, and Vincent 2013; Schmidhuber 2015).

Deep Learning architectures such as deep neural networks, deep belief networks, recurrent neural networks and convolutional neural networks have been applied to a lot of fields, including computer vision, speech recognition, natural language processing, audio recognition, medical image analysis, and so on. They have produced results comparable to and in some cases superior to human expert (Cireşan, Meier, and Schmidhuber 2012).

In our work, we have applied such techniques to examine a dataset concerning London crime data, from January 2008 to December 2016, represented using a CSV file. Elaborating those data through the software RapidMiner[1], we have been able to create a neural net that provided us with useful statistical data about the crimes committed in the districts of London. Applying machine learning techniques, such as Deep Learning, via our neural net, we have been able to make classifications on the crime data, to the aim to discover whenever there is an increase of a particular kind of crime in certain areas of London city.

This can help the police districts into the assignment task of policemen, cars and other resources, thus increasing the attention concerning districts with an increasing crime rate.

In Section 2, we show some work related to ours. In Section 3 we can see a brief introduction to Machine Learning and Deep Learning. In Section 4 it is possible to see an explanation of the dataset we used to build our work. In Section 5, we can see how we created our neural network using rapid miner. In Section 6 it is possible to see how we applied Deep Learning to our neural network and a working example of our strategy. In Section 7 we discuss about our strategy and in Section 8 we concludes our work.

2 Related Work

In (Lin, Chen, and Yu 2017), the authors use Deep Learning to predict Drug-related criminal activity in Taiwan. They improved model performance by accumulating data with different time scales. In their work, they visualize potential crime hotspots on a map and observe whether the created models can identify true hotspots. Differently from them, we do not use a map, we relate to a large CSV file with all the crimes data.

In (McClendon and Meghanathan 2015), the authors use WEKA (we uses RapidMiner), an open-source data mining software, to conduct a comparative study between the violent crime patterns from the Communities and Crime Unnormalized Dataset and actual crime statistical data for the state of Mississippi. Differently from us, they used Linear Regression and Additive Regression to identify crime patterns, instead of the Deep Learning algorithm used by us.

3 Machine Learning and Deep Learning

Machine Learning is the investigation of algorithms that improve naturally through experience. It can be viewed as a

[1]https://rapidminer.com/

subset of man-made "consciousness". ML algorithms fabricate a scientific model dependent on test information, known as "training data", to settle on expectations or choices without being expressly modified to do so(Koza et al. 1996). Machine learning algorithms are utilized in a wide assortment of uses, for example, email filtering and computer vision, where it is troublesome or infeasible to create regular algorithms to play out the needed tasks. Machine learning is closely related to computational statistics, which focuses on making predictions using computers. ML includes computers learning from information given so that they carry out certain assignments. For basic tasks relegated to computers, it is conceivable to program algorithms telling the machine how to execute all steps required to unravel the issue at hand; on the computer side, no learning is required. For more progressed assignments, it can be challenging for a human to physically make the required calculations. In practice, it can turn out to be more viable to assist the machine to develop its claim algorithm, instead of having human software engineers indicate each required step.(Alpaydin 2020)

Early classifications for machine learning approaches separated them into three wide categories, depending on the nature of the "flag" or "input" accessible to the learning framework. These were:

- Supervised learning: The computer is displayed with case inputs and their craved yields, given by an "instructor", and the objective is to memorize a common rule that maps inputs to outputs.
- Unsupervised learning: No predefined classification is provided to the learning algorithm, relying upon its capabilities to discover structure in its input. Unsupervised learning can be an objective in itself (finding covered up designs in information) or a step towards a conclusion (feature learning).
- Reinforcement learning: A computer program interacts with a dynamic environment in which it must perform a certain objective (such as driving a vehicle or playing a game against a rival). This technique is thus concerned with the problem of identifying suitable actions to take in a given situation in order to maximize a reward. So, the learning algorithm is not given examples of optimal outputs, in contrast to supervised learning, but must instead discover them by a process of trial and error (Bishop 2006).

Deep learning is part of a broader family of machine learning methods based on artificial neural networks with representation learning. In particular, deep learning is a class of machine learning algorithms that uses multiple layers to progressively extract higher-level features from the raw input. For example, in image processing, lower layers may identify edges, while higher layers may identify the concepts relevant to a human such as digits or letters or faces(Deng and Yu 2014). We used Deep Learning instead of other methods such as Random Forest and Naïve Bayes because, with fine-tuning, they provide better predictions(Lin, Chen, and Yu 2017).

4 The Dataset

The dataset we used in our work is about London crimes data, from 2008 to 2016[2]. In this dataset (in a CSV form), it is possible to find all crimes committed in the city of London from the year 2008 to the year 2016, subdivided into categories and borough. The dataset has the following data structure:

- LSOA_CODE: Code for Lower Super Output Area in Greater London
- BOROUGH: Common name for London borough.
- MAJOR_CATEGORY: High-level categorization of crime
- MINOR_CATEGORY: Low-level categorization of crime within the major category
- VALUE: Monthly reported count of categorical crime in a given borough
- YEAR: Year of reported counts
- MONTH: Month of reported counts

The dataset is composed of over 13 millions entries, being much too large to set up our neural network. Therefore, we decided to split up our dataset by years, using only the data about 2015/2016 to set-up our neural network and make predictions using deep learning algorithms.

5 Neural Network Creation

As described in the previous section, our whole dataset is too large to be used in the selected Machine Learning software, requiring a very long time for computation, due to our mid-range PC. We decided to split up our dataset by years, using only the crime data from the year 2015/2016 to set-up our neural network. We began loading the data into the selected software, RapidMiner. It has a very nice import functionality, that automatically understands the data type of each column. Later, always using RapidMiner, we split up our dataset, creating a smaller one, that considered only the 2015/2016 crime data.

We started the Auto-Modeling functionality of RapidMiner, having as input our newly created dataset. Their auto-model functionality allows us to simply set-up a neural network in minutes, describing the input data that we want to use. This newborn neural network is a feed-forward one, trained by a back propagation algorithm (multi-layer perceptron). It also shows us the correlations among data; this will be useful to the predictions we are going to calculate.

6 Deep Learning and Predictions

After creating our Neural Network through RapidMiner auto-model functionality, we applied a Deep Learning algorithm on it. Using this software, it just requires a few steps. One simply selects the newborn neural network, and chooses the Deep Algorithm among all those available. It is also possible to customize the deep learning functionalities, through manual settings. We set-up our system in order to

[2]https://www.kaggle.com/jboysen/London-crime

make classifications on our crimes and predictions about the crime increase/decrease in particular areas.

We choose as input data the columns: Borough, Minor_Category, Month and Year, being the more appropriate ones for the dataset we have. As data to be predicted by our neural network with deep learning algorithm, we choose the number of crimes that can potentially be committed in that month/year in the selected borough.

After setting-up our whole system, we launched our Deep Learning algorithm on our neural network. In a certain amount of time, the process ended, showing up interesting results. Those results can be seen through a simple interface, called *Simulator*. In the simulator, you set-up the Borough you are interested, the Minor_Category, the Month and the Year, and it returns the predicted value that can happen in that particular year/month, in that Borough, of that particular kind of crime.

6.1 Simulation example

In our environment, we extracted a particular case using our Simulator, that is a crime prediction. To begin, we extracted real data: all the thefts happened in the borough of Westminster in 12/2016. The result that the simulator returned is 4.813, as it is possible to see in Figure 1, that is the number of thefts happened in Westminster in that particular period. Later, we set-up the Date of our simulator in the year 2019: in this way, we are using Deep Learning to predict the increase/decrease of thefts in Westminster, with respect to the year 2016. From the data we obtained, we can see that the simulator returned a value of 4.730, as it is possible to see in Figure 1, that is lower than the value obtained in 2016, deriving that the thefts are lowering in Westminster borough in December in the following three years.

7 Discussion

By using a neural network, we are able to make previsions on the crime rate increase/decrease in a particular London borough in a particular year/month combination. The main problem of our strategy, is that we are not able to know the reasons underlying the increase or decrease of crime rate and the real happenings of crime events; if, for example, crimes increases due to a particular event in that zone in that place, our system is not able to say the reasons of the increase in the predicted data. The dataset that we have analyzed unfortunately lacks more specific information, such as the precise date of each crime, its geo-localization and the profiles of perpetrators and victims, features that would have made our results more interesting. Actually, it happens what is emphasized in (Pearl and Mackenzie 2018), we are unable to understand the reasons underlying the increase/decrease of crime rates in various areas. Therefore, as future work we intend to create a background knowledge base so as to be able to process ML outcome in order to devise, in a "white box" fashion, a causal explanation of results. For example, an increase of thefts in the houses in the Westminster borough might be related to wealthy houses left unattended during vacations, or a high incidence of scams in some other borough might be related to a high percentage of low-culture elderly population. We believe that Inductive Logic

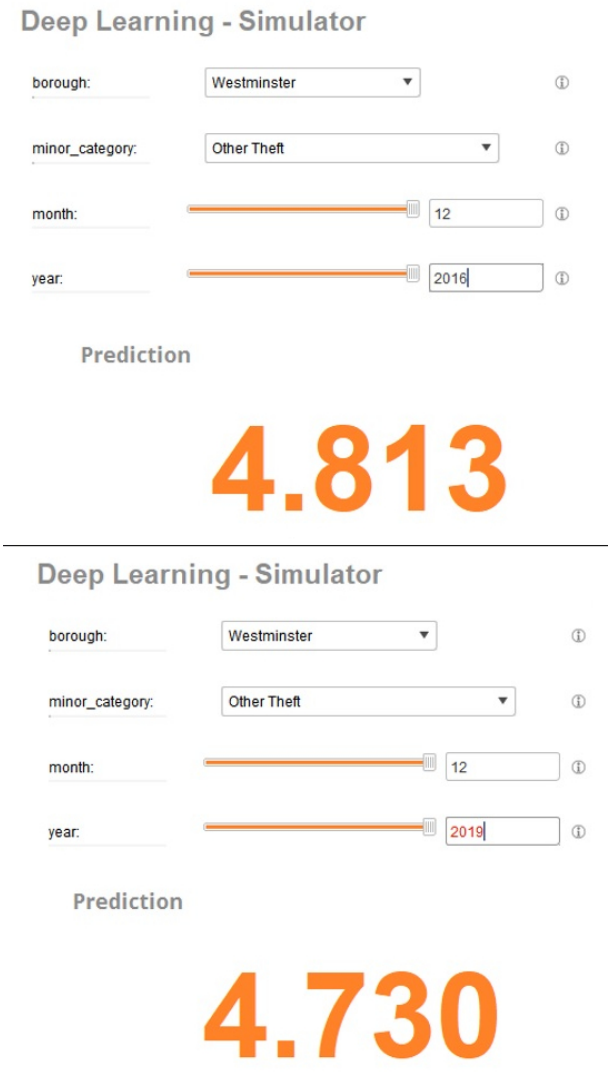

Figure 1: Our Simulation Example

Programming with Answer Set Programming in the background (Brewka, Eiter, and (eds.) 2016; Muggleton 1991; Law, Russo, and Broda 2019), might be suitable tools for the envisaged analysis.

8 Conclusions and Future Work

Our work discusses an interesting way to predict the incidence of crime in particular places, allowing the police to better assign policemen in particular places during a certain period of the year. Thanks to our strategy, it is possible to assign fewer policemen in zones where the crime rate is lowering, and assigning more in zones with increasing crime rate, allowing the Police to do a better resources allocation, useful for crime prevention.

To tackle the problems we saw in Section 7, 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 learns rules, that should represent causal connections extracted from data to understand "why" they increase/decrease in each specific area, assuming that provided data are richer than those we have examined. In complement, forms of reasoning such as Answer Set Programming (ASP) might elicit future plausible scenarios of crime distribution. To do this, however, the datasets to be analyzed should be richer of significant features.

To conclude, our work can be very useful to Police to better understand how to develop their forces into the relative locations, increasing the effectiveness of crimes prevention.

References

Alpaydin, E. 2020. *Introduction to machine learning*. MIT press.

Bengio, Y.; Courville, A.; and Vincent, P. 2013. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence* 35(8):1798–1828. DOI: 10.1109/TPAMI.2013.50.

Bishop, C. M. 2006. *Pattern recognition and machine learning*. springer. ISBN: 0387310738.

Brewka, G.; Eiter, T.; and (eds.), M. T. 2016. Answer set programming: Special issue. *AI Magazine* 37(3).

Cireşan, D.; Meier, U.; and Schmidhuber, J. 2012. Multi-column deep neural networks for image classification. *arXiv preprint arXiv:1202.2745*.

Deng, L., and Yu, D. 2014. Deep learning: Methods and applications. Technical Report MSR-TR-2014-21, Microsoft.

Koza, J. R.; Bennett, F. H.; Andre, D.; and Keane, M. A. 1996. *Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming*. Dordrecht: Springer Netherlands. 151–170. DOI: 10.1007/978-94-009-0279-4_9.

Law, M.; Russo, A.; and Broda, K. 2019. Logic-based learning of answer set programs. In Krötzsch, M., and Stepanova, D., eds., *Reasoning Web. Explainable Artificial Intelligence - 15th International Summer School 2019, Bolzano, Italy, September 20-24, 2019, Tutorial Lectures*, volume 11810 of *Lecture Notes in Computer Science*, 196–231. Springer.

Lin, Y.-L.; Chen, T.-Y.; and Yu, L.-C. 2017. Using machine learning to assist crime prevention. In *2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, 1029–1030. IEEE.

McClendon, L., and Meghanathan, N. 2015. Using machine learning algorithms to analyze crime data. *Machine Learning and Applications: An International Journal (MLAIJ)* 2(1):1–12.

Muggleton, S. 1991. Inductive logic programming. *New generation computing* 8(4):295–318.

Pearl, J., and Mackenzie, D. 2018. *The book of why: the new science of cause and effect*. Basic Books.

Schmidhuber, J. 2015. Deep learning in neural networks: An overview. *Neural networks* 61:85–117. DOI: 10.1016/j.neunet.2014.09.003.

Qualitative Spatial Reasoning for Digital Forensics: A Cardinal Directional Calculus Approach using Answer Set Programming

Yusuf Izmirlioglu[1] , **Esra Erdem**[2]

[1]Sabanci University, Faculty of Engineering and Natural Sciences
[2]Sabanci University, Faculty of Engineering and Natural Sciences
yizmirlioglu@sabanciuniv.edu, esra.erdem@sabanciuniv.edu

Abstract

In our earlier studies, we have introduced a framework (called nCDC-ASP) to reason about cardinal directions, based on Cardinal Directional Calculus, using the computational methods of Answer Set Programming. In this study, we describe an application of nCDC-ASP to digital forensics, and discuss its usefulness with a scenario.

1 Introduction

Digital Forensics is the science of recovering, validating and interpreting of data and artifacts from digital sources for the purpose of facilitating criminal investigation, private/corporate inspection or intelligence. The digital source can be a hard drive, smart phone, camera, server, cloud or a network component. There are 15 legal cases documented in (Brainz 2020) which would not be resolved without the assistance of digital forensics.

The key phases of a digital forensics process are recovery of the state of the device, analyzing its state to find data, artifact and reporting clean and processed information as an evidence to the legal institutions.

However digital sources may provide partial, incomplete or uncertain information about the environment. Consequently, inspectors aim to find out additional information about the true state of the world at the moment of the criminal event from the interrogations of the eyewitnesses and perhaps suspects.

One critical aspect of a criminal investigation is identifying honesty of suspects of a criminal event because they are usually the main actors or observers of the event. In this respect, the inspectors question the suspects and examine whether their statement is consistent or conflicts with the eyewitnesses and digital evidence. Therefore further reasoning is necessary to compare or confirm statement of the suspects with the digital evidence. Traditionally this procedure is manually performed by police officers. In this research, we propose an Artificial Intelligence tool to automate spatial aspect of this procedure with our consistency checking framework.

Clues and evidences relevant for a forensics process are often items such as a knife, gun, shell, drug. Additional information is related to the situation of the event venue and objects, position of the above murder elements among other objects, furniture, clothes inside the same room. In this paper we study reasoning about space in a digital forensics process in terms of locations of objects, people and configuration of a room during the crime. The spatial information typically comes from digital sources, interrogations of eyewitnesses and suspects.

Qualitative spatial reasoning studies representation and reasoning with different aspects of space such as direction, distance, size, shape using coarse, inexact, imprecise terms of human language rather than quantitative data. Over the last three decades qualitative spatial reasoning has been successfully applied to geographical information systems, cognitive robotics, computer graphics, spatial databases, building design and regulation, bioengineering. Qualitative models are useful in contexts where quantitative data is not available due to incomplete knowledge or uncertainity e.g., exploration of an unknown territory. Even if quantitative data is available, qualitative spatial reasoning is also relevant in contexts with human presence. Human agents tend to express spatial relation or configuration by means of qualitative terms such as “left, right, front, back, near, far” for the sake of sociable and convenient communication. Thus qualitative models are more suitable for representing and reasoning about spatial relations in these environments.

In our earlier studies (Izmirlioglu and Erdem 2018), we have introduced a formal framework (called nCDC-ASP) using Answer Set Programming (ASP) (Marek and Truszczyński 1999; Niemelä 1999; Lifschitz 2002), for representing and reasoning about cardinal directions, based on Cardinal Directional Calculus (CDC). CDC has been introduced by (Goyal and Egenhofer 1997; Skiadopoulos and Koubarakis 2004; 2005) to represent and reason with directional relations. In CDC, relative direction of an object with respect to one another is described using cardinal directions, e.g., north/south, east/west, onto which are more natural for verbal descriptions. Depending on the context, analogous terms such as up/below, right/left can be used. CDC can also express overlapping items and parthood relation to some degree.

One of the central problems in CDC and qualitative spatial reasoning literature is checking consistency of a given set of CDC constraints. Consistency checking asks for whether a feasible configuration of the objects exists on the plane which satisfy the given CDC constraints. Suppose

statements of suspects include claims such as “the knife is to the right of the body”, “I saw a gun on the table”, “there was a cell phone on the floor, right back to the teapoy”. Each claim can be represented as a CDC constraint. Suppose also that the detective has some information about locations of the objects in the venue provided by the camera images. Then the detective can test whether the statements make sense or not, by checking the consistency of the corresponding CDC constraints together with the digital data.

nCDC-ASP utilizes ASP for checking the consistency of CDC constraints. In particular, it represents the CDC constraints and the meaning of CDC relations as a program in the expressive formalism of ASP, and then calls the ASP solver CLINGO to check whether the program has an answer set. If the program has an answer set then the CDC constraints are consistent; otherwise, they are inconsistent.

In this study, we discuss how our framework for consistency checking in CDC can be used to resolve honesty of suspects of a criminal event in the sense that the statement of the suspects fits or conflicts with the existing facts and evidence. This framework is capable for handling incomplete knowledge and uncertainty in spatial relations. We refer the reader to (Izmirlioglu and Erdem 2018) for details about nCDC-ASP.

2 Consistency Checking using nCDC-ASP

A brief overview of CDC. Cardinal directional calculus defines qualitative orientation of a spatial object a (the primary or target region) with respect to another object b (the reference region) by cardinal direction relations. The minimum bounding box of an object a, denoted $mbr(a)$, is the smallest rectangle which contains a and has sides parallel to the axes. Sides of $mbr(a)$ are the straight lines $x = \inf_x(a)$, $x = \sup_x(a)$, $y = \inf_y(a)$ and $y = \sup_y(a)$. The minimum bounding rectangle of the reference object divides the plane into nine regions (called tiles) and these tiles define the nine cardinal directions: *north* (N), *south* (S), *east* (E), *west* (W), *northeast* (NE), *northwest* (NW), *southeast* (SE), *southwest* (SW), *on* (O) as in Figure 1(i). By identifying the unique tiles $R_1(b), ..., R_k(b)$ $(1 \leq k \leq 9$ occupied by the primary object a, direction of a with respect to b is shown by the basic CDC relation $R_1{:}R_2{:}...{:}R_k$. For example, according to Figure 1(ii), $a\ E:NE\ b$ since a occupies a region in $E(b)$ and $NE(b)$. A disjunctive CDC relation is a finite set $\delta = \{\delta_1, ..., \delta_o\}$, $o > 1$ of basic CDC relations, intuitively describing their exclusive disjunction. A CDC relation can be basic or disjunctive. A formula of the form $u\ \delta\ v$, where u and v are spatial variables and δ is a CDC relation, is called a CDC constraint. A CDC constraint network C is a set of CDC constraints defined by spatial variables $V = \{v_1, ..., v_l\}$. C is consistent if there exists an instantiation of objects which satisfies all constraints in C.

As shown in (Izmirlioglu and Erdem 2018), checking consistency of a CDC network can be discretized and solved over a grid of size $m \times n$ where $m = n = 2|V| - 1$. In the discrete domain, we impose the following two conditions to ensure that CDC constraints hold. We say that a pair (a, b) of spatial objects on the grid satisfies a basic CDC constraint

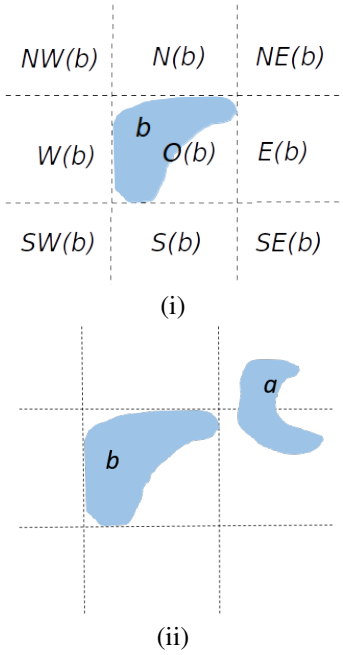

Figure 1: (i) Nine tiles with respect to a reference object b (ii) A basic CDC relation that describes orientations of a with respect to b: $a\ NE{:}E\ b$ (“Some part of a is in $NE(b)$ and the rest of a is in $E(b)$”)

$u\ \delta\ v$ in C if

(C1) $a \cap R^{m,n}(b) \neq \emptyset$ for every single-tile relation R in δ, and

(C2) $a \cap R^{m,n}(b) = \emptyset$ for every single-tile relation R that is not included in δ.

We formulate CDC consistency checking in ASP, as explained in (Izmirlioglu and Erdem 2018).

Represent the input. We represent the given constraint network C in ASP by a set of facts. We describe every basic CDC constraint $u\ \delta\ v$ in C where $\delta = R_1{:}R_2{:}...{:}R_k$, $(k \geq 1)$, by atoms of the form $rel(u, v, r)$ for each single-tile relation r in δ.

$$rel(u, v, R_i) \leftarrow . \tag{1}$$

For example, a basic nCDC constraint $u\ N{:}NE\ v$ is represented by the facts:

$$rel(u, v, N) \leftarrow .$$
$$rel(u, v, NE) \leftarrow .$$

The answer set for the program (1) characterizes the input network C.

Generate assignments of spatial objects to variables. An assignment of a nonempty set of grid cells $(x, y) \in \Lambda_{m,n}$ to every variable $u \in V$ is generated by a set of choice rules as follows:

$$1\{occ(u, x, y) : (x, y) \in \Lambda_{m,n}\} \leftarrow . \tag{2}$$

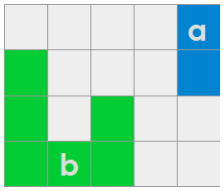

Figure 2: Two objects a, b on the grid satisfies a CDC constraint $u\ N{:}NE\ v$

Note that these choice rules are augmented by a cardinality constraint to ensure that at least one grid cell is assigned to every variable.

Minimum Bounding Rectangle. To check whether a generated assignment satisfies every basic CDC constraint $u\ \delta\ v$ in C, first we identify the minimum bounding rectangle $mbr^{m,n}(v)$ of the spatial object in $D_{m,n}$ assigned to v.

CDC Constraints We need to ensure that the instantiation of objects (by assignment of cells) to variables $u \in V$ satisfies every basic nCDC constraint $u\ \delta\ v$ in C. For this, we add rules to impose that conditions (C1) and (C2) are satisfied. For example, if δ contains the single tile relation N(north), then the rule below ensures condition (C1) for N: u must occupy some grid cells to the north of $mbr^{m,n}(v)$.

$$\leftarrow \{occ(u,x,y) :\ \underline{x} \leq x \leq \overline{x},\ y{>}\overline{y},\ (x,y) \in \Lambda_{m,n}\}0,\\ rel(u,v,N),\ inf_x(v,\underline{x}),\ sup_x(v,\overline{x}),\ sup_y(v,\overline{y}) \quad (u \in V). \tag{3}$$

If δ does not contain N, u must not occupy any grid cells to the north of $mbr_{m,n}(v)$ in accordance with condition (C2).

$$\leftarrow 1\{occ(u,x,y) :\ \underline{x} \leq x \leq \overline{x},\ y{>}\overline{y},\ (x,y) \in \Lambda_{m,n}\},\\ not\ rel(u,v,N),\ existrel(u,v),\ inf_x(v,\underline{x}),\ sup_x(v,\overline{x}),\\ sup_y(v,\overline{y}) \quad (u \in V). \tag{4}$$

For the other 8 single tile relations, we add rules similar to (3) and (4).

Figure 2 shows an example on how the above two set of rules ensure CDC constraints in the network. A pair of objects a, b on the grid satisfies a CDC constraint $u\ N{:}NE\ v$. (C1) imposes that a occupies tiles $NE(b)$ and $E(b)$ of b, (C2) imposes that a does not occupy any other tile of b.

Disjunctive CDC Constraints A disjunctive CDC constraint $u\ \{\delta_1, ..., \delta_o\}\ v$ in C is represented in ASP by the facts

$$disjrel(u,v,i,r) \leftarrow \quad (r \in \delta_i,\ 1 \leq i \leq o). \tag{5}$$

Recall that a pair (a,b) of spatial objects satisfies $u\ \delta\ v$ where $\delta = \{\delta_1, ..., \delta_o\}$, if $a\ \delta_i\ b$ holds for exactly one $\delta_i \in \delta$. Therefore, for every disjunctive CDC constraint $u\ \delta\ v$, we nondeterministically choose $\delta_i \in \delta$, and represent the basic CDC constraint $u\ \delta_i\ v$:

$$1\{chosen(u,v,i) : 1 \leq i \leq o\}1 \leftarrow \tag{6}$$

$$rel(u,v,R) \leftarrow chosen(u,v,i),\ disjrel(u,v,i,R). \tag{7}$$

The rule (6) nondeterministically selects one disjunct $\delta_i \in \delta$, $1 \leq i \leq o$ and $chosen(u,v,i)$ atom indicates its index. (7) generates $rel(u,v,R)$ atoms corresponding to the selected disjunct δ_i.

With the ASP program described briefly above, we can check the consistency of a given CDC constraint network. We call this ASP-based formal framework for reasoning about CDC constraints, as nCDC-ASP.

Inferring Unknown Cardinal Directions When the given CDC network is incomplete, it may be useful to infer the cardinal directions between two spatial objects whose CDC relation is unknown. For a pair of objects u, v where there does not exist a CDC constraint $u\ \delta\ v$ in C, first a basic CDC relation is generated for them:

$$1\{inferrel(u,v,R) : R \in Q\} \leftarrow not\ existrel(u,v). \tag{8}$$

Then, for the inferred CDC relation between (u,v) we add ASP rules similar to (3), (4) in order to ensure conditions (C1) and (C2). Atoms of the form $inferrel(u,v,R)$ in the answer set reveal the unknown cardinal directional relation between objects u, v. Aside from digital forensics, inferring directional relations has applications in other domains such as the meeting scenario or the missing child scenario which are demonstrated in (Izmirlioglu and Erdem 2018).

Default CDC Constraints In various applications, due to dynamic domains with human presence, qualitative spatial relations may have exceptions. For example, the chair is by default in front of the table or the charger is normally attached to the gun. Then it will be desirable to express such commonsense knowledge formally, similar to CDC constraints, to allow nonmonotonic reasoning.

Motivated by such examples, we introduce default CDC constraints as expressions of the form:

$$default\ u\ \delta\ v \tag{9}$$

where $u\ \delta\ v$ is a CDC constraint. We represent this default CDC constraint by a set of facts:

$$defaultrel(u,v,r) \leftarrow \quad (r \in \delta). \tag{10}$$

We define semantics of default CDC constraints in terms of ASP rules. This is possible thanks to the nonmonotonic construct *not* and the aggregates supported by ASP. The default CDC constraint $default\ u\ \delta\ v$ applies if there is no evidence against it:

$$drel(u,v) \leftarrow not\ \neg drel(u,v), defaultrel(u,v,r) \quad (r \in \delta). \tag{11}$$

The evidence against a default constraint $default\ u\ \delta\ v$ can be due to a violation of a CDC constraint. Such a violation can come from an existing CDC constraint between the same (u,v) pair in the network or an inferred CDC constraint between (u,v). If the existing CDC constraint or the inferred CDC relation between (u,v) is different from δ, this would constitute an evidence against the default constraint.

$$\neg drel(u,v) \leftarrow not\ inferrel(u,v,r),\ defaultrel(u,v,r),\\ existinferrel(u,v)\\ \neg drel(u,v) \leftarrow inferrel(u,v,r),\ not\ defaultrel(u,v,r),\\ existDefRel(u,v). \tag{12}$$

The following weak constraint minimizes the evidences against the default constraints to satisfy as many default CDC constraints as possible.

$$\overset{\sim}{\leftarrow} \neg \mathit{drel}(u, v),\ \mathit{existDefRel}(u, v) \quad [1@1, u, v]. \quad (13)$$

Commonsense knowledge about spatial relations often induces vague assumptions which might be incorrect. Hence we represent the default assumption by a weak constraint in (13). However, depending on the nature of the domain and level of incomplete knowledge, some commonsense knowledge constitutes stronger assumption between objects. Consider for example "the charger is by default at the gun" in a criminal investigation. In this case, we impose the corresponding default CDC constraint as a hard constraint as in (Izmirlioglu and Erdem 2018).

$$\leftarrow \neg \mathit{drel}(u, v), \mathit{existDefRel}(u, v). \quad (14)$$

3 An Application of nCDC-ASP to Digital Forensics

The application presented in this section is motivated by the challenges of evidence-based digital forensics (Costantini, Gasperis, and Olivieri 2019), that goes beyond data analysis.

As a case study, we examine the following fictional crime story. A murder has occured at the living room of the Sytles mansion and there are three suspects of the crime. Suppose the true state of the world just after the murder is as in Figure 3 and not fully known by the police officers. When the detective Hercule Poirot arrived at the event venue, the body of the victim had already been transported to the hospital and objects in the room might have been relocated or taken out.

The camera image yields limited information about the event and the situation of the living room at the moment of the crime. Suppose the data from camera image reveal the following:

Data:

Gun is on the Teapoy.
Table is to the right-back of the Teapoy.
Table occupies left side of the Chair.
Body is lying along the front side of the Chair.
Body is to the right-front of the Table.
Cell Phone is to the left or left-back of Hanger.
TV is to the right-front of the Teapoy.
Shells are in front of the Table.
Shells are to the left of the Body.

nCDC Constraint:

Gun O *Teapoy*
Table NE *Teapoy*
Table $NW{:}W{:}SW$ *Chair*
Body $SW{:}S{:}SE$ Chair
Body SE *Table*
CellPhone $\{W, NW\}$ *Hanger*
TV SE *Teapoy*
Shell S *Table*
Shell W *Body*

Notice that there is some uncertainty regarding the position of the cell phone. In addition to above, detective Poirot has the commonsense knowledge about the objects that are

Commonsense Knowledge:

Charger is normally attached to the Gun.
Chair is normally in front of the Table.
Hat is normally at the Hanger.
Umbrella is normally at the Hanger.

nCDC Constraint:

default Charger O *Gun*
default Chair S *Table*
default Hat O *Hanger*
default Umbrella O *Hanger*

Such commonsense knowledge can be encoded using default CDC constraints. Then detective Poirot interviews with the suspects to obtain new clues and also determine whether the suspects are telling the truth or not. Suppose that in their interrogations, the suspects are describing the configuration of the room and position of the objects.
During his interrogation, suspect 1 tells:

> "... probably someone else had visited the victim before because when I entered the room I saw a hat on the floor, it was in front of the table and left of the charger. He was dead and lying along the front of the chair. There were some shells near and right of the teapoy."

In interview with suspect 2, he mentions:

> "...The crime had already happened when I came to the living room. He was dead on the floor. There was a suitcase standing near and in front of the drawer, I saw an umbrella on the hanger, pills to the left and close to the TV and a syringe on top of the drawer."

Lastly, statement of suspect 3 is:

> "...The murder might have been committed by the knife or the gun. The knife was to the back of the body and in front of the chair. His head and body was in blood. The room was untidy, the chair was leaning to the right side of the table. I saw some empty shells, I don't remember exactly but they were either to the right or rear of the teapoy. There was a cell phone on the floor. It was to the left and near to the body and probably belongs to him..."

From the interrogations we obtain the following CDC constraints.

Suspect 1 claims:
Hat S *Table*
Hat W *Charger*
Body $SW{:}S{:}SE$ *Chair*
Shell E *Teapoy*

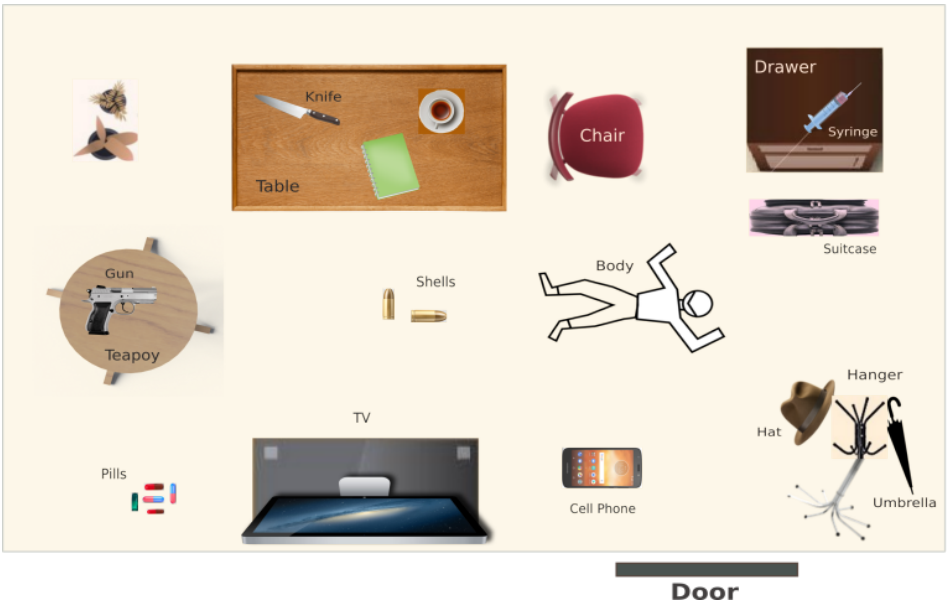

Figure 3: Murder Scene

Suspect 2 claims:
Suitcase S *Drawer*
Umbrella O *Hanger*
Pills W *TV*
Syringe O *Drawer*

Suspect 3 claims:
Knife N *Body*
Knife S *Chair*
Chair E *Table*
Shell $\{E, N\}$ *Teapoy*
CellPhone W *Body*

Given the nCDC constraints C_d obtained from the camera image, and the nCDC constraints $C_{s,2}$ obtained from interrogation of Suspect 2, nCDC-ASP finds out that $C_d \cup C_{s,2}$ is consistent. On the other hand, the nCDC constraints $C_{s,1}$ obtained from interrogation of Suspect 1, and C_d are found inconsistent; this suggests that Suspect 1 is untruthful. Indeed, the reason of inconsistency is that the charger is by default attached the gun and there is no evidence against this strong assumption. Thus if the hat is in front of the table, it cannot be to the left of the charger.

As for suspect 3, the nCDC constraints $C_{s,3}$ in his statement together with the data C_d is consistent. Notice however that the information about position of knife is missing in the camera image. Suppose that detective Poirot later investigates the infrared camera at the rear of the room to search for further evidence. The image of the infrared camera reveals the following:

Knife O *Table*
Cup O *Table*
Notebook O *Table*
Suitcase S *Drawer*.

With the additional data $\hat{C}_d$ above, the detective checks statement of suspect 2 and suspect 3 again. The additional data $\hat{C}_d$ can be augmented to the existing network to test consistency. nCDC-ASP finds that $C_d \cup \hat{C}_d \cup C_{s,2}$ is still consistent whereas $C_d \cup \hat{C}_d \cup C_{s,3}$ is now inconsistent. According to the data the knife is on the table, yet suspect 3 claims it to be between the body and the chair. That is, the new data helps the detective to identify suspect 3 being untruthful. This manner honesty of suspects can be identified.

4 Discussion and Conclusion

We have illustrated how our framework nCDC-ASP for consistency checking can be used for digital forensics, e.g., to resolve honesty of suspects of a criminal event. Our framework can deal with the challenges of incomplete information and uncertainty. Namely this method can be utilized for inspecting spatial aspect of a criminal event and its suspects.

Acknowledgments

This work is partially supported by Cost Action CA17124.

References

Brainz. 2020. 15 criminal cases solved with digital evidence. https://www.brainz.org/15-criminal-cases-solved-digital-evidence/, Last accessed on 2020-06-22.

Costantini, S.; Gasperis, G. D.; and Olivieri, R. 2019. Digital forensics and investigations meet artificial intelligence. *Ann. Math. Artif. Intell.* 86(1-3):193–229.

Goyal, R., and Egenhofer, M. J. 1997. The direction-relation matrix: A representation for directions relations between extended spatial objects. *The annual assembly and the summer*

retreat of University Consortium for Geographic Information Systems Science 3:95–102.

Izmirlioglu, Y., and Erdem, E. 2018. Qualitative reasoning about cardinal directions using answer set programming. In *Proc. of AAAI*.

Lifschitz, V. 2002. Answer set programming and plan generation. *Artificial Intelligence* 138:39–54.

Marek, V., and Truszczyński, M. 1999. Stable models and an alternative logic programming paradigm. In *The Logic Programming Paradigm: a 25-Year Perspective*. Springer Verlag. 375–398.

Niemelä, I. 1999. Logic programs with stable model semantics as a constraint programming paradigm. *Annals of Mathematics and Artificial Intelligence* 25:241–273.

Skiadopoulos, S., and Koubarakis, M. 2004. Composing cardinal direction relations. *Artificial Intelligence* 152(2):143–171.

Skiadopoulos, S., and Koubarakis, M. 2005. On the consistency of cardinal direction constraints. *Artificial Intelligence* 163(1):91–135.

Legal Issues in AI forensics: understanding the importance of humanware

Raffaella Brighi[1] , **Michele Ferrazzano**[2] , **Leonardo Summa**[3]
[1] Associate Professor of Digital Forensics, Alma Human AI - Department of Legal Studies, University of Bologna
[2] Adjunct professor of Information technology and Research fellow - Department of Legal Studies, University of Modena and Reggio Emilia
[3] Alma Human AI - Department of Legal Studies, University of Bologna
raffaella.brighi@unibo.it, michele.ferrazzano@unimore.it, leonardo.summa@studio.unibo.it

Abstract

Nowadays, our most cogent need is to embrace a new vision of the digital forensics field, which requires to be focused on: (a) the harmonization of the legal framework and technical standards; (b) the pursuit of common paths when conducting forensic investigations; and (c) the definition of an epistemological frame of reference. These three elements should be intended as the cornerstone of this change. The growing influence that ICT technology is having on the work of judges and legal professionals now requires a stronger holistic basis—concerning principles, practices, and procedures—of what is available, namely, *humanware*, and what is useful, namely, AI, to achieve and disseminate best practices. Firstly, the full potential of AI calls for a deep insight into its technical implications and into the requirements needed to keep operating in a forensic-based environment, but it also calls for deep understanding by policymakers, who may lack a sense for the ethical and legal implications of AI, while pushing for its deregulation. Therefore, understanding the urgency to act for the development of a strong and well-trained *humanware* is just the baseline in tackling well-known problems in the application of AI technologies (e.g. the reliability and explainability of machine learning methods) in the digital forensics field, as well as in the whole of society.

1 Introduction

In recent times, a debate has been ignited in the juridical world endeavouring to regulate the deployment and the possible applications of artificial intelligence (AI). Having legal decisions supported by AI is an appealing idea that dates back several years (Sartor 1992; Sartor 1998).

Numerous expert systems have been developed in the past, with the aim of autonomously reaching decisions by exploiting the representation of specialized legal knowledge in symbolic form, with logical rules and predefined inferences: the outcomes, however, were less promising than expected. Nevertheless, AI has evolved, owing to highly effective machine learning methods that deploy the knowledge deriving from big data analysis (Russell and Norvig 2009). Consequently, these recent developments have questioned both the introduction of AI technologies in different legal systems and its ethical-legal sustainability be questioned (Floridi et al. 2018).

An evidence of the great potential of these tools can be found in the widespread application of intelligent agents in support of daily and repetitive actions. At the same time, it indicates that the legal consequences of an unregulated use should be taken into account and prevented (Lasagni and Contissa 2020).

Examples include the potential probative interest profiles guarded by intelligent devices, a subject studied by IoT forensics, predictive capabilities and the fallacious discriminatory bias, the effectiveness and usefulness of the results obtained in terms of reliability and, finally, the remedies we should choose in overcoming the limits that have become apparent (Sommaggio and Marchiori 2020).

There is now widespread news, as well as numerous studies, concerning robot-judges (Millar and Kerr 2013), AI systems in a position to predict the potential criminal activities (i.e. so-called predictive policing), or even algorithms assessing an individual's social dangerousness, such as the COMPAS system implemented in U.S. courts to quantify the risk of recidivism, within the frame of predictive justice (Degeling and Berendt 2018). However, although these applications are already in the testing phase, the full potential of AI might be underestimated.

The use of AI in the collection and forensic analysis of digital evidence could be the real breakthrough that can help the justice system to streamline procedures, primarily by shortening the timeframe of investigations. It is evident that digital forensics (DF) faces mounting challenges in terms of accuracy and timeliness in the analysis of a growing amount of data from increasingly diverse sources (Council and others 2009).

Thus, a question arises as to what application of AI may effectively optimize investigation time and ensure the reliability of the results of digital evidence analysis. The aim of the present paper is to answer this question by investigating the sustainable and desirable points of contact between AI applications and the substantive and procedural rules to be observed during investigation activities, though they might differ from the traditional forms.

The keys to a productive dialogue lie in the human factor, in forensic IT experts acquiring sufficient knowledge of these tools, and in legal practitioners becoming sensitized to forensic IT issues (Brighi and Maioli 2016). If AI applications in digital forensics are to be properly regulated, their operating mechanisms need to be fully comprehended, and the boundaries between legally acceptable and unacceptable

consequences must be set, rather than enthusiastically embracing uptake at all costs and shifting the burden of damages to end users, both in the legal area and in our daily interaction with these technologies (Abdelnasser Gamal 2020).
The future of AI is clear now. The challenge is to have these instruments formally accepted in court proceedings by grounding their use in fundamental rights and fair trial principles. This work aims to endorse the role of the human factor in the sedimentation of today's digital transformation by highlighting the friction generated with the legal categories of reference and fostering the development of skills and tools by which to manage such promising technologies. The *raison d'être* of this work is indeed the human-based vision of the coexistence of our modern society with new technologies, rooted in the neutrality of the latter and the fertile *Weltanschauung* that has allowed the development of such revolutionary tools.
We aim to identify the legal issues arising in connection with the adoption of AI by DF and to suggest possible solutions. Section 2 provides an overview of state-of-the-art AI applications in DF investigations and highlights the constraints and benefits that can be derived from their implementation. Section 3 analyses the legal consequences, in terms of compression of the right of defence, violation of the legal principles protecting the fundamental rights of individuals, quantification of the acceptable margin of error regarding findings of guilt, solidity in terms of the logicality and coherence, and verifiability of results capable of satisfying the obligation to justify judicial measures adopted in cooperation, in whole or in part, with AI-based instruments.
Finally, Section 4 illustrates some future guidelines to be followed for the construction of a desirable synergy between techniques and law, between humanware and progress in the field of ICT.

2 The importance of AI into the Digital Forensics field

The last decade has witnessed the conversion of most data, such as books, videos, pictures, and medical information, into digital formats. Laptops, tablets, smartphones, and wearable devices are the major enablers of this digital data transformation and have become a substantial part of our daily lives.
As a result, we are becoming a soft target for many forms of cybercrimes. Digital forensic investigation seeks to recover lost or deliberately deleted or hidden files from a suspect's device. However, due to underdeveloped skills and lack of time, current human capabilities and government resources are insufficient for cybercrime investigations.
Existing digital investigation procedures and practices require time-consuming human interactions, thus slowing down the entire process. Many research projects, studies, and even some professional products have begun to offer solutions based on artificial intelligence to overcome known obstacles.
However, a focus on what AI is would take us away from the purview of this work. Different approaches have been tried in the history of AI which have variously paid attention to the mental models and human reasoning or to human behaviour, in attempt to develop systems that simulate human tasks execution and to build either ideally intelligent systems or systems that employ rational behaviours in order to act properly. For the purposes of our paper, AI can be considered as an instrument capable of conducting and facilitating human tasks.
AI technology is growing day by day, and its widespread use increases the number of malicious activities, with some relevant issues arising about their legal attribution (King et al. 2020). Artificial intelligence programs are called intelligent agents, and they are used to interact with the environment. The agent uses different techniques to identify the environments through its sensors, and then it can take the action needed to achieve the desired state through its sensors. The important aspects in AI technologies are how the sensors are used to collect data and how they map them onto the actuators; this is how the functions within agents can achieve these results.
A rational agent does not limit itself to gathering information but must be able to learn as much as possible by accumulating experience. Machine learning (ML) is a specific part of artificial intelligence that enables computers to learn without being explicitly programmed. For example, a machine learning system is able to find patterns in data and use them to predict the outcome of something it has never seen before. AI technologies afford significant advantages and have a bright future ahead. However, these technologies are also unavoidably used to carry out some serious crimes that can be dangerous for people (King et al. 2020; Ferrazzano 2019). Below is an overview of AI applications in DF investigations, highlighting constraints and benefits.

2.1 ML/AI & Incident Response

Until recently, cyberattacks were dealt with by relying on basic antivirus software or firewall with a list of rules. However, current cyberattacks are sophisticated enough to bypass traditional security measures. This is owed to limited human expertise and efficiency, which in turn can be attributed to several causes: the time required to detect and investigate daily threats, lack of skills, lack of accuracy, failure to detect advanced threats such as advanced persistent threats (APTs), ransomware, or fileless attacks (Ghafir et al. 2018).
AI can efficiently handle cybersecurity threats by rapidly detecting and analysing millions of logs and anomalous events, identifying a malicious file, or recognizing an atypical behaviour from a seemingly harmless data cluster or file. Security strategists can provide current advanced machine learning models with a massive quantity of historical training data, achieving increasingly better security responses when more valuable data are provided.
A practical example that displays who and what could benefit from machine learning is represented by the Security Operations Centers (SOCs). A SOC is a facility that hosts an information security team responsible for continuously monitoring and analysing an organization's security posture: the goal is to detect, analyse, and respond to cybersecurity incidents by using a combination of technology solutions and a strong set of processes. Given the number of sources of rel-

evant data alone, the impracticality of manually reviewing log files is apparent.
This challenging obstacle is traditionally overcome by relying on a system that correlates inputs by dozens of different security products, each monitoring a specific attack vector, so as to notify the SOC about the occurrence of an unusual event.
Since the SOC writes these correlation rules after the occurrence of an incident – in order to be notified of its reoccurrence – there are two main downsides. Firstly, several important events are missed because correlation rules rely on a specific set of inputs. If excessively narrow rules are defined by the SOC, the system will not be triggered by minimally different events. Considering the intra-organization variability in applications, systems, and environments, it is unlikely that two attacks will be identical. Secondly, false positive results can be generated if the rules are not narrow enough: this poses the risk of masking real attacks by generating countless alerts that cannot be readily filtered by the SOC to identify real threats.
Either way, analysts miss attacks in the deluge of data, or they identify them too late. In order to find important security events without generating low value alerts that demand time, attention, and manual remedy, the SOC may leverage AI and ML.
Let us recall that AI is a broad term that refers to algorithms, models, and a field of scientific study. ML is the concept of training a system to perform narrowly focused tasks without using explicit instructions, relying on pattern detection and conclusion inference. It focuses on a specific need.
AI and ML can identify important security events in an organization, with high accuracy, by gathering together data from multiple sources while optimizing the time and experience required in the SOC. To date, many security companies have developed products that work with ML algorithms to try to help companies fight cybercrime [1] [2] (Trifonov et al. 2019; Hasan et al. 2011).

2.2 ML/AI & Forensics Analysis and Evaluation

An increasingly important area in computing, digital forensics frequently requires the intelligent analysis of large amounts of complex data: most challenges currently posed by these needs may be ideally approached through AI. An important issue for AI in the forensic arena is the ability to explain the reasoning process (Krivchenkov, Misnevs, and Pavlyuk 2019).
Two subtypes of AI techniques are recognized: symbolic (techniques reasoning with discrete entities in a knowledge base) and sub-symbolic (techniques in which the knowledge is spread across the representation structure). Expert systems are a common example of symbolic AI techniques: they follow a predefined rule base, and normally rely on a regulated strategy to select which rule to use at any particular moment in time.
Therefore, expert systems can, at any point, provide an explanation of the reasoning for the conclusions obtained, thus permitting an outside entity to review the reasoning process and to recognise any flaws in the reasoning itself (Mitchell 2014).
However, two major drawbacks of symbolic systems can be identified. The first of these drawbacks is that they operate in a closed world: any item that is not part of the rule base cannot be Justified in the reasoning process.
This is a serious issue in a rapidly evolving area such as computing, as rebuilding a rule base *de novo* is a time-consuming task and adding additional rules (a process known as "rule base repair") can damage the original performance.
The second drawback is that expert systems perform poorly with large quantities of data. This is a major disadvantage in digital forensic investigations, where exponentially larger amounts of data need to be investigated. However, techniques such as expert systems might prove to be useful in higher-order situations, such as suggesting the following steps to an investigator, or advising on what an organisation's policy should prefer in a given situation (Costantini, De Gasperis, and Olivieri 2019).
A form of typically symbolic AI that may bypass the disadvantages of expert systems (and other symbolic rule-based systems) is that of case-based reasoners (CBRs). CBRs are built on psychological notions concerning information representation by domain experts themselves.
Most domain experts heavily rely on their past experiences: when faced with an issue, they will draw parallels between current and past situations, thus using first principles to find a solution only when all possible similar cases in their experience have been exhausted. Similarly, a CBR system first collects a large number of cases (and, in digital forensics, the resulting actions), and then resorts to a metric to relate the current situation to one already included in the case base. If a perfect match is found, then the current situation will be managed through the same solution applied in the initial case.
Likewise, if a partially similar match is found, the system may attempt to adapt the action of the matched case to the current situation employing the so called "repair" rules. CBR systems have the advantage of approaching a problem in a way that is familiar to the expert, while coping with large amounts of data, and dealing with entirely unknown situation.
Since the reasoning can be inspected (this case was closest to X, and in X you did Y), CBR system also expose their reasoning process. Consequently, the quality of the cases and the number of different scenarios included in the case base are crucial to determine the performance of CRBs. A further limit of CRBs is that, while they can support the investigation, they might be ill-suited to lower-level activities (i.e. "find all pictures with naked people in them") (Sanchez et al. 2019).

[1] Microsoft uses its own cybersecurity platform, Windows Defender Advanced Threat Protection (ATP), for preventative protection, breach detection, automated investigation and response.

[2] Splunk software has a variety of applications, including IT operations, analytics and cybersecurity. It's designed to identify a client's current digital weak points, automate breach investigations and respond to malware attacks. Products like Splunk Enterprise Security and Splunk User Behavior Analytics use machine learning to detect threats so they can be quickly eliminated

Identifying specific types or clusters of data in an investigation is best handled by a type of AI known as "pattern recognition". The type of pattern recognition that people are most familiar with is perhaps image recognition, where software attempts to identify parts of a picture.
Furthermore, there are many other examples of pattern and image recognition, such as detecting a pattern in a SPAM e-mail, or a pattern in a disk image that might indicate it is part of a sound file. Many of the techniques used rely very heavily on statistics or probabilistic reasoning, or both.
The most complex and accurate forms of image recognition that can be used to locate certain types of picture, rely on the awareness of how human perceptual system works. However, at these tools currently have a high rate of false positives and false negatives (depending on where the thresholds are set), besides being very computationally intensive.

3 Legal and Ethical issues

The relationship between technology and the law recalls the second of Zeno's four paradoxes of movement, that of Achilles and the tortoise. According to this paradox Achilles, representing the law, races against but will never be able to overtake the tortoise, representing technology.
In this endless chase, the law has often tried to model the existing concepts whenever the relevant transformations produced by computer osmosis in legal relations have generated distortions that are no longer tolerable for the legal system itself. Consequently, reinforced protection at European level has become necessary to regulate the processing of personal data. Similarly, we argue that it is necessary to develop a regulatory framework for the investigative uses of technology that guarantees respect for procedural principles and the fundamental rights of individuals.
For this to materialise, it is necessary to become involved in the constant development and updating of computer skills useful for the construction of investigative models that comply with fundamental rights. This is what we call *humanware*, referring to the human factor that intervenes in digital investigations as well as in the relationship with technology.
Focusing on the growth of a more conscious *humanware* by encouraging certified training course for DF examiners, lawyers and judges will limit the potential pathogenic causes – such as discrimination and bias, margins of error, false positives, false negatives – of unlawful decisions based on AI system. Thus, it will be possible to achieve greater respect for fundamental rights, regarding the application of AI-based systems.
In this section, we will examine the repercussions in terms of the substantive and procedural rights generated by the application of AI tools in the formation of digital evidence, with particular attention to the principles that distinguish civil law with an adversarial legal system.

3.1 *Male captum bene retentum*

The legal issue around the usability of illegally acquired evidence is of extreme relevance and known in every legal system. The legal dispute involves a very important question: can testimony constitute fully usable evidence when obtained by illegal means, such as torture?
In this extreme context, two opposing factions can be distinguished: those who claim that such results are also illegitimate—*the fruit of the poisonous tree doctrine*—and those who, on the contrary, save the evidentiary results in the light of the Latin principle of *Male captum bene retentum*.
The rationale behind this latter principle is to safeguard the results of investigations, even if they are achieved through the violation of those procedural rules that protect the fundamental rights of persons subject to judicial ruling.
This theory expresses the problematic synthesis of two opposing requirements that are difficult to reconcile: on the one hand, the need to ensure sources of evidence even by using instruments not typified by procedural rules and, on the other hand, the need to safeguard the guarantees put in place to protect against abuses and violations of internationally recognized fundamental rights. The legal ethical sustainability of AI applications in the DF field cannot prescind from the analysis of this contradiction (Losavio et al. 2019; Abdelnasser Gamal 2020).
Accordingly, it is essential to be aware of the legal effects of the use of such technologies, which cannot accept silent adaptations and advocate the greatest possible sharing in the definition of the criteria, limits, and benefits deriving from the introduction of such technologies into the legal arena. Such a phase transition, with the legal implications of these instruments being carefully assessed, is paramount, lest the function of social defence of the law be transmuted into a contractual relationship supported by the mere criteria of efficiency and usefulness unrelated to its social function (Sanger 2018).
In other words, without such a phase transition, the procedural position of each of us would become as a stock exchange listing, fuelled by the logic of reducing the workload of the courts and ensuring greater efficiency, in comparison with human judgment. And it is precisely in contrast to such a logic that we will have to construct proceedings-sustainable variations of the different AI applications available in the field of digital evidence.
Technological transformation must be reconciled with respect for the fundamental rights of the individual, around which the boundaries of law are drawn: the right to a fair trial, which incorporates the right to an impartial judge; the presumption of innocence until otherwise proven, and the duty of judicial authorities to give reasons for their ruling (Vuille, Lupària, and Taroni 2017).
The question appears Hamletic: how can the need to make judicial processes efficient coexist with thee respect for procedural rules and individual fundamental rights?
The answer is to be found in a more mature symbiosis than the one we are currently experiencing, guided by people's awareness of the instruments, both in sustaining their usefulness and in paying attention to its pathological evolutions.
Public debate should be encouraged to become aware of the legal conscience, which is now weak, in order to raise and stimulate active participation in the formation of judicial practices, while respecting the fundamental rights rec-

ognized at the international level (Quattrocolo et al. 2020).
The first step is to realize the biunivocal character that marks the relationship between *technè* and law, by arising a section in the criminal and civil procedure code dedicated to computer investigations and digital evidence acquisition processes. Specific guidelines and procedures must be provided to ensure compliance with the technical principles of digital forensics and fundamental human rights.

3.2 Beyond a reasonable doubt

When assessing the sustainability of the use of AI-based systems in the DF field, another consideration might arise: the introduction of AI-based technologies into evidence generation is strongly conditioned by the degree of reliability achievable in the design of such systems.
The provocative tone of the question offers an opportunity to reflect on the function of these technologies in legal systems. When using AI-based techniques (ANNs, K-means, NLP, etc.), the result that is obtained is reliable by the measure of the margin of error known for that particular system. The acceptable range of error for a given legal system is to be defined in the same way as the degree of tolerance within which human error is justified (Kotsoglou 2019).
The matter of transparency and justifiability of the choices and results produced is a well-known technical problem and cannot be underestimated when applying AI to legal reasoning. Eliminating the risk attendant on the factors of human error (i.e. prejudices, likes/dislikes, personal beliefs, emotional distress) and their consequent influence on the decision-making process is an appealing concept. However, we eventually accept decisions that are unquestionable because the original mechanism producing the result is unexplored (Grace 2019).
For instance, a crucial aspect of paedopornographic crimes is age determination of the victims. The automation of the processes of identification and attribution of the underage factor would be of extraordinary value (Anda, Le-Khac, and Scanlon 2020).
Nevertheless, attention should be paid to some basic considerations:

- Dataset training: checking the input that data used to train neural networks is fundamental. The chosen model is initially built around a training dataset which is a set of examples used to set parameters for the model (e.g., skin tone, height, etc.). To evaluate whether a model is being trained correctly, it is necessary to take note of the loss: the smaller the loss, the better a model. The loss is calculated on the basis of training and validation and can be interpreted by how well the model is doing for these two sets;
- Accuracy problems: neural networks are ML algorithms that provide the state of the accuracy on many use cases. Frequently, the accuracy of the network we are building is not be satisfactory: 99% accuracy is not equal to 99% success. Legally, a 1% failure rate means not having, beyond any reasonable doubt, the certainty that the output is actually what was expected. When evaluating an ML model, it is useful to establish the so-called high bias and high variance. High bias refers to a scenario where your model "underfits" the example dataset: the model is assumed not to present a precise or representative picture of the relationship between the inputs and the predicted output. Contrarywise, high variance refers to a scenario in which the model "overfits" the dataset: it is so accurate that it is perfectly fitted to your example dataset.While seemingly a good outcome, it is a concerning one, as such models often fail to generalize to future datasets. These models might work properly for prefixed existing data, but not for general uses
- Debug problems: for a result to be demonstratable and reproducible, it is necessary to probe all steps leading to a certain result. Technically, it is difficult to accomplish a similar degree of transparency. Such criticality finds a double explanation: firstly, proceeding with real-time debugging, capable of witnessing step by step the choices made, is virtually impossible; secondly, due to the unpredictability of machine learning algorithms applied in the development of neural networks, it is not always possible to predict the variations suffered by the original mathematical model in the face of new and unknown scenarios.

The margin-of-error question becomes a matter of constitutionality, as the decision-making process must provide comprehensive and coherent reasoning from a legal and logical point of view. The need to reconstruct the logical path in a way that justifies and accounts for the results put out by the instrument clashes with the technical difficulties encountered in the process (Horsman 2019).
Justifying the results obtained requires that these instruments be used in keeping with the need to undergo authoritative measures that can be judged on the merits of their assumptions. This obstacle suggests that the use of these technologies should be limited to an auxiliary support function, of circumstantial rank, which requires the results obtained through their falsification to be evaluated at a time prior to the evaluation.
As to satisfy the gap in terms of the reliability and transparency of AI-based systems, it is essential to recognize the key role played by having a deeper and more sensitive approach to the legal reflections on the usage of digital technologies. In order to achieve this target, we strongly endorse the creation of supervised systems, those who still address interpretability to its own choices; and protecting the rights of all the parties involved in the trial, by opening up to their participation in the execution of technical operations; forging a set of certified IT skills and opening the road to the so called *humanware* in Digital Forensics field. If we do not act upon the paths of a human-centered perspective, we will not be able to take advantage from the application of AI-based systems.

3.3 *Nemo tenetur se detergere*

The amount of information passing every second through digital networks and devices is the preferred source of evidence in criminal proceedings: the techniques available in the field of DF for the detection of crimes and the resolution of legal cases are used on a daily bases DF

experts use a variety of technologies for the detection of crimes and the resolution of legal cases (Opijnen and Santos 2017). This obliges us to reflect, with greater consideration, on the relationship between principles and procedural rules and the new technological frontiers.

The critical profiles are highlighted above all with reference to the violation of the right to confidentiality of correspondence and privacy. The most extreme consequences of this schism develop in procedural systems based on the recognition of the right against self-incrimination, which deserve to be properly regulated.

The pervasiveness of digital investigation, due to the growth of the storage capacities, the distribution of digital services in performing daily activities over which we generate a huge amount of valuable information, and the advent of a new online reality, are now facts shared in the ordinary experience. Investigative techniques are constantly evolving and have had to undergo the transformation dictated by the entry of the digital dimension, that became a new space inside which it is possible to commit and prosecute old and new crimes. Techniques in digital investigations need to continually fit the growth and spread of computer skills in crime commission.

For this reason, they require a regulation that encourages the unfolding of skills that can safeguard the conduct of investigations in the digital field in respect of the right not to self-incriminate. It draws a distinction between the possible investigative scenarios, by setting a minimum level of warranty, such as the faculty to attend to the technical operations or a video recording that repeatable. Even creating an *ad hoc* stage in the trial to guarantee the right of a fair trial by the opening of technical schemes, such as keyword searches, is a good point to envisage a better way for the employment of those rights.

For this reason, we argue that technical and regulatory frameworks should be developed to guarantee internationally recognised fundamental rights, when they are not already established by national legislation (Saleem, Baggili, and Popov 2014). In the current scenario, increasing attention is being paid to respect for procedural guarantees in the processing of digital evidence, not only with regard to the technical requirements of admissibility but also to the limits of usability of the acquired information (Nieto et al. 2019).

On the one hand, studies aimed at raising the thresholds for the protection of the rights at stake are growing; on the other, there is a widespread reluctance to reconsider the centrality of the means of proof offered by DF techniques in ascertaining legally relevant facts (Sunde and Dror 2019; Henseler and [van Loenhout] 2018).

There are numerous attempts to save the regulatory scope of traditional institutes by adapting technological innovations to pre-existing legal concepts, rather than studying their functioning and understanding which legal rationale would be more appropriate for them. Despite the delays accumulated by legislation, there are encouraging signs of development of privacy-preserving architectures in the context of digital investigations: only the artefacts relevant to the crime being prosecuted would be exposed, while excluding any other personal information or information related to other crimes, of which one may become aware by analysing all the stored content (Opijnen and Santos 2017; Verma et al. 2019).

For these reasons, we believe that the defence of fundamental rights cannot find a justifiable compression in the availability of invasive and unregulated means.

4 Prospective proposals

Due to the incremental collection and sharing of Electronically Stored Information (ESI) from different sources, such as the increase and fragmentation of storage capability, the computer specialist's daily workload is evidently increased: it often requires a reactive response in a large data-set, in order to prosecute the crime and preserve the evidence.

AI/ML techniques are well suited to automate traditional tasks, possibly optimizing the time consumption and quality of the forensic process. Examples include classification of relevant evidence, detection of suspicious artefacts, recognition of suspects' faces, age calculation in child sexual exploitation material (Anda, Le-Khac, and Scanlon 2020), in addition to the creation of a framework of intelligent agents to parallelize tasks and ensure particular reliability for each of them, thanks to, for instance, privacy-preserving architecture that enables the access only case-relevant information (Verma et al. 2019).

In this context, we believe that the application of AI in DF is an appealing solution to the current and future challenges of DF, by both overcoming the limits of time shortage and ensuring reliability and admissibility of the digital evidence processed by AI forensics tools.

We also firmly believe that the human factor cannot be replaced by a machine, which is why growing a well-established *humanware* is fundamental to tackling the legal issues relating to the limits of AI in D(Casey 2017). Any digital investigator knows from their daily experience the importance of understanding how an analytic tool approaches evidence, in order to produce a reliable explanation and consequently collect admissible evidence.

This is only the first step in providing better compliance with a digital forensics framework related to the reliability of evidence, achieving reproducible results, and balancing fundamental rights with the trial's needs. The best way to tackle the previously uncovered legal issues is to cast AI in a supporting role in DF tasks.

In spite of that, how could be possibly brought out such a model? Beginning by structuring an architecture dedicated to the running of digital investigations, accessible on every prosecutors' departments. Trough the creation of a dedicated law enforcement agencies, in close interaction with the academic researchers, formed up with qualified training courses to tackle the endless evolving of DF techniques, we could probably be capable to face out the grade of ethical and legal issues caused by the introduction of AI systems into decision-making processes.

In our daily scenario we are searching, almost without any other alternative source, a digital proof even related to ancient crimes in order to find relevant artifacts that prove that

prosecuted crime. Due to this reason, we have a lack of updating regulation and building a fundamental component of a system based on the principles of a fair trial, a *humanware* fact maybe the turning point of this intricate challenge which is balancing fundamental right with the range of Digital Forensics tools based on AI potential.
For these reasons, we believe that the only sustainable solution is fighting for is to face all the ethical problems relating to AI by following a human-centred vision. In this path forward we have to raise a strong background for achieving a truly trustworthy AI ecosystem, also with the help of the EU ethics guidelines for trustworthy AI, which are focused on the development of AI-based tools that allow compliance with all laws and regulations and with ethical principles, and offering a more robust and reliable solution from both a technical and a social perspective.
This will therefore make it possible to develop technical equipment aimed at guaranteeing all of the fundamental rights that may be at risk when it comes to AI (Hamon, Junklewitz, and Sanchez 2020; Commission 2019).

***

Although this article is the result of the authors joint research, the draft (paper) has been divided as it follows: R.Brighi par.1, 3, 4; M.Ferrazzano par.2, 2.1, 2.2; L. Summa par.3.1, 3.2, 3.3.

References

Abdelnasser Gamal, A. 2020. Artificial intelligence and humans. *International Journal of Scientific and Research Publications (IJSRP)* 10:p9970.

Anda, F.; Le-Khac, N.-A.; and Scanlon, M. 2020. Deepuage: Improving underage age estimation accuracy to aid csem investigation. *Forensic Science International: Digital Investigation* 32:300921.

Brighi, R., and Maioli, C. 2016. Un cambiamento di paradigma nelle scienze forensi. dall'armonizzazione tecnico-giuridica a una nuova cornice epistemologica. *Informatica e Diritto* XXIV:217–234.

Casey, E. 2017. The value of forensic preparedness and digital-identification expertise in smart society. *Digital Investigation* 22:1–2.

Commission, E. 2019. Ethics guidelines for trustworthy ai.

Costantini, S.; De Gasperis, G.; and Olivieri, R. 2019. Digital forensics and investigations meet artificial intelligence. *Annals of Mathematics and Artificial Intelligence* 86(1-3):193–229.

Council, N. R., et al. 2009. *Strengthening forensic science in the United States: a path forward*. National Academies Press.

Degeling, M., and Berendt, B. 2018. What is wrong about robocops as consultants? a technology-centric critique of predictive policing. *AI SOCIETY* 33:3:347 – 356.

Ferrazzano, M. 2019. *Autonomous driving e informatica forense: la prova della responsabilità in caso di sinistri*. Giappichelli.

Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F.; Schafer, B.; Valcke, P.; and Vayena, E. 2018. Ai4people—an ethical framework for a good ai society: Opportunities, risks, principles, and recommendations. *Minds and Machines* 28:689–707.

Ghafir, I.; Hammoudeh, M.; Prenosil, V.; Han, L.; Hegarty, R.; Rabie, K.; and Aparicio-Navarro, F. J. 2018. Detection of advanced persistent threat using machine-learning correlation analysis. *Future Generation Computer Systems* 89:349 – 359.

Grace, J. 2019. Machine learning technologies and their inherent human rights issues in criminal justice contexts.

Hamon, R.; Junklewitz, H.; and Sanchez, I. 2020. Robustness and explainability of artificial intelligence.

Hasan, R.; Raghav, A.; Mahmood, S.; and Hasan, M. 2011. Artificial intelligence based model for incident response. 3.

Henseler, H., and [van Loenhout], S. 2018. Educating judges, prosecutors and lawyers in the use of digital forensic experts. *Digital Investigation* 24:S76 – S82.

Horsman, G. 2019. Tool testing and reliability issues in the field of digital forensics. *Digital Investigation* 28:163–175.

King, T. C.; Aggarwal, N.; Taddeo, M.; and Floridi, L. 2020. Artificial intelligence crime: An interdisciplinary analysis of foreseeable threats and solutions. *Science and Engineering Ethics* 26:1:89–120.

Kotsoglou, K. N. 2019. Proof beyond a context-relevant doubt. a structural analysis of the standard of proof in criminal adjudication. *Artificial Intelligence and Law*.

Krivchenkov, A.; Misnevs, B.; and Pavlyuk, D. 2019. Intelligent methods in digital forensics: State of the art. *Lecture Notes in Networks and Systems* 274–284.

Lasagni, G., and Contissa, G. 2020. When it is (also) algorithms and ai that decide on criminal matters: In search for an effective remedy. *European journal of Crime, Criminal Law and Criminal Justice* 3.

Losavio, M.; Pastukov, P.; Polyakova, S.; Zhang, X.; Chow, K.; Koltay, A.; James, J. I.; and Ortiz, M. 2019. The juridical spheres for digital forensics and electronic evidence in the insecure electronic world. *Wiley Interdisciplinary Reviews: Forensic Science*.

Millar, J., and Kerr, I. 2013. Delegation, relinquishment and responsibility: The prospect of expert robots. *SSRN Electronic Journal*.

Mitchell, F. 2014. The use of artificial intelligence in digital forensics: An introduction. *Digital Evidence and Electronic Signature Law Review* 7.

Nieto, A.; Rios, R.; Lopez, J.; Ren, W.; Wang, L.; Choo, K.-K. R.; and Xhafa, F. 2019. *Privacy-aware digital forensics*.

Opijnen, M., and Santos, C. 2017. On the concept of relevance in legal information retrieval. *Artificial Intelligence and Law* 25:65–87.

Quattrocolo, S.; Anglano, C.; Canonico, M.; and Guazzone, M. 2020. *Technical Solutions for Legal Challenges: Equality of Arms in Criminal Proceedings*, volume 20.

Russell, S., and Norvig, P. 2009. *Artificial Intelligence: A Modern Approach*. USA: Prentice Hall Press, 3rd edition.

Saleem, S.; Baggili, I.; and Popov, O. 2014. Quantifying relevance of mobile digital evidence as they relate to case types: A survey and a guide for best practices. *The Journal of Digital Forensics, Security and Law* 9.

Sanchez, L.; Grajeda Mendez, C.; Baggili, I.; and Hall, C. 2019. A practitioner survey exploring the value of forensic tools, ai, filtering, safer presentation for investigating child sexual abuse material (csam). *Digital Investigation* 29:S124–S142.

Sanger, R. 2018. Forensics: Educating the lawyers. *SSRN Electronic Journal*.

Sartor, G. 1992. Artificial intelligence in law and legal theory. *Current Legal theory* 10:1–59.

Sartor, G. 1998. Judicial applications of artificial intelligence. *Artificial Intelligence and Law* 7:157–372.

Sommaggio, P., and Marchiori, S. 2020. Moral dilemmas in the a.i. era: A new approach.

Sunde, N., and Dror, I. 2019. Cognitive and human factors in digital forensics: Problems, challenges, and the way forward. *Digital Investigation* 29.

Trifonov, R.; Yoshinov, R.; Manolov, S.; Tsochev, G.; and Pavlova, G. 2019. Artificial intelligence methods suitable for incident handling automation. *MATEC Web of Conferences* 292:01044.

Verma, R.; Govindaraj Dr, J.; Chhabra, S.; and Gupta, G. 2019. Df 2.0: An automated, privacy preserving, and efficient digital forensic framework that leverages machine learning for evidence prediction and privacy evaluation. *Journal of Digital Forensics, Security and Law* 14(2):3.

Vuille, J.; Lupària, L.; and Taroni, F. 2017. Scientific evidence and the right to a fair trial under article 6 echr. *Law, Probability and Risk* 16(1):55–68.

Digital Forensics & real cases: from Prosecutor's request to solution

Raffaele Olivieri[1] , **Stefania Costantini**[1] , **David Billard**[2] ,
[1]University of L'Aquila
[2]HESSO University of Applied Sciences in Geneva
raffaele.olivieri@gmail.com, stefania.costantini@univaq.it, David.Billard@hesge.ch

Abstract

The Digital Forensics (DF), as any other forensic discipline, is a science that follows rigorous methodologies and procedures could be generalized in steps.

During the activities related to a police investigation, particulary during the DF analysis or Digital Investigation (DI) activities, after the phases of data collection, further elaboration of the data is needed, by the investigators, for the contextualization of the objective elements in the real investigative case. The contextualization is required to search for facts, actions, events (and their sequences), as well as testing investigation hypothesis (verifiable) to be proposed as evidence in court during a trial. Very complex investigations, which often involve an enormous amount of heterogeneous data, represent a huge problem for the human mind when is needed to search the connection between events, facts or to demonstrate the existence of alternative scenario or solution. With considerable frequency, the investigative problem description may seems outline solutions which are non-linear, or seemingly even chaotic, but after a methodic analysis of the case, and its discompose in elementary components, many cases can be represented with a mathematical approach. The shape that the problems take on are typical of known optimization problems, belonging to various classes of complexity theory among which P, NP, or not far beyond, that can be thus expressed and often solvable with reasonable efficiency by using logic programming. Therefore, the aim of this demonstration is to present the formalization of some realistic investigative cases, via the reduction the case to the known optimization problem and find solution via simple logical programs using ASP (Answer Set Programming), and thus show how this approach leads to the formulation of concrete investigative hypotheses. In this way, the European Cost Action CA17124 called DigForASP (Digital forensics: evidence analysis via intelligent systems and practices), wants to delineate the future of the investigations, or simply the data contextualization, defining an implementation of a Decision Support System for investigators, by the integration of many techniques of Artificial Intelligence, Automated Reasoning and Computational Logic, the feasible implementation of intelligent agents as an aid for the human operator (specifically as a means to aid judges, lawyers, police, criminologists, etc.), supporting her/him in the checking of concrete investigative hypotheses.

During the demonstration, it will initially be illustrated how the requests of the judiciary evolved over the last twenty years and how they have become increasingly complex. The complexity arises from the need not only to search for existing data within a digital system, but today more and more frequently to correlate existing data with reasoning to carry out investigative evaluations.

Later it will be shown how an analysis of DF and DI is born and how today it is carried out by investigators with traditional methods.

Finally it will be illustrated some investigative cases approaching the use of logic programming with ASP (*Answer Set Programming*), led to formulation concrete investigative hypothesis.

Investigative cases are usually complicated, and involve a number of factors and several data to be taken into account. A formal explanation of such conclusions cannot in general be provided. After a deep analysis of a great number of DF real cases, as well as general investigations, we have reached the conclusione that many investigative problems can be reduce to computational problems, often to known ones. With this approach the reduction is clearly the analyst's responsibility and the solutions can however be found via the execution of algorithms, whose correctness can be proved.

Efficient Argument-based Inquiry at the Dutch Police

Daphne Odekerken[1,2] , **AnneMarie Borg**[1] , **Floris Bex**[1,3]

[1]Department of Information and Computing Sciences, Utrecht University
[2]National Police Lab AI, Netherlands Police
[3]Tilburg Institute for Law, Technology and Society, Tilburg University
{D.Odekerken, A.Borg, F.J.Bex}@UU.nl

Abstract

We study the dynamic argumentation task of detecting stability: given a specific structured argumentation setting, can adding information change the acceptability status of some propositional formula? Detecting stability is not tractable for every input, but efficient computation is essential in practical applications. We present a sound approximation algorithm that recognises stability for many inputs in polynomial time. This algorithm is currently applied for fraud inquiry at the Dutch National Police - we provide an English demo version that also visualises the output of the algorithm.

1 Introduction

One task of the police is the intake of citizens' reports on crimes: the citizen tells the police what happened; subsequently, additional questions can be asked to determine if the citizen has been the victim or witness of a crime. Certain high-volume crimes can be reported online. This can be as simple as filling out a web form, but can also be a more involved online dialogue with a (possibly artificial) agent. One specific high volume crime that can be reported online at the Dutch National Police is internet trade fraud. This concerns fake web shops and malicious second-hand traders on platforms such as eBay. In (Bex, Peters, and Testerink 2016), an initial sketch was given for an artificial agent handling the intake of internet trade fraud by combining natural language processing with symbolic techniques for reasoning about crime reports. During the subsequent development of the intake agent, we regarded intake as *argument-based inquiry* (Black and Hunter 2009). In this inquiry, defeasible rules representing the laws and practices surrounding trade fraud are combined with the citizen's knowledge of the specific situation they observed, to build arguments for and against the main claim made by the citizen: that they have been the victim of trade fraud.

We present an implemented version of the intake agent, which has been released on the web site of the Dutch Police[1] where it handles the intake of hundreds of fraud reports every day. Because the police web site only shows the Dutch user interface, we provide a demo[2] of an English version that gives more insight in the underlying reasoning. The demo includes a simple chat interface and two algorithms (Testerink, Odekerken, and Bex 2019; Odekerken, Borg, and Bex 2020) for computing the underlying stability status of the conclusions.

2 Intake of fraud complaint reports at the police

The agent's architecture is illustrated in Figure 1. The *information extraction* component uses natural language processing techniques to automatically extract the initial observations from the free text user input (Schraagen and Bex 2019). These observations are then combined with rules concerning trade fraud in the argumentation system to build arguments for and against the claim "fraud". The *stability* component then decides if any additional observations that the citizen could possibly add in the future can change the acceptability status of the "fraud" claim. If not, the dialogue terminates; otherwise a *question policy* component finds the best question to ask given current observations. The stability component is thus an important part of the agent's architecture: it provides a termination criterion which prevents the agent from asking unnecessary questions. If, for example, from the initial observations it is already clear that we are not dealing with fraud because the citizen simply received a product they did not like, the agent will not continue to exhaustively inquire (Black and Hunter 2009) about further details of the situation.

3 Demonstration

We demonstrate an English version of the Dutch system used by the police. The online demo[2] has links to the full police system as well as the simpler examples from our papers (Testerink, Odekerken, and Bex 2019; Odekerken, Borg, and Bex 2020). Furthermore, we provide examples of situations where the algorithms are not complete. Note that the demo focuses on the Stability module, and thus has a simplified chat window in which the user can input their observations.

[1]https://aangifte.politie.nl/iaai-preintake/#/lmioform/AANKOOP

[2]https://nationaal-politielab.sites.uu.nl/estimating-stability-for-efficient-argument-based-inquiry/

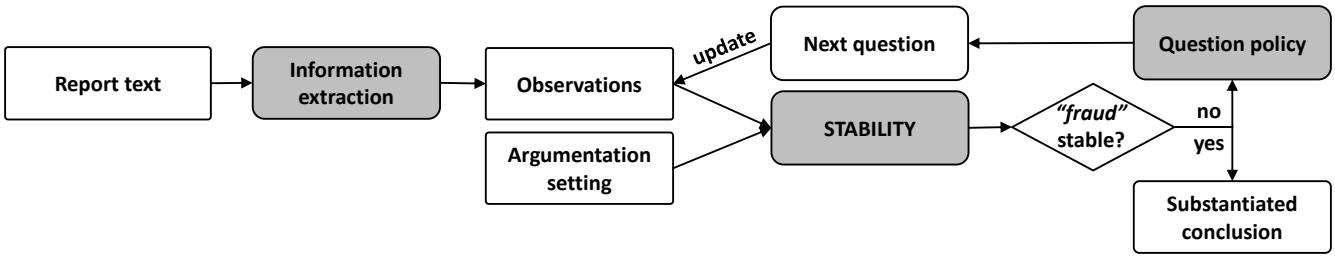

Figure 1: *Overview of the hybrid inquiry agent for the intake of fraud complaints.*

Acknowledgments

This research has been partly funded by the Dutch Ministry of Justice and the Dutch National Police.

References

Bex, F.; Peters, J.; and Testerink, B. 2016. AI for online criminal complaints: From natural dialogues to structured scenarios. In *Artificial Intelligence for Justice Workshop (ECAI 2016)*, 22–29.

Black, E., and Hunter, A. 2009. An inquiry dialogue system. *Autonomous Agents and Multiagent Systems* 19(2):173–209.

Odekerken, D.; Borg, A.; and Bex, F. 2020. Estimating stability for efficient argument-based inquiry. In *Computational Models of Argument: Proceedings of COMMA 2020*.

Schraagen, M., and Bex, F. 2019. Extraction of semantic relations in noisy user-generated law enforcement data. In *13th International Conference on Semantic Computing*, 79–86. IEEE.

Testerink, B.; Odekerken, D.; and Bex, F. 2019. A method for efficient argument-based inquiry. In *Proceedings of the 13th International Conference on Flexible Query Answering Systems*.

Author Index