EasyChair Preprint

№ 5101

Proposed Intelligence Systems Based on Forensics Case Study: Review Paper

Mohammed Hasan Ali and Mustafa Musa Jaber

March 4, 2021

Proposed Intelligence Systems Based on Forensics Case Study: Review Paper

Mohammed Hasan Ali[1], Mustafa Musa Jaber

[1] Computer Techniques Engineering Department, Faculty of Information Technology, Imam Ja'afar Al-sadiq University, Najaf, Iraq

Department of Computer Science, Dijlah University College, Baghdad, Iraq

Abstract: The field of information security has seen shifts a traditional approach to an intelligence system. Moreover, an increasing of researchers to focus on propose intelligence systems and framework based on the forensic case studies because of the limitations of traditional methods such as analysis intensive data manually, intelligence visualization to make evidence more understandable and intelligence system for store data. However, most of these intelligence systems are still facing different limitations. Furthermore, the primary goal of this work analysis popular intelligence system that was used based on forensic. Moreover, propose a new algorithm which it's achieved good results in different other fields.

Keywords: machine learning, forensic, Artificial intelligence

Introduction

In general, the systems with capable of handling a comparison with a different couple degree with intelligence will definitely give positive enhancement by an increase the quality and productivity which represent such a powerful example of this kind of the systems. They have been contributed to develop several fields and sciences such as Intrusion-Detection System (Mohammed Hasan Ali, 2018), Image Processing (Rosten & Drummond, 2006), prediction of water level (Deo & Şahin, 2016). On other hand, digital multimedia currently has become an integral part of our day activities. It has become necessary to secure this content from the illegal use, efficiently detect and reconstruct illegal activities from it. More in general, it is the set of techniques that can be applied to understand how a system has been used or abused to commit mischief. The increasing use of forensic techniques has led to the development several techniques that can make this process difficult (Maggi et al., 2008). However, (Battiato et al., 2012) tasks that were previously subjected to manual inspection are now far beyond the capacities of forensic experts, tools are needed to

support the protection, management, processing, interpretation, and visualization of multimedia data during the various steps of the investigative process.

The community of multimedia researchers have developed several exciting solutions to enhance images, videos and audio include automatic categorization, knowledge extraction and indexing. (Battiato et al., 2012) many researchers mentioned several advantages to adapt, tailor and extend multimedia analysis for forensics. Even though forensic personnel have consciously used past experiences in solving new cases, the idea of applying machine intelligence to support decision-making in forensics is still in its infancy and poses a great challenge (Koc et al., 2012). This work provides an overview of the most popular intelligence systems that have been proposed based forensic. This paper structure is organized as following. In section.2 represents the state of the rat of the main limitations of the forensic that encouraged the researchers to adapted intelligence systems for forensic cases. In section.3 the details of related works that proposed intelligences systems based on forensic. The paper conclusion with Section.4.

Intelligence System Based on Forensic

The previous decades showed that traditional forensic methods in different fields faced many limitations. Moreover, these limitations are works as main motives for researchers to adapt intelligence systems such as artificial intelligence and machine learning to improve such as accuracy, time-consuming and complexity (Aditya et al., 2018). forensic investigation is one of important field of busting cyber criminals which make the evidences that are proved correct scientifically and use them for experiments and analysis Casey, E. (2011). Furthermore, different aspects have been enhanced of forensic such as in (Fatima et al., 2017) mentioned that because of the large number of access data and insufficient attack analysis techniques, the digital crimes represent one of the big problems. Moreover, improve the detection and analysis for intensive data, which need for technique to integrate with these data to collect information from different sources.

Moreover, Rossy (Rossy & Ribaux, 2014) mentioned other problem that criminal investigation faced which it is ineffective collaboration between the partners are still poorly expressed. In addition, (Yeow et al., 2014) Yeow in his work discussed different forensic problem, which called case-based reasoning (CBR). It is representing one of the matured paradigms of artificial intelligence, also they mentioned that used intelligence system to support forensics is represented an infancy and poses a great challenge. Finally, even there are several works proposed intelligence forensic models, which achieve better results in compared with traditional models. This work aim to analysis the most popular intelligence forensic systems which give an opportunity for researchers in the future as shows in next section.

Overview of Related Work

Many researchers mentioned that most of the forensic evidence was analysis based on traditional methods, even though computational forensics can address some of the limitations to these methods (Srihari, 2010). This section provides an analysis for the most popular works that proposed based on different kinds of intelligence systems based forensic. Moreover, the

intelligence methods and algorithms have been improved several forensic aspects as following table (1) shows summaries of related works.

Table.1 Related Work

Ref	Main Methods	Description
(Rossy & Ribaux, 2014)	**Propose Special Framework**	The effective cooperation between partners of the criminal investigation represents an important to the resolution of crimes. Moreover, the authors mentioned these connections between partners still poorly, which represents a main problem of the work. Furthermore, to solve this problem the authors proposed a special framework which suggested based on related work and background. However, proposed methodological is not modeling, which makes it difficult to formation for future works.
(Yeow et al., 2014)	**Information extraction (IE) & Naïve Bayes**	The authors in work proposed a new model of case-based reasoning (CBR) for forensic based on machine learning which represents by proposed Naïve Bayes. In addition, the data that used to evaluate the proposed model is collected from the Srebrenica Historical Project. However, the authors didn't provide a critical analysis for current machine learning algorithms to gives a reason or motive to selected Naïve Bayes instead of other machine learning algorithms or methods. Moreover, the authors didn't show enough details about the dataset such as the number of features.
(Nirkhi et al., 2016)	**Multidimensional Scaling and Hierarchical Clustering**	The online message represents the main case study in this work because these messages include important information and facing different kinds of attacks or could be spying. In particular, the authorship identification is the main perspective of the online messages this work discussed. The data that has been used to evaluate the proposed model is content 619,446 images that belonging to 158 users. First algorithm is hierarchical clustering that used to measure similarity between known and unknown document. Second algorithm multidimensional scaling used to visualization the matrix results of the first algorithm that represents the distance of similarity. However, the representation of results was not clear to recognize the improved of the proposed model. Furthermore, the authors didn’t use to evaluate measurement for final results.
(Fatima et al., 2017)	**Self-Organizing Map (SOM)**	The main problem in this work about analysis an intensive data. Specially, the data that collected from different databases and sources. Moreover, authors proposed a clustering algorithm SOM to reduce the complexity to threat detection criteria. Nevertheless, there several limitations of the proposed model such as the data that used to evaluate the model which it is collected by the authors themselves, and divided into training and testing without mentioned for percentage each part and without validation too.

(Chung et al., 2017)	**Proposed a special tool called a proof of concept tool (CIFT)**	The smart device such as Amazon Echo which is a main case study of this work. Moreover, alexaenable represents a gate for all sounds and voices to acho device. These kinds of devices represent a great source of digital evidence because of, they are used in many places, which can be used to supporting investigations. Furthermore, authors proposed CIFT concept to analysis and identification of native artifacts from both cloud and client based on data called Elastic Stack which include details about web browser and mobile applications. However, the paper didn't include enough information about data that used to evaluate the proposed model. Also, there is not any benchmarks to showed the percentage of improvement.
(Spranger & Labudde, 2014)	**Proposed a special integrated computation development by authors called (Q-A system)**	This work mentioned several methods and systems of forensic texts in related work part. Furthermore, the authors address the main problem that this work tried to solve. Moreover, which it's how the search for information or specific finding among countless documents as handcrafted work is currently a time-consuming. In the result's part, the authors did not use any benchmarks to compare with any other works to show the improved percentage. In addition, addition, they didn't use enough measurement evaluation to show the enhancement of the proposed models.
(Kumar et al., 2018)	**Random Forest classifier**	The main proposed for this work to design a predictive forensic to detect suspicious malware in android applications. The authors used Random Forest classifier to build Forensic Analysis of Mobile device's using of application permissions (FAMOUS). The experiments of the proposed model are divided into two parts. Moreover, first part is compared with different machine learning algorithms based on accuracy. Second part is to compare FAMOUS with a real user's device. There are several limitations about the proposed model such as the partition of dataset without validation part, which is the impact on the accurate the results. Moreover, even the authors compared the proposed classifier with several algorithms. But on other hands, they didn't compare with most machine learning high popularity such as ELM, FLN, which in many case studies improved the results of Random Forest classifier. For the other part of experiment results. FAMOUS achieved only 94.84 accuracy
(Bornik et al., 2018)	**Proposed a new Software Framework**	The technology can increase evidence analysis, which makes it more understanding by both non expert and expert. The authors in this work proposed a new software framework and evaluate based on special 3D images Datasets which used magnetic resonance imaging (MRI) and computed tomography (CT). The authors did not give enough information about the dataset that used to evaluate

		the proposed model, also didn't propose any benchmarks to compare proposed model results with it.
(Maggi et al., 2008)	**Cluster algorithm+ Markov Model**	The authors in this work proposed forensic model based on machine learning to create an anomaly detection system. Moreover, main proposed model includes two phases. First phase represents learning step based on special dataset created by the authors. Second phase represented the machine learning process by a used cluster algorithm. The results of the proposed model didn't show enough enhancement in compared with complexity of the model. Furthermore, the authors did not evaluate the results based on measurement matrix such as Precision and recall.
(Aditya et al., 2018)	**Deep Neural Nets (DNN)**	The authors designed and implemented Adversary Testing Framework (ATF) by used DNN which represents as proposed way to enhance forensic investigations. Moreover, results of proposed model evaluated based on 400 images that divided (assault gun, assault rifle) in general. Finally, even the proposed framework showed good results in compare with benchmarks, still facing some limitations such as output results not easy to understand and analysis from normal users, which not expert, also framework was sensitive to noise.

Table.1 shows that most of the forensic based on intelligence systems achieve better results in compared with traditional methods. However, these kinds of intelligence systems still facing several limitations such as difficult to find a reliable and accurate dataset to evaluate propose systems because of data is include sensitive information, proposes a hybrid model instead of the system based on a single algorithm.

Conclusion

Current days, digital security represents an important aspect of technology field because of it is involved in most of our daily activities. Moreover, build a forensic based on Intelligence system represents a one of important field that the researchers recently focus to develop it by proposed several kinds of systems. This work provided an overview for forensic intelligence systems and analysis most of popular systems. Finally, we provide different limitations for researchers in the future such as need to provides a standard reliable dataset and propose model based on hybrid algorithms instead of model based on single algorithm to reduce the limitation of one algorithm by integrate with other algorithm.

Reference

Aditya, K., Grzonkowski, S., & Lekhac, N. (2018). Enabling Trust in Deep Learning Models: A Digital Forensics Case Study. *Proceedings - 17th IEEE International Conference on Trust,*

Security and Privacy in Computing and Communications and 12th IEEE International Conference on Big Data Science and Engineering, Trustcom/BigDataSE 2018, 1250–1255. https://doi.org/10.1109/TrustCom/BigDataSE.2018.00172

Battiato, S., Emmanuel, S., Ulges, A., & Worring, M. (2012). Multimedia in Forensics, Security, and Intelligence. *IEEE Multimedia*, *19*(1), 17–19. https://doi.org/10.1109/mmul.2012.10

Bornik, A., Urschler, M., Schmalstieg, D., Bischof, H., Krauskopf, A., Schwark, T., … Yen, K. (2018). Integrated computer-aided forensic case analysis, presentation, and documentation based on multimodal 3D data. *Forensic Science International*, *287*, 12–24. https://doi.org/10.1016/j.forsciint.2018.03.031

Chung, H., Park, J., & Lee, S. (2017). Digital forensic approaches for Amazon Alexa ecosystem. *Digital Investigation*, *22*, S15–S25. https://doi.org/10.1016/j.diin.2017.06.010

Deo, R. C., & Şahin, M. (2016). An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. *Environmental Monitoring and Assessment*, *188*(2), 1–24. https://doi.org/10.1007/s10661-016-5094-9

Fatima, H., Satpathy, S., Mahapatra, S., Dash, G. N., & Pradhan, S. K. (2017). Data fusion & visualization application for network forensic investigation - A case study. *2017 2nd International Conference on Anti-Cyber Crimes, ICACC 2017*, (March 2018), 252–256. https://doi.org/10.1109/Anti-Cybercrime.2017.7905301

Koc, L., Mazzuchi, T. A., & Sarkani, S. (2012). Expert Systems with Applications A network intrusion detection system based on a Hidden Naïve Bayes multiclass classifier. *Expert Systems With Applications*, *39*(18), 13492–13500. https://doi.org/10.1016/j.eswa.2012.07.009

Kumar, A., Kuppusamy, K. S., & Aghila, G. (2018). FAMOUS: Forensic Analysis of MObile devices Using Scoring of application permissions. *Future Generation Computer Systems*, *83*, 158–172. https://doi.org/10.1016/j.future.2018.02.001

Maggi, F., Zanero, S., & Iozzo, V. (2008). Seeing the invisible: forensic uses of anomaly detection and machine learning. *Operating Systems Review of the ACM Special Interest Group on Operating Systems (SIGOPS)*, *42*(3), 51–58. https://doi.org/10.1145/1368506.1368514

Nirkhi, S., Dharaskar, R. V., & Thakare, V. M. (2016). Authorship Verification of Online Messages for Forensic Investigation. *Physics Procedia*, *78*, 640–645. https://doi.org/10.1016/j.procs.2016.02.111

Rossy, Q., & Ribaux, O. (2014). A collaborative approach for incorporating forensic case data into crime investigation using criminal intelligence analysis and visualisation. *Science and Justice*, *54*(2), 146–153. https://doi.org/10.1016/j.scijus.2013.09.004

Rosten, E., & Drummond, T. (2006). Machine learning for high-speed corner detection, 1–14.

Spranger, M., & Labudde, D. (2014). Establishing a Question Answering System for Forensic Texts. *Procedia - Social and Behavioral Sciences*, *147*, 197–205.

https://doi.org/10.1016/j.sbspro.2014.07.152

Yeow, W. L., Mahmud, R., & Raj, R. G. (2014). An application of case-based reasoning with machine learning for forensic autopsy. *Expert Systems with Applications*, *41*(7), 3497–3505. https://doi.org/10.1016/j.eswa.2013.10.054