INFLUENCIA DE LA INTELIGENCIA ARTIFICIAL EN LA COMPUTACIÓN FORENSE

INFLUENCE OF ARTIFICIAL INTELLIGENCE IN FORENSIC COMPUTATION

Deimer Antonio Romero Madera[1]
Luis Carlos Tovar Garrido[2]
Paulo Sexto Oyola Quintero[3]

Resumen

El análisis forense digital es el medio utilizado por el investigador cibernético para rastrear al delincuente en caso de que no haya evidencia física. No obstante, la falta de mecanismos adecuados para obtener este objetivo, es un obstáculo que presenta la computación forense. Por tanto, el propósito de realizar este estudio fue determinar la influencia de la inteligencia artificial en la computación forense para resaltar su importancia e identificar ventajas que aporta al realizar un análisis forense digital, donde la investigación fue de tipo cualitativa con enfoque fenomenológico. Como resultado, se obtuvo que la computación forense se ha apoyado en el aprendizaje automático para detectar el comercio y venta de sustancias psicoactivas en redes sociales mediante algoritmos basados en patrones e inferencias sobre los proveedores de sustancias ilícitas.

Palabras claves: seguridad informática, detección de intrusos, evidencia digital, redes sociales, aprendizaje automático.

Abstract

Digital forensic analysis is the means used by the cyber investigator to track the offender in case there is no physical evidence. However, the lack of adequate mechanisms to obtain this objective is an obstacle presented by forensic computing. Therefore, the purpose of this study was to determine the influence of artificial intelligence in forensic computing to highlight its

[1] Estudiante de Ingeniería de sistemas de la Universidad de Cartagena, integrante de grupo de investigación INGESINFO. Mail: dromerom@unicartagena.edu.co
[2] Ingeniero de sistemas de la Universidad del Norte; Magister en Ciencias Computacionales de la Universidad Autónoma de Bucaramanga. Docente Investigador del programa de Ingeniería de Sistemas de la Universidad de Cartagena. Integrante del grupo de Investigación INGESINFO. ltovarg@unicartagena.edu.co
[3] Magister en Educación, Universidad de Cartagena. Docente Investigador del Programa de Administración de Empresas, Facultad de Ciencias Económicas, Universidad de Cartagena. Mail: poyolaq@unicartagena.edu.co

importance and identify advantages that it provides when performing a digital forensic analysis, where the research was of a qualitative type with a phenomenological approach. As a result, it was obtained that forensic computing has relied on machine learning to detect the trade and sale of controlled substances in social networks through algorithms based on patterns and inferences about suppliers of illegal substances.

Keywords: cybersecurity, intrusion detection, digital evidence, social networks, machine learning.

Introducción

Los avances en tecnologías y las tendencias cambiantes en el comportamiento de las personas, han llevado a un incremento en volumen, variedad, velocidad y veracidad de datos disponibles para el análisis forense digital (Quick y Raymond, 2018), el cual es un método de investigación de delitos mediante la recopilación, identificación y examinación de la información del delincuente (Satpathy, Mallick, y Pradhan, 2018), donde el objetivo principal es identificar, analizar, preservar, recuperar y presentar información y juicios específicos relacionados con la información digital (Singh, Agrawal, y Khan, 2018). El análisis forense digital se asocia principalmente con delitos informáticos relacionados con la investigación a través de computadoras y los resultados se utilizan en procedimientos judiciales civiles (Damshenas, Dehghantanha y Mahmoud, 2014).

Del mismo modo, la computación forense utiliza varias fases para mantener un estudio estructurado, lo cual facilita la verificabilidad y reproducibilidad del análisis (Presley, Landry y Black, 2018). Según Yusoff, Ismail y Hassan (2011), las fases de la computación forense son: adquisición, consiste en obtener copias de la información sospechosa que está relacionada con un acontecimiento; preservación, se debe garantizar los datos recopilados con el propósito de que no sean transformados; análisis, consiste en utilizar hardware y software diseñados especialmente para el análisis forense; documentación, se debe documentar las cuestiones críticas y relevantes a los hechos; por último, la fase presentación, consiste en entregar un informe ejecutivo que demuestre de manera resumida los rasgos más importantes de la investigación.

El análisis forense digital es importante porque es el medio utilizado por el investigador cibernético para rastrear al delincuente en caso de que no haya evidencia física (Tri, Riadi y Prayudi, 2018). No obstante, la falta de mecanismos adecuados para obtener este objetivo, es un obstáculo que presenta la computación forense (Stelly y Roussev, 2018).

Asimismo, presenta varios desafíos al recopilar evidencias de ataques en la computación en la nube, como la violación, integridad y confidencialidad de los datos (Neware y Khan,2018). Además, la carencia de capacitación a examinadores, la falta de automatización inteligente y las herramientas actuales de la computación forense, no permite manejar los casos más complejos que se presentan en un caso de investigación forense digital (Palmer, 2018).

Teniendo en cuenta lo anterior, la inteligencia artificial es una posible solución a los problemas y desafíos que presenta la computación forense (Geradts, 2018), debido a que los avances impresionantes en el campo de la inteligencia artificial contribuyen a la automatización del análisis de datos y permiten el aprendizaje automático de tareas específicas (Helbing, Frey, Gigerenzer, Hafen, Hagner, Hofstetter y Zwitter, 2017). Por ello, la finalidad de esta indagación fue determinar la influencia de la inteligencia artificial sobre la computación forense para resaltar su importancia a través de las aplicaciones e identificar ventajas que aporta al realizar un análisis forense digital.

Marco teórico

Aprendizaje automático

El aprendizaje automático es un campo complejo y multidisciplinario de investigación y desarrollo, que incluye métodos teóricos y aplicados de estadística, informática, inteligencia artificial, biología y psicología (Ortiz, 2018). Los avances en el aprendizaje automático han revolucionado muchos campos y han llevado a innovaciones que van desde los autos que conducen hasta el reconocimiento facial, asimismo, es sustancialmente mejor para hacer predicciones, debido a que no impone una estructura innecesaria en los datos (Erel, Stern, Tan y Weisbach, 2018).

Ataque cibernético

Un ataque cibernético es cualquier intento de alterar, deshabilitar, destruir, robar, dañar u obtener acceso no autorizado de un activo mediante redes de computadoras (Kundur, Feng, Liu, Zourntos, y Butler-Purry, 2010), el cual está orientado a las empresas, servicios y personas particulares, con el propósito de obtener recursos económicos al tener acceso a información privada, técnica e institucional y otros recursos de propiedad intelectual (Hathaway, Crootof, Levitz, Nix, Nowlan, Perdue, y Spiegel, 2012), donde los atacantes pueden eludir las defensas del objetivo, mantener el control sobre el medio ambiente, ex filtrar recursos sensibles, ocultar signos de intrusiones en la computadora y enviarlo a un tercero (Kao, Wang, Tsai, y Chen, 2018).

Computación forense

La computación forense se define como el manejo de información, archivos y datos disponibles en herramientas tecnológicas para revelarlos en un proceso legal, los cuales pueden ser computadoras, celulares, asistentes digitales o cualquier dispositivo digital con memoria que se pueda convertir en prueba frente un delito informático (Do, Martini, y Choo, 2015). Asimismo, el propósito de la computación forense radica en determinar las respuestas a los interrogantes: quién, cómo, qué, por qué, cuándo y dónde (Cahyani, Martini, Choo, y Al- Azhar, 2016). De la misma manera, el procesamiento y análisis de imágenes forenses completas, incluidas las de diferentes sistemas distribuidos, se realiza para determinar el valor probatorio y de inteligencia (Quick, y Raymond, 2016).

Inteligencia artificial

La inteligencia artificial es un campo de la informática que se encarga de la creación de máquinas inteligentes que funcionan y reaccionan similares a los seres humanos, la cual se ha transformado en una parte fundamental para la industria de la tecnología por ser altamente técnica y especializada (Dunjko y Briegel, 2018). Los problemas principales de los que se encarga resolver la inteligencia artificial, son: habilidad para manipular objetos, aprendizaje, razonamiento, conocimiento, percepción, planificación y resolución de problemas (Steels y Brooks, 2018).

Metodología

La investigación realizada fue de tipo cualitativa, debido a que la información consultada fue observada y analizada (Tracy, 2013; Leavy, 2017; Bordens y Abbott, 2018). De igual manera, fue descriptivo, debido a que se busca describir un fenómeno en una situación real (Grove y Gray, 2018). El enfoque aplicado fue fenomenológico (Hernández, Fernández y Baptista, 2014), porque el estudio se fundamentó en concepciones, técnicas y estudios no cuantitativos para determinar la influencia de la inteligencia artificial en la computación forense.

Técnicas de recolección de información

La técnica utilizada para la recolección de información en la presente investigación fue la revisión crítica o analítica, la cual recurre a información escrita que pudo haber sido el producto de otras investigaciones (Hurtado, 2012). Asimismo, se utilizó material bibliográfico con características relevante acerca de la relación que existe entre inteligencia

artificial y computación forense, donde la información fue extraída de documentos institucionales de acceso público, de artículos publicados en revistas indizadas y no indizadas, así como de fuentes documentales que relatan los aspectos relacionados con la temática estudiada (Cauas, 2015).

Análisis de los datos

Se aplicó el análisis documental, el cual permite la descripción objetiva y sistemática de elementos de contenido, significado y estructura del documento y su contraste con otros instrumentos de significado similar (Chacón et al., 2013). El análisis de datos se fundamentó en la determinación de la influencia de la inteligencia artificial en la computación forense, donde se tuvo en cuenta los aportes que realiza cada método o campo de la inteligencia artificial a la computación forense, con el propósito de mejorar los diferentes procesos investigativos realizados por la computación forense.

Resultados

En este apartado, se reflejan los resultados hallados a través de la revisión bibliográfica, donde se encontraron los principales aportes de la inteligencia artificial a la computación forense, como se observa en la Tabla 1.

Tabla 1. *Estudios de computación forense aplicando inteligencia artificial*

Técnicas y campos de la Inteligencia Artificial	**Descripción de la técnica**	**Estudios relacionados con la computación forense**
Aprendizaje Automático (Machine Learning)	El aprendizaje automático es el estudio científico de algoritmos y modelos estadísticos que utilizan los sistemas informáticos para realizar con eficacia una tarea específica sin utilizar instrucciones explícitas, sino que se basan en patrones e inferencia.	El aprendizaje autónomo ha permitido detectar el comercio y venta de sustancias controladas a través de Twitter por Los vendedores en línea, donde se estableció la viabilidad de un protocolo de prueba de un algoritmo de aprendizaje automático sin supervisión para detectar tweets de vendedores de opioides en línea ilícitos

		(Mackey, Kalyanam, Klugman, Kuzmenko, Gupta, 2018).
Redes neuronales artificiales (Artificial Neural Networks)	Las redes neuronales consisten en capas de entrada y salida, así como una capa oculta que realiza un conjunto de procesos basados en las unidades de entradas para arrojar algo que la capa de salida puede usar.	Se ha investigado el uso de redes neuronales en bases de datos criminales y ha llegado a varias conclusiones. Se han combinado diversos tipos de información sobre las relaciones entre 22000 delincuentes conocidos. Esto hizo uso de datos de redes sociales, informes policiales y registros de arrestos. Estos delincuentes han sido condenados por diversos delitos, como el uso y el tráfico de drogas, la extorsión, el lavado de dinero y la fabricación de drogas sintéticas (Geradts, 2018).
Redes Bayesianas (Bayesian Networks)	Las redes bayesianas se definen como un modelo gráfico probabilístico que utiliza la inferencia bayesiana para los cálculos de probabilidad, donde su objetivo consiste modelar la dependencia condicional y causalidad, al representar la dependencia condicional por bordes en un gráfico dirigido.	Los investigadores forenses digitales han utilizado las redes bayesianas para razonar sobre la evidencia y cuantificar la confiabilidad y trazabilidad de la hipótesis correspondiente (Liu, Singhal, y Wijesekera, 2016).
Minería de datos (Data Mining)	La minería de datos es el proceso de encontrar anomalías, patrones y correlaciones dentro de grandes volúmenes de datos	Estudios reflejan que, al utilizar el Marco de reducción de datos forenses digitales y la minería de datos y un

	para predecir resultados, donde se usan técnicas de aprendizaje automático, estadísticas e inteligencia artificial.	subconjunto de datos reducido, se puede comprender mejor los datos a un costo sustancialmente menor, en comparación con el almacenamiento de imágenes forenses completas (Quick y Raymond, 2014).

Fuente: autores

De acuerdo a lo planteado, la computación forense se ha apoyado en el aprendizaje automático para detectar el comercio y venta de sustancias controladas en redes sociales mediante algoritmos basados en patrones e inferencias sobre los proveedores de sustancias ilícitas. Conforme a Valenga, Britos, Perversi, Fernández, Merlino, y García (2007), plantean que la fusión de aprendizaje automático, teoría de bases de datos, visualización de datos y minería de datos, está motivada por el valioso crecimiento de los datos, con el propósito de adquirir la mayor cantidad de elementos para constituir políticas de inteligencia criminal conforme a los datos útiles en los diferentes soportes.

Por otro lado, la aplicación de redes neuronales en los procesos forenses, ha permitido agilizar los hallazgos de evidencias digitales en procesos judiciales, donde se ha mezclado información de diferentes fuentes como redes sociales, informes policiales y registros de arrestos para hallar a los criminales. Asimismo, Hutson (2017), plantea que la inteligencia artificial automatiza las decisiones judiciales en un proceso penal, teniendo en cuenta que se deben considerar y validar las decisiones tomadas.

Del mismo modo, las redes bayesianas permiten razonar sobre la evidencia y cuantificar la confiabilidad y trazabilidad de hipótesis forense. Asimismo, la minería de datos mejora la comprensión de los datos forenses digitales en comparación con el almacenamiento de imágenes forenses completas. Según Biedermann, Voisard, y Taroni, (2012), plantean que la literatura pertinente sobre las aplicaciones de las redes bayesianas para la inferencia en la ciencia forense, es bastante escasa, lo cual implica que científicos interesados no puedan familiarizarse con esta temática.

Finalmente, un desafío importante para el análisis forense digital es el continuo crecimiento del volumen de datos capturados y presentados para su análisis, por lo cual la influencia de la inteligencia artificial sobre la computación forense es alta, debido a que le proporciona herramientas que optimizan el análisis de datos forenses. Conforme a Quick y

Raymond (2014), plantean que las soluciones a este tipo de problemas van desde la extracción y reducción de datos, el aumento del poder de procesamiento, el procesamiento distribuido, la inteligencia artificial y otros métodos innovadores.

Conclusiones

A partir de los resultados, se concluye que la inteligencia artificial tiene una alta influencia en la computación forense, debido a que proporciona herramientas que optimizan el análisis de datos forense para facilitar la detección de comercio y venta de sustancias psicoactivas, agilizar los hallazgos de evidencias digitales en procesos judiciales, analizar la evidencia y cuantificar la confiabilidad y trazabilidad de hipótesis forense. Asimismo, la aplicación de inteligencia artificial en la computación forense, permite la detección de lavado de activos, extorsión y comercialización de drogas sintéticas en redes sociales a través de algoritmos de aprendizaje automático.

Referencias bibliográficas

Biedermann, A., Voisard, R., & Taroni, F. (2012). Learning about Bayesian networks for forensic interpretation: An example based on the 'the problem of multiple propositions'. Science & Justice, 52(3), 191-198.

Bordens, K., y Abbott, B. (2018). *Research Design and Methods: A Process Approach*. New York: McGraw-Hill Education.

Cahyani, N., Martini, B., Choo, K., y Al- Azhar, A. (2016). Forensic data acquisition from cloud- of- things devices: windows Smartphones as a case study. *Concurrency and Computation: Practice and Experience*, *29*(14).

Cauas, D. (2015). Definición de las variables, enfoque y tipo de investigación. Bogotá: biblioteca electrónica de la universidad Nacional de Colombia.

Chacón, J. W. B., Herrera, J. C. B., y Villabona, M. R. (2013). Revisión y análisis documental para estado del arte: una propuesta metodológica desde el contexto de la sistematización de experiencias educativas. Investigación Bibliotecológica: archivonomía, bibliotecología e información, 27(61), 83-105.

Damshenas, M., Dehghantanha, A., y Mahmoud, R. (2014). A survey on digital forensics trends. International Journal of Cyber-Security and Digital Forensics, 3(4), 209 - 235.

Do, Q., Martini, B., y Choo, K. (2015). A forensically sound adversary model for mobile devices. PloS one, 10(9).

Dunjko, V., y Briegel, H. J. (2018). Machine learning & artificial intelligence in the quantum domain: a review of recent progress. Reports on Progress in Physics, 81(7).

Erel, I., Stern, L., Tan, C., y Weisbach, M. (2018). Selecting Directors Using Machine Learning. National Bureau of Economic Research, University of Washington, Seattle.

Geradts, Z. (2018). Digital, big data and computational forensics. Forensic Sciences Research, 3(3), 179 - 182.

Grove S. & Gray, J. (2018). Understanding Nursing Research E-Book: Building an Evidence-Based Practice. Elsevier Health Sciences.

Hathaway, O., Crootof, R., Levitz, P., Nix, H., Nowlan, A., Perdue, W., y Spiegel, J. (2012). The law of cyber-attack. California Law Review, 817 - 885.

Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., y Zwitter, A. (2017). Will Democracy Survive Big Data and Artificial Intelligence? We are in the middle of a technological upheaval that will transform the way society is organized. We must make the right decisions now. Scientific American, 1 - 48.

Hernández, R., Fernández, C., y Baptista, P. (2014). Metodología de la investigación. México. Mc Graw Hill.

Hurtado, J. (2012). "El proyecto de investigación". (7ª Edición). Ediciones Quirón. Venezuela.

Hutson, M. (2017). Artificial intelligence prevails at predicting Supreme Court decisions. Science.

Kao, D., Wang, Y., Tsai, F. y Chen, C. (2018). Forensic analysis of network packets from penetration test toolkits. In Advanced Communication Technology (ICACT), 2018 20th International Conference on IEEE, 363-368.

Kundur, D., Feng, X., Liu, S., Zourntos, T., y Butler-Purry, K. (2010). Towards a framework for cyber-attack impact analysis of the electric smart grid. In Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference, 244 - 249.

Leavy, P. (2017). Research Design. Nueva York: The Guildford Press.

Liu, C., Singhal, A., y Wijesekera, D. (2016). A probabilistic network forensic model for evidence analysis. In IFIP International Conference on Digital Forensics, 189-210.

Mackey, T., Kalyanam, J., Klugman, J., Kuzmenko, E., & Gupta, R. (2018). Solution to Detect, Classify, and Report Illicit Online Marketing and Sales of Controlled Substances via Twitter: Using Machine Learning and Web Forensics to Combat Digital Opioid Access. Journal of medical Internet research, 20(4).

Neware, R., y Khan, A. (2018). Cloud Computing Digital Forensic challenges. In IEEE 2nd International conference on Electronics, Communication and Aerospace Technology (ICECA 2018). Coimbatore, India.

Ortiz, G. B. (2018). IBM RPG Software Quality Prediction using Machine Learning Techniques. Universidad César Vallejo.

Palmer, I. N. (2018). Forensic analysis of computer evidence (Doctoral dissertation). University of Illinois at Urbana-Champaign, Illinois, United States.

Presley, S., Landry, J., y Black, M. (2018). Using Project Management Knowledge and Practice to Address Digital Forensic Investigation Challenges. In 2018 KSU Proceedings on Cybersecurity Education, Research and Practice. Georgia, United States.

Quick, D., y Raymond, K. (2014). Impacts of increasing volume of digital forensic data: A survey and future research challenges. Digital Investigation, 11(4), 273-294.

Quick, D., y Raymond, K. (2014). Data reduction and data mining framework for digital forensic evidence: Storage, intelligence, review and archive. Australian Institute of Criminology, 480, 1 - 11.

Quick, D., y Raymond, K. (2016). Big forensic data management in heterogeneous distributed systems: quick analysis of multimedia forensic data. Software: Practice and Experience, 47(8), 1095 - 1109.

Quick, D. y Raymond, K. (2018). Digital forensic intelligence: Data subsets and Open Source Intelligence (DFINTOSINT): A timely and cohesive mix. Future Generation Computer Systems, 78(2), 558 - 567.

Satpathy, S., Mallick, C., & Pradhan, S. K. (2018). Big Data Computing Application in Digital Forensics Investigation and Cyber Security. International Journal of Computer Science and Mobile Applications, 129 - 136.

Singh, N., Agrawal, A., & Khan, R. A. (2018). Voice Biometric: A Technology for Voice Based Authentication. Advanced Science, Engineering and Medicine, 10(1), 1 - 6.

Steels, L., y Brooks, R. (2018). The Artificial Life Route to Artificial Intelligence Building Embodied, Situated Agents. London, England: Routledge.

Stelly, C., y Roussev, V. (2018). Nugget: A digital forensics language. Digital Investigation, 24, 38- 47.

Tracy, S. (2013). Qualitative Research Methods: Collecting Evidence, Crafting Analysis, communicating impact. Malden, USA: Wiley.

Tri, M., Riadi, I., y Prayudi, Y. (2018). Forensics Acquisition and Analysis Method of IMO Messenger. International Journal of Computer Applications, 179(47), 9 - 14.

Valenga, F., Britos, P. V., Perversi, I., Fernández, E., Merlino, H., & García Martínez, R. (2007). Aplicación de minería de datos para la exploración y detección de patrones delictivos en Argentina. In XIII Congreso Argentino de Ciencias de la Computación.

Yusoff, Y., Ismail, R., y Hassan, Z. (2011). Common phases of computer forensics investigation models. International Journal of Computer Science & Information Technology, 3(3), 17-31.