

Aplicación de minería de datos e Internet de las Cosas (IoT) para investigaciones agropecuarias sobre artrópodos

Mariano R. Droz ⁽¹⁾, Carlos E. Alvez ⁽¹⁾, Graciela R. Etchart ⁽¹⁾, Beatriz M. Diaz ⁽²⁾

⁽¹⁾ Facultad de Ciencias de la Administración – Universidad Nacional de Entre Ríos
Av. Monseñor Tavella 1424 – (CP 3.200) Concordia, Entre Ríos, República Argentina

⁽²⁾ Estación Experimental Agropecuaria Concordia – Instituto Nacional de Tecnología Agropecuaria
Ruta Provincial 22 y vías del Ferrocarril – (CP 3.200) Estación Yuquerí, Concordia, Entre Ríos,
República Argentina

{drozm, caralv, getchart}@fcad.uner.edu.ar
diaz.beatriz@inta.gob.ar

RESUMEN

La Minería de Datos permite extraer información oculta y descubrir patrones y conocimientos interesantes a partir de grandes cantidades de datos. Las fuentes de los mismos pueden incluir bases de datos, *Data Warehouses*, la Web, u otros depósitos de información. Así, en el contexto actual, las tecnologías denominadas Internet de las cosas (IoT)¹ se han convertido en un proveedor de datos a gran escala. A partir de ello, se observa una estrecha relación entre la Minería de Datos e IoT, ya que las fortalezas de ambas permiten una complementación beneficiosa. Por ello, es posible encontrar nuevas aplicaciones en diferentes ámbitos. Algunas áreas donde dicha relación se está haciendo evidente, consiste en la agricultura de precisión y en la investigación agropecuaria. En esta última, existen oportunidades y nuevos campos de aplicación, como ser el estudio de la biodiversidad mediante la investigación sobre artrópodos en agroecosistemas. Resultando potencialmente factible aplicar Minería de Datos y utilizar una trampa de caída por tiempo con tecnología IoT, para determinar patrones de comportamiento de artrópodos, en función de factores abióticos relevantes para el campo de estudio.

Palabras clave: Bases de Datos, Minería de Datos, Internet de las cosas (IoT), Agroecosistemas, Artrópodos.

CONTEXTO

Este trabajo es desarrollado en el marco de una Tesis de la Maestría en Sistemas de Información de la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos. A su vez, está vinculado al Proyecto PID “Modelos de Machine Learning para la mejora de la precisión, seguridad y eficiencia en la gestión de datos biométricos” de la Universidad Nacional de Entre Ríos.

Por otra parte, se enmarca en la línea de trabajo “Desarrollo de tecnologías de bajo impacto ambiental aplicadas a la horticultura”, que lleva a cabo el Grupo Hortícola de la Estación Experimental Agropecuaria Concordia del Instituto Nacional de Tecnología Agropecuaria.

1. INTRODUCCIÓN

La Minería de Datos (*Data Mining*) implica descubrir patrones novedosos, interesantes y potencialmente útiles de grandes conjuntos de datos y aplicar algoritmos para la extracción de información oculta. Muchas otras expresiones se usan para hacer referencia a las técnicas de Minería de Datos, entre ellas: descubrimiento de conocimiento en bases de datos (*Knowledge Discovery -mining- in Databases, KDD*), extracción de conocimiento (*knowledge extraction*), análisis de patrones (*data/pattern analysis*), arqueología de datos (*data archeology*) y recolección de información (*information harvesting*) [1], [2].

¹ Estas tecnologías también son denominadas Internet de los Objetos o, en inglés, *Internet of Things*.

Una visión amplia de las técnicas de Minería de Datos permite definir las como el proceso de descubrir patrones y conocimientos interesantes a partir de grandes cantidades de datos, donde las fuentes de los mismos pueden incluir bases de datos, *Data Warehouses*, la Web, u otros depósitos de información o datos [1].

A partir de esta última definición, actualmente es posible determinar que una fuente importante de datos para aplicar las técnicas de Minería de Datos, radica en las tecnologías denominadas Internet de las cosas (IoT).

El IoT puede concebirse como una infraestructura global de la sociedad de la información, que permite ofrecer servicios avanzados mediante la interconexión de objetos (físicos y virtuales), gracias a la interoperatividad de tecnologías de la información y la comunicación (TIC) presentes y futuras [3]. En el paradigma de IoT, muchos de los objetos que nos rodean están conectados de una forma u otra, generando enormes cantidades de datos que deben almacenarse, procesarse y presentarse de forma transparente, eficiente y fácilmente interpretable [4].

En este sentido, los datos recolectados por los dispositivos IoT se utilizan para comprender y controlar entornos complejos, lo que permite una mejor toma de decisiones, mayor automatización, eficiencia, productividad, precisión y reducción de costos [5]. Asimismo, es posible afirmar que este tipo de dispositivos se constituyen como una fuente de datos en permanente expansión en diferentes ámbitos de aplicación.

Por consiguiente, un desafío importante en este escenario, es el análisis oportuno de grandes cantidades de datos para producir conocimiento y decisiones altamente confiables y precisas, para que IoT pueda cumplir con sus expectativas [5].

Como consecuencia de lo anterior, surge la necesidad de hacer que IoT sea más inteligente, introduciendo tecnologías de análisis de datos. En efecto, una de las tecnologías más valiosas en este campo es la Minería de Datos [2], la cual ha tenido su origen a partir de la necesidad de

descubrir automáticamente información valiosa de enormes cantidades de datos y transformar esos datos en conocimiento organizado [1].

De este modo, se observa una estrecha relación entre la Minería de Datos e IoT, ya que la aplicación de la misma en bases de datos recolectados o generados por dispositivos IoT permite y permitirá utilizar efectivamente ambas tecnologías en procura de objetivos comunes.

Las áreas de trabajo de estas tecnologías son vastas y cada vez es posible encontrar nuevas aplicaciones de las mismas en diferentes ámbitos. En lo que concierne a investigaciones agropecuarias, se evidencia un interés creciente en monitorear y recolectar datos sensibles del entorno, como ser: temperatura y humedad ambiente, temperatura y humedad del suelo, radiación ultravioleta (UV), nivel de precipitaciones, entre otros. Así, estos datos podrían ser correlacionados con otros de tipo temporal, de posicionamiento y de rendimiento de cultivos, con el fin de descubrir patrones ocultos, optimizar, predecir y aportar precisión a determinados procesos productivos, área denominada agricultura de precisión.

Investigaciones recientes han obtenido resultados promisorios [6], [7], desarrollando sistemas de riego inteligentes, basados en una red de sensores IoT que recolectan datos relevantes para el problema, aplicándose técnicas de Minería de Datos con el propósito de generar nuevo conocimiento, mejorar la toma de decisiones y el proceso productivo mediante una optimización de la gestión del riego. Otro aporte interesante, consiste en el desarrollo de un prototipo de un sistema de control para el cultivo hidropónico del tomate [8], el cual utiliza tecnología IoT y desarrolla e implementa una *Deep Neural Network* para controlar con mayor precisión el sistema en tiempo real.

Cabe señalar también, que otra línea específica de investigación agropecuaria consiste en el estudio de la biodiversidad. En este campo, alcanzar una mayor comprensión sobre la dinámica de las especies y sobre la estructura y funcionamiento de los ecosistemas resulta ser apremiante, dada la incertidumbre sobre la dimensión de los posibles efectos del cambio climático sobre la misma [9], [10].

En esta línea de investigación, un área de trabajo comprende el estudio sobre artrópodos en agroecosistemas, donde la idea básica subyacente al uso de artrópodos como indicadores de características generales de los ecosistemas es que, la sola presencia (o abundancia) de algunas especies es el producto de una larga y compleja lista de circunstancias que la han hecho posible [11]. Así, la existencia o no de determinadas especies de artrópodos da cuenta de la biodiversidad de un agroecosistema específico.

En este campo de estudio, es habitual la utilización de trampas de caída (*pitfall trap*) y de trampas de caída por tiempo (*time-sorting pitfall trap*). El avance tecnológico también ha permitido la evolución de estas trampas [12], [13], mediante la incorporación de sensores y unidades de registro locales de datos (*datalogger*). Específicamente [13], ha desarrollado una trampa que registra la fecha, la hora, el número del tubo muestreado y la temperatura del suelo. Posteriormente, se relacionan estos datos con la identificación y conteo de los artrópodos capturados en distintos experimentos. A su vez, en su trabajo también se menciona la posibilidad de expandir este tipo de trampas mediante la incorporación de sensores adicionales.

De este modo, surge el interés por desarrollar un prototipo que incorpore nuevos sensores para relevar factores abióticos relevantes para el campo de estudio. Sumado a lo anterior, la posibilidad de dotar de tecnología IoT abre nuevas posibilidades a dichos dispositivos, permitiendo complementar metodologías de investigación específicas de la disciplina, descubriendo y determinando patrones de actividad de los artrópodos del suelo, mediante la aplicación de Minería de Datos en una base de datos específicamente diseñada para este fin.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

El presente trabajo se enmarca en el análisis y aplicación de técnicas de Minería de Datos con fines científicos. Asimismo, cuenta con un enfoque transversal, ya que explorará la relación

entre estas técnicas e IoT, con el propósito de diseñar una base de datos que almacenará datos recolectados por una trampa de caída por tiempo dotada de tecnología IoT, la cual registrará factores abióticos tales como temperatura y humedad ambiente, temperatura y humedad del suelo y radiación ultravioleta (UV).

A partir de ello, específicamente el trabajo apunta a la aplicación de Minería de Datos e IoT para el estudio de la biodiversidad e impacto del cambio climático, mediante la investigación agropecuaria sobre patrones de comportamiento de artrópodos del suelo.

Por otra parte, esta propuesta se encuadra en la línea de investigación “*Bases de Datos*”, establecida como una de las líneas prioritarias dentro de la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos, según las Resoluciones “C.D.” N° 160/11 y 203/11.

Asimismo, se relaciona con las denominadas AgroTIC, según lo expresado en el “*Libro Blanco de la Prospectiva TIC - Proyecto 2020*” [14]. Concretamente, involucra distintas áreas de trabajo, entre ellas: bases de datos y gestión del conocimiento, y sistemas de adquisición y administración de datos incorporados a dispositivos de uso específico.

3. RESULTADOS ESPERADOS

El objetivo general consiste en describir, analizar y aplicar Minería de Datos en una base de datos específicamente diseñada para dispositivos científicos con tecnología IoT utilizables en investigaciones agropecuarias sobre patrones de comportamiento de artrópodos.

Para cumplir con dicho objetivo, se presentarán y clasificarán las principales técnicas de la Minería de Datos. A su vez, se describirán y analizarán las características relevantes del IoT. Luego de ello, se profundizará en la aplicación de estas técnicas a partir de datos obtenidos por dispositivos con tecnología IoT.

A continuación, se creará el diseño conceptual de una trama de caída por tiempo

con nuevos sensores y tecnología IoT apta para investigaciones agropecuarias sobre artrópodos. Posteriormente, se diseñará una base de datos para recolectar los datos generados por dicho dispositivo. Así, la misma será puesta a prueba mediante simulaciones de captura y almacenamiento de datos, las cuales darán paso a una etapa de aplicación de las técnicas de Minería de Datos con el propósito de efectuar una determinación de patrones de comportamiento de artrópodos del suelo en función de factores abióticos.

De este modo, el aporte original del trabajo consiste en aplicar una innovación al utilizar la Minería de Datos para la determinación de patrones de comportamiento de artrópodos del suelo, a partir de datos recolectados por una trampa de caída por tiempo con nuevos sensores y tecnología IoT, respecto a lo cual aún no se observa evidencia científica donde se hayan utilizado ambas tecnologías para dicha finalidad.

4. FORMACIÓN DE RECURSOS HUMANOS

Las tareas se desarrollarán como parte de las actividades planificadas para que uno de los integrantes realice la Tesis de la Maestría en Sistemas de Información de la Facultad de Ciencias de la Administración de la Universidad Nacional de Entre Ríos, acreditada por la Resolución CONEAU N° 1276/12.

A su vez, proporcionará un marco propicio para la formación e iniciación en actividades de investigación, mediante la realización de Proyectos de Investigación Científica, Desarrollo e Innovación Tecnológica en el marco del Programa de Investigación y Desarrollo la Universidad Nacional de Entre Ríos.

De igual modo, será un ámbito pertinente para que estudiantes de la Universidad Nacional de Entre Ríos, de carreras de grado afines, lleven a cabo sus proyectos de Trabajo Final o participen como Becarios de Iniciación en la Investigación.

5. BIBLIOGRAFÍA.

- [1] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2011.
- [2] A. V. Vasilakos, J. Wan, F. Chen, X. Rong, D. Zhang, and P. Deng, “Data Mining for the Internet of Things: Literature Review and Challenges,” *Int. J. Distrib. Sens. Networks*, vol. 11, no. 8, p. 431047, 2015.
- [3] UIT-T, “UIT-T Rec. Y.2060 Descripción general de Internet de los objetos,” *Sect. Norm. las Telecomunicaciones la UIT*, no. 2012-06-15, p. 20, 2012.
- [4] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of Things (IoT): A vision, architectural elements, and future directions,” *Futur. Gener. Comput. Syst.*, vol. 29, no. 7, pp. 1645–1660, 2013.
- [5] F. Alam, R. Mehmood, I. Katib, and A. Albeshri, “Analysis of Eight Data Mining Algorithms for Smarter Internet of Things (IoT),” *Procedia Comput. Sci.*, vol. 58, no. DaMIS 2016, pp. 437–442, 2016.
- [6] A. Goap, D. Sharma, A. K. Shukla, and C. Rama Krishna, “An IoT based smart irrigation management system using Machine learning and open source technologies,” *Comput. Electron. Agric.*, vol. 155, no. September, pp. 41–49, 2018.
- [7] J. Muangprathub, N. Boonnam, S. Kajornkasirat, N. Lekbangpong, A. Wanichsombat, and P. Nillaor, “IoT and agriculture data analysis for smart farm,” *Comput. Electron. Agric.*, vol. 156, no. November 2018, pp. 467–474, 2019.
- [8] M. Mehra, S. Saxena, S. Sankaranarayanan, R. J. Tom, and M. Veeramanikandan, “IoT based

- hydroponics system using Deep Neural Networks,” *Comput. Electron. Agric.*, vol. 155, no. October, pp. 473–486, 2018.
- [9] E. Uribe, “Estudios del cambio climático en América Latina: El cambio climático y sus efectos en la biodiversidad en América Latina,” *Euroclima*, vol. vol 51, no. 3, p. 86, 2015.
- [10] D. U. Hooper, E. C. Adair, B. J. Cardinale, J. E. K. Byrnes, B. A. Hungate, K. L. Matulich, A. Gonzalez, J. E. Duffy, L. Gamfeldt, and M. I. O’Connor, “A global synthesis reveals biodiversity loss as a major driver of ecosystem change,” *Nature*, vol. 486, p. 105, May 2012.
- [11] I. Ribera and G. Foster, “El uso de artrópodos como indicadores biológicos,” *Bol. la Soc. Entomológica Aragon.*, vol. 20, pp. 265–276, 1997.
- [12] S. Buchholz, “Design of a time-sorting pitfall trap for surface-active arthropods,” *Entomol. Exp. Appl.*, vol. 133, no. 1, pp. 100–103, 2009.
- [13] M. S. McMunn, “A time-sorting pitfall trap and temperature datalogger for the sampling of surface-active arthropods,” *HardwareX*, vol. 1, pp. 38–45, 2017.
- [14] G. Baum, A. Artopoulos, C. Aguerre, I. Albornoz, V. Robert, and MINCYT, *Libro Blanco de la perspectiva TIC, Proyecto 2020*. 2009.