

Integración de Minería de Datos y Sistemas Multiagente: un campo de investigación y desarrollo

Ms.C. Guillermo Molero Castillo
Ms.C. María Elena Meda Campaña

En este artículo se presenta un panorama de la integración e interacción de Minería de Datos y Sistemas Multiagente, estos, con el objetivo de dar a conocer alternativas de investigación relacionadas a las dos áreas. Se presenta un ejemplo ilustrativo de la integración de Minería de Datos y Tecnología Multiagente. Se listan algunos temas de estudio que podrían ser de interés para su investigación y desarrollo. Finalmente, se aborda las conclusiones finales, donde se expone, que la Tecnología Multiagente y Minería de Datos han surgido como dos campos de conocimiento que se relacionan entre sí, abriendo espacio para la integración e interacción de ambas áreas.

Palabras clave: Minería de Datos, Agentes, Sistemas Multiagente

RESUMEN

ABSTRACT

This paper presents an overview on the integration and interection of Data Mining and Systems Multiagente. Its aim is to give some research alternative that are related to both areas. It is presente an example to the integration of Data Mining Now a days, the use of Data Mining includes to discovering new information and acquaring knowledge. It is usefull to developing new methods to make the comprehension of different phenomenon easier. In this way, Data Mining is considered as a tool for suppoting the make decisions. Finally, it is a conclusion that the Tecnología Multiagente and Data Mining have emerged as a knowledge field very related, and it is the oportunity to the integration and interection of both areas.

Key words: Data Mining, Agents, Systems Multiagente

Introducción

En la actualidad, se presenta una excelente oportunidad para crear innovadoras herramientas y sistemas de interacción entre Minería de Datos y Sistemas Multiagente, las cuales pueden ofrecer resultados significativos para esta nueva tecnología. Por ejemplo, herramientas para el descubrimiento de conocimiento de fuentes de datos distribuidas, aprendizaje automático y toma de decisiones recurrentes en entornos distribuidos. Actualmente, los sistemas de esta magnitud no son posibles sin la integración de Agentes y Minería de Datos.

En la literatura existen variadas formas de definir Minería de Datos y Sistemas Multiagente, motivo por el cual se resume que Minería de Datos es el proceso de encontrar información útil y de interés en un conjunto de datos; mientras que Sistemas Multiagente es la fusión de varios agentes informáticos que interactúan entre sí, con la finalidad de ejecutar tareas específicas de forma automática y flexible para dar solución a determinados problemas, en virtud de los objetivos establecidos.

En este artículo se presenta un panorama de

la relación actual entre la Minería de Datos y los Sistemas Multiagente, áreas originalmente separadas, que se aborda con el objetivo de dar a conocer la integración e interacción de ambos campos de conocimiento, como oportunidad de realizar investigaciones de innovación y desarrollo, ligadas no sólo al ámbito académico, sino a diversos campos de la actividad humana.

Minería de Datos

Una realidad de la *Minería de Datos* es su papel como tecnología de apoyo, que ha venido a resolver dos grandes retos: trabajar con conjuntos de datos para extraer y descubrir información de interés, y usar técnicas adecuadas para explorar, analizar, comprender e identificar tendencias y comportamientos que faciliten una mejor comprensión de los fenómenos que nos rodean y sirvan de ayuda en la toma de decisiones (Molero, 2008).

En general, el impulso al empleo de Minería de Datos se asocia a una serie de factores como (Larose, 2005):

- El crecimiento exponencial de la recolección de datos y la evolución del poder de cómputo.
- El almacenamiento de los datos en data warehouses, con la finalidad de tener acceso a una base de datos actualizada y confiable.
- La creciente disponibilidad de información en Internet.
- La presión competitiva del mercado en una economía globalizada.
- El desarrollo de herramientas comerciales para llevar a cabo la Minería de Datos.

Estos factores han llevado a considerar a la Minería de Datos como un campo interdisciplinario que involucra diversas disciplinas, como: Tecnología de Bases de Datos, Sistemas para la toma de decisiones, Estadística, Inteligencia Artificial, Aprendizaje Automático, Redes Neuronales Artificiales, Reconocimiento de patrones, Visualización de datos, Computación paralela, entre otras (Hand *et al.*, 2001; Hernández *et al.*, 2004; Sumathi y Sivanandam, 2006).

Diferentes áreas que se relacionan con la Minería de Datos donde se evidencia su aplicación:

Finanzas, análisis de mercado, procesos industriales, medicina, biología, química, telecomunicaciones, seguridad, análisis científico, biometría, análisis ambiental, análisis hidrológicos, entre otros.

Asociación de diferentes disciplinas que se relacionan con la Minería de Datos:

- Tecnología de Bases de Datos
- Estadística
- Redes Neuronales Artificiales
- Aprendizaje Automático
- Computación Paralela
- Inteligencia Artificial
- Reconocimiento de Patrones
- Otras disciplinas

En la actualidad, la necesidad de afrontar nuevos desarrollos de Minería de Datos en entornos distribuidos, ha motivado la investigación de Minería de Datos Distribuida (Moemeng *et al.*, 2009), las misma pone de manifiesto las oportunidades de mejor, debido a la gran cantidad de información descentralizada disponible (Cao, 2009). Por lo que uno de los mayores esfuerzos de la Minería de Datos es en función de mejora el rendimiento de sus técnicas; esto mediante el diseño eficiente de las estructuras de datos y los métodos de cálculo para reducir la complejidad computacional. Esto es debido a la importancia de hacer frente a la reducción de tiempos de respuesta en el proceso de Minería de Datos, procesos que funcionen prácticamente en línea, por ejemplo, respuesta a un fraude con tarjeta de crédito.

Sistemas Multiagente

Los Agentes son entes o programas informáticos que perciben el entorno en el que están situados y a partir de tales percepciones ejecutan acciones de forma automática y flexible para dar solución a problemas específicos previamente establecidos (Jennings *et al.*, 1998).

Son diversos los ámbitos de la ciencia en los que la tecnología de Agentes contribuye: interfaces de usuario, comercio electrónico, recuperación de información, robótica, juegos, educación, simulación, entre otros. Entre las principales características de los Agentes destacan (Botti *et al.*, 2002; Ghedini, 2009):

- Autonomía, es la capacidad de actuar sin intervención humana directa o de otros agentes.
- Reactividad, es la capacidad de reacción ante los estímulos que perciben del entorno en el que están inmersos.

- Interacción, capacidad de interactuar con otros agentes utilizando algún medio de comunicación entre agentes.

- Iniciativa, capacidad de reacción no sólo ante los cambios que se producen en el entorno, sino la iniciativa de cumplir con los objetivos establecidos.

Un Sistema Multiagente es la fusión de varios agentes inteligentes que interactúan entre sí, teniendo como responsabilidad una tarea específica, en virtud de los objetivos establecidos (Baazaoui *et al.*, 2005; Chao y Wong, 2009). Son distintos los campos en los que se pueden aplicar los Sistemas Multiagente, por ejemplo (Botti *et al.*, 2002):

- Ingeniería de Software, para el manejo de clases, herencia, mensajes entre objetos, encapsulamiento, entre otros.

- Inteligencia Artificial, para la representación del conocimiento, manejo del razonamiento y aprendizaje, manejo de sensores, desarrollo de procesos inteligentes, entre otros.

- Sistemas Distribuidos, para la construcción de aplicaciones distribuidas, distribución de datos y procesos, manejo de conectividad, redes y protocolos de comunicación.

Una de las características de un Sistema Multiagente es que el grupo de agentes que lo integran deben trabajar de manera conjunta e individual (Chao y Wong, 2009). De manera conjunta, para cumplir tareas globales que surgen a consecuencia de la búsqueda de solución de un problema general; y de manera individual, porque las tareas globales son descompuestas en subtareas, generando tareas específicas para cada uno de los agentes que participarán en la solución del problema.

Integración de Minería de Datos y Sistemas Multiagente

La integración de tecnologías es un avance natural en el desarrollo de la ciencia. Esta integración puede trascender en una nueva tecnología cuyo poder puede acarrear potenciales beneficios de solución. De manera tal, la integración e interacción de Minería de Datos y Sistemas Multiagente es a consecuencia de problemas específicos que enfrentan cada una de estas áreas. Problemas que surgen a causa del constante avance tecnológico que se vive en la actualidad, por ejemplo:

- Los Sistemas Multiagente se enfrentan al problema de mejorar la capacidad de aprendizaje de los agentes, motivando así su integración con la Minería de Datos, que

permitirá descubrir el conocimiento y mejorar significativamente la capacidad de aprendizaje de los agentes, esto es, mediante la alimentación de algoritmos de razonamiento como: redes neuronales artificiales, razonamiento probabilístico, algoritmos genéticos, árboles de decisión, lógica difusa, entre otros.

- La Minería de Datos se enfrenta a problemas con fuentes de datos distribuidas, puesto que los datos son cada vez más heterogéneos y distribuidos en diferentes lugares, que con frecuencia tienden a seguir creciendo, por ejemplo, bases de datos de una cadena de supermercados. Por lo que, incluir tecnología Multiagente puede mejorar aspectos como: integración de varias fuentes de datos, acceso a aplicaciones distribuidas, interacción con diferentes usuarios y comunicación de diversas aplicaciones.

Estos dos ejemplos ilustran el potencial de la integración de Minería de Datos y Sistemas Multiagente como alternativa de solución a los problemas presentados en ambas áreas.

De acuerdo a Cao (2009), la integración e interacción de Minería de Datos y Sistemas Multiagente son producto de la unión de diferentes elementos como: recursos, infraestructura, conocimiento, aprendizaje, interacción, interfaz, entorno social, aplicación, entre otros (Fig. 1).

Los **recursos** permiten la integración de los datos y la información; la **infraestructura** facilita la interacción con la arquitectura y el proceso de las partes; el **conocimiento** provee la sapiencia del experto humano, el metaconocimiento o el conocimiento recuperado, extraído o descubierto en los recursos; el **aprendizaje** proporciona los métodos y capacidades de aprendizaje; la **interacción** permite la coordinación, cooperación y configuración de comunicación; la **Interfaz** permite el modelado, diseño e implementación de la interfaz de usuario; el **entorno social** proporciona los factores sociales y de organización, por ejemplo, los roles humanos; y la **aplicación** facilita la integración e interacción de las aplicaciones disponibles.

En particular, como una nueva y promisoría área de investigación, la integración entre Minería de Datos y Sistemas Multiagente tiene un futuro expansivo que puede dar lugar a técnicas y aplicaciones innovadoras que vayan más allá de las áreas individuales. Para efectos ilustrativos muestra una aplicación de vanguardia presentada por Chao y Wong (2009) es *i⁺DiaMAS*, actualmente en continuo desarrollo, que es una herramienta



Fig. 1. Elementos de integración e interacción de Agentes y Minería de Datos. Fuente: Adaptado de Cao (2009)

de Minería de Datos basada en estrategia Multiagente, que sirve de apoyo al personal médico para el diagnóstico de casos clínicos de forma objetiva, los cuales son inferidos a través de un árbol de decisión que analiza reglas de clasificación adquiridas de historias clínicas y fuentes distribuidas de datos médicos.

El principio de la arquitectura de *i+DiaMAS* se fundamenta en la distribución de tareas a varios Agentes que tienen la capacidad de procesar tareas específicas, quienes se comunican mediante el envío de mensajes. En total, son cinco los tipos de Agentes significativos que hacen frente a las diversas tareas de Minería de Datos (Fig. 2), los cuales a continuación son brevemente abordados:

- *Agente coordinador*. Constituye el componente principal de *i+DiaMAS*, que desempeña el papel de coordinar las

actividades de los diversos agentes, siendo capaz de dividir una tarea en subtareas que se delegan a agentes específicos, para posteriormente combinar los resultados de éstos y dar respuesta global a una tarea. Además, es responsable de la comunicación de los agentes y la incorporación de nuevos agentes.

- *Agente de preprocesamiento de datos*. Este agente se encarga de procesar datos útiles a partir de la fuente de datos principal. Entre las tareas que realiza destacan: manejo de valores nulos y/o faltantes, discretización de datos, selección de variables, entre otros.

- *Agente de minería de datos*. Es el núcleo de *i+DiaMAS* que abarca el descubrimiento de conocimiento para diagnosticar una enfermedad específica. El aprendizaje se basa en la inferencia de reglas de

clasificación mediante un árbol de decisión que es capaz de manejar nuevos casos médicos de forma incremental.

- *Agente de optimización*. El rol de este agente es optimizar el proceso de aprendizaje de los agentes, con el fin de retroalimentar a cada uno de éstos y mejorar los resultados obtenidos. Por ejemplo, una de las tareas de optimización es la poda de ramas del árbol de decisión que no son significativos en el proceso de clasificación.

- *Agente de apoyo*. Se encarga de interactuar entre los usuarios y el sistema, permite el manejo de archivos de datos, transformaciones de formato, entre otros. Además, es el responsable de la interfaz que recibe las peticiones de los usuarios a través de una pantalla interactiva, para luego entregar la solicitud y posteriormente desplegar los resultados obtenidos.

Para mostrar la bondad de funcionamiento *i+DiaMAS*, se presenta el siguiente escenario: Una vez recibido un nuevo caso médico por el *Agente de apoyo*, es el *Agente coordinador* quien notifica al *Agente de preprocesamiento de datos* que tiene tareas de preproceso de datos para el nuevo caso solicitado, tales como: manejo de valores faltantes, discretización de datos, selección de variables, entre otros. Terminada la tarea, el *Agente coordinador* recoge los resultados y los reúne en una sola fuente de datos, que es entregado al *Agente de minería de datos* para el proceso de clasificación mediante el árbol de decisión.

Por último, el resultado obtenido es entregado al *Agente de optimización* que se encarga de optimizar la información para su entrega a los usuarios a través de la interfaz gráfica establecida en el *Agente de apoyo*. Tal estrategia puede realizarse en paralelo, permitiendo acelerar el funcionamiento del sistema y reducir la complejidad del problema.

En resumen, no sólo en el campo de la Salud hay avances y oportunidades de investigación, sino también en otras áreas de la actividad humana. De acuerdo a Ghedini (2009), en la actualidad se realizan estudios específicos en *Bioinformática*, donde se hacen investigaciones sobre agentes biológicos; así como en *Informática Forense*, en el que se analizan sistemas informáticos con el fin de encontrar posibles evidencias de fraude e ilegalidad.

Tendencia y apertura de campos de investigación

En la actualidad se está empleando tecnología Multiagente en nuevas aplicaciones de Minería

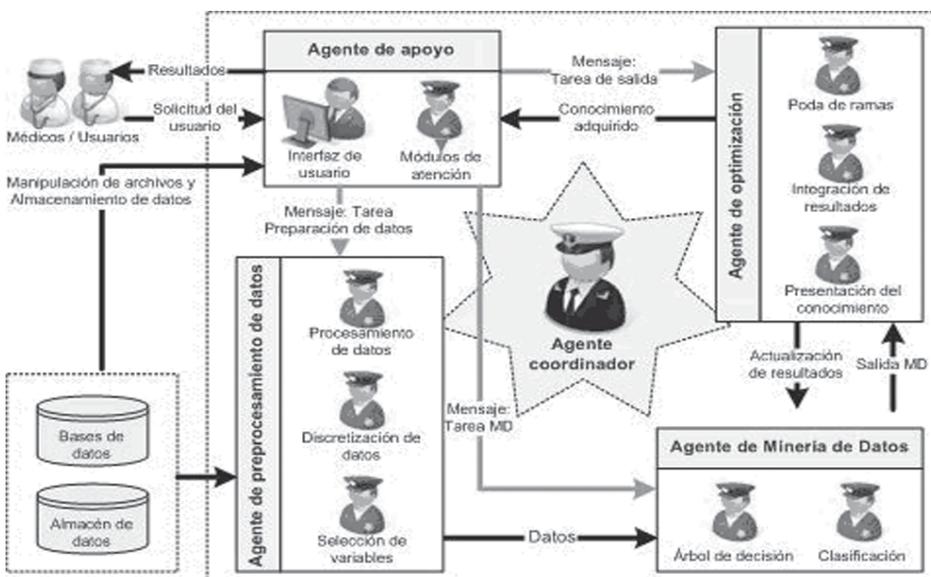


Fig. 2. Esquema general de la herramienta *i+DiaMAS*. Fuente: Adaptado de Chao y Wong (2009)

de Datos, como en Minería de Textos y Minería Web, esto debido al crecimiento exponencial de información disponible en la red (Alam *et al.*, 2009). La Minería de Textos se enfoca al descubrimiento de patrones de interés en un conjunto amplio de información textual, identificando tendencias, desviaciones y asociaciones de interés. Mientras que la Minería Web aborda el problema de descubrimiento de información presentes en la Web, esto mediante agrupaciones, clasificaciones, y predicciones de información de documentos Web. Por ejemplo, organizaciones como Google y Yahoo recogen terabytes de datos relacionados con las actividades del usuario, y se analiza para sus intereses comerciales, como la comercialización de productos, recomendación de sitios de interés, análisis del rendimiento del servidor Web, gestión del ancho de banda, entre otros.

Como un área emergente, el acoplamiento de Minería de Datos y Sistemas Multiagente tiene un futuro expansivo que puede conducir a sistemas con mayor razonamiento y más eficientes, así como, disponer de técnicas innovadoras que vayan más allá de las áreas individuales, superando así las limitaciones de éstas.

La siguiente es una lista de algunos temas de investigación abiertos, impulsada por la Minería de Datos y la tecnología Multiagente, que podrían ser de interés para su investigación y desarrollo:

- Modelado de Minería de Datos y Sistemas Multiagente.

- Arquitectura de integración de Minería de Datos y Agentes; para el acceso a múltiples fuentes de datos; para la preparación de datos; para el procesamiento distribuido; para el procesamiento paralelo; para el aprendizaje distribuido; para el aprendizaje colaborativo; para el descubrimiento de conocimiento; para el proceso y gestión de proyectos; basada en Web; basada en textos.

Es evidente que en los próximos años la integración e interacción entre Minería de Datos y Sistemas Multiagente se convierta en una destacada y desafiante área de investigación y desarrollo, permitiendo explorar nuevos desafíos. En la actualidad existe una comunidad Web dedicada a promover la investigación y desarrollo de la integración de Agentes y Minería de Datos (Moemeng *et al.*, 2009), <http://www.agentmining.org>, que proporciona recursos relacionados a temas de investigación, proyectos de investigación, actividades, talleres, conferencias, enlaces, publicaciones, grupos de investigación, entre otros.

Conclusiones

1- En particular, la integración de Minería de Datos y Sistemas Multiagente promete alcanzar el nivel de una nueva y promisoría área de investigación, que podría convertirse en un campo de conocimiento importante en la ciencia y tecnología.

2- Existe un claro esfuerzo para eliminar las fronteras entre Minería de Datos y Sistemas Multiagente, que es el acoplamiento de ambas tecnologías, impulsada por los desafíos que enfrentan ambas áreas y la necesidad de desarrollar nuevos campos de investigación.

3- La necesidad de integrar tecnología Multiagente en la Minería de Datos, básicamente es a consecuencia de la naturaleza heterogénea y distribuida de las fuentes de datos. Las cuales, por lo general, son enormes volúmenes de datos, que con frecuencia tienden a seguir creciendo.

4- En caso de los Sistemas Multiagente, estos tienen la necesidad de integrarse con la Minería de Datos debido al problema de mejorar la capacidad de aprendizaje de los agentes. Esto se logra mediante la alimentación de técnicas o algoritmos como: redes neuronales artificiales, razonamiento probabilístico, algoritmos genéticos, lógica difusa, árboles de decisión, entre otros.

5- Finalmente, queda claro que la Tecnología Multiagente y Minería de Datos han surgido como dos campos de conocimiento que se relacionan entre sí, abriendo espacio para la integración e interacción de ambas áreas. Esta nueva área de investigación tiene mucho que ofrecer en el marco de investigación y desarrollo, en particular en el contexto de cooperación y apoyo en la toma de decisiones.

Bibliografía

- Alam S., Dobbie G. y Riddle P. (2009). Exploiting Swarm Behaviour of Simple Agents for Clustering Web Users' Session Data. Ed. por Cao L. Data Mining and Multi-agent Integration, editado por Springer-Verlag, USA, pp. 61-75.
- Baazaoui H., Faiz S. y Ben H. (2005). A Framework for Data Mining Based Multi-Agent: An Application to Spatial Data. World Academy of Science, Engineering and Technology, 5, USA, pp. 22-26
- Botti V., Carrascosa C., Julián V. y Rebollo M. (2002). Sistemas Multiagente. Universidad Politécnica de Valencia. <<http://www.upv.es/sma/teoria/agentes/agentes.pdf>>. consultado: 10 de marzo de 2010
- Cao L. (2009). Introduction to Agent Mining Interaction and Integration. Ed. por Cao

L. Data Mining and Multi-agent Integration, editado por Springer-Verlag, USA, pp. 3-36.

- Chao S. y Wong F. (2009). A Multi-Agent Learning Paradigm for Medical Data Mining Diagnostic Workbench. Ed. por Cao L. Data Mining and Multi-agent Integration, editado por Springer-Verlag, USA, pp. 177-186.
- Ghedini C. (2009). Towards the Integration of Multiagent Applications and Data Mining. Ed. por Cao L. Data Mining and Multi-agent Integration, editado por Springer-Verlag, USA, pp. 37-46.
- Hand D., Mannila H. y Smyth P. (2001). Principles of Data Mining, editado por The Massachusetts Institute of Technology Press., edit. USA. pp. 546.
- Hernández J., Ramírez M. J. y Ferri C. (2004). Introducción a la Minería de Datos. Pearson Educación. Editorial Pearson Prentice Hall, España, pp. 680.
- Jennings N., Sycara K. y Wooldridge M. (1998). A Roadmap of Agent Research and Development. Autonomous Agents and Multi-Agent Systems, 1, 275-306. Kluwer Academic Publishers, Boston, USA.
- Larose D. (2005). Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley & Sons, Inc., USA, pp. 222.
- Moemeng C., Gorodetsky V., Zuo Z., Yang Y. y Zhang C. (2009). Agent-Based Distributed Data Mining: A Survey. Ed. por Cao L. Data Mining and Multi-agent Integration, editado por Springer-Verlag, USA, pp. 47-58.
- Molero G. (2008). Desarrollo de modelo de técnicas de Minería de Datos. Maestría en Ciencia e Ingeniería de la Computación, UNAM, México, pp. 163
- Sumathi S. y Sivanandam S. (2006). Introduction to Data Mining and its Applications. Studies in Computational Intelligence, 29, editado por Springer-Verlag, USA, pp. 828.

Recibido: 12 de febrero de 2010.
Aprobado en su forma definitiva:
5 de junio de 2010

Ms.C. Guillermo Molero Castillo
Universidad de Guadalajara, Zapopan, Jalisco
País: México
Correo electrónico : <g.molero@gmail.com>

Ms.C. María Elena Meda Campaña
Universidad de Guadalajara, Zapopan, Jalisco
País: México
Correo electrónico : <emed@cucea.udg.mx>
