

# Interacción humano computadora y minería de datos para la generación y representación de conocimiento útil

## *Human-computer interaction and data mining for generation and representation of useful knowledge*

Guillermo Gilberto Molero Castillo  
Edgard Iván Benítez Guerrero  
Carmen Mezura Godoy

*En este trabajo se hace una proyección de la combinación de dos áreas que ofrecen condiciones ideales para abordar nuevos desafíos con el objetivo de crear nuevos algoritmos y sistemas de interacción para el descubrimiento de conocimiento a través del análisis de datos. Datos que provienen de sensores, dispositivos móviles, imágenes, redes sociales, vídeos digitales, registros de compra, transacciones bancarias, cómputo móvil y ubicuo, entre otros. Dar sentido a estos datos sigue siendo un reto fundamental, por lo que, uno de los grandes desafíos es el análisis de estos complejos volúmenes de datos, que requieren de nuevas soluciones, eficientes y fáciles de usar para su manipulación y entendimiento. En este sentido, la combinación sinérgica de procesos, enfoques y métodos de dos áreas originalmente separadas, minería de datos e interacción humano computadora, tiene un impacto no sólo en el ámbito académico, sino en la sociedad, que podría convertirse en un campo de conocimiento importante en la ciencia y tecnología. El propósito es impulsar investigaciones para el desarrollo de nuevos algoritmos y sistemas interactivos centrados en el usuario, con el objetivo principal de mejorar la interacción humana, la representación y visualización de nuevos patrones de datos válidos y potencialmente útiles, previamente desconocidos, para dar soporte a la toma de decisiones. En este caso teniendo como punto de partida el reconocimiento de patrones mediante el análisis visual de datos.*

**Palabras clave:** *diseño centrado en el usuario, generación de conocimiento, interacción humano computadora, minería de datos, tecnología interactiva.*

### RESUMEN

### ABSTRACT

*In this work, a combination of two areas is proposed that offer ideal conditions to approach new challenges with the objective of creating new algorithms and interaction systems for the discovery of novel knowledge through data analysis. Data that comes from sensors, mobile devices, social networks, images, digital videos, purchase records, banking transactions, mobile and ubiquitous computing, among others. Making sense of these data remains a fundamental challenge, so one of the great challenges is the analysis of these complex data volumes, which require new, efficient and easy to use solutions for their manipulation and understanding. In this sense, the synergistic combination of processes, approaches, and methods of two originally separated areas, data mining and human-computer interaction, has an impact not only on academia but on society, which could be a knowledge field important in the science and technology. The purpose is to promote research for the development of new algorithms and user-centered interactive systems, with the main objective of improving human interaction, representation, and visualization of new data patterns of valid and potentially useful, as support in decision-making. In this case, the starting point is pattern recognition through visual data analysis.*

**Keywords:** *data mining, knowledge generation, human-computer interaction, interactive technology, user-centered design.*

## Introducción

El continuo crecimiento exponencial de la recolección y almacenamiento de datos y la creciente disponibilidad de información en Internet, junto con la evolución del poder de cómputo, hacen que se considere a la minería de datos (Data Mining o DM, por sus siglas en inglés) como una disciplina de apoyo para el análisis inteligente de datos (Graham, 2011; Larose, 2014), que tiene como propósito resolver dos grandes retos (Molero, 2014): a) trabajar con conjuntos de datos para extraer y descubrir información de interés y b) usar técnicas adecuadas para analizar e identificar tendencias y comportamientos que faciliten una mejor comprensión de los fenómenos que ocurren en el entorno y sirvan de ayuda en el proceso de la toma de decisiones.

Una realidad actual de la DM es su empleo en diversos campos de aplicación, como (Hand *et al.*, 2001; Sumathi y Sivanandam, 2006; Graham, 2011): educación, seguridad, análisis de mercado, procesos industriales, biología, química, medicina, biometría, telecomunicaciones, interacción humano computadora, entre otros. Es precisamente en la interacción humano computadora (*Human Computer Interaction* o HCI, por sus siglas en inglés) donde a diario se generan miles y miles bytes de información, que superan terabytes e incluso petabytes. Información que proviene de sensores, dispositivos móviles, redes sociales, imágenes, vídeos digitales, registros de compra, transacciones bancarias, señales GPS (sistema de posicionamiento global), cómputo móvil y ubicuo, entre otros. A toda esta información se conoce en la actualidad como grandes datos (Molero, 2014), cuyo desafío a partir de esta fuente masiva de datos es el reconocimiento de patrones que son obtenidos a partir de procesos de agrupamientos, reglas de asociación, correlaciones, clasificaciones, regresiones y otros, donde cada patrón queda asignado a una categoría o clase que representa a un conjunto de elementos que comparten características similares.

Sin embargo, dar sentido a estos datos sigue siendo un reto fundamental, por lo que, uno de los grandes desafíos es el análisis de estos complejos volúmenes de datos, que requieren de nuevas soluciones, eficientes y fáciles de usar para su manipulación y entendimiento. En consecuencia, con las crecientes

expectativas de los usuarios finales, los enfoques tradicionales de DM a menudo carecen de una buena interpretación visual por lo que existe el riesgo de entregar resultados poco satisfactorios. Tal es así que para hacer frente a esta creciente avalancha de datos son útiles nuevos enfoques basados en el diseño centrado en el usuario.

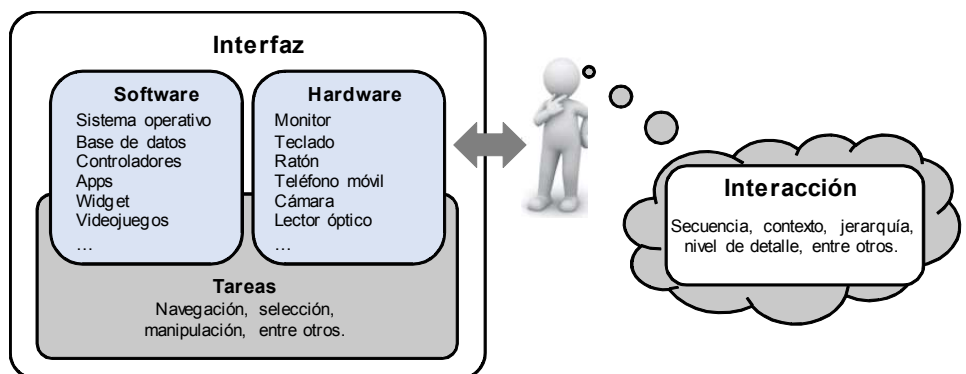
El diseño centrado en el usuario es ampliamente usado en distintas áreas. Algunos ejemplos de estos son: el diseño de equipos médicos, la creación de vehículos, el diseño de dispositivos móviles, la creación de productos de consumo, entre otros. Recientemente este diseño centrado en el usuario se está proyectando también en la minería de datos. Precisamente, uno de los trabajos recientes fue realizado por Horberry *et al.* (2015), en el cual se describe la colaboración que tiene el diseño centrado en el usuario en la minería de datos, donde los estudios realizados reflejaron que el uso del diseño centrado en el usuario puede ser un proceso exitoso y eficaz en la minería de datos, siendo la característica clave el enfoque incesante en los usuarios y sus tareas a lo largo del proceso de diseño iterativo.

En este sentido, realizar investigaciones en pro de permitir la interacción humana en el reconocimiento de patrones válidos, novedosos y útiles es crucial, pero resulta contraproducente desarrollar modelos eficientes cuando se ve reducida la participación del usuario final. Por lo que, una combinación sinérgica de procesos y enfoques de dos áreas originalmente separadas, como la interacción humano computadora y la minería de datos (HCI-DM), ofrece condiciones ideales para

el desarrollo de nuevos algoritmos y herramientas eficientes centradas en el usuario, con el objetivo principal de mejorar la interacción humana, la representación y visualización de nuevos patrones de datos válidos y potencialmente útiles, previamente desconocidos, para dar soporte a la toma de decisiones. Este trabajo proporciona una proyección práctica sobre la HCI-DM con el propósito de motivar y estimular nuevas investigaciones de esta promisoría área de investigación que tiene un futuro expansivo que puede dar lugar a sistemas de interacción con mayor razonamiento y más eficientes, así como técnicas innovadoras que vayan más allá de las áreas individuales.

## Interacción humano computadora

La HCI es un área transversal entre la ingeniería, psicología, ergonomía y diseño, que investiga la teoría, diseño, implementación y evaluación de la forma en que los seres humanos utilizan e interactúan con los dispositivos tecnológicos (Kim, 2015; Reeves, 2015). Se investiga además la forma en que la tecnología influye en el trabajo y las actividades humanas (Dix, 2009), así como el diseño de nuevas tecnologías que permitan la interacción de los usuarios y los dispositivos digitales (Grudin, 1992; Dix, 2009; Kim, 2015). El concepto de interacción se distingue de otro término similar, interfaz. En términos generales, la interacción se refiere a un modelo abstracto por el cual los seres humanos interactúan con el dispositivo para realizar una determinada tarea, mientras que una interfaz es el desarrollo de hardware o software para llevar a cabo la interacción (Figura 1).



**Figura 1.** Conceptos del modelo de interacción e interfaz de usuario

**Fuente:** Adaptado de Kim (2015).

En la actualidad, la tecnología interactiva (software y hardware), ya sean teléfonos móviles, computadoras, electrodomésticos, sistemas de navegación, sensores, entre otros, forman parte de la vida diaria, las cuales se utilizan para conectarse con amigos y familia, crear y gestionar música personal y álbumes de fotos, explorar nuevos medios de comunicación, realizar tareas a través de interfaces, entre otras actividades. Como resultado, la investigación actual se centra cada vez más en la creación de experiencias de usuario en lugar de interfaces. Este cambio de enfoque implica un salto en la conceptualización, el uso nuevas metodologías y las mejores prácticas de la HCI, como el diseño de interacción y el diseño centrado en el usuario, que se centran en la forma de diseñar la tecnología interactiva para que sea fácil de entender y agradable de usar (Dix, 2009).

Un aspecto clave en el diseño es el concepto de usabilidad, que por lo general se define en términos de eficiencia, eficacia y satisfacción (Chalmers y Galani, 2004; Barkhuus y Polichar, 2011). Además, el concepto de facilidad de uso significa que los productos resultantes sean eficaces para la realización de una determinada tarea, garanticen la seguridad y conduzcan a una conclusión correcta de la tarea. Por tanto, usabilidad e interacción eficiente se traducen en una mayor productividad.

En este sentido, el enfoque actual de la HCI, con énfasis en el diseño, va más allá de utilizar la tecnología para resolver un problema, sino se trata de crear una experiencia tecnológica funcional, intuitiva e incluso agradable para las personas que lo utilizan. El objetivo es diseñar sistemas eficientes, sencillos y amenos a la hora de utilizarlos, ya sea como herramientas interactivas en forma de aplicaciones, interfaces, productos digitales, sitios de redes sociales, entre otros, que mejoran la usabilidad, aumentan el aprendizaje, facilitan la comunicación y enriquecen la vida de los usuarios.

En este contexto, el concepto de experiencia de usuario (User eXpert o UX, por sus siglas en inglés) no solo abarca la integridad funcional y la alta usabilidad y estética de una aplicación interactiva, sino también su integración en el estilo de vida e incluso en la creación de uno nuevo (Kim, 2015), por ejemplo, la invención, o redescubrimiento,

del ratón es una pieza clave en la revolución del uso de la computadora personal, permitiendo un manejo fácil e intuitivo comparado con el anterior sistema de comandos a través del teclado. Otro claro ejemplo es la evolución de las interfaces en la historia de la computación, es decir, de terminal y teclado se pasó a interfaces gráficas de usuario controladas por el ratón, y actualmente se utilizan también interfaces basadas en el contacto, esto es, orientadas al tacto, como los teléfonos inteligentes.

Por tanto, debido a esta naturaleza multidisciplinaria de la HCI, los usuarios contribuyen en su éxito. Por el lado humano son relevantes la comunicación, el diseño gráfico, la lingüística, las ciencias sociales, la psicología y los factores humanos, tales como la satisfacción del usuario al hacer uso de la computadora. Por el lado de los dispositivos, las técnicas de gráficos por computadora, sistemas operativos, lenguajes de programación y entornos de desarrollo son relevantes. Además, son relevantes los métodos de ingeniería y diseño (Hewett *et al.*, 2015).

Una de las razones para el aumento del diseño en la HCI es la necesidad de la comprensión de las necesidades humanas (Holzinger, 2002; Ebert *et al.*, 2012), que se asocia con aspectos de percepción, cognición, inteligencia e interacción con cualquier tipo de información en cualquier tipo de dispositivo (Hooper y Dix, 2013). Además, en la actualidad las fuentes de datos proporcionan nuevas oportunidades de estudio de diferentes fenómenos, como es en el caso de la comprensión de la interacción de las personas, por ejemplo, las redes sociales de Facebook y Twitter; las tendencias de negocios, por ejemplo, clientes de Amazon y Netflix; y para prevenir enfermedades, por ejemplo, interacción de las redes de proteínas.

## Minería de datos

La minería de datos es un paso esencial de un proceso más amplio cuyo objetivo es el descubrimiento de conocimiento (Knowledge Discovery o KD, por sus siglas en inglés). El KD es el proceso de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia comprensibles a partir de los datos (Fayyad *et al.*, 1996; Klösgen y Zytkow, 2002). Para Vazirgiannis *et al.* (2003) el KD es un proceso general de descubrir conocimiento dividido en una secuencia de pasos, donde la DM es considerada como uno de los pasos más importantes en todo el proceso (Figura 2), esto es, el corazón del KD mismo (Wong y Leung, 2002).

El proceso de KD recurre a dos aspectos fundamentales: la iteración y la interacción, dado que estas acciones permiten iniciar o retomar algunos de los pasos anteriores para obtener un conocimiento de calidad. En términos generales, las fases del KD son:

- integración y recopilación de datos, donde se definen las fuentes de datos disponibles y se recopilan volúmenes de éstos, ya sean de registros, eventos, bases de datos, u otras fuentes
- selección, limpieza y transformación, en la que se eligen las variables de estudio, se eliminan redundancias en los datos y, en caso de requerirse, se transforman para obtener una vista de datos minable
- minería de datos, en la cual se elige uno o más algoritmos, de acuerdo a los objetivos del proyecto, para llevar a cabo el proceso de adquisición de patrones de datos
- evaluación e interpretación, una vez aplicado el algoritmo, los resultados de

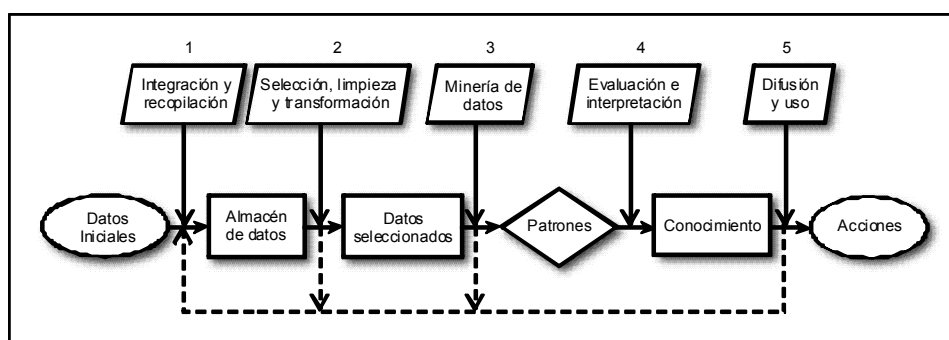


Figura 2: Proceso de descubrimiento de conocimiento a través de minería de datos

éste son evaluados e interpretados con base en mediciones que garanticen que los patrones de datos extraídos sean realmente significativos

- e) difusión y uso del conocimiento, mediante la cual el nuevo conocimiento adquirido se incorpora en algún procedimiento para su difusión y uso por los usuarios finales

En la actualidad, la DM presenta variadas definiciones. Hand *et al.* (2001) la consideran como el análisis de conjuntos de datos observados, a menudo extensos, para encontrar relaciones insospechadas y resumir los datos en forma comprensible y útil para el usuario de la información. Por su parte Berry y Linoff (2004) la define como la exploración y análisis de grandes cantidades de datos para descubrir reglas y patrones significativos; mientras que para Graham (2011) es el arte y la ciencia del análisis inteligente de datos, cuyo objetivo es descubrir conocimiento de interés a partir de conjuntos de datos.

Estas definiciones tienen en común la extracción de información potencialmente útil (patrones, asociaciones, tendencias o relaciones entre los datos) para los usuarios finales. Otras definiciones consideran además del proceso de descubrir patrones válidos, la exploración y análisis de series extensas de datos. Debido a las diversas formas de definir la DM, se considera que el objetivo de este campo de conocimiento es el proceso de convertir conjuntos de datos en información útil, novedosa, comprensible y explotable. En general, el impulso al empleo de DM se asocia a una serie de factores (Graham, 2011; Larose, 2014; Molero, 2014):

- a) la evolución del poder de cómputo
- b) el crecimiento exponencial de la recolección de datos
- c) el almacenamiento de información en una base de datos actualizada y confiable
- d) la creciente disponibilidad de información en Internet
- e) la presión competitiva del mercado en una economía globalizada
- f) el desarrollo de herramientas libres y comerciales para llevar a cabo la DM.

Estos factores han llevado a considerar a la DM como un campo interdisciplinario, que puede aplicarse en distintas áreas de conocimiento, como (Hand *et al.* 2001; Sumathi y Sivanandam, 2006; Graham, 2011): finanzas, seguridad, análisis de mercado, procesos industriales, medicina, biología, química, telecomunicaciones, análisis científico, biometría, análisis ambiental, análisis hidrológicos, cómputo móvil y ubicuo, interacción humano computadora, entre otras.

La DM al ser un proceso para extraer conocimiento útil mediante el análisis de los datos, ésta recurre a modelos y métodos que permiten encontrar relaciones, patrones o reglas inferidas previamente desconocidas. Los modelos utilizados son el descriptivo y el predictivo. Los modelos descriptivos permiten identificar patrones que se utilizan para resumir y explicar el comportamiento de los datos; mientras que los modelos predictivos permiten estimar valores futuros de alguna variable de interés con base en datos históricos. A su vez cada uno de estos modelos cuentan con un amplio número de métodos, por ejemplo, algunos métodos clásicos no supervisados, que consisten en adquirir conocimiento a partir de los datos sin requerir influencia externa, son (Hand *et al.* 2001; Berry y Linoff (2004); Hernández *et al.*, 2004; Larose, 2014): algoritmos de agrupamiento (K-medias, Bietápico, Cobweb, redes de Kohonen y otros), medidas de distancias, funciones de similitud, reglas de asociación, análisis de componentes principales, entre otros. Mientras que algunos métodos clásicos supervisados, que consisten en aprender mediante el control de un supervisor que determina la respuesta que se desea generar del sistema, son (Hand *et al.* 2001; Berry y Linoff 2004; Hernández *et al.*, 2004; Larose, 2014): árboles de decisión (ID3, C4.5, CART, CHAID, QUEST y otros), redes bayesianas, regresión (lineal logarítmica y logística), máquinas de soporte vectorial, reglas de inducción (algoritmo LEMI), entre otros. Otros métodos avanzados son algoritmos evolutivos, redes neuronales artificiales, lógica difusa y entropía gráfica.

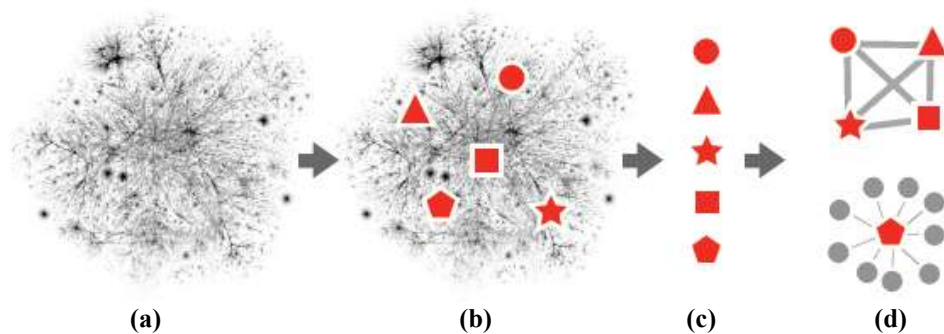
## Proyección práctica de la HCI y DM

La combinación de tecnologías es un avance natural en el desarrollo de la ciencia. Esta combinación puede trascender en

una nueva tecnología cuyo poder puede acarrear potenciales beneficios de solución. De manera tal, la combinación natural de la HCI y DM es a consecuencia de la necesidad de desarrollar nuevos algoritmos y herramientas eficientes centradas en el usuario, con el objetivo principal de mejorar la interacción humana y la representación y visualización de nuevos patrones de datos válidos, potencialmente útiles, para dar soporte a la toma de decisiones.

Históricamente, a causa del constante avance tecnológico, ambas disciplinas han permitido el desarrollo de métodos para extraer información útil de los datos; sin embargo, éstas se han desarrollado de manera separada. Por ejemplo, la HCI se centra en métodos de interacción y visualización que son aprovechados por usuarios finales y la DM se centra en métodos escalables y automáticos. Por lo que, como una destacada área de investigación, la combinación de HCI y DM tiene un futuro expansivo que puede dar lugar a métodos y aplicaciones innovadoras que vayan más allá de ambas áreas. Para efectos prácticos, se presenta de manera gráfica la superposición de nodos (clientes) y aristas (llamadas) de un número amplio de clientes de una empresa de telecomunicaciones (Figura 3a). El objetivo es detectar anomalías en las llamadas telefónicas que realizan sus clientes. Para esto se emplea métodos de DM para analizar el comportamiento de los clientes cuyas llamadas sean diferentes al resto (Figura 3b). Ante esto, los métodos de detección de anomalías identifican a los clientes con comportamiento distinto, pero sin explicar el por qué (Figura 3c), puesto que los algoritmos son como cajas negras que dicen «qué» pero no «por qué», es decir, ¿Por qué esos clientes presentan un comportamiento distinto?

Una forma de explicar ese comportamiento es a través de una representación visual de los patrones que son representativos pero que con frecuencia carecen de una explicación. Por lo que, al hacer uso de una interfaz gráfica para la visualización de los patrones, ésta sirve de apoyo para explicar el comportamiento y conexiones de los nodos detectados (Figura 3d). Por ejemplo, para este caso ilustrativo se revela que los primeros cuatro nodos (patrones) forman un grupo en forma de grafo completo, cuyos elementos, por el grosor de las aristas, presentan una conexión fuerte. Esto representa que estos cuatro clientes han



**Figura 3:** Combinación de HCI y DM para la detección de anomalías en llamadas telefónicas

realizado reiteradas llamadas uno a otro, evidenciándose con esto una posible anomalía, puesto que no se ajusta a un comportamiento normal. Mientras que al expandir las conexiones del quinto y último patrón se revela que éste es el centro de un grafo tipo estrella. Esto indica que el cliente podría ser un agente de ventas que ha estado ocupado realizando un amplio número de llamadas.

En el ejemplo, la DM es la que ayuda en la identificación de patrones y la HCI apoya en una mejor visualización, interacción y explicación de los patrones obtenidos. Por lo que, combinando lo mejor de ambas disciplinas se pueden lograr nuevos algoritmos escalables y sistemas que ayuden a las personas a entender e interactuar con los patrones adquiridos a través de una exploración gráfica de la información.

Un análisis visual puede aportar nuevos conocimientos sobre los patrones ocultos en los datos, que es exactamente el objetivo del descubrimiento de conocimiento. Los grafos son intuitivamente más informativos que las palabras o frases. Por lo que, en los últimos años la teoría de grafos es una de las líneas de investigación más estudiadas en la DM (Chaiken *et al.*, 2008). La teoría de grafos proporciona herramientas para mapear las estructuras de datos y para encontrar nuevas conexiones (relaciones) entre los objetos o nodos (Chakrabarti *et al.*, 2004). Mediante estos grafos se pueden representar y dar sentido a diversos problemas en los dominios que van desde la bioinformática (estructuras biológicas) hasta las redes de citas de literatura científica, científica, redes sociales, vínculos entre las páginas Web, redes de información personal, entre otros. Esta riqueza en la representación de información los convierte en una herramienta crucial para los métodos

de DM y la extracción de conocimiento (Russell *et al.*, 1993).

En este sentido, es importante impulsar las investigaciones sobre algoritmos y herramientas de DM basadas en grafos o minería de grafos (Graphs Mining o GM, por sus siglas en inglés), con la cual se puede formular y resolver problemas reales de la vida diaria. Sin embargo, un componente crítico en la representación y despliegue de la información es la interfaz de usuario, es decir, la «ventana» a través de la cual el usuario interactúa con el algoritmo y con los patrones obtenidos. Por lo que, la propuesta de una interfaz sencilla e intuitiva de visualización podría ayudar a los usuarios a entender los resultados que el sistema produce, por ejemplo, una interfaz de usuario que muestre el comportamiento de posibles fraudes con tarjetas de crédito. Un esquema básico para la combinación de la HCI y DM se muestra en la Figura 4, donde mediante un enfoque híbrido se integran tres componentes principales que permiten la interactividad para el resumen, detección de anomalías y visualización de patrones.

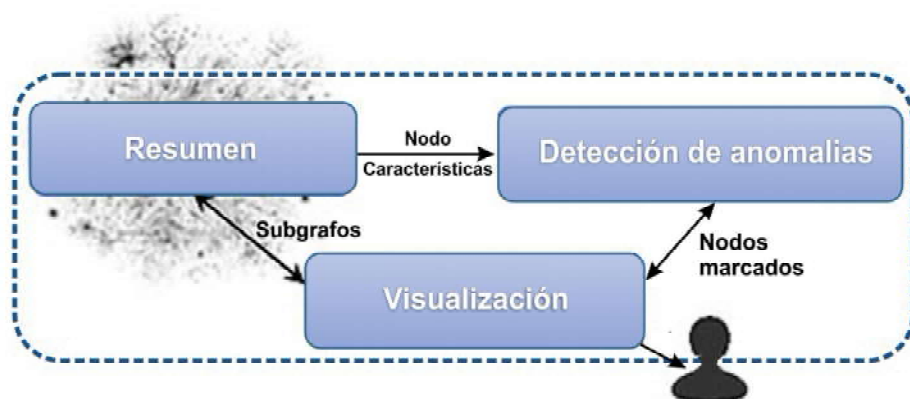
El resumen proporciona los algoritmos escalables de DM basada en grafos para

desplegar y calcular estadísticas de los nodos del grafo y sus relaciones, así como el grado de su distribución. La detección de anomalías provee métodos para la detección automática de patrones y detectar anomalías en el grafo, tales como nodos con muchos contactos, pero pocas interacciones entre éstos, como es el caso de agentes de ventas. Estos nodos identificados son puntos de partida para el análisis, pues representan información potencialmente nueva y significativa. La visualización proporciona a los usuarios un entorno interactivo para visualizar y explorar los nodos del grafo. Es importante resaltar que para analizar los patrones identificados (puntos de partida) se necesita utilizar un enfoque mixto de visualización, interacción con el usuario y algoritmos de aprendizaje, esto con el propósito de guiar al usuario a explorar de forma incremental e interactiva los nodos y subgrafos de la red para darle sentido a los datos. Esta tarea implica una exploración detallada y la formulación de una o varias hipótesis.

Una aproximación de una interfaz de usuario para el reconocimiento de patrones a través de la minería de grafos es:

- un panel de configuración para el usuario con el objetivo de mejorar la legibilidad de la visualización
- un panel de filtrado para controlar los tipos de nodos a mostrar
- un panel de grupo para permitir al usuario crear, renombrar y borrar grupos (subgrafos)

En la Figura 5 podemos observar una aproximación de una interfaz de usuario para el reconocimiento de patrones a través de la minería de grafos, se presenta:



**Figura 4:** Componentes básicos para la interactividad entre la HCI y DM

- a) los operadores que contienen funciones para cargar la fuente de datos
- b) un panel de configuración para el diseño de la minería de datos con el objetivo de mejorar la legibilidad de la visualización
- c) un panel de filtrado para controlar los tipos de nodos a mostrar
- d) un panel de grupo para permitir al usuario crear, renombrar y borrar grupos (subgrafos) a través de una interfaz interactiva

En general, la secuencia de los operadores, en el panel de diseño, permite hacer la configuración de la minería de datos basada en grafos, teniendo para esto funciones para habilitar y deshabilitar los operadores, con el propósito de mejorar la usabilidad y la experiencia del usuario. Es importante destacar que a medida que se necesitan incluir nuevos operadores en la herramienta, el grafo permite anexar nuevos nodos haciendo que el software sea escalable, donde cada nodo representa una salida.

## Algoritmos y herramientas

Uno de los algoritmos más utilizados para modelos de DM basadas en grafos es la propagación de creencias (Belief Propagation o BP, por sus siglas en inglés). Originalmente este algoritmo fue propuesto para el cálculo de las distribuciones marginales

exactas en los árboles (Yedidia *et al.*, 2003), posteriormente se aplicó en grafos como un algoritmo de aproximación, basado en razonamiento bayesiano, para inferir las probabilidades en los nodos mediante una matriz de propagación (Pearl, 2014; Hadoop, 2015). En la actualidad, existen investigaciones sobre BP en paralelo para múltiples núcleos de memoria compartida (González *et al.*, 2009). Sin embargo, en estos trabajos se asume que los grafos encajan en una sola memoria principal, ya sea en un equipo de cómputo o cluster, pero en realidad se necesitan hacer más investigaciones para abordar esta problemática, puesto que los grafos grandes no caben en una sola memoria principal. Otros algoritmos de la minería de grafos, conocidos también como algoritmos conocidos también como algoritmos basados en la propagación, son: HITS (Kleinberg, 1999) y PageRank (Brin y Page, 2012).

Una herramienta para la minería de grafos es MapReduce, que es un marco de programación en paralelo para el procesamiento de datos a escala Web. Esta herramienta tiene dos ventajas (Dean y Ghemawat, 2008; Lämmel, 2008):

- a) la distribución de los datos, la tolerancia a fallos y el equilibrio de carga de forma automática
- b) la utilización del concepto de programación funcional. Por otro lado, Hadoop es la

versión de código abierto de MapReduce (Hadoop, 2015), que es una herramienta para realizar tareas de minería de grafos, que utiliza su propio sistema de archivos distribuidos (Hadoop Distributed File System o HDFS, por sus siglas en inglés), y proporciona un lenguaje de alto nivel llamado Pig (Olston *et al.*, 2008). Debido a su excelente escalabilidad y facilidad de uso, Hadoop es ampliamente utilizado para la DM a gran escala (Kang *et al.*, 2010). Otras variantes relacionados al sistema MapReduce son (Pike *et al.*, 2005; Chaiken *et al.*, 2008; Grossman y Gu, 2008): SCOPE, Sphere y Sawzall.

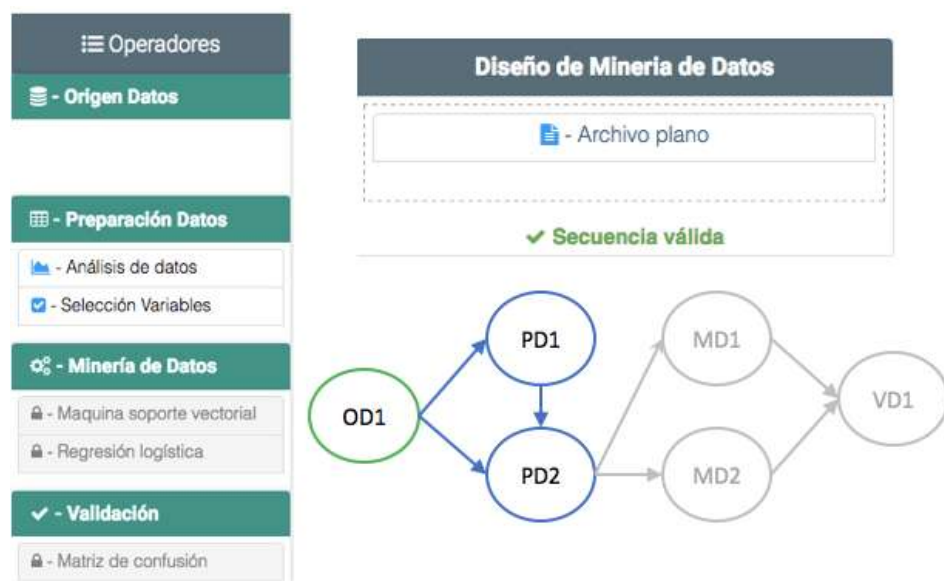
## Visualización y exploración gráfica

La visualización de datos está relacionada al despliegue de gráficas en representaciones de dos y tres dimensiones, tomando en cuenta aspectos de ampliación, enfoque, vistas de contexto y agrupamiento. Es importante señalar que estas gráficas son del orden de cientos y miles de nodos. Algunos sistemas y bibliotecas de software para la visualización son: Graphviz, Walrus, Otter, Prefuse, entre otros; pero éstos sólo realizan despliegue gráfico, sin ninguna funcionalidad de detección de valores atípicos.

Por su parte, para la exploración existe una serie de herramientas que han sido desarrolladas para apoyar en la vista rápida de la información. Estos incluyen WebBook y Web-Forager (Card *et al.*, 1996), que utilizan una comparación para encontrar y gestionar páginas Web. Butterfly (Mackinlay *et al.*, 1995) destinada a acceder a artículos en redes de citación; y Webcutter (Maarek *et al.*, 1997), que colecta y presenta colecciones de URL en un árbol. Algunas investigaciones estudian la forma de apoyar la exploración local de los gráficos. Estos enfoques generalmente apoyan la idea de iniciar con un pequeño subgrafo y expandir los nodos para mostrar sus vecinos y ayudar a identificar vecindarios útiles para ampliarlos.

## Conclusiones

Las tecnologías interactivas implican mucho más que cambios de funciones (widgets) en las interfaces de usuario. Las aplicaciones tecnológicas, por



**Figura 5:** Interfaz de usuario para el reconocimiento de patrones a través de la minería de grafos

ejemplo, dispositivos móviles, tabletas y electrodomésticos, y las prácticas de diseño, por ejemplo, investigación, creación de prototipos y evaluación, son igualmente importantes en la cultura a la interacción.

El punto de vista o la experiencia del usuario tiene que ver con la forma en que los usuarios de la tecnología interactúan con ésta y, por lo tanto, su diseño es aún más crítico, puesto que todo se convierte en digital incrustados en servicios interactivos que hacen de la vida cotidiana más emocionante, eficiente y conveniente.

La necesidad de combinar la HCI y DM es básicamente a consecuencia de la avalancha de datos que surgen en la ciencia, gobierno y empresas; las cuales, por lo general, son enormes volúmenes, que con frecuencia tienden a seguir creciendo. Por lo que, se necesita desarrollar nuevos algoritmos eficientes e interfaces de apoyo que permitan a los usuarios entender y darle sentido a esta información para tomar mejores decisiones con base en el análisis visual de patrones.

Encontrar patrones de datos no es suficiente, gran parte del trabajo es entender por qué suceden ciertos fenómenos. Esto representa un problema serio que necesita ser resuelto. Para el caso de la HCI es crucial el diseño de nuevos métodos de visualización de datos, con el propósito de ofrecer al usuario nuevas formas viables de identificación y explicación de patrones obtenidos. Mientras que para la DM es importante impulsar investigaciones sobre minería de datos basada en grafos, con la cual se pueden formular y resolver problemas reales de la vida diaria a través de la representación de grafos.

La asociación de HCI y DM tiene un impacto no sólo en el ámbito académico, sino en la sociedad. El propósito es impulsar investigaciones para el desarrollo de nuevos algoritmos y sistemas interactivos centrados en el usuario. En este caso teniendo como punto de partida el reconocimiento de patrones mediante el análisis visual de datos.

En particular, la combinación de HCI y DM promete alcanzar el nivel de una promisorio área de investigación, que podría convertirse en un campo de conocimiento importante en la ciencia y tecnología. Actualmente existe un claro esfuerzo para

eliminar las fronteras entre la HCI y DM, que es la combinación de ambos dominios, impulsada por los desafíos que enfrenan ambas áreas y la necesidad de desarrollar nuevos campos de investigación.

Finalmente, en este trabajo se establece como punto de referencia la combinación natural de dos áreas que ofrecen condiciones ideales para abordar nuevos desafíos con el objetivo de crear nuevos algoritmos y sistemas de interacción inteligentes para el descubrimiento de conocimiento novedoso a través del análisis de datos.

## Agradecimiento

Este trabajo forma parte del proyecto «Infraestructura para agilizar el desarrollo de sistemas centrados en el usuario» financiado por el Consejo Nacional de Ciencia y Tecnología (CONACYT) en el marco del programa de Cátedras CONACYT (Ref. 3053).

## Referencias bibliográficas

- Barkhuus L. y Polichar V. (2011). Empowerment through seamfulness: smart phones in everyday life. *Personal and Ubiquitous Computing*, 15(6), 629-639.
- Berry M. y Linoff G. (2004). *Data Mining Techniques: for marketing, sales, and customer relationship management*. 2da Ed. Wiley Publishing, Inc., pp. 643, ISBN: 0-471-47064-3, Indiana, Estados Unidos.
- Brin S. y Page L. (2012). The anatomy of a large-scale hypertextual Web search engine. *Computer networks*, 56(18), 3825-3833.
- Card S., Robertson G. y York W. (1996). The WebBook and the Web Forager: an information workspace for the World-Wide Web. In *SIGCHI*, ACM, pp. 111.
- Chaiken R., Jenkins B., Larson P., Ramsey B., Shakib D., Weaver S. y Zhou J. (2008). SCOPE: easy and efficient parallel processing of massive data sets. *VLDB*, 1(2), 1265-1276.
- Chakrabarti D., Papadimitriou S., Modha

D. y Faloutsos C. (2004). Fully automatic cross-associations. In *ACM SIGKDD*, pp. 79-88.

- Chalmers M. y Galani A. (2004). Seamless interweaving: heterogeneity in the theory and design of interactive systems. In *Proceedings of the CDIS*. ACM, pp. 243-252.
- Dean J. y Ghemawat S. (2008). MapReduce: simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107-113.
- Dix A. (2009). Human-computer interaction. *Encyclopedia of Database Systems*. Ed. Springer, pp. 1327-1331, ISBN: 978-0-387-35544-3.
- Ebert A., Gershon N. y van der Veer G. (2012). Human-Computer Interaction. *KI - Künstliche Intelligenz*, 26(2), 121-126.
- Fayyad U., Shapiro G. y Smyth P. (1996). From Data Mining to Knowledge Discovery in Databases American Association for Artificial Intelligence. 17, 3, 37-54.
- González J., Low Y. y Guestrin C. (2009). Residual splash for optimally parallelizing belief propagation. *Conference on Artificial Intelligence and Statistics*, pp. 177-184.
- Graham W. (2011). *Data Mining with Rattle and R. The Art of Excavating Data for Knowledge Discovery*. Ed. Springer Science Business Media, pp. 374, ISBN 978 1 4419 9889 7, New York, Estados Unidos.
- Grossman R. y Gu Y. (2008). Data mining using high performance data clouds: experimental studies using sector and sphere. In *ACM SIGKDD*, pp. 920-927.
- Grudin J. (1992). Utility and usability: research issues and development contexts. *Interacting with computers*, 4(2), 209-217, DOI:10.1016/0953-5438(92)90005-z.
- Hadoop (2015). What Is Apache Hadoop? Hadoop information. [<http://hadoop.apache.org>], última consulta 28.11.2016.

- Kang U., McGlohon M., Akoglu L. y Faloutsos C. (2010). Patterns on the connected components of terabyte-scale graphs. *IEEE International Conference on DM*, pp. 875-880.
- Hand D., Mannila H. y Smyth P. (2001). *Principles of Data Mining*. Ed. The Massachusetts Institute of Technology Press, pp. 546, ISBN: 0-262-08290-X, Massachusetts, Estados Unidos.
- Hernández J., Ramírez M. y Ferri C. (2004). *Introducción a la Minería de Datos*. Pearson Educación. Ed. Pearson Prentice Hall, pp. 680, ISBN: 84-205-4091-9, Madrid, España.
- Hewett T., Baecker R., Card S., Carey T., Gasen J., Mantel M., Perlman G., Strong G. y Verplank W. (2015). *ACM SIGCHI Curricula for Human-Computer Interaction*, pp. 162, ISBN: 0-89791-474-0, New York, Estados Unidos.
- Holzinger A. (2002). *Multimedia Basics Learning. Cognitive Basics of Multimedia Information Systems*, 2. Laxmi-Publications, New Delhi.
- Hooper C. y Dix A. (2013). *Web science and human-computer interaction: forming a mutually supportive relationship*. *Interactions*, 20(3), 52-57.
- Horberry T., Burgess-Limerick R. y Steiner L. (2015). *Human Centred Design for Mining Equipment and New Technology*. In *Proceedings 19th Triennial Congress of the IEA*, 9, pp. 1-6.
- Kim G. (2015). *Human-Computer Interaction Fundamentals and Practice*. Ed. CRC Press, pp. 184, ISBN: 978-1-4822-3389-6, New York, Estados Unidos.
- Kleinberg J. (1999). *Authoritative sources in a hyperlinked environment*. *Journal of the ACM*, 46(5), 604-632.
- Klösgen W. y Zytlow J. (2002). *Handbook of Data Mining and Knowledge Discovery*. Oxford University Press, Inc., pp. 980, ISBN: 0 19 511831 6, New York, Estados Unidos.
- Lämmel R. (2008). *Google's MapReduce programming model-revisited*. *Science of computer programming*, 70(1), 1-30.
- Larose D. (2014). *Discovering Knowledge in Data: An Introduction to Data Mining*. 2da Ed. John Wiley & Sons, Inc., pp. 336, ISBN: 978-0-470-90874-7, New Jersey, Estados Unidos.
- Maarek Y., Jacovi M., Shtalham M., Ur S., Zernik D. y Ben-Shaul I. (1997). *WebCutter: a system for dynamic and tailorable site mapping*. *Computer networks*, 29(8), 1269-1279.
- Mackinlay J., Rao R. y Card S. (1995). *An organic user interface for searching citation links*. In *SIGCHI*, ACM, pp. 67-73.
- Molero G. (2014). *Clasificador bayesiano para el pronóstico de la supervivencia y mortalidad de casos de cáncer de mama en mujeres de origen hispano*. Universidad de Guadalajara, tesis doctoral, pp. 155, México.
- Olston C., Reed B., Srivastava U., Kumar R. y Tomkins A. (2008). *Pig latin: a not-so-foreign language for data processing*. In *ACM SIGMOD*, pp. 1099-1110.
- Pearl J. (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.
- Pike R., Dorward S., Griesemer R. y Quinlan S. (2005). *Interpreting the data: Parallel analysis with Sawzall*. *Scientific Programming*, 13(4), 277-298.
- Reeves S. (2015) *Human-computer interaction as science*. In *Critical Alternatives*, 5th Decennial Aarhus Conference, pp. 1-12, Aarhus, Dinamarca.
- Russell D., Stefik M., Pirulli P. y Card S. (1993). *The cost structure of sensemaking*. In *Proceedings of the INTERACT3 and CHI*, Ed. ACM, pp. 269-276.
- Sumathi S. y Sivanandam S. (2006). *Introduction to Data Mining and its Applications*. Studies in Computational Intelligence, 29. Ed. Springer Verlag, pp. 828, ISBN: 3-540-34350-4, Alemania.
- Vazirgiannis M., Halkini M. y Gunopulos D. (2003). *Uncertainty Handling and Quality Assessment in Data Mining*. AIKP. Ed. Springer, pp. 226, ISBN: 1-85233-655-2, Heidelberg, Alemania.
- Wong M. y Leung K. (2002). *Data Mining Using Grammar Based Genetic Programming And Applications*. Genetic Programming Series. Ed. Kluwer Academic Publishers, pp. 213, ISBN: 0-792-37746-X, Estados Unidos.
- Yedidia J., Freeman W. y Weiss Y. (2003). *Understanding belief propagation and its generalizations*. *Exploring artificial intelligence in the new millennium*, 8, 236-239.

Recibido: 2 de febrero de 2017  
Aprobado en su forma definitiva:  
2 de marzo de 2017

---

**Guillermo Gilberto Molero Castillo**  
Universidad Veracruzana. Xalapa, Veracruz,  
CONACYT. México  
Correo-e.: ggmolero@conacyt.mx

**Edgard Iván Benítez Guerrero**  
Universidad Veracruzana. Xalapa, Veracruz,  
CONACYT. México  
Correo-e.: edbenitez@uv.mx

**Carmen Mezura Godoy**  
Universidad Veracruzana. Xalapa, Veracruz,  
CONACYT. México  
Correo-e.: cmezura@uv.mx

---