



La Minería de Datos educativos en los contextos mediados por tecnologías en la Educación 4.0

Fabiola Ocampo Botello^{1}, Roberto De Luna Caballero¹, José Jaime Esqueda Elizondo²*

¹Escuela Superior de Cómputo del Instituto Politécnico Nacional, ²Universidad Autónoma de Baja California

*focampob@ipn.mx

RESUMEN

La Minería de Datos educativos (MDE) es una disciplina que permite identificar relaciones entre datos, patrones de comportamiento y modelos de generalización y/o predicción, datos que se producen debido a la interacción de los estudiantes con la computadora, aspecto que resulta de gran utilidad para comprender el comportamiento de los estudiantes y los entornos en los que aprenden. Considerando lo anterior, en este artículo se describen las diversas áreas de investigación presentes en la literatura que señala la importancia de aprovechar la información producida por la interacción antes mencionada, dicho análisis puede ser enfocado desde uno o más de los siguientes puntos de vista: 1) predicción, 2) agrupamiento, 3) minería de relaciones, 4) surgimiento de datos para el análisis humano y 5) descubrimiento de modelos, esta investigación es una respuesta a la motivación de conocer las características de uno de los pilares de estudio y formación de especialistas en la educación 4.0.

Palabras claves: análisis de datos, minería de datos, educación 4.0, minería de datos educativos, agrupamiento.

ABSTRACT

Data Mining in Education (DME) is a discipline that allows to identify relations between data, behavioral patterns and generalization/prediction models, data that are produce due to the interaction of the students with computer, aspect that results very useful to comprehend de behavior of students and the environments in which their learn. Given the above, in this article are described the different areas of current investigations in literature

that points the importance to take advantage of information produced by the interaction mention above, this analysis might be focused from one or more of the following points of view: 1) prediction, 2) groupings, 3) relationship mining, 4) distillation of data for human judgment and 5) discovery with models, this investigation is a reply to the motivation of know the characteristics of one of the pillars of studying and specialist training for education 4.0.

Key words: data analysis, data mining, education 4.0, educational data mining, clustering.

INTRODUCCIÓN

El análisis de datos educativos ya no es sólo la búsqueda y la generación de reportes de información mediante consultas a los datos almacenados, con lo que se pueden conocer las preferencias de estudio de los aprendices. Las nuevas formas de interacción que tienen los estudiantes con los dispositivos digitales, las tecnologías orientadas a aplicaciones educativas y los variados medios de comunicación e interconexión que tienen los discentes, generan datos de forma permanente que es necesario analizar para comprender los entornos en los que se produce el aprendizaje y el descubrimiento de modelos de generalización, patrones de comportamiento, relaciones de datos y predicción de conductas que permitan crear una visión y el reconocimiento de los estudiantes inmersos en la educación 4.0. con la intención de conocer el contexto presente con visos futuros.

En este artículo se presenta una descripción de las diversas áreas de la minería de datos educativos (MDE) que permite a los docentes conocer los datos que se generan en el entorno en el que se forman los aprendices.

Se presenta un estudio descriptivo que consideró una revisión bibliográfica para responder la pregunta de investigación: ¿Cuáles son las características de los diversos modelos de la minería de datos educativos que permiten conocer la información que generan los aprendices en la educación 4.0?

DESARROLLO

La educación 4.0 provoca cambios estructurales en las formas de evaluación, la interacción de forma vertical y horizontal entre alumnos y tutores académicos, la modificación del rol del profesor, ya no como la única fuente de información a la que los estudiantes pueden acudir en busca de orientación, sino como un facilitador del aprendizaje del alumno, así como la interconexión cada vez más presente y permanente de comunicación mediada por tecnologías. Establece Fisk (citado en Aziz Hussin, 2018), que la educación 4.0 se caracteriza por el aprendizaje que se genera en cualquier momento y lugar, por lo que el e-learning, y el aprendizaje invertido proporcionan la oportunidad de que esto suceda, ya que el alumno aprende a su propio ritmo y tienen la libertad de elegir las técnicas y momentos que mejor le favorezcan para el aprendizaje. Lo anterior, crea un nuevo escenario que es necesario explorar e investigar para entender el comportamiento y reconocer la forma en que los alumnos aprenden, lo cual es posible mediante el análisis de datos que se generan mediante la interacción continua que el estudiante tiene con tales dispositivos, lo cual es el tema central de este artículo.

La importancia del análisis de datos históricos para crear modelos de generalización, identificación de relaciones entre variables y patrones de comportamiento de datos con la intención de aplicarse a muestras de datos futuras comprende la llamada minería de datos.

La minería de datos se utiliza para extraer información con sentido y desarrollar relaciones significativas entre variables contenidas en un almacén de datos (Sahu, Shirma & Gondhalakar, 2011) mediante un proceso para extraer conocimiento útil y comprensible, previamente desconocido, en grandes cantidades de datos que se encuentran en diversos formatos (Witten & Frank, 2000 citado en Hernández, Ramírez y Ferri, 2004).

Hernández, Ramírez y Ferri (2004) mencionan que el conocimiento descubierto mediante esta disciplina debe tener las siguientes cuatro características:

- 1) Validez. Los patrones de la relación entre variables descubiertos deben ser válidos para muestras futuras, para datos nuevos, con cierto grado de certidumbre, y no solamente para aquellos datos que se utilizaron para generar tales patrones.

- 2) Novedoso. El conocimiento descubierto debe aportar algo desconocido para el usuario.
- 3) Potencialmente útil. Los patrones descubiertos deben producir acciones que aporten algún beneficio para el usuario.
- 4) Comprensible. El conocimiento descubierto debe ser entendido por el usuario. La comprensibilidad (también conocido como interpretabilidad) se refiere a qué tan bien los humanos captan el modelo generado (Rokach & Maimon, 2015).

Si bien, la minería de datos ha sido aplicada en diversas áreas de estudio como en mercadotecnia para identificar las preferencias de compras de los clientes con base en sus perfiles de compra (Larose & Larose, 2015), en la banca comercial o agencias de préstamos para identificar las características de los clientes que pagarán o no un crédito (Dunham, 2002), la creación de catálogos de productos considerando las características de los clientes (Hernández, Ramírez y Ferri, 2004), por mencionar.

Debido a las características propias de las tareas y métodos de la minería de datos, esta área de estudio se ha utilizado de manera relativamente reciente en el ámbito educativo, Menciona Moreno (2017) que la minería de datos ha sido aplicada en el sector educativo con el propósito de determinar los factores clave que impulsan el éxito y el fracaso de los alumnos en un contexto escolar.

En el área de educación, la minería de datos surgió del análisis de los registros de la interacción estudiante-computadora, el aprendizaje individual a partir de software educativo, el aprendizaje colaborativo asistido por computadora, las pruebas adaptativas y los factores asociados con el fracaso o la no retención de los estudiantes en los cursos (Baker & Yacef, 2009), la cual la ha convertido en un área de estudio, llamada Minería de Datos Educativos (*Educational Data Mining, EDM*).

Baker & Yacef (2009) mencionan que

“La minería de datos educativos (MDE) es una disciplina emergente, que se ocupa de desarrollar métodos para explorar los tipos de datos que provienen de entornos educativos y utilizar métodos para comprender mejor a los estudiantes y los entornos en los que aprenden" (p. 4).

En lo referente al área de educación, los mismos autores plantean que el análisis de datos puede ser abordado desde una o varias de las siguientes áreas de estudio: 1)

predicción, 2) agrupamiento, 3) minería de relaciones, 4) surgimiento de datos para el análisis humano y 5) descubrimiento de modelos.

Los cuales se describen a continuación:

- 1) La predicción como área de estudio del análisis de datos en educación, permite predecir el valor o comportamiento de una variable en función de variables conocidas. La predicción consiste en intentar conocer resultados futuros a partir de modelar los datos actuales (Rodríguez y Díaz, 2009).

Este tipo de técnicas resultan de gran utilidad en aquellos casos en los cuales se desea predecir el resultado que se tendrá, por ejemplo, del aprovechamiento de un nuevo grupo de alumnos en función del referente históricos del comportamiento de otros de su mismo tipo.

- 2) El agrupamiento, generalmente se logra determinando la similitud entre los datos en atributos predefinidos, los datos más similares se integran en grupos, también llamados clústeres (Dunham, M. H., 2002). Un grupo (clúster) es una colección de registros que son similares entre sí y diferentes a los registros de otros clústeres (Larose & Larose, 2015).

En un contexto educativo, el método de agrupamiento descubre, por ejemplo, grupos de estudiantes con características similares que pudieran integrarse en un equipo de trabajo, lo cual permitiría comprender por qué algunos equipos de trabajo trabajan mejor que otros.

- 3) La minería de relaciones en educación, en este artículo se expresará en dos sentidos: la minería de reglas de asociación y la minería de correlación.

- La minería de reglas de asociación se basa en la identificación de características que “van juntas”, es decir, en descubrir asociaciones que pueda ser expresada mediante reglas que cuantifiquen la relación entre dos o más atributos o características (Larose & Larose, 2015).

En un contexto educativo, estas reglas darían cuenta de la relación que se genera entre los diversos recursos digitales o pedagógicos que los alumnos usan en sus procesos de aprendizaje, lo cual representaría un patrón de comportamiento al estudio.

- Respecto a la minería de correlación, Bennet, Briggs & Triola (2011) establecen que una correlación se presenta entre dos variables cuando valores mayores de una variable van con valores consistentes de otra variable o cuando valores grandes de una variable corresponden de manera consistentes con valores menores de otra variable.
En un contexto educativo, este tipo de técnica identifica el comportamiento de la relación de dos variables en términos de la intensidad y sentido que guardan en el desempeño de un estudiante. Por ejemplo, se podría conocer la relación que tiene el desempeño académico y la motivación del alumno.
- 4) El surgimiento de datos para el análisis humano se relaciona con la estadística y visualización (Baker & Yacef, 2009).
- 5) El descubrimiento de modelos, es una categoría inusual, pero que poco a poco se va volviendo más popular, por ejemplo, en aquellos casos en los que se desea conocer cómo los diversos tipos de comportamiento de los discentes incide de forma diferente en el aprendizaje de los estudiantes y cómo las variaciones en el diseño de los tutores inteligentes afectan el comportamiento de los estudiantes a lo largo del tiempo (Baker & Yacef, 2009).

La investigación en la MDE, sigue creciendo e incursionando en diversas áreas como la web, en la cual, el estudio se centra en el análisis de datos más variados, más voluminosos y que se generan a gran velocidad y que sin lugar a dudas aportará información relevante para caracterizar los espacios en los que los alumnos actualmente aprenden.

DISCUSIÓN

La industria 4.0, como todo proceso industrial, modificó la forma en que las personas se relacionan, se comunican y establecen relaciones de trabajo y estudio; esta ha sido considerada la revolución industrial que evolucionó de forma vertiginosa provocando el surgimiento de diversas áreas de estudio de la llamada educación 4.0, siendo una de ellas, la minería de datos educativos.

La minería de datos es un tema de actualidad, debido a que tiene una gran cantidad de aplicaciones en la educación, que van desde la descripción de un contexto en general,

hasta el reconocimiento de características teóricas y el descubrimiento de relaciones de variables o patrones de comportamiento que no sería posible conocer sin aplicar técnicas de minería de datos, inteligencia artificial, estadística y recientemente el big data que permite la visualización y análisis de datos complejos que se generan a gran volumen, velocidad y variedad.

La MDE presenta diversas diferencias con la minería de datos tradicional, una de ellas son los datos mismos, ya que en la MDE estos son muy variados y dinámicos debido a los cambios que se presentan en la oferta educativa en cada periodo escolar, aunado a que los estudiantes no realicen los mismos ejercicios o actividades y que consecuentemente no se tengan las mismas variables.

CONCLUSIONES

Respecto a la MDE, las conclusiones en este artículo versan en tres sentidos:

Es un campo multidisciplinario que permite conocer los contextos actuales en los cuales se desenvuelve el aprendiz, la identificación de los datos que se generan por la interacción en los diversos canales de comunicación que existen entre el estudiante y los recursos tecnológicos para crear un referente de la forma en que se produce el aprendizaje y con ello comprender otros aspectos como las preferencias de los recursos pedagógicos que prefieren los discentes en su aprendizaje, la predicción de un comportamiento que pudiera afectar su desempeño y la identificación de estudiantes con característica afines para crear planes de atención.

El surgimiento de un área de estudio extensa y multidisciplinaria que no sólo permita identificar modelos de generalización, patrones de comportamiento y relación entre variables, sino el aporte de la interpretación y tendencia que permita crear comunidades de aprendizaje.

La integración de variadas técnicas y métodos de análisis, cuya naturaleza es necesario conocer para aplicarla de forma adecuada en la MDE y es en este punto donde la intuición de los profesores tiene un papel muy importante.

LITERATURA CITADA

- Aziz Hussin, A. (2018). Education 4.0 Made Simple: Ideas for teaching. *International Journal of Education & Literacy Studies*. 6(3). 92–98. Doi: <http://dx.doi.org/10.7575/aiac.ijels.v.6n.3p.92>
- Baker, R. & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*. Article 1, Vol 1, No 1, Fall 2009. 3–16.
- Bennet, Briggs & Triola. (2011). *Razonamiento estadístico*. Pearson. México.
- Dunham, M. H. (2002). *Data mining: introductory and advanced topics*. Prentice Hall.
- Sahu, H., Shurma, S. & Gondhalakar, S. (2011). A Brief Overview on Data Mining Survey. *International Journal of Computer Technology and Electronics Engineering (IJCTEE)*, vol.1. 114–121.
- Hernández, J., Ramírez, M.J. y Ferri, C. (2004). *Introducción a la Minería de datos*. Editorial Pearson.
- Larose, D. & Larose, Ch. (2015). *Data Mining and Predictive Analytics*. Second Edition. Wiley.
- Moreno S., (2017). Científico de datos: codificando el valor oculto e intangible de los datos. *Revista Digital Universitaria*. Vol. 18, Núm. 7, septiembre-octubre 2017.
- Rokach, L. & Maimon, O. (2015). *Data Mining with decision trees. Theory and Applications*. Second Edition. World Scientific Publishing Co. Pte. Ltd.
- Rodríguez, Y. y Díaz, A. (2009). Herramientas de Minería de Datos. *Revista Cubana de Ciencias Informáticas [en línea]* 2009, 3 (Julio-Diciembre). Disponible en: <http://google.redalyc.org/articulo.oa?id=378343637009>