

USO DE TÉCNICAS DE MINERÍA DE DATOS EN LA ENSEÑANZA DEL ALGEBRA LINEAL

Lenniet Coello Blanco, Olga Lidia Perez Gonzalez, Ángela Mercedes Martín Sánchez

Universidad de Camagüey (Cuba), Universidad Autónoma de Santo Domingo (República Dominicana)

lenniet.coello@reduc.edu.cu

Palabras clave: algebra lineal, minería de datos, sistema tutorial inteligente

Key words: lineal algebra, data mining, intelligent tutorial system

RESUMEN: La presente investigación surge de la necesidad de dar solución a la baja promoción de los estudiantes de primer año de las carreras de ciencias técnicas de la Universidad de Camagüey, en la asignatura de Algebra Lineal. En esta materia el estudiante requiere de altos niveles de abstracción y un buen razonamiento lógico-matemático. Se decidió, entonces desarrollar un Sistema Tutorial Inteligente (STI) usando técnicas de Minería de datos que apoye al educando durante su estudio independiente, interviniendo como un tutor particular que actúe de acuerdo a las necesidades del alumno. Como resultado se pretende obtener un sistema para ser utilizado por estudiantes y profesores cuyo objetivo principal es proporcionar al educando la ayuda pedagógica adecuada en apoyo al proceso de enseñanza-aprendizaje de la asignatura del Algebra Lineal.

ABSTRACT: The present investigation arises of the necessity of giving solution to the drop promotion of the students of first year of the careers of technical sciences of the University of Camagüey, in the subject of Lineal Algebra. In this matter the student requires of high levels of abstraction and a good logical-mathematical reasoning. Was decided develop a Intelligent Tutorial System (ITS) using technical of Data Mining that supports to the educating during their independent study, intervening as a particular tutor that acts according to the student's necessities. As a result it is sought to obtain a system to be used by students and professors whose main objective is to provide to the educating the pedagogic help adapted in support to the process of teaching learning of the subject of the Lineal Algebra.

■ INTRODUCCIÓN

El desarrollo de las Tecnologías de la Información y las Comunicaciones (TIC) aumenta infinitamente las capacidades para procesar, almacenar y transmitir información y con el avance de las mismas, germina una nueva sociedad en la que el manejo de la misma es sinónimo de progreso. Además, la utilización de las TIC tiene grandes ventajas: interés, motivación, interacción, continua actividad intelectual, desarrollo de la iniciativa, mayor comunicación entre profesores y alumnos.

La Educación Superior, por ser la última etapa de la formación profesional del pregrado complejiza el rigor de sus materias y el sistema de evaluación, además, el dinamismo de la vida universitaria acorta los plazos de tiempo para el estudio. Como parte de esta tendencia mundial, Cuba ha incrementado la utilización de las tecnologías en este sector, lo que posibilita que en el país pueda considerarse una sociedad de la información en desarrollo.

En el campo de las Ciencias de la Computación, la Inteligencia Artificial (IA) surge como una de las ramas de estudio más recientes y promisorias, cuyo objetivo es el de entender la naturaleza de la inteligencia a través del diseño de sistemas computacionales que la exhiban (Salgueiro, 2005).

Una de las críticas que se oyen más a menudo respecto a la Inteligencia Artificial (IA), es que las máquinas no se pueden considerar inteligentes hasta que no sean capaces de aprender a hacer cosas nuevas y a adaptarse a las nuevas situaciones, en lugar de limitarse a hacer aquellas actividades para las que fueron programadas. En vez de preguntarse si una computadora es capaz de “aprender”, resulta mucho más clarificador intentar describir a qué actividades se refiere exactamente cuándo se dice “aprender” y cuáles mecanismos se pueden utilizar para llevar a cabo dichas actividades.

(Simon, 1983) (Simón, 1983) se refiere al aprendizaje como cambios en el sistema para desarrollar tareas a partir de las mismas condiciones de un modo más eficiente y eficaz cada vez.

El Aprendizaje Automático es el área de la Inteligencia Artificial que se ocupa de desarrollar técnicas capaces de aprender, es decir, extraer de forma automática conocimiento subyacente en la información (Mitchell, 1997). Constituye junto con la estadística el corazón del análisis inteligente de los datos (Ruiz, 2006). Los principios seguidos en el aprendizaje automático y en la minería de datos son los mismos: la máquina genera un modelo a partir de ejemplos y lo usa para resolver el problema.

En el desarrollo de la presente investigación se realizó un seguimiento de los alumnos de la Facultad de Informática de la Universidad de Camagüey a través del análisis de los reportes semestrales con el fin de conocer el motivo de la baja promoción en la asignatura de Álgebra Lineal, se evidenció que en los estudiantes existen muchas dificultades en cuanto a la comprensión y asimilación de los tópicos de Diagonalización, por lo que en la actualidad los profesores se encuentran en la búsqueda y elaboración de estrategias destinadas a la atenuación del problema.

La influencia y dependencia que tiene en varias áreas dentro y fuera del álgebra lineal constituye una de las principales razones para crear un sistema que respalde la enseñanza-aprendizaje de la Diagonalización. Los temas de Diagonalización son una vía de simplificar las extensas soluciones algebraicas a través de sus conceptos y métodos. Los valores y vectores propios son esenciales para plantear y resolver problemas de física e ingeniería relacionados con sistemas dinámicos, oscilatorios, teoría general de la estabilidad mecánica cuántica entre otros. (Coello, Casas, Pérez, & Caballero, 2015)

La Diagonalización es un proceso que involucra todos los temas anteriores en el plan de estudio de la asignatura: sistemas de ecuaciones lineales, matrices y determinantes, espacios vectoriales, sub-espacios, transformaciones lineales, núcleo, imagen, entre otros. Diremos que una matriz es diagonalizable si todos sus valores propios son distintos. Además cuando se da esta situación diremos que esta matriz y su matriz diagonal son semejantes (Coello et al., 2015)

El Álgebra, dado lo abstracto de su contenido, necesita de una herramienta que permita la total apropiación del conocimiento y la creación de habilidades. El álgebra enlaza los contextos gráficos, geométricos y tabulares, haciendo del lápiz y papel una actividad estática y no como se haría con el uso de las TIC que lo hace más dinámico. (Coello et al., 2015). Por este motivo se pensó en el desarrollo de un Sistema Tutorial Inteligente (STI), una técnica de IA capaz de adaptarse a los conocimientos previos del estudiante y a partir de esto enseñar adaptándose continuamente a su evolución particular.

Este sistema cumpliría la tarea de entrenar adaptando diferentes estrategias o modalidades de enseñanza, enfatizando en el tópico de Diagonalización del Algebra Lineal. Un tutor inteligente es un sistema de software que utiliza técnicas de inteligencia artificial para representar el conocimiento e interactúa con los estudiantes para enseñárselo (VanLehn, 1988).

Tomando en cuenta los argumentos anteriormente expuestos, el objetivo es un sistema que, aprovechando los recursos disponibles en la actualidad permita disminuir las tareas del profesor. Para ello desarrollaremos un sistema tutorial inteligente que se nutre de las técnicas del aprendizaje automático o Minería de dato para mejorar la experiencia de aprendizaje desde la perspectiva del estudiante

■ DESARROLLO

Minería de Datos

Desde hace más de dos décadas se vienen desarrollando y utilizando complejos algoritmos para la extracción de patrones útiles en grandes conjuntos de datos. Gran parte de esta información representa transacciones o situaciones que se han producido, siendo útil no sólo para explicar el pasado, sino para entender el presente y predecir la información futura. En muchas ocasiones, el método tradicional de convertir los datos en conocimiento consiste en un análisis e interpretación realizada de forma manual por especialistas en la materia estudiada. Esta forma de actuar es lenta, cara y altamente subjetiva. De hecho, la enorme cantidad de datos desborda la capacidad humana de comprenderlos y el análisis manual hace que las decisiones se tomen según la intuición de los especialistas.

A finales de la década de los 80, la creciente necesidad de automatizar todo este proceso inductivo abre una línea de investigación para el análisis inteligente de datos. Al conjunto de métodos matemáticos y técnicas software para análisis inteligente de datos y búsqueda de regularidades y tendencias en los mismos, aplicados de forma iterativa e interactiva, se denominaron técnicas de Minería de Datos o Data Mining (DM). Su nombre proviene de las similitudes encontradas entre buscar valiosa información de negocio en grandes bases de datos y minar una montaña para encontrar una veta de metales valiosos. Su tarea fundamental es la de encontrar modelos inteligibles a partir de los datos, y para que el proceso sea efectivo debe ser automático, generando patrones que ayuden en la toma de decisiones beneficiosas para la organización.

Para obtener conclusiones válidas y útiles al aplicar minería de datos, es necesario complementar este proceso con una adecuada preparación de los datos previa al proceso de minería y un análisis posterior de los resultados obtenidos. Así, podemos afirmar que el proceso de minería de datos pertenece a un esquema más amplio, denominado extracción o descubrimiento de conocimiento en bases de datos (KDD, Knowledge Discovery in Databases). Se puede intuir que el KDD es un campo multidisciplinar, donde las principales áreas contribuyentes son el Aprendizaje Automático, las Bases de datos y la Estadística.

El Aprendizaje Automático es el área de la Inteligencia Artificial que se ocupa de desarrollar técnicas capaces de aprender, es decir, extraer de forma automática conocimiento subyacente en la información. Constituye junto con la estadística el corazón del análisis inteligente de los datos. Los principios seguidos en el aprendizaje automático y en la minería de datos son los mismos: la máquina genera un modelo a partir de ejemplos y lo usa para resolver el problema. (Ruiz, 2006).

Dentro de la Minería de Datos se incluyen las actividades siguientes Clasificación, Estimación, Predicción, Determinar grupos afines o reglas de asociación, Clustering (agrupamiento), Descripción y visualización. Las tres primeras, es decir la Clasificación, Estimación y Predicción se agrupan bajo del nombre de MD Directa. Por su parte, las tres últimas reciben el nombre de MD Indirecta. En la Minería de Datos Directa (MDD) se tiene claro el objetivo, mientras que en la Minería de Datos Indirecta (MDI) no se sabe aún a ciencia cierta qué resultados se quieren obtener. (Linoff, 2000).

En este trabajo se muestran las ideas puestas en prácticas de crear un STI con técnicas de Minería de Datos: Algoritmo de agrupamiento y Algoritmo de clasificación para emular al tutor humano y además que proveer al estudiante de cierta flexibilidad para la elección del tipo de tutorizado más adecuado.

Sistema Tutor Inteligente con técnicas de Minería de datos para la asignatura de Álgebra Lineal (STIAL).

En la actualidad, los tutores inteligentes permiten brindar un seguimiento eficaz del proceso enseñanza–aprendizaje, puesto que brindan una alternativa de tutoría personalizada para el alumno a través de técnicas de enseñanza, tales como: aprendizaje por reforzamiento y ejercitación, búsqueda interactiva de conocimiento, aprendizaje por descubrimiento y proceso de construcción de conocimiento.

Los profesores humanos no responden a las preguntas o dan ayuda de acuerdo a un guión. En lugar de eso, disponen de conocimiento sobre el área que enseñan, tiene una intuición sobre lo que sabe (o debería saber) el alumno, y tiene conocimientos sobre estrategias para enseñar. Esas tres fuentes de conocimiento se utilizan de manera conjunta para determinar en cada momento como interactuar con el alumno para conseguir que aprenda (Martin, 2007).

Por tanto un STI es un tipo de agente inteligente caracterizado por su habilidad para tomar decisiones y la ejecución de acciones sin la intervención de ninguna persona que monitoriza el funcionamiento del sistema. Así pues, un STI tendría que, en primer lugar, evaluar el conocimiento que el aprendiz tiene sobre la materia. A continuación, ha de presentar información acorde a ese nivel (por ejemplo, definiciones, material gráfico o auditivo). Y posteriormente ha de evaluar si la persona ha asimilado el conocimiento nuevo que se pretendía que el usuario hubiese adquirido (Puga & García, 2008).

Entre sus ventajas se encuentra: Constituyen una fuente de materiales de enseñanza, proporcionan problemas para el estudiante pueda alcanzar un determinado nivel de conocimiento, controlan el nivel

de dificultad de los problemas de manera que el estudiante se enfrente a ejercicios adecuados a sus necesidades, contribuyen a planificar la instrucción y, a la vez, ayuda a los estudiantes a aprender a aprender. (Durán & Costaguta, 2007)

Módulo Pedagógico o Tutor

Este módulo es el responsable de aplicar una estrategia pedagógica y seleccionar el material de aprendizaje acorde al estudiante. Integra el conocimiento acerca del método de enseñanza, las técnicas didácticas y del dominio a ser enseñado. El módulo tutor se puede descomponer en dos submódulos: Analizador de Perfil y Protocolos Pedagógico.

El analizador de perfil permite determinar, a partir de los datos almacenados en el módulo estudiante y a través de las herramientas que proporciona la Minería de Datos, tales como los algoritmos de agrupamiento, obtener la relación de los estilos de aprendizaje dominantes en los estudiantes. Para obtener el diagnóstico de los estilos de aprendizaje predominantes en los estudiantes de Ingeniería Informática el trabajo se sustentó en las investigaciones presentadas por (Durán y Costaguta, 2007) y por (Rivero et.al, 2013). El instrumento utilizado para recolectar los datos de los estudiantes de informática fue el Test de Estilos de Aprendizaje propuesto por Felder y Silverman en 1988 (Felder & Silverman, 1988). Para realizar este diagnóstico, se optó por el algoritmo de agrupamiento FarthestFirst por tratarse de un problema de k centros donde se pretende que la máxima distancia entre una tupla y su centroide sea mínima. El Algoritmo FarthestFirst se clasifica como un algoritmo de agrupamiento no jerárquico. Comienza seleccionando aleatoriamente una instancia que pasa a ser el centro del clúster. Se calcula la distancia entre cada una de las instancias y el centro. La distancia que se encuentre más alejada del centro más cercano es seleccionada como el nuevo centro del clúster. Este proceso se repite hasta alcanzar el número de clusters o grupos buscado. Los objetos que componen cada subgrupo deben ser bastante similares entre ellos y lo suficientemente disimilares con respecto a los ejemplos del otro subgrupo. Cada centroide o prototipo devuelto por el algoritmo de agrupamiento representan a la mayoría de los estudiantes de informática encuestados, con lo que quedaron identificados dos grupos con sus correspondientes estilos de aprendizaje. Grupo 1: Visual, Intuitivo, Activo y Secuencial y Grupo 2: Visual, Sensitivo, Activo y Global.

Cuando el estudiante accede al sistema, luego de realizar el test lo primero que hace es recibir su categorización. Aquí implementamos un algoritmo de clasificación que devuelve al estudiante su perfil de aprendizaje, que no es más que una de las dos clases o grupos que vimos anteriormente. La clasificación supervisada consiste en el proceso de asignar a una entrada concreta, el nombre de una clase a la que pertenece. Las clases entre las que puede elegir el procedimiento de clasificación supervisada se pueden describir de gran cantidad de formas. La clasificación supervisada constituye una parte importante de muchas de las tareas de resolución de problemas y en su forma más simple se presenta como una tarea de reconocimiento (Caballero, 2007). Para el STIAL se decidió implementar el Naive Bayes que es un algoritmo sencillo, fácil de implementar, sin parámetros que ajustar y que a pesar de su simplicidad es sin duda fuerte y uno de los clasificadores más utilizados. Además, diversos estudios demuestran que sus resultados son competitivos con otras técnicas (redes neuronales y árboles de decisión entre otras) en muchos problemas y que incluso las superan en algunos otros. El fundamento principal del clasificador Naive Bayes es la suposición de que todos los atributos son independientes conocido el valor de la variable clase. Su denominación proviene de la hipótesis de que las variables predictivas son condicionalmente independientes dada la variable a

clasificar y con esto ya queda definida una estructura, por lo que sólo se tienen que aprender las probabilidades de los valores de los atributos dada la clase (Chávez, 2008).

Módulo del Estudiante

Este módulo tiene por objetivo realizar el diagnóstico cognitivo del alumno y es identificada como la tarea más compleja en el desarrollo de STI. Puede afirmarse que el modelo del estudiante es un problema de investigación que debe enfocarse desde todas sus aristas con el fin de obtener una representación de las características del estudiante completa y precisa. Algunos autores, toman en consideración características tales como: el estilo de aprendizaje, el nivel de conocimiento, la información personal o la combinación de algunas de ellas (Rodríguez & Fernández, 2000). Cuanto aprenda un estudiante en una clase dependerá de la habilidad innata y de su preparación previa, pero además de la compatibilidad entre su estilo de aprendizaje y el estilo de enseñanza de su instructor.

Por otra parte el determinar las características del perfil de aprendizaje de nuestros estudiantes para, en función de ellas, adecuar las estrategias de enseñanza y detectar el estilo de aprendizaje dominante en nuestro grupo de estudiantes puede favorecer al utilizar el método de enseñanza más adecuado, con el objeto de lograr un resultado óptimo (Rodríguez & Fernández, 2000).

Para realizar este diagnóstico en el STIAL el estudiante apenas se autentica deberá llenar un cuestionario. Se utilizarán las planillas de estilos de aprendizaje recreadas de Felder y Silverman (Felder & Silverman, 1988) que son herramientas para la toma de datos de alta validez y confiabilidad que han demostrado su aptitud a la hora de clasificación de los estilos por su autor y otros investigadores que las utilizaron.

Módulo del Dominio

El módulo del dominio, denominado también por muchos autores como módulo experto, proporciona los conocimientos del dominio. En el caso del STIAL decidimos en esta primera versión tutoriar el tema de la Diagonalización quedando el sistema abierto a la inserción de otros temas del Álgebra Lineal.

El módulo del Dominio satisface dos propósitos diferentes. En primer lugar, presentar la materia de la forma adecuada para que el alumno adquiera las habilidades. En segundo lugar, debe ser capaz de evaluar las respuestas del estudiante y en dependencia de esta brindarle al estudiante o bien una ayuda o bien una estimulación o reto.

Los materiales didácticos con las estrategias pedagógicas adecuadas para adaptarse a los modelos de estudiantes. Como las dos clases de estudiante son visuales las formas de mostrarle el conocimiento debía poseer diagramas y gráficos. Luego si el estudiante es sensitivo se decidió además que el medio de enseñanza a utilizar debía ser un video, donde se mostraran gráficos para estimular la vista y que además de estimulará el oído. Si el estudiante por el contrario es intuitivo se le presentara un ppt que es la misma presentación en la que se apoya el video pero que incluirá más conceptos y argumentos para estimular a través de ideas y lecturas como lo prefieren este estilo de aprendizaje.

Como las dos clases de estudiante son activos la forma de procesar la información es a través de ejercicios a los que les ven conexiones inmediatas con el mundo real. Luego si el estudiante es global los ejercicios requieren de un procesamiento con una visión integral si por el contrario el estudiante es secuencial los ejercicios mostrados necesitan una progresión lógica de pasos incrementales pequeños.

El propósito del STIAL es adaptarse a la forma en que el estudiante percibe la información y las formas en la que las procesa. Además debe brindarle ayuda cuando cometa errores, proveerle el conocimiento que necesita para solucionar los ejercicios con explicaciones.

■ CONCLUSIONES

El Sistema Tutor Inteligente para el Algebra Lineal puede ser usado en la educación y el entrenamiento del estudiante como apoyo al proceso de enseñanza-aprendizaje en el tema de Diagonalización.

Con el objetivo del trabajo cumplido, se provee al campo de los Sistemas Tutores Inteligentes de una nueva herramienta redundando esto en una ganancia, no solo para el desempeño del STI en sí mismo, sino en el estudiante, que es el componente humano fundamental que hace útil al sistema y le brinda identidad. Así se pretende realizar un aporte al sugerir técnicas de agrupamiento como el FarthestFirst y de clasificación como el Naive Bayes para facilitar la selección del protocolo pedagógico adecuado. Con su implementación se pronostica un uso de los recursos materiales y humanos con más eficiencia, además de la búsqueda de métodos y soluciones para una mayor comprensión de la asignatura.

■ REFERENCIAS BIBLIOGRÁFICAS

- Caballero, Y. (2007). Acerca del Aprendizaje Automatizado, la Selección de Características y la Edición de conjuntos de Entrenamiento. (Tesis de Doctorado), Universidad de Camagüey, Camagüey.
- Coello, L., Casas, L., Pérez, O. L., & Caballero, Y. (2015). Redes neuronales artificiales en la producción de tecnología educativa para la enseñanza de la diagonalización. *Revista Academia y Virtualidad*, 8(1), 12-20.
- Chávez, M. (2008). Modelo de redes bayesianas en el estudio de secuencias genómicas y otros problemas biomédicas. (Tesis de Doctorado), Universidad Central "Marta Abreu" de las Villas, Villa Clara.
- Durán, E., & Costaguta, R. (2007). Minería de datos para descubrir estilos de aprendizaje. *Revista Iberoamericana de Educación*, 2(42).
- Felder, R. M., & Silverman, L. K. (1988). Learning Styles and Teaching Styles in Engineering Education. *Engr. Ed*, 78(7), 674-681.
- Linoff, B. (2000). *Mastering Data Mining* Wiley (Ed.)
- Martin, P. P. G. (2007). Modelo de enseñanza basada en casos: de los tutores inteligentes a los videojuegos. (Tesis de Grado de Doctor), Universidad Complutense de Madrid, Madrid.
- Mitchell, T. (1997). *Machine Learning*: McGraw-Hill Science.
- Puga, J. L., & García, J. G. (2008). Sistemas de Tutorización Inteligente Basados en Redes Bayesianas. *Revista Electrónica de Metodología Aplicada*, 13(1), 13-25.

- Rodríguez, L. I. M., & Fernández, Y. d. A. (2000). Sistema Tutorial Inteligente de apoyo al proceso de enseñanza-aprendizaje de la asignatura de Sistemas Operativos. (Tesis de Grado), Universidad de las Ciencias Informáticas, Habana.
- Ruiz, R. (2006). Heurísticas de selección de atributos para datos de gran dimensionalidad. (Tesis Doctoral para optar al grado de Doctor en Informática), Universidad de Sevilla, Sevilla.
- Salgueiro, F. (2005). Sistemas Inteligentes para el Modelado del Tutor (Tesis de Grado en Ingeniería Informática), Universidad de Buenos Aires, Buenos Aires.
- Simon, H. A. (1983). Why should machines learn? In R. S. Michalski, J. G. Carbonell & T. M. Mitchel (Eds.), Machine Learning, An Artificial Intelligence approach. Palo Alto, CA.
- VanLehn, K. (1988). Student Modelling. . In H. N. J. L. E. Associates (Ed.), Foundations of Intelligent Tutoring systems (pp. 55-78).