

# Minería de Datos Educativa en Ambientes Virtuales de Aprendizaje

Constanza R. Huapaya, Francisco A. Lizarralde, Graciela M. Arona, Stella M. Massa

Grupo de Investigación en Inteligencia Artificial (integrante del CIMEPB)

Departamento de Matemática-Facultad Ingeniería

Universidad Nacional de Mar del Plata

Juan B. Justo 4302. 7600. Mar del Plata

223-4816600 int 259

[{huapaya,flizarra,grarona,smassa}@fi.mdp.edu.ar](mailto:{huapaya,flizarra,grarona,smassa}@fi.mdp.edu.ar)

## Resumen

En la actualidad, los **Ambientes Virtuales de Aprendizaje (AVAs)** almacenan una gran cantidad de datos sobre las actividades de los estudiantes cuando estos toman un curso y usualmente esta información es utilizada para monitorear características del curso. Esta información se presenta comúnmente en formato tabular que, dependiendo de la cantidad o del tipo de datos, a veces resulta difícil de interpretar.

Esta vasta cantidad de datos no puede ser inspeccionada manualmente. Esta es la razón por la cual el uso de la **Minería de Datos Educativa (MDE)** es muy apropiada para descubrir información “escondida” en las bases de datos de un AVA. Los métodos pueden ser aplicados para explorar, visualizar, y analizar datos de e-learning con la finalidad de identificar patrones útiles aplicables a la evaluación de la actividad del usuario en la web y descubrir más profundamente como aprenden los estudiantes.

Los métodos nos permiten revelar nuevo, interesante y útil conocimiento basado en datos de la actividad del estudiante. En particular, nos interesa investigar la información para estimar los perfiles cognitivos del estudiante a partir de los datos recolectados. Estos perfiles aportarán

conocimiento al AVA para entregar material remedial (como Objetos de Aprendizaje personalizados) para los estudiantes con bajo rendimiento.

**Palabras clave:** Minería de datos educativos, Ambientes virtuales de aprendizaje, educación en ingeniería.

## Contexto

La investigación descrita en esta presentación se desarrolla en el marco del Proyecto “Ambientes virtuales de aprendizaje para la enseñanza de la ingeniería” perteneciente al Departamento de Matemática de la Facultad de Ingeniería (integrante del Centro de Investigación en Procesos Básicos, Metodologías y Educación (CIMEPB, Facultad de Psicología UNMDP) y presentado ante la Universidad Nacional de Mar del Plata.

## Introducción

Los Ambientes Virtuales de Aprendizaje (AVAs) [1,2,3,4,5] (en inglés, **Virtual Learning Environments** también conocidos como **Learning Management Systems** o **Learning Support Systems**) facilitan el intercambio de información y comunicación entre los participantes de los cursos (docentes y estudiantes).

Si bien existen diferentes definiciones de AVAs, estos pueden describirse como una colección de herramientas informáticas integradas que facilitan la administración del aprendizaje en línea, proveyendo un mecanismo de entrega, seguimiento del estudiante, evaluación y acceso a los recursos

Existen grandes expectativas sobre el aprendizaje con AVAs por su *versatilidad* y *flexibilidad*. Se describen como versátiles porque se consideran apropiados para todo tipo de contenidos y temas. Flexibles, porque permiten el acceso en cualquier tiempo y en cualquier lugar. Al aplicar las nuevas tecnologías en los procesos educativos se modifican las condiciones de la enseñanza y el aprendizaje. Los roles y las habilidades tradicionales del estudiante y del docente deben cambiar.

Los estudiantes que interactúan con un AVA deben tener un rol más activo. Los programas de educación basados en Web requieren un mayor esfuerzo, implicación y habilidades del estudiante para gestionar información. Desde esta perspectiva, el estudiante debe evolucionar para convertirse en el principal responsable de su propio proceso de aprendizaje.

Para los docentes, trabajar en un ambiente de este tipo no significa llevar su tarea habitual de un aula física a una virtual. Se requiere de docentes preparados para innovar en su estilo de enseñanza, para pensar en posibilidades y realidades, en definitiva, enseñar en un salón de clase transformado en un ambiente de aprendizaje potenciado por la web.

La interacción entre un sistema AVA y sus usuarios produce una gran cantidad de datos y ésta información personal es almacenada en bases de datos. Esta información comprende datos sobre la interacción misma (tiempo de acceso, materiales visitados, etc), resultados de la actividad académica (resultado de ejercicios, notas de tests, etc.), datos personales, etc.

La Minería de Datos [6] es un área multidisciplinaria en la cual convergen diversos paradigmas computacionales como la programación lógica, redes neuronales artificiales, lógica difusa, inducción por reglas, etc. y trabaja con métodos como clustering, estadística, clasificación, minería de textos, visualización, etc. La aplicación de la Minería de Datos sobre los sistemas de enseñanza/aprendizaje originaron la Minería de Datos Educativa (MDE) y datan de hace unos años [7,8,9,10,11,12,13]. MDE explora, visualiza y analiza la información de sistemas computacionales de aprendizaje.

El proceso de la Minería de Datos Educativa convierte a los datos adquiridos desde un AVA en información valiosa que puede poseer gran impacto en la práctica e investigación educativa. En particular, nos referimos a la búsqueda automática de conocimiento valioso en los datos, i.e., el descubrimiento de patrones de comportamiento interesantes para la investigación.

Los datos originales de los estudiantes deben ser transformados para que estos sean compatibles con los que aceptan los métodos o algoritmos de la Minería de Datos. Por ejemplo, esta transformación incluye:

- **Limpieza de los datos:** consiste en la eliminación de datos irrelevantes
- **Identificación del usuario:** permite reconocer la asociación entre el estudiante y su actividad en las páginas del sistema
- **Identificación de la sesión:** se referencia las páginas del curso que el estudiante visitó
- **Transformación de los datos:** cálculo de nuevos atributos a partir de los existentes.
- **Integración de los datos:** es la integración y sincronización de datos desde fuentes heterogéneas.

- **Reducción de datos:** se reduce la dimensión de los datos a fin de un mejor procesamiento

En nuestra investigación, el AVA que se usará es MOODLE [14]. En la figura 1 se aprecia la página principal del sistema en desarrollo:

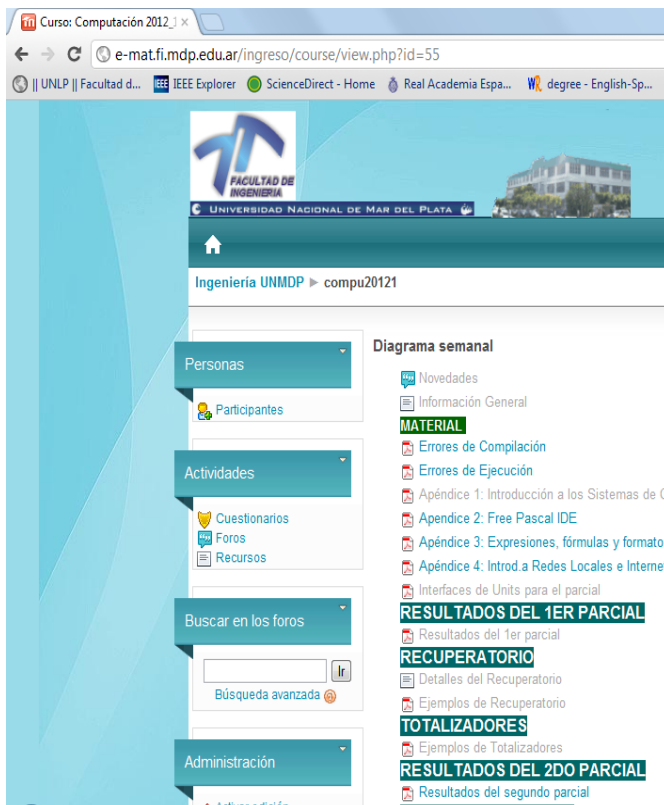


Figura 1: front-end de un curso en MOODLE

Uno de los objetivos perseguidos por esta investigación es aplicar los algoritmos de la Minería de Datos Educativa a fin de construir una estimación del estado cognitivo del estudiante [15,16,17,18,19,20,21,22]. En especial se usará los métodos de la Lógica Difusa y clustering difuso para construir los perfiles cognitivos [23,24,25].

Se usará KEEL (Knowledge Extraction based on Evolutionary Learning) como herramienta para procesar los datos de MOODLE. KEEL (ver figura 2) es un framework cuyo objetivo es evaluar algoritmos evolutivos con fines educativos y de investigación. Se incluyen regresión, clasificación, clustering, minería de

patrones. Asimismo, KEEL contiene algoritmos clásicos de extracción, técnicas de pre procesamiento, y algoritmos de aprendizaje basados en Inteligencia Computacional como aprendizaje basado en reglas evolutivas (Pittsburgh, Michigan and IRL, etc.), y modelos híbridos tales como sistemas difuso/genéticos, redes neuronales evolutivas, etc.



Figura 2: Página inicial de KEEL

El uso repetitivo de la información generada por MOODLE creará perfiles dinámicos a lo largo de un curso determinado (p.e. asignaturas de las carreras de Ingeniería de la Universidad Nacional de Mar del Plata). Este proceso nos permitirá crear cursos mejorados a través de la adaptación al nivel de aprendizaje de los estudiantes. En la figura 3 se aprecia el ciclo de mejoramiento.

## Líneas de investigación y desarrollo

- *Construcción de un perfil cognitivo del estudiante basado en los datos filtrados por técnicas de Minería de Datos Educativa: la obtención de perfiles individuales y grupales permitirán la*

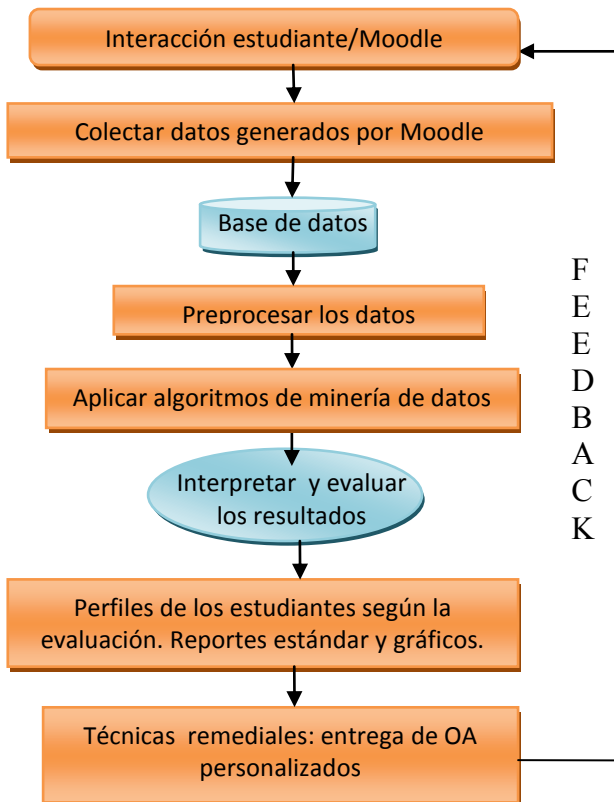


Figura 3: ciclo de mejoramiento de un curso

adaptación del material instruccional (Objetos de aprendizaje, tests, etc.)

- *Análisis de las herramientas de Minería de Datos más apropiadas para los datos recolectados:* se estudiarán las diversas técnicas de la Minería de Datos a fin de adaptarlas a nuestros objetivos.
- *Adquisición de información desde un AVA:* decisión sobre las bases de datos a usar, generados por la acción del estudiante.

## Resultados y Objetivos

- Se ha desarrollado un modelo de perfiles cognitivos considerando una variable lingüística de salida (nivel de conocimiento) y tres variables lingüísticas de entrada (progresión de notas, nivel de aprobación de las pruebas y nota final respecto a la media del curso). El modelo se completa con 27

reglas difusas que capturan la experticia de los profesores. Se espera un mejoramiento incremental de los perfiles con nuevas técnicas de Minería de Datos y nuevas variables lingüísticas.

- Se ha desarrollado un espacio tridimensional que se genera a partir de la evaluación de tres habilidades cognitivas de cada estudiante: comprensión del problema, planteo general de la solución y eficiencia de la solución. Posteriormente, se aplicaron diversas técnicas de clustering y se analizaron los resultados a fin de encontrar los perfiles del grupo. Se aplicarán nuevas técnicas de KEEL para mejorar los resultados.
- Se desarrollará de cursos propios de las carreras de Ingeniería basados en MOODLE y KEEL basados en MDE.

## Formación de Recursos Humanos

El equipo de trabajo está integrado por especialistas en Informática, en Psicología y estudiantes de Ingeniería y Psicología. Actualmente se está desarrollando tres tesis de la Maestría en Tecnología Informática Aplicada a Educación de la Facultad de Informática de la Universidad Nacional de La Plata y una tesis de doctorado de la misma Facultad.

## Referencias

- [1] Dillenbourg P. Schneider D., Synteta P.(2002) Virtual Learning Environments”, *Proceedings of the 3rd Congress on Information and Communication Technologies in education*, Rhodes, Kastaniotis Editions, Greece, 3-18.
- [2] Schneider, D. Frété, C & Synteta, P. (2002) Community, Content and Collaboration Management Systems: socio-constructivist scenarios for the masses?, *Proceedings of ED-MEDIA 2002 World Conference on Educational Multimedia, Hypermedia &*

- Telecommunications*, June 24-29, 2002; Denver, Colorado, USA.
- [3] Reiser, R.A. Dempsey, J.V (2007). Trends and issues in instructional design and technology. Segunda Edición, Prentice Hall.
- [4] Rovai, A. P., Barnum, K. T. (2003). On-Line Course Effectiveness: An Analysis of Student Interactions and Perceptions of Learning. *Journal of Distance Education*. Vol. 18, No 1, 57-73
- [5] Lipponen, L.; Rahikainen, M.; Hakkarainen, K.; Palonen, T. (2002). *Effective participation and discourse through a computer network: Investigating elementary students' computer supported interaction*. *Journal of Educational Computing Research*, 27, (353-382).
- [6] I. H. Witten and E. Frank (2005). *Data Mining: Practical Machine Learning Tools and Techniques*, 2ed. Morgan Kaufmann.
- [7] Hammouda, K., Kamel, M. (2006). Data Mining in e-learning. *E-Learning Networked Environments and Architectures: A Knowledge Processing Perspective*, Samuel Pierre (Ed.), Springer Book Series: Advanced Information and Knowledge Processing (pp. 1-28).
- [8] Romero C, Ventura S. y García E. (2008) *Data Mining in course management systems: Moodle case study and tutorial*. *Computers & Education*. Volume: 51, (1), 368-384
- [9] Romero, C., Ventura, S., & Bra, P. D. (2004). Knowledge discovery with genetic programming for providing feedback to courseware author. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, 14(5), 425-464.
- [10] Romero, C., & Ventura, S. (2006). *Data mining in e-learning*, Southampton, UK: Wit Press.
- [11] Romero C y Ventura S. (2007) Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications* 33 (2007) 135-146
- [12] Romero, C., Porras, A.R., Ventura, S., Hervás, C., Zafra, A. (2006). Using sequential pattern mining for links recommendation in adaptive hipermedia educational systems. *In International Conference Current Developments in Technology-Assisted Education*. Sevilla, Spain (pp.1016-1020).
- [13] Jeong H. y Biswas G. (2008). *Mining Student Behavior Models in Learning-by-Teaching Environments*. The 1st International Conference on Educational Data Mining. 127-136.
- [14] Moodle. (2007) <http://moodle.org/>
- [15] Chenn-Jung Huang, Ming-Chou Liu, San-Shine Chu, Chih-Lun Cheng.(2007). An intelligent learning diagnosis system for Web-based thematic learning platform. *Computers & Education* 48 , 658-679
- [16] Stathacopoulou R., Grigoriadou M., Samarakou M., Mitropoulos D.(2007) Monitoring students' actions and using teachers' expertise in implementing and evaluating the neural network-based fuzzy diagnostic model. *Expert Systems with Applications* 32 , 955-975.
- [17] Grigoriadou M., Kornilakis H., Papanikolaou K. y Magoulas G. (2002) *Fuzzy Inference for Student Diagnosis in Adaptive Educational Hypermedia*. *Methods and Applications of Artificial Intelligence*. Lecture Notes in Computer Science, , Volume 2308/2002.
- [18] Mazza, R., Dimitrova, V. (2004). Visualising student tracking data to support instructors in web-based distance education. *In International world wide web conference*, New York, USA (pp. 154-161).
- [19] Gress, C. L.; Fior, M.; Hadwin, A. F.; Winne, P. H. (2007) *Measurement and assessment in computer-supported collaborative learning*. *Computers in Human Behavior*, DOI:10.1016/j.chb.2007.05.012.
- [20] Chen, G., Liu, C., Ou, K., Liu, B., (2000). Discovering decision knowledge from web log portfolio for managing classroom processes by applying decision tree and data cube technology. *Journal of Educational Computing Research* 23(3), 305-332.
- [21] VanLehn K. Student modeling. En *Foundations of Intelligent Tutoring Systems*, eds.: Polson M.C. y Richardson J.J. Lawrence Erlbaum. 1988.
- [22] Brusilovsky, P., Peylo, C., (2003). Adaptive and intelligent web-based educational systems. *International Journal of Artificial Intelligence in Education*, 13, 156-169.
- [23] Zadeh L.A (1965). Fuzzy sets. *Information and control* 8. p 338-353
- [24] Zadeh L.A.. The concept of a linguistic variable and its application to approximate reasoning . I-III *Information Sci* 8, , 199-250 y 301-357. 1975
- [25] Dubois D. y Prade H. What are fuzzy rules and how to use them. *Fuzzy sets and systems*, 84:169-185. 1996.
- [26] Alcalá, J. del Jesús, M.J., Garrell, J.M., Herrera, F., Hervás, C., Sánchez, L. (2004). Proyecto KEEL: Desarrollo de una Herramienta para el Análisis e Implementación de Algoritmos de Extracción de Conocimiento Evolutivos. *Tendencias de la Minería de Datos en España*. Eds. Giradles, J., Riquelme, J.C., Aguilar, J.S. (pp. 413-423).
- [27] J. Alcalá-Fdez · L. Sánchez · S. García · M. J. del Jesús · S. Ventura · J. M. Garrell · J. Otero · C. Romero · J. Bacardit · V. M. Rivas · J. C. Fernández · F. Herrera. (2009) KEEL: a software tool to assess evolutionary algorithms for data mining problems. *Soft Computing* 13:307-318.