

# Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basada en web

Cristóbal Romero Morales, Sebastián Ventura Soto, Cesar Hervás Martínez

Departamento de Informática y Análisis Numérico.  
Escuela Politécnica Superior. Universidad de Córdoba  
14071 Córdoba  
cromero@uco.es, sventura@uco.es, chervaz@uco.es

## Resumen

En este artículo se va a realizar un resumen del estado actual en la investigación sobre la aplicación específica de técnicas de minería de datos a los sistemas de enseñanza a distancia basados en web o sistemas de e-learning. Tanto la minería de datos como los sistemas de enseñanza basados en web son áreas muy de actualidad y con un gran crecimiento, por lo que su unión está despertando interés entre los investigadores y las empresas de ambas áreas. De entre los diferentes métodos y técnicas existentes de minería de datos, nos hemos centrado principalmente en la minería de utilización web, concretamente en clasificación y agrupamiento, descubrimiento de reglas de asociación y secuencias de patrones, ya que son las técnicas más utilizadas actualmente en sistemas de e-learning, aunque también se describen algunas otras técnicas diferentes como minería de contenidos y minería de texto, detección de anomalías, etc.

## 1. Introducción

Los sistemas basados en Web son cada vez más, la tecnología más utilizada para la educación a distancia, debido a la facilidad de utilización y disponibilidad de las herramientas para navegar por el Web y la facilidad del desarrollo y mantenimiento de los recursos Web. El desarrollo actual de los sistemas de enseñanza basada en Web o sistemas de e-learning [14], se ha incrementado exponencialmente en los últimos años y esto también ha motivado la aplicación de técnicas de minería de datos o descubrimiento de conocimiento [39] como herramientas para poder mejorar el aprendizaje en los sistemas de e-learning [38]. Estas técnicas de minería de datos o data mining ya se han aplicado con éxito en sistemas de comercio electrónico o e-commerce,

para comprender el comportamiento de clientes en línea de sistemas de comercio electrónico y poder incrementar las ventas [28]. Estas herramientas inteligentes utilizan técnicas de extracción de conocimiento o minería de datos para descubrir información útil para poder mejorar el sistema. Aunque los métodos de descubrimiento de información utilizados en ambas áreas (e-commerce y e-learning) son similares, los objetivos finales tienen matices totalmente diferentes debido a que en e-commerce el objetivo es guiar a los clientes durante la compra para maximizarla, mientras que en e-learning el objetivo es guiar a los estudiantes durante su aprendizaje para maximizarlo. Por lo tanto, cada uno tiene unas características específicas que requieren de un tratamiento diferente dentro del problema de minería de Web.

Las principales aplicaciones de las técnicas de minería de datos en educación, son como sistemas de personalización [28], sistemas recomendadores [15], sistemas de modificación [24], sistemas de detección de irregularidades [6], etc. debido a sus capacidades para [38]: el descubrimiento de patrones de navegación regulares e irregulares, realización de clasificaciones de alumnos y de los contenidos, construcción adaptativa de planes de enseñanza, descubrimiento de relaciones entre actividades, diagnóstico incremental de los estudiantes, etc.

La aplicación de técnicas de minería de datos en educación se puede ver desde dos puntos de vista u orientaciones distintas:

- **Orientado hacia los autores.** Con el objetivo de ayudar a los profesores y/o autores de los sistemas de e-learning para que puedan mejorar el funcionamiento o rendimiento de estos sistemas a partir de la información de utilización de los alumnos. Sus principales aplicaciones son: obtener una mayor realimentación de la enseñanza, conocer más

sobre como los estudiantes aprenden en el web, evaluar a los estudiantes por sus patrones de navegación, reestructurar los contenidos el sitio web para personalizar los cursos, clasificar a los estudiantes en grupos, etc.

- **Orientado hacia los estudiantes.** Con el objetivo de ayudar o realizar recomendaciones a los alumnos durante su interacción con el sistema de e-learning para poder mejorar su aprendizaje. Sus principales aplicaciones son: sugerir buenas experiencias de aprendizaje a los estudiantes, adaptación del curso según el progreso del aprendiz, ayudar a los estudiantes dando sugerencias y atajos, recomendar caminos más cortos y personalizados, etc.

También es importante indicar que aunque el área de la minería de datos aplicada a educación es muy reciente, ya cuenta con un número importante de investigadores, y muestra de ello son las múltiples contribuciones publicadas en diferentes congresos internacionales (ICCE, ICALT, ITS, Elearn, PAKDD, GECCO, UM, AH, WISE, ISDE, etc.) y revistas (IJEL, IEEE Education, UMUAI, etc.), y que su número se está incrementando enormemente cada año, indicando la gran importancia que esta teniendo y el interés que está despertando. También indicar que aunque la mayoría de las investigaciones se están realizando dentro de las Universidades, ya se están comenzando a interesarse algunas empresas (SPSS, TEMIS, Giunti Interactive Labs, etc.) dado idea de su potencial aplicación comercial.

El objetivo de este artículo es de presentar el estado de la investigación en esta área, para ello, primero se realiza una introducción a la enseñanza a distancia y a la minería de datos basada en web, a continuación se describe las principales técnicas de minería de datos aplicadas a e-learning y por último se presentan algunas conclusiones y las líneas más actuales de investigación.

## 2. Educación basada en web

El desarrollo de las nuevas tecnologías de la educación y la comunicación han hecho posible la utilización de Internet y más concretamente la WWW (World Wide Web) en la educación a distancia, dando lugar a la denominada Educación basada en Web o e-learning [14]. En la actualidad existen por todo el mundo miles y miles de

centros de enseñanza a distancia tanto públicos como privados, orientados a todos los niveles y tipos de educación: primaria, secundaria, superior, especial, adaptada, etc. Cada uno de ellos utiliza un sistema o plataforma de enseñanza basado en web [9] que puede ser: o bien un sistema propio desarrollado específicamente por ellos mismos, o bien uno de los múltiples sistemas comercial existentes como: Web-CT, Virtual-U, TopClass, etc. o de libre distribución como: ATutor, ILIAS, Moodle, etc. Estos sistemas proporcionan servicios útiles para la enseñanza a distancia como son herramientas para la comunicación sincrónica y asíncrona, herramientas para la gestión de materiales de aprendizaje y herramientas para la gestión, seguimiento y evaluación de los estudiantes.

Un problema de la mayoría de los sistemas para la enseñanza basados en web anteriores es que los cursos que proporcionan no son más que una red de páginas web estáticas a través de las que navegan todos los estudiantes. Para solucionar este problema se han desarrollado los Sistemas Hipermedia Adaptativos Basados en Web [8] que son un nuevo tipo de sistemas educativos que provienen de la evolución de los Sistemas Tutores Inteligentes (STI) y de los Sistemas Hipermedia Adaptativos (SHA), y que comparten con ellos características tales como: aumento de la interacción con los usuarios y adaptación de los contenidos a las necesidades de estos. Para ello, construyen un modelo del alumno y lo utilizan durante la interacción con dicho usuario para adaptarse a sus necesidades. Algunos ejemplos de Sistemas Hipermedia Adaptativos basados en Web para educación [8] son: Interbook, DCG, ELM-ART, CALAT, AHA!, etc.

Por último indicar también la existencia y el incremento en la utilización de múltiples estándares de e-learning [23]: IMS, ADL SCORM, AICC, IEEE LTSC, etc. que además de permitir la interoperabilidad entre distintos sistemas, permiten la reutilización de contenidos educativos, y también facilitan la incorporación de diferentes técnicas adaptativas.

### 3. Minería de datos web

La minería de datos es el proceso de descubrimiento de conocimiento para encontrar información no trivial, previamente desconocida y potencialmente útil de grandes repositorios de datos [Klosgen]. La minería de datos es un área multidisciplinar donde convergen diferentes paradigmas de computación como son la construcción de árboles de decisión, la inducción de reglas, las redes neuronales artificiales, el aprendizaje basado en instancias, aprendizaje bayesiano, programación lógica, algoritmos estadísticos, etc. Las principales tareas y métodos de la minería de datos son: clasificación, agrupamiento, estimación, modelado de dependencias, visualización y descubrimiento de reglas. La minería de datos es en realidad uno de los pasos que comprenden el proceso de descubrimiento de conocimiento, que está compuesto por:

- **Preprocesamiento.** Consiste en la recogida o extracción de los datos, limpieza de datos, discretización, selección de los atributos e integración de datos.
- **Minería de datos.** Consiste en la selección de los algoritmos de minería de datos a utilizar y la aplicación de dichos algoritmos sobre los datos.
- **Postprocesamiento.** Consiste en la interpretación, evaluación de los resultados obtenidos y la utilización del conocimiento descubierto.

Un caso particular de la minería de datos es la minería de Web o web mining [27], que como el propio nombre indica consiste en la aplicación de técnicas de minería de datos para extraer conocimiento a partir de datos de la Web. Se pueden distinguir tres tipos de minería de Web:

- **Minería de contenidos web.** Es el proceso de extraer información a partir de los contenidos de los documentos Web.
- **Minería de estructura web.** Es el proceso de descubrir información a partir de la estructura de la Web.
- **Minería de utilización web.** Es el proceso de descubrir información a partir de los datos de utilización de la Web.

De los estos tres tipos de minería de Web, el que más se ha utilizado para el descubrimiento de información en los sistemas de enseñanza basada

en web es la minería de utilización Web o web usage mining.

### 4. Técnicas de minería web más utilizadas en sistemas de e-learning

Las técnicas más utilizadas en la minería de datos aplicada a los sistemas de e-learning son: clasificación y agrupamiento, descubrimiento de reglas de asociación, y análisis de secuencias. A continuación, se van a detallar los principales trabajos de investigación agrupados dentro de estos tres tipos de técnicas, aunque algunos de los investigadores no sólo utilizan una única técnica sino varias.

#### 4.1. Clasificación y agrupamiento

Las técnicas de clasificación y agrupamiento o clustering consisten en [3] la habilidad intelectual para ordenar o dividir fenómenos complejos (descritos por conjuntos de objetos descritos por datos altamente dimensionales) en pequeños y comprensibles unidades o clases que permiten un mejor control o comprensión de la información. Su aplicación a sistemas de e-learning permite agrupar a los usuarios por su comportamiento de navegación, agrupar a las páginas por su contenido, tipo o acceso y agrupar los comportamientos de navegación similares. A continuación describimos algunos trabajos de aplicación de minería de datos en e-learning.

La utilización de técnicas de agrupamiento es utilizada por Gord McCalla y Tiffany Tang [31] para formar clusters o grupos de usuarios basándose en su comportamiento de navegación. Utilizan un algoritmo de clustering basado en largas secuencias generalizadas. También proponen incluir un sistema recomendador inteligente dentro de un sistema de aprendizaje basado en web evolutivo capaz de adaptarse no sólo a sus usuarios sino también a la Web abierta. El sistema puede encontrar contenidos relevantes en la web y puede personalizar y adaptar sus contenidos basándose en observaciones del sistema y por las propias valoraciones acumuladas dadas por los estudiantes. Otro trabajo que también emplea agrupamiento es el realizado por Elena Gaudioso y Luis Talavera [32] que analizan los datos obtenidos de cursos basados en sistemas e-learning y utilizan técnicas de clustering

similares al modelo probabilístico de Naive Bayes para descubrir patrones que reflejan comportamientos de los usuarios. Su objetivo es utilizar la minería de datos para dar soporte a la tutoría en comunidades de aprendizaje virtual.

La utilización conjunta de clustering con otras técnicas como secuenciación es realizada por Julia Miguillón y Enric Mor [20] para analizar el comportamiento de navegación de los usuarios para la personalización de e-learning. Utilizan clustering de estudiantes para intentar extender las capacidades de secuenciación de algunos sistemas estándares de manejo de aprendizaje como SCORM para incluir el concepto de itinerario recomendado. Los autores Erkki Sutinen y otros [29] proponen un modelo híbrido que combina técnicas de minería de datos y de aprendizaje de máquinas para la construcción de una red bayesiana para describir el proceso de aprendizaje de los estudiantes. Su objetivo es clasificar a los estudiantes para poder ofrecerles diferentes guías dependiendo de sus habilidades y otras características. Esta tarea se realiza con la categorización y clustering de los estudiantes dependiendo de sus habilidades o conocimiento. Finalmente el trabajo realizado por Jing Luan [16] utiliza técnicas de minería de datos en educación superior y propone la utilización conjunta de predicción y clustering dentro de una herramienta de soporte de decisiones permitiendo a la universidad anticiparse a las necesidades de los estudiantes.

#### 4.2. Reglas de asociación

Las reglas de asociación [1] descubren relaciones entre atributos de un conjunto de datos que superan unos determinados umbrales. Su aplicación más típica ha sido en los sistemas de comercio electrónico para informar sobre las preferencias de compra de los clientes. Su aplicación a sistemas de e-learning permite descubrir relaciones o asociaciones entre distintas páginas Web visitadas. A continuación se describen algunos trabajos de aplicación de minería de datos en e-learning que utilizan esta técnica.

Uno de los pioneros de la utiliza técnicas de minería web en sistemas de e-learning es Osmar Zaïne [38] y actualmente propone utilizar agentes recomendadores [37] para recomendar actividades de aprendizaje en línea o atajos en un curso web

basándose en los historiales de acceso y mejorar el proceso de aprendizaje en línea. Concretamente utiliza minería de reglas de asociación para entrenar al agente recomendador y construir un modelo que representa el comportamiento de acceso o asociaciones entre actividades de aprendizaje en línea. Otro trabajo que analiza los ficheros log de entornos de aprendizaje web utilizando técnicas de minería de reglas de asociación y filtrado colaborativo, es el realizado por Feng-Hsu Wang [34] para descubrir patrones de navegación útiles y proponer un modelo de navegación. El modelo de navegación consiste en dos tipos de relaciones: relaciones de asociación y relaciones de secuencia entre documentos.

La utilización de métodos heurístico como regresión lineal en combinación con reglas de asociación es propuesta por Freyberger y otros [10] para buscar buenos modelos de transferencia de aprendizaje de estudiantes a partir de los ficheros logs de interacción de sistemas tutores inteligentes. El modelo de transferencia relaciona preguntas con los componentes del conocimiento necesarios para responderlas correctamente. Las reglas de asociación guían el proceso de búsqueda para encontrar modelos de transferencia que predicen el éxito de los estudiantes

También se están utilizando técnicas de softcomputing, por ejemplo Pao-Ta Yu y otros [36] proponen la utilización de reglas de asociación difusas para descubrir relaciones entre patrones de comportamiento de los estudiantes, incluyendo el tiempo de acceso, números de páginas leídas, preguntas contestadas, mensajes leídos y enviados, etc. Mediante monitorización y análisis el sistema indicará mediante umbrales de alarma los resultados a los estudiantes y al instructor. Un trabajo más orientado a los sistemas hipermedia adaptativos es el de Romero y Ventura [25] que utilizan minería de reglas como técnica de descubrimiento de información útil para los autores de este tipo de cursos con el objetivo de poder realizar mejoras tanto del contenido, como de la estructura de los cursos y de su adaptación. Concretamente proponen la utilización de algoritmos evolutivos multiobjetivo para el descubrimiento de relaciones importantes a partir de los datos de utilización (tiempos de acceso a páginas, aciertos y fallos, y niveles de conocimiento) de los estudiantes. Otro trabajo que también emplea algoritmos evolutivos es el realizado por Behrouz Minaei-Bidgoli y William

F. Punch [17] para realizar un análisis de asociación para predecir el rendimiento de los estudiantes. Utilizan clustering de recursos web valorados y descubrimiento de reglas de asociación interesantes mediante algoritmos genéticos para optimización de minería de datos [18] con el objetivo es clasificar a los estudiantes para predecir su clasificación final basándose en las características extraídas de los ficheros logs.

#### 4.3. Análisis de secuencias

El análisis de secuencias o secuencia de patrones [2] es una técnica de minería de datos que descubre secuencias dentro de un conjunto de datos. Al igual que las reglas de asociación, también se han aplicado en sistemas de comercio electrónico para descubrir secuencias de acciones de los clientes. Su aplicación a sistemas de e-learning permite analizar secuencias de páginas visitadas durante una sesión o en distintas sesiones de un mismo usuario. A continuación se describen algunos trabajos de aplicación de minería de datos en e-learning que utilizan esta técnica.

El análisis de patrones de navegación en entornos de aprendizaje basado en web es utilizado por Karin Becker y otros [7] dentro de una herramienta de minería de utilización web para el análisis de patrones y preprocesado de datos de utilización de entornos de aprendizaje basados en web. Las técnicas de descubrimiento de patrones utilizadas son asociación y secuencia, y la secuencia de patrones describen accesos a páginas relaciones en un orden específico. Otro propuesta de análisis de caminos es el realizado por Sung Ho Ha y otros [12] que utilizan el análisis de caminos transversales de páginas web para la personalización de la enseñanza y la asociación de páginas web para la estructuración de conocimiento. El descubrimiento de caminos se realiza desde dos puntos de vista: caminos individuales y caminos agregados que incluye el proceso de clustering utilizando mapas autoorganizativos de Kohonen.

El análisis de los patrones de comportamiento es realizado por Claus Pahl [22] para el análisis de la interacción en sistemas e-learning. Para ello utiliza técnicas de minería Web como clasificación, patrones secuenciales y series temporales para el análisis y la evaluación de las interacciones de los estudiantes con los contenidos de los sistemas de

e-learning. El análisis de los patrones de comportamiento permite describir los caminos de aprendizaje ideales dentro de la topología de un curso. Otro trabajo sobre el análisis y la minería de la información sobre el comportamiento del aprendizaje en entornos basados en SCORM es el realizado por Wei Wang y otros [35] para ayudar a los profesores a comprender los motivos del buen aprendizaje o no de los alumnos. Utilizan primero técnicas de minería de secuencias de patrones, posteriormente agrupan a los estudiantes mediante clustering y finalmente generan un árbol de actividades personalizadas.

#### 5. Otras técnicas de minería de datos aplicadas en e-learning

Existen otras técnicas distintas de minería de datos utilizada en sistemas de e-learning como minería de contenidos y minería de texto, detección de irregularidades, razonamiento basado en casos, etc. A continuación se van a describir algunas investigaciones que utilizan este tipo de técnicas.

La utilización de minería de contenido es utilizada por Qing Li y otros [30] en la construcción automática de libros electrónicos a partir de una jerarquía de temas especificados por el usuario. Para ello utilizan técnicas de minería de contenido para identificar el tipo de página y el tema del contenido, y técnicas de minería de secuencia de patrones frecuentes para descubrir nuevos conceptos en las páginas web.

La utilización de minería de texto o text mining es utilizado por Stefan Geibler y Roberto Vaccaro [11] para la extracción automática de información e indexado de componentes dentro del contexto de manejo de conocimiento de e-learning. El resultado es un conjunto útil de de metadatos automáticamente extraídos de recursos de tipo texto para dar un soporte rápido y exacto a la recuperación y el indexado. Otra propuesta de utilización de minería de texto en educación y publicidad es la de Dunja Mladenic y otros [19] para potenciar una aplicación de construcción de ontologías/taxonomía a partir de un conjunto de documentos planos, realizar búsquedas en la base de documentos y tratar problemas específicos del lenguaje.

La detección de irregularidades es utilizada por Maomi Ueno [33] para detectar en línea irregularidades en perfiles de procesos de

aprendizaje a partir de los tiempos de respuesta de los contenidos de sistemas e-learning. El método de detección de alumnos está basado en distribuciones predictivas bayesianas. El profesor y que utiliza el profesor para asistir a los alumnos que presenten alguna dificultad. Otros autores que utilizan la detección del mal uso por parte de los estudiantes de sistemas tutores inteligentes son Baker y otros [5] para detectar comportamientos de obtención de respuestas correctas y avances en el currículo utilizando la ayuda y las realimentaciones. Para ello utilizan un modelo de aprendizaje de máquinas de respuesta latente para identificar los alumnos que juegan o hacen un mal uso del sistema y poder darles una respuesta apropiada para que no sigan hacia un pobre aprendizaje.

La inferencia de variables de aprendizaje no observables a partir de los comportamientos recogidos de los estudiantes al utilizar un sistema tutor inteligente es propuesta por Arroyo y otros [4] utilizando una red bayesiana de correlaciones entre actitudes, percepciones y comportamientos de los estudiantes para poder inferir las actitudes positivas y negativas de los estudiantes a partir de sus comportamientos, con el objetivo de poder responder de forma adaptada a las necesidades de los estudiantes.

La utilización de técnicas de minería de datos y de razonamiento basado en casos para enseñanza a distancia es propuesta por Ruimin Shen y otros [26] que además de utilizar algoritmos para clasificar a los estudiantes en clases dependiendo de sus acciones y de descubrir reglas de asociación entre diferentes puntos de conocimiento, utilizan razonamiento basado en casos para personalizar la interacción. Los casos son concretamente un conjunto de preguntas y sus correspondientes respuestas, de forma que ante una pregunta de un alumno, se seleccionan las N preguntas más similares y se le proporciona las respuestas al alumno. Otro trabajo que utiliza razonamiento basado en casos es el de Heraud [13] que propone la generación de la estructura de enlaces de un entorno de aprendizaje adaptativo basado en web utilizando los logs de la interacción de los estudiantes. En concreto descompone los logs en episodios de aprendizaje y utiliza razonamiento basado en casos para ofrecer a los alumnos una ayuda contextual con una estructura de enlaces adaptados del curso.

## 6. Conclusiones

A lo largo de este artículo, se han descrito las principales técnicas de minería de datos aplicadas a entornos de enseñanza basada en Web y se han presentado las principales investigaciones y trabajos desarrollados. Como se ha podido ver la gran mayoría de estos trabajos se encuentran dentro del área de la minería de utilización web, realizando clasificación y agrupamiento, descubrimiento de reglas de asociación o secuencias de patrones. Aunque también se están comenzando a emplear otras técnicas de minería de datos como minería de texto, detección de irregularidades y razonamiento basado en casos. Todas estas aplicaciones se pueden diferenciar en dos grupos distintos dependiendo del tipo de sistemas de enseñanza sobre el que se aplica, por un lado a los entornos de enseñanza a distancia basados en web tradicionales, es decir, sistemas de e-learning que no incorporan técnicas de inteligencia artificial en los cursos [38] y por otro lado los Sistemas Hipermedia Adaptativos Basados en Web para Educación y Sistemas Tutores Inteligentes basados en Web [21].

La minería de datos aplicada a sistemas tradicionales de e-learning suele utilizar solamente la información proporcionada por los ficheros log capturados por los servidores web y las preferencias personales de los estudiantes. Existe una relación con los sistemas de comercio electrónico al utilizar la misma fuente de datos y suelen utilizar las mismas técnicas de minería de datos pero adaptadas a los entornos educativos.

Por otro lado la minería de datos aplicada a SHA y STI para educación basados en Web, suele disponer de mayor información en los ficheros logs de la interacción entre el estudiante y el sistema sobre actividades realizadas, aciertos y fallos, notas y niveles de conocimiento, etc. Además la capacidad de estos sistemas para la enseñanza adaptada, hacen que la aplicación de minería de datos se diferencie más de la aplicada en comercio electrónico, al estar mucho más orientados al problema concreto del proceso de aprendizaje de los estudiantes.

Finalmente, algunas de las líneas de investigación futuras o donde más se está trabajando en la actualidad sobre la aplicación de minería de datos en educación basada en web son:

- Facilidad de utilización de los algoritmos de minería de datos. Desarrollo de herramientas

más fáciles e intuitivas de utilizar, orientadas para ser utilizadas por personas no expertas en minería de datos, sino en educación.

- Integración de algoritmos de minería de datos dentro de las propias herramientas autor de construcción y mantenimiento de los cursos, para la mejora automática de los sistemas.
- Integración de algoritmos de minería en los interfaces de usuario de los cursos y de las plataformas dentro de sistemas agentes recomendadores, para realización de sugerencias sobre rutas, actividades, etc.
- Desarrollo de algoritmos de minería de datos específicos para problemas relacionados con la enseñanza y el aprendizaje a distancia utilizando entornos hipermedia adaptativos y sistemas tutores inteligentes basados en web.
- Aplicación de nuevas técnicas de minería de datos como minería de semántica web, minería de datos multimedia, minería de datos distribuida, etc.

### Agradecimientos

Este trabajo ha sido financiado por el MCYT a través del proyecto TIC2002-04036-C05-02 y de fondos FEDER.

### Referencias

- [1] Agrawal, R.; Imielinski, T.; Swami, A. Mining association rules between sets of items in large databases. ACM SIGMOD Conference on Management of Data. pp. 207-216. 1993.
- [2] Agrawal, R.; Srikant, R. Mining sequential patterns. Int. Conf. Data Engineering, pp. 3-14. Taiwan. 1995.
- [3] Arabie, P.; Hubert, J.; De Soete, G. Clustering and Classification. World Scientific Publishers. 1996.
- [4] Arroyo, I.; Murray, T.; Woolf B.P.; Beal, C.R. Inferring Unobservable Learning Variables from Students' Help Seeking Behavior. Int. Conf. on Intelligent Tutoring Systems. pp 782-784. 2004.
- [5] Baker, R.S.; Corbett, A.T.; Koedinger, K.R. Detecting Student Misuse of Intelligent Tutoring Systems. Int. Conf. on Intelligent Tutoring Systems. Pp. 531-540. 2004.
- [6] Barnett, V.; Lewis, T.; Outliers in Statistical Data. John Wiley & Sons. 1994.
- [7] Becker, K.; Marquardt, C.G.; Ruiz, D.D. A Pre-Processing Tool for Web Usage Mining in the Distance Education Domain. pp. 78-87. 2004.
- [8] Brusilovsky, P. Adaptive and Intelligent Technologies for Web-based Education. Special Issue on Intelligent Systems and Teleteaching, Künstliche Intelligenz, 4, 19-25. 1999.
- [9] Flate, M. Online education and learning management systems. Global e-learning in a Scandinavian perspective. Oslo. NKI Forlaget. 2003.
- [10] Freyberger, J.; Heffernan, N.T.; Ruiz, C. Using Association Rules to Guide a Search for Best Fitting Transfer Models of Student Learning. Int. Conf. on Intelligent Tutoring Systems. 2004.
- [11] GeiBler, S.; Vaccaro, R.; Automatic Information Extraction and Indexing in the context of E-Learning Knowledge Management. Int. conf. on Data Mining, Text Mining and their Business Applications. Greece. 2005.
- [12] Ha, S.H.; Bae, S.M.; Park, S.C. Web mining for distance education. APAN Conference. Beijing. 2000.
- [13] Heraud, J.M.; A Web Adaptive Learning Environment where the link structure is generated by experience reuse. Int. Conf. on Intelligent Tutoring Systems. Brazil. 2004.
- [14] Horton, W. Designing Web-Based Training. John Wiley&Sons. 2000.
- [15] Li, J.; Zaiane, O.R. Combining Usage, Content and Structure Data to Improve Web Site Recommendation. Int. Conf. on Electronic Commerce and Web Technologies. Spain. 2004.
- [16] Luan, J. Data Mining, Knowledge Management in Higher Education, Potential Applications. Workshop Associate of Institutional Research International Conference. Toronto. 2002.
- [17] Minaei-Bidgoli, B.; Punch, W.F. Predicting student performance: an application of data mining methods with the educational web-based system LON-CAPA. IEEE Frontiers in Education. Pp 1-6. 2003.
- [18] Minaei-Bidgoli, B.; Punch, W.F. Using Genetic Algorithms for Data Mining Optimization in an Educational Web-Based System. GECCO 2003: 2252-2263.

- [19] Mladenic, D.; Grobelnik, M.; Jermol, M. Exploiting Text Mining in Publishing and Education. Workshop on Data Mining Lessons Learned. ICML. Pp. 34-39. 2002.
- [20] Mor, E.; Minguillón J. E-learning Personalization based on Itineraries and Long-term Navigational Behavior. World Wide Web Conference. pp. 264-265. New York. 2004.
- [21] Mostow, J. Some useful design tactics for mining ITS data. Workshop on Analyzing Student-Tutor Interaction Logs to Improve Educational Outcomes. Int. Conf. on Intelligent Tutoring Systems. 2004.
- [22] Pahl, C. Data Mining Technology for the Evaluation of Learning Content Interaction. International Journal on E-Learning IJEL. 2004.
- [23] Paramythis, A.; Loidl-Reisinger, S. Adaptive Learning Environments and e-Learning Standards. Electronic Journal of e-Learning, 2 (1), pp. 181-194. 2004.
- [24] Perkowitz, M.; Etzioni, O. Adaptive web sites: Automatically synthesizing web pages. National Conference on Artificial Intelligence. WI. 1998.
- [25] Romero, C.; Ventura, S.; de Bra, P. Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware Author. User Modeling and User-Adapted Interaction. Vol. 14. No. 5. 2005.
- [26] Shen, R.; Han, P.; Yang, F. Data Mining and Case-based Reasoning for Distance Learning. International Journal of Distance Education Technologies, Vol. 1(3), pp. 46-58. 2003.
- [27] Scime, A. Web Mining: Applications and Techniques. Idea Group. 2004.
- [28] Srivastava, J.; Mobasher, B.; Cooley, R. Automatic Personalization Based on Web Usage Mining. Communications of the Association of Computing Machinery. pp. 142-151. 2000.
- [29] Sutinen, E.; Hämäläinen, W.; Suhonen, J.; Toivonen, H. Data Mining in Personalizing Distance Education Courses. Conference on Open Learning and Distance Education. Hong Kong. 2004.
- [30] Tang, C.; Lau, R.; Li, Q.; Yin, H.; Li, T.; Kilis, D. Personalized Courseware Construction based on Web Data Mining. pp. 204-211. 2000.
- [31] Tang, T. Y.; McCalla, G. Evaluating A Smart Recommender for an Evolving E-Learning System. Proceedings of Canadian Artificial Intelligence Conference. Canada. 2004.
- [32] Talavera, L.; Gaudioso, E. Mining student data to characterize similar behavior groups in unstructured collaboration spaces. Workshop on Artificial Intelligence in CSCL. ECAI. pp. 17-23. 2004.
- [33] Ueno, M.; Online Outlier Detection System for Learning Time Data in E-Learning and Its Evaluation. Pp. 248-253. 2004.
- [34] Wang, F. On Analysis and Modeling of Student Browsing Behavior in Web-Based Asynchronous Learning Environments. Int. Conf. on Web-based Learning. Pp. 69-80. 2002.
- [35] Wang, W.; Weng, J.; Su, J.; Tseng, S. Learning Portfolio Analysis and Mining in SCORM Compliant Environment. IEEE Frontiers in Education Conference. 2004.
- [36] Yu, P.; Own, C.; Lin, L. On the Learning Behavior Analysis of Web Based Interactive Environment. ICCE. 2001.
- [37] Zaiane, O.R. Building a Recommender Agent for e-Learning Systems. International Conference on Computers in Education. New Zealand. pp 55-59. 2002.
- [38] Zaiane, O.Z. Web Usage Mining for a Better Web-Based Learning Environment. Conference on Advanced Technology for Education. pp 60-64. Alberta. 2001.
- [39] Zytlow J.; Klosgen W., Handbook of Data Mining and Knowledge Discovery. Oxford University Press. 2001.