

Las Técnicas de Clustering en la Personalización de Sistemas de e-Learning

Roberto A. Farias, Elena B. Durán, Saritha G. Figueroa

Facultad de Ciencias Exactas y Tecnologías (FCEyT) – Universidad Nacional de Santiago del Estero (U.N.S.E.)
Santiago del Estero, Argentina

E-mail: fariasroberto@gmail.com, eduran@unse.edu.ar, sarithaf@unse.edu.ar

Abstract

One of the principal problems of e-learning is that many of the students that begin a course are frustrated, they do not feel motivated, they are not able to assimilate the knowledge and, then, they abandon it. One of the main reasons is that most courses are designed without keeping in mind that the students learn in different way, because each one possesses a learning style that characterizes him. Therefore, it is desirable that an e-learning system presents certain degree of adaptability that allows it to personalize teaching strategies according to the preferences and necessities of each student; and particularly, to the student's learning style.

A method based on Cluster Analysis to detect the dominant student's learning style is proposed in this article. This method considers student's interactions with the system to recognize, not only his dominant learning style, but also the changes in this style along the e-learning course. The method was evaluated, with favorable results, in the frame of an e-learning course for student of Computer Science.

Key Words: e-learning, personalization, learning styles, clustering analysis.

Resumen

Uno de los grandes problemas de la educación a distancia soportada por computadora (e-learning) radica en que muchos de los estudiantes que inician un curso, se frustran, se sienten desmotivados, no logran asimilar los conocimientos y en consecuencia lo abandonan. Una de las principales razones es que la mayoría de los cursos se diseñan sin tener en cuenta que los estudiantes aprenden de diferente manera, ya que cada uno posee un estilo de aprendizaje que lo caracteriza. Por lo tanto, es deseable que un sistema de e-learning presente un cierto grado de adaptabilidad que le permita personalizar la estrategia de enseñanza de acuerdo a las preferencias y necesidades de cada alumno; y en particular, al estilo de aprendizaje que presenta cada estudiante.

En este artículo se propone un método basado en técnicas de Análisis de Cluster para detectar el estilo de aprendizaje dominante del estudiante. Este método considera las interacciones del alumno con el sistema para poder reconocer, no sólo su estilo de aprendizaje dominante, sino también los cambios en este estilo a lo largo del curso de e-learning. El método fue evaluado, con resultado favorable, en el marco de un curso a distancia para alumnos de Informática.

Palabras Clave: e-learning, personalización, estilo de aprendizaje, análisis de cluster.

1. INTRODUCCIÓN

Actualmente, la tecnología informática y de comunicaciones ha provocado un cambio radical en la forma de relacionarnos y conseguir información, pero en los sistemas educativos esto no es suficiente. Tenemos que ser capaces no sólo de transmitir información, sino también de lograr la asimilación efectiva de conocimiento.

Las aplicaciones e-learning actuales no se adaptan completamente a las necesidades y a los estilos de aprendizaje del alumno, por lo que se producen soluciones que no facilitan el proceso de aprendizaje del estudiante. Por consiguiente, es muy importante brindar a los cursos de e-learning la capacidad de personalizar sus contenidos de acuerdo a las preferencias y necesidades de los estudiantes, lo cual sólo es posible si se mantiene un modelo del alumno, que es la imagen que el sistema tiene del estudiante. Esta capacidad de adaptación hace que las aplicaciones e-learning puedan contribuir en los nuevos procesos de enseñanza; ya que al estar basados en modelos centrados en el estudiante, favorecen un aprendizaje significativo y activo; por esto han tenido gran difusión en el campo de la educación asistida por computación.

Para poder llevar a cabo la personalización en un contexto educativo, se necesita adquirir cierta comprensión del estudiante como así también de las tareas que son importantes para el aprendizaje [20]. Es en este contexto que adquiere singular importancia el modelo del alumno que se define como el componente de un sistema inteligente que representa el estado actual de un estudiante [30], y más precisamente como una representación de algunas características, tales como conocimiento a priori del dominio (background), experiencia, estilos de aprendizaje, estilo cognitivo, etc.

Diversas investigaciones [2; 11; 3; 22; 29; 26; 28; 20; 21; 24] han demostrado que los estudiantes aprenden con más efectividad cuando se les enseña según sus estilos de aprendizaje predominantes. Según Felder [6], se debe ser consciente de las diferencias que tienen los estudiantes para procesar la información, con el fin de poder ofrecer materiales pedagógicos dinámicos adaptados a preferencias particulares de aprendizaje. Por lo tanto, un buen curso de e-learning, deberá tener en cuenta que hay que proporcionar diferentes vías para que todos los estudiantes con distintos estilos de aprendizaje puedan escoger su propio camino.

Existen dos alternativas para identificar el estilo de aprendizaje de un alumno. La primera se lleva a cabo por medio de una prueba o test inicial que realiza el alumno al comienzo del curso. El problema de estas pruebas es que los alumnos tienen una tendencia a escoger respuestas arbitrariamente, debido a que los cuestionarios son demasiados extensos, o no se dan cuenta de las consecuencias o los usos futuros de los mismos. Además, los resultados obtenidos pueden ser inexactos y pueden no reflejar los estilos reales de aprendizaje. Por otro lado, el estilo de aprendizaje, una vez identificado, no cambia hasta la próxima vez que el alumno realice nuevamente el test. La segunda alternativa consiste en la aplicación de técnicas de aprendizaje de máquina, que permiten que los estilos de aprendizaje del alumno se mantengan actualizados a lo largo del tiempo. En definitiva, se trata de implementar técnicas de aprendizaje de máquina para reconocer en los cursos de e-learning los estilos de aprendizaje de cada uno de los alumnos y, a partir de allí adaptar la estrategia de enseñanza.

El presente artículo documenta la investigación realizada con el fin de detectar automáticamente el estilo de aprendizaje de los estudiantes. Para ello se propone un método basado en técnicas de Análisis de Cluster aplicable en entornos e-learning. El Análisis de Cluster se emplea en este trabajo para agrupar comportamientos similares mientras el alumno interactúa con el Sistema. Los datos suministrados al algoritmo de clustering se obtienen a partir de los registros de transacciones (log), que contienen las tareas ejecutadas por los estudiantes en el sistema y su participación en actividades tales como salas de chat y foros. Una vez obtenidos los clusters, se espera que el cluster

dominante determine el estilo de aprendizaje del estudiante el cual será enmarcado dentro de las cuatro dimensiones propuestas por el modelo de aprendizaje de Felder y Silverman [7].

En las siguientes secciones se describen algunos aspectos de los estilos de aprendizaje, y en particular, se presenta el modelo de estilos de aprendizaje de Felder y Silverman. Luego, se realiza una breve introducción a las técnicas de clustering, y se citan algunos trabajos relevantes relacionados con el tema. A continuación, se describe el método propuesto para detectar el estilo de aprendizaje dominante en cada alumno. Seguidamente, se describe la experiencia realizada para evaluar la precisión del método y se documentan los resultados obtenidos. Finalmente, se presentan las conclusiones del trabajo realizado.

2. LOS ESTILOS DE APRENDIZAJE

Todos los estudiantes aprenden de forma diferente, es decir que cada uno tiene sus propios métodos o estrategias de aprendizaje, diferentes trasfondos de conocimientos, fortalezas y debilidades que le son propias, niveles de motivación distintos, etc. Estas diferencias en la forma de aprender son las que definen el estilo de aprendizaje de cada persona.

Existen un gran número de definiciones para estilos de aprendizaje. Keefe [16] define: “los estilos de aprendizaje son los rasgos cognitivos, afectivos y fisiológicos que sirven como indicadores relativamente estables, de cómo los alumnos perciben interacciones y responden a sus ambientes de aprendizaje”.

En 1988, Richard Felder y Linda Silverman [7] formularon un modelo de estilo de aprendizaje diseñado para capturar la mayoría de las diferencias importantes de estilos de aprendizaje en alumnos de ingeniería y proveer una buena base a los profesores para formular métodos de enseñanza dirigidos a las necesidades de todos los estudiantes. En este modelo, los estilos de aprendizaje, se clasifican en cuatro dimensiones que se relacionan con las respuestas a cuatro preguntas respectivamente:

1. *Dimensión percepción:* ¿Qué tipo de información perciben preferentemente los estudiantes? Básicamente, los estudiantes perciben dos tipos de información: información externa o sensitiva a la vista, al oído o a las sensaciones físicas; e información interna o intuitiva a través de memorias, ideas, lecturas, etc. En esta dimensión se consideran los estilos extremos de sensitivos – intuitivos.
2. *Dimensión entrada:* ¿A través de qué modalidad sensorial se percibe más efectivamente la información cognitiva? Los estudiantes pueden recibir la información externa en formatos visuales, mediante cuadros, diagramas, gráficos, demostraciones, etc.; o en formatos verbales, mediante sonidos, expresión oral y escrita, fórmulas, símbolos, etc. En esta dimensión se consideran los estilos extremos de visuales – verbales.
3. *Dimensión procesamiento:* ¿Cómo progresa el estudiante en su aprendizaje? El progreso de los estudiantes sobre el aprendizaje implica un procedimiento secuencial que necesita progresión lógica de pasos incrementales pequeños, o entendimiento global que requiere de una visión integral. En esta dimensión se consideran los estilos extremos de secuenciales – globales.
4. *Dimensión comprensión:* ¿Con qué tipo de organización de la información está más cómodo el estudiante a la hora de trabajar? La información se puede procesar mediante tareas activas, a través de compromisos en actividades físicas o discusiones; o a través de la reflexión o introspección. En esta dimensión se consideran los estilos extremos de activos – reflexivos.

Tal como se puede apreciar, cada dimensión se extiende entre dos polos opuestos, pero también se contemplan los estilos balanceados; es decir, aquellos que comparten características de ambos extremos.

En esta investigación se ha seleccionado este modelo de estilo de aprendizaje por considerarlo el más apropiado para estudiantes de Ciencias de la Computación, población con la que se realizó la evaluación del método propuesto.

3. LAS TÉCNICAS DE CLUSTERING

Se denomina clustering al “proceso de agrupar un conjunto de objetos físicos o abstractos en clases de objetos similares” [15]. Un cluster es una colección de objetos de datos que son similares a otros dentro del mismo cluster y son distintos a los objetos de otros clusters. El clustering es un ejemplo de aprendizaje no supervisado en el cual se usan métodos que intentan encontrar particiones naturales de patrones [19]. Se trata de una técnica descriptiva dentro de la Minería de Datos (MD), muy útil para identificar patrones que explican o resumen los datos mediante la exploración de las propiedades de los mismos.

En la actualidad, el análisis de clustering ha jugado un rol muy importante en una amplia variedad de áreas tales como: reconocimiento de patrones, análisis de datos espaciales, procesamiento de imágenes, cómputo y multimedia, análisis médico, economía, bioinformática y biometría, principalmente. Esto ha hecho posible que el análisis de clustering se considere como una de las mejores técnicas para obtener conocimiento y realizar exploraciones en los datos. Mediante clustering, es posible identificar una región densa y esparcida y, por consiguiente, descubrir patrones de distribución global y correlaciones interesantes entre patrones de datos [15].

Un problema de análisis de clustering parte de un conjunto de casos u objetos cada uno de los cuales está caracterizado por un grupo de variables. A partir de esta información se trata de obtener grupos de objetos, de tal manera que los objetos que pertenecen a un grupo sean similares entre sí y diferentes entre los distintos grupos. Expresado en términos de variabilidad, hablaríamos de minimizar la variabilidad dentro de los grupos pero al mismo tiempo maximizar la variabilidad entre los distintos grupos.

Supongamos los siguientes casos o conjunto de objetos:

	C_1	...	C_i	...	C_k
O_1	X_{11}	...	X_{1i}	...	X_{1k}
...
O_j	X_{j1}	...	X_{ji}	...	X_{jk}
...
O_n	X_{n1}	...	x_{ni}	...	X_{nk}

Donde $O = O_1, \dots, O_n$ representa el conjunto de objetos. Se trata de dividir O en k clústers, C_1, \dots, C_k , de tal forma que $\bigcup_{i=1}^k x_i = O$. En particular, O es un conjunto de n objetos, como personas, con k variables (también denominadas medidas o atributos), como edad, peso, altura, género, raza, etc. La estructura tiene la forma de una matriz de $n \times k$ (n objetos por k variables).

Una vez planteado el problema, las actividades del análisis de clustering incluyen los siguientes pasos [14]:

1. *Representación de patrones.* Se refiere al establecimiento del número de clases, número de patrones, y el número, tipo y tamaño de las características disponibles para el algoritmo de clustering.
2. *Definición de proximidad.* La proximidad de los patrones es usualmente medida por una función distancia entre un par de datos.
3. *Clustering.* La etapa de agrupamiento puede desarrollarse en un gran número de formas. Se pueden utilizar agrupamientos de clusters jerárquicos, particionales y otros más que abarcan métodos probabilísticos o de teoría de grafos.
4. *Abstracción de datos.* Es el proceso de extraer una representación simple y compacta del conjunto de datos.
5. *Verificación de resultados.* Consiste en validar el análisis de cluster realizado evaluando los resultados obtenidos.

4. ANTECEDENTES

Existen numerosos trabajos orientados al modelaje del alumno considerando sus estilos de aprendizaje. En estos trabajos se aplican diversos modelos de estilos de aprendizaje y diversas técnicas para reconocer los diferentes estilos en los alumnos.

En [10], se evalúa la precisión de las Redes Bayesianas para detectar el estilo de aprendizaje de un estudiante en un sistema de educación a distancia. La Red Bayesiana modela aspectos diferentes de comportamiento del estudiante mientras éste interactúa con el sistema. De esta manera, es posible inferir los estilos de aprendizaje del alumno a partir de los comportamientos modelados. En el trabajo de referencia las Redes Bayesianas fueron evaluadas en el contexto de un curso e-learning de Inteligencia Artificial y se empleó el modelo de aprendizaje propuesto por Felder y Silverman.

En [32] se propone un algoritmo genético para identificar automáticamente los estilos de aprendizaje de estudiantes, de acuerdo a las acciones que realiza mientras asiste a un curso con modalidad e-learning. Más precisamente, el algoritmo propuesto apunta a detectar cuál es la combinación de acciones que el estudiante normalmente lleva a cabo mientras atiende una unidad académica. En este trabajo, cada cromosoma de la población inicial representa una combinación posible de acciones. Cada vez que el alumno finaliza una unidad académica se genera una nueva población de cromosomas. Luego, usando una función de mapeo, el mejor cromosoma en una población determina el estilo de aprendizaje del alumno.

En [9] se emplea el sistema 3DE (Design, Development and Delivery Electronic Environment for Educational Multimedia), y demuestran que al combinar e-learning con los estilos de aprendizaje es posible personalizar la enseñanza y en consecuencia mejorar el rendimiento individual. El sistema 3DE, permite crear de manera automática o guiada cursos adaptados a las necesidades de cada estudiante. Para ello, en primer lugar, cada estudiante debe realizar un test que analice sus estilos de aprendizaje. El test que utiliza este sistema se basa en el modelo de Honey y Mumford [13].

Mistral [25] es otra herramienta de autor para el desarrollo de cursos a distancia, que automatiza completamente el proceso de creación, mantenimiento, enseñanza/aprendizaje y administración de cursos a distancia, basándose en las mismas tecnologías que los Sistemas Hipermmedia Adaptativos

(SHA) y los Sistemas Tutor Inteligentes (STI). Un aspecto especial de esta propuesta, que incide en la capacidad de adaptación al usuario, es la posibilidad de diagnosticar el conocimiento y los estilos de aprendizaje del alumno para así poder elegir la mejor estrategia de enseñanza según sus necesidades, y el mecanismo de evaluación más adecuado a su conocimiento. Este sistema diagnostica los estilos de aprendizaje de los alumnos según el inventario de aprendizaje de David Kolb [17; 18].

MAS-PLANG (MultiAgent System PLANG) [1], es un sistema multiagente diseñado para ofrecer características de adaptabilidad a la plataforma USD (Unidades de Soporte a la Docencia) utilizada para dar soporte a la docencia a través de la web. Este sistema, transforma el entorno educativo de la USD en un sistema adaptativo teniendo en cuenta los estilos de aprendizaje para categorizar a los alumnos. Se basa en el modelo de Felder y Silverman.

Otro de los antecedentes vinculados al problema es el trabajo de Conejo et al. [4]; en el cual se propone utilizar para el modelado del alumno, teorías bien fundamentadas y ampliamente comprobadas que garanticen el funcionamiento óptimo del sistema en todas las situaciones posibles. En concreto, se propone el uso de la *teoría de la probabilidad* como marco teórico. Además, se establece una relación entre el problema del modelado del alumno con la teoría de los *test adaptativos informatizados* (TAI), desarrollado dentro del campo de la psicometría, y que pese a su capacidad demostrada para mejorar el proceso de diagnóstico, tanto en precisión como en tiempo, no ha sido aún utilizada dentro del campo de los STI. Estas investigaciones se encuentran en torno al sistema SIETTE, que es una herramienta web basada en la teoría de la respuesta al ítem unidimensional que cumple con dos objetivos: (a) permitir que los profesores definan de una forma muy sencilla un test adaptativo informatizado; y (b) permitir que los alumnos realicen los test definidos y sean evaluados por el sistema, todo ello a través de la Web.

Es importante destacar que en ninguno de los antecedentes encontrados se aplica la técnica de clustering para identificar el estilo de aprendizaje de los estudiantes.

5. MÉTODO PROPUESTO

En líneas generales, el método consiste en analizar las transacciones realizadas por el alumno, en el curso e-learning, con el propósito de determinar su estilo de aprendizaje dominante. En particular, a fin de reconocer un patrón en el comportamiento del alumno que permita posteriormente determinar su estilo de aprendizaje, se aplica el análisis del cluster a los registros del log.

En la definición de las etapas del método se siguen las fases del proceso de descubrimiento de conocimiento (KDD). KDD es un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos [27]. Teniendo en cuenta este proceso se proponen tres fases.

Fase de Pre-procesamiento o preparación de los datos: Se compone de las siguientes tareas:

- *Selección.* Consiste en elegir las Bases de Datos (BD) relevantes con respecto a las transacciones realizadas por los alumnos en su interacción con el sistema.
- *Limpieza:* Para cada una de las BD se eliminan los datos irrelevantes, dejando únicamente los atributos o campos que se consideran de importancia para realizar el proceso de clustering.
- *Transformación:* Esta tarea consiste en la conversión de datos en un modelo analítico donde los datos se transforman o consolidan en forma apropiada para la fase de MD. Cabe aclarar que

aunque las BD relacionales son la fuente para la mayoría de las aplicaciones de MD, muchas técnicas, en particular el análisis de cluster, no son capaces de trabajar con toda la BD, sino con una sola tabla a la vez. A partir de las transformaciones realizadas a los datos, se obtiene una vista minable, conformada por una única tabla sobre la cual se aplicará el proceso de MD.

Fase de Procesamiento: En esta fase se aplica el proceso de MD para cada alumno. Particularmente, al tratarse de una tarea descriptiva mediante la cual se tratan de identificar patrones que caractericen el comportamiento de cada uno de los alumnos, se aplican técnicas de análisis de cluster sobre la vista minable obtenida.

Con el propósito de alcanzar la mayor precisión posible se ha optado por la implementación del algoritmo FarthestFirst, el cual minimiza la distancia entre una tupla y su centroide dando como resultado clusters bien concentrados.

Para la concreción de esta fase se emplea el software WEKA que contiene las herramientas necesarias para realizar tareas de clustering.

Fase de post-procesamiento: Esta fase incluye la interpretación y evaluación de los patrones obtenidos en el proceso de MD.

Para la interpretación de los patrones se definen, previamente, un conjunto de reglas que permitan mapear o enmarcar los resultados obtenidos dentro de las cuatro dimensiones del modelo de aprendizaje de Felder y Silverman.

EVALUACIÓN DEL MÉTODO

El método propuesto fue evaluado en el contexto de un curso sobre “Sistemas de supervisión y control de procesos”, desarrollado en el marco de la asignatura Sistemas de Información II, de la carrera Licenciatura en sistemas de Información, de la F.C.E. y T., de la U.N.S.E. Esta asignatura corresponde a la línea curricular de Sistemas y está orientada fundamentalmente a brindar a los alumnos herramientas para el diseño de Sistemas en Tiempo Real. La población de estudiantes que realizó el curso totalizaba 22 alumnos.

Para la implementación del curso e-learning se empleó la plataforma educativa MOODLE [23]. El curso, fue diseñado de tal manera de ofrecer a cada uno de los alumnos diferentes vías para que, conforme a su estilo de aprendizaje, cada uno pueda escoger su propio camino para concretar la apropiación de conocimiento. En particular, el curso fue diseñado considerando las cuatro dimensiones del modelo de estilos de aprendizaje de Felder y Silverman. Tal es así, que los contenidos teóricos fueron planteados en el curso con dos modalidades. Una mediante texto narrativo y la otra, más esquemática, mediante una presentación en Power Point. Además, se incorporaron archivos adicionales con ejemplos explicativos sobre cada tema teórico. En cuanto a las actividades prácticas, se diseñaron diferentes actividades para cada tema. Para ello se han seguido los lineamientos formulados en [5], ya que estos son propiciadores de una mayor compatibilidad entre el estilo de aprendizaje de los alumnos y el de enseñanza de los profesores. Atendiendo a estos lineamientos, se han contemplado distintas modalidades y el uso de las diferentes tecnologías ofrecidas por Moodle, especialmente las interactivas y las colaborativas. Esto permitió que los alumnos pudieran elegir aquella modalidad de ejercitación con la que se sintieran más cómodos, según su estilo de aprendizaje.

En la fase de Pre-procesamiento, de las 186 tablas generadas por el sistema, sólo se seleccionaron 5, que son las que contienen datos relevantes con respecto a las transacciones realizadas por los alumnos en su interacción con el sistema. Para cada una de las tablas se eliminaron los datos irrelevantes, dejando únicamente los atributos o campos que fueron considerados de importancia

para realizar el proceso de clustering. Luego se definieron los nuevos atributos, que permitirían caracterizar el comportamiento de los alumnos en cada sesión. Los mecanismos de conversión, definidos para la obtención de los nuevos atributos, fueron codificados en un *script* que recorre, para cada alumno, la tabla de registros transaccionales identificando todas las sesiones de usuario, y por cada sesión, el valor de cada atributo de acuerdo a una serie de criterios pre-definidos. Tanto los nuevos atributos, como los criterios establecidos para la asignación de valores, se definieron teniendo en cuenta las cuatro dimensiones del modelo de estilos de aprendizaje de Felder y Silverman. A modo de ejemplo se cita un atributo, con sus respectivos valores y criterios de asignación, correspondiente a la dimensión activo / reflexivo.

Atributo: Foro.

Descripción: Este atributo describe la participación del alumno en cualquiera de los foros definidos en el curso.

Valores: post, view, ambos, no usa.

Criterios para asignar cada valor:

- *Post.* Se asigna este valor si durante la sesión el alumno ha ingresado a cualquiera de los foros definidos para el curso principalmente para iniciar una nueva discusión o para agregar una respuesta a una discusión existente.
- *View.* Se asigna este valor si durante la sesión el alumno ha ingresado a cualquiera de los foros definidos en el curso principalmente para visualizar lo que han agregado otros usuarios.
- *Ambos.* Se asigna este valor cuando durante la sesión el alumno ha ingresado a cualquiera de los foros para visualizar el aporte de otros usuarios y para agregar una discusión o una respuesta sin lograrse reconocer ninguna preferencia por una u otra acción.
- *No Usa.* Se asigna este valor si durante la sesión el alumno no ha accedido a ninguno de los foros definidos para el curso.

A partir de los nuevos atributos definidos en la etapa de transformación, se generó la vista minable conformada por una única tabla (Tabla 1) sobre la cual se aplicó el proceso de MD. Cada registro en la tabla corresponde a una sesión del alumno identificado por *Iduser*.

Tabla 1. Vista Minable

Mdl_sesiones
Id_sesion
Iduser
Wiki
Foro
Chat
Glosario
Acceso a la Teoría
Acceso a la Práctica
Análisis de la Información
Tipo de Material
Material de Entrada
Participación en el Foro

El último paso, fue convertir los registros de la tabla a un formato de archivo con extensión *.arff* que luego fue utilizado por la herramienta WEKA [31] en la fase de Procesamiento. Al finalizar la fase de Pre-Procesamiento se obtuvo un archivo *.arff* para cada alumno del curso.

En la fase de Procesamiento, y dada las características del problema, se decidió aplicar técnicas de tipo descriptivo, en particular análisis de cluster para descubrir un patrón en el comportamiento de cada alumno y a partir de éste, determinar su estilo de aprendizaje dominante. El algoritmo de clustering seleccionado fue el FarthestFirst [12] por tratarse de un problema de k centros donde se pretende que la máxima distancia entre una tupla y su centroide sea mínima. Para la realización de esta fase, se suministraron al software WEKA cada uno de los archivos .arff obtenidos para cada alumno.

En la fase de postprocesamiento, se definieron un conjunto de reglas para determinar el valor correspondiente a cada una de las cuatro dimensiones del modelo de Felder y Silverman. Estas reglas fueron aplicadas a los valores de los atributos presentes en el cluster dominante de cada alumno. A continuación se transcribe una de las reglas definidas:

Si el Acceso a la teoría es “secuencial”, y el Acceso a la práctica es “secuencial” o “ambos”, y el análisis de la información es “detallado” o “ambos”, entonces el valor de la dimensión de comprensión será “secuencial”.

Al finalizar esta fase, se obtuvo el estilo de aprendizaje dominante de cada alumno.

A fin de evaluar la precisión del método propuesto, se compararon los resultados obtenidos por el test ILS (Index of Learning Style) [8] (realizado al inicio del curso por los alumnos) con los resultados obtenidos mediante el método propuesto. A continuación, se presenta una tabla (Tabla 2) donde se muestra el estilo de aprendizaje obtenido para cada alumno mediante la aplicación de ambos métodos.

Tabla 2: Resultados obtenidos mediante ILS y mediante el Análisis de Cluster.

Alu	Procesamiento		Percepción		Entrada		Comprensión	
	Visual/Verbal		Sensorial/Intuitivo		Visual/Verbal		Secuencial/Global	
	ILS	Clustering	ILS	Clustering	ILS	Clustering	ILS	Clustering
1	Equilibrado	Reflexivo	Sensorial	Sensorial	Equilibrado	Visual	Equilibrado	Mixto
2	Equilibrado	Reflexivo	Sensorial	Sensorial	Equilibrado	Mixto	Equilibrado	Mixto
3	Activo	Mixto	Sensorial	Sensorial	Visual	Mixto	Equilibrado	Secuencial
4	Equilibrado	Reflexivo	Sensorial	Sensorial	Equilibrado	Mixto	Equilibrado	Mixto
5	Equilibrado	Reflexivo	Equilibrado	Sensorial	Visual	Visual	Equilibrado	Mixto
6	Equilibrado	Mixto	Sensorial	Sensorial	Visual	Visual	Global	Mixto
7	Equilibrado	Reflexivo	Equilibrado	Sensorial	Equilibrado	Mixto	Equilibrado	Mixto
8	Equilibrado	Mixto	Sensorial	Sensorial	Equilibrado	Visual	Secuencial	Secuencial
9	Reflexivo	Reflexivo	Sensorial	Sensorial	Visual	Verbal	Equilibrado	Mixto
10	Equilibrado	Mixto	Sensorial	Sensorial	Equilibrado	Mixto	Equilibrado	Secuencial
11	Activo	Reflexivo	Equilibrado	Intuitivo	Visual	Mixto	Secuencial	Secuencial
12	Activo	Reflexivo	Equilibrado	Sensorial	Equilibrado	Mixto	Equilibrado	Mixto
13	Equilibrado	Reflexivo	Sensorial	Sensorial	Visual	Visual	Equilibrado	Mixto
14	Equilibrado	Reflexivo	Sensorial	Mixto	Equilibrado	Mixto	Equilibrado	Mixto
15	Activo	Reflexivo	Sensorial	Sensorial	Equilibrado	Mixto	Equilibrado	Mixto
16	Equilibrado	Reflexivo	Sensorial	Mixto	Equilibrado	Visual	Secuencial	Mixto
17	Equilibrado	Mixto	Equilibrado	Sensorial	Visual	Visual	Secuencial	Secuencial
18	Equilibrado	Mixto	Equilibrado	intuitivo	Visual	Mixto	Equilibrado	Secuencial
19	Activo	Reflexivo	Equilibrado	Sensorial	Visual	Mixto	Secuencial	Mixto
20	Equilibrado	Activo	Equilibrado	Sensorial	Equilibrado	Mixto	Equilibrado	Mixto
21	Equilibrado	Reflexivo	Sensorial	Mixto	Visual	Verbal	Equilibrado	Mixto
22	Reflexivo	Reflexivo	Sensorial	Sensorial	Equilibrado	Mixto	Equilibrado	Mixto

La evaluación se realizó teniendo en cuenta la siguiente métrica [10]:

$$Precisión = \frac{\sum_{i=1}^n Sim(LS_C, LS_{ILS})}{n}$$

En donde *Sim* es 1 si los valores obtenidos con la técnica de clustering y con el ILS son iguales, 0 si son opuestos, y 0.5 si uno es equilibrado o mixto y el otro un valor extremo (en este caso, el error es pequeño comparado a la situación en la cual los resultados son opuestos); y *n* es el número de estudiantes analizados (22 alumnos).

Para la dimensión de Procesamiento se obtuvo una precisión de 57%, para la dimensión de Percepción y la dimensión de Entrada una precisión de 75%, y para la dimensión de Comprensión una precisión de 86%.

Tal como se puede apreciar, la menor precisión se ha obtenido en la dimensión de Procesamiento. Analizando las inconsistencias en esta dimensión y los datos empleados para el clustering, hemos detectado que los alumnos han hecho poco uso de las herramientas colaborativas tales como chat, wiki, foro y glosario. Investigando las causas se ha detectado que el grupo de estudiantes posee poca experiencia en el uso de herramientas colaborativas como foro, wiki y glosario; y durante el curso han empleado otras herramientas de comunicación externas al sitio, como por ejemplo *messenger* en lugar del chat definido en el curso.

Con respecto a las dimensiones restantes (Entrada, Percepción y Comprensión) hemos obtenidos muy buenos resultados en cuanto a la precisión del método propuesto, principalmente en la dimensión de percepción (86%). Podríamos afirmar, que en estas tres dimensiones, es posible utilizar el Análisis de Cluster para determinar el estilo de aprendizaje de un alumno con una alta probabilidad.

6. CONCLUSIONES

Tal como hemos mencionado al comienzo del artículo, los sistemas e-learning actuales presentan muchas falencias relacionadas con la adaptación a las características distintivas de los alumnos, lo cual conlleva a una educación deficiente con alumnos que no logran asimilar los conocimientos.

Con el propósito de permitir a los sistemas de e-learning personalizar los contenidos pedagógicos y la estrategia de enseñanza a las necesidades y preferencias de los alumnos, se ha propuesto en este artículo una solución alternativa al problema mediante el empleo de Técnicas de Análisis de Cluster para reconocer el estilo de aprendizaje dominante de cada alumno.

Del análisis de los resultados obtenidos se puede concluir que el método propuesto constituye una herramienta muy efectiva para determinar el estilo de aprendizaje de los estudiantes en las dimensiones de entrada, percepción y comprensión. Con respecto a la dimensión de procesamiento, la efectividad del método está condicionada por la experiencia de los estudiantes en el manejo de herramientas colaborativas tales como wiki, foro, y glosario.

El conocimiento obtenido mediante el análisis de cluster podrá ser utilizado por los sistemas de e-learning para la construcción de un modelo de estudiante que refleje el estilo de aprendizaje y permita ofrecer una enseñanza personalizada, fortaleciendo la formación del alumno en cuanto a sus capacidades y habilidades que lo destacan, permitiéndole reanimar, estimular o perfeccionar aquellas poco desarrolladas.

7. REFERENCIAS

- [1] Aguilar M., Peña C. I., Fabregat R. (2002). SMIT: un agente sintético antropomórfico para un entorno virtual de aprendizaje. *III Conferencia internacional sobre educación, formación y nuevas tecnologías Virtual Educa 2002*.
- [2] Alonso C. M., Gallego D. J., Honey P. (1999). *Estilos de aprendizaje: procedimientos de diagnóstico y mejora*, 4ª Edición, Editoriales Bilbao, Mensajero.
- [3] Carver C. A., Howard R. A. y Lavelle E. (1996). Enhancing student learning by incorporating learning styles into adaptive hypermedia, en *Proceedings of ED-MEDIA'96 World Conference on Educational Multimedia Hypermedia*, 118-123, Boston, USA.
- [4] Conejo R., Millán E., Pérez De La Cruz J. L. y Trella M. (2001). Modelado del alumno: un enfoque bayesiano. *Revista Iberoamericana de Inteligencia Artificial*. No.12 (2001), 50-58.
- [5] Durán E. y Costaguta R. (2008). Una experiencia de enseñanza adaptada al estilo de aprendizaje de los estudiantes en un curso de Simulación. *Formación Universitaria*, Vol. 1(1), 19-28.
- [6] Felder R. M. y Brent R. (2005). Understanding Student Differences, *Journal of Engineering Education*, 94 (1), 57-72.
- [7] Felder R. M. y Silverman L. K. (1988). Learning and teaching styles in engineering education, *Journal of Engineering Education*, 78(7), 674-681.
- [8] Felder R. M. y Soloman B. (2006). *Index of Learning Style Questionnaire*. Disponible en <http://www.engr.ncsu.edu/learningstyles/ilsweb.html>.
- [9] Gallego Rodríguez, A. y Martínez Caro E. (2003). Estilos de aprendizaje y e-learning. Hacia un mayor rendimiento académico. *RED. Revista de Educación a Distancia*, N°7, 1-10
- [10] Garcia P., Amandi A., Schiaffino S. y Campo M. (2007). Evaluating Bayesian networks' precision for detecting students' learning styles. *ACM*, Vol. 49, Issue 3 (November 2007), 794-808.
- [11] Gilbert J. E. y Ham C. Y. (1999). Adapting instruction in search of a significant difference, en *Journal of Network and Computer applications*, Vol 22.
- [12] Hochbaum, Shmoys (1985). A best possible heuristic for the k-center problem. *Mathematics of Operations Research*. 10(2):180-184
- [13] Honey, P.; Munford, A. (1986). *Using your learning styles*. Peter Honey Publications.
- [14] Jain A. K., Murty M.N., and Flynn P. J. (1999) Data clustering: a review. *ACM Computing Surveys*, 31(3):264–323.
- [15] Kaufmann M., Han J., Kamber M. (2000). *Data Mining Concepts and Techniques*. Elsevier. 2ª Edición.
- [16] Keefe, J. W. (1987). *Learning Style Theory and Practice*. National Association of Secondary School Principals. Reston, VA, (EEUU).
- [17] Kolb, D. (1984). *Experiential Learning. Experience as the source of learning and Development*. Prentice Hall P T R, Englewood Cliffs, New Jersey. 1984
- [18] Kolb, D. (1985). *Learning-style Inventory*. Boston: McBer & Company.
- [19] Nils J. N. (2006). *Introduction Machine Learning*. Stanford University. Stanford (EEUU).
- [20] Papanikolaou K. A., Grigoriadou M., Kornilakis H. y Magoulas G. D. (2003). Personalizing the Interaction in a Web-based Educational Hypermedia System: the case of INSPIRE. *User Modelling and User – Adapted Interaction*, 13, 213-267, Kluwer Academic Publishers. Netherlands.
- [21] Paredes P. y Rodríguez P. (2002). Considering sensing-intuitive dimension to exposition-exemplification in adaptive sequencing, en *Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based System (AH2002)*, Málaga. España.

- [22] Peña C.I. y Marzo J.L. (1999). Adaptive Intelligent Agent Approach to Guide the Web Navigation on the Plan-G Distance Learning Platform. *IEEE Colloquium. Lost in the Web – Navigation on the Internet*. London, Netherland.
- [23] Rice W. (2006). *Moodle: e-learning course development: a complete guide to successful learning using Moodle*, Packt Publishing, Birmingham, Reino Unido.
- [24] Rumetshofer H. y Wöb W. (2003). An Approach for adaptable learning systems with respect to psychological aspects, en *Proceedings of the 2003 ACM Symposium on Applied Computing*, 558-563. ACM.
- [25] Salcedo Lagos P., Labraña C., Farrán Leiva Y. (2002). Una Plataforma Inteligente de Educación a Distancia que incorpora la Adaptabilidad de Estrategias de Enseñanza al Perfil, Estilos de Aprendizaje y Conocimiento de los Alumnos. *XXVIII Conferencia Latinoamericana de Informática - CLEI 2002*, Montevideo, Uruguay.
- [26] Stash N. y P. De Bra. (2004). Incorporating Cognitive Styles in AHA! (The Adaptive Hipermedia Architecture), en *Proceedings of the IASTED International Conference – WEB-BASED EDUCATION*, 16-18, Innsbruck, Austria.
- [27] Steinbach M., Tan P., and Kumar V. (2006). *Introduction to Data Mining*. Addison Wesley.
- [28] Stern M. y Woolf P. (2000). Adaptive content in an online lecture system, en *Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-based systems*, 291-300, Toronto, Italia.
- [29] Triantafillou E., Pomportsis A. y Georgiadou E. (2002). AES-ES: Adaptive Educational System base on cognitive styles, en *Proceedings of the 2nd International Conference on Adaptive Hypermedia and Adaptive Web Based System (AH2002)*, 10-20, Malaga, España, Mayo 2002.
- [30] VanLehn K., Polson. M.C., and Richardson, J.J., (eds). (1988). *Student modeling in Foundations of Intelligent Tutoring Systems*, cap 3., 55-78, Lawrence Erlbaum Associates publishers.
- [31] Witten I. H. y Frank E. (1999). *Data Mining. Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann publisher.
- [32] Yannibelli V., Godoy D., y Amandi A. (2006). A Genetic Algorithm Approach to Recognize Students' Learning Styles. *Interactive Learning Environments*, Vol. 14, Issue 1 April 2006, 55 – 78.