

# Replanificación de actividades en cursos virtuales personalizados con árboles de decisión, lógica difusa y colonias de hormigas

## Re-Planning of activities in virtual course personalized with decision trees, fuzzy logic and ant colonies

Karol Lina López Bedoya M.Sc., Néstor Darío Duque Méndez Ph.D., & Darwin Brochero Bueno M.Sc.  
Grupo de Investigación GAIA.

Universidad Nacional de Colombia Sede Manizales

karol-lina.lopez.1@ulaval.ca, ndduqueme@unal.edu.co, darwin.brochero.1@ulaval.ca

Recibido para revisión 10 de septiembre de 2010, aceptado 03 de enero de 2011, versión final 11 de febrero de 2011

**Resumen**— Las técnicas de adaptación en los cursos virtuales se han enfocado en la planificación estática, la cual se basa en la configuración de varios modelos de un curso, para permitir la elección del más adecuado bajo algún parámetro determinante. En este artículo se propone un modelo de replanificación que permite la selección dinámica e inteligente de actividades, soportado en la conformación de una matriz de actividades clasificadas por el nivel de dificultad y el grado de participación de tres recursos de enseñanza: texto, gráficos y medios audiovisuales.

A partir de esta configuración y la interacción de seis agentes inteligentes y cinco bases de datos se desarrollan dos lineamientos inteligentes para la selección dinámica de las actividades a presentar al estudiante; por un lado se tiene inicialmente un sistema basado en Lógica Difusa alimentado por el conocimiento experto desarrollado en investigaciones de los estilos de aprendizaje de Felder-Silverman y por el otro, se tiene una fase de refinamiento que obedece al seguimiento del proceso de enseñanza/aprendizaje mediante un modelo supervisado que se basa en Árboles de Decisión, Lógica Difusa y Colonias de Hormigas (ATAFIS: Adapting decision Trees and Ant colony in Fuzzy Inference System). Los resultados encontrados en series generadas aleatoriamente demuestran la robustez y eficiencia del sistema en la simulación del nivel de comprensión, base de la replanificación de actividades propuesta.

**Palabras Clave**— Replanificación adaptativa, Cursos Virtuales, Árboles de Decisión, Sistemas Difusos, Colonias de hormigas.

**Abstract**— The adaptive techniques in virtual courses have focused in the static planning, which is based on the configuration of several models of a course, to allow the choice most appropriate under a determinant parameter.

This article proposes a re-planning model that allows dynamic and intelligent selection of activities supported in the formation of an array of activities grouped by level of difficulty and degree of participation of three teaching media resources: text, graphics and

audiovisual. From this setting and the six interacting intelligent agents, two lines intelligent dynamic selection of activities to introduce the student was develop. On the one hand there is initially a Fuzzy Logic based system powered by the expertise developed investigations of the learning styles of Felder-Silverman and on the other, there is a phase of refinement that reflects the monitoring of the teaching / learning through a supervised model is based on Decision Trees, Fuzzy Logic and Ant Colonies (ATAFIS). The results found in randomly generated sets demonstrate the robustness and efficiency of the simulation the level of understanding, based on the replanning of activities proposed.

**Keywords**— Adaptive Replanning, Virtual Courses, Decision Trees, Fuzzy Systems, Ant Colonies.

### I. INTRODUCCIÓN

La adaptación en cursos virtuales debe realizarse mediante una determinada estrategia que permita relacionar elementos educativos tales como objetivos, estrategia pedagógica, contenidos, recursos y actividades con las características particulares de cada estudiante, con el fin de realizar la planificación inicial del curso y posteriormente, mediante un seguimiento controlado, determinar si es necesaria o no una replanificación. Los diferentes elementos involucrados y las características heterogéneas del estudiante, hacen que la estrategia de planificación y replanificación se presente compleja, pero la aplicación de diferentes técnicas inteligentes permite obviar gran parte de las dificultades sin sacrificar el objetivo [7].

La mayoría de los cursos virtuales adaptativos basan su funcionamiento en la conformación de varios modelos de un mismo curso para asociarlo a cada estudiante de acuerdo con

algunas de sus características y además, tienen una capacidad de diagnóstico e interactividad muy limitada [13].

Evaluando varios relevantes sistemas educativos inteligentes y adaptativos basados en la Web (AIWBES: Adaptive and intelligent Web-based educational systems)[15]; por ejemplo, KBS-Hyperbook [14], MetaLinks [17], ActiveMath [8], ELM-ART II [9] COLER [2], SICAD [6] y SQLT-Web [4]; el análisis del nivel de comprensión y la replanificación de actividades no están contemplados explícitamente en ninguno de ellos.

El modelo de replanificación de actividades en cursos virtuales propuesto en esta investigación, involucra seis agentes inteligentes y cinco bases de datos o tablas para la manipulación de la información del modelo.

El Agente Administrador de Cursos AAC se encarga de organizar las unidades expuestas (actividades o explicaciones) que los profesores entregan al sistema, el Agente de Recomendación de Actividades ARA tiene como función la elección inteligente de la actividad a presentar al estudiante con base en la información del Agente de Retroalimentación AR (el cual hace el seguimiento académico y psicopedagógico del estudiante); el Agente de Interfaz AI es el encargado de la interacción con el estudiante y el Agente de Modelación del Historial AMH tiene como función simular la información almacenada del seguimiento del estudiante. Finalmente se tiene el Agente de Evaluación Adaptativa AEA, encargado de elegir la forma de evaluar a cada estudiante.

El sistema inteligente del ARA, está basado en un módulo de Lógica Difusa (FLM); donde el grado de personalización se perfecciona en dos fases, en la fase inicial, el FLM se define a partir de investigaciones que relacionan la participación óptima de recursos de enseñanza con alguno de los estilos de aprendizaje de Felder y Silverman, y en la fase posterior, es necesario el seguimiento del estudiante (a cargo del AR) y las simulaciones del AMH, proceso que involucra la minería de información con Árboles de Decisión, Lógica Difusa y optimización con Colonias de Hormigas.

En síntesis el funcionamiento del modelo se puede resumir como la presentación de actividades al estudiante con base en el nivel de comprensión simulado, el nivel de dificultad y la disponibilidad de recursos.

## II. ARQUITECTURA DEL SISTEMA

En esta sección se describe la arquitectura del sistema y la interacción entre los distintos agentes que intervienen en la tarea de elección inteligente de actividades educativas. En la sección 2.1 se presenta la nomenclatura utilizada en el modelo, en la sección 2.2 se muestra la concepción general del modelo, en la sección 2.3 se expone el funcionamiento general del modelo y en la sección 2.4 se describen las funciones de cada agente.

### 2.1 Nomenclatura utilizada en el modelo

Para una mejor comprensión del lector, se presenta a continuación la nomenclatura utilizada por el modelo de replanificación propuesto.

Unidad Expuesta (UE): representa el tema a explicar al estudiante fundamentado en tres contextos: teoría (T), ejemplos (X) y ejercicios (E). Las UE pueden ser compuestas e incluir varios de estos elementos.

Recursos de enseñanza (RE): medios u elementos multimedia con los cuales se diseñan las actividades. Se tuvieron en cuenta tres tipos: textos (TX), gráficos (GR) y medios audiovisuales (AV). Es importante resaltar que para el desarrollo del modelo, los profesores deben determinar la participación de estos recursos en cada actividad diseñada, por ejemplo para la actividad "x", que puede ser T, X o E, se tienen las siguientes proporciones de recursos en su conformación: TX = 35%, GR = 45% y AV = 20%, como es lógico, los recursos deben sumar el 100% en una actividad.

Estilos de Aprendizaje (EA): definidos de acuerdo con la propuesta de Felder y Silverman [12] la cual determina cuatro dimensiones activo/reflexivo, sensitivo/intuitivo, secuencial/global y visual/verbal, y se pueden evaluar mediante un cuestionario de 44 preguntas diseñado por Felder y Soloman [11].

Nivel de comprensión (NC): escala en la cual se mide o se simula el grado de comprensión de un estudiante ante una actividad presentada. Para el caso, esta escala se enmarcó entre 0 y 10, donde 10 refleja la comprensión total de una actividad.

Nivel de Dificultad (ND): este parámetro categórico discrimina tres niveles que los profesores asignan a cada actividad diseñada: bajo, medio o alto (fácil, normal o avanzado).

Secuencia de actividades (SA): representa el orden, con respecto al contexto (T, X, E), para presentar las actividades. En este caso, el orden está determinado por el estilo de aprendizaje del estudiante [16], por ejemplo, un estudiante con un estilo de aprendizaje sensitivo comprende mejor si se le presentan primero los ejercicios, luego los ejemplos y finalmente la teoría.

Asimismo se cuenta con seis agentes inteligentes y cinco bases de datos, así: Agente Administrador de Cursos (AAC), Agente de Recomendación de Actividades (ARA), Agente de Retroalimentación (AR), Agente de Interfaz (AI), Agente de Modelación del historial (AMH), Agente de Evaluación adaptativa (AEA), Base de datos de Unidades Expuestas (BUE), Base de datos de Cuentas de profesores (BCP), Base de datos del Perfil del Estudiante (BPE), Base de datos de Cuentas de Estudiantes (BCE) y Base de datos de Evaluaciones Adaptativas (BEA).

### 2.2 Concepción del modelo MORE

El modelo de replanificación de actividades o Modelo de Optimización de Recursos de Enseñanza (MORE) se basa

principalmente en la configuración inteligente de una matriz de actividades para generar n cursos a partir de procesos combinatorios, obteniendo un alto nivel de personalización en el proceso de enseñanza.

Los resultados obtenidos por Sabine Graf [16] y Carver et al. [1] sobre la relación entre los estilos de aprendizaje (EA), los recursos de enseñanza (RE), la secuencia de actividades (SA) y el nivel de comprensión del estudiante (NC), guiaron la consecución del modelo hacia el aprovechamiento de reglas del tipo "si x es a entonces y es b", por lo cual se eligió la utilización de los Sistemas de Inferencia Difusa.

De tal forma el MORE toma las anteriores variables como elementos de replanificación continua de los cursos; sin embargo, el modelo de simulación del NC tiene en cuenta solamente los EA y/o los RE puesto que la SA y el ND son referentes de la selección final de la actividad a presentar al estudiante.

Es importante resaltar que el NC real, es la respuesta que el sistema obtiene del estudiante después de presentada una actividad. De ésta manera, el mecanismo de personalización

del MORE se fundamenta en el seguimiento sistemático a cada estudiante, teniendo en cuenta las características de las actividades presentadas y el NC alcanzado por el estudiante.

El proceso de replanificación de actividades, proporciona una mejor personalización en el proceso enseñanza/aprendizaje a través del seguimiento del estudiante, tarea que es ejecutada mediante Árboles de Decisión y la aplicación de reglas condicionales en un Sistema de Inferencia Difusa optimizado con Colonias de Hormigas.

### 2.3 Funcionamiento del MORE

La Figura 1 muestra el funcionamiento del MORE y se detallan a continuación los pasos:

1. El profesor entra al sistema mediante la validación de su membresía en la base de datos de cuentas de profesores (BCP), establece el umbral para considerar aceptable la comprensión de una actividad y los umbrales de ascenso y descenso del nivel de dificultad a presentarle al estudiante, además define los índices para la evaluación del curso.

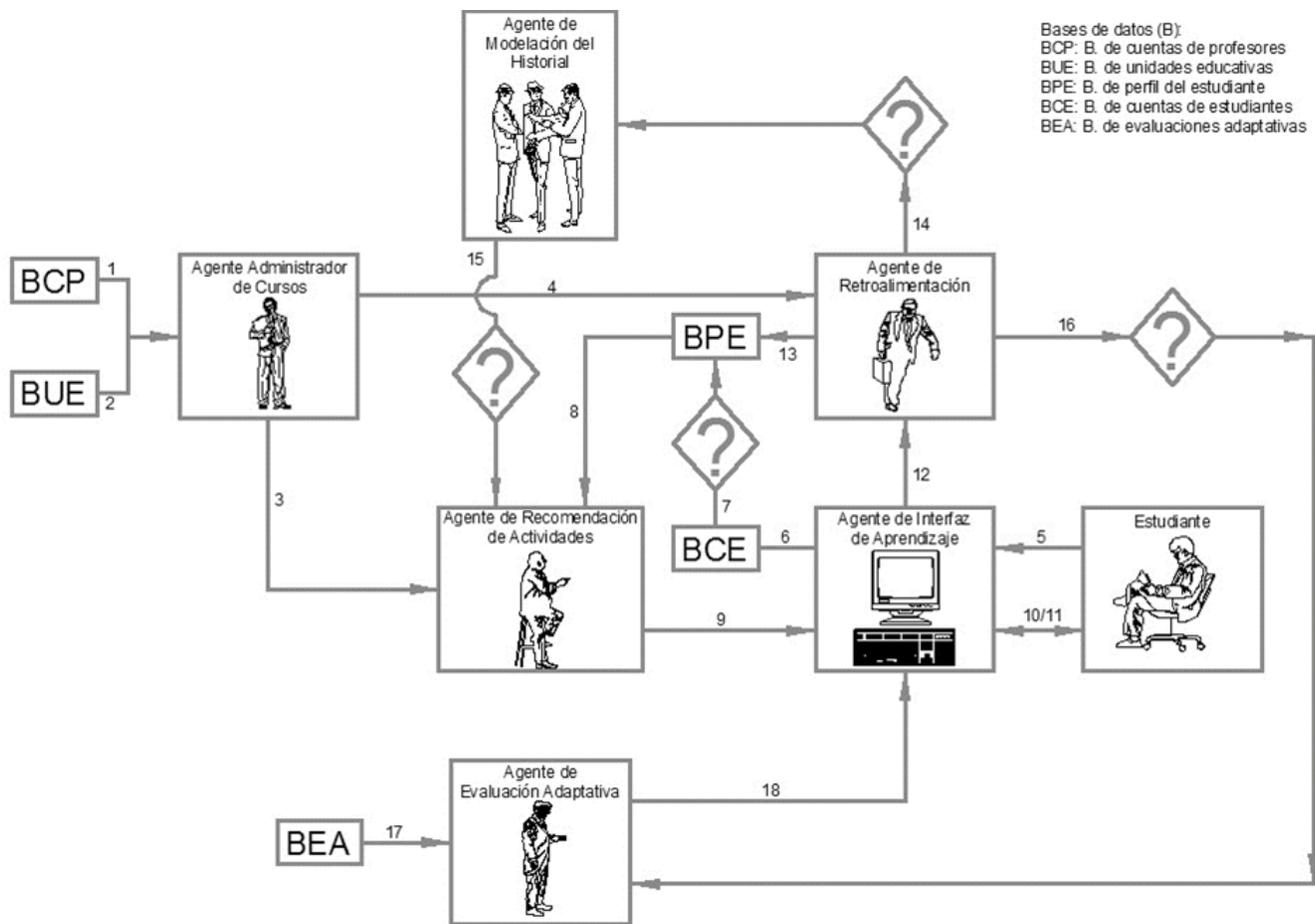


Figura 1. Arquitectura del sistema de replanificación adaptativa de actividades

2. El profesor ingresa las actividades educativas teniendo en cuenta los siguientes aspectos: UE a la que pertenece, contexto del curso, estimación del nivel de dificultad asociado y la participación de recursos de enseñanza.
  3. El Agente Administrador de Cursos (AAC) organiza la información entregada por los profesores y la entrega al Agente de Recomendación de Actividades (ARA).
  4. El AAC entrega al Agente de Retroalimentación (AR) los índices para realizar las evaluaciones, pruebas a cargo del Agente de evaluación Adaptativa (AEA).
  5. El estudiante inicia la interacción con el Agente de Interfaz (AI).
  6. La entrada del estudiante al sistema es verificada en la base de datos de cuentas de estudiantes (BCE) para continuar el proceso.
  7. El sistema debe corroborar si es la primera vez que el estudiante ingresa al sistema, de ser así se realiza un test académico para definir la secuencia de currículo académico y el nivel de dificultad para presentar la actividad inicial. Asimismo se realiza el test psicopedagógico para calificar los estilos de aprendizaje. Si las pruebas ya han sido realizadas al estudiante, el sistema se enfoca en actualizar la secuencia del currículo. La base de datos del perfil del estudiante (BPE) almacena la información actualizada del estudiante.
  8. La información académica y psicopedagógica del estudiante es entregada al Agente de Recomendación de Actividades (ARA).
  9. Con base en una metodología de selección dinámica de actividades, que tiene en cuenta la disponibilidad de recursos, las reglas de secuencia con respecto a los estilos de aprendizaje, nivel de dificultad previamente definido para el estudiante y los niveles de comprensión simulados por el sistema inteligente, el ARA entrega una actividad educativa al AI.
  10. El AI muestra al estudiante la unidad elegida por el ARA. Posteriormente el estudiante activa la opción para continuar con otra actividad y el AI pide al estudiante evaluar el nivel de comprensión de la unidad presentada.
  11. El sistema continúa el proceso si el estudiante responde la pregunta sobre la comprensión de la actividad presentada. La respuesta debe estar dentro de los límites establecidos por el administrador del sistema.
  12. La respuesta del estudiante es transmitida al AR, el cual almacena la historia académica y psicopedagógica del estudiante.
  13. El AR almacena en la base del perfil del estudiante (BPE) la información actualizada del estudiante.
  14. El AR verifica si se cumple la condición necesaria para pasar la información al Agente Modelador del Historial del estudiante (AMH). Esta condición se puede establecer en términos de la cantidad de patrones positivos (comprensión) y patrones negativos (sin comprensión) contenidos en el historial.
  15. Si se cumple la sentencia de paso para el AMH se ejecutan los modelos de simulación guiados hacía un análisis individual del estudiante.
  16. El AR, una vez se han cumplido las metas de comprensión de ciertas UE (información suministrada por los profesores), comunica al Agente de Evaluaciones Adaptativas (AEA) la necesidad de presentar una evaluación de conocimientos al estudiante.
  17. De la Base de Datos de Evaluaciones Adaptativas (BEA), configurada de forma similar a la Base de datos de Unidades Expuestas (BUE), se eligen las evaluaciones correspondientes a las Unidades educativas desarrolladas.
  18. El AEA entrega al AI la evaluación finalmente considerada, por ejemplo la evaluación que mayor similitud muestre con las preferencias observadas en el estudiante. Esta evaluación puede propiciar una modificación en la secuencia de currículo.
- El proceso básico continúa desarrollándose de manera cíclica desde el paso 1 hasta el paso 13, dando cabida a los demás agentes de replanificación con base en las sentencias de condicionamiento que se establezcan para cada uno.

## 2.4 Funciones de los agentes del MORE

### 2.4.1. Agente Administrador de Cursos (ACC)

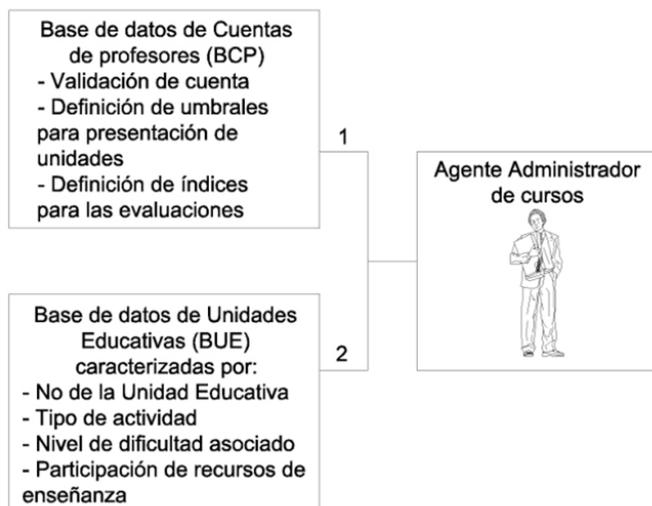


Figura 2. Agente Administrador de Cursos

El agente administrador de cursos (ACC) (Figura 2) es el encargado de la interacción entre el MORE y el profesor. El profesor, después de validar su membresía en el sistema, puede definir y modificar los umbrales de replanificación y los índices para las evaluaciones, además, puede establecer la base de unidades expuestas BUE cuya configuración matricial se puede observar en la Ecuación 1.

$$BUE = \begin{bmatrix} CUE_i & \dots & CUE_n \\ CTA_i & \dots & CTA_n \\ CND_i & \dots & CND_n \\ TX_i & \dots & TX_n \\ GR_i & \dots & GR_n \\ AV_i & \dots & AV_n \end{bmatrix} \quad \text{Equation 1}$$

Donde CUE es la codificación de la UE, CTA la codificación del contexto (Teoría, Ejemplos o Ejercicios), CND es la codificación del Nivel de Dificultad (Bajo, Medio y Alto); TX, GR y AV representan los porcentajes de participación de los recursos: Texto, Gráficas y medios Audiovisuales respectivamente.

Igualmente se definen los umbrales de comprensión de actividad (UC), de aumento de nivel de dificultad (UA) y el de disminución de del nivel de dificultad (UD); también los Índices para las evaluaciones adaptativas.

### 2.4.2. Agente de Interfaz (AI)

Este agente interactúa directamente con el estudiante, expone al aprendiz la actividad que el ARA considera más conveniente y permite la comunicación de los agentes humanos con el sistema.

Cuando un estudiante ingresa por primera vez debe realizar dos pruebas, una académica y otra psicopedagógica; de lo contrario, solo debe realizar la prueba académica, la cual tiene dos fines: el primero establecer la secuencia del currículo personalizado para el estudiante y el segundo, determinar el nivel de dificultad para presentar la primera actividad.

Par este caso en particular, la prueba psicopedagógica se enfoca en la evaluación de los estilos de aprendizaje mediante el test ILS formulado por Felder y Soloman [11]. En este test

|     |    |   |   |   |   |       |   |   |   |   |   |   |    |     |   |    |     |
|-----|----|---|---|---|---|-------|---|---|---|---|---|---|----|-----|---|----|-----|
| ACT | 11 | 9 | 7 | 5 | 3 | X     | 1 | 1 | 3 | 5 | 7 | 9 | 11 | REF |   |    |     |
|     |    |   |   |   |   | <---> |   |   |   |   |   |   |    |     |   |    |     |
| SEN |    | X |   |   |   |       | 1 | 1 | 3 | 5 | 7 | 9 | 11 | INT |   |    |     |
|     |    |   |   |   |   | <---> |   |   |   |   |   |   |    |     |   |    |     |
| VIS |    |   |   |   |   |       | X | 1 | 1 | 3 | 5 | 7 | 9  | VRB |   |    |     |
|     |    |   |   |   |   | <---> |   |   |   |   |   |   |    |     |   |    |     |
| SEQ |    |   |   |   |   |       |   |   | X | 1 | 1 | 3 | 5  | 7   | 9 | 11 | GLO |

Figura 3. Ejemplo de Resultado del Test Felder-Soloman

se realizan 44 preguntas con el fin de caracterizar al estudiante en las cuatro dimensiones de los estilos de aprendizaje (activo/reflexivo, sensitivo/intuitivo, secuencial/global y visual/verbal). Un ejemplo del resultado del test puede verse en la figura 3.

Entre 1 y 3: Hay un balance entre las dos dimensiones de la escala. Entre 5 y 7: Existe una preferencia moderada por una dimensión de la escala, se asume que el estudiante aprenderá más fácilmente si el ambiente de enseñanza favorece esta dimensión. Entre 9 y 11: Se tiene una preferencia muy fuerte por una dimensión de la escala, se pueden tener dificultades de aprendizaje en un ambiente que no soporte esta preferencia.

### 2.4.3. Agente de Recomendación de Actividades (ARA)

Este agente es el encargado de elegir la actividad a presentar al estudiante con base en la simulación del Nivel de Comprensión (NC) de las actividades contenidas en la BUE y en la disponibilidad de recursos para un Nivel de Dificultad (ND) y un contexto dado (T, X, E) (Figura 4).

Como se muestra en la Figura 4, la selección de actividades se centra en el NC simulado por el sistema para cada estudiante; sin embargo, si la BUE no contiene actividades con las características iniciales del estudiante (SA y ND) se establecen jerárquicamente dos sentencias que permitan encontrar la actividad más adecuada a presentar, por un lado se tiene como alternativa la variación del contexto elegido para presentar la actividad (T, X, E) y ante la dificultad repetida de disponibilidad de recursos se tiene como última opción la modificación del ND.

La simulación del NC es un proceso basado en Lógica Difusa que obedece a dos fases independientes: la inicial, ejecutada por el ARA, se basa en investigaciones sobre estilos de aprendizaje [1,12], y la segunda, desarrollada por el Agente de Modelación del Historial (AMH), como resultado de un seguimiento sistemático del proceso observado en cada estudiante.

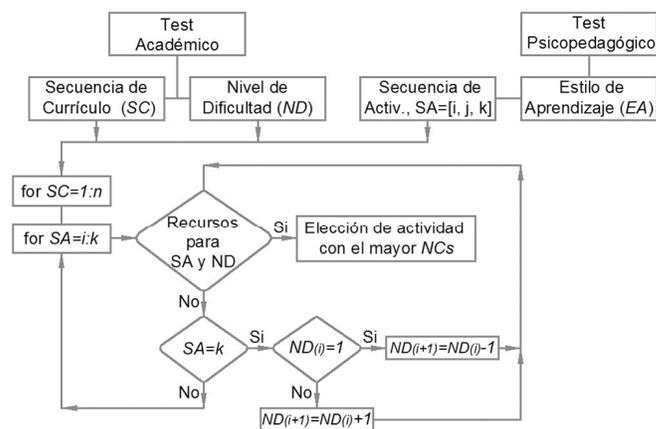


Figura 4. Mecanismo básico de selección de actividades

El modelo inicial de Lógica Difusa (FLM) se establece a partir de los siguientes criterios:

**Definición de las variables difusas:** Se tienen siete entradas correspondientes a las cuatro dimensiones de los EA y la participación de los tres recursos de enseñanza. Como salida única del sistema se define el nivel de comprensión de las actividades presentadas al estudiante (NC).

**Establecimiento de los conjuntos difusos:** Se tienen dos conjuntos difusos por cada variable de entrada y tres para la variable de salida (NC).

En las figuras 5, 6 y 7 se presentan las funciones de pertenencia o membresía que rigen a los conjuntos difusos de estilos de aprendizaje, recursos de enseñanza y nivel de comprensión respectivamente.

Es importante resaltar el dominio de las anteriores figuras puesto que los EA fueron reescalados entre 1 y 23 para facilitar la interpretación de los resultados del test psicopedagógico (Figura 3). Las abscisas en la Figura 6 representan el porcentaje de participación de los recursos y en la Figura 7 la escala del NC representado entre cero y diez.

Las funciones de membresía se establecieron como sigmoideas simétricas, a excepción de la correspondiente al NC medio, la cual es una función gaussiana de dos parámetros.

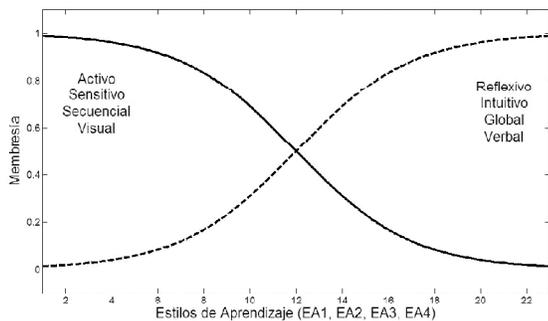


Figura 5. Funciones de membresía para los Estilos de Aprendizaje

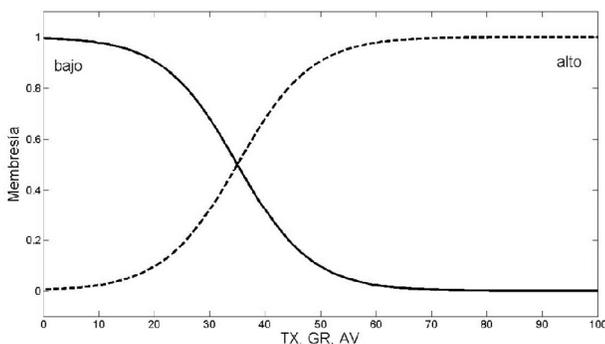


Figura 6. Funciones de membresía para recursos de enseñanza

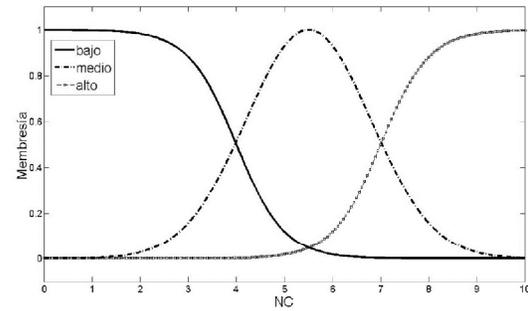


Figura 7. Funciones de membresía para el nivel de comprensión (NC)

**Reglas del sistema:** Con base en el número de variables de entrada (7) y los conjuntos difusos por cada una de ellas (2), se tiene que el FLM estaría idealmente definido por 128 reglas (27); no obstante, con base en los registros de Carver et al. [1], se dedujeron 48 reglas producto de la falta de información que relacione la incidencia simultánea de varios EA en la participación óptima de los recursos de enseñanza (RE). De igual forma se descartan las reglas donde la participación de los RE para todos los medios es baja o alta.

**Configuración del Sistema de Inferencia Difuso (FIS, Fuzzy Inference System)** Acrónimo utilizado en la literatura inglesa para denotar los sistemas de Inferencia Difusa: El módulo del tipo Mamdani está complementado con los siguientes operadores: intersección (valor mínimo), unión (valor máximo), implicación (valor mínimo), agregación (valor máximo) y defuzificación (valor máximo).

El ejercicio para definir la estructura del FIS consistió en la creación de 100.000 patrones aleatorios para las variables de entrada y consecuentemente analizar el espectro de respuesta en 120 combinaciones de operadores con los estadísticos básicos de la serie (mínimo, máximo, rango, media, desviación, asimetría, curtosis y coeficiente de variación).

En la Figura 8 se muestran los histogramas de los estadísticos de las simulaciones para distintas estructuras (combinaciones de operadores), se puede observar que más del 75% de las estructuras conservan valores mínimos en las simulaciones del Nivel de Comprensión (NC) por debajo de 4, en cambio un 60% aproximadamente de las estructuras alcanza a simular valores máximos mayores o iguales a un NC igual a 9.

En cuanto a la medida del valor central de las series, se tiene que más del 90% de las estructuras simulan series con valores promedios entre 4.5 y 5.5 indicando un pivote central adecuado para la modelación del Nivel de Comprensión.

Para el análisis del espectro de respuesta del modelo, en este caso la dispersión de la serie, se tienen como principales referentes el rango, las desviaciones y el coeficiente de variación. En cuanto al rango las estructuras empiezan a mostrar falencias puesto que idealmente deberían simular valores desde 0 hasta 10; sin embargo, apenas un 25% de las configuraciones de FIS son capaces de representar este espectro de respuestas.

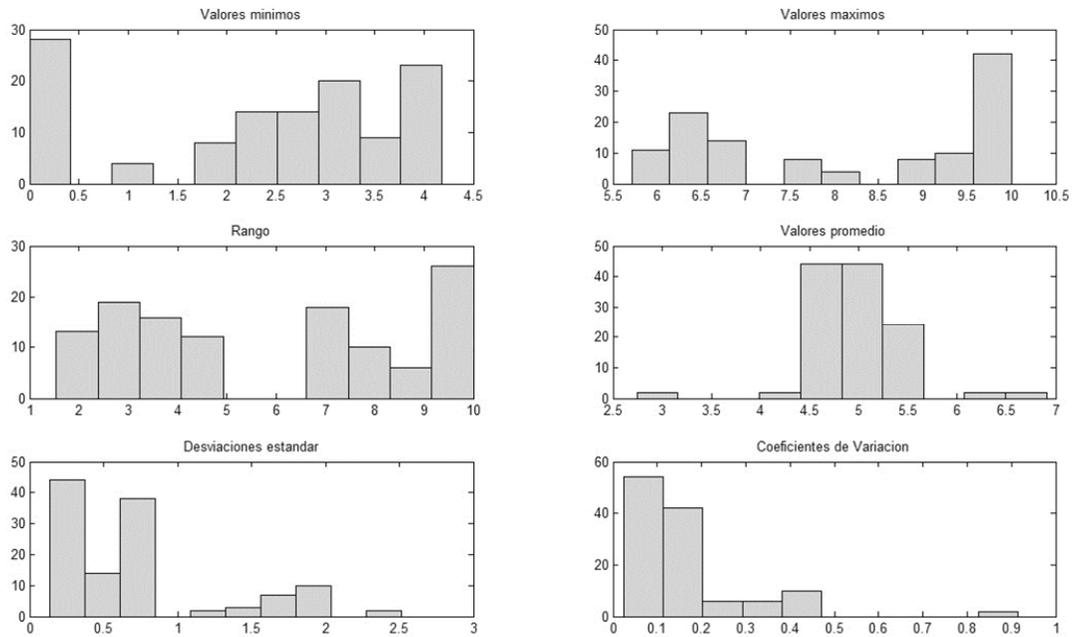


Figura 8. Resultados de las combinaciones de estructuras

De tal forma el análisis de los modelos se reduce a tomar en cuenta las estructuras que reportan un espectro de respuesta con rango mayor o igual a 8 (35% de los modelos), así, el estudio de la mejor estructura se centra en el análisis de la desviación estándar y el coeficiente de variación puesto que el rango se ve altamente influenciado por los valores extremos de las series.

Con respecto a la desviación estándar de las series es determinante la baja dispersión de información que se muestra en el histograma, ya que alrededor del 80% de las series tienen desviaciones menores a la unidad, lo cual representa una concentración de las respuestas que no es adecuada para el modelo de simulación del Nivel de Comprensión. Asimismo el 90% de los coeficientes de variación son menores a 0.2 indicando que la mayoría de estructuras tienen bandas de respuestas muy estrechas.

En definitiva la mejor estructura del FIS se evalúo con base en los siguientes criterios: modelos con rangos del nivel de comprensión simulado superiores a 8, valores promedios entre 5.1 y 6.9 (valores que corresponden a  $\pm 15\%$  de 6, umbral definido para la comprensión de actividades) y valor mínimo del criterio de elección (CE) ecuación 2:

$$CE = 0.2CA + 0.4 \frac{Curtosis-3}{3} + 0.4 \frac{maxCV-CV}{maxCV}$$

Equation 2

La formulación del CE se basa en los siguientes criterios: el coeficiente de asimetría, CA, debe ser cercano a cero (simetría del histograma), el coeficiente de apuntamiento, Curtosis, debe ser similar a 3 (curtosis de la curva normal) y el coeficiente

de variación debe estar cercano al máximo evaluado; los factores de ponderación 0.2 para la asimetría y 0.4 tanto para la curtosis como para el coeficiente de variación, representan la importancia arbitrariamente consignada en cada uno de los estadísticos.

#### 2.4.4. Agente de Retroalimentación (AR)

Este agente ejecuta las labores de seguimiento de cada estudiante almacenando las características de las UE presentadas y los resultados obtenidos (NC), conformando la matriz del historial académico del estudiante.

Adicionalmente, se encarga de actualizar el ND para presentar la nueva actividad al estudiante, a través de sentencias condicionales que establecen la necesidad de bajar, mantener igual o subir el ND de la próxima actividad de acuerdo con los umbrales definidos en el AAC.

La comunicación con el AAC también define la activación del Agente de Evaluación Adaptativa (AEA) con base en los índices de evaluación entregados por el profesor.

Asimismo actualiza el perfil académico del estudiante y lo comunica al ARA. Por último, la comunicación con el Agente Modelador del Historial (AMH) se ejecuta luego de cumplir algún criterio definido en función de la simulación del NC. Para el MORE se propone tener en cuenta el número de patrones observados en la matriz del historial académico del estudiante.

#### 2.4.5. Agente Modelador del Historial (AMH)

Este agente se encarga de la simulación de la matriz del historial académico del estudiante entregada por el AR. La

simulación basada en la interacción entre los recursos de enseñanza como variables de entrada y el Nivel de Comprensión observado (NCo) como variable de salida, se realiza mediante Árboles de Decisión (DT), Sistemas de Inferencia Difusa (FIS) y Optimización con Colonias de Hormigas (ACO).

A continuación se muestra paso a paso el modelo de simulación del historial académico del estudiante.

- Pre-procesamiento del historial: En la mayoría de los casos, el historial del estudiante contendrá grupos de iguales entradas con diferentes NCo, lo cual corresponde a casos donde el sistema se ve obligado a presentar la misma actividad (puede ser por falta de disponibilidad de recursos) hasta obtener la comprensión del estudiante. El modelo MORE maneja la hipótesis de comprensión de una actividad en función de los recursos de enseñanza y no de procesos de repetición, es decir, el modelo toma como patrón a modelar el que representa la primera presentación.
- Categorización de las variables: Como el objetivo último es obtener un FIS ajustado a los valores observados en el historial (modelo supervisado), se deben obtener reglas de comportamiento entre variables de entrada y salida, por lo cual se proponen dos modelos diferenciados en el número de estados (dos o tres) en los cuales se categorizan las variables de entrada. La variable de salida (NC) conserva la misma estructura de tres estados (bajo, medio y alto).
- Minería de datos con Árboles de Decisión (DT): Con la información clasificada se procede a realizar la minería de datos con DT tipo ID3.
- Extracción de las reglas del DT: Posterior a la aplicación del DT y al análisis del número de instancias correctamente clasificadas se debe extraer el conjunto de reglas semánticas que definen la relación de las variables del problema.
- Definición del Sistema de Inferencia Difusa (FIS): Con base en los límites de la transformación de las series continuas en categóricas se establece la forma aproximada de los conjuntos difusos; asimismo las funciones sigmoideas y las gaussianas definen los grados de pertenencia (siguiendo los lineamientos del FLM inicial).
- Aplicación de las colonias de hormigas (ACO): Con la definición de los intervalos en los que se mueven los parámetros de las distintas funciones de membresía, se ejecuta el modelo de optimización combinatoria ACO en funciones continuas [5, 10].
- Elección del mejor modelo: Finalmente el AMH debe elegir el mejor modelo para entregar al ARA, para lo cual el MORE considera los siguientes estimadores en orden de importancia: índice de la eficiencia simple del modelo (efm), coeficiente de correlación lineal ( $r^2$ ) y la raíz del error medio cuadrático (rmse).

El efm propuesto en este trabajo resalta la importancia de la

simulación en un esquema general binario de comprensión o incompreensión, definido con base en el umbral (Uc). En síntesis el efm se calcula como el porcentaje de aciertos al comparar las series simuladas y observadas bajo el esquema general binario.

#### 2.4.6. Agente de Evaluación Adaptativa (AEA)

La evaluación adaptativa que se plantea teóricamente para este modelo, consiste en mantener una estructura de selección de pruebas similar a la conformada en el ARA teniendo como principales referentes el ND y las preferencias observadas en el historial en cuanto a los recursos de enseñanza.

En lugar de una base de unidades expuestas (BUE) se tendría una base de evaluaciones adaptativas (BEA) con una conformación estructural similar, a excepción de la codificación del tipo de actividad que en este caso no aplicaría.

Es importante resaltar que la evaluación adaptativa es una herramienta que permitiría ejecutar pruebas de conocimientos acordes con las preferencias del estudiante; sin embargo, los resultados de la evaluación estarían directamente ligados con la necesidad de replantear la secuencia de currículo más no la replanificación de actividades en sí misma.

### III. VALIDACIÓN DEL MORE

Una base de datos con las variables que utiliza el MORE requiere de un programa de educación específico, de tal forma el planteamiento teórico expuesto solo podrá ser verificado bajo un escenario real de aplicación; no obstante, con el fin de visualizar la forma de operación se configuraron series sintéticas de información a partir de números aleatorios uniformemente distribuidos.

#### 3.4. Configuración del MORE

Para probar el funcionamiento del MORE se generó una BUE ficticia como plataforma, asimismo se eligió una secuencia supuesta de currículo y los resultados del test psicopedagógico y académico. Además se establecieron los umbrales para la evaluación del NC y el ND, también se programaron respuestas aleatorias del estudiante en cuanto al NC alcanzado luego de la presentación de cada actividad.

El FLM inicial se estableció con base en los lineamientos anteriormente descritos, de tal forma el proceso a desarrollar se centra especialmente en la modelación del historial académico.

- Preprocesamiento de la información: El historial académico contenía 49 patrones que luego de la filtración para remover patrones repetidos se consolidó con 43 patrones a simular.
- Categorización de las variables: En la serie con dos etiquetas, llamada SC2S, los recursos de enseñanza (RE) de las series originales (SO) son clasificados con nivel de participación Bajo (1) o Alto (2) a partir del valor límite de 35.

Para la serie de tres etiquetas, denominada SC3S, la serie original de RE (SO) se clasifica en Bajo (1), Medio (2) y Alto (3) con base en los valores límite 15 y 30.

Para la variable de salida NCo, con dominio entre cero y diez, se tienen tres estados: Bajo (1), Medio (2) y Alto (3) con base en los límites 4 y 7.

- Minería de datos con Árboles de Decisión (DT). La complejidad de un DT se puede estimar previamente a partir de la combinación posible de etiquetas, de tal forma para la serie SC2S la combinación es de solo 8 valores (23) y para la serie SC3S el resultado es de 27 valores (33). Dada la sencillez relativa del árbol se decidió formar la estructura del mismo con el algoritmo ID3, el cual se caracteriza por evitar las simplificaciones del DT.
- Extracción de las reglas. Las respuestas de un DT se representan a partir de sentencias condicionales que relacionan las variables del problema.

Para el ejemplo de la serie SC2S la extracción de conocimiento estaría dada por una sentencia tal como: Si los recursos

audiovisuales (AV) son bajos (1) y los recursos Gráficos (GR) son bajos (1) entonces el nivel de comprensión observado (NCo) será bajo (1).

Un aspecto fundamental a revisar antes de incorporar las reglas extraídas del DT al FIS se centra en analizar los resultados del árbol DT con base en los índices de correcta clasificación y los valores de la matriz de confusión.

- Definición del Sistema de Inferencia Difusa (FIS). La estructura operativa del FIS tipo Mamdani sigue los mismos lineamientos del FLM inicial; sin embargo, la clasificación de las series en dos (SC2S) y tres estados (SC3S) obliga a configurar dos FIS: el FIS2S y el FIS3S.
- Aplicación de las colonias de hormigas (ACO) . La optimización se concentra en la evaluación del FIS con respecto a la variación de los parámetros que definen las funciones de membresía; para ilustrar la metodología considérese la Figura 9 donde se muestra la variación en la definición de los conjuntos difusos de entrada del FIS3S a partir del movimiento de los parámetros.

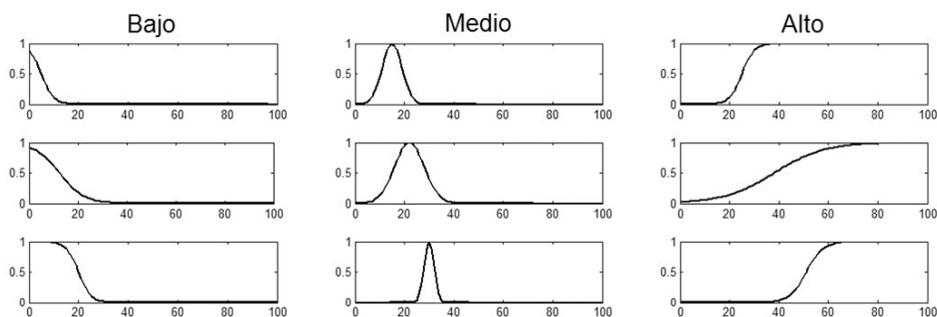


Figura 9. Ejemplo de variación de las funciones de membresía.

La configuración de parámetros del ACO obedece a experimentos realizados en otras investigaciones con funciones multi-mínimas tales como la función de Goldstein-Price [3,5], así, para la optimización se establecieron los siguientes parámetros: Número de hormigas = 20, número de exploradoras = 5, número de mensajes en la pila = 10, valor para establecer cercanía entre hormigas = 0.01.

- Elección del mejor modelo. La elección del mejor modelo depende del criterio de estimación utilizado, como se expresó anteriormente, en el MORE se utilizan tres criterios que son: efm, r2 y rmse.

#### IV. RESULTADOS Y ANÁLISIS

##### 4.1. FIS para la serie SC2S (FIS2S)

De la tarea de minería de datos con DT se tienen seis reglas condicionales a partir de las cuales se alcanza una correcta clasificación de casos observados en tan solo un 51%.

Este problema se simplifica bajo la suposición de iguales funciones de membresía para los RE, asimismo supone una simetría entre las funciones sigmoideas para los conjuntos difusos bajo y alto, así, el problema queda reducido a encontrar los dos parámetros óptimos de la función sigmoideal.

Para evaluar preliminarmente la influencia de los parámetros en las respuestas del modelo se efectuó un análisis combinatorio de parámetros regido por la técnica conocida como fuerza bruta (Figura 10).

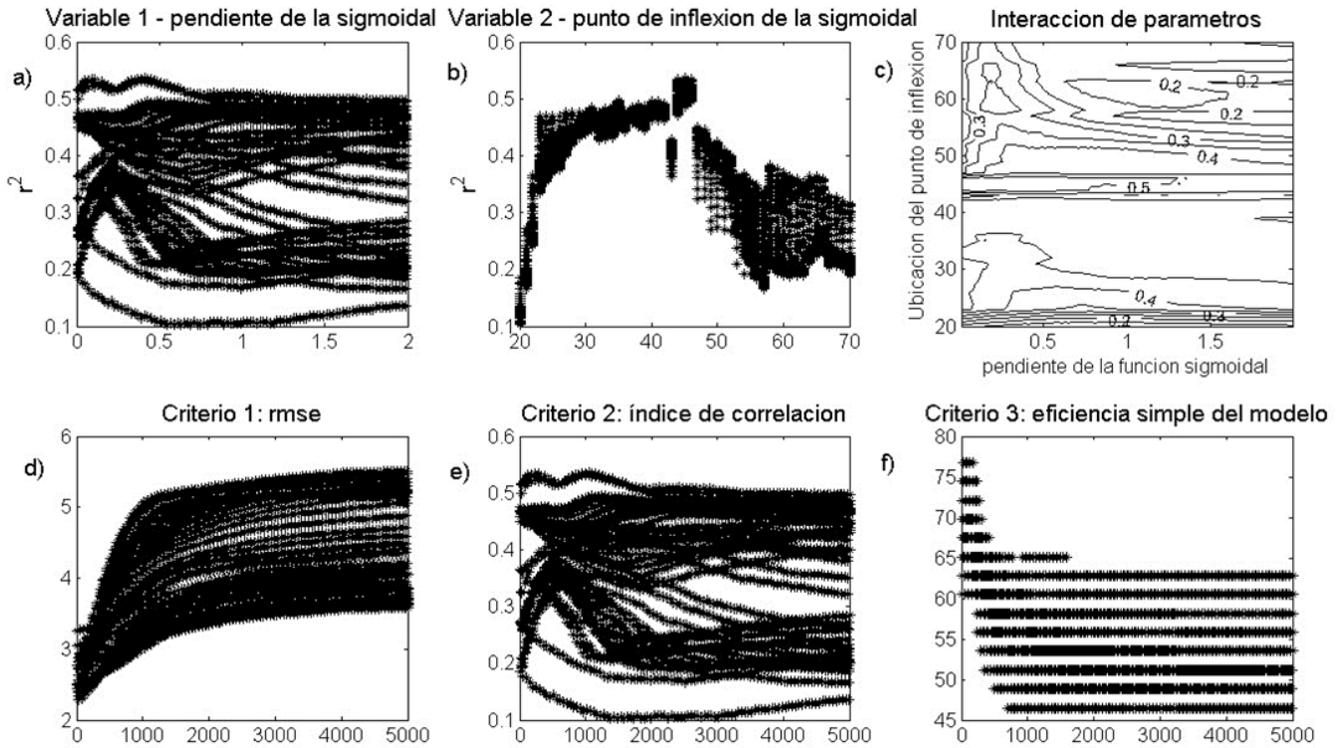


Figura 10. Sensibilidad del modelo y criterios de ajuste.

Las Figuras 10a, 10b y 10c están en función del índice de correlación lineal ( $r^2$ ), por lo cual se puede aducir una aproximación regular del modelo si se tiene en cuenta que los valores máximos obtenidos oscilan alrededor de 0.5 y la correlación lineal ideal es igual a 1.

En cuanto al parámetro 1 que define la pendiente en la sigmoidal, no se detecta un valor mínimo global, inclusive este parámetro no tiene una tendencia clara que indique un valor óptimo global (Figura 10a).

Para el parámetro 2 (definición del punto de inflexión de la sigmoidal), la Figura 10b muestra una forma pseudo-parabólica negativa con un máximo global identificable en la abscisa correspondiente a una participación de recursos del 45%.

La insensibilidad paramétrica, definida como el bajo efecto en el modelo con la variación de algún parámetro, visualizada inicialmente para la variable que define la pendiente de la sigmoidal, se ratifica en el análisis de contornos de la función objetivo con respecto a los dos parámetros (Figura 10c), donde se destaca una curva plana para el contorno correspondiente a 0.5 (máximo valor alcanzado para el  $r^2$ ) que indica la baja influencia del parámetro 1.

Por otro lado, se evaluaron distintos criterios de ajuste de la simulación, en la Figura 10d se muestra la evaluación del rmse, mostrando una amplia franja de valores entre 3.5 y 5.5 que no

representan un buen ajuste; a pesar de esto, la nube de puntos converge en un valor cercano a 2.5 que define los modelos con mejor aproximación bajo la concepción de este criterio.

Asimismo, la Figura 10e muestra la sensibilidad del modelo a la definición de los parámetros ya que se obtienen correlaciones con un espectro de respuesta entre 0.1 y 0.53. En esta figura se destaca la presencia de dos máximos locales en la banda superior que define las máximas correlaciones lineales.

De la misma forma, la Figura 10f, correspondiente a la eficiencia simple del modelo, muestra que el FIS2S alcanza eficiencias del 77% con respecto a la evaluación de comprensión o incomprensión del estudiante.

En síntesis, para el problema hipotético planteado, se observa la insensibilidad del parámetro 1 (pendiente de la sigmoidal), por lo cual el problema queda definido con base en el parámetro 2 (posición del punto de inflexión de la sigmoidal). En la Figura 11 se presenta el comportamiento del criterio elegido para el ajuste  $r^2$ , con base en los parámetros del modelo 1 y 2, demostrando una vez más la insensibilidad del parámetro 1 ante la proyección constante de la figura en este sentido de la función, asimismo se destaca otra vez la cresta de la función objetivo cuando el parámetro 2 es igual a 45.

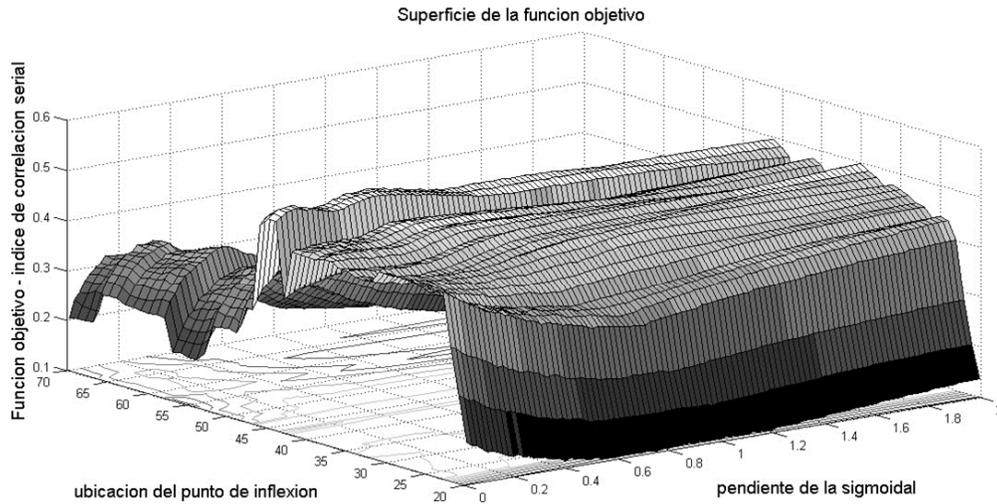


Figura 11. Superficie de la función objetivo.

En la Figura 12 se presentan los mejores conjuntos de parámetros en función del criterio de ajuste.

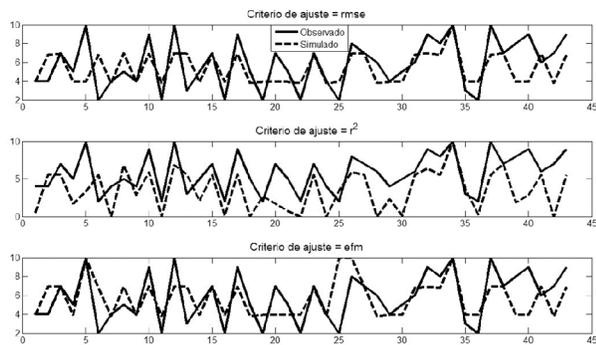


Figura 12. Mejores conjuntos de parámetros en distintos criterios de ajuste.

#### 4.2. Aplicación del ACO en el FIS2S

La fortaleza del ACO se centra en los problemas combinatorios de alto grado; no obstante, se efectuaron evaluaciones para observar las fortalezas del algoritmo medidas en tiempo computacional y numero de iteraciones.

En la Figura 13 se presenta el seguimiento de una de las soluciones encontradas con el ACO, el ejemplo tiene 4 hormigas y 3 exploradoras por individuo. La Figura 13 a) muestra que en el ciclo 1 una de las hormigas se ubica cerca de la solución óptima, posteriormente para el ciclo 5 se muestra como tres hormigas convergen cerca del óptimo global (Figura 13 b) y finalmente en el ciclo 20 llegan dos hormigas a la solución óptima, una tercera queda muy cerca y la cuarta sigue su exploración en la parte baja del dominio de la función (Figura 13c).

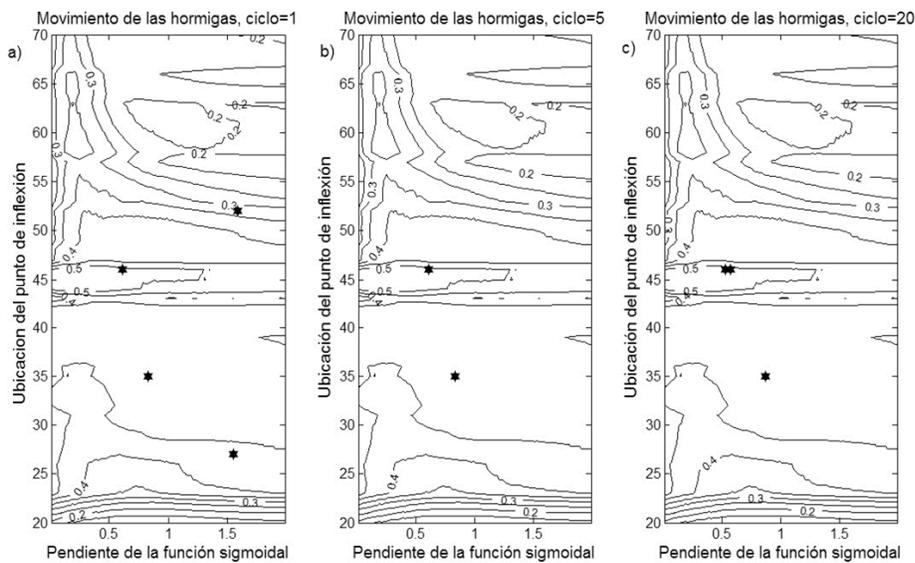


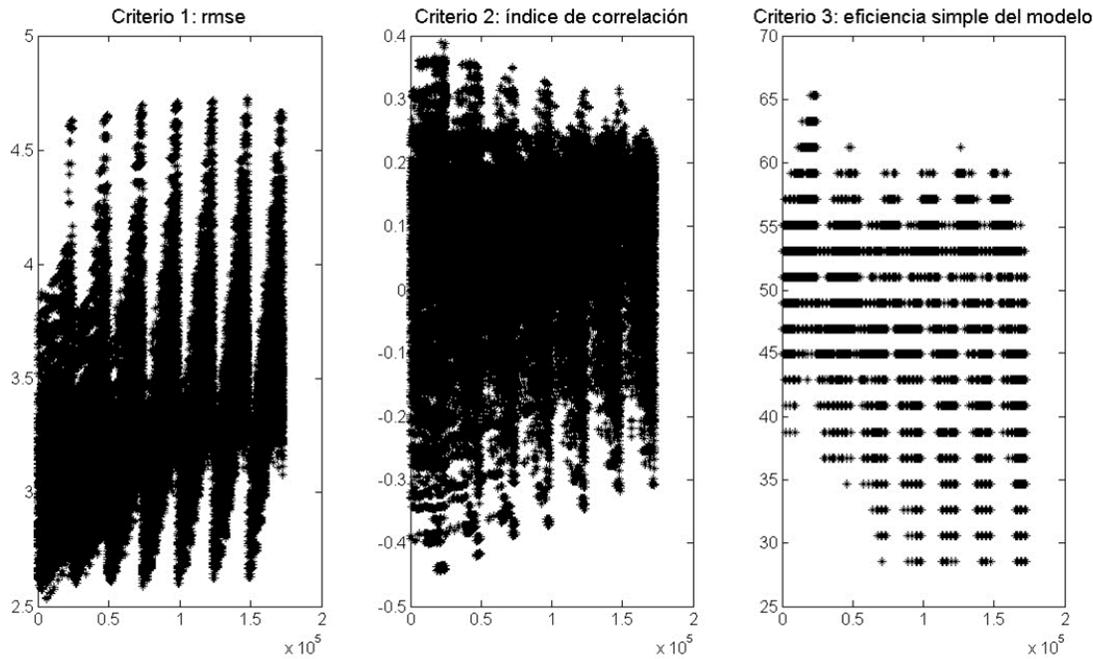
Figura 13. Movimiento de las hormigas en la optimización del FIS2S.

Es importante resaltar el aislamiento de una de las hormigas en el proceso, ya que refleja la capacidad que tiene el algoritmo de explorar varias regiones simultáneamente evitando caer rápidamente en óptimos locales.

En materia de eficiencia del algoritmo, en la Tabla 1 se muestra que el factor de compresión en cuanto al número de

**Tabla 1.** Comparación de resultados para FIS2S.

| Criterio de evaluación | Fuerza Bruta |      |          |        | ATAFIS |          |        |
|------------------------|--------------|------|----------|--------|--------|----------|--------|
|                        | Mín          | Max  | No Eval. | t(seg) | Mejor  | No Eval. | t(seg) |
| <i>rmse</i>            | 2.28         | 5.51 |          |        | 2.28   | 234      | 1.06   |
| $r^2$                  | 0.1          | 0.53 | 5100     | 36.96  | 0.53   | 210      | 1.02   |
| <i>efm</i>             | 46.5         | 76.7 |          |        | 69,8   | 173      | 0,82   |



**Figura 14.** Ajuste del modelo FIS<sub>3s</sub>

#### 4.4. Aplicación del ACO en el FIS3S

Aunque el FIS3S mostró problemas en el acoplamiento de los DT con el FIS, de nuevo la metaheurística ACO demostró la fortaleza en encontrar soluciones aproximadas a las mejores posibles en tiempos relativamente breves.

En la Tabla 2 se presentan los resultados obtenidos mediante combinaciones configuradas por fuerza bruta y mediante la metodología ATAFIS (Adapting decision Trees and Ant Colony in Fuzzy Inference System).

evaluaciones es en promedio del 4% (206/5100), y en cuanto al tiempo empleado en encontrar la solución es del orden del 2.7% (1/37).

#### 4.3. FIS para la serie SC3S (FIS3S)

Luego de aplicar los DT se obtienen 17 reglas condicionales con un índice de correcta clasificación de la información igual al 74%, condición que supone mayor precisión del FIS3S en comparación con el FIS2S.

La Figura 14 muestra que los mejores ajustes no representan fortalezas del modelo con respecto a la tarea inicial de los DT (rmse mayores a 2.5, correlaciones menores a 0.4 y eficiencias simples menores al 65%).

**Tabla 2.** Comparación de resultados para FIS3S.

| Criterio de evaluación | Fuerza Bruta |      |          |        | ATAFIS |          |        |
|------------------------|--------------|------|----------|--------|--------|----------|--------|
|                        | Mín          | Max  | No Eval. | t(seg) | Mejor  | No Eval. | t(seg) |
| <i>rmse</i>            | 2.6          | 4.9  |          |        | 2.8    | 3028     | 24.3   |
| $r^2$                  | -0.4         | 0.4  | 172872   | 24180  | 0.3    | 3032     | 25.2   |
| <i>efm</i>             | 27.9         | 62.8 |          |        | 58.2   | 3031     | 24.5   |

En materia de eficiencia del algoritmo, en la Tabla 2 se muestra que el factor de compresión, en cuanto al número de evaluaciones es en promedio del 1,8% (3030/172872), y en cuanto al tiempo empleado en encontrar la solución es del orden del 0,1% (3030/24180).

## V. CONCLUSIONES

Si bien el modelo de replanificación de actividades MORE se basa inicialmente en investigaciones focalizadas en los estilos de aprendizaje, la fase de refinamiento del modelo, ejecutada por el agente modelador del historial académico (AMH), se centra únicamente en los niveles de comprensión (NC) observados y los recursos de enseñanza utilizados en las actividades, así, el MORE se configura como un proceso de análisis personalizado en un ambiente gobernado por la diversidad y complejidad en torno a las tareas de enseñanza.

La base de unidades expuestas (BUE) como plataforma del MORE debe conformarse siguiendo metodologías precisas en lo referente al cálculo del nivel de dificultad asociado a cada actividad (ND) y a la participación estimada de los recursos de enseñanza (RE) en cada presentación.

La tarea dinámica de elección de actividades, ejecutada por el agente de recomendación de cursos (ARC), está basada en el ND y el NC simulado, por tanto la eficiencia de esta labor se ve condicionada por la variedad de actividades en la BUE.

En cuanto a la manipulación del ND para presentar las actividades, tarea realizada por el agente de retroalimentación (AR), en el MORE se proponen sentencias condicionales basadas en umbrales definidos por los profesores; no obstante, es importante evaluar otras metodologías dada la importancia de la variación de este parámetro.

En cuanto al modelo de seguimiento del historial académico, evaluado con series generadas aleatoriamente, se obtuvieron resultados aceptables en cuanto a la eficiencia computacional lograda con el ATAFIS. No obstante, la presunción de la existencia de patrones en el aprendizaje los estudiantes, facilita la precisión del modelo propuesto.

La metodología ATAFIS, se destaca como un proceso de refinamiento y optimización combinatoria que se genera a partir de los resultados iniciales obtenidos con Árboles de Decisión (DT); no obstante, la complejidad del modelo ATAFIS en el acoplamiento de varias técnicas de inteligencia artificial debe ser verificada, puesto que se obtuvieron resultados para el FIS3S de menor precisión que los logrados primariamente con los DT.

Es importante replantear, a partir de la puesta en marcha del modelo, los criterios que definen el modelo inicial de Lógica Difusa, inclusive sería interesante programar como etapa posterior del MORE, la simulación de los resultados de grupos de estudiantes con algún criterio de selección (cluster) para aprovechar los historiales en la presentación inicial de los cursos.

## RERERENCIAS

- [1] Carver C. A., Howard R. A., and Lane W. D., 1999. Enhancing student learning through hypermedia courseware and incorporation of student learning styles. *IEEE Transactions on Education*, 42(1):33–38.
- [2] Constantino G. M. A., Suthers D., and Escamilla De Los Santos J. G., 2003. Coaching web-based collaborative learning based on problem solution differences and participation. *International Journal of Artificial Intelligence in Education*, 13:263 – 299.
- [3] Coy C. C. A., 2005. Implementación en hidrofónica de un método de optimización matemática basado en la colonia de hormigas. Technical report, Pontificia Universidad Javeriana, Facultad de ingeniería.
- [4] Mitrovic A., 2003. An intelligent SQL tutor on the web. *International Journal of Artificial Intelligence in Education*, 13:173 – 197.
- [5] Brochero D., 2008. Predicción de lluvia en una región colombiana con un sistema experto basado en conocimientos indígenas.
- [6] Duque M. N. D., Jimenez B.J.A., and D.O. Carranza., 2007. Adaptive planning system for virtual courses. *IEEE Conference, Electron-ics, Robotics and Automotive Mechanics, CERMA*, (25-28):572 – 578.
- [7] Duque M. N. D., Ovalle C. D., Vicari R., and Azambuja S. R., 2008. Modelo inteligente genérico para adaptatividad de cursos virtuales. Technical report, CINTED-UFRGS Novas Tecnologias na Educacao.
- [8] Melis E., Andrés E., Büdenbender J., Frischauf A., Gogvadze G., Libbrecht P., Pollet M., and Ullrich C., 2001. Activemath: A web-based learning environment. *International Journal of Artificial Intelligence in Education*, 12(4):385–407.
- [9] Weber G. and Specht M., 1997. User modeling and adaptive navigation support in www-based tutoring systems. Technical report, In Anthony Jameson, Cecile Paris, and Carlo Tasso (Eds.) *Proceedings of the Sixth International Conference on User Modeling, UM97*, Springer, Wien New York.
- [10] Dréo J. and Siarry P., 2002. A new ant colony algorithm using the heterarchical concept aimed at optimization of multimimima continuous functions. Technical report, Université de Paris XII Val-de-Marne.
- [11] Felder R. M. and Soloman B. A., 1997. Index of learning styles questionnaire. <http://www.engr.ncsu.edu/learningstyles/ilsweb.html>.
- [12] Felder R. M. and Silverman L. K., 1998. Learning and teaching styles in engineering education. *Engr. Education*, 78(7):674–681.
- [13] Virvou M. and Moundridou M., 2003. Analysis and design of a web-based authoring tool generating intelligent tutoring systems. *Computers & Education*, 40:157–181.
- [14] Henze N. and Nejd W., 2001. Adaptation in open corpus hypermedia. *International Journal of Artificial Intelligence in Education*, 12:325–350.
- [15] Brusilovsky P. and Peylo C., 2003. Adaptive and intelligent web-based educational systems. *International Journal of Artificial Intelligence in Education*, 13:156–169.
- [16] Graf S., 2007. Adaptivity in Learning Management Systems Focussing on Learning Styles. Phd thesis, Vienna University of Technology Faculty of Informatics.
- [17] Murray T., 2003. Metalinks: Authoring and affordances for conceptual and narrative flow in adaptive hyperbooks. *International Journal of Artificial Intelligence in Education*, 13:199–233.

**Karol Lina López Bedoya.** Estudiante de Doctorado en Ingeniería Eléctrica e Informática. Universidad Laval, Québec, Canadá. Maestría en Ingeniería - Automatización Universidad Nacional de Colombia. Manizales. Ingeniera Electricista. Universidad Nacional de Colombia. Manizales. e-mail: karol-lina.lopez.1@ulaval.ca

**Néstor Darío Duque Méndez.** Ph.D. Doctorado en Ingeniería (c). Universidad Nacional de Colombia - Medellín, UNCM, Antioquia, Colombia. Maestría en Ingeniería de Sistemas. Universidad Nacional de Colombia. Medellín. Especialista en Sistemas. Universidad Nacional de Colombia. Manizales. Ingeniero Mecánico. Universidad Tecnológica de Pereira. Profesor Universidad Nacional de Colombia. Sede Manizales. Coordinador del Grupo de Investigación en Ambientes Inteligentes Adaptativos GAIA. Miembro de Comité editorial da Revista de Engenharia e Tecnologia. Universidade Estadual de Ponta Grossa, Paraná. Brasil; miembro del Comité Evaluador. Revista de Ingeniería. Universidad Tecnológica de Bolívar y del Comité Editorial Revista Gerencia Tecnológica Informática GTI. Edición 3 del año 2009. Parte del Comité Técnico de I Simposio Pluridisciplinar sobre Diseño, Evaluación y Descripción de Contenidos Educativos Reutilizables. SPDECE 2004. Universidad de Alcalá. España. Director Día Internacional de Seguridad en Computo 1999-2004  
e-mail: ndduqueme@unal.edu.co

**Darwin Brochero Bueno.** Estudiante de Doctorado en Ingeniería Civil. Universidad Laval, Québec, Canadá. Maestría en Ingeniería - Recursos hidráulicos. Universidad Nacional de Colombia. Bogotá. Ingeniera Electricista. Universidad Nacional de Colombia. Manizales. e-mail: darwin.brochero.1@ulaval.ca