

Sistemas Inteligentes Aplicados a los Procesos de Evaluación

Laura C. Díaz, Carlos A. Bartó

Departamento Computación, Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de Córdoba
lcd_ic@yahoo.com.ar
cbarto@gmail.com

Resumen

El objetivo de esta presentación es mostrar los procesos que involucran la implementación de un Sistema Virtual Inteligente para la Evaluación de asignaturas relacionadas con Programación. Si bien es incipiente el avance hacia la incorporación definitiva del Sistema Tutor Inteligente elegido para ser aplicado en la asignatura Informática de las carreras de Ingeniería de la Facultad de Ciencias Exactas Físicas y Naturales (FCEFYN), los avances en esa dirección conducen a repensar aspectos significativos que se pusieron en evidencia durante los procesos de prueba para su implementación. Se muestran los resultados alcanzados y las líneas de acción sugeridas. Este es parte de un proyecto más integral, que un equipo de Docentes Investigadores desarrolla actualmente en esta Unidad Académica.

Palabras clave: *evaluación, sistema inteligente, tutor virtual, enseñanza de programación.*

Introducción

Múltiples son los aspectos que despiertan el interés por el estudio de las Tecnologías de Información y Comunicación (TICs) en estos tiempos en que la construcción de saberes constituye un desafío importante para la Educación Superior. (Zabalza M., 2013)

Las mejoras que generan las estrategias centradas en la incorporación de dichas tecnologías, se refieren tanto al impacto en los indicadores de calidad en los procesos de aprendizaje y acreditación, como también a aspectos relacionados con la apropiación del conocimiento por parte de la sociedad. En ese

marco encuentran sentido las acciones desarrolladas desde este proyecto de investigación.

El objeto de esta presentación, a su vez forma parte de un conjunto de acciones orientadas a incorporar Sistemas Tutores Inteligentes (STI) en la enseñanza de la Programación como estrategia para mejorar el rendimiento académico de los estudiantes de Ingeniería. Específicamente se refiere a la mejora en las instancias de evaluación a los estudiantes para la acreditación de la asignatura Informática en las carreras de Ingeniería de la FCEFYN de la Universidad Nacional de Córdoba.

La mejora a introducir se materializa en dos sentidos: el primero a través de acciones de retroalimentación mejor articuladas con las dificultades reales del estudiante, el segundo hacia los resultados para expresar con mayor exactitud el nivel de conocimiento y habilidades adquirido por el estudiante.

Se propuso a SIETTE (Bartó, Díaz: 2012) como STI para implementar el proceso de evaluación y además se aplicó la teoría de modelos mentales (Johnson-Laird: 1993) para construir los modelos de retroalimentación que nutren la base de conocimiento del STI para mejorar el proceso.

Los Sistemas Tutores Inteligentes aplicados a evaluación

Una de las estrategias que utilizan los STI en las instancias de evaluación, consiste en el uso de los denominados test adaptativos informatizados (TAI) los cuales, partiendo de un conocimiento previo, se actualizan con los resultados obtenidos de su aplicación a estudiantes. A través de un TAI se obtiene como resultado la estimación del nivel de conocimiento (o denominado de forma

genérica, rasgo latente) del alumno. Según la teoría de los TAI, este resultado es independiente del test utilizado. Sea cual fuere el TAI administrado, para un alumno determinado, el nivel de conocimiento inferido debe ser el mismo (siempre que no medie, entre administraciones del test, proceso de aprendizaje alguno). Además, los TAI requieren un número menor de preguntas por test con respecto a los convencionales. Esto es debido a que, generalmente, este tipo de test muestra las preguntas de una en una, y cada pregunta es seleccionada en función de la estimación que hasta ese momento ha realizado el test del conocimiento del alumno.

Los TAI se basan en aplicar un algoritmo de evaluación, el cual se sustenta principalmente en una teoría psicométrica denominada Teoría de Respuesta al Ítem (TRI). Según esta teoría, la respuesta que un alumno da a una pregunta está relacionada con el nivel de conocimiento que posee. Esta relación se cuantifica mediante una o más funciones de densidad de probabilidad denominadas curvas características del ítem.

Antes de que un TAI esté operativo, es necesario calibrar sus preguntas. El procedimiento de calibración consiste en determinar las curvas características de cada pregunta. En general, los procedimientos que existen con este fin son muy costosos, puesto que requieren disponer de una muestra poblacional de alumnos de tamaño considerable (como mínimo del orden de la centena). A estos alumnos se les administra un test con las preguntas que se desea calibrar, de forma no adaptativa, y a partir de esos resultados se aplica un algoritmo de calibración.

El STI SIETTE y la Teoría de Respuesta al Ítem

SIETTE es un tutor inteligente que cuenta con un módulo para la evaluación con TAI. Con este tutor los profesores pueden definir sus tests y los estudiantes pueden tomarlos on-line. Los test se generan con las especificaciones definidas por el profesor y se adaptan, esto implica que las preguntas son elegidas

inteligentemente para ajustar el nivel de conocimiento del estudiante.

La Teoría de Respuesta al Ítem, basa su desarrollo en la presunción de la relación funcional entre los valores de la variable que miden los ítems –rendimiento académico alcanzado- y la probabilidad de acertar éstos. Dentro de los distintos modelos matemáticos para la representación de esta relación, la más utilizada es la función logística bi o tri parametrizada. El modelo logístico de dos parámetros no incluye la probabilidad de acierto al azar.

SIETTE, a través de las denominadas curvas características del ítem –CCI-, considera tres parámetros: a –índice de discriminación-, b- grado de dificultad del ítem- y c -probabilidad de acertar al azar-. Estos valores se calculan a partir de los datos obtenidos al aplicar los ítems a una muestra amplia y representativa de estudiantes (Conejo, R., Guzmán, E., Millán E. y otros: 2004).

El índice de discriminación es un valor proporcional a la pendiente de la curva. Un valor alto indica que la probabilidad de que un alumno con un rasgo latente estimado mayor que la dificultad del ítem acierte es mayor. Cuanto más discriminante es un ítem mejor contribuye a una estimación más precisa del conocimiento del alumno al corresponderse con una varianza menor.

El grado de dificultad corresponde al valor para el cual la probabilidad de responder correctamente al ítem es la misma que de responder de forma incorrecta, descontando el factor de adivinanza. Analíticamente representa el grado en el que la curva está desplazada a la izquierda (tendencia a la facilidad) o a la derecha (tendencia a la dificultad) con respecto al eje de abscisas.

El tercero representa la probabilidad de que un alumno sin conocimiento ninguno responda correctamente a la pregunta. Mediante este parámetro el modelo contempla el caso en el que un alumno responde de forma aleatoria al ítem y acierta.

En la TRI se asume que los ítems constituyen una sola dimensión, esto es, que mediante su aplicación se persigue la evaluación de un

único concepto y además suponen la independencia local, es decir que la respuesta a uno de ellos no puede estar condicionada a la respuesta dada a otros ítems. Por lo tanto, en el diseño de las herramientas evaluadoras es importante contar con alguna ontología que posibilite las categorizaciones necesarias para materializar esta condición. Esto es posible a través de la construcción de redes y/o mapas conceptuales.

Desarrollo del Trabajo

En una primera instancia se adoptó la discriminación de conceptos o ítems del libro de texto de la asignatura para el aprendizaje en lenguaje Python (Marzal, A y García, I: 2003). Actualmente se trabaja para construir a partir de ella una ontología en forma de red conceptual utilizando la herramienta Cmaps.

El proceso de evaluación de la asignatura actualmente consta de tres instancias: las Evaluaciones Conceptuales, la Especificación de Algoritmos y los Trabajos Prácticos.

Las Evaluaciones Conceptuales, adaptables a SIETTE, están organizadas actualmente en sesenta y tres ítems de igual peso, materializados en preguntas de tipo opción múltiple, distribuidos en siete test que se aplican a lo largo del desarrollo de la asignatura, con lo cual es posible la retroalimentación para el aprendizaje. Los contenidos abarcan todos los aspectos conceptuales desarrollados. El estudiante debe alcanzar un mínimo de sesenta por ciento (60 %) de rendimiento como condición excluyente para la acreditación.

Inicialmente se detectaron los modelos mentales erróneos de las estudiantes a partir de los resultados de experimentos realizados durante el primer cuatrimestre 2012. Esta información (Díaz L., Marangunic R., Bartó C.: 2013) constituyó la base del diseño del modelo de preguntas utilizado tanto en SIETTE como en MOODLE para poner en funcionamiento la segunda etapa del plan de trabajo propuesto para el primer cuatrimestre 2013. Los resultados, publicados en RUEDA (Díaz, Bartó, 2013) evidenciaron la necesidad de continuar alimentando la base de

conocimiento del tutor y además, de realizar una primera prueba piloto de evaluaciones inteligentes utilizando TAI.

Así, las acciones que se proponen para continuar en esta línea de investigación son:

1° Poner a punto el funcionamiento de SIETTE para la asignatura, aplicando evaluaciones de tipo opción múltiple sobre un nuevo grupo piloto de alumnos elegidos entre los que cursan Informática correspondientes a las carreras de Ingeniería Civil, Mecánica, Mecánica Electricista, Industrial, Aeronáutica y Química.

2° Alimentar la base de conocimiento del STI para la calibración de las Curvas características de Respuesta al Ítem (CCI).

3° Desarrollar actividades de formación con los docentes de la asignatura, del tipo seminario- taller, con el fin de consolidar la incorporación del STI.

4° Analizar la viabilidad de incorporar al STI como herramienta de entrenamiento para el estudiante.

5° Detectar automáticamente, mediante Redes Neuronales Artificiales, patrones asociados a los Modelos Mentales de los Estudiantes (MME) y su rendimiento académico.

Aspectos Metodológicos

Para implementar la segunda etapa del plan (1° cuatrimestre 2013) fue necesario diseñar y aplicar los nuevos instrumentos para las evaluaciones conceptuales tanto en SIETTE como en MOODLE.

Se estructuraron las siete evaluaciones con nueve ítems cada una, distribuidas a lo largo del libro de texto.

Las preguntas –ítems- se diseñaron con cinco opciones de respuestas, cuatro de ellas correspondían a los modelos mentales erróneos detectados en las evaluaciones tomadas durante 2012. Para ambas plataformas se diseñaron las preguntas con retroalimentación conforme a los MME, con lo cual se pretendía dar al estudiante una explicación de sus conceptos erróneos, facilitando así la disminución de la brecha

entre el modelo conceptual y su modelo mental.

Esta tarea implicó la preparación de sesenta y tres semillas que alimentaron al generador wgen (Cebollada y Verdaguer, 2013) que construye las instancias de preguntas diferentes a partir del producto cartesiano resultante de las combinaciones de valores de variables definidas por el usuario en cada semilla. Se generaron alrededor de 30 preguntas promedio, conceptualmente iguales, con diferentes valores de entrada y salida. Se pretendía enfrentar a los estudiantes a evaluaciones del mismo tipo y nivel de complejidad, con idénticos modelos mentales erróneos pero con valores diferentes.

Para cada semilla –ítem- se escogió una instancia –pregunta con valores determinados de entrada y salida- para cargar en SIETTE, mientras que todas las demás instancias fueron destinadas a MOODLE.

Se definió el tamaño de la muestra en alrededor del 10% del universo de alumnos del primer cuatrimestre para formar el grupo testigo. Se eligieron al azar los alumnos en cada una de las seis comisiones de la asignatura cuyo responsable era miembro del equipo de investigación, ello en razón de que los docentes que las aplicaban debían estar familiarizados con SIETTE.

Se aplicaron cada una de las evaluaciones conceptuales simultáneamente sobre el grupo testigo (SIETTE) y sobre el grupo control (MOODLE) conformado por los demás estudiantes, alrededor de cincuenta estudiantes para el primero y quinientos para el segundo.

Se realizó el análisis estadístico, prueba de hipótesis y test CHI-Cuadrada, siguiendo las pautas metodológicas del análisis correspondiente al periodo 2012, con la finalidad de validar la representatividad de los ítems aplicados en SIETTE en relación con los homólogos en MOODLE.

Posteriormente se determinaron los tres parámetros de las curvas características para cada uno de los ítems, tanto para SIETTE como para MOODLE considerando en esta primera aproximación al nivel de habilidad o rasgo latente como la calificación obtenida en

el test. El proceso de calibración de las curvas características para MOODLE no fue finalizado. Sin embargo, actualmente no es un objetivo inmediato ya que se piensa en la incorporación de las TAI sólo en la instancia de entrenamiento, para lo cual se cuenta con la calibración de la CCI que realiza SIETTE.

El uso de este STI como práctica habitual de los estudiantes como instancia previa a la acreditación, implicará el logro de uno de los grandes objetivos del proyecto de investigación: incorporar evaluaciones inteligentes en esta asignatura.

Las restantes acciones están previstas para llevarse a cabo durante 2014 y 2015.

Resultados

Se ha dividido esta sección en subsecciones, las dos primeras se refieren al tratamiento de los resultados obtenidos en 2013 y la tercera a las acciones requeridas para dar consolidar la incorporación de las evaluaciones inteligentes en la asignatura Informática durante 2014 y 2015. Se incluyen algunas precisiones significativas que tendrán incidencia a la hora de tomar decisiones en el rumbo de la investigación.

Análisis estadístico:

Del mismo modo en que se llevó a cabo para el 2012, se realizó el análisis confirmatorio, mediante test de hipótesis, del grupo testigo de preguntas –ítems- que se utilizó en SIETTE (Díaz, Bartó: 2013). Atendiendo al tipo de variable del calificador, se escogió un test de hipótesis de diferencia entre proporciones y una prueba Chi Cuadrada como respaldatoria del primero.

Los resultados muestran que sólo en el 65 % de los casos se acepta la hipótesis nula, con lo cual se sugiere hacer una revisión de: 1° la heterogeneidad de las preguntas –instancias de cada semilla- ya que pueden estar ocasionando diferencias significativas en sus niveles de complejidad y 2° la distribución de los individuos y el tamaño de las dos muestras presentadas para la prueba para que el tamaño

de los subgrupos resultantes permita la comparación.

Para la muestra del testigo, alrededor de 50 individuos, se aplicó una única pregunta proveniente de una instancia de la semilla. En tanto que en el grupo control, alrededor de 500 individuos, se aplicaron todas las instancias de la semilla, totalizando entre 18 y 48 preguntas diferentes según el ítem. Los grupos sobre los que se aplicaron cada una de las preguntas tenían un tamaño promedio oscilando entre 10 y 25 individuos.

Si bien la homogeneidad del grupo testigo es innegable y la distribución de preguntas entre los individuos del grupo control es aleatoria e independiente, no son suficientes estas condiciones para esperar que los resultados entre ellos resulten comparables.

En el gráfico 1 se muestran los valores de proporción para cada uno de los grupos, donde es posible observar lo anteriormente expresado.

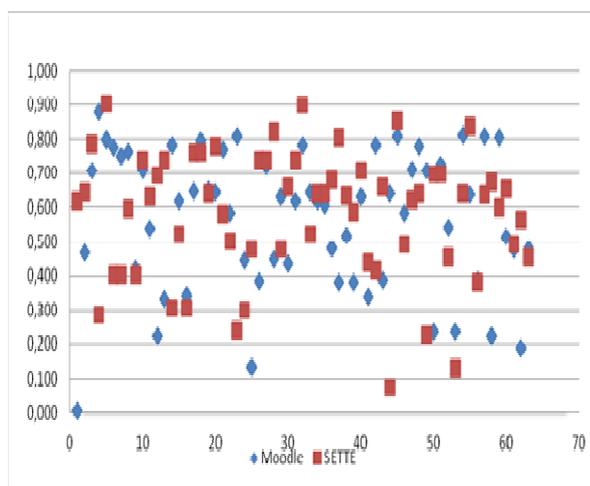
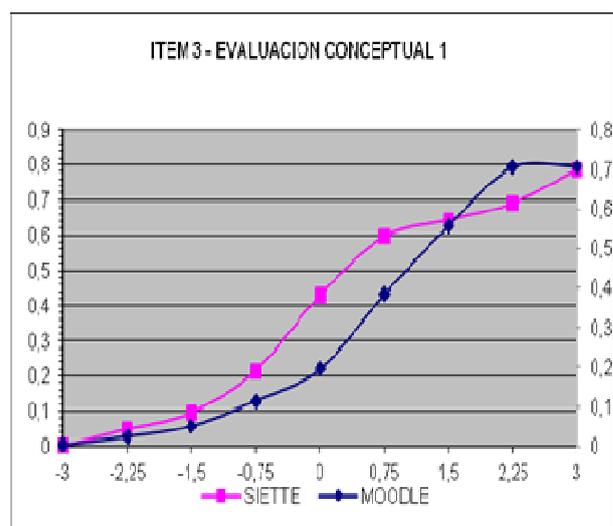


Gráfico 1: Porcentaje de aciertos para los 63 ítems en MOODLE y en SIETTE

Avances en la Determinación de los parámetros de las curvas características:

Sujeto a las consideraciones ya mencionadas y con la finalidad de reproducir el proceso de calibración de CCI de SIETTE para los resultados obtenidos bajo la plataforma MOODLE, se determinaron los parámetros de las curvas para los sesenta y tres ítems de ambos grupos. Sólo a modo de ejemplo, se

muestran dos curvas para un mismo ítem, una del grupo control y la otra del testigo. Como para los demás ítems, en este la curva de MOODLE presenta mayor discriminación, además, en general no se aprecia correlación significativa para el nivel de dificultad ni para el parámetro de acierto al azar. Todas las curvas del grupo control resultaron más suaves. En el gráfico 3 se muestra una salida del analizador de SIETTE.



ITEM	a	b	c
3 MOODLE	0,453	1,252	-0,05893
3 SIETTE	0,3338	-0,009656	-0,221

Gráfico 2: Aproximación de Curvas Característica para un ítem

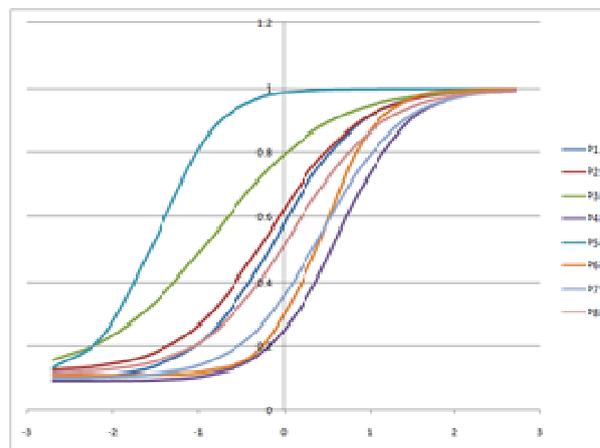


Gráfico 3: CCI - Evaluación Conceptual 7. Fuente: Analizador de SIETTE

Incorporación de evaluaciones inteligentes en la asignatura Informática:

Finalizado el experimento anterior, es posible hacer un análisis de factibilidad de la incorporación de evaluaciones inteligentes en forma definitiva.

En primer lugar, resta aún poner en marcha los TAI. Esto se llevará a cabo durante el cuatrimestre actual para una comisión en algunas instancias de evaluación.

Por otra parte, es necesario pensar una propuesta que contemple equilibradamente aspectos estratégicos y operacionales del STI. Esto implica responder a los siguientes interrogantes: ¿se implementarán para entrenamiento o para acreditación?; en cada caso, ¿cuáles son los requerimientos informáticos?; ¿cuáles son las limitaciones del acceso al servidor web de SIETTE? ¿Es conveniente contar con una plataforma con servidor propio? ¿Qué ajustes y mejoras restan por hacer al sistema de evaluaciones automáticas para hacerlo operativo? ¿Cuál es la capacitación requerida para los docentes?

Atendiendo a estos aspectos, el equipo de investigación actualmente está desarrollando un proceso de formación dirigido a los docentes de la asignatura con la finalidad de incorporar al STI como estrategia para el entrenamiento de los estudiantes. De los resultados del taller se desprenderán las mejoras y las líneas de acción para dar continuidad a estas acciones.

Por otra parte, en relación con el último objetivo de descubrir patrones de comportamiento de los estudiantes en relación con sus modelos mentales erróneos puestos en evidencia en las sucesivas evaluaciones conceptuales. Se aplicarán Redes Neuronales Artificiales del tipo aprendizaje no supervisado de mapas autoorganizados.

Conclusiones

Si bien en el desarrollo de esta presentación se evidencia que el volumen de resultados no es suficiente para el logro de los objetivos presentados; con las acciones realizadas se ha atendido a la razón primera de este proyecto de investigación: mejorar el rendimiento

académico de los estudiantes de Informática de las carreras de Ingeniería. La incorporación de preguntas con opción múltiple con retroalimentación en el primer cuatrimestre 2013, para las cuales las respuestas erróneas correspondían a los MME detectados por este equipo de investigación en el experimento realizado durante 2012, facilitó un mejor aprendizaje.

La puesta en marcha de SIETTE para nuestro curso de programación es otro objetivo en avance, aunque está pendiente la aplicación de las evaluaciones con TAI en esta plataforma. Además, a partir de los resultados del experimento, se cuenta con mejor información para avanzar en la detección de patrones de aprendizaje y rendimiento académico mediante el uso de inteligencia computacional, aplicando Redes Neuronales Artificiales, esto posibilitará un gran avance en esta dirección, ya que los resultados anteriores sólo predecían rendimiento académico a partir de la información referida a test predictores tomados al inicio de la cursada y a la continuidad del estudiante.

Si bien los resultados obtenidos hasta el momento no son definitivos, nos replantean la factibilidad de profundizar el trabajo en SIETTE y de desarrollarlo para MOODLE, aspectos ambos a definir durante el desarrollo del taller de capacitación actualmente en ejecución.

Referencias

- Bartó, C. A., and Díaz, L. C. "Intelligent Systems Applied to Computer Engineering Teaching". *Revista IEEE Latin American Transactions*, 2012.
- Carretero, Mario y otros, *Introducción a la psicología cognitiva*. Aiqué Grupo Editor, 2005.
- Cebollada y Verdaguer, M., "Generador de ejercicios sobre web para la enseñanza de la programación, basados en evaluación objetiva y respuesta corta para el LMS MOODLE". I Jornadas Nacionales, III Jornadas de la UNC, Experiencias e Investigación en Educación a Distancia y Tecnología Educativa, PROED, 2013.

Conejo, R., Guzmán, E., Millán E. y otros “SIETTE: A Web-Based Tool for Adaptive Testing”. *International Journal of Artificial Intelligence in Education* 14 (2004) 1 -33 IOS Press, 2004.

Díaz, L. C., Algorry, A. M, Eschoyez, M., Marangunic R. and Bartó, C. A., “Actions towards the application of intelligent systems in computer education”. *Revista IEEE Latin American Transactions*, 2012.

Díaz, L., Marangunic, R., Bartó, C., “Hacia La Detección De Los Modelos Mentales De

Los Estudiantes De Programación”. I Jornadas Nacionales, III Jornadas de la UNC, Experiencias e Investigación en Educación a Distancia y Tecnología Educativa, PROED, 2013.

Johnson-Laird, Phillip N. *Mental Models, Deductive Reasoning, and the Brain*. Department of Psychology, University of Princeton. N. J. U.S.A. 1993.

Marzal, Andrés y García, Isabel. “Introducción a la programación com Python”. 2003.