



# ESTUDIO DE LOS PRINCIPALES MÉTODOS PARA ESTIMAR EL CONOCIMIENTO LATENTE DE LOS ESTUDIANTES EN SISTEMAS E-LEARNING

STUDY OF THE MAIN METHODS TO ESTIMATE THE LATENT KNOWLEDGE OF STUDENTS IN E-LEARNING SYSTEMS

Ing. Lisset Salazar Gómez

[lsgomez@uci.cu](mailto:lsgomez@uci.cu)

Universidad de las Ciencias Informáticas, Cuba

M.Sc. Angel Alberto Vazquez Sánchez

[aavazquez@uci.cu](mailto:aavazquez@uci.cu)

Universidad de las Ciencias Informáticas, Cuba

## Resumen

Aumentar el conocimiento de los estudiantes es la meta primaria de la educación. Por tanto, la medición del conocimiento adquirido por los estudiantes adquiere vital importancia en el proceso educativo. La Estimación del Conocimiento Latente es una de las tareas predictivas de la Minería de Datos Educativos que está encargada de determinar en qué medida un estudiante posee un conocimiento o habilidad determinada en un momento dado. En el presente trabajo se realiza un estudio de los principales métodos empleados para estimar el conocimiento latente entre los que se encuentran Bayesian Knowledge Tracing, Performance Factor Analysis e Item Response Theory.

**Palabras clave:** Minería de Datos Educativos, Estimación del Conocimiento Latente, Bayesian Knowledge Tracing, Performance Factor Analysis, Item Response Theory

## Abstract

The main objective of education is to increase the knowledge of the students. Therefore, measuring the knowledge acquired by students is of vital importance in the educational process. Latent Knowledge Estimation is one of the predictive tasks of Educational Data Mining, which is responsible for determining, up to what extent, the students have a knowledge or specific skill. In this paper, a study of the main methods used to estimate the latent knowledge is made; among those methods are Tracing Bayesian Knowledge, Performance Factor Analysis and Item Response Theory.

**Keywords:** Educational Data Mining, Latent Knowledge Estimation, Bayesian Knowledge Tracing, Performance Factor Analysis, Item Response Theory

## 1. Introducción

En los sectores de consumo, se han recogido y analizado datos desde principios de 1990 para informar a las empresas sobre el comportamiento del cliente y sus preferencias. Una reciente tendencia en la educación ha tratado de emplear una analítica similar para mejorar la enseñanza y el aprendizaje en los cursos a nivel institucional (Johson, 2014).

La educación siempre ha tenido la capacidad de producir grandes cantidades de datos, comparado con cualquier otra industria, debido a que los estudios académicos requieren muchas horas de trabajo escolar y de tareas durante muchos años. Esto añadido con las interacciones con materiales produce una gran cantidad de datos (Aghabozorgi, Mahroei, Dutt, T. Y. Wah, and Herawan, 2014). Además, en la

actualidad las instituciones educacionales están generando cantidades enormes de datos, desde evaluaciones o notas de exámenes hasta números de matrículas o admisiones mientras realizan evaluaciones y admisiones en línea respectivamente (Wassan, 2015).

El potencial del uso de datos para mejorar los servicios, la retención de los estudiantes y el éxito de los mismos es evidente claramente (Johson, 2014). Debido a esto se ha establecido la aplicación de técnicas de Minería de Datos Educativos (MDE) para resolver diferentes problemáticas surgidas en el contexto educacional (Romero y Ventura, 2010) (Baker and K. Yacef, 2009).

Se puede definir la MDE como el empleo de las herramientas tecnológicas, algoritmos y las estrategias



de análisis de información utilizadas por la minería de datos, pero dentro de un contexto educativo para la búsqueda, análisis y la extracción de patrones de conocimiento, donde resuelvan problemas que mejoren el proceso de enseñanza-aprendizaje a partir de modelos predictivos de forma cualitativa y cuantitativa (Román, Sánchez-Guzmán y García, 2014).

La minería de datos y los softwares de analíticas ayudan a identificar y aplicar acercamientos pedagógicos para analizar grandes cantidades de actividades de estudiantes y profesores almacenadas en los servidores web de instituciones educacionales para visualizar y predecir el rendimiento de estudiantes, generando recomendaciones para los estudiantes, modelos de estudiantes, desarrollando softwares didácticos (courseware), planeando y planificando otras actividades (Wassan, 2015).

Una de las principales tareas de la minería de datos es la predicción, cuyo objetivo es desarrollar modelos que permitan inferir un aspecto simple de los datos. Y en la MDE, los clasificadores y las regresiones son los tipos modelos de predicción más comunes. El área de la estimación de conocimiento latente (ECL) es de particular importancia dentro de la MDE, y el trabajo en esta área en gran parte emerge del Modelado de Usuarios, la Inteligencia Artificial en Educación, y las tradiciones de las Métricas Psicométricas/Educacionales (Larsson y White, 2014).

Aumentar el conocimiento de los estudiantes es la meta primaria de la educación. Por tanto, si el conocimiento puede ser medido, puedes saber dónde lo estás haciendo mejor, puedes informar a los instructores (o cualquier otro interesado en el proceso) sobre el mismo y además puedes realizar decisiones pedagógicas automáticas (Baker y Corbett, 2014).

Inferir el conocimiento que dominan los estudiantes puede ser útil para varios objetivos, por ejemplo, puede ser una entrada significativa para otros tipos de análisis, puede ser útil para decidir cuándo avanzar un estudiante en el currículum (Larsson y White, 2014) o intervenir en otras vías (Roll, Alevan, McLaren, y Koedinger, 2007), y además puede constituir información útil para los instructores (Feng y Hefferman, 2007).

Para ello, existen diferentes algoritmos que permiten estimar el conocimiento latente de los estudiantes a partir del comportamiento sobre una habilidad, dentro de los algoritmos existentes se estará centrando el estudio en 3 algoritmos fundamentales como el Bayesian Knowledge Tracing (BKT), Performance Factor Analysis (PFA) e Item Response Theory (IRP).

Estos algoritmos son utilizados en los Sistemas de

Tutores Inteligentes (STI) o Sistemas Adaptativos Hipermedias Educacionales (AEHS, por sus siglas en inglés). Estos tipos de sistemas son una alternativa al acercamiento de “solo ponlo en la web”, tratando de adaptar la docencia a las necesidades de cada estudiante en particular. Los mismos permiten que se modele la enseñanza, el aprendizaje, la comunicación y el dominio del conocimiento del especialista y el entendimiento del estudiante sobre ese dominio.

Mediantes estos sistemas utilizados en la educación, se generan grandes cantidades de datos debido a que incorpora técnicas de IA (Inteligencia Artificial) a fin de crear un ambiente que considere los diversos estilos cognitivos de los alumnos que utilizan el programa (Giraffa, Nunes, y Viccari, 1997).

## 2. Materiales y métodos

Dentro de los métodos utilizados en la investigación se encuentra el Histórico - Lógico, el empleo de este método permitió llevar a cabo el estudio crítico del proceso de estimación del conocimiento latente, además permitió constatar teóricamente cómo han evolucionado los algoritmos para la estimación del conocimiento latente.

Otros de los métodos es la revisión documental, donde se detecta, obtiene y consulta la bibliografía y otros materiales donde se hayan aplicado los algoritmos de estimación latente. Se extrae y recopila la información relevante y necesaria que sirva para dar solución a la investigación. Esta revisión debe ser selectiva, puesto que cada año se publican en diversas partes del mundo miles de artículos de revistas, periódicos, libros y otras clases de materiales en las áreas del conocimiento. Con la revisión se logra clasificar las más importantes y recientes literaturas que sirven de referencia al trabajo.

La Estimación del Conocimiento Latente (LKE, por sus siglas en inglés) es una de las tareas predictivas de la MDE. Este tipo de tarea también es conocida como Inferencia de Conocimiento (KI, por las siglas de Knowledge Inference) o como Seguimiento del Conocimiento (KT, por las siglas de Knowledge Tracing); (Baker e Inventado, 2014); (Baker, 2015).

En (Steiner, Kickmeier-Rust y Albert, 2014) definen la LKE como la evaluación del conocimiento o habilidades que posee un estudiante. Para (Baker e Inventado, 2014) en LKE el conocimiento de los estudiantes sobre habilidades específicas es evaluado mediante sus patrones de exactitud en esas habilidades (y en ocasiones también otras habilidades). Según (Ritter, Harris, Nixon, Dickison, Murray y Towle, 2009) el KT modela la comprensión del estudiante mediante una colección de componentes de conocimiento lo que

permite estimar el conocimiento latente que posee.

Los principales objetivos de la inferencia de conocimiento en este contexto son (Baker, 2015):

- Medir qué sabe un estudiante en un momento específico.
- Medir qué componentes de conocimientos relevantes un estudiante conoce en un momento específico.

Los modelos usados para la estimación del conocimiento latente en MDE vienen de dos fuentes fundamentales: las tomadas de los acercamientos psicométricos tradicionales y las investigaciones en el modelado de usuario (en la inteligencia artificial) (Larusson y White, 2014).

Existen un amplio rango de algoritmos para LKE; los dos más usados actualmente son el Bayesian Knowledge Tracing (Corbett y Anderson, 1995) y Performance Factors Analysis (Pavlik, Cen y Koedinger, 2009) (cada uno con sus respectivas variantes y/o actualizaciones de mejora, por ejemplo (Beck, Chang, Mostow, y Corbett, 2008; Pardos y Heffernan, 2010; Baker, Corbett, y Alevan, 2008; Baker, Goldstein, and Heffernan, 2011) se ha comprobado que ambos tienen rendimientos equiparables en un número de análisis (Pardos, Gowda, Baker y Heffernan, 2012). Otro método que permite estimar el conocimiento que posee un estudiante es conocido por Item Response Theory (Clement A. Stone, 2015).

A continuación se realiza un estudio de los métodos relacionados anteriormente.

### 2.1 Modelo de rastreo del conocimiento (Knowledge Tracing)

El modelo de Rastreo o Seguimiento del Conocimiento (KT, por sus siglas en inglés) es una técnica de modelado de estudiantes ampliamente usada (Figura 1). Está basada en una red Bayesiana dinámica de dos estados, donde el rendimiento del estudiante es la variable observable y el conocimiento de los estudiantes es la variable latente. El modelo toma el rendimiento del estudiante y lo usa para estimar los niveles de conocimiento del estudiante. Como parte del entrenamiento del modelo dos parámetros de rendimiento son estimados: cometer un deslíz y adivinar, los cuales median entre el conocimiento del estudiante y el rendimiento del mismo (Gong, Beck y Heffernan, 2011).

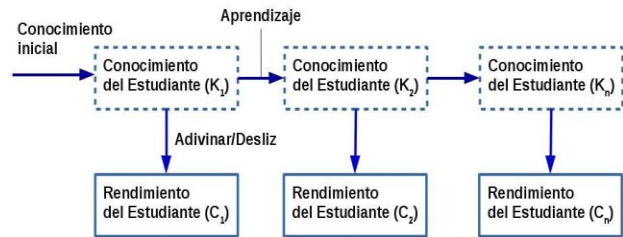


Figura 1. Modelo de Rastreo del Conocimiento.

Fuente: (Gong, Beck y Heffernan, 2011)

El algoritmo que implementa este modelo es conocido como Rastreo Bayesiano del Conocimiento (más conocido por Bayesian Knowledge Tracing, o BKT por sus siglas en inglés).

#### 2.2.1 Rastreo Bayesiano del Conocimiento (Bayesian Knowledge Tracing)

Este es el método clásico para medir con exactitud habilidades definidas en un entorno de aprendizaje en línea. Fue propuesto inicialmente por Richard Atkinson en los 60s, pero ha sido más estudiado y articulado por Albert Corbett y John Anderson a partir de 1995 (Baker, 2015).

El objetivo principal es medir cuán bien un estudiante conoce una habilidad o componente de conocimiento (KC, por sus siglas en inglés) en un momento específico, basado en su historial de rendimiento con esa habilidad o KC.

Según (van De Sande, 2013) este modelo viene en dos formas diferentes. La primera forma en la que aparece BKT es usando un Modelo Oculto de Markov (Doleck, Basnet, Poitras y Lajoie, 2015) el cual predice la probabilidad de aplicar correctamente una habilidad como la función de un número previo de oportunidades para aplicar esa habilidad y en los parámetros del modelo. La segunda forma es un algoritmo de seguimiento del conocimiento (KT) el cual usa el rendimiento del estudiante en cada oportunidad de aplicar una habilidad y actualiza la probabilidad condicional que el estudiante ha aprendido esa habilidad.

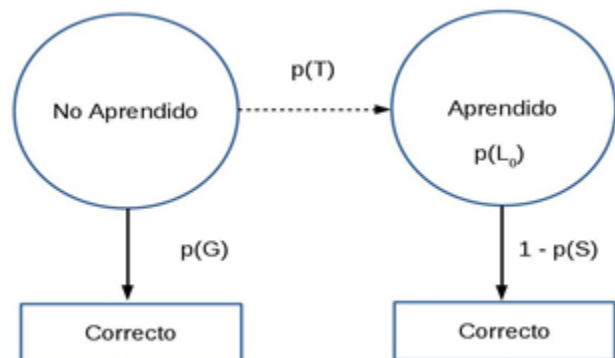


Figura 2. Modelo Oculto de Markov adaptado a BKT.  
Fuente: (Baker, 2015).

Las asunciones principales para este algoritmo son (Baker, 2015):

- Cada elemento debe involucrar solamente una característica o habilidad latente.
- Cada habilidad tiene cuatro parámetros. Desde esos parámetros, y el patrón de éxitos y fallos que el estudiante ha tenido en cada habilidad relevante hasta el momento podemos computar el conocimiento latente ( $P(L_j)$ ) y la probabilidad de que el estudiante responda correctamente el elemento ( $P(C_j)$ ).
- Es modelo de dos estados (cada habilidad está aprendida o no).
- En la solución de problemas el estudiante puede aprender la habilidad en cada oportunidad de aplicarla.
- El estudiante no olvida una habilidad una vez que la ha aprendido.

### 2.2.2 Rastreo Bayesiano del Conocimiento (Bayesian Knowledge Tracing)

El modelo BKT tiene cuatro parámetros (van De Sande, 2013) (Figura 2):

$P(L_0)$  - Es la probabilidad inicial de que el estudiante conoce una habilidad particular.

$P(G)$  - Es la probabilidad de que adivine (guess) correctamente, si el estudiante no conoce la habilidad.

$P(S)$  - Es la probabilidad de que cometa un desliz (slip), si el estudiante conoce la habilidad.

$P(T)$  - Es la probabilidad de aprender la habilidad si el estudiante no conoce la habilidad. Se asume que esta probabilidad es constante en el tiempo.

Se define el paso  $j$  como la  $j$ -ésima oportunidad para un estudiante de aplicar una habilidad dada. Sea  $P(L_j)$  la probabilidad de que un estudiante domine la habilidad en el paso  $j$ . De acuerdo con el modelo,  $P(L_j)$  puede ser determinado de la previa oportunidad por:

$$P(L_j) = P(L_{j-1}) + P(T)(1 - P(L_{j-1})) \quad (1)$$

En este modelo la probabilidad de que el estudiante responda correctamente la oportunidad es:

$$P(C_j) = P(G)(1 - P(L_j)) + (1 - P(S))P(L_j) \quad (2)$$

En las ecuaciones (1) y (2) se define el modelo oculto de Markov, donde  $P(L_j)$  es la variable "oculta".

Luego de realizar varias transformaciones la expresión (2) puede ser reescrita como:

$$P(C_j) = 1 - P(S) - Ae^{-\beta j} \quad (3)$$

Donde  $A = (1 - P(S) - P(G))(1 - P(L_0))$  y  $\beta = -\log(1 - P(T))$ .

### 2.2.3 Algoritmo de BKT

Para predecir  $P(L_j)$  para un estudiante individual se puede emplear el algoritmo de KT. En este caso buscamos  $P(L_j|O_j)$ , la probabilidad de que el estudiante ha aprendido la habilidad luego de completar el paso  $j$  dado el rendimiento del estudiante  $O_j$  en los pasos previos, donde  $O_j = \{o_1, o_2, \dots, o_j\}$  es el rendimiento del estudiante en las primeras  $j$  oportunidades y  $o_i$  puede ser correcto o incorrecto (van De Sande, 2013). Estas probabilidades condicionales responden a la recurrencia:

$$P(L_j|O_j) = 1 - \frac{(1 - P(T))[1 - P(L_{j-1}|O_{j-1})]P(G)}{P(G) + (1 - P(S) - P(G))P(L_{j-1}|O_{j-1})}, \quad o_j = \text{correcto} \quad (4)$$

$$P(L_j|O_j) = 1 - \frac{(1 - P(T))[1 - P(L_{j-1}|O_{j-1})](1 - P(G))}{1 - P(G) - (1 - P(S) - P(G))P(L_{j-1}|O_{j-1})}, \quad o_j = \text{incorrecto} \quad (5)$$

Para ajustar los parámetros se han usado diferentes enfoques, entre los que se encuentran Maximizar la Expectación (EM, por sus siglas en inglés) y un acercamiento por fuerza bruta (BF, por sus siglas en inglés). En EM se procesa el rendimiento de un estudiante como una pieza de evidencia con un orden temporal, y usa esta evidencia para la etapa de expectativa donde se calcula la probabilidad esperada; entonces se computa los parámetros que maximizan esta probabilidad y se ejecutan estos dos pasos de forma iterativa hasta que finalmente encuentra el mejor ajuste para estos parámetros. Este método no garantiza encontrar un global, dado que puede quedar en un máximo local. En el enfoque por fuerza bruta, por otro lado, intenta minimizar la Suma del



Error Cuadrado (SSE). Como originalmente los parámetros para BKT son continuos, no existe forma de componer un espacio de búsqueda finito, lo que es necesario para realizar una búsqueda exhaustiva. Para mitigar esta dificultad se limita el espacio de búsqueda considerando solamente dos lugares decimales de precisión, y de esta forma se reduce el espacio de búsqueda de infinito a 994 conjuntos de parámetros posibles para cada habilidad. En el algoritmo cada parámetro comienza en 0.01 y se incrementa en 0.01 en cada iteración. Finalmente, para cada iteración se encuentra el conjunto de parámetros que resulte en la menor SSE (Gong, Beck y Heffernan, 2011).

El algoritmo fue utilizado en un STI (el sistema utiliza 2 modelos fundamentales, modelo del alumno y modelo del profesor) en el área de proceso de diagnóstico del alumno mediante test es donde se utiliza el algoritmo. El diagnóstico es sin duda uno de los procesos más importantes dentro de cualquier STI, ya que la calidad del modelo del alumno dependerá la capacidad de adaptación del sistema (CONEJO et al, 2001). Respecto a las relaciones entre variables para medir el conocimiento, se considera que dominar un nodo de conocimiento tiene influencia causal en dominar aquellos nodos de conocimiento del nivel inmediatamente anterior en la jerarquía de granularidad que estén con él relacionados. En cuanto a la relación entre los nodos de conocimiento y las preguntas, se considera que poseer el conocimiento tiene influencia causal en responder adecuadamente a las preguntas. La red bayesiana resultante, donde los nodos se han etiquetado de la siguiente forma: A representa al nodo asignatura, cada  $T_i$  representa un nodo tema, cada  $C_i$  representa un nodo concepto y cada  $P_i$  un nodo pregunta tipo test, figura 3. Existe una parte que soporta el proceso de diagnóstico, en el que se determina a partir de las respuestas del alumno el conjunto de conceptos que conoce/no conoce, y la parte que soporta el proceso de evaluación, en el que a partir de los resultados obtenidos en el proceso anterior se determina una medida del grado de conocimiento alcanzado por el alumno, tanto en la asignatura como en cada uno de los temas de los que consta.

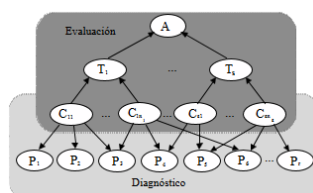


Figura 3. Red bayesiana para el diagnóstico mediante

test. Fuente: (CONEJO et al, 2001)

## 2.2 Análisis de los factores de rendimiento (Performance Factors Analysis)

Este modelo constituye una alternativa a BKT, trabajando en dirección de algunas limitaciones de este algoritmo; pero no alcanza todas las buenas características de BKT. Fue propuesto en 2009 por Pavlik, Cen y Koedinger, 2009 en "Performance Factors Analysis--A New Alternative to Knowledge Tracing.", Surge a partir de la reconfiguración del algoritmo de Análisis de los factores de aprendizaje (LFA, por sus siglas en inglés) y es mayormente conocido por el nombre en inglés Performance Factors Analysis (PFA) (Gong, Beck y Heffernan, 2011)

Mide cuanta habilidad latente tiene un estudiante mientras está aprendiendo. Pero la expresa en términos de exactitud la próxima vez que la habilidad es encontrada. No es una expresión directa de la cantidad de habilidad latente, excepto por esta probabilidad de responder correctamente (Baker, 2015).

Entre las ventajas que presenta este modelo se encuentran (Pavlik Jr, Cen, y Koedinger, 2009):

- Este modelo ha sido más ampliamente investigado como una solución al problema de múltiples rendimientos en los componentes de conocimiento.
- Modelos como este han sido empleados frecuentemente para predecir la duración de cada acción, dado que produce un valor estimado real de la fuerza para cada KC.

Este algoritmo puede ser usado en contextos donde (Baker, 2015):

- Es necesario estimar el conocimiento de un estudiante en un tópico X.
- Donde los estudiantes pueden aprender en cada elemento (debido a la ayuda, a la retroalimentación, etc).
- Cuando existen una secuencia de elementos que son puntuados de forma dicotómica (por ejemplo, el estudiante adquiere una nota de 0 o 1 en cada elemento).

Este modelo asume que el "rendimiento" es indicativo del aprendizaje de los estudiantes por dos razones. La primera, respuestas correctas son un fuerte indicativo de que la fortaleza actual ya es alta, por tanto, dar una respuesta correcta va a ayudar al modelo a incrementar la fuerza estimada. Y, en segundo lugar,



las respuestas correctas pueden llevar a mayor aprendizaje que las respuestas incorrectas (Pavlik Jr, Cen, y Koedinger, 2009). La sensibilidad a las respuestas incorrectas permite actuar como indicador y estimar el aprendizaje en el sentido inverso.

Los parámetros principales de este método son (Baker, 2015); (Pavlik Jr, Cen, y Koedinger, 2009):

- Cada elemento puede tener múltiples habilidades latentes o componentes de conocimiento (representado por KC).
- Cada habilidad tiene un índice de aprendizaje exitoso y un índice de fallo en el aprendizaje  $p$ .
- El parámetro  $\beta$  indica la dificultad para los KCs.

A partir de estos parámetros, y del número de éxitos y fallos que el estudiante ha tenido en cada habilidad relevante hasta el momento, podemos computar la probabilidad  $P(m)$  de que el estudiante responderá correctamente el ítem:

$$m(i, j \in KCs, s, f) = \beta + \sum_{j \in KCs} \gamma_j s_{i,j} + \rho_j f_{i,j} \quad (6)$$

$$P(m) = \frac{1}{1 + e^{-m}} \quad (7)$$

Donde  $s$  es el seguimiento de éxitos anteriores para el KC para el estudiante, y  $f$  es el seguimiento de los fallos anteriores para el KC.

El parámetro  $\beta$  se propone en (Pavlik Jr, Cen, y Koedinger, 2009) para tres elementos diferentes:

- Ítem
- Tipos de Ítem
- Habilidades

Al igual que en BKT los métodos usados para ajustar los parámetros son EM y fuerza bruta (Gong, Beck y Heffernan, 2011). Pero a diferencia de BKT la familia de modelos logísticos de descomposición del aprendizaje está basada en la forma de la regresión logística estándar, por lo que se asegura que el procedimiento alcance un máximo global, por tanto, existe un único ajuste óptimo de los parámetros.

Una variante del algoritmo PFA es Learning Factor Analysis (LFA) (Cen, Koedinger, y Junker, 2006) que se utiliza en un modelo cognitivo con un conjunto de reglas de producción o habilidades codificadas en tutores inteligentes para modelar como los estudian-

tes resuelven problemas. Por lo general, es generado por el refinamiento iterativo e intercambio de ideas entre los expertos en el tema, programadores y científicos cognitivos.

LFA tiene tres componentes: un modelo estadístico que cuantifica las habilidades, los factores de dificultad que pueden afectar el rendimiento del estudiante en el currículo del tutor, y una búsqueda combinatoria que realiza la selección del modelo.

Basado en un conjunto de asunciones se desarrolla un modelo estadístico de regresión logística múltiple

$$\ln\left(\frac{p}{1-p}\right) = \sum \alpha_i X_i + \sum \beta_j Y_j + \sum \gamma_j Y_j T_j \quad (8)$$

Donde,

- $P$  es probabilidad de obtener un ítem correctamente,
- $X$  las covariables para estudiantes,
- $Y$  las covariables para las habilidades,
- $T$  las covariables para el número de oportunidades de practicar las habilidades,
- $Y T$  las covariables para la interacción entre la habilidad y el número de oportunidades para esa habilidad,
- $\alpha$  es el coeficiente para cada estudiante, por ejemplo, la intercepción de estudiante,
- $\beta$  es coeficiente para cada regla, por ejemplo, el intercepto de producción,
- $\gamma$  el coeficiente para la interacción entre una producción y sus oportunidades, por ejemplo.
- El factor de dificultad se refiere específicamente a la propiedad del problema que causa dificultad en los estudiantes. Al evaluar la diferencia de rendimiento en pares de problemas que varían por un factor a la vez, podemos identificar los componentes de conocimiento ocultos que se pueden utilizar para mejorar un modelo cognitivo. Los factores de dificultad han sido usados para evaluar empíricamente un número pequeño de modelos alternativos.

El objetivo de la búsqueda combinatoria es realizar una selección de modelo en el espacio de modelos de regresión logística. Para este objetivo se utiliza el algoritmo de búsqueda  $A^*$ . Para limitar el espacio de búsqueda se emplea una heurística para organizar cada nodo y visitar los nodos en el orden que estima

la heurística. Estas heurísticas son el Criterio de Información de Akaike (AIC) [32] y el Criterio de Información Bayesiano (BIC) (Zhao,2013)

### 2.3 Teoría de respuesta al ítem

La Teoría de Respuesta al Ítem (más comúnmente conocido por *Item Response Theory*, IRT) es una de las teorías de prueba que pueden ser aplicadas para evaluar datos para describir como estimar, inferir y predecir una característica, rasgo o habilidad particular que una persona pueda hacer a partir de las respuestas a los ítems de un examen (Clement A. Stone, 2015).

Los modelos IRT involucran una clase de modelos matemáticos que pueden ser usados para predecir las respuestas de una persona para cada ítem de un examen mediante el uso de las características de la persona y las características de los ítems. Las características de la persona refleja uno o más constructos latentes siendo medidos por un instrumento (examen, encuesta o escala), y el nivel o estatus particular de la persona en cada constructo siendo medido es asumido para afectar el desempeño de la prueba (Clement A. Stone, 2015).

El objetivo principal de IRT es saber en qué medida una persona posee alguna característica latente. Y es típicamente usada para medir el conocimiento actual de un estudiante sobre un tópico determinado (Baker, 2015).

Para predecir o modelar las respuestas a los ítems, los modelos IRT consisten en uno o más parámetros latentes de la persona y uno o más parámetros del ítem. Estos parámetros, en conjunción con una función matemática, como la distribución normal o función logística (caso general de la función sigmoide), puede ser usada para modelar la probabilidad de una respuesta  $u_j$  al ítem  $j$ . Esta respuesta puede ser, por ejemplo, *correcta* o *incorrecta*, o la aceptación de una opción particular de respuesta. Una expresión general para la clase de modelos IRT es la siguiente:

$$P_j(U_j = u_j | \omega_j, \theta) \quad (8)$$

donde  $\theta$  es un vector de las características de la persona reflejando diferentes atributos siendo medidos, y  $\omega_j$  es un vector de parámetros de rasgos característicos para el ítem  $j$  que se asume que afecta el rendimiento.

El modelo IRT para un ítem puntuado de forma dicotómica (0,1; incorrecto, correcto; No, Si) es mostrado típicamente usando una curva de respuesta al ítem o curva característica del ítem (ICC, por sus siglas en

inglés). En la figura se muestra una ICC para un ítem puntuado de forma dicotómica, donde  $P_j(U_j = 1 | \omega_j, \theta)$  es la función de respuesta del ítem con una habilidad latente asumida para determinar el rendimiento del ítem. En la figura, el eje x corresponde con la habilidad latente ( $\theta$ ) siendo medida, y el eje y se corresponde con la probabilidad de contestar correctamente el ítem. Como se puede apreciar la función matemática refleja una relación monótona creciente entre la probabilidad de respuesta y el estatus o nivel de la habilidad siendo medida (Clement A. Stone, 2015).

En la ICC, el conjunto de parámetros del ítem ( $\omega_j$ ) para el modelo específico de IRT determina la forma de la función. El parámetro *pendiente* (*slope*) para ICC es usado para describir cuán rápido la probabilidad de respuesta cambia (sensibilidad) como una función de cambios en el rasgo latente  $\theta$ . Un segundo parámetro es el *umbral* (*threshold*), está en la misma escala que la habilidad latente  $\theta$  y describe donde la función de respuesta está centrada en la escala de  $\theta$ . La *asíntota inferior* (*lower asymptote*) puede también ser incluida para modelar la probabilidad de la respuesta de los estudiantes con bajo  $\theta$ . En el contexto educacional, estos tres parámetros (pendiente, umbral y asíntota inferior) son a menudo referidos como parámetros de *discriminación*, *dificultad* y *adivinación* (Clement A. Stone, 2015).

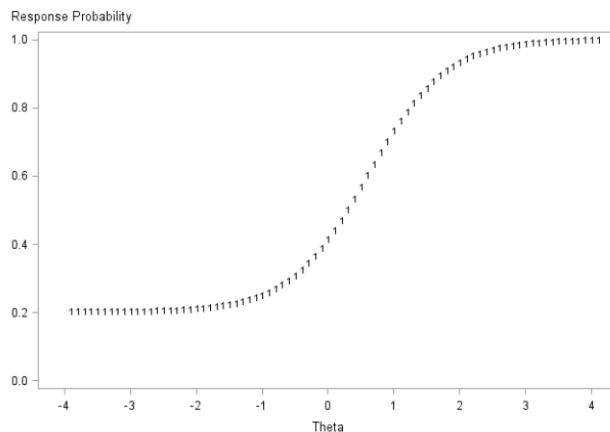


Figura 4. Ejemplo de una ICC para un ítem puntuado de forma dicotómica. Fuente: (Clement A. Stone, 2015).

Para respuestas puntuadas de forma dicotómica generalmente se utilizan tres modelos IRT: Modelos de 1, 2 y 3 parámetros. En la forma general para estos tres modelos, la probabilidad de que un estudiante reciba una respuesta (1 = correcta como con-



traría a 0 = Incorrecta) para el ítem  $j$ -ésimo es comúnmente modelado por una función logística:

$$P_j(\theta) = P_j(U_j = 1 | \omega_j, \theta) = \Psi(z_j) = c_j + (1 - c_j) \frac{e^{z_j}}{1 + e^{z_j}} \quad (9)$$

Donde  $\Psi(z_j)$  es la función logística acumulativa y  $z_j = D a_j (\theta - b_j)$  es una desviación logística;  $D$  es la constante de escalado (comúnmente se utiliza los valores 1 y 1.702 de manera indistinta);  $\theta$  referencia al parámetro de la persona (la habilidad examinada); y  $\omega_j$  referencia el vector de posibles parámetros de ítem para el ítem  $j$ , y  $c_j$  es la asíntota inferior para el parámetro  $j$  (Clement A. Stone, 2015); (DeMars, 2010).

El modelo de 3 parámetros (3PL) incluye los tres parámetros del ítem ( $a_j$ ,  $b_j$  y  $c_j$ ) y se aplica la fórmula expresada en Text para calcular la probabilidad de responder correctamente el ítem (Baker, 2015); (Clement A. Stone, 2015); DeMars, 2010).

El modelo de 2 parámetros (2PL) no incluye  $c_j$ , y es apropiado para aplicaciones donde la asíntota inferior se asume que sea 0 y que los estudiantes adivinen sea insignificante. Se considera una versión restringida del modelo 3PL. El valor de la probabilidad de contestar correctamente el ítem puede ser expresado entonces de la forma:

$$P_j(\theta) = \frac{1}{1 + e^{-a_j(\theta - b_j)}} \quad (10)$$

Finalmente, el modelo de 1 parámetro no incluye  $c_j$  y estima una pendiente común a través de todos los ítems ( $a_j = 1$ ). Este modelo es conocido como modelo Rasch, y por supuesto es el modelo más popular (Baker, 2015); (Clement A. Stone, 2015); DeMars, 2010). El valor de probabilidad de que el ítem sea respondido correctamente es:

$$P_j(\theta) = \frac{1}{1 + e^{-1(\theta - b_j)}} \quad (11)$$

El método más utilizado para ajustar los parámetros es maximizar la expectación.

Una manera en que fue utilizado el algoritmo IRT fue

en un STI, que utiliza una arquitectura de 3 modelos (modelo del dominio, modelo del estudiante y modelos pedagógicos), (Guzmán; Conejo, 2005); donde en el modelo del estudiante hacen inicialmente una recuperación de la información del estudiante, y del mismo hacen una recuperación y selección de los ítems que se van a evaluar (que se encuentra en un banco de preguntas), califican las respuesta asignando 1: respuesta correcta y 0: respuesta incorrecta, a partir de esa calificación entonces se aplica el algoritmo IRT donde estiman el nivel de conocimiento del estudiante, y por último generan los resultados de evaluación. Con este sistema propuesto facilita la evaluación de los estudiantes y se percibirá qué habilidades no tiene vencidas.

### 3. Resultados y discusión

Para la evaluación de los algoritmos se decide basarse en los criterios siguientes:

- Capacidad de estimar el conocimiento latente
- Cantidad de habilidades por ítem
- Métodos de ajuste de parámetros del modelo
- Métricas para la medición de resultados

En la Tabla 1 se muestra un resumen comparativo de los tres algoritmos estudiados respecto a los criterios escogidos. Como se puede apreciar todos estos algoritmos comparten características comunes como son el uso de métricas similares como Área debajo de la curva (AUC), Error Cuadrático Medio (RMSE).

Un elemento que los diferencia concretamente es la capacidad de estimación del conocimiento latente, siendo únicamente el algoritmo BKT quien calcula un estimado de la habilidad latente, así como la probabilidad de responder correctamente al próximo ítem con la presencia de dicha habilidad; el resto solamente calcula un estimado de la posibilidad de que se conteste correctamente el ítem asumiendo que es una medida de la posesión de la habilidad asociada con ese ítem.

Por otro lado, el algoritmo BKT posee un método de ajuste de parámetros (Probabilidad Empírica) que en "Learning bayesian knowledge tracing param-eters with a knowledge heuristic and empirical probabilities," demostró tener probabilidad predictiva similar a EM, pero que evita los modelos degenerados y trabaja considerablemente más rápido que el EM.

Criterio	Bayesian Knowledge Tracing	Performance Factors Analysis	Item Response Theory
----------	----------------------------	------------------------------	----------------------





Capacidad de estimar el conocimiento latente	Expresa la probabilidad de que el estudiante domine la habilidad latente, y además muestra la probabilidad de que el estudiante responda correctamente la próxima vez que enfrente un problema donde deba aplicar dicha habilidad.	No es una expresión directa de la cantidad de habilidad latente, excepto por la probabilidad de responder correctamente	Predice la respuesta de una persona ante un ítem determinado.
Cantidad de habilidades por ítem	Una habilidad por ítem	Múltiples habilidades por ítem	Una habilidad por ítem.
Métodos de ajuste de parámetros del modelo	<ul style="list-style-type: none"> <li>• Maximización de la expectativa</li> <li>• Fuerza Bruta</li> <li>• Probabilidad Empírica</li> </ul>	<ul style="list-style-type: none"> <li>• Maximización de la expectativa</li> <li>• Fuerza Bruta</li> </ul>	<ul style="list-style-type: none"> <li>• Maximización de la expectativa</li> </ul>
Métricas para la medición de resultados	<ul style="list-style-type: none"> <li>• Media del error absoluto (MAE)</li> <li>• Error Cuadrático Medio (RMSE)</li> <li>• A'</li> <li>• Área debajo de la Curva (AUC)</li> </ul>	<ul style="list-style-type: none"> <li>• Error Cuadrático Medio (RMSE)</li> <li>• Área debajo de la curva (AUC)</li> </ul>	<ul style="list-style-type: none"> <li>• Curva Característica del Ítem</li> <li>• Error Cuadrático Medio (RMSE)</li> <li>• Área debajo de la curva (AUC)</li> </ul>
Observaciones		Existe un único ajuste óptimo de los parámetros (Baker, Goldstein y Heffernan, 2011)	No es utilizado a menudo en el aprendizaje en línea donde el conocimiento de los estudiantes está cambiando a medida que es evaluado (Baker, 2015).

Tabla 1. Comparativa entre los diferentes algoritmos.

En (Gong, Beck y Heffernan, 2010) se muestra que no existen diferencias significativas respecto a la exactitud de las predicciones entre los modelos BKT y PFA.

Por otro lado, como se muestra en las observaciones del cuadro comparativo el algoritmo IRT no es utilizado a menudo en el aprendizaje en línea (Baker, 2015) el cual constituye una de las tendencias actuales en el e-learning.

#### 4. Conclusiones

Esta revisión de los métodos de estimación del conocimiento latente ha cubierto los tres acercamientos principales: Bayesian Knowledge Tracing, Performance Factors Analysis e Items Response Theory Esta es un área de interés dentro de la Minería de datos educacionales. Desde cada uno de los métodos estudiados se determina la probabilidad de que un estudiante domine una habilidad o componente

del conocimiento dado, y la probabilidad de que el estudiante sea capaz de responder correctamente la próxima vez que se enfrente a un problema donde deba aplicar dicha habilidad o componente del conocimiento.

Además, cada uno de estos algoritmos necesita que se ajusten previamente aquellos parámetros necesarios, siendo los métodos más empleados para esta tarea maximizar la expectativa, fuerza bruta y minimizar la suma del error cuadrado. Por tanto, con la estimación del conocimiento latente haciendo uso de un algoritmo se dará apoyo al proceso de enseñanza – aprendizaje permitiéndole al profesor verificar si un estudiante posee una habilidad determinada o buscar otra vía de aprendizaje para que el estudiante se apropie del conocimiento. Muchos de estos algoritmos son utilizados en sistemas e-learning principalmente en STI, donde son utilizados en su mayoría en el modelo del estudiante, permitiendo predecir si un



estudiante domina una habilidad no. Otros STI combinan los algoritmos, principalmente unen IRT con BKT, donde cada cual aporta al proceso de enseñanza en su conjunto. Por ejemplo, en uno de los sistemas emplean IRT para estimar el nivel de habilidad. Las redes bayesianas se emplean para seleccionar el próximo ítem a responder.

## 5. Referencias bibliográficas

- Johnson L., S. A. Becker, Estrada V., and Freeman A., NMC Horizon Report: 2014 Higher Education. 2014.
- S. Aghabozorgi, H. Mahrooian, A. Dutt, T. Y. Wah, and T. Herawan, "An Approachable Analytical Study on Big Educational Data Mining," in *Computational Science and Its Applications--ICCSA 2014*, Springer, 2014, pp. 721–737.
- J. T. Wassan, "Discovering Big Data Modelling for Educational World," *Procedia - Soc. Behav. Sci.*, vol. 176, pp. 642–649, 2015.
- C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," *Syst. Man, Cybern. Part C Appl. Rev. IEEE Trans.*, vol. 40, no. 6, pp. 601–618, 2010.
- R. S. J. D. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *JEDM-Journal Educ. Data Min.*, vol. 1, no. 1, pp. 3–17, 2009.
- A. B. Román, D. Sánchez-Guzmán, and R. García, "Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo," *Lat. Am. J. Phys. Educ. Vol.*, vol. 7, no. 4, p. 662, 2014.
- J. A. Larusson and B. White, *Learning Analytics*. Springer, 2014.
- R. S. Baker and A. T. Corbett, "Assessment of robust learning with educational data mining," *Res. Pract. Assess.*, vol. 9, 2014.
- I. Roll, V. Alevan, B. M. McLaren, and K. R. Koedinger, "Can Help Seeking Be Tutored? Searching for the Secret Sauce of Metacognitive Tutoring.," in *AIED*, 2007, vol. 2007, pp. 203–210.
- M. Feng and N. T. Heffernan, "Towards live informing and automatic analyzing of student learning: Reporting in assistent system," *J. Interact. Learn. Res.*, vol. 18, no. 2, pp. 207–230, 2007.
- R. S. Baker and P. S. Inventado, "Educational data mining and learning analytics," in *Learning Analytics*, Springer, 2014, pp. 61–75.
- R. S. Baker, *Big Data and Education*, 2nd ed. New York, NY: Teachers College, Columbia University, 2015.
- C. M. Steiner, M. D. Kickmeier-Rust, and D. Albert, "Learning Analytics and Educational Data Mining: An Overview of Recent Techniques," *Learn. Anal. Serious Games*, p. 6, 2014.
- S. Ritter, T. K. Harris, T. Nixon, D. Dickison, R. C. Murray, and B. Towle, "Reducing the Knowledge Tracing Space.," *Int. Work. Gr. Educ. Data Min.*, 2009.
- A. T. CORBET and J. R. ANDERSON, "Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge," *User Model. User-adapt. Interact.*, vol. 4, pp. 253–278, 1995.
- P. I. Pavlik Jr, H. Cen, and K. R. Koedinger, "Performance Factors Analysis--A New Alternative to Knowledge Tracing.," *Online Submiss.*, 2009.
- J. E. Beck, K. Chang, J. Mostow, and A. Corbett, "Does help help? Introducing the Bayesian Evaluation and Assessment methodology," in *Intelligent Tutoring Systems*, 2008, pp. 383–394.
- Z. A. Pardos and N. T. Heffernan, "Modeling individualization in a bayesian networks implementation of knowledge tracing," in *User Modeling, Adaptation, and Personalization*, Springer, 2010, pp. 255–266.
- R. S. J. d Baker, A. T. Corbett, and V. Alevan, "More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing," in *Intelligent Tutoring Systems*, 2008, pp. 406–415.
- R. S. J. D. Baker, A. B. Goldstein, and N. T. Heffernan, "Detecting learning moment-by-moment," *Int. J. Artif. Intell. Educ.*, vol. 21, no. 1–2, pp. 5–25, 2011.
- Z. A. Pardos, S. M. Gowda, R. Sj. Baker, and N. T. Heffernan, "The sum is greater than the parts: ensembling models of student knowledge in educational software," *ACM SIGKDD Explor. Newsl.*, vol. 13, no. 2, pp. 37–44, 2012.
- X. Z. Clement A. Stone, *Bayesian Analysis of Item Response Theory Models Using SAS*. SAS Institute, 2015.
- Y. Gong, J. E. Beck, and N. T. Heffernan, "How to construct more accurate student models: Comparing and optimizing knowledge tracing



- and performance factor analysis,” *Int. J. Artif. Intell. Educ.*, vol. 21, no. 1–2, pp. 27–46, 2011.
- B. van De Sande, “Properties of the bayesian knowledge tracing model,” *JEDM-Journal Educ. Data Min.*, vol. 5, no. 2, pp. 1–10, 2013.
- T. Doleck, R. B. Basnet, E. Poitras, and S. Lajoie, “Towards examining learner behaviors in a medical intelligent tutoring system: A Hidden Markov Model approach,” in *Advance Computing Conference (IACC), 2015 IEEE International*, 2015, pp. 329–332.
- C. DeMars, *Item Response Theory*. Oxford University Press, USA, 2010.
- L.M.M GIRAFFA; NUNES, M. A.; VICCARI, R.M. *Multi-Ecological: an Learning Environment using Multi-Agent architecture*. Proc. MASTA'97: Coimbra: DE-Universidade de Coimbra, 1997.
- GUZMÁN, Eduardo; CONEJO, Ricardo. *Un modelo de evaluación cognitiva basado en Tests Adaptativos Informatizados para el diagnóstico en Sistemas Tutores Inteligentes*. Unpublished doctoral dissertation, Dpto. Lenguajes y Ciencias de la Computación. ETSI Informática., Universidad de Málaga, 2005.
- enfoque--CONEJO, Ricardo, et al. *Modelado del alumno: un enfoque bayesiano*. *Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial*, 2001, vol. 5, no 12.
- H. Cen, K. Koedinger, and B. Junker, “Learning factors analysis-a general method for cognitive model evaluation and improvement,” in *Intelligent tutoring systems*, 2006, vol. 4053, pp. 164–175.
- S. Hu, “Akaike information criterion,” *Cent. Res. Sci. Comput.*, vol. 93, 2007.
- X.-M. Zhao, “Bayesian Information Criterion (BIC),” *Enycl. Syst. Biol.*, p. 73, 2013.
- W. J. Hawkins, N. T. Heffernan, and R. S. J. D. Baker, “Learning bayesian knowledge tracing parameters with a knowledge heuristic and empirical probabilities,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8474 LNCS, Springer, Cham, 2014, pp. 150–155.
- Y. Gong, J. E. Beck, and N. T. Heffernan, “Comparing knowledge tracing and performance factor analysis by using multiple model fitting procedures,” in *Intelligent tutoring systems*, 2010, pp. 35–44.

**Fecha de recepción:** 20 de mayo de 2018

**Fecha de aceptación:** 9 de junio de 2018