

# **ALGORITMO PARA LA CLASIFICACIÓN DE PREGUNTAS Y EVALUACIÓN DEL NIVEL DE CONOCIMIENTO EMPLEANDO REDES BAYESIANAS Y LA TEORÍA DE RESPUESTA AL ÍTEM.**

**Ing. Reynel Fals de Pedro, [rfals@uci.cu](mailto:rfals@uci.cu), UCI, Cuba**

**Ing. Yenisleidy Piloto Lastra (P), [ypiloto@uci.cu](mailto:ypiloto@uci.cu), UCI, Cuba**

## **Resumen**

Este algoritmo surge como necesidad de clasificar las preguntas en las multimedia educativas de forma inteligente. Utiliza la Teoría de Respuesta al Ítem (Item Response Theory, IRT) para determinar la probabilidad de que sea respondida una pregunta o ítem de manera correcta según la dificultad, grado de discriminación y probabilidad de adivinación. Luego con esta probabilidad de que sea correcta la respuesta se elabora una red bayesiana que determinará la secuencia de contenidos o temas a cursar a posteriori según el nivel de habilidad esperado. Se emplearon materiales sobre IRT, redes bayesianas y la teoría clásica de respuesta al ítem.

**Palabras claves:** *Teoría de respuesta al ítem, redes bayesianas, multimedia.*

## **INTRODUCCIÓN**

Este algoritmo procede de los Test Adaptativos Informatizados (TAI) que son mucho más generales. Es el resultado de haberlo simplificado y especificado en la dirección de las multimedia educativas. Hay trabajos posteriores con solo IRT y con solo redes bayesianas aplicando la teoría clásica de evaluación. Eva Millán (Millán, y otros, 2001) ha desarrollado un trabajo muy interesante en esta rama de unir el IRT y las redes bayesianas, pero a veces usa IRT donde viola los supuestos planteados por esta teoría, por ejemplo ella incluye preguntas que tienen relación con dos o más contenidos violando el supuesto de *Unidimensionalidad*<sup>1</sup> del ítem. Esto determina que no se tenga un correcto uso de la teoría y si los supuestos no se ajustan a la hipótesis, la tesis no es válida. Por lo que podemos arribar a la conclusión de que el algoritmo tal y como lo concibe Millán tiene ciertos problemas con el IRT en los supuestos que son subsanados en este algoritmo.

---

<sup>1</sup> Este supuesto consiste en que en una prueba todos los ítems están midiendo una y sólo una característica de los examinados.

El objetivo de este trabajo es el de obtener un algoritmo eficaz y de una relativa facilidad de implementación para aplicar a los software educativos, principalmente para las multimedia creadas por nuestras casas de software para ponerla en manos de los estudiantes de todos los niveles de enseñanza. Porque no es tan sólo un algoritmo que muestra la habilidad de un examinado, también lo va guiando y le muestra la respuesta correcta en caso de que la respuesta esté incorrecta. Claro que esto puede ser opcional en dependencia de en que tipo de software será usado el algoritmo, si es un test donde no importe que se sepa las respuestas, aunque no es el principal objetivo, se puede eliminar este paso de mostrar la respuesta correcta en caso de incorrección. La principal diferencia de este algoritmo es que adapta los TAI a la idea de los tutores inteligentes.

## **1 DESARROLLO**

### **1.1 *Inteligencia Artificial***

Desde hace mucho tiempo, el hombre, se ha empeñado en crear un semejante, no solo en el aspecto externo, sino en el razonamiento y los sentimientos. El monstruo de la obra de Mary Shelley, Frankenstein es una expresión de este anhelo. El surgimiento de las computadoras abrió la puerta a lo que por mucho tiempo estuvo vedado al estrecho marco de los libros de Ficción: la creación de un ente inteligente, la inteligencia artificial. Esta rama de la ciencia pronto ganó popularidad entre los círculos científicos. Algunos de ellos dieron diferentes definiciones como:

“Intento de reproducir (modelar) la manera en que las personas identifican, estructuran y resuelven problemas difíciles” (Miller, y otros, 1982).

“La interesante tarea de lograr que las computadoras piensen [...] (Haugeland, 1985).

“El arte de crear máquinas con capacidad de realizar funciones que cuando son hechas por personas requieren de inteligencia”. (Kurzweil, 1990)

“Un campo de estudio que se enfoca a la explicación y emulación de la conducta inteligente en función de procesos computacionales”. (Schalkoff, 1990)

“Estudio de los mecanismos de la inteligencia y las tecnologías que lo sustentan”. (Newell, y otros, 1988)

“El estudio de cómo lograr que las computadoras realicen tareas que, por el momento, los humanos hacen mejor”. (McGraw Hill. Rich, y otros, 1991)

Es en esencia la manera en que se debe resolver un problema real - tedioso para el hombre- mediante un modelo computacional.

### **1.2 *Sistema Experto***

Cualquier sistema asistido por computadoras que con intervención humana ejecute automáticamente tareas de conocimiento en un dominio de aplicación engorroso y donde se empleen las destrezas, experiencias, juicios adquiridos

previamente por parte de expertos y conocimiento heurístico derivado de datos históricos acumulados por el hombre en la observación del mundo que lo rodea y no a partir de fuentes bibliográficas, se le denomina Sistema Experto (SE).

Un SE está compuesto fundamentalmente por el Motor de Inferencia (MI), cuya función es la comprobación de la hipótesis y se implementa generalmente como un Software que realiza las operaciones de búsqueda, auxiliándose de algoritmos que operan sobre la Base de Conocimiento (BC) del dominio de la aplicación. Sobre la BC se apoya el MI para la búsqueda de soluciones, por tanto, la magnitud y precisión de la información que almacena la BC sobre el dominio del problema tienen un peso decisivo en la certeza de la respuestas. (Lago Aguilar, 2004).

### **1.3 Sistemas expertos basados en probabilidades.**

Un SEBP se considera un tipo de SE cuya BC de conocimiento está conformada por elementos de un espacio probabilístico o conocimiento abstracto y el MI se basa en el Marco de Inferencia Bayesiano a través de la estrategia conocida como razonamiento probabilístico o inferencia probabilística.

En un SEBP se asume la probabilidad como una medida intuitiva de incertidumbre, en la que la distribución de *probabilidad conjunta*<sup>2</sup> (DPC) de las variables se usa para describir las relaciones de dependencia entre ellas y se sacan conclusiones con fórmulas muy conocidas de la Teoría de las Probabilidades.

Se usan rasgos y clases para la caracterización de los elementos de población de objetos cualesquiera. Entiéndase por clase cada posible solución del problema y por rasgo, cada característica que presenten las clases. La membresía de un objeto a determinada clase está dada por los rasgos que caracterizan a la clase y este puede pertenecer a una y sola una de ellas. Al seleccionar un objeto de la población se determina también a que clase pertenece y cuáles son los rasgos que caracterizan al mismo. Las frecuencias de aparición de dichos objetos para los posibles valores de los rasgos y clases a que pertenecen, es lo que conforma precisamente la BC de los SEBP. Las relaciones entre los rasgos de los objetos de la población se describen mediante su función de probabilidad conjunta y es por ello, que dicha función forma parte de lo que se llama conocimiento. (Lago Aguilar, 2004)

### **1.4 Teorema de Bayes.**

Si la probabilidad de que ocurra A es  $P(A)$ , llamamos  $P(A|B)$  a la probabilidad de que ocurran los sucesos A y B (ambos).  $P(A|B)$  a la probabilidad de que ocurra A cuando sabemos que ha ocurrido B ( se denomina probabilidad condicionada). Se

---

<sup>2</sup> Dado un n – plus  $(X_1, X_2, \dots, X_n)$  de variables aleatorias, se llama distribución de probabilidad conjunta a la función  $F[x_1, x_2, \dots, x_n] = Prob [X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n]$ . Dicha probabilidad no puede calcularse en términos de las distribuciones individuales de  $X_1, X_2, \dots, X_n$ , a menos que haya independencia.

ha determinado que la probabilidad de ocurrencia de A y B es igual al producto de multiplicar la probabilidad de B por la de A una vez ocurrido B:

$$P(A \cap B) = P(B) P(A|B) = P(A) P(B|A)$$

Supongamos que A la dividimos en tres subconjuntos (A1, A2, A3) como muestra la Figura 1:

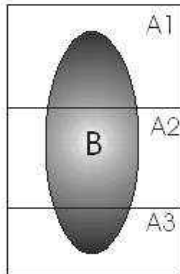


Figura 1

Estos sucesos son *mutuamente excluyentes*<sup>3</sup> y constituyen todas las posibles situaciones, nótese que  $P(A_1) + P(A_2) + P(A_3) = 1$ , porque siempre ocurrirá un suceso A. El suceso B se puede producir en cualquiera de las tres situaciones.

Si recordamos la ecuación anterior:

$$P(A_1 \cap B) = P(B) P(A_1|B) = P(A_1) P(B|A_1)$$

Luego:

$$P(A_1|B) = \frac{P(B|A_1) P(A_1)}{P(B)} \quad \text{Ecuación 1}$$

Pero:

$$P(B) = P(B|A_1) P(A_1) + P(B|A_2) P(A_2) + P(B|A_3) P(A_3) = \sum_{i=1}^3 P(B|A_i) P(A_i) \quad \text{Ecuación 2}$$

Por tanto:

$$P(A_1|B) = \frac{P(B|A_1) P(A_1)}{\sum_{i=1}^3 P(B|A_i) P(A_i)} \quad \text{Ecuación 3}$$

Lo anteriormente expuesto es en esencia el Teorema de Bayes, sólo hay que aplicarlo a cada suceso A y extenderlo a la cantidad de sucesos A que se tenga en cada caso. Aquí se ha reducido a uno particular con el fin de hacerlo más comprensible, mediante un ejemplo. (M. Molinero, 2002)

<sup>3</sup> Que no pueden ocurrir dos o más de ellos a la vez.

### **1.5 Redes Bayesianas.**

Las redes bayesianas se conocieron a partir de los años '80 y son una metodología para la construcción de sistemas expertos. Pero exhaustivamente hablando una red bayesiana es: Un grafo acíclico dirigido en el que cada nodo representa una variable y cada arco una dependencia probabilística, en la cual se especifica la probabilidad condicional de cada variable dados sus padres; la variable a la que apunta el arco es dependiente (causa-efecto) de la que está en el origen de éste. La topología o estructura de la red nos da información sobre las dependencias probabilísticas entre las variables pero también sobre las independencias condicionales de una variable (o conjunto de variables) dada otra u otras variables. Dichas independencias, simplifican la representación del conocimiento (menos parámetros) y el razonamiento (propagación de las probabilidades). El obtener una red bayesiana a partir de datos, es un proceso de aprendizaje que se divide en dos etapas: el aprendizaje estructural y el aprendizaje paramétrico (Pearl, 1988). La primera de ellas, consiste en obtener la estructura de la red bayesiana, es decir, las relaciones de dependencia de una estructura dada e independencia entre las variables involucradas. La segunda etapa, tiene como finalidad obtener las probabilidades a priori y condicionales requeridas a partir de una estructura dada.

### **1.6 Ítem Response Theory**

La primera meta de la educación y la medición psicológica es la determinación de cuanto de cada rasgo latente posee una persona. Desde que la mayoría de las investigaciones han lidiado con variables como la lectura, las matemáticas y habilidades aritméticas, el término genérico habilidad es usado dentro de la *Ítem Response Theory* (IRT) para referirse a muchos rasgos latentes. (B. Baker, 2001) Los modelos IRT se basan en la probabilidad de ser respondido un ítem o pregunta de acuerdo al nivel de habilidad de un examinado y a otros parámetros que posee el ítem en cuestión. Hay modelos de varios parámetros pero tres de ellos han sido los más estudiados y los más populares.

1. Modelo logístico de un parámetro.
2. Modelo logístico de dos parámetros.
3. Modelo logístico de tres parámetros.

Los parámetros se relacionan de acuerdo a una curva característica que se llama *curva característica de un ítem* (CCI)<sup>4</sup>. Los *niveles de habilidad* no tienen cotas ni

---

<sup>4</sup> Es una expresión matemática que relaciona la probabilidad de responder correctamente un ítem con la habilidad medida por la prueba y las características del ítem.

superiores ni inferiores<sup>5</sup>, pero siempre se establecen estas cotas para darle un sentido práctico a la teoría.

La IRT parte de varios supuestos. Los principales son:

1. *Unidimensionalidad*: en una prueba todos los ítems están midiendo una y sólo una característica de los examinados.
2. *Independencia local*: dado un nivel de habilidad, las respuestas a los ítems no pueden estar correlacionadas entre sí. Si hay correlación entre preguntas, ésta sólo se explica por habilidad.
3. Que todos los alumnos hayan tenido *experiencias educacionales similares*.
4. Que la prueba no haya sido apurada.
5. Que no haya *efectos de contexto* no controlados: algunas preguntas se comportan de modo diferente según la posición que tengan en la prueba. (Dussaillant, 2003)

### 1.6.1 El modelo logístico de tres parámetros.

Este modelo presenta la CCI siguiente:

$$P(\theta) = c_i + \frac{1 - c_i}{1 + e^{D a_i (\theta - b_i)}} \quad i=1, 2, 3 \dots n \quad \text{Ecuación 1}$$

Donde:

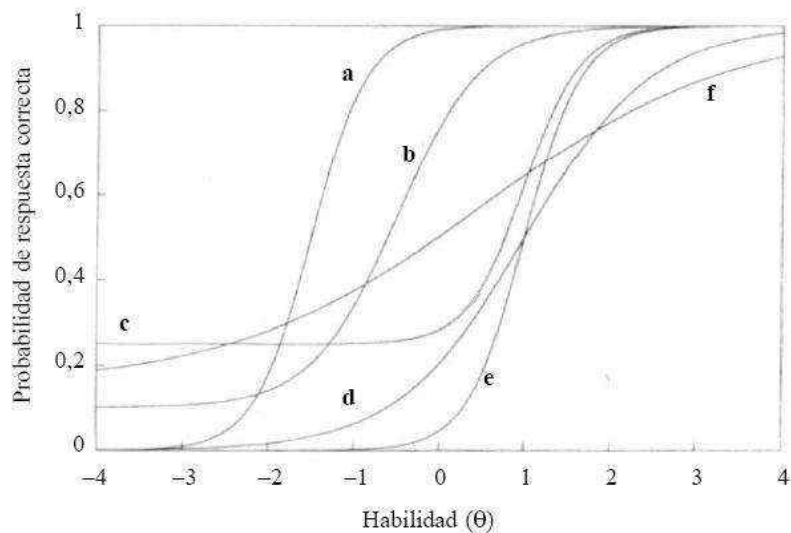
- $n$  es la cantidad de ítems a responder.
- $P(\theta)$  es la probabilidad de responder la pregunta un examinado de nivel de habilidad  $\theta$ .
- $C_i$  es la probabilidad de adivinación del ítem  $i$ . En ciertas preguntas de selección múltiple el ítem puede ser respondido al azar sin tener conocimiento cierto sobre él<sup>6</sup>.
- $b_i$  es el nivel de dificultad del ítem  $i$ .
- $a_i$  es el grado de discriminación del ítem  $i$ . La discriminación es la medida de un ítem para determinar el nivel de habilidad entre dos examinados con habilidades similares.
- $D$  “es una constante que se introduce para que la función logística se acerque lo más posible a una normal<sup>7</sup>.” (Dussaillant, 2003)

<sup>5</sup> Tiene valores que van desde  $-\infty$  hasta  $+\infty$ .

<sup>6</sup> En una pregunta donde las respuestas son  $n$  y sólo una es la correcta la probabilidad de responder correctamente es  $1/n$ .

<sup>7</sup> Normalmente siempre se usa 1.7

CURVAS CARACTERÍSTICAS DE ÍTEMS SEGÚN EL MODELO LOGÍSTICO DE TRES PARÁMETROS



**Figura 1**

Los parámetros correspondientes a las curvas de la **Figura 2** son:

- Ítem a: Discriminación ( $a$ ) = 1,8 Dificultad ( $b$ ) = -1,5 Parámetro  $c$  = 0,0
- Ítem b: Discriminación ( $a$ ) = 1,2 Dificultad ( $b$ ) = -0,5 Parámetro  $c$  = 0,1
- Ítem c: Discriminación ( $a$ ) = 1,8 Dificultad ( $b$ ) = 1,0 Parámetro  $c$  = 0,25
- Ítem d: Discriminación ( $a$ ) = 0,8 Dificultad ( $b$ ) = 1,0 Parámetro  $c$  = 0,0
- Ítem e: Discriminación ( $a$ ) = 1,8 Dificultad ( $b$ ) = 1,0 Parámetro  $c$  = 0,0
- Ítem f: Discriminación ( $a$ ) = 0,4 Dificultad ( $b$ ) = 0,5 Parámetro  $c$  = 0,15 (Dussailant, 2003).

Como se puede observar la discriminación ( $a$ ) influye en la pendiente de la curva. El nivel de dificultad ( $b$ ) está dado por el punto de inflexión de la curva. La probabilidad de adivinación ( $c$ ) permite una asíntota diferente de cero en la CCI. Por ejemplo:

1. La curva **a** tiene una mayor pendiente que la **f**, por sus diferencias con respecto al grado de discriminación.
2. La **d** tiene el punto de inflexión más a la derecha que **a** por tener diferentes niveles de dificultad.
3. La **c** pasa por encima de 0.2 en su asíntota y sin embargo **e** claramente tiene su asíntota en 0 en dependencia de la probabilidad de adivinación.

### 1.6.2 Estimación del nivel de habilidad.

Para estimar el nivel de habilidad de un examinado conociendo los parámetros del ítem se le aplica la maximización de la fórmula:

$$L(u_1, u_2, u_3, \dots, u_n | \theta) = \prod_{j=1}^n P_j^{u_j} Q_j^{1-u_j} = P_1^{u_1} Q_1^{1-u_1} \times P_2^{u_2} Q_2^{1-u_2} \times \dots \times P_n^{u_n} Q_n^{1-u_n} \quad \text{Ecuación 2}$$

Donde:

- n es la cantidad de ítems respondidos.
- $u_j$  es el valor dicotómico de ser respondido el ítem j (1 correcto, 0 incorrecto).
- $P_j$  es la probabilidad de que el ítem j sea respondido correctamente.
- $Q_j$  es la probabilidad de que el ítem j no sea respondido correctamente.

$$(Q_j = 1 - P_j)$$

Por ejemplo, si tenemos un estudiante que rindió una prueba que constaba de tres ítems, y contestó correctamente el primero ( $u_1 = 1$ ), incorrectamente el segundo ( $u_2 = 0$ ) y correctamente el tercero ( $u_3 = 1$ ), su habilidad se obtendría maximizando (Dussaillant, 2003):

$$P_1^1 Q_1^0 \times P_2^0 Q_2^1 \times P_3^1 Q_3^0 = P_1 \times Q_2 \times P_3 \quad \text{Ecuación 3}$$

Basta con encontrar los valores de  $\theta$  para los cuales la ecuación 4.6 sea máxima. Este sería el nivel de habilidad del examinado.<sup>8</sup>

### 1.7 Test Adaptativo Informatizado TAI

Es un test administrado por un ordenador donde la selección de la siguiente pregunta a plantear y la decisión de finalizar el test se realizan de una forma dinámica basándose en la estimación actual del nivel de conocimiento del alumno. (Millán, 2004)

Los elementos básicos de un TAI son:

- **Modelo de respuesta** asociado a cada pregunta: Es la forma en que se espera sea respondido el ítem.
- **Colección de ítems**: O también conocido como banco de preguntas, contiene un conjunto de ítems y sus modelos de respuestas correspondientes almacenada en una estructura determinada.
- Método de **selección de preguntas**<sup>9</sup>.

<sup>8</sup> Ver Anexo I.

<sup>9</sup> Ver 2.7.2 Método de selección de preguntas. - 82 -

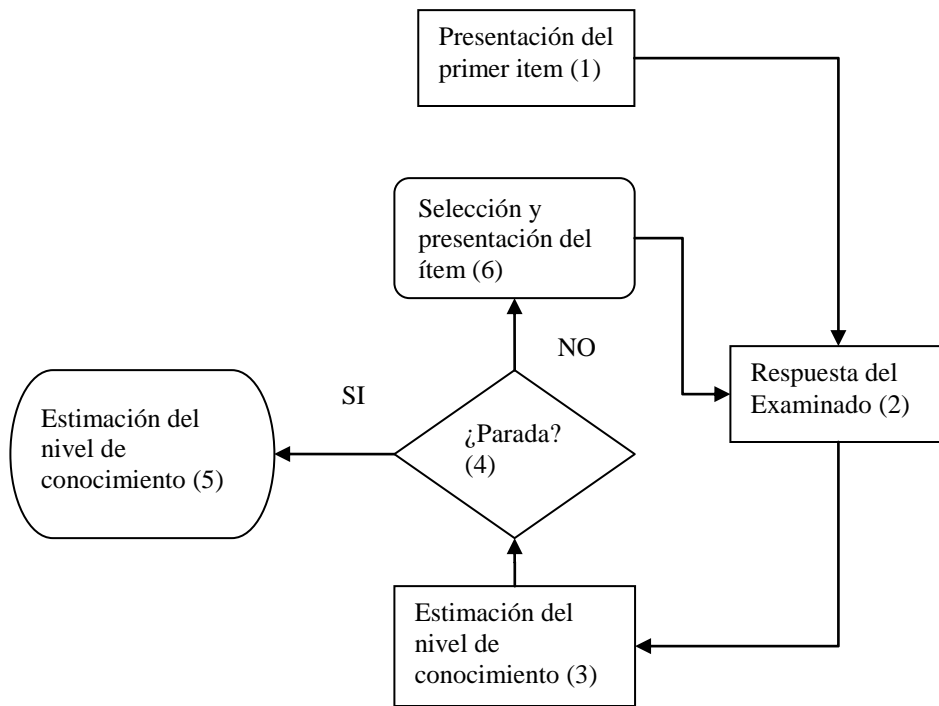


- Método de **estimación del nivel de conocimiento**: Existen dos métodos fundamentales, la *teoría clásica* y la *teoría de respuesta al ítem (IRT)* (Dussaillant, 2003).

**Criterio de finalización**: Es el criterio que se tiene en cuenta para determinar el fin en la iteración de un procedimiento o algoritmo.

### 1.7.1 El Algoritmo TAI

El Algoritmo TAI es el siguiente:



Los procesos 1 y 6 acceden a un *banco de preguntas*<sup>10</sup>, y mediante un *método de selección de preguntas* se realiza la más indicada. El proceso 1 tiene que escoger en dependencia del *nivel inicial*<sup>11</sup> que se determine. El examinado responde en el proceso 2. La respuesta es comparada con el *modelo de respuesta* que está asociado a cada pregunta y por tanto forma parte del banco de preguntas. El proceso 3 estima el nivel de conocimiento del examinado según un *método de puntuación* y verifica el *criterio de parada* en el proceso 4, en caso de que se haya cumplido el criterio se estima el nivel de conocimiento por última vez. Si todavía el criterio de parada no se cumple, se pasa a al proceso 6, repitiéndose el ciclo.

<sup>10</sup> Conjunto de preguntas o ítems.

<sup>11</sup> Es un nivel normalizado que depende del nivel de enseñanza y asignatura del test.

### **1.7.2 Método de selección de preguntas.**

Existen dos criterios principales para la selección de preguntas:

- **Aleatorio.**
- **Adaptativo.**
  - Basados en información.
    - Basados en la ganancia de información.
    - Basados en los conceptos de sensibilidad/especificidad.
  - Condicionados (Millán, 2004).

El aleatorio es seleccionando una pregunta al azar, no es muy pedagógico y puede que no lleguen a ponerse las preguntas correctas, por ejemplo si se ponen las 5 preguntas más difíciles esto arroja que el estudiante no sabe nada porque todas las respuestas dadas son erróneas respecto al modelo de respuesta, sin embargo puede ser que el examinado tenga un nivel un poco por debajo al que determinan estos 5 ítems y no se puede determinar el nivel de conocimiento del mismo.

El adaptativo presenta una gran ventaja, para una mejor comprensión del nivel de un examinado, ya que se orienta en ese sentido variando el nivel de dificultad de las preguntas.

### **1.8 Un algoritmo para enseñar.**

Partiendo de las ideas antes señaladas, se presenta la necesidad de construir un algoritmo que permita enseñar y evaluar a los estudiantes.

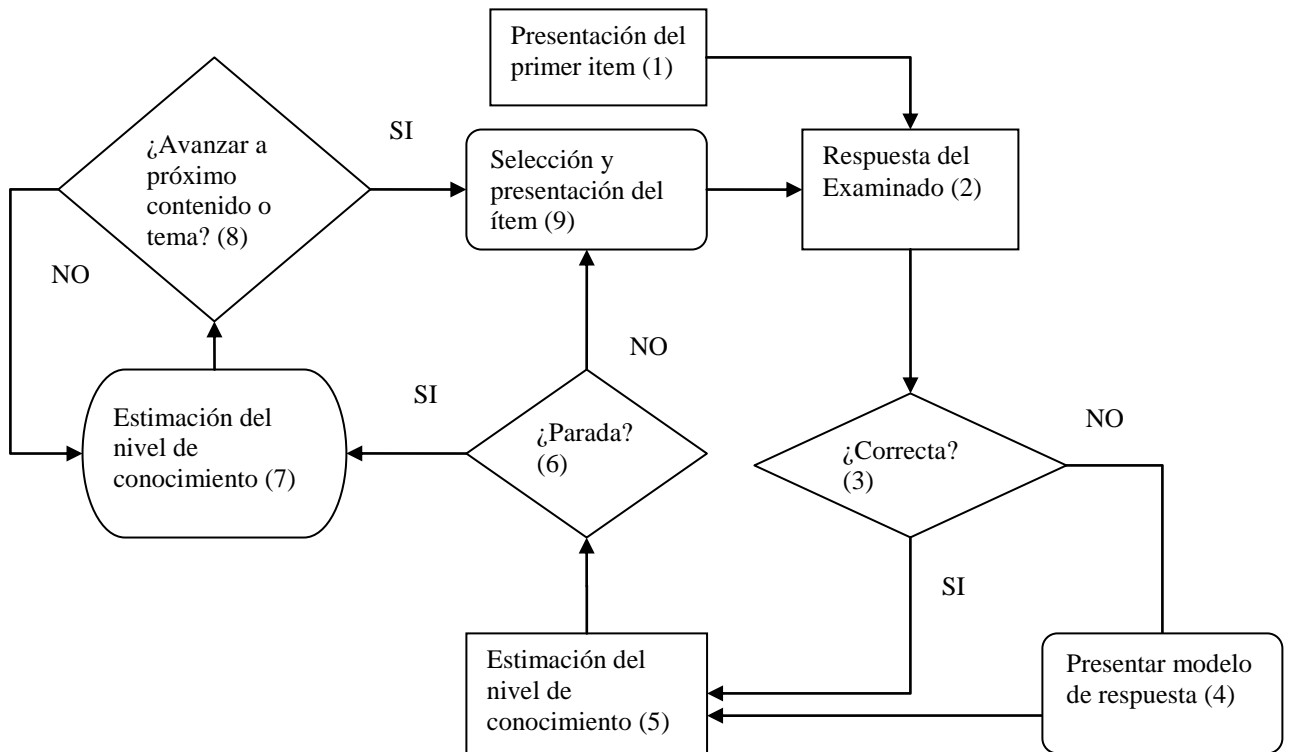
Proponemos dividir las asignaturas en temas (**T**) y estos a su vez en contenidos (**C**), en cada contenido habrá preguntas (**P**) para responder, aunque sólo se responderán las que se estime conveniente para saber si el estudiante domina o no un contenido. En esencia esto se conoce como *criterio de parada*<sup>12</sup>.

Para este reto se emplean IRT para estimar el nivel de habilidad. Las redes bayesianas se emplean para seleccionar el próximo ítem a responder, esta predicción se hace más difícil en el caso de que ya el criterio de parada de un contenido nos da la señal de que debemos abandonar ese contenido porque se domina o no y pasar al próximo contenido o no. En el caso de que se siga hasta un nuevo contenido entonces se debe ir a un ítem correspondiente a este nuevo contenido. La forma de seleccionar este ítem es el desafío de la red porque el nivel de habilidad del conocimiento anterior puede influir en cierta medida sobre el contenido nuevo. Se calcula cuanto debe saber del nuevo contenido el estudiante y se ajusta la red a ese nivel de habilidad, seleccionándose el ítem adecuado.

---

<sup>12</sup> Ver 2.7 Test Adaptativo Informatizado TAI

Representado en un diagrama de flujo el algoritmo es el siguiente:



Nótese que es muy parecido al algoritmo TAI<sup>13</sup> de hecho es una extensión del mismo adaptándolo a las condiciones de un tutor inteligente.

Al inicio se escoge en el proceso 1 un ítem de forma normalizada, de un banco de preguntas<sup>14</sup> según la capacidad del estudiante medio de ese nivel de enseñanza. El examinado responde y si no es correcta según el modelo de respuesta del ítem en cuestión, se presenta en el proceso 4 este modelo al estudiante, pero de una forma que le sea comprensible al estudiante<sup>15</sup>. De esta manera garantizamos que el estudiante se nutra del sistema y esto puede influir en su nivel de habilidad. Se pasa al proceso 5 donde la IRT nos brinda un nuevo nivel de conocimiento en dependencia de la habilidad anterior, si la habilidad anterior dio como resultado que el estudiante fallara, se disminuye el nivel de habilidad de forma tal que la

<sup>13</sup> Ver 2.7 Test Adaptativo Informatizado TAI.

<sup>14</sup> Idem.

<sup>15</sup> Quizás el modelo sean una serie de números que sólo tiene sentido para el programa.

probabilidad de responder correctamente ( $P_{\theta}$ ) sea lo más próxima posible a 0.5, es decir que la desviación típica mínima respecto a la media de 0.5. La red bayesiana se ajusta a la selección de este ítem y se compara el criterio de parada<sup>16</sup> y se determina si se cumplieron los objetivos del contenido o si se saben que no se van a cumplir ya. En cualquiera de los dos casos se pasa al proceso 7 que es el último del contenido o tema (contenido y tema no es lo mismo, se usa la palabra o porque puede ser el fin de un contenido tanto como de un tema, siempre el fin de un tema es un último contenido y el fin de un contenido es una última pregunta con la que se estima el nivel de conocimiento teniendo en cuenta todas las antes respondidas) donde se estima el nivel de conocimiento del contenido por la Ecuación 6. Se considera si se va a avanzar hasta el próximo contenido o tema en caso de que los haya, en ese caso se escoge a que contenido se pasará, el más probable de la red bayesiana, la red se conformó buscando como contenido más probable al que más le aportara el tema antes cursado, hace una evaluación con la probabilidad de que con los conocimientos que se saben tiene el estudiante pueda dominar el nuevo contenido y el nivel de habilidad demostrado en los contenidos anteriores, calculando que ítem del nuevo contenido comenzar<sup>17</sup>, esto lo hace en el proceso 9 que está como “sobrecargado” para un contenido y para entre contenidos actuar de una manera similar con más datos y más requerimientos. En caso de que no se avance hasta el próximo contenido o tema se estima el nivel de conocimiento completo de los contenidos o temas dando una nota del tema o de la asignatura según sea el caso. Esta forma de evaluación de nivel de habilidad del tema o la asignatura está actualmente en estudio para darle el mayor carácter científico posible. Algo que no se debe olvidar nunca es escoger bien los ítems que integrarán el banco de preguntas así como el modelo de respuesta de cada ítem. Para eso se recomienda estudiar a fondo los supuestos del ítem.

La *evaluación real* es un tema que aún no nos ocupa, pero puede verse un párrafo alusivo a esta en el Anexo I.

## 2 VALORACIÓN ECONÓMICA Y APOORTE SOCIAL

Este algoritmo puede usarse en las multimedias educativas para todos los niveles. Es un algoritmo de fácil comprensión y puede darle al estudiante una fuerte motivación ya que pone a su altura las preguntas que puede responder, no lo hace sentirse aturdido por preguntas que quizás le sean muy fáciles o muy difíciles. Puede emplearse para mejorar la educación a distancia. Puede ser empleado ampliamente en la industria del software para la exportación o para el programa educativo cubano, así como en la creación de CDs educativos para incentivar la cultura general integral del pueblo cubano, dando un aporte aunque modesto

---

<sup>16</sup> Ver 2.7 *Test Adaptativo Informatizado TAI*.

<sup>17</sup> La forma en que se calculará el próximo ítem esta en estudio todavía.

significativo a la idea de hacer del pueblo cubano uno de los pueblos mas cultos e instruidos del mundo.

Como no se tiene un algoritmo definitivo para la industria del software educativo este se podrá extender a cada uno de los que se haga en Cuba en incluso automatizar en un futuro la implantación de este algoritmo en los software, como por ejemplo un generador que dado los ítems con sus parámetros divididos en contenidos y temas haga todo el programa de selección y calificación de los ítems.

### 3 CONCLUSIONES Y RECOMENDACIONES

Se ha ajustado un algoritmo (TAI) para las necesidades específicas del software educativo y todavía se está trabajando para mejorar las formas de selección y calificación por medio del IRT y las redes bayesianas. Recomendamos usar este algoritmo cuando esté del todo especificado y probado en gran escala en la Isla para que pueda ser fácil la creación del software educativo. Este algoritmo permite hacer una comparación inequívoca de todos los estudiantes en un nivel de enseñanza y pueden convertirse en un criterio de selección para el paso del estudiante, al siguiente nivel de enseñanza.

### 4 BIBLIOGRAFÍA

**B. Baker, Frank. 2001.** *The basics of item response theory. Second edition.*

United States of America : ERIC Clearinghouse on Assessment and Evaluation, 2001.

**Dussailant, Francisca. 2003.** *Técnicas de medición en pruebas de admisión a las universidades.* ,Chile : Estudios Públicos., 2003.

**Haugeland, John. 1985.** *Artificial Intelligence: The Very Idea.* Cambridge : s.n., 1985.

**Kurzweil, Ray. 1990.** *The Age of Intelligent Machines.* s.l. : Association of American Publishers, 1990.

**Lago Aguilar, Eduardo. 2004.** *Sistema experto en inferencia probabilística.* Santa Clara : Universidad Central "Marta Abreu" de las Villas, 2004.

**M. Molinero, Luis. 2002.** El método bayesiano en la investigación médica. [En línea] 2002. [Citado el: ] <http://www.seh-lelha.org/bayes1.htm> .

**McGraw Hill, Rich E y Knight, K. 1991.** *Inteligencia Artificial: Una nueva síntesis.* 1991.

**Millán, Eva. 2004.** *Diseño y evaluación de test adaptativos basados en redes bayesianas.* Málaga, España : Departamento de Lenguajes y Ciencias de la Computación, Facultad de Informática, 2004.

**Millán, Eva, y otros. 2001.** *Uso de redes bayesianas en test adaptativos computarizados.* Málaga, España. : Departamento de Lenguajes y Ciencias de la Computación, Facultad de Informática, 2001.

- Miller, R. A. y Pople, H. E. 1982.** *INTERNIST-i, an experimental computer-based diagnosis consultant for general internal medicine.* New England : Journal of Medicine, 1982.
- Newell y A.Simon, H. 1988.** *GPS: A program that simulates human thought.* San Mateo : A. Collins and E. E. Smith, 1988.
- Pearl, J. 1988.** *Probabilistic reasoning in intelligent systems: networks of plausible.* San Mateo, California : s.n., 1988.
- Schalkoff, Robert J. 1990.** *Artificial Intelligence: An Engineering Approach.* 1990.

## 5 ANEXOS

### 5.1 ANEXO I

“La escala  $\theta$  original puede ser alterada mediante transformaciones lineales del parámetro que no alteran el cálculo de las probabilidades de responder correctamente los ítems (si es que los parámetros de los ítems son transformados también a su vez mediante el mismo proceso). Otra transformación de la escala de  $\theta$ , y quizás una de las más importantes y utilizadas, es la transformación a puntaje verdadero ( $\tau$ ). En pocas palabras, el puntaje verdadero de un examinado con habilidad  $\theta$  corresponde a la suma de las probabilidades de responder correctamente los ítems, evaluada en el valor de  $\theta$  correspondiente. Si graficamos los puntajes verdaderos ( $\tau$ ) para todo el rango de posibles  $\theta$ , lo que estamos haciendo es graficar la suma de las CCI utilizadas en las pruebas. La curva resultante se llama curva característica de la prueba (CCP). Mediante esta curva se mapea la relación existente entre la habilidad  $\theta$  del examinado y su puntaje verdadero  $\tau$ . El puntaje verdadero ( $\tau$ ) puede ser considerado como una relación no lineal de  $\theta$ . Ya que  $P_j(\theta)$  está entre 0 y 1, entonces el rango para el puntaje verdadero está entre 0 y  $n$  (con  $n$  = número total de ítems que conforman la prueba). La transformación de  $\theta$  a puntaje verdadero ( $\tau$ ) tiene implicancias importantes. Primero que nada, se eliminan los puntajes negativos. La nueva escala, además, se mueve en el rango de 0 a  $n$ , que es más fácil de interpretar, aunque el puntaje  $\tau$  obtenido por este método no corresponde a número de respuestas correctas, con lo que su interpretación se puede prestar a malos entendidos. La última implicancia de la transformación de  $\theta$  a  $\tau$  es que el  $\tau$  de un examinado cuya habilidad  $\theta$  es conocida puede ser computado para un set de ítems que nunca le han sido administrados.”

Tomado textualmente de (Dussailant, 2003)