

Un Modelo Cognitivo para la Determinación de Habilidad en la Validación de Proposiciones Lógicas Aplicado a CAT's

Emma Lucio, Joel Suárez, and Félix Castro

Centro de Investigación en Tecnologías de Información y Sistemas,
Universidad Autónoma del Estado de Hidalgo,
Pachuca de Soto, Hidalgo, México
{elm_2303@hotmail.com, jsuarez@uaeh.edu.mx, fcastro@lsi.upc.edu}

Resumen En este trabajo se propone un modelo cognitivo para determinar en forma automática las complejidades de reactivos enfocados a medir habilidades en la validación de proposiciones lógicas con cierta estructura. El análisis parcial del modelo cognitivo propuesto se realiza en forma estadística aplicando la prueba de Mantel para verificar la similaridad entre la métrica de co-ocurrencias en el modelo cognitivo y la métrica de los resultados experimentales. Los resultados arrojados al utilizar el modelo cognitivo propuesto indican la estrecha relación entre éste y las respuestas de los estudiantes que participaron en el experimento.

Key words: Proposiciones Lógicas; Evaluación Adaptable Computarizada; Prueba de Mantel

1. Introducción

La evaluación adaptable computarizada (CAT por sus siglas en inglés) ha mostrado ser un instrumento sumamente valioso para la determinación de habilidades en diferentes áreas del conocimiento. Prueba de ello es que actualmente se aplica en diferentes entornos educativos para estimar habilidades en idiomas, en el manejo de operaciones aritméticas y en procesos de razonamiento lógico. En términos generales, un sistema de evaluación adaptable computarizado consiste esencialmente de un depósito de reactivos sobre algún tema en específico y un administrador de reactivos. Se asume que los reactivos en el depósito han sido calibrados previamente y que por ello, para cada reactivo en el depósito, se conoce de antemano una serie de parámetros que están relacionados con la complejidad y la capacidad discriminatoria de éste, entre otros posibles parámetros cuya presencia depende del tipo de modelo psicométrico que se emplee para determinar el nivel de habilidad del examinando. Por su uso frecuente, la tasa de exposición de un reactivo puede llegar a ser alta, por lo que se requiere de una política de sustitución del mismo que haga que en todo momento se cuente con un depósito de reactivos confiable. Sin embargo, lograr este objetivo resulta costoso si se emplean técnicas tradicionales de construcción de reactivos para

evaluación adaptable computarizada. Por esta razón, algunos investigadores en el campo han propuesto alternativas novedosas de construcción de reactivos para evaluación adaptable computarizada [1,9,4,10]. Estas técnicas conservan algunas de las características de las técnicas usuales, pero incorporan el concepto de generación automática del reactivo.

2. Aspectos Teóricos

La generación automática de reactivos determina la complejidad de un reactivo basándose en la existencia de un modelo cognitivo que especifica algunos de los procesos mentales que intervienen en la solución de un problema específico. En este trabajo se aborda el tema de la validación de proposiciones lógicas que contienen a lo más dos proposiciones simples conectadas con operadores de negación, conjunción, disyunción, implicación, reducción y equivalencia. Podemos definir este tipo de problemas a través de una gramática libre de contexto relativamente simple como la siguiente

$$\begin{aligned}
\langle \text{proposition} \rangle &::= \langle \text{negation} \rangle \{ \langle \text{operator} \rangle \langle \text{negation} \rangle \}_0^2 \\
\langle \text{negation} \rangle &::= \langle \text{simpleprop} \rangle | \sim \langle \text{simpleprop} \rangle \\
\langle \text{operator} \rangle &::= \wedge | \vee | \rightarrow | \leftarrow | \leftrightarrow \\
\langle \text{simpleprop} \rangle &::= p
\end{aligned} \tag{1}$$

donde se define el símbolo de negación como \sim y p es cualquier proposición simple. Por ejemplo, la proposición $p_1 \rightarrow p_2 \wedge \sim p_3$ es válida en el lenguaje definido por estas reglas de producción, como se ilustra en la siguiente derivación

$$\begin{aligned}
\langle \text{proposition} \rangle &\Rightarrow \langle \text{negation} \rangle \langle \text{operator} \rangle \langle \text{negation} \rangle \langle \text{operator} \rangle \langle \text{negation} \rangle \\
&\Rightarrow \langle \text{simpleprop} \rangle \langle \text{operator} \rangle \langle \text{negation} \rangle \langle \text{operator} \rangle \langle \text{negation} \rangle \\
&\Rightarrow p_1 \rightarrow \langle \text{simpleprop} \rangle \langle \text{operator} \rangle \langle \text{negation} \rangle \\
&\Rightarrow p_1 \rightarrow p_2 \wedge \sim \langle \text{simpleprop} \rangle \\
&\Rightarrow p_1 \rightarrow p_2 \wedge \sim p_3
\end{aligned}$$

2.1. Modelo Cognitivo

El modelo cognitivo considera diferentes aspectos relacionados con la forma en que el examinando resuelve el problema de validar una proposición lógica con las propiedades especificadas anteriormente. Atiende aspectos aparentemente simples como el uso del operador de negación y la relación que éste guarda cuando aparece en una proposición en combinación con los otros cinco operadores binarios ya introducidos por la gramática (1).

En particular, especifica la complejidad de un reactivo proposicional en términos de diferentes características que incluyen la presencia o no de cualquiera de los operadores, la jerarquía que éstos tienen, el número de operadores en el reactivo y las posibles interpretaciones que el examinando asigna a las proposiciones

simples en la proposición compuesta. Con esta información se define un vector de características de dieciocho componentes. Para ilustrar de manera concreta lo anterior, consideremos el siguiente ejemplo: las proposiciones p, q, r son respectivamente dadas por los enunciados 'El metalenguaje es un lenguaje que explica otro lenguaje', 'El elemento neón en su estado natural es un gas' y 'Gran Bretaña es miembro de la Comunidad Europea'. Interpreta la proposición $p \leftarrow q \wedge \sim r$. Aquí el vector de características es (1 1 0 0 1 0 1 1 1 0 1 0 0 1 0 1 1 1), el cual se construye siguiendo el diagrama mostrado por Figura 1 y considerando los contenidos temáticos presentes en las proposiciones, así como la precedencia de los operadores. El modelo cognitivo incluye un paso previo que resuelve la negación explícita, incluyéndola implícitamente en la proposición. Por ejemplo \sim 'Gran Bretaña es miembro de la Comunidad Europea' se transforma en 'Gran Bretaña **no** es miembro de la Comunidad Europea'.

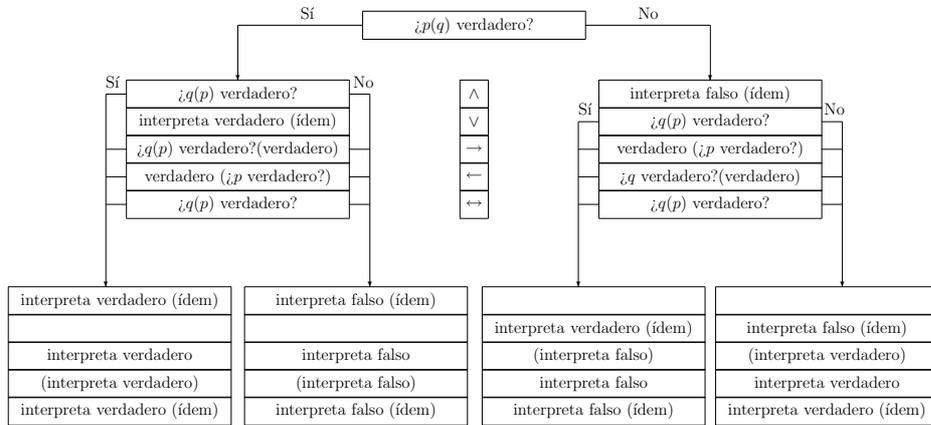


Figura 1. Modelo cognitivo para validación de proposiciones lógicas generadas por gramática (1), después de integrar negaciones en proposiciones.

2.2. Distancia entre Reactivos y Similitud entre Métricas

Dada una evaluación que contiene reactivos proposicionales como los señalados anteriormente, se construye una matriz W cuyas filas denotan a cada uno de los reactivos en la evaluación y cuyas columnas señalan la presencia o ausencia de las componentes del vector de características, cuyos valores son proporcionados de acuerdo al modelo cognitivo propuesto. Estos reactivos son agrupados siguiendo una métrica de co-ocurrencias que es simplemente dada por el producto matricial $C = WW^T$, donde el superíndice T señala la transpuesta de la matriz. Sin embargo, esta agrupación debe ser comparada con una agrupación empírica dada por los evaluandos y que se define a través de una métrica de co-ocurrencias que toma en cuenta las opiniones que ellos vierten acerca de la

presencia o ausencia de algunas de las características en el correspondiente reactivo. Si D_i es una matriz cuyas filas representan a cada uno de los examinandos y cuyas columnas se refieren a cada una de las características que definen la complejidad de un reactivo i , con valores dados por la opinión propia del evaluando, entonces la métrica de co-ocurrencias

$$D_{ij} = \sum_{k,l} (D_i D_j')_{k,l} \quad (2)$$

especifica la distancia que guarda el reactivo i en relación con el reactivo j . Esta es la matriz de proximidad de la muestra y el interés radica en saber qué tan similar es esta métrica empírica con la métrica de co-ocurrencias definida para el modelo cognitivo. Para determinar la similaridad entre ambas métricas se emplea la estadística o prueba de Mantel además de determinar el coeficiente de correlación de Pearson r entre las componentes de las métricas de co-ocurrencias. Por la simetría de ambas matrices, esta correlación toma en cuenta solamente a los elementos en la parte triangular superior y es dada por la expresión siguiente

$$r = \frac{1}{d-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left(\frac{C_{ij} - \tilde{C}}{\sigma_C} \right) \left(\frac{D_{ij} - \tilde{D}}{\sigma_D} \right) \quad (3)$$

donde n es el número de reactivos considerados, d es el número de proximidades en la parte triangular superior de cada una de las matrices C y D , σ_C y σ_D son las desviaciones estándar de estos dos conjuntos de proximidades, respectivamente.

3. Evaluación Experimental

En el experimento participaron 77 estudiantes (37 mujeres y 40 hombres entre los 19 y 21 años) de la Universidad Tecnológica de Xicotepec de Juárez (UTXJ) pertenecientes al tercer cuatrimestre de la Carrera de Técnico Superior Universitario en Tecnologías de la Información y Comunicación. El instrumento de evaluación consistió de un examen con 30 reactivos, que fueron generados en forma automática empleando la gramática libre de contexto dada por Ecuación (1). Algunos de estos reactivos son mostrados en el Cuadro 1. Los examinandos interpretan la proposición compuesta (asignan un valor de verdad a la proposición compuesta) basándose en el conocimiento que tienen en lógica proposicional. Como complemento al examen, se aplicó un cuestionario para recabar información acerca de la opinión del examinando sobre la presencia o ausencia de las características que definen la complejidad de cada reactivo. Estas características son las mismas que las empleadas en el modelo cognitivo.

3.1. Análisis y Discusión de Resultados

Se hace hincapié en la demostración parcial de que el modelo cognitivo propuesto es útil para especificar, con base en componentes, las complejidades de reactivos con los que se miden las habilidades interpretativas de proposiciones

Cuadro 1. Primer grupo de cuatro reactivos aplicados a muestra de 77 evaluandos para interpretar cada una de las proposiciones compuestas conociendo las proposiciones simples que las forman. El examen consiste de tres grupos de diez reactivos, completando un total de 30 reactivos.

Reactivos			
No.	Proposición simple	Símbolo	Proposición compuesta
1.	El planeta de nuestro sistema solar más cercano al sol es mercurio. La primera ley de Newton habla sobre la acción y la reacción.	p q	$p \vee \neg q$
2.	Windows es el nombre de un sistema operativo. La antartida se encuentra en el polo norte.	p q	$p \leftrightarrow q$
3.	La conjunción acepta solamente un operando. La obra 'El Quijote' fue escrita por Miguel de Cervantes Saavedra.	p q	$\neg p \wedge \neg q$
4.	La suma de N unos es igual a N + 1.	p	$\neg p$

lógicas de la forma $p \square q \square r$. Para demostrar que la prueba es válida es necesario que se determine que existe una correlación entre las complejidades predichas por el modelo y aquéllas obtenidas de manera usual; es decir, calibrando los reactivos como tradicionalmente se hace en evaluación adaptable computarizada [14]. Para un mejor entendimiento y validación de los resultados obtenidos, en esta sección se muestra una discusión en el nivel de similaridad y las pruebas estadísticas de normalidad de la prueba de Mantel.

3.2. Similaridad

Los resultados de la encuesta para cada uno de los 30 reactivos se agrupan en igual número de matrices D_i , mientras que las características predichas por el modelo cognitivo se agrupan en una sola matriz W . Cabe aclarar que la similaridad entre las métricas es dada por Ecuación (3) y se complementa con pruebas estadísticas de normalidad de la muestra y nivel de significancia calculado mediante prueba estadística de Mantel, además de la determinación del valor del coeficiente de correlación a través de estadística bootstrap.

El coeficiente de correlación r resultante de aplicar la Ecuación (3) a las matrices de co-ocurrencias C y D tiene un valor 0.58, mientras que el coeficiente de correlación entre las medias de las co-ocurrencias de la muestra y la proximidad dada por las co-ocurrencias de W tiene un valor 0.89, lo cual es ilustrado en la Figura 2. En ambos casos se indica una correlación aceptable entre los valores proporcionados por ambas métricas, por lo que se puede afirmar que ambas son similares. La aplicación de la prueba de Mantel en la verificación de la similitud entre las matrices C y D arroja un p -valor de 0,01 % para el caso en que se realizan numéricamente 10,000 permutaciones diferentes y de 0,05 % para el caso en que el número de permutaciones diferentes es igual a 20,000. Debe recordarse

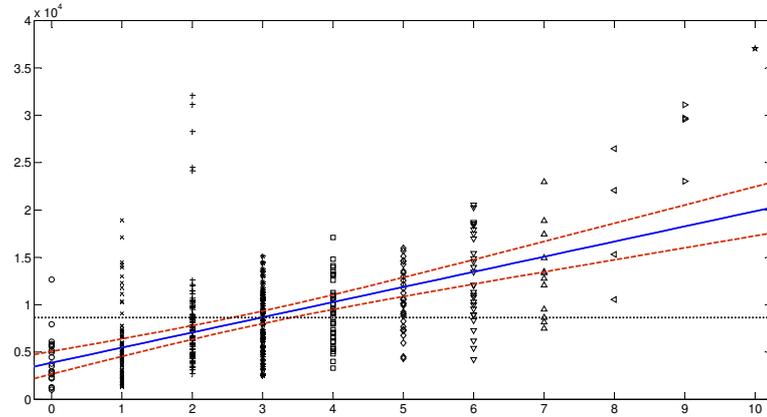


Figura 2. Correlación de proximidades entre reactivos según la muestra (eje vertical) versus proximidades entre reactivos según el modelo cognitivo (eje horizontal). Cada valor de proximidad entre reactivos según el modelo cognitivo se encuentra en una o más entradas de la matriz de proximidad C . Para este valor y sus entradas en C , se buscan los valores de proximidad en las mismas entradas de la matriz de proximidad D , y éstos se grafican sobre el eje de ordenadas.

que se tienen 77 examinandos con 30 reactivos cada uno, lo que arroja una matriz de co-ocurrencias C de orden 30, llevando a un número total de permutaciones mayor que 265×10^{30} . Este número es sumamente grande como para que quede plenamente justificado el cálculo numérico de solamente un subconjunto del total de permutaciones posibles.

3.3. Pruebas Estadísticas de Normalidad

La variable de interés es la distancia entre reactivos y sus valores empíricos se obtienen a través de la métrica de co-ocurrencias empleada para el estudio de campo (Ecuación (2)) o bien en forma directa a través del producto matricial $D_i D_j^T$. Los estadísticos, en este trabajo, se refieren solamente a cualquiera de estos casos. En la primer forma, para cada valor de distancia entre reactivos i y j según el modelo cognitivo, C_{ij} , que se grafica sobre el eje horizontal (eje de las abscisas), corresponde más de un valor posible de distancia entre el mismo par de reactivos, D_{ij} , obtenido de la muestra, el cual se grafica sobre el eje vertical (eje de las ordenadas). La Figura 2 muestra la distribución de los datos experimentales para cada valor entero en el intervalo $[0, 10]$. Esto es así debido a que estos extremos definen los valores mínimo y máximo de co-ocurrencias reportados por la selección automática de la batería de reactivos. Utilizando esta información, la normalidad de los datos experimentales se verifica por la co-ocurrencia predicha por el model cognitivo. La Figura 3 ilustra la normalidad de los datos experimentales para cada distancia predicha por el modelo cognitivo.

4. Conclusiones

Los resultados experimentales obtenidos sugieren que en el proceso de responder un reactivo algunos examinandos consideran superfluas algunas características; es decir, no las consideran necesarias durante el proceso de validación de una proposición lógica, o bien no coinciden en sus apreciaciones acerca de las características necesarias con el objeto de obtener el valor de verdad de una proposición, o bien muchos de ellos consideran que estas características resultan redundantes y que probablemente ya se encuentran consideradas en algunas de las características. Más concretamente, el hecho es que las co-ocurrencias que asignan los examinandos a cada par de reactivos no son más allá de diez, esto significa que globalmente ellos opinan que se requiere fijar la atención a lo más en diez características para poder asignar el valor de verdad a una proposición.

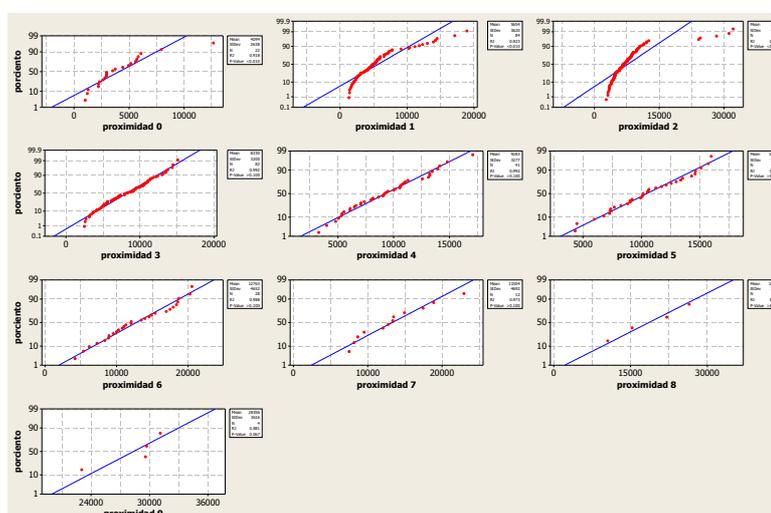


Figura 3. Evidencia gráfica de normalidad de co-ocurrencias determinadas por sujetos del muestreo de acuerdo a Ecuación 2 sin la eliminación de outliers y por cada valor de proximidad predicha por el modelo cognitivo.

Con la idea de validar el modelo cognitivo en ambientes educativos diferentes al presentado en este artículo, se está realizando un experimento en la Universidad Autónoma del Estado de Hidalgo (UAEH) en el que los evaluandos no son sólo estudiantes sino también profesores expertos en el área. No es el objetivo del presente trabajo el determinar qué características son éstas, sin embargo un trabajo futuro es utilizar herramientas de machine learning para reducir la dimensionalidad mediante la identificación de las variables más relevantes. Investigaciones pasadas de los autores en contextos de evaluación en ambientes educativos, han demostrado la viabilidad de utilizar métodos de determinación

de la importancia de variables mediante t-GTM [3] y CR-FIR [2], por lo tanto un primer paso es realizar un análisis utilizando tales algoritmos. Por otro lado, el utilizar t-GTM permite neutralizar el impacto negativo de los outliers, generando resultados más realistas sin la necesidad de eliminarlos durante la etapa de modelado del sistema.

Agradecimientos. Los autores desean agradecer profundamente las facilidades y soporte proporcionados para la realización de este proyecto por parte del Consejo de Ciencia y Tecnología del Estado de Hidalgo (Proyecto FOMIX 2008/97071), la UAEH y la UTXJ.

Referencias

1. Bejar, I.I.: A generative analysis of a three dimensional spatial task. *Psychological Measurement*. (1990)
2. Nebot, A., Mugica, F., Castro, F.: Causal relevance to improve the prediction accuracy of dynamical systems using inductive reasoning. *International Journal of General Systems*. 38(3), 331–358 (2009)
3. Castro, F., Vellido, A., Nebot, A., Minguiñón, J.: Finding relevant features to characterize student behavior on an e-learning system. In: *International Conference on Frontiers in Education: Computer Science and Computer Engineering (FECS 2005)*, Las Vegas, Nevada, USA, 210–216 (2005)
4. Revuelta, J., Ponsoda, V.: Un Test Adaptativo Informatizado de Análisis Lógico Basado en la Generación Automática de Ítems. *Psicothema*. 10, 709–716 (1988)
5. Freeman, L.C., Freeman, S.C., Romney, A.K.: The implications of social structure for dominance hierarchies in red deer, *Cervus elaphus*. *Animal Behavior*. 44, 239–245 (1992)
6. Harding, M.R., Sokal, R.M.: Classification of the European language families by genetic distance. *Proc. Natl. Acad. Sci. USA Population Biology*. 85, 9370–9372 (1988)
7. Dow, M.M., Cheverud, J.M., Jonathan, S.F.: Partial Correlation of Distance Matrices in Studies of Population Structure. *American Journal of Physical Anthropology. Proc. Natl. Acad. Sci. USA Population Biology*. 72, 343–352 (1987)
8. Jáuregui, A.L., Oliden, P.E.: Formulación y validación de un modelo logístico lineal para la tarea de adición y sustracción de fracciones y números mixtos. *Psicothema*. 14(4), 802–809 (2002)
9. Embretson, S.E.: *Applications of Cognitive Design Systems to Test Development*. Ed Plenum Press, New York (1994)
10. Tapia, J., Olea, J.: *Modelos de evaluación de los conocimientos matemáticos*. Madrid: CIDE, Madrid (1997)
11. Foster, I., Kesselman, C.: *The Grid: Blueprint for a New Computing Infrastructure*. Morgan Kaufmann, San Francisco (1999)
12. Van der Linden, W.: *Handbook of modern item response theory*. Springer-Verlag, New York (1997)
13. Fischer, G.H., Laming, D.: *Contributions to Mathematical Psychology, Psychometrics, and Methodology*. Springer-Verlag, New York (1993)
14. Muñoz, J.: *Introducción a la Teoría de Respuesta a los Ítems*. Pirámide, Madrid (1997)
15. Thode, H.C.: *Testing for Normality*. Marcel Dekker, New York (2002)