

Algoritmo mixto mínima entropía-máxima información para la selección de ítems en un test adaptativo informatizado

David Aguado García, Carlos Santa Cruz, José R. Dorronsoro y Víctor J. Rubio Franco
Universidad Autónoma de Madrid

El objetivo del estudio que presentamos es comparar la eficacia como estrategia de selección de ítems de tres algoritmos diferentes: a) basado en máxima información; b) basado en mínima entropía; y c) mixto mínima entropía en los ítems iniciales y máxima información en el resto; bajo la hipótesis de que el algoritmo mixto, puede dotar al TAI de mayor eficacia. Las simulaciones de procesos TAI se realizaron sobre un banco de 28 ítems de respuesta graduada calibrado según el modelo de Samejima, tomando como respuesta al TAI la respuesta original de los sujetos que fueron utilizados para la calibración. Los resultados iniciales muestran cómo el criterio mixto es más eficaz que cualquiera de los otros dos tomados independientemente. Dicha eficacia se maximiza cuando el algoritmo de mínima entropía se restringe a la selección de los primeros ítems del TAI, ya que con las respuestas a estos primeros ítems la estimación de θ comienza a ser relevante y el algoritmo de máxima información se optimiza.

Item selection algorithms in computerized adaptive testing. The aim of this paper is to compare the efficacy of three different item selection algorithms in computerized adaptive testing (CAT). These algorithms are based as follows: the first one is based on Item Information, the second one on Entropy, and the last algorithm is a mixture of the two previous ones. The CAT process was simulated using an emotional adjustment item bank. This item bank contains 28 graded items in six categories, calibrated using Samejima (1969) Graded Response Model. The initial results show that the mixed criterium algorithm performs better than the other ones.

La mayor ventaja de un TAI consiste en la mayor eficiencia en la estimación del nivel de rasgo o habilidad del evaluado con un menor número de elementos de medida que los tests convencionales (Chang y Ying, 1996). Dos son los aspectos básicos de un TAI para conseguir el objetivo anteriormente enunciado. De un lado el método de estimación de las puntuaciones del evaluado y, de otro, el criterio de selección del ítem que se presentará al sujeto en cada momento para construir la estrategia adaptativa.

Diferentes estrategias han sido presentadas como criterio de elección de ítems en el transcurso del TAI. De entre ellas las más comunes son los criterios basados en la Máxima Información del Ítem y los procedimientos Bayesianos de estimación de varianza posterior (Owen, 1975). En la última década diferentes nuevos métodos han sido propuestos como desarrollo de éstos (Veerkamp y Berger, 1997; van der Linden, 1995; Chang y Ying, 1996; Davey y Parshall, 1995), con el objetivo de optimizar el número de ítems presentados en función de la precisión final de la estimación del nivel de rasgo o habilidad.

En estas líneas se presentan datos comparativos de la administración de un TAI sobre un banco unidimensional de respuesta gra-

duada de 28 ítems para la evaluación del ajuste emocional de las personas, utilizando como estrategia de selección de ítems a) criterio de máxima información; b) criterio de mínima entropía y c) criterio mixto de mínima entropía y máxima información.

La información del ítem es referida siempre a un nivel de rasgo o habilidad determinado.

Y se define para un modelo politómico por:

$$I_i(\theta) = \sum_{x=0}^{n_i-1} \frac{[P'_{x_i}(\theta)]^2}{P_{x_i}(\theta)}$$

De modo que en la integración del criterio de máxima información en el proceso TAI es necesaria la θ estimada para seleccionar el ítem más informativo. De este hecho se derivan los problemas básicos del procedimiento (Davey y Parshall, 1995; van der Linden, 1995; Chang y Ying, 1996) relacionados con la poca precisión de la estimación de θ en los primeros momentos del TAI. Por un lado, para la elección del primer ítem del TAI no hay estimación del nivel de rasgo. Y, por otro, no es posible la estimación de θ mediante máxima verosimilitud si el evaluado responde siempre en las categorías de respuesta extremas. Para el primer problema diferentes estrategias han sido utilizadas de entre las que destaca el procedimiento de asignar aleatoriamente una θ al sujeto entre unos valores predeterminados, y tomar ésta como primera estimación del nivel de rasgo o habilidad del evaluado. Respecto al segundo problema, el procedimiento más común ha consistido en la

utilización de un algoritmo step-size (Dodd, De Ayala y Koch, 1995). Ambos problemas se maximizan cuando se dispone de un banco de ítems pequeño, o se requiere que el TAI tenga muy pocos elementos.

En este trabajo se presenta un algoritmo para la selección de ítems que trata de superar los anteriores déficits. La entropía (Shannon, 1949) es una cuantificación imparcial de la bondad de un ítem, basada en las características de las CCI's por sí mismas, no teniendo en cuenta el nivel previo de estimación de la habilidad del sujeto. Este indicador mide el poder de discriminación del ítem y ofrece valores pequeños cuando las CCI's son estrechas y separadas. Por el contrario el indicador se incrementa cuando la distribución de probabilidades comienza a ser más homogénea. La entropía del ítem, para un nivel dado de θ , se define como:

$$S_i(\theta) = - \sum_{x=0}^{n_i-1} P_{x_i}(\theta) \ln P_{x_i}(\theta)$$

y su valor incondicional viene dado por:

$$\{S_i\} = \int_{-\infty}^{\infty} S_i(\theta) f(\theta) d\theta$$

Al comienzo de la administración del TAI, la estimación de θ , como se ha comentado anteriormente, es sumamente imprecisa. Es más, en el primer ítem del test es desconocida, de modo que criterios de selección del ítem que no tengan en cuenta dicha estimación pueden resultar más eficientes en los primeros momentos del TAI que criterios de Máxima Información que introducen en la selección del ítem la estimación previa de θ . Por el contrario, según avanza el TAI la estimación de la habilidad se realiza con mayor precisión, de modo que las estrategias que contemplan ésta resultarán más eficientes.

Método

Muestra

395 sujetos estudiantes de psicología en la Universidad Autónoma de Madrid con un rango de edad que variaba desde los 20 años hasta los 25, situándose la moda en 22 años.

Instrumentos

Banco de ítems extraído de un Cuestionario de Personalidad elaborado ad hoc con fines de Selección de Personal por el Servicio de Psicología Aplicada de la Facultad de Psicología de la U.A.M. El banco de ítems se compone de los 28 ítems de la escala de ajuste emocional del mencionado cuestionario, con una opción de respuesta graduada desde 1 (Totalmente de Acuerdo) hasta 6 (Totalmente en Desacuerdo). El banco se encuentra calibrado según el modelo de respuesta graduada de Samejima (1969) con el programa Parscale 3.0 (Muraki y Bock, 1996).

Procedimiento

Se realizaron un conjunto de simulaciones sobre los vectores de respuesta reales obtenidos por los sujetos de la muestra en la administración en lápiz y papel de la escala. La estimación del nivel

de θ de cada uno de los sujetos se realizó siempre mediante máxima verosimilitud.

Los criterios de selección de ítems utilizados fueron el criterio de Máxima Información y el Criterio de Entropía definidos anteriormente, junto con un criterio mixto que implementa ambas estrategias.

En el caso de Máxima Información el procedimiento TAI comienza con una estimación de θ aleatoria entre -1 y $+1$; y en caso de respuestas extremas que impidan la estimación por máxima verosimilitud se implementa un procedimiento de stepsize variable.

En el caso de Mínima Entropía el procedimiento TAI utiliza en todo momento como criterio para la selección del ítem siguiente la minimización de la función de entropía.

Por último en el caso Mixto, el procedimiento TAI comienza con un criterio de Mínima Entropía para elegir el primer ítem a administrar al sujeto, y, para el segundo y siguientes se utiliza el criterio de Máxima Información.

Se comparó mediante la correlación de Pearson la θ estimada por máxima verosimilitud en cada momento del proceso TAI (en las tres condiciones anteriores) respecto de la estimación de θ por máxima verosimilitud teniendo en cuenta todo el vector de respuestas de los sujetos.

Resultados

En la figura 1, en el eje de abscisas se representa el número de ítems utilizados en el proceso TAI para realizar la estimación de la habilidad de los sujetos. En el eje de ordenadas se representa la correlación existente entre la estimación de θ realizada con los 28 ítems del banco y la estimación de θ realizada en diferentes momentos del proceso TAI (es decir teniendo en cuenta diferente cantidad de ítems). Esta correlación es presentada en tres series diferenciadas en función del método de selección de ítems utilizado.

Como puede observarse, en los tres casos (máxima información, mínima entropía y criterio mixto) como era de esperar, la correlación aumenta progresivamente con el número de ítems utilizado acercándose a la unidad y mostrando básicamente los mismos resultados las tres estrategias a partir de la presentación de 15 ítems.

Sin embargo, hasta la presentación de los 15 ítems algunas diferencias deben ser señaladas.

Tomando la estimación de θ realizada con únicamente los tres primeros ítems administrados a los sujetos, se observa cómo el criterio de máxima información es menos eficiente que el criterio de

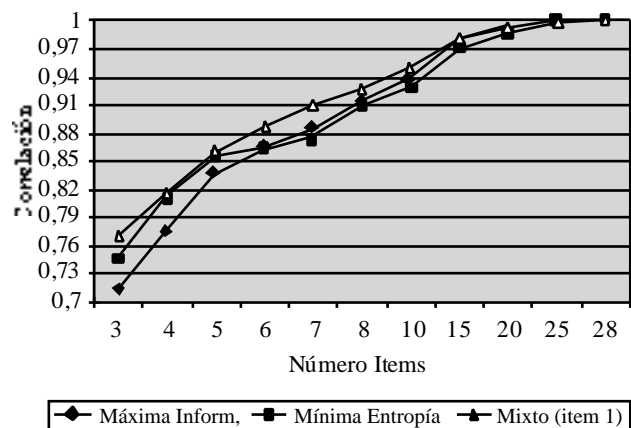


Figura 1. Rendimiento TAI en función del método de selección de ítems

mínima entropía, y éste lo es menos que el criterio mixto. Esto es debido a que, como preveíamos, con el criterio de máxima información al utilizar la habilidad estimada de los sujetos, y no ser ésta sumamente precisa al administrar únicamente los primeros elementos, la selección de los ítems no es la más adecuada para proporcionar una estimación de θ más precisa. Sin embargo, con el criterio de mínima entropía, aún produciéndose una selección de ítems más adecuada que con máxima información, se produce un rendimiento inferior que con el criterio mixto. Esto puede ser debido a que en el criterio mixto para la selección de los ítems segundo y tercero sí se utiliza la información del nivel de rasgo estimado mientras que en el criterio de mínima entropía no es tenido en cuenta. Por tanto, esta diferencia entre el criterio mixto y el de mínima entropía se debe fundamentalmente al buen funcionamiento del primer ítem administrado (seleccionado en ambos casos mediante mínima entropía). Por ello, la ubicación del nivel de θ de los sujetos tras la administración de este primer ítem proporciona una buena estimación como para comenzar a utilizarla en la selección del ítem siguiente (criterio de máxima información).

Dicha situación se reproduce básicamente cuando nos fijamos en los resultados obtenidos al analizar la estimación de θ tras 4 y 5 ítems, lo que supone un 17% del banco.

Por otro lado, al llegar a la administración de 6 elementos (20% del banco), el criterio mixto se sigue beneficiando del buen efecto del primer ítem, y el criterio de máxima información se recupera respecto del criterio de mínima entropía igualándose los resultados de ambos. Lo que parece indicar que, progresivamente, al ser cada vez más precisa la estimación de θ el criterio de máxima información selecciona ítems con mayor aporte de precisión.

Este efecto de mayor efectividad del criterio mixto, junto con la disminución de la efectividad del criterio de mínima entropía y aumento del de máxima información, va progresivamente disminuyendo al aumentar el número de ítems utilizados, de modo que al realizar la estimación de θ con los 15 primeros ítems (50% del banco) la precisión de los tres criterios comienza a igualarse significativamente.

Conclusiones

Los resultados presentados en este breve estudio muestran la viabilidad de implementación de un nuevo algoritmo de selec-

ción de ítems en un proceso adaptativo. En este sentido, el trabajo indica cómo algoritmos basados en criterios de entropía, en los cuáles no es tenida en cuenta la estimación de θ para la selección del ítem, pueden ser utilizados en los primeros momentos de desarrollo del TAI en el que dicha estimación es sumamente imprecisa. Sin embargo, conceptualmente, tal y como hemos presentado el algoritmo de mínima entropía, como algoritmo único de selección de ítems no produce un proceso adaptativo, ya que no incorpora ninguna información proveniente de la respuesta del sujeto.

Por ello, es necesario realizar una consideración fundamental. El algoritmo de mínima entropía, bien debe ser estudiado en un contexto mixto de selección de ítems, o bien debe incorporar información sobre la respuesta de los sujetos para producir un algoritmo realmente adaptativo.

En este sentido, dos líneas de trabajo han de ser desarrolladas. De un lado el estudio del criterio mixto: mínima entropía y máxima información, analizando en qué momentos del TAI la estimación de la θ del sujeto es suficientemente precisa como para cambiar del algoritmo de entropía al de máxima información. Y, de otro lado, la incorporación en el algoritmo de entropía la información derivada de la respuesta de los sujetos a cada ítem. De este modo un sujeto al responder a un ítem concreto, en función de la opción elegida, restringe el rango de θ posible para él. Y, es sobre ese rango de θ sobre el que debería aplicarse el criterio de mínima entropía y no sobre todo el rango posible (como por ejemplo, para la elección del primer ítem).

Un aspecto colateral ha de ser señalado. El algoritmo de entropía proporciona una solución satisfactoria para la elección del primer ítem a administrar en el TAI. No obstante, dicha solución entra en conflicto con criterios de exposición de ítems. Según el algoritmo, el mejor ítem del banco es el primero que se utiliza, y ese mejor ítem es siempre el mismo, de modo que el TAI comenzaría siempre con el mismo elemento de medida.

Por otro lado, el grado de ganancia del criterio mixto respecto del de máxima información reposa fundamentalmente en la bondad del ítem con menor entropía del banco. Cuanto «peor-mayor» sea esta entropía del «mejor ítem» del banco, menor diferencia existirá entre ambos criterios, por el contrario cuanto «menor-mejor» sea la entropía del «mejor ítem» del banco mayor ganancia respecto al criterio de máxima información se obtendrá.

Referencias

- Chang, H. y Ying, Z. (1996). A global information approach to Computerized Adaptive Testing. *Applied Psychological Measurement* 20, 213-229.
- Davey, T. y Parshall, C.G. (1995). New algorithms for item selection and exposure control with Computerized Adaptive Testing. *Paper presented at the annual meeting of the American Educational Research Association*, April 18-22, San Francisco.
- Dodd, B.G., De Ayala, R.J. y Koch, W.R. (1995). Computerized Adaptive Testing with Polytomous items. *Applied Psychological Measurement* 19, 5-22.
- Muraki, E. Y Bock, R.D. (1996). *Parscale 3.0*. Chicago: Scientific Software International, Inc.
- Owen, R.J. (1975). A bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association* 70, 351-356.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, Monograph Supplement 17.
- Shannon, C.E. (1949). A mathematical theory of communication. Part III. *Bell System Technical Journal*, XXVIII, 623-656.
- van der Linden, W.J. (1995). Bayesian Item Selection in adaptive testing. *Paper presented at the annual meeting of the Psychometric Society*. Minneapolis MN.
- Veerkamp, W.J. y Berger, M.P. (1997). Some new item selection criteria for adaptive testing. *Journal of educational and behavioral statistics* 22, 203-226.