|  |  |
| --- | --- |
| **Contenido**  En este artículo se hace una revisión amplia, pero introductoria, a la evaluación de modelos y los distintos métodos que existen para ello. Se enfatiza la existencia de tres grandes grupos para la evaluación de modelos: los métodos basados en teoría, las simulaciones y los métodos aplicados.  Los métodos basados en teoría implican el desarrollo de medidas formales para la evaluación de modelos con base a criterios bien justificados. Suelen ser los que más información proporcionan respecto de la adecuación de los modelos, pero no son factibles en el caso de modelos complejos compuestos por muchos parámetros. Los métodos de simulación implican el cómputo de métodos que exploran la relación entre los modelos y los datos obtenidos, son menos informativos pero considerablemente mejores para evaluar modelos complejos. Por su parte, los métodos aplicados evalúan de manera directa la habilidad de los modelos para predecir datos nuevos o diferentes. Son los más accesibles, sin embargo, no son claros en términos de exponer los aspectos específicos en que los modelos fallan o destacan. | **Método**  El artículo hace énfasis en las aproximaciones basadas en teoría, desde donde desarrolla tres aproximaciones:  1.- La perspectiva Bayesiana. Surge de la teoría de la probabilidad y da preferencia al modelo que corresponda con mayor probabilidad (likelihood) con los datos. Para ello recurre al factor de Bayes que no es otra cosa que la razón entre las verosimilitudes con que los modelos favorecen los datos.  2.- El principio de Minimun-Description-Lenght. Surge de la teoría de la información y da preferencia a los modelos que mejor aprehenden los datos y patrones observados. El método que en años recientes ha atraído mayor atención se conoce como Normalized Maximun Likelihood (NML)  3.- Los métodos de Predicción Secuencial. Como su nombre sugiere, surge de la filosofía clásica que asume que el mejor modelo es aquel que mejor predice datos futuros con base en los previamente observados. Para ello se calcula el Error de Predicción Acumulado, utilizando los datos obtenidos para estructurar nuevas predicciones y cotejarlas con lo observado. |
| **Desarrollo**  Los métodos previamente mencionados se desarrollan formalmente de la siguiente forma:  1.- Razón de Bayes    Se compara la verosimilitud con que los datos se ajustan a los modelos a comparar. Tomando en cuenta que para los parámetros, se marginaliza.  2.- NML    Para cada modelo se computa la razón entre la verosimilitud de ajuste del modelo con los datos y la verosimilitud de ajuste a cualquier otro conjunto posible de datos.  3.- Error de Predicción Acumulado.    Se utilizan los datos del primer momento para predecir el segundo y se compara con los datos reales. El proceso se repite y se suman las diferencias. | **Aplicación**  Estoy muy emocionada con este primer artículo. Constituye un tutorial, relativamente amigable, que explora de manera amplia las características, ventajas y desventajas de los distintos tipos de aproximaciones que existen para la evaluación de modelos en psicología.  Esto me permitirá continuar mi trabajo en el estudio de la percepción en situaciones de incertidumbre, evaluando el grado en que distintos modelos (en particular, el modelo propuesto por la Teoría de Detección de Señales y un modelo de inferencia bayesiana) se ajustan a los datos obtenidos.  Además, facilita y amplía mi comprensión de la estadística Bayesiana en un sentido más amplio: tanto en la construcción de modelos de actualización diacrónica de información, como en la evaluación de modelos con base en la razón de verosimilitud de su ajuste a los datos observados. |
| **Referencia:**  Shiffrin, R., Lee, M., Kim, W., Wagenmakers, E. (2008). A Survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods Cognitive Science. 1248-1284. | |
| **Contenido**  Aun viviendo en un mundo de procesos estocásticos (cuyas variables varían de forma aleatoria), los organismos somos capaces de adaptarnos a partir de mecanismos computacionales que permean la percepción de probabilidades simples, no estacionarias. Los modelos de aprendizaje adaptativo más comunes apelan a la regla delta (e.g. Rescorla y Wagner), donde se computa la discrepancia entre lo observado y lo predicho, ajustando sistemáticamente las estimaciones en favor de los resultados observados, en un algoritmo recursivo con cuatro características principales; (1) el ajuste depende solo del resultado actual, (2) su complejidad computacional no aumenta con la experiencia (3) eficiencia para estimar parámetros estocásticos y (4) neurobiológicamente plausibles.  Sin embargo, en experimentos donde el parámetro a estimar p (i.e. la probabilidad de que se observe uno de dos posibles eventos), se extrae de un proceso Bernoulli, un modelo de regla delta asumiría una tasa de aprendizaje baja incompatible con los cambios abruptos en p. Encontrándose en su lugar tasas de aprendizaje que varían dependiendo el desarrollo de la tarea. | **Método y Resultados**  Se plantean experimentos de estimación de probabilidad en que el valor del parámetro no es estacionario y los sujetos reportan sus estimados, observación a observación, permitiendo evaluar las respuestas transitorias del sistema perceptual ante cambios abruptos en lo observado.  Resultados: (1) Los sujetos ofrecen estimados razonablemente acertados acerca del rango de sus posibles valores (sin sobre o subestimar), (2) detectan y se adaptan a los cambios. (3) la detección es rápida para cambios substanciales, (4) realizan grandes cambios abruptos en el estimado, pero (5) también pueden mantener un único estimado por varios ensayos. Estos resultados evidencian la diferencia entre la estimación ensayo a ensayo sugerida por los modelos de regla-delta y los patrones que generan los sujetos, que dan un seguimiento acertado y de forma “escalonada” del parámetro subyacente.  Este patrón persiste aun añadiendo un umbral de cambio (donde la actualización depende de que la diferencia observada exceda cierto valor) |
| **Discusión**  Se propone un modelo Bayesiano de estimación del punto de cambio basado en dos características principales: (1) se actualiza el estimado sólo cuando hay evidencia de que el estimado actual es erróneo y (2) se compara Bayesianamente entre tres diferentes explicaciones, (a) el estimado errado está basado en supuestos errados, (b) ha habido un cambio real en *p* y (3) el último punto de cambio fue innecesario.    El modelo propuesto cuenta solo con dos parámetros libres; T1, el criterio de decisión en la evidencia a favor de un problema con la percepción en curso, y T2, el criterio de decisión en la posterior en favor de un cambio. Con esta simplicidad el modelo explica un par de fenómenos observados: (1) una mayor tasa de detección en las primeras dos sesiones, sugiriendo una creencia inicial y (2) la precisión con la que los sujetos estimaban la probabilidad a través de algoritmos que computan el estimado (ensayo a ensayo) del observador ideal y un medida apropiada del error, diferencia entre estimado real del sujeto y la predicción dada por el modelo del observador ideal dada la secuencia observada hasta entonces. | **Aplicación**  Todas las conclusiones a las que llega Gallistel se obtienen a partir del modelo de detección de señales. La sensibilidad de los sujetos a los cambios en el parámetro p, se evaluó a partir de la tasa de Hits y de Falsas Alarmas cometidos.  Lo que rescato de este artículo no es sólo la expansión del modelo al estudio de la percepción en un sentido más abstracto, sino también una ampliación en la definición de lo que se considera un Hit y una Falsa Alarma. En este artículo, se define un Hit como la primer ocasión en que el sujeto indica ‘detecté un cambio’ después de que ocurre un cambio real y una Falsa Alarma como todos los cambios indicados después de este primer momento, sin que haya ocurrido un cambio real en p. A primera vista, esto representa un problema para la metodología clásicamente propuesta en tanto que ambas tasas están definidas en relación a conjuntos distintos de estímulos (Hay menos oportunidades para cometer Hits en relación a las muchísimas Falsas Alarmas que podrían contabilizarse), sin embargo, sigue siendo posible trazar la respectiva curva ROC que permea la evaluación de la sensibilidad de los sujetos ante la tarea de detección propuesta. |
| **REFERENCIA**  Gallistel, C., Liu, Y., Krishan, M., Miller, R. (2014). The perception of probability. American Psychology Association. Pp 96-123. | |
| **Contenido**  El artículo enfatiza el papel de la utilidad en la Teoría de Detección de Señales en el estudio de la toma de decisiones en situaciones donde hay tanto incertidumbre perceptual como riesgo, permitiendo explicar con mayor claridad la relación entre la precisión y la optimización, así como una relación inversa entre la sensibilidad y los sesgos de respuesta.  Pensando en términos de utilidad, la SDT considera tres parámetros que caracterizan la ejecución del sistema en una tarea de detección: (1) Los pagos o consecuencias de acertar o errar; (2) La tasa base con que aparece el estímulo-señal de acuerdo al participante; (3) La similaridad que exista entre el estímulo de interés y cualquier otro parecido (ruido) que a su vez puede surgir de manera interna o externa al agente en cuestión. Con base en esta información el sistema establece un criterio que determina a partir de cuánta evidencia sensorial se va a juzgar la presencia o ausencia de la señal. | **Conflicto**  La versión utilidarizada de la SDT no del todo nueva. No obstante, proporciona observaciones teóricas importantes: (1) Existen contextos donde la Máxima Precisión entra en conflicto con la Utilidad Maximizada, de tal forma que la primera se sacrifica y (2) Plantea una relación funcional entre los sesgos y la sensibilidad.  Se propone una función de utilidad que permite evaluar la optimalidad del criterio en cada punto dependiendo de estos tres parámetros, permitiendo predecir ‘el mejor criterio posible’ en función a las condiciones específicas. Sin embargo, los sujetos pueden ‘malestimar’ alguno de estos parámetros y terminar actuando de manera subóptima. Con base en ello se subraya la necesidad de evaluar la cualidad óptima de la ejecución de los participantes, sin limitarnos a la descripción paramétrica de sus respuestas. |
| **Propuesta**  Dado que se asume debería existir una relación funcional entre los sesgos y la sensibilidad, los observadores con baja sensibilidad ‘deberían’ mostrar más sesgo. (i.e. A menor sensibilidad, más liberal el sesgo), ya que los errores se vuelven más probables y las conductas extremas ayudan a evadirlos.  Se propone una nueva ecuación para el trazo de una Línea de Respuesta Óptima que, en relación a las condiciones particulares de la tarea y como extensión de la relación funcional sesgo-sensibilidad ya descrita, sirve como un referente para evaluar qué tan buena resulta la ejecución de los participantes en relación a la tarea. La LOR arroja un valor óptimo para Beta (típicamente definida únicamente en función a las condiciones de la tarea), pero despejando para d’ y C (la distancia entre la intersección de las distribuciones y el criterio observado), obtenemos un valor óptimo que incorpora la importancia de la discriminabilidad de los estímulos.  ->  Cada combinación particular de Tasas bases y Pagos tiene una LOR exclusiva. Se interpreta la distancia entre el punto definido por la sensibilidad y el sesgo observados en el sujeto y la LOR, como un indicador de qué tan bien el observador ajusta sus sesgos para acomodarse óptimamente a sus niveles de sensibilidad. | **Aplicación**  Con base en lo sugerido por el artículo, quise poner a prueba la existencia de una relación inversa entre la discriminabilidad de la señal con respecto del ruido, y el peso que dan los participantes a las condiciones de la tarea.  Para ello, propuse utilizar una ilusión óptica ampliamente estudiada (i.e. Los círculos de titchner, o ilusión de Ebbinghaus) para establecer dos niveles de discriminabilidad en una tarea de detección. La idea de este primer experimento era corroborar (1) que existiera una diferencia en términos de d’ entre los dos niveles de ‘dificultad’ (a.k.a. discriminabilidad) propuestos por la literatura y (2) que dicha discrepancia correlacionara con valores distintos para el parámetro de sesgo C.  En específico: que cuando las personas respondieran a una tarea particularmente difícil (d’ baja), mostraran una mayor susceptibilidad a responder de acuerdo a sus sesgos (C grande), y viceversa. |
| Lynn, S., Barrett, F. (2014) “Utilizing” Signal detection theory. Association for Psychological Science. Pp. 1663-1673 | |