

Modelo: una máquina de inferencia con dominio

IDE-015 · Ideas fundamentales

Ensayos conceptuales para El universo y sus observadores

Autor: Juan Segurola

Versión: v0.8 ES · 19 de septiembre de 2026

Estado: Versión de publicación

Forma del artículo: Ensayo conceptual

Escala / contexto: Modelos analíticos, estadísticos y computacionales en todas las ciencias

Límite: Los modelos como máquinas de inferencia y cómo se evalúan. La relación de sustitución en sí corresponde a IDE-012; el modelo de medida dentro de una cadena metrológica corresponde a IDE-014

Trabajos relacionados: IDE-016, IDE-012, IDE-014

Perfil epistémico: Conocimiento establecido con límites interpretativos

Licencia en publicación: Todos los derechos reservados

Resumen del artículo

Un modelo es una estructura deliberadamente simplificada construida para sostener inferencias sobre otra cosa. Su valor nunca es absoluto: es condicional a una pregunta, un dominio y un conjunto de supuestos, y un modelo excelente para una inferencia puede no servir para otra construida sobre el mismo sistema.[5,6] El aforismo de Box de que todos los modelos son falsos pero algunos son útiles se cita sin parar y se lee sin cuidado; lo que sostuvo fue que el científico debe preguntarse si un modelo ilumina y no si es verdadero, y atender a los supuestos gravemente equivocados y no a la falsedad universal.[1] Este artículo expone qué debe declarar un modelo, por qué el acuerdo con los datos es evidencia más débil de lo que parece, cómo se eligen modelos cuando varios ajustan, y qué hacer cuando modelos buenos discrepan.

En 60 segundos

Todo modelo omite, y lo que omite lo elige la pregunta para la que se construyó, así que lo primero que un modelo debe declarar es su propósito y su dominio.[6] Ajustar los datos es la evidencia más débil que puede ofrecer. Un modelo numérico de un sistema natural abierto no puede verificarse en sentido lógico, porque el acuerdo con las observaciones es compatible con muchas estructuras de modelo y con errores que se compensan; Oreskes y colaboradores sostienen que tales modelos pueden confirmarse en cierto grado y nunca establecerse como verdaderos.[2] La parsimonia tiene una versión precisa y no estética: los criterios de información estiman la pérdida predictiva esperada y penalizan parámetros porque la flexibilidad innecesaria degrada la predicción sobre datos nuevos, no porque la simplicidad sea elegante.[3,4] La inferencia multimodelo evalúa conclusiones sobre un conjunto de modelos plausibles y puede identificar las que permanecen estables en ese conjunto.[4] Y un modelo que ajusta por razones

equivocadas, con parámetros que absorben el efecto de un mecanismo omitido, fallará justo donde se extrapole.

Por qué importa esta pregunta

Los modelos son el lugar donde de hecho se hacen las afirmaciones científicas, y los fallos más frecuentes no son errores dentro del modelo sino errores sobre él: usarlo fuera de su dominio, tomar el acuerdo con los datos por prueba y olvidar qué mecanismos quedaron fuera. Cada uno de ellos es invisible si el propósito y los supuestos del modelo no se enunciaron de entrada. Lo que está en juego es práctico, porque las decisiones de política e ingeniería descansan en extrapolaciones, y una extrapolación es precisamente el uso para el que la evidencia del ajuste pasado es más débil.

1. Un modelo se construye para una pregunta

Un modelo conserva algunos rasgos de su objetivo y descarta otros. Eso no es un compromiso lamentable sino la fuente de su utilidad: una representación que lo conservara todo sería tan inmanejable como el propio sistema. Lo que vuelve defendible la selección es que se hizo respecto de una pregunta, de modo que los rasgos omitidos son aquellos de los que la pregunta no depende. [5,6]

La representación y la similitud entre modelo y objetivo dependen del propósito.[5] Dos personas pueden escribir las mismas ecuaciones y estar afirmando cosas distintas, porque han asignado papeles distintos a los mismos términos.

La forma práctica de esto es una declaración. Un modelo debe enunciar para qué sirve, qué representa, qué excluye deliberadamente y en qué rango de condiciones se espera que valga. Un modelo sin eso no es más general: es menos comprobable, porque no hay condición enunciada bajo la cual pudiera detectarse que se aplica mal.

En qué punto estamos

Declarar propósito, exclusiones y dominio es lo que hace detectable una mala aplicación. Esa declaración cambia además lo que significa realmente la frase conocida sobre que todos los modelos son falsos.

2. Falso pero útil, dicho con precisión

El artículo de Box de 1976 suele reducirse a cinco palabras, y el argumento que las rodea es más útil que el eslogan. Escribiendo sobre modelos estadísticos de fenómenos naturales, sostuvo que, siendo falsos todos esos modelos, el científico no puede obtener uno correcto por elaboración excesiva y debe buscar en cambio una descripción económica que capture los rasgos importantes.[1] La recomendación es buscar los supuestos gravemente equivocados y no constatar que todos lo están.

Conviene conservar ese alcance. El asunto de Box son los modelos empíricos idealizados y no toda estructura formal, y leído como afirmación irrestricta de que nada puede describirse nunca correctamente el aforismo se convierte en una tesis filosófica y no en la metodológica que él defendió. Dentro de su alcance la distinción hace trabajo real: la falsedad universal no autorizaría nada, porque sería igual de cierta para el mejor y para el peor modelo disponible, mientras que si una simplificación concreta es consecuente para una inferencia concreta es una pregunta con respuesta, y la respuesta se halla perturbando el supuesto y mirando si la conclusión se mueve.

Box subrayó también la iteración. Un modelo se critica contra los datos, se diagnostican sus fallos y se modifica, lo que convierte la construcción de modelos en un proceso y no en un producto.[1] Los residuos son la parte productiva de ese lazo, porque un patrón sistemático en lo que el modelo no explicó es información sobre lo que quedó fuera.

En qué punto estamos

El argumento de Box, sobre modelos estadísticos de fenómenos naturales, es que ninguna elaboración produce uno correcto, así que el trabajo consiste en hallar qué supuestos están equivocados de forma consecuente para la inferencia en curso, e iterar contra los residuos. La falsedad universal por sí sola no excusa nada. La pregunta siguiente es qué establece realmente el acuerdo con los datos una vez ajustado un modelo a ellos.

3. Ajustar no es confirmar

Oreskes, Shrader-Frechette y Belitz sostuvieron que los modelos numéricos de sistemas naturales abiertos no pueden verificarse ni validarse en el sentido que esas palabras tienen en lógica y en ingeniería del software. Los sistemas naturales nunca están cerrados, los resultados del modelo nunca son únicos y el acuerdo con la observación es compatible con más de una estructura de modelo.[2]

El mecanismo es fácil de enunciar. Los parámetros se ajustan para encajar; un parámetro puede absorber el efecto de un mecanismo ausente; y dos errores pueden cancelarse en el rango de los datos. El modelo ajustado reproduce entonces las observaciones mientras codifica un relato equivocado de por qué salieron así, cosa indetectable desde el ajuste y visible solo fuera del rango ajustado.

Lo que tales modelos sí pueden es confirmarse en cierto grado por el número y la variedad de observaciones que reproducen, y usarse heurísticamente para explorar consecuencias de supuestos.[2] Es un estatuto epistémico real y no una descalificación, pero más débil de lo que sugiere el vocabulario de la validación, y la diferencia importa sobre todo en la extrapolación, que es el uso que la evidencia menos sostiene.

En qué punto estamos

El acuerdo con los datos no verifica un modelo de un sistema abierto, porque los parámetros pueden absorber mecanismos ausentes y los errores pueden cancelarse en el rango ajustado. La confirmación admite grados y proviene de la variedad de observaciones reproducidas, no de la bondad del ajuste. Queda la cuestión práctica de cómo elegir cuando varios modelos ajustan aceptablemente.

4. Elegir entre modelos

Añadir parámetros siempre mejora el ajuste a los datos disponibles, así que el ajuste no puede ser el criterio. La idea de Akaike fue definir el objetivo como el rendimiento predictivo esperado sobre datos nuevos del mismo proceso, y mostrar que la verosimilitud maximizada es un estimador sesgado de él, con un sesgo relacionado con el número de parámetros estimados.[3] El criterio resultante penaliza la complejidad por una razón enunciada y no por principio.

Eso convierte la parsimonia de preferencia estética en argumento predictivo. Un modelo con flexibilidad innecesaria ajusta el ruido igual que la señal, y sus predicciones sobre datos nuevos

empeoran por ello. Burnham y Anderson desarrollan el aparato práctico, incluida la comparación de modelos candidatos sobre un conjunto común de datos y la ponderación de modelos por su apoyo relativo.[4]

Dos cautelas acompañan a la herramienta. Los criterios comparan los candidatos ofrecidos y no dicen nada sobre un modelo mejor ausente del conjunto, así que un modelo bien apoyado puede ser el mejor de un campo pobre. Y la comparación solo es válida entre modelos ajustados exactamente a los mismos datos, requisito fácil de violar sin darse cuenta.[4]

En qué punto estamos

La elección de modelo no puede apoyarse en el ajuste, porque el ajuste siempre mejora con parámetros, así que los criterios estiman el rendimiento esperado sobre datos nuevos y penalizan la complejidad por esa razón. La parsimonia pasa a ser un argumento predictivo y no un gusto. Los criterios ordenan solo los candidatos ofrecidos, lo que plantea qué hacer cuando varios modelos plausibles sobreviven juntos.

5. Cuando modelos buenos discrepan

La inferencia multimodelo evalúa conclusiones sobre un conjunto de modelos plausibles y puede identificar las que permanecen estables en ese conjunto. Una simulación molecular, una ecuación de medio continuo y un sustituto estadístico pueden acertar cada uno en su dominio sin ser intercambiables, porque responden preguntas distintas a costes distintos.[5,6] La discrepancia entre ellos informa en vez de avergonzar.

Eso da también la manera honesta de comunicar un resultado de modelado. Enuncia la pregunta, el dominio, los supuestos que se variaron, las conclusiones que fueron estables y las que no. Un solo número de un solo modelo, presentado sin esa estructura, es la forma en que se usan mal casi todos los resultados de modelado.

En qué punto estamos

Un resultado que se invierte al cambiar una elección estructural es un hecho sobre el modelo. IDE-016 recoge una propiedad que a los modelos se les pide cuantificar y que resiste una medida única: la complejidad.

Lo que sabemos y cómo lo sabemos

- La representación y la similitud entre modelo y objetivo dependen del propósito.[5,6]
- Box sostuvo, sobre modelos estadísticos de fenómenos naturales, que la elaboración no puede producir uno correcto y que la tarea es hallar qué supuestos están equivocados de forma consecuente, procediendo iterativamente contra los residuos.[1]
- Los modelos numéricos de sistemas abiertos no pueden verificarse; el acuerdo con los datos es compatible con múltiples estructuras y con errores que se compensan, así que la confirmación admite grados.[2]
- Los criterios de información penalizan parámetros porque la flexibilidad innecesaria degrada la predicción sobre datos nuevos, y ordenan solo los modelos candidatos realmente ofrecidos. [3,4]

Lo que todavía no sabemos

- Cómo caracterizar la fiabilidad predictiva de modelos de dimensión muy alta cuyo número efectivo de parámetros no está bien definido, donde los argumentos clásicos de penalización no se aplican sin más.
- Cómo ponderar entre sí modelos estructuralmente distintos cuando no se ajustan a una verosimilitud común, que es la situación habitual en los conjuntos multimodelo climáticos y epidemiológicos.
- Si la robustez en un conjunto de modelos plausibles es evidencia sobre el sistema o evidencia sobre los supuestos compartidos de la comunidad que produjo el conjunto.

Qué podría hacernos cambiar de opinión

- Un procedimiento demostrado que estableciera un modelo numérico de un sistema natural abierto como verdadero y no como confirmado en cierto grado, superando las objeciones de no unicidad y de errores que se compensan, invalidaría el estatuto epistémico estándar de esos modelos.[2]
- Evidencia reproducible de que los modelos seleccionados por criterios de información predicen peor datos nuevos del mismo proceso que los seleccionados por ajuste dentro de muestra, en entornos variados y con conjuntos de datos correctamente emparejados, socavaría la justificación predictiva de la parsimonia.[3,4]
- Un caso trabajado en el que una conclusión robusta en toda estructura de modelo plausible resultara falsa sobre el sistema, con el fallo atribuido a un supuesto compartido y no al azar, debilitaría el análisis de robustez como evidencia.[4,5]

Si solo recuerdas una cosa

Un modelo es una máquina de inferencia con pregunta, dominio y exclusiones enunciados; ajustar los datos es la evidencia más débil que puede ofrecer.

Artículos relacionados

- IDE-016 — Complejidad
- IDE-012 — Representación
- IDE-014 — Medición

Referencias

- [1] George E. P. Box. Science and Statistics. Journal of the American Statistical Association 71(356), 791-799 (1976). DOI 10.1080/01621459.1976.10480949.
- [2] Naomi Oreskes; Kristin Shrader-Frechette; Kenneth Belitz. Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences. Science 263, 641-646 (1994). DOI 10.1126/science.263.5147.641.
- [3] Hirotugu Akaike. A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control 19(6), 716-723 (1974). DOI 10.1109/TAC.1974.1100705.
- [4] Kenneth P. Burnham; David R. Anderson. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed. Springer (2002).
- [5] Michael Weisberg. Simulation and Similarity: Using Models to Understand the World. Oxford University Press (2013).

[6] Ronald N. Giere. How Models Are Used to Represent Reality. *Philosophy of Science* 71(5), 742-752 (2004). DOI 10.1086/425063.