

Proyecto 42: tras el “conductoma” humano

Laura Vargas-Parada

17 de diciembre de 2018

“Voy a comenzar con una historia de ciencia ficción”, dijo Christopher Stephens a la audiencia, el pasado 28 de noviembre, al iniciar la ponencia donde presentó algunos de los resultados de su investigación dentro de la 3.^a Semana de la Complejidad organizada por el Centro de Ciencias de la Complejidad (C3) de la UNAM.

La historia, explicó Stephens, coordinador de Ciencia de los Datos del C3 e investigador del Instituto de Ciencias Nucleares de la UNAM, se publicó en la serie *The Hitchhiker Guide to Galaxy* del escritor y guionista británico Douglas Adams y habla sobre una civilización avanzada que construyó una computadora, *Deep Thought*, para calcular la respuesta a la pregunta fundamental: el por qué del sentido de la vida, el universo y todo lo demás. Eventualmente, la computadora llegó a una respuesta: 42.

La máquina explicó que para comprender qué significa el 42 había primero que comprender el verdadero sentido de la pregunta original. Los que preguntaron en primera instancia realmente no sabían qué era lo que estaban preguntando por lo que se requeriría una computadora todavía más poderosa para calcular la pregunta para la respuesta.

“Ahora les mostraré mi 42”, dijo Stephens. Un modelo predictivo para obesidad basado en más de 2,000 variables que están siendo estudiadas en distintas poblaciones de la UNAM con apoyo de un grupo multidisciplinario de investigadores. El modelo parece dar buenos resultados predictivos, pero resulta ser equivalente a un 42 porque los investigadores aún no pueden

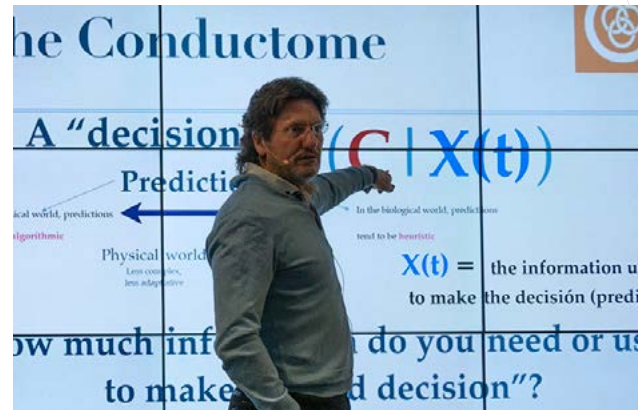
comprender cómo estas variables se relacionan, en el detalle más fino, con los miles de factores directos e indirectos que influyen en la toma de decisiones.

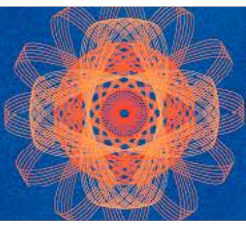
Tras el conductoma

A pesar de la enorme inversión en recursos humanos y financieros, la epidemia de obesidad continúa en aumento, y junto con ella, la incidencia de las enfermedades crónicas asociadas.

En la década de 1980, sólo 7% de los mexicanos eran obesos. Ese número se triplicó para 2016 (20.3%), de acuerdo al *Institute for Health Metrics and Evaluation* de la Universidad de Washington, EUA. Por su parte, un análisis publicado en la revista *Public Health Nutrition* estima que para 2050 la proporción de hombres y mujeres obesos en el país llegará al 54 y 37% respectivamente.

Hasta ahora, las iniciativas implementadas han tenido poco efecto en revertir la tendencia lo que para Stephens y su equipo de colaboradores significa “que aún falta algo fundamental en nuestra comprensión de las causas y efectos”.





Con el supuesto de que la obesidad es una enfermedad causada por comportamientos particulares, Stephens y su equipo de trabajo proponen el concepto de “conductoma”: un conjunto completo de factores que influyen en el comportamiento humano y en la toma de decisiones. Han denominado a su proyecto *Proyecto 42*. El objetivo a largo plazo es obtener y modelar datos que permitan comprender mejor el conductoma humano con el fin de realizar pronósticos sobre el comportamiento.

“Aunque el conductoma es enormemente multifactorial”, explicó Stephens, “la revolución en la ciencia de datos y los recientes desarrollos tecnológicos asociados nos permiten poder comenzar a construirlo”.

Un modelo predictivo para la obesidad

Los primeros resultados del modelo predictivo que presentó Stephens en su conferencia se basan en datos de un estudio de 1,076 académicos y no académicos de la UNAM. Se analizaron 2,524 variables incluyendo variables genéticas, epidemiológicas (incluyendo historia personal, historia familiar, nutrición, estilo de vida, conocimientos de salud, autoevaluación de la salud) y antropométricas/fisiológicas (como niveles de colesterol, insulina, hemoglobina, triglicéridos, etc.). Las variables se compilaron a través de cuestionarios y entrevistas personales así como mediante estudios de laboratorio. Para elaborar los cuestionarios los investigadores se apoyaron en cuestionarios nacionales como ENSANUT.

Los datos personales incluyen preguntas desde a qué se dedica la persona o dónde vive, a variables sociales como cuántos amigos tiene, historia personal de salud, cómo se siente, preguntas sobre percepción de su salud e incluso sobre qué cosas sabe esa persona.

“Algunas de las preguntas que hacemos pueden sonar un poco raras”, explicó Stephens a la audiencia. Por ejemplo, cuántas calorías hay en un vaso con agua. Este tipo de pregunta resulta muy predictiva para saber quién tiene obesidad, pero el problema que tienen los investigadores, su 42 retomando la analogía del principio, es que aún no tienen suficiente información para comprender por qué esta pregunta en particular resulta predictiva.

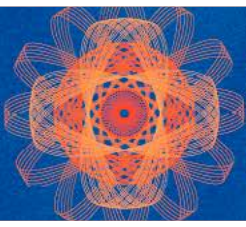
Con respecto a la población universitaria, el análisis arroja un dato interesante, los académicos tienen una tasa de obesidad del 14% mientras que en los trabajadores no académicos la tasa de obesidad alcanza el 35-40%. Esto es, los académicos tienen 3 veces menos obesidad que los no académicos. Nuevamente, esta primera aproximación indica algo, pero la explicación puede estar en muchísimas variables: el nivel de estudios, los recursos socioeconómicos, la edad, etc. Para los investigadores, el conductoma está esencialmente en tratar de determinar cómo todas esas variables se relacionan con el comportamiento.

Para estudiar la base de datos los investigadores utilizan un algoritmo de aprendizaje automático (*machine learning*). “Lo que hacemos es tomar todas las variables y analizarlas con este algoritmo para identificar las combinaciones de variables que se presentan en personas obesas y no obesas”, dijo Stephens. En ese estudio el aprendizaje automático es fundamental para combinar múltiples variables y comparar sus contribuciones.

La revolución de los datos

Para los investigadores, los primeros resultados obtenidos con la base de datos que han logrado compilar y los análisis asociados, sugieren que es factible la construcción iterativa del conductoma, a la vez que permite identificar las principales oportunidades y desafíos.

La propuesta del grupo de investigación es que hay esencialmente un conductoma para cada decisión que tomamos. Y este conductoma está determinado por la influencia que múltiples factores tienen en particular para dicha acción. Aún no se cuenta con toda esa información pero la revolución de datos está ayudando a generarlos rápidamente.



Otras pregunta que los científicos tratan de resolver con la construcción del conductoma son: ¿qué nos lleva a tomar buenas o malas decisiones? ¿qué información es necesaria y qué información, acaso, es suficiente para decidir? ¿las malas decisiones están codificadas en nuestros genes? (esto es, si adaptaciones que fueron útiles evolutivamente en algún momento ahora nos resultan perjudiciales).

Stephens destaca que es importante enfatizar que “aunque los datos son una condición necesaria para construir el conductoma, no son suficientes, requieren nuevos marcos conceptuales y metodológicos para el análisis de estos datos”.

Es ahí donde entran las ciencias de la complejidad con su aproximación integral a los problemas y la disponibilidad de nuevas técnicas de modelado como el aprendizaje automático (*machine learning*) y el modelado basado en agentes.

Definitivamente, concluyó Stephens, “los enfoques reduccionistas y disciplinarios tradicionales de la ciencia” resultan inadecuados para enfrentar este tipo de problema.

Los investigadores esperan publicar los primeros resultados del “conductoma” de la obesidad durante 2019.

