

Meeting Transcription

Meeting started: Nov 30, 2024, 5:57:58 PM

Meeting duration: 14 minutes

Meeting participants: Mateo Builes Duque

[View original transcript at Tactiq.](#)

Transcript

00:00 Mateo B.D.: el trabajador de un banco alemán robó el dinero de la caja fuerte y desapareció sin dejar rastro La policía sospechaba que se dirigía a Londres pero les llevaba días de ventaja era inútil enviar la fotografía del ladrón por correo postal porque no llegaría a tiempo para capturarlo con el dinero encima. Por suerte en esa época ya existía la línea telefónica Pero esto planteaba un problema, se puede enviar una imagen por teléfono, o sea, se puede grabar un audio que permita reconstruir una copia casi exacta de la imagen al escucharlo? Cuánto duraría se puede inventar un código para que dure menos cuánto exactamente? no es un problema sencillo, estamos acostumbrados a comunicarnos con textos imágenes audios y vídeos y sabemos que todos ellos transmiten información, pero da la sensación de que son informaciones diferentes que no se pueden transformar entre sí sin embargo, en 1948 apareció un artículo que cambió nuestra forma de ver el problema la teoría de la información de Shannon desde entonces sabemos que cualquier tipo de mensaje sin importar su formato tiene una cantidad precisa de información que se puede medir usando la misma unidad el bit para calcular la información hay que seguir una fórmula muy extraña, pero esa fórmula consigue unificar todos los medios de comunicación Bajo la misma teoría y es muy útil en otras ramas de la ciencia desde los agujeros negros al ADN Pero Qué significa esta fórmula Y por qué ha cambiado nuestra manera de entender la comunicación? Bueno, empecemos por el principio.

01:40 Mateo B.D.: La teoría de la información nació para mejorar el envío de mensajes y una buena forma de verlo, es pensar que un mensaje contiene la

respuesta a una pregunta que le has hecho a alguien a veces la pregunta tendrá dos respuestas posibles. A veces tendrá cientos de respuestas Y a veces habrá tantas respuestas como combinaciones de letras, te puedas imaginar Eso sí, al hacer la pregunta preparas una lista mental con todas las posibles respuestas Y sabes que el contenido del mensaje será una de ellas.

02:13 Mateo B.D.: Shannon pensaba que la información de un mensaje se debía medir como la cantidad de preguntas de sí o no que necesitarías para averiguar Su contenido y a cada respuesta la llamó bit. Por ejemplo si un mensaje contiene Cuál es mi planeta favorito del sistema solar sabrás que la respuesta está entre ocho candidatos y Podrías hacerme preguntas para descartar algunos. Sí el nombre de mi planeta termina no, no tiene un aspecto azulado Y sí, sí tiene tiene anillos tres preguntas y adivinado y sea cual sea el planeta que elijas al principio siempre lo puedes encontrar con tres preguntas.

02:50 Mateo B.D.: Por lo tanto esta información tiene un valor de 3 bits. Cuántas preguntas, necesitas para encontrar al candidato correcto de una lista Bueno lo más Sensato es hacer preguntas que dividan al grupo por la mitad? Y puedes hacer preguntas de este estilo hasta que solo te quede una opción la opción correcta. El problema Entonces se puede plantear como Cuántas veces puedes dividir por dos el grupo inicial hasta quedarte con un solo candidato? Y Oye puedes hacer las cuentas a mano, pero ya existe una operación para esto el logaritmo en base 2 y si sé que la palabra logaritmo puede asustar Así que imagínatelo como una tecla en la calculadora, una tecla que te dice el número de preguntas que necesitas para quedarte con un solo candidato.

03:35 Mateo B.D.: Y de momento Esta es la información que tiene un mensaje. Por poner ejemplos hay 10.000 números posibles para el pin de tu tarjeta necesitaría unas 13,3 preguntas para averiguarlo la información de tu pin es de 13,3 bits. Al escribir un tweet imagina que puedes usar 100 caracteres distintos necesitaría unas 6,6 preguntas para averiguar Cuál es el carácter concreto que quieres usar. Si el mensaje tiene 280 caracteres Necesito unas 1860 preguntas para averiguarlo o sea, 1860 bits de información y lo mismo para una imagen pequeña compuesta de píxeles, si cada Pixel puede ser de ocho colores distintos necesito tres preguntas por píxel y multiplicándolo por el número de píxeles en fin Necesito unas 29.000 preguntas para averiguar el aspecto exacto de la imagen.

04:30 Mateo B.D.: Y esto ya nos enseña Cómo enviar una imagen por teléfono Oye escucha, pinta una muestra con 8 tonos de gris, el color del primer Pixel está en la parte derecha de la muestra, Vale y de los colores que quedan también está a la derecha y de los que quedan sigue a la derecha. para el color del segundo Pixel está en la parte derecha de una vez contestadas toda la información de la imagen estará codificada en estas 28.800 respuestas con ellas puede reconstruir la imagen original de forma exacta y podrías enviarlas por voz o con puntos y rayas al mandar dos sonidos por segundo la imagen llegaría a Londres en unas 4 horas y Esto fue muy a grandes rasgos lo que pasó en 1907 con ayuda de una máquina en París la policía hizo llegar la imagen a Londres unas horas antes que el ladrón y fue detenido en la misma Estación Esta es la teoría de la información más básica que podemos imaginar cualquier mensaje tiene una cantidad precisa de información que se puede codificar en forma de sí es y no es de puntos y rayas o de ceros y unos Y aunque esta fórmula es muy convincente Shannon advirtió una paradoja muy sutil Y si quisiéramos enviar una fotografía completamente Blanca tendríamos que enviar las 28.800 respuestas como antes o podríamos decir Oye todos los píxeles son blancos con los dos mensajes podemos recuperar la imagen original Así que contienen la misma información, pero Por qué uno es mucho más corto que el otro como se explica Resulta que no todas las preguntas de sí o no son iguales a veces dividen al grupo de candidatos por la mitad y a veces a veces no no todas las respuestas nos aportan la misma información.

06:14 Mateo B.D.: Este detalle se puede arreglar diciendo que la información de una respuesta debe tener en cuenta los candidatos que había antes de hacer la pregunta y también los que quedan después de contestarla. Mira con planetas queda más claro que te digan que el nombre de mi planeta favorito sí acaba en o te hace pasar de 8 candidatos a solo cuatro. Según la calculadora te aporta un bit de información.

06:38 Mateo B.D.: Y los cálculos serían prácticamente los mismos si te contestará que no. Ahora bien, Si te dijera que mi planeta favorito sí tiene koalas pasarías de 8 candidatos a solo uno esa respuesta te aporta los tres bits de información de golpe, o sea es como si te dijera directamente que he elegido la tierra y antes hubieras necesitado tres preguntas para averiguar esto, por eso son tres bits. En

cambio Si te digo que no tiene koalas, seguirás dudando entre 7 candidatos Así que te apor to muy poca información.

07:08 Mateo B.D.: Lo importante es que esta fórmula tiene en cuenta que no todas las respuestas son igual de informativas lo que la hace mucho más versátil. Además, la fórmula está muy relacionada con la probabilidad de las respuestas, si eligieras el planeta al azar, solo una de cada ocho veces tendría koalas y siete de cada ocho veces no los tendría estos son los números que hemos introducido en la fórmula de Shannon pero invirtiendo su orden.

07:37 Mateo B.D.: Por eso es frecuente encontrar la fórmula escrita de esta manera y los números cuadran bastante, si sabes al 100% cual va a ser la respuesta a una pregunta, no te aporte información porque ya sabías la respuesta antes de preguntar. Si la probabilidad de una respuesta es del 50% entonces obtienes un bit de información. Y Cuanto más improbable sea la respuesta más información te aportará. Por eso algunos dicen que esta fórmula mide la sorpresa y un pequeño ejemplo con la lotería deja claro, por qué si te dicen algo que era poco probable te generan muchas sorpresa y te dan mucha información.

08:15 Mateo B.D.: Y cuando te dicen algo que ya te podías imaginar ni es sorprendente ni te aporta Demasiada información. Y este es el punto clave de Shannon porque en la vida cotidiana los mensajes se agrupan formando estructuras más o menos probables, no elegimos el número pin al azar Y es probable que alguna vez hayas usado el 1, 2, 3 4 Si escribes el Twitter en español sabrás que hay letras mucho más frecuentes que otras y es probable que la que más se repita sea la e. Y las imágenes no son píxeles distribuidos al azar, sino que hay colores más frecuentes que otros en la fotografía del ladrón, por ejemplo, la mitad de los píxeles son blancos Así que a la hora de preguntar por el color de un píxel lo mejor es preguntar si es blanco porque la mitad de las veces averiguarás el color con una sola respuesta.

09:01 Mateo B.D.: si te contestaran que no deberías preguntar por el color negro en esta situación, la mitad de las veces acertarías el color con Solo dos respuestas a lo que voy es que si conoces la probabilidad de las respuestas, puedes organizar mejor tus preguntas con este método cada color necesita una cantidad distinta de preguntas, pero los colores que más se repiten son los que Las preguntas requieren y el proceso general es mucho más rápido, Cuántas

preguntas, necesitamos para codificar la fotografía del Ladrón de esta forma Esta es la pregunta del millón, solo algunos colores aparecen en la fotografía y cada uno en una proporción distinta.

09:37 Mateo B.D.: Como conocemos estas proporciones podemos calcular la sorpresa de cada uno. Uno de cada dos píxeles es blanco, así que nos da una información de un bit, o sea, lo podemos encontrar haciendo Solo una pregunta. El Pixel de color negro es un poco más raro y nos da una información de dos bits porque lo podemos encontrar haciendo dos preguntas. En fin, cada color necesita un número distinto de preguntas según nos dice la fórmula de la sorpresa.

10:04 Mateo B.D.: Y como cada uno necesita un número distinto de preguntas, lo mejor es hacer una especie de media ponderada. Una de cada dos veces necesitaremos Solo una pregunta para adivinar el color del Pixel Porque será blanco. Una de cada cuatro veces necesitaremos dos preguntas, Porque será negro, Si contamos todas estas contribuciones. Veremos que para averiguar el color de un píxel necesitamos de media unas dos preguntas. O sea, en el primer capítulo necesitábamos 28.800 preguntas para enviar la información de la imagen y ahora solo 19.200 por eso la información real de la imagen es de 19.200 bits y se podría haber enviado a Londres en dos horas y media con las preguntas adecuadas.

10:49 Mateo B.D.: lo que quiero remarcar es que este cálculo de Shannon funciona con cualquier tipo de mensaje sin importar su formato o si lo puedes entender Si conoces la probabilidad de los símbolos que aparecen en el mensaje, puedes calcular la información que contiene con este método calculas la sorpresa, haces una media ponderada y el resultado se puede resumir en una fórmula como esta es ligeramente más complicada que la que usábamos en el primer capítulo, pero es mucho más potente aún así en matemáticas no nos interesan las fórmulas solo porque calculen algo sino porque resumen una idea en unos pocos símbolos la idea de buscar algo dividiendo a los candidatos entre dos o la de aprovechar que hay candidatos más probables que otros para ahorrarte preguntas, pues Hemos llegado a un punto donde tenemos dos fórmulas para hablar de informaciónCuál es la información de verdad coloquialmente, podríamos decir que las fórmulas nos dan datos e información puedes tener una gran cantidad de datos pero Ca información en ellos como pasaba con la imagen

de color blanco los dos métodos nos permiten recuperar el mensaje original pero la información de Shannon nos dice Cuál es la cantidad óptima de preguntas que necesitamos por eso la información suele ser más pequeña que los datos Porque está optimizada.

12:05 Mateo B.D.: Normalmente en un mensaje cualquiera hay más datos que información esta diferencia se debe a que la información está repetida y es redundante piensa en un mensaje de texto, si puedo quitar letras y el mensaje se sigue entendiendo es porque la información de esas letras está repartida por el resto del mensaje. Está repetida a veces está redundancia. Es útil. Si cometes un error al escribir un mensaje la redundancia, te ayuda a detectarlo y a que puedas entender el mensaje original y si no te interesa la redundancia, porque te hace perder espacio, puedes eliminar las partes repetidas de un mensaje y comprimirlo hasta el límite de Shannon Cómo hacen los algoritmos de compresión del ordenador, se comprime el tamaño, pero la información no se pierde.

12:46 Mateo B.D.: Y bueno, quizá debería comentaros un par de cosas porque el envío de mensajes no es tan bonito como lo he pintado en el vídeo a veces la probabilidad de un símbolo cambia según los símbolos que lo rodean lo que complica los cálculos, pero permite comprimir los datos de forma más eficiente. Y lo más seguro es que tengas que reservar un trozo del mensaje para explicar Con qué preguntas has codificado los síes y los noes Pero estos son problemas de cálculo la idea que hay detrás sigue siendo la misma y por eso la teoría de la información ha sido un éxito en el mundo matemático y tecnológico y es increíble todas las aplicaciones inesperadas que tiene desde los concursos de la tele de toda la vida.

13:24 Mateo B.D.: Pasando por la física porque se parece mucho a la fórmula de la entropía. hasta el ADN y en eso justamente es en lo que estoy trabajando ahora con la gente de adentro en una investigación que me flipa la versión resumida es que en el ADN está codificada nuestra información genética en un mensaje muy largo compuesto por cuatro símbolos, como es un mensaje podemos usar la teoría de la información para ver qué partes del código no aportan mucha información porque aparecen en el 100% de la población y también, qué partes del genoma son las más raras las que tienen muchas sorpresas Y de alguna manera hacen a cada individuo raro y único y ahora tú puedes participar en esta

experiencia en adentro, puedes conseguir un test de ADN de los que se hacen desde casa y con el que

[View original transcript at Tactiq.](#)