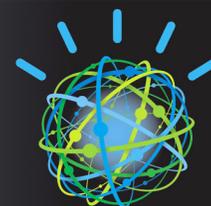


La Era de los Sistemas Cognitivos: Una mirada al interior de IBM Watson y ¿Cómo funciona?



IBM WATSON™

Redguides
para Líderes Empresariales

Rob High



- Sepa cómo los sistemas cognitivos, tal como IBM Watson, pueden transformar la manera de pensar, actuar y operar de las organizaciones
- Entienda las capacidades de procesamiento de lenguaje natural y más de IBM Watson
- Vea cómo las respuestas basadas en evidencias pueden impulsar mejores resultados



Visión general ejecutiva

IBM® Watson™ representa un primer paso en sistemas cognitivos, en nueva era de computación. Watson se basa en la era actual de computación programática, pero difiere de maneras significativas. La combinación de las siguientes capacidades que hace a Watson única:

- ▶ *El procesamiento de lenguaje natural*, al ayudar a entender las complejidades de datos no estructurados, constituyen en gran parte los 80 por ciento de los datos en el mundo de hoy
- ▶ *Generación y evaluación de hipótesis* mediante la aplicación de la analítica avanzada para analizar y evaluar un panel de respuestas con base en evidencias relevantes solamente
- ▶ *Aprendizaje dinámico* mediante la ayuda para mejorar el aprendizaje con base en los resultados con el fin de ser más inteligente con cada iteración e interacción

Aunque ninguna de estas capacidades por sí solas son exclusivas de Watson, la combinación ofrece una solución de gran alcance:

- ▶ Ir más allá de los límites de la computación programática
- ▶ Pasar de la dependencia de los datos estructurados y locales al desbloqueo del mundo de datos globales, no estructurados
- ▶ Pasar de decisión basada en los árboles, de aplicaciones deterministas a sistemas probabilísticos que co-evolucionan con sus usuarios
- ▶ Pasar de la búsqueda por palabra clave que brinda una lista de lugares donde una respuesta podría (o no) ubicarse, a un medio intuitivo, conversacional de descubrir un conjunto de respuestas calificadas de confianza

Sistemas cognitivos, tales como IBM Watson, pueden transformar la manera de las organizaciones pensar, actuar y operar en el futuro. Esta publicación de IBM Redguide™ describe cómo Watson combina el procesamiento de lenguaje natural, aprendizaje dinámico y evaluación y generación de hipótesis para dar respuestas directas, con base en la confianza.

De qué lenguaje se trata y por qué es difícil para las computadoras entender

Lenguaje es la expresión de ideas. Es el medio por el cual transmitimos una comprensión de las cosas entre las personas. Es la forma como transmitimos el miedo, la esperanza, la historia y las direcciones para el futuro. Algunos dicen que es lo que solemos pensar, especular e imaginar. Está en la base de nuestra cognición, nuestra capacidad de comprender el mundo que nos rodea o, por lo menos, la base de nuestra capacidad de manipular e intercambiar esa comprensión.

Y es increíblemente imprecisa.

Nuestro lenguaje está lleno de insinuaciones, idiosincrasias, expresiones idiomáticas y ambigüedades. Tenemos narices que corren, y pies que huelen. ¿Cómo puede una *pequeña posibilidad* y *escasas posibilidades* ser lo mismo, pero un *hombre sabio* y un *sabelotodo* ser opuestos? ¿Cómo puede una casa *quemar* como *incendiar*? ¿Por qué *completamos* un formulario al *rellenarlo*?

Y aún, puede ser increíblemente preciso.

Transmitimos tanto significado, y logramos tanta colaboración en medio de todas las dificultades con el lenguaje. De alguna manera, podemos ver a través de las lagunas, incoherencias y contradicciones, la irregularidad, y la falta de claridad y comprenderse entre sí con mucha precisión.

Esta diferencia entre precisión y exactitud es importante. *Precisión* es la exactitud mecánica o científica que se puede encontrar en un pasaje de texto. Podemos determinar si una palabra específica existe en un pasaje con un grado de precisión superior. *Precisión* es el grado en que un pasaje infiere que otro pasaje podría ser considerado como verdadero por la gente sensata.

¿Qué pasa si, cuando decimos “2 + 2,” nos referimos a la disposición automovilística, como a dos asientos delanteros y dos traseros?

La respuesta a “2 + 2” es precisamente 4. La matemática nos enseña este hecho. También nos enseña que, independientemente de cuántos ceros usted coloque después del decimal para representar mayor precisión, la respuesta siempre será 4. Pero que sucede cuando decimos “2 + 2,” no queremos decir que se tome literalmente como una fórmula matemática, sino como una expresión idiomática para una expresión automovilística, como en *dos asientos delanteros* y *dos asientos traseros*. ¿Qué sucedería si un psicólogo utiliza “2 + 2” para referirse a una familia con *dos padres* (*padre y madre*) y *dos hijos*? En esos otros contextos, la respuesta *cuatro* tal vez no sea una interpretación precisa de lo que estamos intentando transmitir con el lenguaje.

De hecho, para responder con precisión a una pregunta, a menudo se debe considerar el contexto disponible para la pregunta. Sin información probatoria suficiente, es difícil responder con precisión a una pregunta, aun cuando pueda responder con precisión a elementos en la pregunta de forma literal.

Procesamiento de lenguaje natural superficial

Muchos sistemas de lenguaje natural han intentado enfatizar la precisión dentro de los límites de normas específicas bien hechas. Por ejemplo, el análisis de sentimiento busca, a menudo, un conjunto de palabras específicas y sus sinónimos en sitios de medios de comunicación social. Estos sistemas por consiguiente, sin una nueva evaluación del contexto en el que esas palabras están siendo utilizadas, contar el número de veces que estas palabras son co-ubicadas con alguna marca en la misma frase. Por ejemplo, toma la frase, “... pasa por el IBM Donut Store para un café esta mañana, ello fue excelente ...” y después afirma que la colocación del nombre de la marca y el “óptimo” plazo son una indicación de un

sentimiento positivo. Sin embargo, tenga en cuenta que si el resto de la frase es, "..., fue excelente escuchar que un nuevo Fictional Coffee Shop abrirá pronto, de modo que no siento la tentación de comer donuts todas las mañanas" Entonces, el sistema podría pasar por alto el sentimiento de que no se trata de IBM Donut Store. Denominamos a este concepto *procesamiento de lenguaje natural superficial (NLP)* porque, aunque podría ser bastante preciso en su foco más estrecho, no es muy exacto.

No obstante, también es importante comprender que actualmente NLP superficial tiene un importante rol en muchos sistemas. Si su objetivo es crear una evaluación de tendencias de sentimientos estadísticamente sobre grandes cantidades de información, la falta de precisión para cada ejemplo individual, probablemente no sea un problema. Suponiendo que hay aproximadamente tantos falsos positivos como falsos negativos sobre un conjunto de muestra suficientemente grande que se anulan entre sí. Y si la agrupación de recuentos cancelados se mantiene relativamente constante a través de los conjuntos de muestra con el transcurso del tiempo, el resto de los datos no cancelados obtiene la información de la tendencia relevante estadísticamente. Por lo tanto, los costos de procesamientos adicionales que se requieren para precisión extra de cualquier instancia podrían ser injustificados.

El procesamiento de lenguaje natural superficial puede ser bastante preciso dentro de su foco más estrecho, pero no es muy preciso.

Sin embargo, cuando los casos individuales importan, los sistemas diseñados para ser más precisos sin centrarse en altos niveles de exactitud tienden a ser frágiles. Es decir, se realizan bien dentro de los parámetros estrechos de su diseño pretendido, pero no se desempeñan bien cuando los parámetros cambian. Nosotros comparamos estos sistemas al uso en la construcción del ladrillo tendido. Los ladrillos son fuertes y fácil de construir con ellos. Durante décadas y siglos, hemos perfeccionado la técnica de construcción de ladrillos tendidos para que fuese bastante precisa. Hemos sido capaces de construir estructuras relativamente grandes, ornadas y duraderas. Sin embargo, aunque los edificios de ladrillo tienen una fuerte resistencia de carga, no tienen buena resistencia a rupturas. Caen fácilmente durante terremotos y no soportan grandes distribuciones. Y después de un cierto punto, su resistencia de carga fallará también.

Usted puede observar estas mismas limitaciones en algunos productos de consumo de hoy. Por ejemplo, usted podría utilizar su asistente personal de activación de voz favorito y decir, "Encuéntreme una pizza." como contrapartida, obtiene un local listado de los restaurantes de pizza, que es exactamente lo que deseaba. Ahora dice, "No encuentre pizza para mí." Disfrutará nuevamente una lista local de pizzerías. Igualmente, si dice "Encuentre una pizza cerca" o "Encuentre pizza más lejos", se devuelven las mismas listas locales. El punto es que estos sistemas fueron diseñados de acuerdo con un conjunto específico de reglas y se busca para combinaciones específicas de palabras clave para determinar la respuesta a producir. Estos sistemas no saben cómo distinguir entre las cosas para las que no hay reglas. Podrían ser precisos, pero o necesariamente muy exactos.

Procesamiento de lenguaje natural profundo

Para superar las limitaciones de la construcción de ladrillo, se pasó a usar el acero y el hormigón armado para edificios más grandes. Del mismo modo, estamos asistiendo a un cambio en las técnicas de construcción natural, para procesamiento del lenguaje cuando se necesita extrema precisión. Estas técnicas incorporan mucho más contexto a la evaluación de la pregunta. Nos referimos a este concepto como *procesamiento de lenguaje profundo*, lo que a veces es llamado de *Deep Question-Answering (DeepQA)* cuando el problema es sobre responder las preguntas de lenguaje natural.

Estamos viendo un cambio en las técnicas de construcción para el procesamiento de lenguaje natural, cuando se necesita de precisión

IBM Watson es un sistema NLP profundo. Se logra precisión al tratar de evaluar tanto contexto como sea posible. Se pone ese contexto, tanto dentro del pasaje de la pregunta como de la base de conocimiento (llamado de un *corpus*) que está disponible para encontrar respuestas.

Al prepararse para la demostración del concurso, Jeopardy!, Watson se hizo la siguiente pregunta (pista) a partir de la categoría Lincoln Blogs:

“Secretario de Tesoro americano. Chase acaba de presentar esto a mí por tercera vez – imagine qué compañero, de esta vez yo'm lo acepto.”

Primero, observe la abreviación, “Secret.”, lo que se debe considerar *Secretario*. Además note que *Secretario* aquí no significa alguien que toma dictado y administra una agenda. Los términos combinados *Secretario del Tesoro americano* son importantes aquí como un nombre y un rol. Por lo tanto, para responder a esta pregunta, Watson tenía que encontrar un pasaje que implicase la presentación y aceptación de algo entre Chase y Lincoln, Secretarios del Tesoro americanos (la categoría de la pista). Sin embargo, también observe que la categoría no dice “Presidente Lincoln” necesariamente. La respuesta correcta resultó ser “¿Qué es una renuncia?”.

Al describir este ejemplo en una escuela primaria, momentos después de la retransmisión de IBM Watson participando en JEOPARDY!, un estudiante de quinto grado propuso la idea “¿Qué es una solicitud de amistad?” como posible respuesta.

Sin contexto, estaríamos perdidos.

La respuesta de este estudiante es interesante en parte porque dice mucho sobre el grado en que los medios sociales han penetrado profundamente en el tejido de la sociedad de la nueva generación. No obstante, también es educativo debido a que incluso se puede tomar como una respuesta equitativamente razonable a la pista. Pero, sabemos que esta respuesta es inexacta porque tenemos el contexto histórico. Sabemos que Facebook no estaba disponible en el siglo XIX. Fíjese que el contexto es lo que nos ha permitido aumentar la precisión del sistema en la producción de esta respuesta. Sin ese contexto, estaríamos perdidos.

Vale la pena enfatizar el punto e el que nosotros, como seres humanos, tenemos poca dificultad de procesamiento de nuestro lenguaje, aunque nos confundamos en la ocasión. Por lo general, resolvemos el significado de la información que hemos escrito mucho mejor que los computadores lo suelen hacer.

Tenemos una cualidad innata acerca de cómo eliminar la ambigüedad del lenguaje que queremos capturar y aprovechar en los sistemas informáticos. Este concepto ha sido un objetivo clave de la comunidad de la inteligencia artificial en los últimos cuatro decenios. Y, en gran medida, hemos sido capaces de incrementar la precisión del procesamiento del lenguaje. Pero es sólo con Watson que, finalmente, podemos romper el nivel de precisión que se necesita para que los sistemas de información funcionen bien en el mundo real del lenguaje natural en general.

También una fuerza impulsora enorme busca resolver este problema. Estamos pasando por una explosión de producción de datos. El noventa por ciento de todos los datos en el mundo se produjo en los últimos dos años. Se espera que esta tendencia crezca a medida que nos interconectamos y nos instrumentamos más sobre nuestro mundo. Y el 80 por ciento de toda la información en el mundo es la información no estructurada, la cual incluye textos como literatura, informes, artículos, trabajos de investigación, tesis, emails, blogs, tweets, foros, chats y mensajes de texto. Precisamos computadoras que puedan entender este aluvión de información, de modo que podemos sacar más provecho de ella.

IBM Watson entiende el lenguaje

La navegación efectiva a través del flujo actual de información no estructurada requiere una nueva era de la informática que llamamos *sistemas cognitivos*. IBM Watson es un ejemplo de un sistema cognitivo. Se puede separar el lenguaje humano para identificar inferencias entre pasajes de texto con una precisión similar a la humana, a una velocidad y escala que son mucho más rápidos y mucho más rápidos de lo que cualquier persona puede hacerlo por sí misma. Ello puede gestionar un alto nivel de precisión cuando se trata de entender la respuesta correcta a una pregunta.

Sin embargo, Watson no comprende realmente las palabras individuales del lenguaje. Sino que entiende las características del lenguaje que son utilizadas por las personas. A partir de esas características, se puede determinar si un pasaje de texto (que denominamos una *pregunta*) se deduce otro pasaje de texto (que denominamos una *respuesta*), con un alto nivel de precisión bajo nuevas circunstancias.

En la presentación del concurso JEOPARDY!, Watson tuvo que determinar si la cuestión, “Jodie Foster llevó esta casa por su rol en ‘Silencio de los Inocentes’” dedujo la respuesta “Jodie Foster ganó un Oscar por su rol en ‘Silencio de los Inocentes’”. En este caso, *llevar algo para casa* dedujo *ganar un Oscar*, pero no siempre. Algunas veces *llevar algo para casa* infiere un resfriado, comestibles o todo tipo de cosas. Por el contrario, no lleva siempre para casa las cosas que usted gana. Por ejemplo, podría ganar un contrato de trabajo, pero no es algo que lleve para casa.

El contexto es importante. Las restricciones espaciales y temporales importan. Todos estos conceptos se añaden con la finalidad de habilitar un sistema cognitivo para comportarse con las características de los seres humanos. Y, para volver a un punto anterior, un enfoque basado en normativas podría precisar un número infinito de reglas para capturar cada caso que podríamos encontrar en el lenguaje.

La navegación efectiva a través del flujo actual de información no estructurada requiere una nueva era de la informática llamada sistemas cognitivos.

Watson distingue la pregunta y las posibles respuestas en el corpus, y entonces lo examina y el contexto de la declaración en cientos de maneras. Después Watson utiliza el resultado para ganar un grado de confianza en su interpretación de la pregunta y posibles respuestas.

Pero hay que retroceder un poco. ¿Cómo Watson obtiene sus respuestas a las preguntas? A través de los siguientes procesos:

1. Cuando una pregunta se presenta por primera vez a Watson, éste la analiza para extraer las características principales de la misma.
2. Éste genera un conjunto de hipótesis buscando por todo el corpus los pasajes que tienen alguna posibilidad de contener una respuesta de valor cierto.
3. Realiza una comparación profunda del idioma de la pregunta y de cada posible respuesta mediante la utilización de algoritmos de varios razonamiento.

Esta etapa es desafiante. Existen cientos de algoritmos de razonamiento, cada uno de los cuales hace una comparación distinta. Por ejemplo, algunos observan la coincidencia de términos y sinónimos, algunos analizan las características espaciales y temporales, y otros observan las fuentes relevantes de la información contextual.

4. Cada algoritmo de razonamiento produce una o más puntuaciones, indicando hasta qué punto la respuesta potencial se infiere por medio de la pregunta basada en el sector de enfoque específico de ese algoritmo.
5. Cada puntuación resultante es entonces ponderada contra un modelo estadístico que captura lo bien que el algoritmo hizo en establecer las inferencias entre dos pasajes similares para ese dominio durante el período de “capacitación” de Watson. Ese modelo

estadístico puede ser utilizado a continuación para resumir un nivel de confianza que Watson tiene acerca de la evidencia que se deduce de la respuesta candidata por medio de la pregunta.

- Watson repite este proceso para cada una de las respuestas candidatas hasta que pueda encontrar las respuestas que surgen como siendo las más fuertes.

Figura 1 ilustra cómo Watson obtiene el resultado de una respuesta para una pregunta.

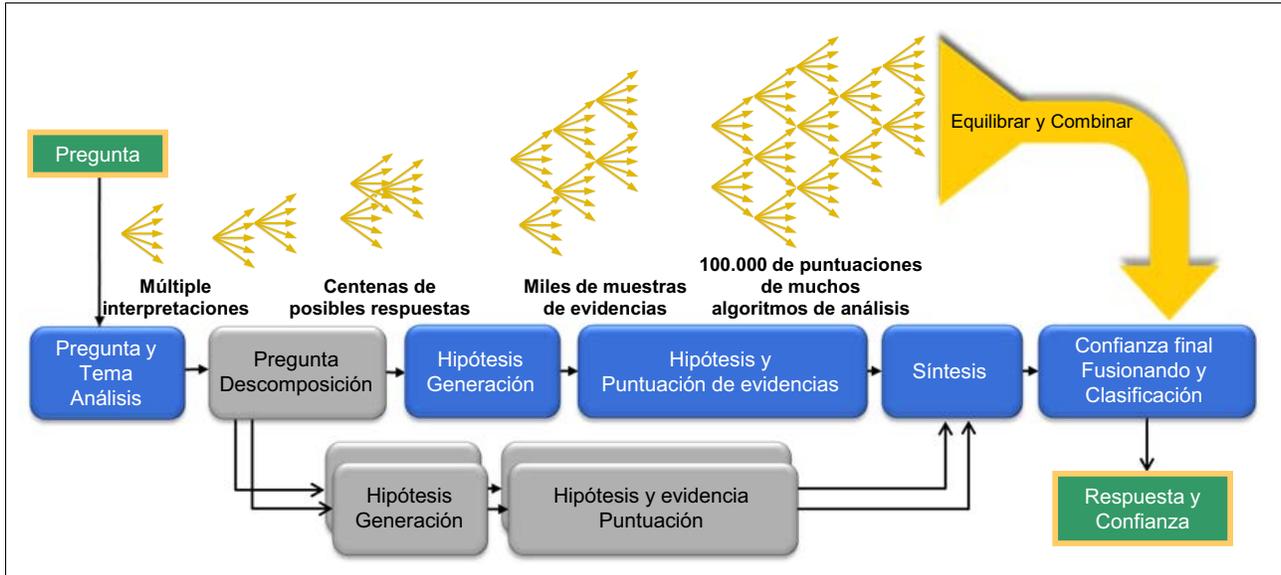


Figura 1 Cómo Watson obtiene como resultado una respuesta para una pregunta

De la importancia primordial a la operación de Watson hay un *corpus de conocimiento*. Este corpus comprende todos los tipos de conocimiento no estructurados, como libros de texto, guías, manuales prácticos, preguntas frecuentes, planes de beneficios y noticias. Watson absorbe el corpus, pasando a través de todo el cuerpo de contenido para obtenerlo de una manera con la que es más fácil trabajar. El proceso de asimilación también cuida del contenido. Es decir, se centra en el corpus, en si incluye el contenido adecuado, revisando los artículos o páginas que están fuera de fecha, que son irrelevantes, o que provengan potencialmente de fuentes poco fiable.

Algunos de los algoritmos de razonamiento se centran en características espaciales y temporales del pasaje, lo que es crítico para desambiguar una tremenda cantidad de lo que los humanos dicen y escriben. Cuando decimos, "Encuentre pizza," se da por hecho que nos referimos a algo cercano. Pero lo que es cercano también es relativo. En otros casos, las relaciones espaciales se muestran relativa a los marcadores geográficos, por ejemplo, un barrio de una ciudad o un estado de un país. Del mismo modo, las características temporales también están presentes en el contexto de la mayor parte de lo que escribimos. Cuando decimos, "Compre queso en el almacén del camino a casa," se infiere un marco temporal. Presumiblemente, el escritor y el destinatario poseen un entendimiento contextual de cuando estén en el camino para casa.

Debe ser hecha una evaluación espacial y temporal tanto en la pregunta como en la respuesta candidata.

La declaración, "En mayo de 1898, Portugal celebró el aniversario de 400 años de la llegada de este explorador a India," demuestra tanto las dimensiones espaciales como las temporales. La celebración se llevó a cabo en Portugal, pero el evento que se celebraba fue la llegada del explorador a India. ¿Sugiere la declaración que el explorador fue de Portugal a la India? ¿Estuvo él alguna vez en Portugal? Observe que la celebración se llevó a cabo en

1898, pero el evento ocurrió 400 años antes. De modo que el evento sucedió realmente en 1498. El pasaje que brindó la respuesta a la pregunta mencionada anteriormente, “El 27 de mayo de 1498, Vasco da Gama desembarcó en Kappad Beach.” La evaluación espacial y temporal se tuvo que desempeñar tanto en la pregunta como en los pasajes de las respuestas candidatas.

El contexto se deriva a partir tanto de la información inmediata como del conocimiento que se encuentra disponible de manera más amplia. Watson puede obtener la información inmediata a partir del título de un documento, otros pasajes en un documento o la base de datos de la fuente a partir de donde se originó. El contexto también puede venir de manera más amplia a partir de una historia compartida. Recuerde que sabíamos eso “¿Qué es una Solicitud de Amistad?” probablemente fue una respuesta incorrecta a la pista en Lincoln Blogs. El motivo se debe a que compartimos un contexto histórico común, que nos cuenta cuándo ciertas cosas sucedieron con relación el uno al otro. Sabemos que Facebook fue creado hace muy poco, pero sabemos que Abraham Lincoln vivió hace unos 150 años, mucho antes que Facebook se volviese popular. El contexto y el raciocinio nos ayudan a crear una base cognitiva para el procesamiento del lenguaje.

La comprensión del lenguaje es sólo el principio

Definimos los sistemas cognitivos como la aplicación de las características parecidas a las humanas para transmitir y manipular las ideas. cuando se combina con las cualidades inherentes de la computación digital, pueden ayudar a resolver problemas con mayor precisión, más resiliencia, y sobre una escala masiva a través de grandes masas de información.

Podemos descomponer un sistema cognitivo como teniendo varios elementos clave (Figura 2). Las cajas medianas sombreadas indican las capacidades actuales de sistemas cognitivos. Las cajas sombreadas más claras indican las capacidades futuras de los sistemas cognitivos.

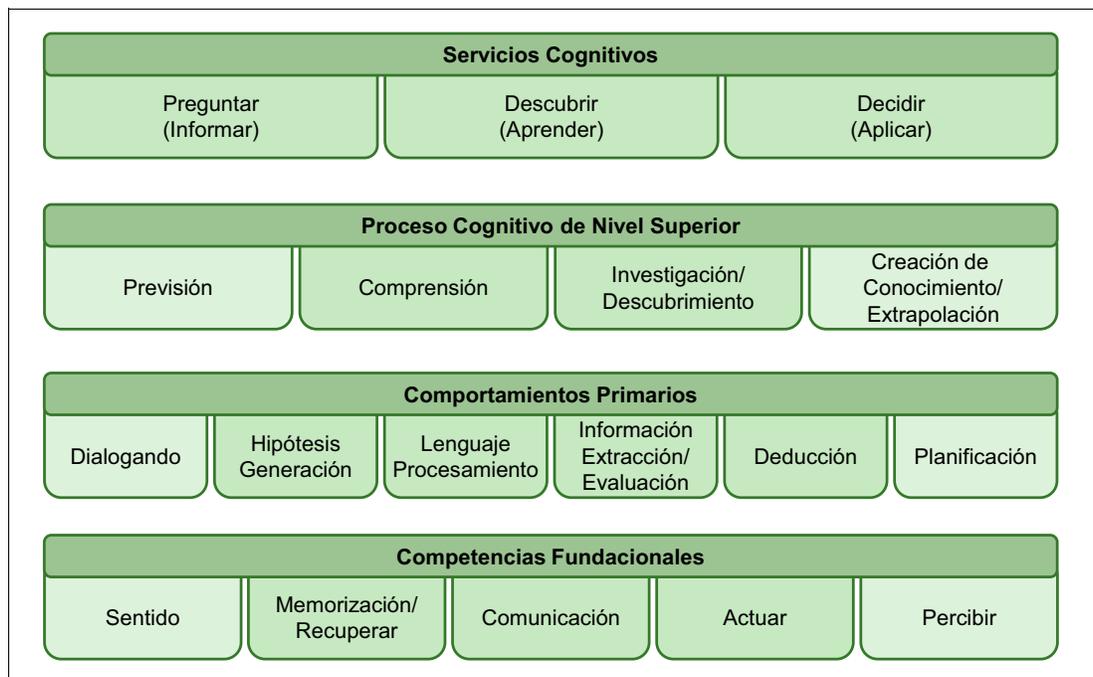


Figura 2 Elementos de un sistema cognitivo

Similar a los humanos, los sistemas cognitivos tiene una manera de reunir, memorizar y rememorar la información, que es el equivalente de la memoria humana. Los sistemas cognitivos también tienen una capacidad básica de comunicarse y actuar. Estas capacidades

se organizan por medio de ciertas construcciones de comportamiento, como los ejemplos a continuación:

- ▶ La capacidad de crear y probar las hipótesis
- ▶ La capacidad de separar y crear inferencias acerca del idioma
- ▶ La capacidad de extraer y evaluar la información útil (como fechas, lugares y valores)

Estas técnicas son fundamentales, sin las que ni computadores ni humanos podrían determinar la correlación correcta entre preguntas y respuestas.

Los procesos cognitivos de orden superior pueden aprovechar los comportamientos fundamentales para lograr un nivel de comprensión. Para comprender algo que nos exige ser capaz de dividirlo, hasta elementos más finos que se comportan de maneras bien ordenadas dentro de una escala determinada. Así como las cosas que funcionan en física a escalas humanas no son como las cosas que funcionan a escalas cósmica o subatómica. Asimismo sistemas cognitivos se diseñan para trabajar a escalas humanas, aunque por colecciones enormes de seres humanos. Como tal, la comprensión del lenguaje comienza con la comprensión de las reglas más adecuadas del idioma, no sólo la gramática formal, pero las convenciones gramaticales informales de uso diario.

Al igual que los humanos, los sistemas cognitivos son impulsados a comprender las cosas por la descomposición de las expresiones de una idea y luego combinar eso con el contexto.

Sin embargo, como lo hacen los humanos, los sistemas cognitivos son conducidos para entender conceptos por descomposición de las expresiones de una idea y luego combinar los resultados con el contexto y la probabilidad de que ciertos términos en el pasaje se utilizan de una manera determinada. Y, como con los seres humanos, nuestra confianza es proporcional a la evidencia que respalda esas probabilidades y el número de algoritmos de raciocinio que tenemos disponible para probar nuestras hipótesis.

Después de establecer un cierto nivel de comprensión, la descomposición del problema en comparación con su probable intención, los sistemas cognitivos puede recomponer los elementos de diversas formas, cada uno de los cuales puede ser probado para imaginar nuevos conceptos. Estas combinaciones se pueden utilizar para llevar a un nuevo descubrimiento y conocimiento, que nos ayudan a encontrar respuestas a las preguntas y hacer realidad las preguntas que nunca pensamos formular.

A continuación, podemos utilizar estas capacidades para solucionar los problemas que se ajustan a ciertos patrones comunes. Podemos hacer preguntas que generan respuestas. Podemos utilizar el sistema para descubrir nuevo conocimiento y desarrollar conceptos que no reconocimos anteriormente. Y podemos usar estos sistemas para apoyar decisiones acertadas o por lo menos ayudar a las personas en las decisiones que que necesitan tomar.

A medida que los sistemas cognitivos crecen copiosamente, esperan obtener la capacidad de percibir.

A medida que los sistemas cognitivos crecen copiosamente, se espera obtener la capacidad de percibir. Esperamos hacer más que sólo leer el texto, sino ver, oír y sentir, para tener un conocimiento básico del entorno. Y esperamos que estos sistemas sean capaces de percibir información, como de reconocer las formas y condiciones cambiantes que además informará su contexto y la capacidad de inferir y razonar. También esperamos que se adopte comportamientos de orden superior y procesos cognitivos, como para tener un diálogo, planificar las diferentes estrategias para la solución de problemas, y obtener previsión y extrapolarla en nuevos conocimientos.

En resumen, los sistemas cognitivos asimilarán muchos de los comportamientos que a los humanos les parece “natural,” y los aplicarán en escala masiva para ayudar a las personas a solucionar los problemas que a menudo, en la actualidad, se encuentran fuera de su alcance.

Estamos comenzando una nueva era. En esta época, las computadoras van más allá de sólo la realización de tareas rutinarias de procedimiento más eficiente, para emplear la cognición humana y hacer a la gente más inteligente acerca de lo que hace.

Los problemas llegan de distintas maneras

A medida que avanzamos con IBM Watson, ¡le descubrimos otras utilidades. En la literatura clásica “Pregunta a Watson,” un usuario formula a Watson una pregunta (o provee una pista, un registro de paciente etc.), de lo que Watson obtiene una respuesta. Watson confía en que la pregunta infiere la respuesta y la evidencia que apoya la respuesta. Watson ha encontrado un hogar en campos como el Diagnóstico Oncológico, Gestión de Utilización (es decir, una aprobación previa de la cobertura del seguro para procedimientos médicos programados), Análisis de Crédito y la investigación básica.

Watson ayuda siempre que un profesional necesita asistencia para obtener la información más relevante a su problema de espacio.

Al formular otras preguntas más importantes, usted puede pensar acerca de sus problemas empresariales de una manera totalmente nueva.

Una de las mayores revelaciones sobre Watson es que, mediante la utilización de Watson para ayudar a responder preguntas, usted puede darse cuenta de que fundamentalmente está formulando las preguntas equivocadas. Cuando Watson responde a sus preguntas, incluso respondiendo correctamente, usted puede darse cuenta de que necesita formular otras cuestiones, mejores y más importante para ayudar a considerar su problema empresarial de una manera totalmente nueva. Usted comienza a pensar de una manera que ayudará a entender las amenazas y las oportunidades competitivas en su mercado, lo que que nunca se le había ocurrido antes.

Estas aplicaciones de tipo descubrimiento se mejoran aun más con el trabajo que IBM está haciendo ahora en los laboratorios de Investigación y Desarrollo de Software de IBM. Los recientes avances en el encadenamiento de inferencia (determinando que *esto* que deduce *que*, lo que infiere algo más etc.) están creando un conocimiento más profundo. Sabiendo que la diabetes causa hiperglucemia es importante. No obstante, dando el siguiente paso para deducir que entre la hiperglucemia que causa ceguera es más importante cuidar del paciente en su totalidad. Estos tipos de inferencias multinivel pueden ser capturados como un gráfico de inferencia desde el que se puede observar un amplio espectro de consideraciones ulteriores. Más importante aún, la convergencia en el gráfico es una forma poderosa de obtener más inferencias significativas, tales como respuestas que pueden revelar un conocimiento más profundo y consecuencias ocultas. Al agrupar los valores de confianza precedentes, se pueden agregar y establecer una mayor confianza en una respuesta como la respuesta preferida a la pregunta.

Podemos producir inferencias invertidas, en efecto, descubriendo preguntas a las respuestas que nunca fueron hechas.

Además, podemos producir inferencias inversas, que en la práctica significa que descubrimos las preguntas a las respuestas que nunca se habían hecho. Determinando que un paciente que tiene una historia de temblores en reposo y una “rostró inexpressivo” podría deducirse que tiene la enfermedad de Parkinson’. Sin embargo, determinar que el paciente aun tiene dificultad para caminar podría revelar también el daño al sistema nervioso de la sustancia negra, que podría haber sido perdido sin la indicación de las preguntas no formuladas previamente.

IBM está invirtiendo en este tipo de mejoras importantes en Watson que creemos que van a dar lugar a nuevos avances en materia de salud, finanzas, centros de contacto, gobiernos,

industria química y un planeta más inteligentes. Estos tipos de avances pueden ayudar a lanzarnos hacia una era de sistemas cognitivos.

En muchas soluciones, Watson está siendo aprovechado en otras formas de computación más tradicionales, tales como análisis estadísticos, normas y procesamiento de negocios, colaboración y presentación de informes para resolver problemas empresariales. Por ejemplo, considerar la idea de que IBM combine otros análisis estadísticos con la habilidad de Watson de responder a preguntas acerca de posibles eventos que puedan indicar un riesgo para una inversión. IBM puede ayudar a mejorar sus procesos de riesgo y valorización para las instituciones financieras. Del mismo modo, la visión que obtenemos acerca de las respuestas de los clientes a través de PNL profundo que puede sugerir cambios en el comportamiento de compra y utilización que de otro modo podrían no ser evidentes en los datos estructurados. En la industria de la salud, Watson está siendo utilizado para ayudar a las compañías de seguros en los procesos de pre-aprobación de tratamientos como parte de los procesos de Gestión de Utilización.

La precisión se aprueba a través de la generalización

A medida que IBM continúa evolucionando y desarrollar estos tipos de sistemas cognitivos, debemos tener cuidado. Nos encontramos en un momento clásico que los seres humanos enfrentan todo el tiempo, que se trata sobre especializarse o generalizar. Podemos especializarnos en tecnología NLP para un dominio específico, concentrándonos, por ejemplo, solamente en las características lingüísticas de ese dominio. Este enfoque es tentador e incluso podría ser necesario, en las primeras fases de evolución, asegurar la viabilidad de la tecnología. De cualquier modo, es probable que este enfoque nos lleve de regreso a la época de la construcción con ladrillos. Si la capacidad de adaptarse con la destreza de los seres humanos hace que los sistemas cognitivos sean especiales, debemos generalizar. Tenemos que reconocer y aprovechar las inferencias a partir de un conjunto más amplio de variación lingüística, bajo un amplio conjunto de circunstancias, como cambios de nuestros conocimientos, como los cambios de contexto, y como el cambio lingüístico contemporáneo.

Al utilizar este enfoque, podemos adaptarnos más fácilmente a los nuevos y mayores problemas. Ya estamos aplicando a Watson a la industria de servicios de la salud y financieros, que tiene los siguientes beneficios:

- ▶ Reúne las ventajas de Watson a diferentes dominios con problemas de alta importancia.
- ▶ Desarrolla los algoritmos de procesamiento de lenguaje de Watson para manejar un conjunto más amplio de variación lingüística.

Este enfoque permite una fácil adaptación a otros dominios y mejora la utilidad de Watson para nuestras aplicaciones de dominio ya existentes.

Nos encontramos en un momento clásico, ya sea para especializarse como sacar conclusiones generales.

Las aplicaciones de cuidado de la salud son interesantes ya que a menudo necesitan de precisión y exactitud. Se necesita precisión para interpretar adecuadamente el texto en una descripción de la salud de un paciente para inferir la condición del mismo. Sin embargo, la directriz de National Comprehensive Cancer Network (NCCN) para cáncer de mama se debe justificar por medio de la presencia de los términos precisos en el registro de la salud del paciente. Entonces se necesita más precisión para encontrar la evidencia que soporte el tratamiento.

Estamos en el comienzo de una nueva era de la informática, que es menos precisa, pero mucho más exacta.

Cada vez que nos encontramos con una anomalía lingüística (algo en el lenguaje que nunca hemos visto antes), se tomará una decisión acerca de si el problema es único para el dominio o es característica de un conjunto más amplio de problemas lingüísticos. Siempre que sea posible, volvemos a nuestro algoritmo principal para determinar si se puede generalizar el algoritmo con el fin de reconocer y evaluar la nueva situación. Al igual que con los seres humanos, mediante la utilización de este enfoque, podemos trazar nuestro conocimiento a las nuevas experiencias, y por lo tanto, incrementar la base contextual para el sistema.

El resultado esperado es que se incrementa la precisión, el alcance y la escala:

- ▶ La precisión de la inferencia lingüística (obtener la respuesta correcta, por la razón correcta, en el momento oportuno)
- ▶ Ámbito del problema de espacio
- ▶ Escalando las cantidades masivas de datos y preguntas, en mucho más dominios

Esperamos observar aun más valor en las próximas soluciones para “mejores medidas”, análisis de opiniones sociales, refinamiento de petróleo y químicos, y muchas otras aplicaciones. Estamos justo en el comienzo de una importante nueva era de la informática que está menos centrada en la precisión, pero mucho más exacta. Es una era de aplicación de comportamiento similar al del ser humanos para problemas de computación a gran escala. Es la era de los sistemas cognitivos.

Otros recursos para obtener más información

Para obtener más información acerca de Watson, consulte el website IBM Watson en:
ibm.com/innovation/us/watson/index.shtml

El autor que escribió este manual

Este manual fue escrito por Rob High en colaboración con Organización de Soporte Técnico Internacional (ITSO).



Rob High es un IBM Fellow, Vicepresidente y Chief Technology Officer, IBM Watson Solutions, IBM Software Group. Él tiene la responsabilidad general de conducir la estrategia técnica y liderazgo de pensamiento de IBM Watson Soluciones. Como un asociado clave del equipo de liderazgo de IBM Watson Solutions, Rob trabaja en colaboración con los equipos de ingeniería, investigación y desarrollo de Watson. Rob tiene 37 años de experiencia en programación y ha trabajado con supervisores de transacción con base en los componentes por los últimos 21 años, incluso SOMObject Server, Component Broker y el más reciente IBM WebSphere® Application Server. Anteriormente Rob había trabajado como Arquitecto Jefe para la fundación WebSphere con la responsabilidad de los arquitectos de WebSphere Application Server y los productos relacionados que se integran en aquella base de tiempo de ejecución.

¡Ahora usted puede convertirse en un autor publicado, también!

Aquí hay una oportunidad de resaltar sus habilidades, crecer en su carrera, y convertirse en un autor publicado—¡todo al mismo tiempo! Únase a un proyecto de residencia de ITSO y ayude a escribir un libro en su campo de conocimiento, mientras perfecciona su experiencia utilizando las tecnologías de última generación. Sus esfuerzos ayudarán a incrementar la aceptación del producto y la satisfacción del cliente, así como también a expandir su red de contactos técnicos y relaciones. Las residencias duran desde dos a seis semanas, y usted puede participar tanto en persona como un residente remoto trabajando desde su domicilio.

Para obtener más información acerca del programa de residencia, navegue por el índice de residencias, y solicite online en:

ibm.com/redbooks/residencies.html

Permanezca conectado a IBM Redbooks

- ▶ Síguenos en Facebook:
<http://www.facebook.com/IBMRedbooks>
- ▶ Síguenos en Twitter:
<http://twitter.com/ibmredbooks>
- ▶ Búsquenos en LinkedIn:
<http://www.linkedin.com/groups?home=&gid=2130806>
- ▶ Explore las nuevas publicaciones de IBM Redbooks®, residencias y cursos prácticos con boletín informativo semanal de IBM Redbooks:
<https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm>
- ▶ Manténgase al tanto de las publicaciones de Redbooks con RSS Feeds:
<http://www.redbooks.ibm.com/rss.html>

Avisos

Esta información se ha desarrollado para productos y servicios ofrecidos en los EE.UU..

IBM puede no ofrecer los productos, servicios o dispositivos tratados en el presente documento en otros países. Consulte a su representante IBM local, para información adicional sobre los productos y servicios disponibles en su área. Cualquier referencia a un producto, servicio o programa IBM, no pretende declarar ni implica que solo puedan utilizarse productos, servicios o programas de IBM. En su lugar, puede utilizarse cualquier producto, servicio o programa funcionalmente equivalente que no infrinja cualquier derecho de propiedad intelectual de IBM. Sin embargo, es de responsabilidad del usuario's evaluar y verificar la operación de cualquier producto programa o servicio que no sea de IBM.

IBM puede tener patentes o solicitudes de patentes pendientes de aplicaciones que tratan los asuntos descritos en el presente documento. La posesión de este documento no le otorga ninguna licencia sobre estas patentes. Puede enviar consultas sobre licencias, por escrito, a:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

El siguiente párrafo no se aplica al Reino Unido u otros países donde dichas disposiciones sean incompatibles con la legislación local: INTERNATIONAL BUSINESS MACHINES CORPORATION SUMINISTRA LA PRESENTE PUBLICACIÓN "COMO ESTÁ" SIN GARANTÍA DE NINGUNA CLASE, EXPRESA O IMPLÍCITA, INCLUYENDO, PERO SIN LIMITACIÓN, LAS GARANTÍAS IMPLÍCITAS DE NO VULNERACIÓN, COMERCIALIZACIÓN Y ADECUACIÓN PARA UN PROPÓSITO PARTICULAR. Algunos Estados no permiten la exclusión de garantías expresas o implícitas en ciertas transacciones, por lo tanto, esta declaración puede no aplicarse a su caso.

Esta información puede incluir imprecisiones técnicas o errores tipográficos. Periódicamente se hacen cambios a la presente información; dichos cambios se incorporarán en nuevas ediciones de la publicación. IBM puede realizar mejoras y/o cambios en el producto(s) y/o el programa(s) descrito en esta publicación a cualquier momento, sin previo aviso.

Cualquier referencia en esta información a sitios web no IBM se proporcionan únicamente para su comodidad y de ninguna manera constituyen un aval de dichos sitios web. Los materiales de estos websites no forman parte de los materiales para este producto de IBM y el uso de esos Websites se encuentra bajo su propio riesgo.

IBM puede utilizar o distribuir cualquier información que suministre de cualquier modo que considere adecuado sin incurrir en ninguna obligación para usted.

Los datos de rendimiento contenidos aquí se han determinado en un entorno controlado. Por lo tanto, los resultados obtenidos en entornos operativos diferentes pueden variar significativamente. Algunas mediciones pueden haberse realizado en sistemas en desarrollo y no existe ninguna garantía de que dichas mediciones serán las mismas en sistemas generalmente disponibles. Además, algunas medidas pueden haber sido estimadas a través de la extrapolación. Los resultados actuales pueden variar. Los usuarios del presente documento deben verificar los datos aplicables a sus entornos particulares.

La información sobre productos no IBM se ha obtenido de los proveedores de dichos productos, de sus anuncios publicados o de otras fuentes públicamente disponibles. IBM no ha probado dichos productos y no puede confirmar la exactitud de rendimiento, compatibilidad u otras afirmaciones relacionadas a productos no IBM. Preguntas sobre las capacidades de los productos no IBM deben dirigirse a los proveedores de dichos productos.

La presente información contiene ejemplos de datos e informes utilizados en las operaciones de negocio diarias. Para ilustrarlos de la manera más completa posible, los ejemplos incluyen nombres de personas, empresas, marcas y productos. Dichos nombres son ficticios y cualquier semejanza con los nombres y las direcciones utilizadas por una empresa real es pura coincidencia.

LICENCIA DE COPYRIGHT:

La presente información contiene programas de aplicación de muestra en el lenguaje de origen, que ilustran las técnicas de programación en diferentes plataformas operativas. Los programas de ejemplo se pueden copiar, modificar y distribuir en cualquier forma sin ningún pago a IBM, para fines de desarrollo, utilización, marketing o distribución de programas de aplicación compatibles con la interfaz de programación de aplicaciones de la plataforma operativa para la cual los programas de ejemplo están escritos. Estos ejemplos no han sido completamente probados bajo todas las condiciones. Por lo tanto, IBM no puede garantizar ni hacer cualquier afirmación sobre la confiabilidad, capacidad de servicio o función de dichos programas.

Este documento, REDP-4955-00, fue creado o actualizado en April 9, 2013.



Marcas registradas

IBM, el logotipo IBM e ibm.com son marcas comerciales o registradas de International Business Machines Corporation en los Estados Unidos, otros países o ambos. Estos y otros términos registrados de IBM están marcados en su primera aparición en esta información con el adecuado símbolo (® o ™), indicando el registro en los EE.UU. o marca registrada de derecho consuetudinario, de propiedad de IBM en el momento que se publicó dicha información. Dichas marcas registradas también pueden ser marcas registradas o marcas registradas de derecho consuetudinario en otros países. Una lista actual de marcas registradas de IBM está disponible en la Web en ibm.com/legal/copytrade.shtml



Los siguientes términos son marcas registradas de International Business Machines Corporation en los Estados Unidos, otros países o ambos:

IBM Watson™
IBM®

Redbooks®
Redguide™

Redbooks (logotipo) ®
WebSphere®

Los siguientes términos son marcas registradas de otras compañías:

¡PELIGRO! es una marca registrada de Jeopardy Productions, Inc. Todos los derechos reservados.

Los nombres de otra compañía, producto o servicio pueden ser marcas registradas o marcas de servicios de terceros.