Inteligencia Artificial: desarmar la caja negra de los chatbots

Presentación

La Inteligencia Artificial está entre nosotros: una horda de aplicaciones y herramientas, bots y asistentes. Cada semana, cada día, aparecen nuevas, en una sucesión acelerada que parece no tener fin. Incluso aunque quisiéramos (¡qué insensatez!) ya es imposible escapar.

La usamos sin mucha duda, a veces sin saberlo. Los sabemos usar, en el mejor de los casos, pero difícilmente sepamos cómo funcionan. No hay que flagelarse, así es la cosa. Los algoritmos de Inteligencia Artificial se caracterizan por ser, en su vasta mayoría, **cajas negras**. Les damos un "input", como la ubicación de un lugar al que queremos ir desde donde estamos, o la primera intervención en una conversación con un chatbot, y nos devuelve un "output", la ruta más rápida, o una respuesta a nuestro texto. Pero es muy difícil saber qué pasó en el medio. Esto no es solamente un problema de no conocer el código, ni de cuestiones de propiedad intelectual de las herramientas: es la naturaleza de los algoritmos de aprendizaje automático, que hoy dominan la Inteligencia Artificial.

En este curso, vamos a desarmar una de las cajas negras más usadas y relevantes de estos tiempos: los chatbots. Vamos a entender los mecanismos internos que hacen que estos agentes inteligentes funcionen y los riesgos que presentan. Empezando por el concepto de aprendizaje automático supervisado y siguiendo el camino de las conexiones de las redes neuronales artificiales, vamos a entender cómo un software puede procesar el texto en lenguaje natural (¿no era que eran todos unos y ceros?) y conocer las tecnologías que permitieron la aparición de chatbots más modernos que hoy se usan para otras tareas. Finalmente, juntaremos de nuevo todas las piezas para crear nuestro chatbot personal.

Clase 1. Enseñar con el ejemplo. Aprendizaje automático supervisado

¿Cómo hacemos para que una computadora haga lo que queremos? Fácil, la programamos. Esto suele interpretarse como la acción de indicarle la serie de pasos a seguir para llegar a un objetivo. No está mal, y funciona para un gran abanico de tareas. Así, podemos programar una computadora para que sume dos números, multiplique matrices, o haga cosas increíblemente complejas. Pero si queremos una computadora que se fije el clima que va a hacer durante el día, mientras nos propone recetas para las viandas de los chicos a la mañana (suspira), a la vez que convierte nuestros dictados en mails escritos para distintos contactos (tal vez en idiomas diferentes), sugiriendo de paso, como si no le costara nada, cambios y correcciones de estilo, puede ser que estemos frente a un problema demasiado grande para expresar paso por paso. La Inteligencia Artificial, y en particular el *machine learning* (aprendizaje automático) cambian el paradigma de programación, y pasan a programar computadoras a través de ejemplos; o dicho de otra manera, a partir de datos.

Así, son capaces de generar algoritmos que resuelven tareas de un nivel de complejidad impensable en el formato de programación tradicional.

En esta primera clase, vamos a describir el rol del *machine learning* en el contexto más grande de la inteligencia artificial, y a entender cómo funciona este método aparentemente mágico de programar. Vamos a describir sus ventajas y revelar sus peligros. Con ejemplos prácticos, vamos a entender el funcionamiento de los algoritmos de *machine learning* más atractivos y versátiles, las redes neuronales artificiales.

Contenidos:
Inteligencia artificial como un campo interdisciplinario amplio.
Aprendizaje automático. Las tareas fundamentales de ML.
Aprendizaje supervisado. Clasificación y regresión.
Conjunto de entrenamiento y conjunto de evaluación.
Redes neuronales.
Sobreajuste.
Ejemplo de visión de computadora.

Clase 2. Datos sin fin. Procesamiento de datos secuenciales.

Hay gente que cuando empieza a hablar, no para. Si quisiéramos registrar ese discurso, y de alguna forma trabajarlo con un algoritmo de IA, nos enfrentaríamos con un problema diferente al de la clase pasada. ¿Cómo podemos trabajar con datos que no tienen una extensión predefinida? Datos que, para colmo, tienen una dependencia temporal (no es lo mismo decir "comer para vivir" que "vivir para comer"). Este problema está en la base de una rama de la Inteligencia Artificial que se llama procesamiento del lenguaje natural, y que busca dotar a las computadoras de la habilidad para, justamente, entender el lenguaje.

En esta clase, vamos a sumergirnos en el Procesamiento del Lenguaje Natural y estudiar las distintas tareas que podemos resolver con una computadora. Vamos a convertir las palabras en objetos más comprensibles para nuestras amigas las máquinas y veremos cómo al hacerlo de manera astuta estamos ya recogiendo parte de la comprensión con la que queremos inducir a una máquina. Pero también generando sesgos que se empiezan a colar en los algoritmos. Finalmente, vamos a entender cómo reconvertir nuestras redes neuronales de la última clase para que trabajen con datos secuenciales. Para terminar, vamos a componer un soneto.

Contenidos:
Datos secuenciales. Series temporales y texto. Problemas vec2vec, seq2vec, seq2seq
Procesamiento del lenguaje natural. Traducción / Análisis de sentimientos / etc.
Tokenización y embeddings (word2vec)
Redes neuronales recurrentes (RNN). LSTM.
Ejemplo de generación de texto shakespeariano.
Ejemplo de Análisis de sentimientos
Problemas de memoria y de tratamiento de series largas

Clase 3. "Transfórmense y avancen". La revolución de los Transformers.

Hasta 2017, las IA que se usaban para procesar lenguaje eran como el protagonista de la película *Memento*, de Christopher Nolan. Podías tener un intercambio totalmente razonable, siempre y cuando la charla no se extendiera mucho. Después de eso, se empezaba a perder coherencia ("*I take it I've told you about my condition*"). Así, una red recurrente o LSTM, como las que vimos en la clase anterior, no podía "leer" frases muy largas y ni pensar en trabajar con documentos enteros. Pero esto cambió con la aparición de los mecanismos de atención y de la arquitectura basada exclusivamente en este método: los *Transformers*. Comenzó así una revolución en el campo del procesamiento del lenguaje natural y se allanó el terreno para los chatbots y aplicaciones que tenemos hoy en día.

En esta clase, vamos a presentar estas nuevas formas de procesar la información, vamos a entender cómo logran estos algoritmos conservar la memoria de lo que leen (la respuesta te sorprenderá). Además, presentaremos los grandes modelos de Lenguaje (LLMs) basados en esta tecnología que están en la base de los chatbots.

Contenidos:
El mecanismo de atención para aumentar la memoria de las RNN.
Attention is all you need. La revolución de los Transformers.
Grandes modelos de lenguaje. GPT, Llama.
Ejemplos de multitareas con LLMs.

Clase 4. La caja negra se abre.

Abrimos la caja negra de un chatbot. Empezamos a sacar pedazos y encontramos que lo más grande e importante que hay en la caja son los Transformers. Pero no solo está eso. Hay una serie de interfaces y capas que hacen que el chatbot se comporte… bueno, como un chatbot. Su comportamiento de asistente amistoso y gentil, siempre listo para darnos una mano, viene dado por una serie de especializaciones y capas adicionales que tenemos que agregarle a los grandes modelos de lenguaje que vimos la semana pasada.

En esta clase, vamos a estudiar qué es lo que hace de un chatbot un chatbot. Vamos a comprender qué flexibilidad tienen estas interfaces y repasar sus peligros. Además, vamos a presentar una metodología reciente para hacer que las respuestas dadas por los chatbots sean más fiables y ajustadas al contenido de un conjunto definido de documentos.

Al finalizar, deberíamos ser capaces de volver a poner todas las piezas en su lugar, cerrar la caja negra, darle al interruptor de encendido y tener nuestro propio chatbot.

Contenidos:
Componentes de un chatbot
Modelos pre-entrenados y personalización
Peligros y limitaciones
Mejorando la fiabilidad de las respuestas
Construyendo tu propio chatbot