



Universidad Nacional Autónoma de México

Facultad de Ciencias

Inteligencia Artificial | 7003

Examen Parcial 1 | Introducción y Agentes

Sosa Romo Juan Mario | 320051926

24/02/24



1. Explica brevemente qué es el problema del significado (The Problem of Meaning) en la inteligencia artificial. Cita tus fuentes (1 pt.).

La siguiente explicación se basa mayoritariamente en el artículo de Froese y Taguchi, 2019. La manera más sencilla para mí de entenderlo es observar la diferencia entre lo que representa una operación como $x=2+2$ para nosotros y para las computadoras (sistemas inteligentes). De entrada, si no sabemos álgebra, este conjunto de símbolos no significa nada; es decir, solo son letras con números relacionados por un símbolo raro. Usualmente, nos enseñan cómo funcionan este tipo de operaciones con ejemplos de la vida real. Por ejemplo, la operación $x=2+2$ significa que si tienes 2 manzanas y tu amigo tiene otras 2 manzanas, juntos tienen 4 manzanas, o lo que es lo mismo, x vale 4.

Es obvio, entonces, que para nosotros la operación por sí misma básicamente no tiene sentido, o mejor dicho, tiene sentido una vez que entendemos su significado y lo abstraemos usando matemáticas. Ahora, podemos comparar este comportamiento con el de una máquina al ver esta operación. Para la máquina, esta operación nativamente tiene "sentido": $x=2+2$ no es más que una serie de, digamos, tokens que al final del día están asociados con un comportamiento preprogramado que le indica qué hacer. Incluso así, el programa no sabe realmente que está sumando dos números; más bien, sigue una serie de pasos hasta el nivel de los transistores, manipulando su memoria hasta llegar al resultado. Como se menciona en el artículo citado, si lo que se está sumando son manzanas, personas u horas, esto no hace ninguna diferencia para la ejecución del programa. Sin embargo, es obvio que para nosotros sí hace diferencia: no solo podemos tomar decisiones en base a lo que estamos viendo, sino que además podemos preguntarnos "¿qué significa lo que estamos haciendo?".

Esto último es lo que más caracteriza la diferencia entre nosotros y estos sistemas. Al final del día, el comportamiento y la lógica de estos sistemas, por más "inteligentes" que sean, no constituyen un entendimiento fundamental de los sucesos, sino algo más parecido a una tabla extremadamente grande de posibles respuestas y métodos matemáticos para intentar obtener la mejor de ellas. Lo cual, por cierto, no es algo necesariamente malo, ya que cuando se trata de cuestiones no subjetivas o que dependen más de la sintaxis que del significado, estos sistemas suelen ser más eficientes.

2. Explica los tres paradigmas principales de la I.A., así como sus características principales (1 pt.).

■ **Simbólico**

En esencia este enfoque busca conseguir "inteligencia" mediante el uso de símbolos y reglas. Es útil si queremos hacer cosas más determinísticas o más estructuradas.

Es un enfoque más limitado en términos de escalabilidad, pues se tienen que de cierta forma saber de manera previa lo que se quiere que el sistema sepa.

Russell y Norvig, 2020

■ Estadístico

Como su nombre lo dice, este paradigma se apoya de metodos probabilisticos y estadisticos como la regresión para intentar corregir o determinar la incertidumbre, incluyen metodos como los de aprendizaje supervisado, no supervisado y de refuerzo.

En si son bastante utiles para ver patrones en grandes cantidades de datos y son lo que se implementa para modelos economicos o los conocidos algoritmos de redes sociales.

En escencia este enfoque tiene 2 problemas principales, el primero y el mas grande (no solo aplica a este tipo de IA) es la cantidad y calidad de los datos, en si, al ser modelos hechos para procesar muchos datos, conseguirlos y ver que sean validos es complicado. El segundo problema es el de la caja negra, a diferencia del enfoque anterior en donde las cosas son relativamente mas simples de entender y simular, en este paradigma, al no necesitar mostrar o explicar los pasos si no mas bien el resultado y el modelo, se vuelve bastante mas complicado entender su funcionamiento.

Bishop, 2006

■ Neuronal

Finalmente, tenemos el paradigma neuronal, como su nombre lo indica, esta corriente se basa de como funcionan los cerebros organicos, especialmente el de los humanos; en su centro tiene la idea de computar a traves de una red de unidades independientes distribuídas que reciben una serie de señales y sacan otra serie de señales tras procesarlas.

Ademas de las unidades que llaman neuronas, existen capas de entrada, ocultas y de salida, que son agrupaciones de neuronas con comportamientos especificos. Una parte buena de este enfoque es que pierde gran parte de la estructura que los otros 2 necesitan, volviendose mas flexible y permite enfocarse mas en las entradas y salidas, lo que a su vez significa que otra vez tenemos el problema de la caja negra y esta vez aun mas intenso.

Este enfoque es el mas popular actualmente, a mi parecer por su gran escalabilidad, ademas, estos modelos son capaces de tratar con datos no estructurados, haciendolos mucho mejores en tareas de traduccion o de procesamiento de imágenes.

Goodfellow et al., 2016

Finalmente, me gusta agregar que aunque son 3 enfoques diferentes, en la práctica es muy comun utilizar una combinación para aprovechar las ventajas de cada uno y cubrir lo que se puede de los otros.

3. **Explica la diferencia entre I.A. débil e I.A. fuerte. Cita tus fuentes (0.5 pt.).**
4. **Investiga y explica brevemente de qué se trata el juego de la imitación de Alan M. Turing. Cita tus fuentes (0.5 pt.).**
5. **Investiga en qué informe se declara el fracaso del programa de traducción de máquina, y explica brevemente las razones que en él se esbozan. Cita tus fuentes (1 pt.).**
6. **Explica brevemente qué es un agente racional en el contexto de inteligencia artificial. Preferiblemente incluye un diagrama en tu descripción (1 pt.).**
7. **Menciona dos diferencias entre la función de rendimiento y el programa del agente (1 pt.).**

-
8. Considera el mundo de la aspiradora constituido por dos celdas. Si el agente utiliza el siguiente algoritmo para desempeñar sus funciones:

Algoritmo 1 Mundo de la Aspiradora

```
1 función DECIDE(celda, sucia) regresa acción
2   if ((celda = A or celda = B) and sucia = 1) then
3     return Limpia
4   end if
5   if (celda = A and sucia = 0) then
6     return Izquierda
7   end if
8   if (celda = B and sucia = 0) then
9     return Derecha
10  end if
11 end función
```

indica el tipo de agente en el que podría clasificarse (1 pt.).

9. Menciona tres categorías en las que podemos clasificar los entornos de trabajo (0.5 pt.).
10. Describe dos situaciones en las que el agente basado en utilidad tiene ventaja sobre el agente basado en objetivo (0.5 pt.).
11. Explica la noción de aprendizaje para agentes (1 pt.).
12. Para el mundo de un robot de servicio de cafetería, enuncia: sus sensores, sus efectores, su ambiente, su entorno de trabajo. Además, propón una medida de rendimiento para que su agencia sea racional. Además, indica qué tipo de agente sería mejor implementar para su servicio: basado en modelo, que aprende, dirigido por tabla, o reactivo simple (1 pt.).

Referencias

- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Froese, T., & Taguchi, S. (2019). The Problem of Meaning in AI and Robotics: Still with Us after All These Years. *Philosophies*, 4(2), 14-. <https://doi.org/10.3390/philosophies4020014>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th). Pearson.