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 ïnteligentes"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 ïnteligencia"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 tiene 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 indica, este paradigma se apoya en métodos probabilísticos y estadísticos, como la regresión, para intentar corregir o determinar la incertidumbre. Incluye métodos como los de aprendizaje supervisado, no supervisado y de refuerzo.

  En sí, son bastante útiles para detectar patrones en grandes cantidades de datos y se aplican en modelos económicos o en los conocidos algoritmos de redes sociales.

  En esencia, este enfoque presenta dos problemas principales. El primero, y el más grande (no solo para este tipo de IA), es la cantidad y calidad de los datos; al tratarse de modelos diseñados para procesar grandes volúmenes de información, conseguir datos y verificar que sean válidos es complicado. El segundo problema es el de la caja negra: a diferencia del enfoque anterior, en el que los procesos son relativamente más simples de entender y simular, en este paradigma, al no ser necesario mostrar o explicar cada paso sino solo el resultado y el modelo, se vuelve bastante más complejo comprender su funcionamiento.

  Bishop, 2006
- **Neuronal**
  Finalmente, tenemos el paradigma neuronal. Como su nombre lo indica, esta corriente se basa en el funcionamiento de los cerebros orgánicos, especialmente el de los humanos; en su núcleo se plantea la idea de computar a través de una red de unidades independientes distribuidas, que reciben una serie de señales y generan otra serie de señales tras procesarlas.

  Además de las unidades, denominadas neuronas, existen capas de entrada, ocultas y de salida, que son agrupaciones de neuronas con comportamientos específicos. Una ventaja de este enfoque es que pierde gran parte de la estructura rígida que requieren los otros dos, volviéndose más flexible y permitiendo enfocarse en las entradas y salidas. Esto, a su vez, implica que nuevamente surge el problema de la caja negra, y en este caso, aún más marcado.

  Este enfoque es el más popular actualmente, a mi parecer, por su gran escalabilidad; además, estos modelos son capaces de tratar con datos no estructurados, lo que los hace mucho mejores en tareas de traducción o procesamiento de imágenes.

  Goodfellow et al., 2016

Finalmente, me gusta agregar que, aunque son tres enfoques diferentes, en la práctica es muy común utilizar una combinación para aprovechar las ventajas de cada uno y compensar las limitaciones de los otros.

3. **Explica la diferencia entre I.A. débil e I.A. fuerte. Cita tus fuentes (0.5 pt.).**

   De manera concisa, la inteligencia artificial débil abarca solo una parte de la inteligencia general, mientras que la inteligencia artificial fuerte, en teoría, cubre todo el espectro de la capacidad cognitiva, haciéndola indistinguible de cualquier experto en cualquier área. Además, este tipo de inteligencia probablemente deba ser capaz de resolver el problema del significado, pues debe poder adaptarse a este tipo de situaciones también.

   Searle, 1980 Russell y Norvig, 2020

4. **Investiga y explica brevemente de qué se trata el juego de la imitación de Alan M. Turing. Cita tus fuentes (0.5 pt.).**

   También conocido como la prueba de Turing, es un experimento mental propuesto por Alan M. Turing en su artículo *Computing Machinery and Intelligence*. La idea es que, para probar la consciencia o su ausencia en las máquinas, se sitúan una máquina, una persona y una segunda persona

que actuará como juez; de esta forma, mediante preguntas y respuestas, el mediador, representado en la figura 'C', debe determinar quién es la computadora y quién es la persona. Si el sistema es capaz de engañar a un alto porcentaje de los jueces, se le considera que ha pasado la prueba de Turing.

Turing, 1950

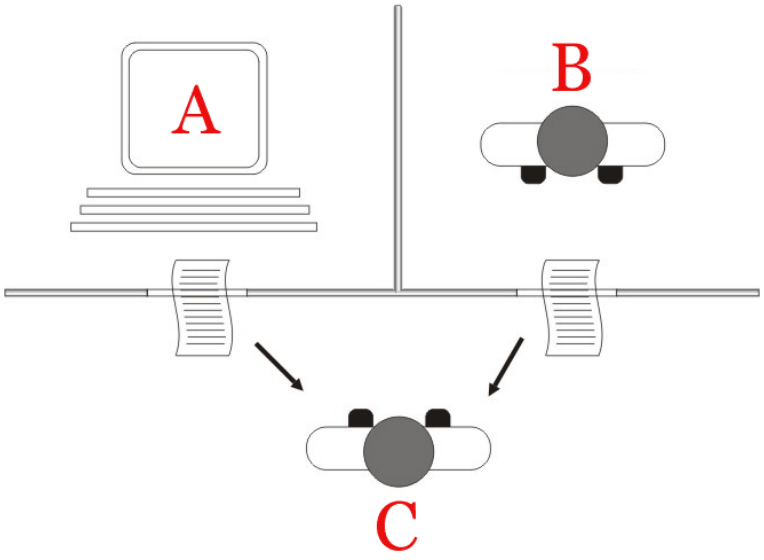

Figura 1: Margallo, s.f.

Aunque el hecho de que el sistema pase el test, en realidad, no nos indica más que su capacidad para comunicarse de manera coherente. Aun así, es altamente probable que nos encontremos ante un escenario similar al experimento del cuarto chino de John Searle, en el que la computadora realmente no tiene idea de lo que está haciendo, pero es capaz de engañar al mediador mediante tácticas inteligentes, generalmente basadas en métodos estadísticos y neuronales.

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.).**

   El informe al que se refiere la pregunta es el elaborado por el *Automatic Language Processing Advisory Committee* (ALPAC) en 1966, donde se revisaron los avances en este tipo de inteligencia artificial y se concluyó que los resultados no eran satisfactorios, recomendándose redirigir el enfoque y los recursos hacia otras áreas de investigación.

   De manera sencilla, el programa fracasó porque no era capaz de generar traducciones de buena calidad, especialmente en comparación con aquellas realizadas por traductores profesionales. Además, se mencionaron grandes limitaciones tanto en hardware como en software de la época, y se destacó que el lenguaje natural es demasiado ambiguo y complejo. En esencia, al comparar los resultados obtenidos con la inversión realizada, el proyecto resultaba insostenible. Este fracaso marcó el inicio del primer ïnvierno de la IA", lo que redujo significativamente el financiamiento para la investigación en el área, además de que los métodos disponibles no eran lo suficientemente avanzados.

   Machine Translation: A Report of the Automatic Language Processing Advisory Committee (ALPAC), 1966

6. **Explica brevemente qué es un agente racional en el contexto de inteligencia artificial. Preferiblemente incluye un diagrama en tu descripción (1 pt.).**

   Un agente racional es una entidad que busca maximizar su función de utilidad; existe en un entorno y es capaz de percibirlo con sus sensores y, posteriormente, actúa sobre él mediante sus actuadores. De manera simple, un agente racional es una entidad que realiza las acciones que más sentido tienen para cumplir sus objetivos.

   Para ilustrarlo de manera más gráfica, se incluye un diagrama creado por ICCSI:

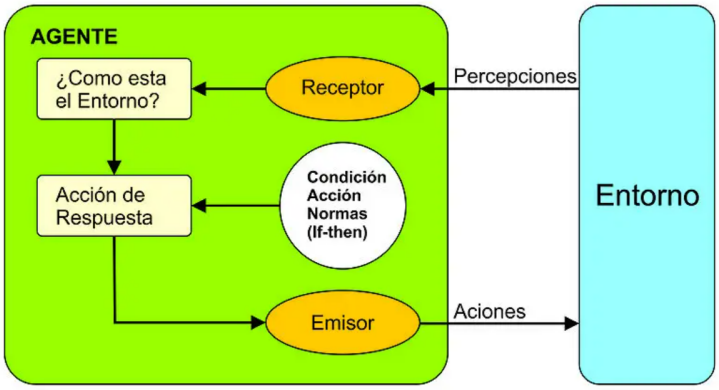

Figura 2: ICCSI, 2023

   Como podemos ver, la definición de agente es lo suficientemente amplia para ser utilizada en diversos contextos.

   Russell y Norvig, 2020

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 siguientealgoritmo 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.

ICCSI. (2023). ¿Qué es un agente software racional? [Imagen disponible en el sitio web de ICCSI]. https://iccsi.com.ar/wp-content/uploads/que-es-un-agente-software-racional-1.webp

Machine Translation: A Report of the Automatic Language Processing Advisory Committee (ALPAC) (inf. téc.). (1966). U.S. Department of Health, Education, y Welfare. Washington, D.C.

Margallo, J. A. S. (s.f.). Test de Turing [File: Test_de_Turing.jpg, CC BY 2.5].

Russell, S., & Norvig, P. (2020). *Artificial Intelligence: A Modern Approach* (4th). Pearson.

Searle, J. R. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, *3*(3), 417-424.

Turing, A. M. (1950). Computing Machinery and Intelligence. *Mind*, *59*(236), 433-460.