Inteligencia Artificial
Introducción a la representación del conocimiento

Primavera 2007

profesor: Luigi Ceccaroni

Representación del conocimiento

- El conocimiento ha de permitir guiar a los mecanismos de IA para obtener una solución más eficiente:
  - ¿Cómo escoger el formalismo que nos permita hacer una traducción fácil del mundo real a la representación?
  - ¿Cómo ha de ser esa representación para que pueda ser utilizada de forma eficiente?

Información y conocimiento

- Llamaremos **información** al conjunto de datos básicos, sin interpretar, que se usan como entrada del sistema:
  - los datos numéricos que aparecen en una analítica de sangre
  - los datos de los sensores de una planta química
- Llamaremos **conocimiento** al conjunto de datos que modelan de forma estructurada la experiencia que se tiene sobre un cierto dominio o que surgen de interpretar los datos básicos:
  - la interpretación de los valores de la analítica de sangre o de los sensores de la planta química para decir si son normales, altos o bajos, preocupantes, peligrosos...
  - el conjunto de estructuras de datos y métodos para diagnosticar a pacientes a partir de la interpretación del análisis de sangre, o para ayudar en la toma de decisiones de que hacer en la planta química

Información y conocimiento

- Los sistemas de IA necesitan diferentes tipos de conocimiento que no suelen estar disponibles en bases de datos y otras fuentes clásicas de información:
  - Conocimiento sobre los objetos en un entorno y posibles relaciones entre ellos
  - Conocimiento difícil de representar de manera sencilla, como intencionalidad, causalidad, objetivos, información temporal, conocimiento que para los humanos es *de sentido común*
- Intuitivamente podemos decir

  Conocimiento = información + interpretación

¿Qué es una representación del conocimiento (RC)?

- Representación del conocimiento o esquema de representación, Davis et al. (MIT), 1993:

1. Un **substituto** de lo que existe en el mundo real o imaginario.
2. Un **conjunto de cometidos ontológicos**: ¿En qué términos hay que pensar acerca del mundo?
3. Una **teoría fragmentaria del razonamiento inteligente**:
   - ¿Qué es la inteligencia?
   - ¿Qué se **puede** inferir de lo que se conoce?
   - ¿Qué se **debería** inferir de lo que se conoce? (¿Qué inferencias son **recomendadas**?)
4. Un **medio para la computación eficiente**: ¿Cómo se debería organizar la información para facilitar la manera de pensar y razonar?
5. Un **medio de expresión humana**: Un lenguaje que las personas usan para hablar entre ellas y con las máquinas.

[1] Es un substituto

- Es un substituto de los objetos del mundo real o imaginario.
- Las operaciones sobre la RC substituyen acciones en el mundo.
- El mismo razonamiento es un substituto de la acción.
- Inversamente, las acciones pueden sustituir el razonamiento.

Preguntas

- ¿Un substituto de qué? → **semántica**

- ¿Hasta que punto exacto como substituto? → **fidelidad**
  - Más fidelidad no es automáticamente mejor
  - La fidelidad perfecta es **imposible**
    - Las mentiras son inevitables
    - Las inferencias incorrectas son inevitables

Terminología y perspectiva

- Inferencia = obtención de nuevas expresiones desde expresiones previas

- Tecnologías de representación del conocimiento (TRCs):
  - Reglas
  - Marcos
  - Lógica
  - Redes semánticas
  - Ontologías
  - …

[2] Conjunto de cometidos ontológicos (COs)

- Los substitutos son inevitablemente imperfectos → La selección de una RC inevitablemente hace un CO
  - Determina lo que se puede conocer, enfocando una parte del mundo y desenfocando las otras
- El cometido ocurre incluso al nivel de las *técnicas de representación del conocimiento* (TRCs)
  - Diagnosis en forma de *reglas* vs. *marcos*

- An ontological commitment is an agreement to use a vocabulary (i.e., ask queries and make assertions) in a way that is consistent (but not complete) with respect to the theory specified by an ontology. We build agents that commit to ontologies. We design ontologies so we can share knowledge with and among these agents.

Tom Gruber <gruber@ksl.stanford.edu>

[3] Fragmento de una teoría de razonamiento inteligente

- ¿Cuáles son todas las inferencias que está permitido hacer?
- ¿ Cuáles inferencias están especialmente impulsadas?

¿Cuáles son todas las inferencias que está permitido hacer?

- Ejemplos
  - Lógica formal clásica: inferencias bien fundamentadas
  - Reglas: inferencias plausibles
  - Ontologías: expectaciones, valores por defecto
- Estilos de respuestas
  - **Precisas**, formuladas en términos de un lenguaje formal
  - **Imprecisas**, no basadas en un lenguaje formal
    - Utilidad del pluralismo

¿Qué inferencias están especialmente impulsadas?

- Ejemplos
  - Ontologías: propagación de valores, enlaces
  - Reglas: encadenamiento, asociaciones
  - Lógica: proposiciones subordinadas (lemas), grafos de conexión
- Explosión combinatoria
  - Necesidad de guía sobre lo que se debería hacer, no sólo lo que se puede hacer

[4] Medio para la computación pragmáticamente eficiente

- Razonar con una RC significa hacer algún tipo de computación
- El componente pragmático
  - ¿Cómo se puede organizar la información para facilitar el razonamiento?
  - Ejemplos:
    - Marcos: *triggers*, jerarquías taxonómicas
    - Lógica: *theorem provers* de grafos de conexión

[5] Medio de expresión y comunicación

- ¿Cómo se expresan las personas acerca del mundo?
- ¿Cómo se comunican las personas entre ellas y con el sistema de razonamiento?
- RC como medio de expresión:
  - ¿Cómo es de general, preciso? ¿Proporciona una expresividad adecuada?
- RC como medio de comunicación:
  - ¿Cómo es de **transparente**? ¿Los humanos pueden entender lo que se está diciendo?
  - ¿Se pueden **generar** las expresiones que interesan?

¿Qué debería ser una RC?

- Los cinco roles tienen importancia.
- Los roles caracterizan el *núcleo* de una representación.
- Hay que tenerlos en cuenta cuando se crea una RC.
- Cada RC es sólo una de muchas posibles aproximaciones a la realidad.

¿Cómo debería ser una RC?

- Pragmática en su visión de validez y eficiencia
- Fuerte en el cometido ontológico
- Pluralista en la definición de las inferencias posibles
- Efectiva en recomendar inferencias y organizar la información
- Efectiva como medio de comunicación
- Rica en abstracciones que corresponden con las tareas
- Capaz capturar la riqueza del mundo natural

RC desde el punto de vista informático

- **Estructuras de datos**: representan el dominio y el problema de manera estática.
- **Procedimientos**: manipulan las estructuras de manera dinámica.
  - **Operaciones**: procedimientos para crear, modificar o destruir las representaciones o sus elementos.
  - **Predicados**: procedimientos para acceder a campos concretos de información.

Tipos de conocimiento

- **Declarativo** (o no procedimental): lógica
  - Pero la lógica puede representar el mismo tipo de procedimientos de un lenguaje de programación. (¡!)
  - Diferencia primaria: la lógica requiere relaciones o predicados explícitos para expresar la secuencia, mientras que los lenguajes procedimentales dependen de la secuencia implícita en la estructura del programa.
- **Procedimental**: funciones, reglas de producción, lenguajes de programación convencionales

Conocimiento declarativo

- Conocimiento relacional simple
  - Conjunto de relaciones del mismo tipo que las de una base de datos
- Conocimiento heredable
  - Estructuración jerárquica
  - Relaciones *es-un* (clase-clase), *instancia-de* (clase-instancia)
  - Herencia de propiedades y valores
    - *Herencia simple y múltiple*
    - *Valores por defecto*
- Conocimiento inferible
  - Descripción vía lógica tradicional
  - Ejemplo de uso: en la resolución

Nivel de granularidad

- ¿A qué nivel de detalle se tiene que representar el mundo?
- **Primitivas**
  - Ej.: relación de parentela
  - Primitivas: madre, padre, hijo, hija, hermano, hermana
  - “Forest es abuelo de Kora”
  - “Iain es primo de Ricardo”
- Si la representación es de **muy bajo nivel**, las inferencias son muy simples, pero ocupan **mucho espacio**.

El *frame problem* (1)

El *frame problem* (1)

Nihil omnino fit sine aliqua ratione

El *frame problem* (1)

Nihil omnino fit sine aliqua ratione
Nothing at all can happen without some reason

El *frame problem* (1)

Nihil omnino fit sine aliqua ratione
Nothing at all can happen without some reason

El *frame problem* (1)

Nihil omnino fit sine aliqua ratione
Nothing at all can happen without some reason

- Ejemplo:
  - El semáforo se pone verde cada minuto par
  - El semáforo se pone rojo cada minuto impar

El *frame problem* (1)

Nihil omnino fit sine aliqua ratione
Nothing at all can happen without some reason

- Ejemplo:
  - El semáforo se pone verde cada minuto par
  - El semáforo se pone rojo cada minuto impar
- **Persistencia**
  - Un programa puede determinar qué pasa en los puntos de tiempo discretos en que el semáforo cambia de color.
  - Pero se necesita más información para determinar qué pasa en los intervalos entre cambios de color.

El *frame problem* (2)

- La declaración en castellano que “el semáforo se queda en verde” o “en rojo” no es necesariamente capturada por los axiomas de un programa.
- Se necesitan dos axiomas de **persistencia** adicionales:
  - Son una manera engorrosa de decir algo que debería ser obvio.
  - Desafortunadamente, los ordenadores no reconocen lo obvio, a menos que no se les diga explícitamente cómo.
  - Estos axiomas engorrosos son la solución de un caso especial del más general *frame problem*.

El *frame problem* (y 3)

- La aproximación general al *frame problem* se basa en el principio de razón suficiente de Leibniz: *Nothing at all can happen without some reason*
  - Es un axioma de meta-nivel que se usa para generar uno o más axiomas de nivel más bajo para cada ejemplo concreto.
    - Implica que el color del semáforo debería quedarse igual a menos que no haya alguna razón para que cambie.