

Pauta Solemne 1 - Inteligencia Artificial

(1er Semestre 2013)

Universidad Diego Portales

Profesor: Alejandro Figueroa.

Ayudante: Nicolás Olivares.

Fecha: Viernes 10 de Mayo de 2013.

Introducción (25 puntos)

- 1. Se dice que el objetivo de la Inteligencia Artificial es imitar el comportamiento humano. Discuta esta aseveración (5 puntos)**

Es verdad que la IA en cierta medida busca imitar algunas tareas que el ser humano realiza, pero es también cierto que gran parte de la investigación se desarrolla ya sea en aumentar la eficiencia de tareas que ya sabemos como realizar, o bien, para encontrar soluciones a problemas que nosotros no sabemos la solución. Es decir, la IA no sólo tiene como objetivo imitar el comportamiento biológico del ser humano, sino que además, desempeñar las tareas que éste puede hacer de una manera más eficiente, y realizar otras que le son imposibles realizar.

- 2. ¿Por qué no es posible jerarquizar máquinas de acuerdo a su coeficiente intelectual? (5 puntos)**

La inteligencia humana se mide en términos del coeficiente intelectual. En especial, a la velocidad y las condiciones humanas de aprendizaje. Las máquinas no están sometidas a las mismas condiciones. Además, existen muchos aspectos que el CI no mide, que podrían afectar la comparación, por ejemplo, tareas manuales. El CI está hecho para humanos, no para máquinas. Por lo demás, aún no hay nivel en la IA para desarrollar máquinas "que aprendan" sistemáticamente como lo hacen los humanos.

3. Compare la inteligencia humana vs. la artificial (5 puntos)

- a. Los programas computacionales tienen muchísima memoria y velocidad, pero sus habilidades están en sintonía con los mecanismos intelectuales que los diseñadores de los programas utilizaron. Es decir, el algoritmo no puede ir más allá para lo cual el ser humano lo diseñó, aún no manifiestan creatividad colectiva.
- b. Los seres humanos desarrollan algunas habilidades de acuerdo a la edad. Aún cuando la IA ha intentado hacer máquinas que desarrollen aprendizaje con el tiempo, se está muy lejos de que sea similar al aprendizaje humano tanto sobre todo en flexibilidad de ámbito/contexto.
- c. Cuando un humano tiene mejor performance en una tarea que una máquina, ya sea completándola o demorando mucho menos, es un indicador de que los diseñadores del programa carecen del entendimiento de los mecanismos intelectuales requeridos para hacer la tarea de manera eficiente.

4. Describa cinco ramas de la Inteligencia Artificial (10 puntos).

- a. **IA lógica** corresponde a los hechos y objetivos que un programa debe saber o manejar acerca del mundo en el cual debe desenvolverse. Normalmente, son expresados por sentencias en algún lenguaje lógico matemático.
- b. **Búsqueda** intenta explorar espacios grandes posibilidades de manera eficiente (poco uso de recursos y rápido)
- c. **Reconocimiento de patrones** aprende regularidades en datos observados los cuales permiten aprender a tomar decisiones en datos nuevos o instancias del problema no vistas antes.
- d. **Representación de conocimiento** intenta buscar modelos eficientes y útiles para almacenar hechos que puedan ser utilizados por otras aplicaciones.
- e. **Inferencia** intenta deducir “hechos” de algunos “hechos” ya existentes. Muchas técnicas existen como el razonamiento no monotónico.
- f. **Planificación** parte con un conjunto de “hecho” acerca del mundo real, un conjunto de hechos acerca de una situación especial y una definición de un objetivo. El objetivo es el diseño de una estrategia para alcanzar el objetivo.
- g. **Epistemología** es el estudio de los tipos de conocimientos que son necesarios para resolver problemas en el mundo.

Búsqueda Combinatoria (50 puntos)

1. Enuncie las diferentes estrategias (ecuaciones) que se utilizan para "calentar" la búsqueda en Simulated Annealing. (5 puntos)

Existen varios mecanismos de enfriamiento α :

- a. Descenso constante de temperatura.
- b. **Geométrico**: $t_{i+1} = \alpha t_i$ $\alpha \in [0.8, 0.99]$
- c. **Boltzmann**: $t_i = t_0 / (1 + \log(i))$
- d. **Cauchy**: $t_i = t_0 / (1 + i)$
- e. **Ludi y Mees**: $t_{i+1} = t_i / (1 + \beta t_i)$, con β muy pequeño.

2. ¿Cuáles son los dos factores que influyen en la decisión de intensificar y/o diversificar en Simulated Annealing? Ejemplifique (15 puntos).

Por un lado tenemos la calidad de la solución, como referencia tenemos el historial, es decir sabemos de cual solución partimos, y donde estamos, por ende, sabemos si la solución que tenemos ahora aparenta ser de buena o mala calidad. Si las soluciones aparentan ser malas, es una buena estrategia moverse a otras malas soluciones con el objetivo de diversificar la búsqueda. Si la solución es buena con respecto al historial, entonces es bastante tentador quedarse explorando las cercanías del punto.

Al explotar las cercanías de un punto es necesario que el movimiento detecte los nuevos patrones que generan las buenas soluciones y los explote. En caso de una mala solución, es importante que el movimiento detecte cuales son los patrones que son más promisorio diversificar.

Por otra parte, tenemos el número de iteraciones, si la búsqueda está muy avanzada se intenta mejorar lo que ya se tiene, en cambio si la búsqueda está en sus comienzos es necesario diversificar más, de manera de conocer o encontrar patrones de soluciones promisorios.

Finalmente, vale la pena recalcar que Simulated Annealing tiene un parámetro que es un mecanismo de control para la diversificación y la intensificación. Ese mecanismo puede llegar a ser muy complejo si es necesario.

3. ¿En qué sentidos se parecen los algoritmos de las colonias de abejas y las colonias de hormigas? (15 puntos).

Primero, ambas son técnicas constructivas. Es decir, generan las soluciones paso a paso, y cada paso toma en cuenta los pasos anteriores para dar el siguiente. Las hormigas ven esta influencia en la lista de vecinas, mientras que las abejas lo ven cuando comparan las soluciones parciales al regresar al panal.

Segundo, ambos algoritmos son de búsqueda paralela, es decir procesan varias soluciones en la misma iteración que pueden o no ser independientes. Tercero, ambas tienen un mecanismo de explotación y exploración. Las abejas al escoger ser fieles a su ruta o no, las hormigas en el factor q_0 .

Cuarto, la estrategia de reclutamiento de las abejas, es un método de colaboración global, ya que una abeja después que decide abandonar su solución tiene una visión global de lo que hacen todas las abejas que son aptas para reclutar. En las hormigas, el mecanismo de colaboración global está dado por la feromona y la respectiva regla de actualización global.

Quinto, uno puede pensar también que en el forward pass las abejas se abren a la exploración (dependiendo del movimiento), y en el backward pass se dedican a la colaboración. En las hormigas, vemos que el forward pass es completo hasta el momento de la colaboración.

4. ¿En qué sentidos se parecen los algoritmos de Hill-Climbing y Simulated Annealing? (15 puntos).

En casi todo debido a que Simulated Annealing es un Hill-Climbing mejorado, en su esencia. Ambos tienen o pueden utilizar métodos similares de generación de soluciones iniciales, e.g. aleatoria y vecino más cercano. Ambos son técnicas que utilizan movimientos para mejorar soluciones, es decir una vecindad. Ambas tienen la idea del descenso de gradiente, sólo que Simulated Annealing permite hacer un "re-start" como el Hill-Climbing Multi-Start con cierta probabilidad. Ambos no tienen memoria (e.g., una lista tabú). Ambos no son algoritmos de búsqueda paralela, y en ambos no hay colaboración entre soluciones para mejorar la búsqueda.

Recuperación de Información (25 puntos)

1. ¿Qué es recuperación de Información? (5 puntos)

Recuperación de Información es la búsqueda de material (documentos) de una naturaleza no estructurada (texto), contenidos en una colección de documentos enorme (típicamente en varias máquinas), que satisface una necesidad.

La frase "*datos no estructurado*" se refiere a datos que no tienen una estructura clara, fácil de identificar para una máquina. Algo estructurado sería una base de datos relacional. En los motores de búsqueda modernos puede ser archivos multimedia.

2. Describa Precisión, Recall y MRR (10 puntos)

Precisión: La fracción de los documentos retornados al usuario que son relevantes a su necesidad de información. Es decir, de los documentos que ve el usuario, cuales satisfacen su necesidad de información, independiente si los ve o no.

Recall: que fracción de los documentos relevantes en el corpus fueron retornados al usuario. Este valor es difícil de estimar, ya que por lo general, no se sabe cuántos son los documentos relacionados cuando la colección es muy grande.

MRR: el valor recíproco de la posición en el ranking del primer documento relevante a la consulta. Se calcula como el promedio obtenido a través de varias consultas. Es importante que el documento relevante aparezca lo más alto del ranking posible.

Nótese que todas estas medidas son dependientes del usuario, ya que la misma consulta puede tener diferentes expectativas dependiendo de los usuarios. Además, cada una de estas métricas provee una visión diferente de los resultados, es decir, de la calidad de las respuestas. Por ende, un motor de búsqueda puede ser mejor en un aspecto, y otro en otro.

3. ¿Qué características debe tener un sistema óptimo de Recuperación de Información? (10 puntos).

- a. Formas de determinar el diccionario, y una recuperación robusta en términos de errores tipográficos y malas elecciones de palabras.
- b. Aumentar el índice de manera de capture la proximidad de términos en los documentos.
- c. Registrar la influencia de una palabra en el documento, no sólo la presencia o ausencia de este. Más peso a los documentos que tienen más veces el término.
- d. Jerarquizar eficientemente los resultados encontrados de acuerdo a la consulta.