

Solemne 2 - Sistemas Inteligentes

martes 20 de Septiembre 2015

Profesor: Alejandro Figueroa

Ayudantes: Alexander Espina - Daniel Palomera

- Está prohibido el uso de teléfonos celulares durante el desarrollo de la prueba.
- La prueba debe responderse con un lápiz de tinta indeleble, de lo contrario no hay opción a correcciones.
- Cualquier alumno que sea sorprendido intentando copiar será calificado con una nota 1.
- Está prohibido conversar durante la prueba. Recuerde que su compañero puede estar concentrado y el ruido puede perturbarlo en el desarrollo de su prueba.
- Utilice sólo las hojas entregadas para escribir sus respuestas.

Pregunta 1

Considere un medio de prensa como www.biobiochile.cl que tiene agrupadas sus noticias en cerca once grupos. El portal tiene una alta afluencia de público ya sea mediante aparatos móviles o de escritorio. Para poder los avisos comerciales es fundamental contar con información de los perfiles de los usuarios. La siguiente tabla muestra la frecuencia con la que postean diez usuarios en las diversas categorías del medio:

Usuario	Nacional	Internacional	Economía	Deporte	Sociedad	Tecnología	Cultura	TV
A	10	11	12	15	9	11	12	13
B	1	1	1	3	1	1	1	1
C	1	2	3	4	4	3	2	1
D	2	3	4	5	5	4	3	2
E	2	2	2	2	2	2	2	2
F	120	110	115	70	50	90	23	24
G	45	12	72	92	52	66	44	92
H	23	34	63	29	70	22	81	82
I	26	89	91	67	40	52	61	82
J	89	48	59	34	26	90	78	56

Un conglomerado de empresas desea contratar servicios publicitarios en Radio Bío Bío, quien cobra por usuario. Las empresas del conglomerado tienen interesados en todas las categorías mostradas en la tabla, y tiene presupuesto sólo para escoger sólo a cinco miembros. Proponga un indicador a este conglomerado y de acuerdo a este indicador señale a qué miembros debería ofrecerle alguno de sus productos.

Desarrollo:

Usuario	Nacional	Internacional	Economía	Deporte	Sociedad	Tecnología	Cultura	TV	Suma	Entropía
A	10	11	12	15	9	11	12	13	93	2.9842741137
B	1	1	1	3	1	1	1	1	10	2.8464393447
C	1	2	3	4	4	3	2	1	20	2.8464393447
D	2	3	4	5	5	4	3	2	28	3.5571580048
E	2	2	2	2	2	2	2	2	16	3
F	120	110	115	70	50	90	23	24	602	2.8023933974
G	45	12	72	92	52	66	44	92	475	2.8490662131
H	23	34	63	29	70	22	81	82	404	2.8253645346
I	26	89	91	67	40	52	61	82	508	2.9068279404
J	89	48	59	34	26	90	78	56	480	2.893516404

Un indicador que nos permite saber lo focalizado de los usuarios es la entropía. De acuerdo a este, podemos ver que los más cercanos a cero son F, H, B, C, G.

Pregunta 2

Compare la complejidad de un problema combinatorio si tiene 60 variables, de las cuales:

- todas son binarias.
- 20 son binarias, 20 pueden ser solo 0 y 20 pueden ser -2, -1, 0 ó 1.
- 30 pueden ser -1, 0 ó 1 y las otras 30 sólo el valor 0.
- ¿Qué sucede en el último caso si se agrega la restricción que las treinta variables que pueden ser -1, 0 y 1 deben tener el mismo valor?

Desarrollo:

Caso a), como todas son binarias, la complejidad es 2^{60} . En el caso b), tenemos veinte binarias, por ende 2^{20} , 20 con un sólo valor factible 1^{20} y otras 20 pueden tomar cuatro valores: 4^{20} . Entonces, para el caso b, tenemos $2^{20} \cdot 4^{20} \cdot 1^{20} = 2^{60}$. Para el caso C, tenemos $3^{30} \cdot 1^{30} = 3^{30}$. Ergo, la complejidad de a) y b) es la misma, en cambio c) es menos complejo ya que 3^{30} es $(3^{1/2})^{60}$ que es 1.73^{60} . En el caso de que las treinta variables tengan el mismo valor, sólo nos permite tres posibles soluciones, sin importar el número de variables.

Pregunta 3

Compare 2-opt con 3-opt. Como referencia, la pauta considera dos similitudes y tres diferencias.

2-opt	3-opt
Es una heurística que encuentra una solución aproximada.	Es una heurística que encuentra una solución aproximada.
Mejoran una solución dada	Mejoran una solución dada
Corta dos arcos e intercambia los destinos	Corta tres arcos e intercambia los destinos
Una forma de reconexión	Dos formas de reconexión
5% de la cota de Help-Karp	3% de la cota de Help-Karp

Pregunta 4

Comparar Backtracking, Fuerza Bruta y Branch and Bound.

Fuerza Bruta	Backtracking	Branch and Bound
Técnica completa, encuentra todas las mejores soluciones global.	Técnica completa, encuentra todas las mejores soluciones global.	Técnica completa, encuentra todas las mejores soluciones global.
Prueba todas las combinaciones, tanto válidas como inválidas.	Se mueve sólo en el espacio de soluciones válidas.	Se mueve sólo en el espacio de soluciones válidas.
Genera soluciones completas	Evalúa soluciones parciales	Evalúa soluciones parciales.
No necesita funciones de	No necesita funciones de	Necesita funciones de

acotamiento.	acotamiento.	acotamiento.
Es generalmente el más lento.	Generalmente, más rápido que Fuerza Bruta, y más lento que Branch and Bound.	Generalmente, más rápido que backtracking y que fuerza bruta.
Necesita un ordenamiento de las variables.	Necesita un ordenamiento de las variables.	Necesita un ordenamiento de las variables.
Necesita un ordenamiento del dominio de cada variable.	Necesita un ordenamiento del dominio de cada variable.	Necesita un ordenamiento del dominio de cada variable.
-	Recursivo, i.e., Recorre el espacio en profundidad	Recursivo, i.e., Recorre el espacio a lo ancho.

Pregunta 5

Comparar Algoritmos Genéticos y Ant Colony Systems. Liste cuatro aspectos donde toman estrategias diferentes, y tres de manera similar.

Algoritmos Genéticos	Ant Colony Systems
Técnica evolutiva	Técnica constructiva
Propagación de genes (herencia)	Feedback positivo
Exploración aplicada mediante un operador de mutación.	Exploración mediante una distribución de probabilidad aplicada de acuerdo un parámetro q_0 .
Explotación basada en el cruzamiento de individuos.	Explotación mediante la selección del mejor movimiento de acuerdo a la regla de transición.
Una iteración consiste en el procesamiento de toda la población de individuos.	Una iteración consiste en el procesamiento de toda la población de hormigas.
Necesita la generación de una población de "n" individuos para inicializar el algoritmos.	Necesita una solución inicial de referencia generada por un mecanismo externo.
Búsqueda paralela: población de "n" individuos.	Búsqueda paralela: población de "m" hormigas.

Pregunta 6

En las tareas del curso, se crearon modelos de minería de datos utilizando la bolsa de palabras. Para esto, se supuso que todas las palabras tenían un impacto beneficioso en la clasificación. Diseñe una heurística Hill-Climbing que

busque mejorar la selección de features de este modelo mediante la eliminación/agregación sistemática de algunas palabras.

Desarrollo:

Función objetivo: el desempeño del clasificador utilizando el subconjunto putativo de palabra, por ejemplo en términos de accuracy.

Representación: Un vector binario de N componentes indicando si la n-ésima palabra en el diccionario debe o no ser incluida en el modelo.

Movimiento: Intercambiar el valor de cada una de las componentes de 0/1, y calcular su desempeño de manera independiente. Escoger la mejor como movimiento.

Punto de partida: Cualquier solución generada aleatoriamente de N ceros y unos.

Condición de término: Ninguno de los “N” swaps provoca una mejora a la solución actual.