

Solemne 2 – Sistemas Inteligentes (Pauta)

Martes 22 de noviembre de 2016

Profesor: Alejandro Figueroa

Ayudante: Alexander Espina

- 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.
- La nota 4.0 se alcanza con 60 de los 100 puntos que tiene la prueba.

Pregunta 1 (20 puntos)

Indique si cada una de las siguientes aseveraciones es verdadera (V) o falsa (F). No es necesario justificar las falsas. Cada respuesta correcta vale 1 punto, mientras que las incorrectas descuentan 0.5 puntos. Letras que no sean nítidas serán consideradas malas.

1. ..F.. PSO es una técnica constructiva evolutiva que desarrolla una búsqueda paralela para resolver problemas de optimización discretos.
2. ..V.. En un ACS tradicional, el nivel de feromona de los arcos perteneciente a soluciones poco interesantes va a tener a converger al mínimo posible.
3. ..V.. Entre más genes tiene un cromosoma de representación binaria, más probable es que sea afectado por la mutación.
4. ..F.. En BCO siempre se tiene que cumplir que NC debe ser divisor de N.
5. ..V.. La memoria atributiva de Tabú Search puede provocar que el algoritmo pase por alto el óptimo global.
6. ..F.. En ACS la tasa de evaporación y de reforzamiento de feromona deben ser iguales para mantener el equilibrio del algoritmo.
7. ..V.. La regla de transición de estados en ACS permite controlar la diversificación y la intensificación de la búsqueda.
8. ..F.. En BCO, las abejas reclutadas van a continuar la misma ruta que la abeja que las recluto.
9. ..V.. BCO al igual que ACS es una técnica constructiva.
10. ...V.. La feromona es un mecanismo de feedback positivo que permite la colaboración entre los agentes en un ACS.
11. ...F.. ACS, BCO, los Algoritmos Genéticos, Búsqueda Tabú y PSO son algoritmos de búsqueda paralela.

- 12...F.. En los algoritmos genéticos, la mutación se focaliza en explotar alguna zona del espacio de búsqueda mientras que el cruzamiento de explorar nuevas zonas de soluciones.
- 13...V.. En PSO, las distintas partículas tienden a converger a la misma zona del espacio de búsqueda al final de un largo proceso iterativo.
- 14...F.. La matriz binaria de incidencias entre documentos y sus respectivos términos es densa.
- 15...V.. Uno de los grandes problemas de una máquina de recuperación de información es que existe una brecha entre las consultas que envía un usuario y el o los documentos más relevantes para aquella consulta.
- 16...V.. El pilar fundamental con el cual se construye un índice invertido son las tuplas: identificador del documento, término.
- 17...F.. Es factible calcular el F1-Score para todas las consultas hechas a un motor de búsqueda web cualquiera (por ejemplo, Google, Yahoo! Y Bing).
- 18...F.. El MRR se ve influenciado por la posición de todos los documentos relevantes devueltos por un motor de búsqueda.
- 19...F.. La lematización transforma una palabra mediante el corte de a una lista de sufijos y prefijos.
- 20...V.. Un problema con el modelo binario de recuperación de información es que no permite el matching de sinónimos.

Pregunta 2 (40 puntos)

1. Detalle dos situaciones complejas en las cuales Hill-Climbing puede encontrarse durante la búsqueda de la mejor solución. (5 puntos)

Básicamente, hay tres opciones producto de la visibilidad local que tiene el operador de movimiento: 1) el algoritmo se encuentra en una planicie donde no hay diferencia entre los puntos a los cuales se puede mover, ergo no hay ni mejor ni peor soluciones en la vecindad; 2) la cima que sube es muy delgada y empinada, por ende pasa de un lado al otro de la “montaña” reiteradamente, haciendo su convergencia mucho más lenta; finalmente 3) el movimiento cayó en un óptimo local, no hay mejor solución en la vecindad, por ende se quedó estancado en un lugar muy alejado al óptimo global.

2. ¿Cuál es la función del parámetro “u” en BCO? (5 puntos)

El parámetro “u” indica el número del forward pass. A medida que “u” crece, las abejas se vuelven menos bravas, es decir son más leales a la solución que han generado hasta el momento.

3. De dos parecidos entre los algoritmos genéticos y PSO. (5 puntos)

El primer parecido consiste en que en PSO se tienen varias partículas recorriendo el espacio de búsqueda en paralelo. En el GA, tenemos varios individuos. En PSO, las partículas intercambian información para generar partículas más cerca del óptimo, en la misma línea, el GA intercambia patrones de las soluciones en el cruzamiento.

4. Compare la complejidad de un problema combinatorio si tiene 30 variables, de las cuales (25 puntos):

- a. Todas son discretas con valores $\{+1, -1\}$
- b. Si las primeras 15 variables son positivas, entonces las 16 últimas son negativas. Además, se sabe que si las primeras 15 variables son negativas, las últimas 15 serán positivas.
- c. Las primeras 16 variables pueden ser $\{+1, -1, -2, -3\}$ y el resto son cero.
- d. Todas las variables deben tener el mismo valor y hay dos opciones.
- e. Diez variables pueden tomar cuatro valores diferentes, otras diez dos y las otras diez sólo un valor. En este escenario, el orden no importa.

En el primer caso, la complejidad está dada por 2^{30} , en cambio en el segundo 2^{15} ya que al fijar el primer grupo, el segundo queda definido. En el tercer caso, dieciséis variables pueden tomar cuatro valores, las otras están definidas, por lo que nos da 4^{16} combinaciones. Las cuales pueden ser re-escritas como 2^{32} . En el cuarto caso sólo tenemos dos opciones. En el quinto caso, tenemos $4^{10} \cdot 2^{10} \cdot 1^{10} = 2^{30}$. Conclusión: en orden de complejidad tenemos: c), a)=e), b) y d).

Pregunta 3 (40 puntos)

1. **Diseñe un algoritmo genético que maximice la función $f(x) = x^2$, donde x puede variar entre 0 y 31¹. (20 puntos)**

Representación: n Cadena de 5 bits

Función de adaptación: $f(a_4 \dots a_0) = (2^4 a_4 + \dots + 2^0 a_0)^2$

Operadores genéticos: Mutación de un punto y Cruce de un punto

Selección: Ruleta

Política de reemplazo: AG generacional

2. **Si una colección de documentos, e.g., noticias, contiene 3600 tokens. Estime paso a paso la frecuencia de la palabra más recurrente de esta colección. Asuma parámetros estándar (o bien aproximaciones que faciliten el cálculo). Grafique en escala logarítmica la curva Zipf para un número determinado de términos frecuentes. (20 puntos)**

¹<http://dis.unal.edu.co/~fgonza/courses/2004-I/CompEvol/introGA.pdf>

Utilizando la ley de Heap $M=kT^b$ ($k=50$ y $b=0.5$), tenemos que el tamaño del diccionario es $50 * 3600^{1/2}=3000$ elementos. (5 puntos)

Utilizando la ley de Zipf tenemos que $X*(1+1/2+1/3+...+1/3000)=3600$. El valor de la sumatoria está entre 8 y 9, más precisamente es 8,58374989. Lo que nos da que X es 419,397122. La palabra más frecuente ocurre 419 veces aproximadamente. (5 puntos)

posición	log posición	frecuencia estimada	log frecuencia
1	0	419	2,62221402
2	0,30103	209,5	2,32118403
3	0,47712125	139,666667	2,14509277
4	0,60205999	104,75	2,02015403
5	0,69897	83,8	1,92324402
6	0,77815125	69,8333333	1,84406277
7	0,84509804	59,8571429	1,77711598
8	0,90308999	52,375	1,71912404
9	0,95424251	46,5555556	1,66797151
10	1	41,9	1,62221402
11	1,04139269	38,0909091	1,58082134
12	1,07918125	34,9166667	1,54303278
13	1,11394335	32,2307692	1,50827067
14	1,14612804	29,9285714	1,47608599
15	1,17609126	27,9333333	1,44612276
16	1,20411998	26,1875	1,41809404
17	1,23044892	24,6470588	1,3917651
18	1,25527251	23,2777778	1,36694152
19	1,2787536	22,0526316	1,34346042
20	1,30103	20,95	1,32118403

