Solucionando el problema de la mochila en el Modelo Sticker

Mario J. Pérez–Jiménez & Fernando Sancho–Caparrini

Ernesto Mancebo

Universidad de Sevilla

Febrero 2020

Contenido

Modelo Sticker

- Modelo propuesto por Sam Roweis.
- Inspirado en Cadenas de ADN.
- Basado en operaciones de filtrado.
- Distinguido por utilizar Memoria de Acceso Aleatorio y la manera de representar la data.

Concepto de Candena de Memoria

Las cadenas de memoria (también llamadas moléculas σ) son cadenas de forma (n, k, m), tal que $n \geq k.m$, siendo n la longitud de la cadena, k la cantidad de sub-cadenas , y m la longitud de cada sub-cadena p.
Cada región m tiene un estado *"apagado"/"encendido"* o *"0"/"1"*

Concepto de Tubo

Un Tubo (T) consiste en un multiconjunto de complejos de memoria o de moléculas σ del mismo tipo.

Se puede ilustrar de la manera: $\begin{bmatrix} \sigma_0 \\ \sigma_i \\ \sigma_{i+1} \end{bmatrix}$

Operaciones sobre las moleculas

- *mezclar*(T_1, T_2)
- *separar*(T, i)
- *encender*(T, i)
- *apagar*(T, i)
- *leer*(T)

Concepto de Librería

- Complejo de memoria de forma (k, l) tal que $(1 \leq k \leq l)$.
- Se inicializan las primeras $k - l$ regiones *"encendidas"/"apagadas"* en todas sus posibles combinaciones.

SubRutina Ordenado por Cardinalidad

Partiendo de los siguientes elementos:

- $A = \{1, \cdots, p\}$
- $B = \{b_1, \cdots, b_s\} \subseteq A$
- $F = \{D_1, \cdots, D_t\} \subseteq P(A)$

Se busca ordenar F con respecto a su cardinalidad en B, o en otras palabras, la cantidad de elementos que conincidan de manera $B \cap D_i$.

Codificando los subconjuntos F

Sea σ una molécula para el tubo T_0, la misma se codifica de forma $T_0 = \{\{\sigma : |\sigma| = p \wedge \exists j(\chi_{D_j} = \sigma)\}\}$; tal que χ_{D_j} es la función característica en A, siendo $(\chi_{D_j}(i) = 1$ si $i \in D_j$ de lo contrario $\chi_{D_j} = 0$ si $i \in A - D_j)$

Resultado esperado

- Una vez concluido el paso i, se tendrán $i+1$ tubos, empezando en T_0.
- Se tendrá $\forall\sigma(\sigma \in T_j \rightarrow |\sigma \in \{b_1, \cdots, b_i\}| = j)$.

Algoritmo

1: **procedure** $Cardinal\_Sort(T_0, B)$
2: **for** $i = 1$ to s **do**
3: $(T_0, T_0') \leftarrow separar(T_0, b_i)$
4: **for** $j = 0$ to $i - 1$ **do**
5: $(T_{j+1}', T_j'') \leftarrow separar(T_j, b_i)$
6: $T_j \leftarrow mezclar(T_j', T_j'')$
7: **end for**
8: $T_i \leftarrow T_i'$
9: **end for**
10: Return $[T_0, ..., T_s]$
11: **end procedure**

Traza

A fin de ilustrar el comportamiento:

- $A : \{0, 1, 2, 3, 4, 5, 6\}$
- $B : \{1, 2, 4\}$
- $F : \{\{2, 6\}, \{3\}, \{4\}, \{2, 4\}\}$

Los elementos que cumplen $B \cap D_j$ en F son:
$\{\{2, 6\}, \{3\}, \{4\}, \{2, 4\}\}$.

F codificado y procesado

Codificando F para llevarlo a un tubo tendremos:

$$\begin{bmatrix} 0,0,1,0,0,0,1 \\ 0,0,0,1,0,0,0 \\ 0,0,0,0,1,0,0 \\ 0,0,1,0,1,0,0 \end{bmatrix}$$

Tras la ejecución de *Cardinal_Sort* tendremos:

$$\begin{bmatrix} T_0 : \{\{0,0,0,1,0,0,0\}\} \\ T_1 : \{\{0,0,0,0,1,0,0\},\{0,0,1,0,0,0,1\}\} \\ T_2 : \{\{0,0,1,0,1,0,0\}\} \\ T_3 : \{\} \\ T_4 : \{\} \\ T_5 : \{\} \\ T_6 : \{\} \end{bmatrix}$$

Otras Notaciones

- $Cardinal\_Sort(T_0)$ cuando $B = A$.
- $Cardinal\_Sort(T_0, l, k)$ cuando $B = \{l, l+1, \cdots, k\}$.

Subrutina Llenado en Paralelo

El propósito de la subrutina de llenado es manipular en el tubo T_0 las moléculas en $(\sigma(i))$ y codifiquen su valor con respecto a f en el sub-conjunto A, esto se define como:

- $A = \{1, \cdots, p\}$
- $r \in \mathbb{N}$
- $f : A \to \mathbb{N}$

Delimitación de Regiones

Podemos considerar la idea de segmentar cada molécula σ en T_0 en posibles regiones como:

- $(A\sigma) = \sigma(1) \cdots \sigma(p)$
- $(L\sigma) = \sigma(p+1) \cdots \sigma(p+r)$
- $(F\sigma) = \sigma(p+r+1) \cdots \sigma(p+r+q_f)$
- $(R\sigma) = \sigma(p+r+q_f+1) \cdots$

Notaciones

Denotamos que si $B \subseteq A$, decimos que $f(B) = \sum_{i \in B} f(i)$. Tal suma nos indica el espacio a reservar en el complejo de memoria T_0 para codificar $f(B)$.
Asimismo, definimos $q_f = f(A)$, tal que $A_i = \{0, \cdots, i\}$ siendo $(1 \leq i \leq p)$.
Por último, tenemos que r corresponde a la región comprendida entre p y $f(B)$.
Todos estos ingredientes nos dan soporte para definir el tubo T_0 de moléculas σ cuyo $k \geq p + r + q_f$.

Algoritmo

1: **procedure** $Parallel\_Fill(T_0, f, p, r)$
2: **for** $i = 1$ to p **do**
3: $(T^+_{i,0}, T^-_i) \leftarrow separar(T_{i-1}, i)$
4: **for** $j = 1$ to $f(i)$ **do**
5: $T^+_{i,j} \leftarrow encender(T^+_{i,j}, p + r + f(A_{i-1}) + j)$
6: **end for**
7: $T_i \leftarrow mezclar(T^+_{i,f(i)}, T^-_i)$
8: **end for**
9: Return T_0
10: **end procedure**

Traza

A fin de ilustrar el comportamiento, estudiemos lo siguiente:

- $A : \{1, 2, 3, 4\}$
- $B : \{2, 3, 4\}$
- $T_0 : \{[2], [4], [3]\}$

Para el tubo T_0 lo codificaríamos de la siguiente forma:

$$\begin{bmatrix} [0,1,0,0,0,0,0,0,0,0,0,0,0,0,0] \\ [0,0,0,1,0,0,0,0,0,0,0,0,0,0,0] \\ [0,0,1,0,0,0,0,0,0,0,0,0,0,0,0] \end{bmatrix}$$

T_0 procesado

Una vez el tubo T_0 sea procesado por $Parallel\_Fill$, tendremos:

$$\begin{bmatrix} 0,1,0,0,1,1,1,0,0,0,0,0,0,0 \\ 0,0,0,1,1,1,1,1,1,1,1,1,1,1 \\ 0,0,1,0,1,1,1,1,1,1,0,0,0,0 \end{bmatrix}$$

Problema de Suma de Subconjuntos

Se busca determinar si existe en B un subconjunto cuyo valor equivalga a k.
En ese sentido, definimos $A = \{1, \cdots, p\}, k \in \mathbb{N}, w : A \to \mathbb{N}$, siendo w la función peso tal que $k \leq w(A) = q_w$.

Algoritmo

1: **procedure** $Subset\_Sum(p, w, k)$
2: $\quad q_w = \sum_{i=1}^{p} w(i)$
3: $\quad T_0 \leftarrow$ Libería$(p + q_w, p)$
4: $\quad T_1 \leftarrow Parallel\_Fill(T_0, w, p, 0)$
5: $\quad T_k \leftarrow Cardinal\_Sort(T_1, p + 1, p + q_w)[k]$
6: $\quad leer(T_k)$
7: **end procedure**

Traza

Teniendo como ejemplo el T_0 anterior:

$$\begin{bmatrix} 0,1,0,0,1,1,1,0,0,0,0,0,0,0 \\ 0,0,0,1,1,1,1,1,1,1,1,1,1,1 \\ 0,0,1,0,1,1,1,1,1,1,0,0,0,0 \end{bmatrix}$$

Para un $k = 10$, una vez el T_0 sea procesado por $Subset\_Sum$ tendríamos como salida: $0,0,0,1,1,1,1,1,1,1,1,1,1,1$

Diseño

Los recursos utilizados para este diseño son: $4 + 2q$ tubos y $2p + q + 1 + \frac{q.(q+3)}{2}$ pasos.

Problema de la mochila acotado

Es un problema considerado NP-completo, en el que se busca recolectar una serie de objetos cuyo peso sea menor que k y valor mayor o igual al k'. En ese sentido, consideramos los valores: $A = \{1, \cdots, p\}$ un conjunto no vacío, $w : A \to \mathbb{N}$ una función que codifica el peso para un A, $\rho : A \to \mathbb{N}$ una función que codifica el valor para un A, asimismo, $k, k' \in \mathbb{N}$, tal que $k \leq w(A) = q_w$ y $k' \leq \rho(A) = q_\rho$.

Algoritmo (I)

1: **procedure** $Bounded\_Knapsack(p, w, \rho, k, k')$
2: $q_w = \sum_{i=1}^{p} w(i); q_\rho = \sum_{i=1}^{p} \rho(i);$
3: $T_0 \leftarrow$Libería$(p + q_w + q_\rho, p)$
4: $T_0 \leftarrow Parallel\_Fill(T_0, w, p, 0)$
5: $T_0 \leftarrow Cardinal\_Sort(T_0, p + 1, p + q_w)$
6: $T_1 = \emptyset$
7: $\ldots$
8: **end procedure**

Algoritmo (II)

1: **procedure** $Bounded\_Knapsack(p, w, \rho, k, k')$
2: $\cdots$
3: **for** $i = k'$ to q_ρ **do**
4:
$T_1 \leftarrow merge(T_1, Cardinal\_Sort(T_0, p + q_w + 1, p + q_w + q_\rho)[i])$
5: **end for**
6: $leer(T_1)$
7: **end procedure**

Diseño

Los recursos utilizados para este diseño son: $5 + 2.max\{q_w, q_\rho\}$ tubos y $4p + k - k' + \frac{q_w.(q_w+5)+q_\rho.(q_\rho+7)}{2} + 1$ pasos.

Problema de la mochila no acotado

Este se distingue de la anterior implementación en el resultado a mostrar, pues ordena los tubos de salida y elije aquel tubo cuyo valor (definido por la función $\rho(B)$) sea el máximo.
Matemáticamente sería: $\rho(B) = max\{\rho(C) : C \subseteq A \wedge w(C) \leq k\}$.

Algoritmo (I)

1: **procedure** $Unbounded\_Knapsack(p, w, \rho, k, k')$
2: $q_w = \sum_{i=1}^{p} w(i); q_\rho = \sum_{i=1}^{p} \rho(i);$
3: $T_0 \leftarrow$Librería$(p + q_w + q_\rho, p)$
4: $T_0 \leftarrow Parallel\_Fill(T_0, w, p, 0)$
5: $T_0 \leftarrow Cardinal\_Sort(T_0, p + 1, p + q_w)$
6: $T_1 = \emptyset$
7: **for** $i = 1$ to k **do**
8: $T_1 \leftarrow mezclar(T_1, Cardinal\_Sort(T_0, p + 1, p + q_w)[i])$
9: **end for**
10: $\cdots$
11: **end procedure**

Algoritmo (II)

1: **procedure** $Unbounded\_Knapsack(p, w, \rho, k, k')$
2: $\quad \ldots$
3: $\quad T_0 \leftarrow Parallel\_Fill(T_1, \rho, p, q_w)$
4: $\quad i = q_\rho; t = 0$
5: $\quad$ **while** $i \geq 1 \wedge t == 0$ **do**
6: $\quad\quad T' \leftarrow Cardinal\_Sort(T_0, p + q_w + 1, p + q_w + q_\rho)[i]$
7: $\quad\quad$ **if** $T' \neq \emptyset$ **then**
8: $\quad\quad\quad leer(T')$
9: $\quad\quad\quad t = 1$
10: $\quad\quad$ **else**
11: $\quad\quad\quad i = i - 1$
12: $\quad\quad$ **end if**
13: $\quad$ **end while**
14: **end procedure**

Diseño

Los recursos utilizados para este diseño son: $5 + 2.max\{q_w, q_\rho\}$ tubos y $4p + k - k' + \frac{q_w.(q_w+5)+q_\rho.(q_\rho+9)}{2}$ pasos.

Conclusiones

- Aplicaciones en ciberseguirdad.
- Aplicaciones en codificación de la data.
- Modelo de estructura de datos.

Bibliografía

- Pérez-Jiménez, M. J., & Sancho-Caparrini, F. (2002, 2002//). *Solving Knapsack Problems in a Sticker Based Model*. Paper presented at the DNA Computing, Berlin, Heidelberg.
- Dasgupta, B., Taghipour, H., Rezaei, M., Esmaili, H. (2013). *Solving the 0/1 Knapsack Problem by a Biomolecular DNA Computer*, Advances in Bioinformatics.
- Roweis, S., Winfree, E., Burgoyne, R., Chelyapov, N., Goodman, M., Rothemund, P., & Adleman, L. (1998). A Sticker-Based Model for DNA Computation. *Journal of computational biology : a journal of computational molecular cell biology, 5*, 615-629. doi:10.1089/cmb.1998.5.615