



Modelos de enfriamiento en recocido simulado

Jeffry Chavarría Molina

jchavarría@itcr.ac.cr
Escuela de Matemática
Instituto Tecnológico de Costa Rica

Juan José Fallas Monge

jfallas@itcr.ac.cr
Escuela de Matemática
Instituto Tecnológico de Costa Rica

Recibido: Setiembre 29, 2015

Aceptado: Diciembre 22, 2015

Resumen. Se realiza una recopilación de los modelos de enfriamiento más utilizados en el algoritmo de recocido simulado. Se muestra una comparación del rendimiento de los modelos en el contexto del problema combinatorio de particionamiento de datos cuantitativos. Además, se propone un modelo empírico alternativo para acelerar el modelo geométrico, el cual es el más comúnmente utilizado en la práctica.

Palabras clave: particionamiento, datos cuantitativos, modelos de enfriamiento, sobrecalentamiento simulado.

Abstract. A collection of the most used cooling schedules for simulated annealing is studied. A comparison is made of the performance of the models on a combinatorial problem for quantitative data. In addition, an alternative empirical model is proposed to accelerate the geometric model most frequently used in practice.

KeyWords: clustering, quantitative data, cooling schedules, simulated annealing.

1.1 Introducción

El algoritmo de recocido simulado (*simulated annealing*, en inglés) fue propuesto por primera vez en 1983 por Kirkpatrick, Gellat y Vecchi, y dos años más tarde fue planteado de forma independiente por V. Černý ([24], [31]). Este algoritmo tiene su base en Metalurgia, específicamente en el proceso de recocido (*annealing*) utilizado para diseñar materiales de mayor dureza. Según Babu & Murty [8], en

Modelos de enfriamiento en recocido simulado. Jeffry Chavarría, Juan J. Fallas

Derechos Reservados © 2016 Revista digital Matemática, Educación e Internet (<http://tecdigital.tec.ac.cr/revistamatematica/>)

una primera etapa un material, como el caso del vidrio, es llevado a una alta temperatura, generando que sus partículas se muevan aleatoriamente. Posteriormente, se entra en un proceso de enfriamiento controlado que permite a las partículas reorganizarse de una manera adecuada, logrando un equilibrio térmico en cada paso de decrecimiento de la temperatura.

Esta idea fue aprovechada por Kirkpatrick et al. [24] y V. Černý, para proponer el algoritmo SA que ha sido ampliamente utilizado para abordar problemas de optimización combinatoria. Este algoritmo utiliza primordialmente la *regla de aceptación de Metropolis* para decidir si en el problema de optimización tratado se acepta o no, bajo una probabilidad, una nueva candidata a solución del problema. En efecto, si en el estado i el sistema se caracteriza por una energía¹ E_i , entonces una perturbación del sistema que genera una nueva energía E_{i+1} es aceptada² si $E_{i+1} - E_i \leq 0$. Sin embargo, el éxito del SA se basa en aceptar nuevos estados aun cuando su energía E_{i+1} no cumpla con la condición $E_{i+1} - E_i \leq 0$. En este caso, la aceptación del estado $i + 1$ se hace solo bajo la probabilidad $P = \exp\left(\frac{E_i - E_{i+1}}{T^*}\right)$ [4, 8, 16], en donde T^* es un parámetro de control, usualmente denominado *temperatura*. En resumen, en el algoritmo de SA la aceptación de nuevas soluciones se rige mediante la regla:

$$P_{T^*}(i + 1) = \begin{cases} 1 & \text{si } f(i + 1) \leq f(i) \\ \exp\left(\frac{f(i) - f(i+1)}{T^*}\right) & \text{si } f(i + 1) > f(i), \end{cases}$$

donde $f(i)$ y $f(i + 1)$ denotan los valores que toma la función objetivo en la iteración i e $i + 1$, respectivamente. Como lo explican Aarts et al. [1], inicialmente, cuando la temperatura es alta, se aceptan soluciones significativamente peores que la solución de mayor calidad que ha encontrado el algoritmo hasta el momento. Por otra parte, cuando el parámetro que controla la temperatura es pequeño, se aceptan solo soluciones que mejoren o que deterioren la calidad, siempre que ese deterioro sea pequeño. Finalmente, cuando la temperatura tiende a cero, solo se aceptan nuevas soluciones si éstas mejoran la solución de mayor calidad que ha encontrado el algoritmo hasta el momento.

Matemáticamente, el algoritmo SA puede ser modelado usando la teoría de cadenas finitas de Markov. Una cadena de Markov es una secuencia de eventos en los que la probabilidad del siguiente evento en la secuencia, depende únicamente del evento anterior [1].

Si Ω denota el conjunto de todos los posibles resultados de un proceso de muestreo y $X(k)$ es una variable aleatoria que denota el resultado obtenido en el k -ésimo evento, entonces la probabilidad de transición para dos posibles resultados $i, j \in \Omega$ está dada por

$$P_{ij}(k) = P\{X(k) = j | X(k - 1) = i\}.$$

Esto es, la probabilidad de obtener el resultado j en el evento k , está condicionada únicamente a haber obtenido el resultado i en el evento $k - 1$. En caso que el conjunto de posibles resultados del experimento sea finito, entonces se dice que la cadena de Markov es finita [1]. El algoritmo 1.1 muestra la estructura general del algoritmo SA. Dado que el sistema debe alcanzar su estabilidad térmica para

¹En el contexto de un problema de optimización combinatoria, se entiende la energía presente en el estado i como el valor que toma la función objetivo durante la i -ésima iteración.

²Esto en el contexto de un problema de minimización, en el cual tiene sentido aceptar este estado $i + 1$ dado que el valor de la función objetivo, representado por E_{i+1} , es menor que el valor en la iteración anterior, y por ende el algoritmo está encontrando una solución de mayor calidad al problema.

cada valor de la temperatura T^* , antes de hacer que ésta decrezca por algún modelo de enfriamiento, entonces se utiliza el parámetro L que representa el largo de la cadena finita de Markov [26]. En la práctica, este parámetro indica el número de soluciones generadas para una temperatura T^* fija (ver línea 6 en el algoritmo 1.1).

Algorithm 1.1: Algoritmo de SA

Entrada: T_f : temperatura final, M : número máximo de iteraciones,
 L : largo de la cadena finita de Markov.

- 1: $s \leftarrow$ solución inicial generada en forma aleatoria.
- 2: $T^* \leftarrow T_{\text{inicial}}$ (temperatura inicial).
- 3: $k = 0$
- 4: **mientras** Criterio de parada sea falso **y** $k < M$ **hacer**
- 5: $k \leftarrow k + 1$
- 6: **para** $j \leftarrow 1$ hasta L **hacer**
- 7: $s' \leftarrow$ Generar una nueva solución mediante una perturbación de s .
- 8: $\Delta E \leftarrow f(s') - f(s)$
- 9: **si** $\Delta E \leq 0$ **entonces**
- 10: $s \leftarrow s'$
- 11: **si no**
- 12: **si** Aleatorio[0,1] $< \exp\left(\frac{-\Delta E}{T^*}\right)$ **entonces**
- 13: $s \leftarrow s'$
- 14: **fin si**
- 15: **fin si**
- 16: **fin para**
- 17: $T^* \leftarrow g(T^*)$ // g denota la función de enfriamiento seleccionada.
- 18: **fin mientras**
- 19: **retornar** s^* , donde s^* es la mejor solución obtenida durante el proceso.

Al igual que en el proceso de recocido, parte del éxito del SA es la rapidez con que se realice el enfriamiento del sistema. Un enfriamiento muy rápido podría no dar tiempo para que las partículas se reorganicen de una forma adecuada, generando eventuales inconsistencias en los materiales. Por ejemplo, un descenso muy brusco en la temperatura aplicada a un vidrio podría generar presencia de burbujas que harían al material más frágil y más propenso a fracturarse. Sin embargo, tampoco se puede recurrir a modelos de enfriamiento tan lentos que aunque favorezcan la calidad y dureza del material, no harán que esté listo en un tiempo razonable para la industria. En resumen, se quiere materiales de alta calidad y dureza, en el menor tiempo posible. Este proceso contextualizado a la optimización combinatoria, implicaría que descensos muy bruscos en la temperatura provocan que el algoritmo no explore adecuadamente el espacio de soluciones factibles. Esto generará que el algoritmo de SA retorne soluciones de mala calidad, que pudieron haber sido mejoradas si se hubiera descendido la temperatura de una forma más controlada. Sin embargo, un enfriamiento muy lento provocará mayores tiempos de ejecución para el algoritmo. En efecto, entre más lento se enfríe el sistema se tienen mayores posibilidades de encontrar mejores soluciones, pero generando tiempos de ejecución más altos [21].

Modelos de enfriamiento

Como ya se indicó, el rendimiento del algoritmo SA depende fuertemente del modelo de enfriamiento seleccionado; un buen modelo podría permitir encontrar buenas soluciones para diversos problemas de optimización combinatoria. Muchos modelos de enfriamiento teóricos han sido probados, pero generalmente han tenido un rendimiento pobre en aplicaciones prácticas. Por lo tanto, en muchos casos se recurre a razonamientos heurísticos y empíricos para proponer modelos que permitan balancear los tiempos de ejecución con la calidad de las soluciones obtenidas [27].

En su artículo Kirkpatrick et al. [24] utilizan un modelo exponencial para realizar el enfriamiento. Específicamente, enfrían el sistema acorde con la fórmula $T_{k+1} = T_0 \cdot \alpha^k$, con $\alpha = 0.90$, $k \in \mathbb{N}$ y T_0 la temperatura inicial. Nótese que esta forma exponencial equivale a la fórmula geométrica $T_{k+1} = \alpha \cdot T_k$, dado que $T_{k+1} = \alpha \cdot T_k = \alpha^2 \cdot T_{k-1} = \dots = T_0 \cdot \alpha^k$. Este modelo es el más utilizado para realizar el enfriamiento al aplicar el algoritmo de SA en diversos problemas de optimización; su simpleza y efectividad puede que sean las razones de esa preferencia. Se recomienda un valor para α cercano a uno, debido a que esos valores de α generan un decrecimiento más lento del modelo geométrico y por ende el sistema tendrá más oportunidad de alcanzar un equilibrio en cada uno de los estados de la temperatura. Kim et al [23] indican que hay dos modelos ampliamente aceptados para hacer el enfriamiento de la temperatura en el algoritmo de SA; el primero de ellos el modelo logarítmico

$$T_k = \frac{T_0 \log(k_0)}{\log(k + k_0)}, \quad k_0 > 1.$$

El segundo, el modelo geométrico $T_k = \alpha T_{k-1}$, $0 < \alpha < 1$. Se tiene que el modelo logarítmico garantiza la convergencia, pero es muy lento en el proceso; por esta razón, el modelo geométrico es el más utilizado en la práctica [23, 31, 33]. Fácilmente se pueden encontrar un gran número de publicaciones en las que se utiliza el algoritmo de SA en diversas disciplinas y que emplean el modelo geométrico para realizar el enfriamiento. Por ejemplo, Rutembar [29] (Electrónica), Wong & Fung [34] (Economía), Babu & Murty [8] (Clustering), Anderson & McGeehan [6] (Sistemas de Comunicación), Wong & Yin-Wa-Wong [35] (Investigación de Operaciones), Barradas et al. [10] (Ingeniería de los Materiales), Svergun [30] (Biología), Kim et al. [23] (Electrónica), Kannan et al. [22] (Computación), Fidanova [18] (Computación), Anagnostopoulos et al. [5] (Optimización Combinatoria), Phillip et al. [28] (Meteorología), Bandyopadhyay et al. [9] (Optimización Multiobjetivo) y Vořechovský & Novák [33] (Probabilidad).

Existen otros modelos de enfriamiento cuyo uso no es tan frecuente como el modelo geométrico. Sin embargo, algunos de ellos tienen una importancia teórica pues han permitido establecer pruebas rigurosas de la convergencia del algoritmo SA. Por ejemplo, Hajek [20] y Geman & Geman [19] presentan desarrollos rigurosos basados en el modelo logarítmico

$$T_{k+1} = \frac{c}{\log(k+1)},$$

para c constante. En su experimento, Geman & Geman [19] utilizaron $c = 3$ y $c = 4$. Por su parte, Osman & Christofides [27] realizan una clasificación de los modelos de enfriamiento. En particular detallan en el modelo propuesto por Aarts & Van Laarhoven [3], el cual establece que

$$T_{k+1} = T_k \cdot \left\{ 1 + \frac{T_k \cdot \ln(1 + \delta)}{3 \cdot \sigma_k} \right\}^{-1},$$

donde δ es una constante cercana a cero (los autores utilizan $\delta = 0.1$) y σ_k es la desviación estándar de los valores que toma la función objetivo en la k -ésima cadena de Markov. Este modelo es empleado

en Zhu et al. [36], Dekkers & Aarts [17] y Van Laarhoven et al. [32] para realizar el enfriamiento en aplicaciones del algoritmo SA.

Posteriormente, Osman & Christofides [27] citan el modelo de Lundy & Mees [25] que establece que la temperatura puede ser disminuida como:

$$T_{k+1} = \frac{T_k}{1 + \beta \cdot T_k}.$$

Si T_0 es la temperatura inicial, T_f es la temperatura final y M es el número total de iteraciones, entonces se puede seleccionar [27]:

$$\beta = \frac{T_0 - T_f}{M \cdot T_0 \cdot T_f}.$$

Este modelo es utilizado en Connolly [13] como parte del algoritmo de SA aplicado al problema de asignación cuadrática (QAP, por sus siglas en Inglés) y en Connolly [14] como técnica para encontrar buenas soluciones a problemas generales codificados en forma binaria.

Basados en el modelo de Lundy & Mees [25], Osman & Christofides [27] proponen un modelo de enfriamiento, haciendo que el valor de β no sea constante en el proceso, sino que dependa del número de iteración. En efecto, proponen

$$T_{k+1} = \frac{T_k}{1 + \beta_k \cdot T_k}, \text{ tal que } \beta_k = \frac{T_0 - T_f}{(\alpha + \gamma \cdot \sqrt{k}) \cdot T_0 \cdot T_f},$$

donde α y γ son constantes. En particular, proponen $\gamma = n$ (n es el número de individuos usado en el contexto del particionamiento de datos; para más detalles consulte la sección III del artículo) y $\alpha = K \cdot N_{feas}$, donde K es el número de clases y N_{feas} es el número de soluciones que fueron aceptadas en el ciclo asociado a la cadena de Markov. Es claro que si $\gamma = 0$, entonces el modelo propuesto por Osman & Christofides [27] coincide con el de Lundy & Mees [25] (basta tomar $\alpha = M$).

Por otra parte, un modelo de enfriamiento alternativo es propuesto por Atiqullah [7], el cual establece calcular la temperatura T_k en la k -ésima cadena de Markov como

$$T_k = T_0 \cdot a^{-\left[\frac{k}{f \cdot k_{\max}}\right]^b},$$

donde a y f son parámetros de control y k_{\max} es el máximo número de cadenas que serán ejecutadas. Nótese que en la última cadena, cuando $k = k_{\max}$ y $T_k = T_f$ (T_f es la temperatura final), entonces

$$b = \frac{-\ln\left(\ln\left(\frac{T_0}{T_f}\right) / \ln a\right)}{\ln f}.$$

Luego, Atiqullah [7] recomienda tomar $a = 2$ y $f = \frac{1}{3}$, lo cual indica que la temperatura será reducida de T_0 a $\frac{T_0}{2}$ en aproximadamente la tercera parte del número total de cadenas de Markov que serán ejecutadas.

Cohn & Fielding [12] realizaron un análisis teórico relacionado con el desempeño de algunos modelos de enfriamiento en SA y de algunas de sus propiedades. Adicionalmente, estos autores examinan los modelos de Aarts & Van Laarhoven [1] (con $\delta = 0.1$, que es el valor usado por los proponentes del modelo), Lundy & Mees [25] (tomando valores de β muy pequeños, que oscilan entre 2.546×10^{-7} y 1.498×10^{-9}), el modelo logarítmico propuesto por Hajek [20] y Geman & Geman [19] (con $c \in \{45, 150, 250, 450, 500, 650\}$), y el modelo geométrico usado por Kirkpatrick et al. [24] (con valores para α entre 0.98700 y 0.99670). Este ejercicio es desarrollado en el contexto del *Problema del agente*

viajero (TSP). La estrategia que siguen los autores para la comparación es fijar una cantidad N de iteraciones para todos los modelos, y controlar los parámetros experimentalmente para que se ajusten a ese número de iteraciones.

El objetivo del presente artículo es diseñar una comparación similar a la realizada por Cohn & Fielding [12], pero en el contexto del particionamiento de datos cuantitativos. Además de los modelos comparados en Cohn & Fielding [12], se analizará el modelo propuesto por Atiqullah [7]. Se busca analizar una eventual ventaja práctica del modelo geométrico tan vastamente usado hoy en día, con respecto a los demás modelos. Una diferencia sustancial en el análisis será que, a diferencia de como lo hacen Cohn & Fielding [12] que fijan el número de iteraciones, en nuestro caso, para una tabla de datos específica, cada modelo enfriará el sistema desde la temperatura inicial T_0 , hasta una temperatura final T_f . Esto con el objetivo de poder ver la evolución del enfriamiento, sin cortar la ejecución con un número de iteraciones. Es claro que hay modelos que enfrían más lentamente que otros, y justamente por ello es que se planifica dejar a cada modelo completar el proceso correspondiente (siempre que esto sea posible en términos prácticos). Con ello se tendrá fundamento para poder ver qué tan lento o qué tan rápido un modelo mueve el sistema desde T_0 a T_f , y qué tanta diferencia en términos de calidad y tiempo de ejecución, representa enfriar con algún modelo lento, en contraposición de usar modelos más rápidos. En resumen, se busca poder ver en el contexto del SA aplicado al particionamiento de datos cuantitativos, la relación tiempo versus calidad, al emplear diversos modelos para el enfriamiento del sistema.

1.2 Particionamiento de datos cuantitativos y las tablas de datos

En este problema combinatorio se considera el conjunto de individuos $X = \{x_1, \dots, x_n\}$, $x_i \in \mathbb{R}^p$, y se quiere construir una partición P de X en K grupos, K conocido a priori. Una partición es un conjunto $P = \{C_1, \dots, C_K\}$ tal que $C_l \subset X$ y $C_l \neq \emptyset$ para todo l , $l \neq l' \Rightarrow C_l \cap C_{l'} = \emptyset$ y $X = \bigcup_{l=1}^K C_l$. La implementación computacional de este problema considera la tabla de datos de tamaño $n \times p$, caracterizada por n individuos y p variables cuantitativas independientes, que se representa en la tabla 1.1. El valor x_{ij} corresponde al valor que toma x_i en la j -ésima variable cuantitativa v_j .

Ind/Var	v_1	v_2	...	v_p
x_1	x_{11}	x_{12}	...	x_{1p}
x_2	x_{21}	x_{22}	...	x_{2p}
\vdots	\vdots	\vdots	\vdots	\vdots
x_n	x_{n1}	x_{n2}	...	x_{np}

Tabla 1.1: Tabla de datos de tamaño $n \times p$.

Si g_l denota el centro de gravedad de la clase l (g_l corresponde al promedio de los individuos que pertenecen a la clase l , esto es $g_l = \frac{1}{|C_l|} \cdot \sum_{x_i \in C_l} x_i$, donde $|C_l|$ denota la cardinalidad de dicha clase), el problema de particionamiento de los individuos x_1, \dots, x_n en K clases, puede formularse como la

minimización de la función

$$W(P) = \frac{1}{n} \sum_{l=1}^K \sum_{x_i \in C_l} \|x_i - g_l\|^2,$$

donde $\|\cdot\|$ corresponde a la norma inducida por la métrica euclídea clásica, tal que

$$\|x_i - g_l\|^2 = d^2(x_i, g_l) = (x_i - g_l)^T (x_i - g_l).$$

La función $W(P)$ se denomina la *inercia intraclase* y permite cuantificar el agrupamiento de los individuos en todas las clases a la vez. Cuanto menor sea el valor de $W(P)$, los individuos pertenecientes a una misma clase son más similares entre sí, indicando por ende una mayor similitud a lo interno de las clases.

1.2.1 Las tablas de datos

Para el análisis se utilizarán ocho tablas de datos cuantitativos extraídas de repositorios internacionales disponibles en <http://archive.ics.uci.edu/ml/> (University of California) y <http://cs.joensuu.fi/sipu/datasets/> (University of Eastern Finland). El valor de W reportado para cada tabla corresponde al valor mínimo de la inercia intraclase que se ha encontrado para esa tabla en el número de clases indicado, y que será tomado como valor de referencia para medir el rendimiento del algoritmo con cada modelo de enfriamiento. A continuación se detallan brevemente las tablas de datos.

- **Iris de Fisher:** Compuesta por $n = 150$ objetos (150 flores de tres especies, por lo que se toma $K = 3$), que son caracterizadas en $p = 4$ variables cuantitativas: largo y ancho del sépalo, y largo y ancho del pétalo. Se considera $W_{\text{Fisher}} = 0.5214$.
- **Wine Quality:** Son dos tablas que versan sobre las variantes roja y blanca del *vinho verde* (vino producido en Minho, zona al noroeste de Portugal). La primera, *winequality-red* (WQ-Red), consta de $n = 1599$ muestras de vino rojo caracterizadas en $p = 11$ atributos cuantitativos. Por su parte, la tabla *winequality-white* (WQ-White) se compone de $n = 4898$ muestras de vino blanco, descritas en esos mismos atributos. En ambos casos se considera $K = 3$, basados en Cortez et al. [15] quienes identificaron tres grupos principales basados en las preferencias sensoriales de los catadores de vino que participaron en el experimento. Para WQ-Red se considera $W_{\text{WQ-Red}} = 247.20$ y para WQ-White se toma $W_{\text{WQ-White}} = 560.41$.
- **Glass:** Está compuesta por $n = 214$ instancias, que corresponden a muestras de seis clases ($K = 6$) de vidrios caracterizadas en $p = 9$ atributos cuantitativos (cantidad presente en cada muestra de Mg, K, Ca, Ba, entre otros elementos químicos). En este caso $W_{\text{Glass}} = 1.57$.
- **S-Set:** Corresponde a un conjunto de cuatro tablas de datos sintéticos, denominadas S1, S2, S3 y S4. Cada tabla tiene 5000 individuos y quince clases. La diferencia entre ellas es el grado de solapamiento entre las clases. En este conjunto se considera: $W_{S_1} = 1783523123.37$, $W_{S_2} = 2655821898.14$, $W_{S_3} = 3377914369.87$ y $W_{S_4} = 3140628447.25$. La Figura 1.1 muestra el conjunto de datos S-Set. En particular se puede apreciar el grado de intersección que existe entre las quince clases en S3 y S4. Esto convierte a este conjunto de datos en uno de alta complejidad para efectos de su particionamiento óptimo.

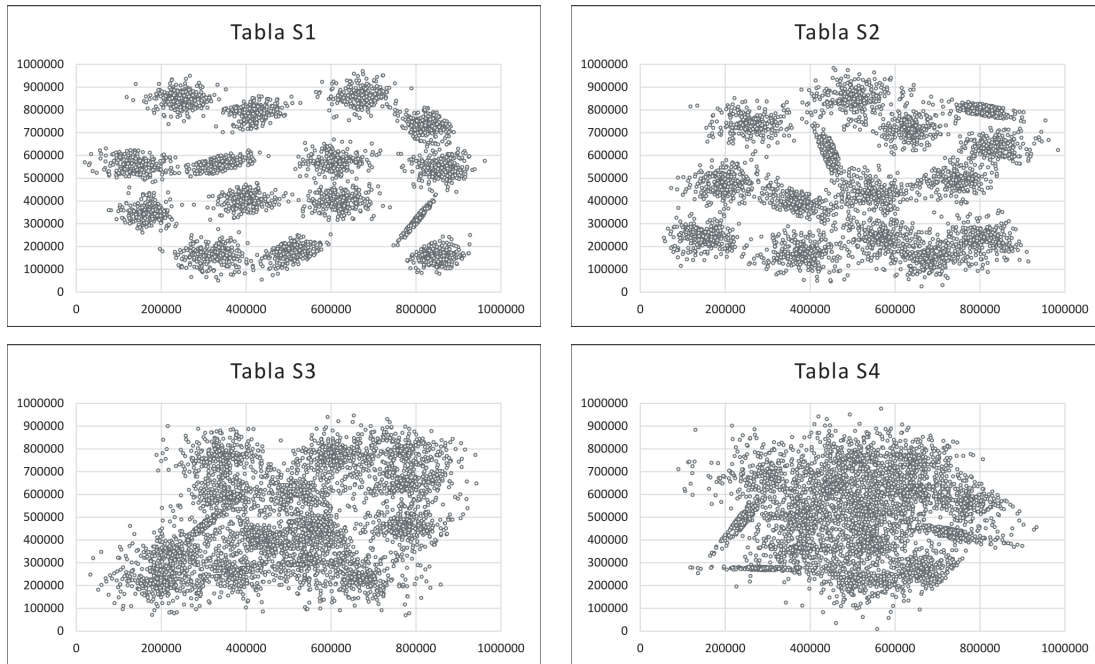


Figura 1.1: Tablas de datos del conjunto S-Set.

1.2.2 Sobre el cálculo de las temperaturas inicial y final

Una de las principales etapas del SA corresponde al cálculo de la temperatura inicial. Si la temperatura inicial es demasiado alta, la búsqueda tendrá un comportamiento muy aleatorio; y si la temperatura es muy baja, entonces la exploración tendrá un comportamiento de búsqueda local, con pocas posibilidades de exploración del espacio [31]. Por ello, debe procurarse un balance en el cálculo de dicho valor, dependiendo de la naturaleza del problema y de los valores que tome la función objetivo. Para efectos de este análisis, para cada tabla de datos se realizó el cálculo de la temperatura inicial basado en el concepto de *criterio de aceptación* indicado en Aarts & Korst [1], y que también es señalado en de los Cobos et al. [16] y Bernábe et al. [11], que se sustenta en el criterio de Metropolis. En este criterio, la probabilidad de aceptación en el caso que $f(j) > f(i)$, se analiza mediante mediante la comparación del valor exponencial

$$\text{Exp}\left(\frac{f(i)-f(j)}{T^*}\right),$$

con un número aleatorio generado con una distribución uniforme en el intervalo $]0,1[$ [2]. Ese aleatorio se puede interpretar como una *tasa de aceptación* χ_0 , con $0 < \chi_0 < 1$, y visualizarlo como el porcentaje que se utilizará para aceptar (bajo una probabilidad) las primeras soluciones en el SA. En síntesis, si $f(j) > f(i)$ y si se fija una tasa de aceptación χ_0 , se tiene que

$$\chi_0 = \text{Exp}\left(\frac{f(i)-f(j)}{T^*}\right) \iff T^* = \frac{f(i)-f(j)}{\ln(\chi_0)}. \quad (1.1)$$

Con base en lo anterior, para cada tabla de datos se generaron de manera aleatoria λ soluciones factibles del problema (en el contexto del particionamiento de datos, esto significa λ posibles particiones de los datos). A cada solución S_i se le construyó un vecino S_i^* tal que $W_{S_i} < W_{S_i^*}$ (condición impuesta sobre la inercia intraclase de la solución y su vecino). Finalmente, basados en la fórmula (1.1), los valores de T_0 y T_f se definieron como el valor del cociente (utilizando un valor χ_0 para el cálculo de T_0 y otro

diferente para el caso de T_f):

$$\frac{\sum_{i=1}^{\lambda} (W_{S_i} - W_{S_i^*})}{\lambda \cdot \ln(\chi_0)}.$$

Dado que en un inicio se desea favorecer una búsqueda de naturaleza aleatoria, entonces para el cálculo de T_0 se utilizó una tasa de aceptación alta, que en este caso se fijó en $\chi_0 = 0.96$. Por otra parte, cuando las temperaturas son bajas solo se desea aceptar soluciones que generen una mejora al sistema. En tal caso, para el cálculo de T_f teóricamente se debe seleccionar $\chi_0 = 0$; sin embargo, en virtud de la fórmula (1.1), en la práctica se selecciona un número positivo suficientemente pequeño, que en este análisis se fijó en $\chi_0 = 10^{-10}$. En ambos procesos se utilizó $\lambda = 10000$. Finalmente, para todas las pruebas y ejecuciones que se presentan en el artículo, se fijó el valor de L (el largo de la cadena finita de Markov) en 100.

1.3 Resultados numéricos

La Tabla 1.2 muestra, para cada conjunto de datos, los valores de T_0 y T_f que fueron calculados para el análisis, con la estrategia descrita. La tabla 1.3 resume los modelos de enfriamiento que fueron implementados en el algoritmo de SA, en el contexto del particionamiento de datos cuantitativos. Así mismo, muestra los parámetros que fueron utilizados en cada modelo.

Tabla 1.2: Resumen de las temperaturas inicial (T_0) y final (T_f), calculadas para las tablas de datos.

Tabla	T_0	T_f
Fisher	0.1853	0.0003
WQ-Red	1.3019	0.0023
WQ-White	0.4307	0.0007
Glass	0.1222	0.0002
S1	46199124	82287
S2	42068053	74274
S3	31721021	56340
S4	23032647	40827

Tabla 1.3: Modelos de enfriamiento por comparar y valores de los parámetros utilizados.

Modelo	Fórmula	Parámetros
Geométrico	$T_{k+1} = T_0 \cdot a^k$	$a = 0.99$
Logarítmico	$T_{k+1} = \frac{c}{\log(k+1)}$	$c = 4$
AartsVanLaarhoven	$T_{k+1} = T_k \cdot \left\{ 1 + \frac{T_k \cdot \ln(1+\delta)}{3 \cdot \sigma_k} \right\}^{-1}$	$\delta = 0.1$
LundyMees	$T_{k+1} = \frac{T_k}{1 + \beta \cdot T_k}$	$\beta = \frac{T_0 - T_f}{M \cdot T_0 \cdot T_f}$
OsmanChristofides	$T_{k+1} = \frac{T_k}{1 + \beta_k \cdot T_k}$	$\beta_k = \frac{T_0 - T_f}{(\alpha + \gamma \cdot \sqrt{k}) \cdot T_0 \cdot T_f}$, $\gamma = n$ y $\alpha = K \cdot N_{feas}$
Atiqullah	$T_k = T_0 \cdot a^{-\left[\frac{k}{f \cdot k_{max}}\right]^b}$	$b = \frac{-\ln\left(\ln\left(\frac{T_0}{T_f}\right) / \ln a\right)}{\ln f}$, $a = 2$ y $f = \frac{1}{3}$

Además de los modelos indicados, en este artículo se propone una estrategia empírica de aceleración del modelo geométrico, basado en el uso de diferentes valores del parámetro α , dependiendo del rango de variación de la temperatura. Esta estrategia nos ha rendido buenos resultados prácticos y, por lo tanto, se incluye con el objetivo de compararla con los otros modelos clásicos. En este caso, se propone realizar el enfriamiento geométrico condicionalmente, de la siguiente manera:

$$T_{k+1} = \begin{cases} T_0 \cdot (0.98)^k & \text{Si } T_k > T_0 - 0.50 \cdot (T_0 - T_f) = \frac{T_0 + T_f}{2} \\ T_0 \cdot (0.99)^k & \text{Si } T_k < T_f + 0.30 \cdot (T_0 - T_f) \\ T_0 \cdot (0.985)^k & \text{En otro caso.} \end{cases}$$

Esta metodología establece enfriar el sistema un poco más rápido al principio (usando $\alpha = 0.98$), durante el primer 50% del rango de disminución de la temperatura total (nótese que el rango total corresponde a la diferencia $T_0 - T_f$). Por otra parte, la cola de la temperatura en el modelo, esto es en el último 30% del rango total, se enfría más lentamente usando $\alpha = 0.99$. Finalmente, nótese que en la franja media determinada entre los porcentajes anteriores, la temperatura se disminuye con un valor intermedio de $\alpha = 0.985$. Para efectos del artículo, este modelo se denominará GeoAcelerado, en función de que corresponde a un aceleramiento del modelo geométrico.

Las Figuras 1.2 y 1.3 se incluyen para ilustrar la forma en que realizan el enfriamiento los siete modelos mencionados, utilizando para ello la tabla de datos WQ-Red. El modelo logarítmico se muestra en una figura aparte, dado que los valores que genera al inicio son mucho más altos que los generados por los otros modelos³, produciendo un problema de visualización del comportamiento de los demás.

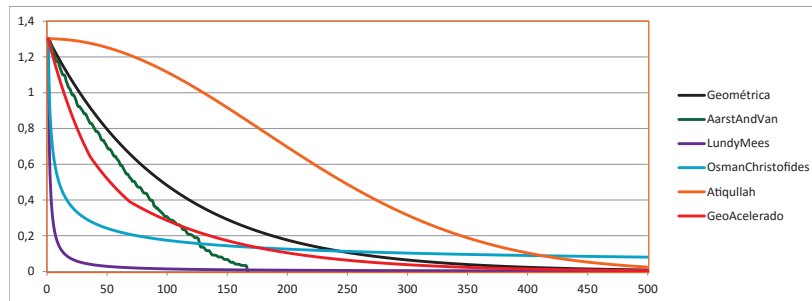


Figura 1.2: Comportamiento de los modelos en la Tabla WQ-Red.

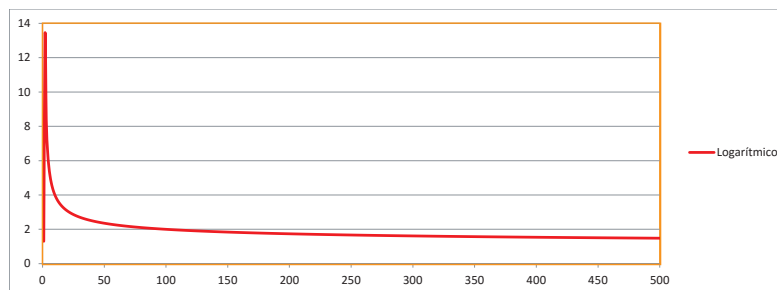


Figura 1.3: Comportamiento del modelo logarítmico en la Tabla WQ-Red.

³Dado que $\log(k+1) < 1$ para $0 < k < 9$, $k \in \mathbb{N}$, entonces en las primeras iteraciones el modelo logarítmico genera un calentamiento al sistema. Este fenómeno se aprecia en la Figura 1.3, en el que $T_0 = 1.3019$ y $T_1 = 4/\log(2) \approx 13.2877$.

La Tabla 1.4 presenta para cinco de los modelos, el tiempo promedio en segundos (de un total de 1000 ejecuciones), la cantidad de iteraciones que fueron requeridas y el porcentaje de rendimiento. En cada caso, esta última característica representa el número de veces que el algoritmo acertó el valor de referencia de la inercia intraclase (ver la sección 1.2.1), de un total de 1000 corridas. No se incluyen resultados del modelo logarítmico, ni el de OsmanChristofides, debido a su lenta convergencia⁴. Ambos modelos requirieron una cantidad significativa e impráctica de iteraciones, para hacer que la temperatura enfriará desde T_0 hasta T_f .

Con base en los resultados de la Tabla 1.4, se nota que existen otros modelos de enfriamiento, diferentes al exponencial, que generaron resultados bastante aceptables en el experimento presentado en este artículo.

Tabla 1.4: Resumen del comportamiento de los modelos en las diferentes tablas de datos.

	Fisher	WQ-Red	WQ-White	Glass	S1	S2	S3	S4
Geométrico								
Tiempo	0.091	1.477	4.295	0.271	3.773	3.866	3.849	3.957
Iteraciones	640	631	639	639	630	631	631	631
Rendimiento	100%	100%	100%	100%	99.6%	100%	78.7 %	1.3%
AarsVanL								
Tiempo	0.033	0.343	0.982	0.085	1.205	1.867	1.942	3.147
Iteraciones	200-240	140-160	110-180	170-220	170-220	240-360	290-400	450-700
Rendimiento	100%	82.3%	100%	92.7%	82.3%	99.8 %	63.8%	0.8%
LundyMees								
Tiempo	0.002	1.470	0.066	0.067	3.965	3.789	3.829	3.846
Iteraciones	15	631	10	160	630	631	631	631
Rendimiento	100%	98.9%	100%	100%	95.6%	99.8%	81.6 %	1.4%
Atiqullah								
Tiempo	0.002	1.485	0.066	0.228	3.856	3.455	3.926	3.933
Iteraciones	15	632	10	525	630	550	631	631
Rendimiento	100%	100%	100%	100%	99.4%	100%	69 %	0.3%
GeoAcelerado								
Tiempo	0.083	1.341	4.067	0.249	3.505	3.496	3.521	3.538
Iteraciones	588	579	588	587	578	579	579	579
Rendimiento	100%	99.8%	100%	100%	98.2%	100%	82.6 %	1.1%

En particular el modelo de Lundy & Mees [25] generó resultados muy similares a los determinados con el modelo geométrico, que inclusive superó en velocidad en tablas como WQ-Red, WQ-White y Glass. Este modelo cuenta con la ventaja que se puede fijar a priori la cantidad de iteraciones que se quieren efectuar para enfriar el sistema desde T_0 hasta T_f . Este aspecto permite controlar, de cierta forma, la velocidad de convergencia del algoritmo. El modelo de Atiqullah [7] tiene esa misma característica, controlada con el parámetro $k_{m\acute{a}x}$. Al igual que con el modelo de Lundy & Mees [25], esto permitió lograr mejores tiempos de ejecución en algunas tablas. Por ejemplo, pueden observarse los tiempos reportados para el modelo de Atiqullah [7] en las tablas Fisher y WQ-White, en comparación con el modelo geométrico. La posibilidad de controlar el número de iteraciones en los modelos de Lundy &

⁴En general, los modelos logarítmicos carecen de una importancia práctica, como efecto de su lenta velocidad de convergencia. Su relevancia consiste en permitir establecer pruebas teóricas, como por ejemplo la demostración de la convergencia asintótica del algoritmo SA.

Mees [25] y Atiqullah [7], ofrece la opción de mejorar el rendimiento del algoritmo, aun cuando esto implique aumentar los tiempos de ejecución. Por ejemplo, en la tabla S3 se exploró el comportamiento de estos modelos haciendo el número de iteraciones igual a 1100. El modelo de Lundy & Mees [25] pasó de un rendimiento del 81.6% a 93.7%, en un tiempo promedio de 6.665 segundos. Por su parte, el modelo de Atiqullah [7] aumentó de 69% a 85.1%, en un tiempo promedio de 6.809 segundos.

Por otra parte, el modelo de Aarts & Van Laarhoven [1] generó resultados ligeramente de menor calidad, pero fue el modelo más rápido para realizar el enfriamiento. Puede que este aspecto de velocidad sea de interés dependiendo del experimento que se esté desarrollando. Finalmente, es notorio que el modelo empírico GeoAcelerado preserva el rendimiento del modelo geométrico, pero hay cierta ganancia en los tiempos de ejecución. Inclusive, en la Tabla 1.4 se puede notar que para S3 fue el modelo que más veces atinó el valor referencial de la inercia. En general, los resultados en el experimento favorecen nuestra tesis de que si se va a enfriar geoméricamente, entonces es viable realizarlo por rangos en la temperatura.

1.4 Conclusiones

Se evidencia en el artículo que el modelo geométrico es el modelo más comúnmente utilizado, dada la gran cantidad de trabajos citados que aplican el algoritmo de SA y que realizan el enfriamiento geoméricamente. Con los resultados expuestos en la Tabla 1.4, se constata en este experimento la sencillez del modelo geométrico y su alto rendimiento. Sin embargo, se hizo notar la existencia de otros modelos que también presentaron un rendimiento similar al geométrico. En particular, el modelo de Lundy & Mees [25] se perfiló como una alternativa muy adecuada para realizar el enfriamiento. En el experimento se constató la poca importancia práctica que tiene el modelo logarítmico, dado que realiza el enfriamiento en un número muy significativo de iteraciones, y empleando para ello largos tiempos de ejecución. De manera similar sucedió con el modelo propuesto por Osman & Christofides [27].

Finalmente, en caso de querer utilizar un modelo geométrico, se reitera la opción de realizar el enfriamiento utilizando diferentes valores para el parámetro α . En particular se recomienda la fórmula:

$$T_{k+1} = \begin{cases} T_0 \cdot (0.98)^k & \text{Si } T_k > T_0 - 0.50 \cdot (T_0 - T_f) = \frac{T_0 + T_f}{2} \\ T_0 \cdot (0.99)^k & \text{Si } T_k < T_f + 0.30 \cdot (T_0 - T_f) \\ T_0 \cdot (0.985)^k & \text{En otro caso.} \end{cases}$$

Bibliografía

- [1] Aarts, E. & Korst, J. (1990). *Simulated annealing and boltzmann machines: A stochastic approach to combinatorial optimization and neural computing*. John Wiley & Sons, Chichester.
- [2] Aarts, E.; Korst, J. & Laarhoven, P. (1997). *Local search in combinatorial optimization*. John Wiley & Sons Ltd., New York.
- [3] Aarts, E. & Van Laarhoven, P. (1985). "Statistical cooling: a general approach to combinatorial optimization problems", *Philips Journal of Research* 40(1): 193226.

- [4] Abbass, H.; Sarker, R. & Newton, C. (2002). *Data mining: A heuristic approach*. Idea Group Publishing, Hershey.
- [5] Anagnostopoulos, A.; Michel, L. & Van Hentenryck, P. (2006). "Simulated annealing based wireless sensor network localization", *Journal of Scheduling* 9: 177193.
- [6] Anderson, H. & McGeehan, J. (1994). "Optimizing microcell base station locations using simulated annealing techniques", en *Vehicular Technology Conference*, IEEE, Reino Unido.
- [7] Atiqullah, M. (2004). "An efficient simple cooling schedule for simulated annealing", en: A. Laganá; M. Gavrilova; V Kumar; Y Mun; C. Tan & O. Gervasi (Eds.), *Computational Science and Its Applications ICCSA 2004*, Springer Berlin Heidelberg, Assisi, Italy.
- [8] Babu, P. & Murty, N. (1994). "Simulated annealing for selecting optimal initial seeds in the k-means algorithm", *Indian Journal Pure and Applied Mathematics* 25(1y2): 8594.
- [9] Bandyopadhyay, S.; Saha, M.; Maulik, U. & Deb, K. (2008). "A simulated annealing-based multiobjective optimization algorithm: Amosa", *Transactions on Evolutionary Computation* 12(3): 269283.
- [10] Barradas, N.; Jeynes, C. & Webb, R. (1997). "Simulated annealing analysis of rutherford backscattering data", *Applied Physics Letters* 71(2): 291293.
- [11] Bernábe, M.; Espinosa, J. & Ramírez, J. (2009). "Evaluación de un algoritmo de recocido simulado con superficies de respuestas", *Revista de Matemática: Teoría y Aplicaciones* 16(1): 159177.
- [12] Cohn, H. & Fielding, M. J. (1999). "Simulated annealing: Searching for an optimal temperature schedule", *SIAM Journal On Optimization* 9(3): 779802.
- [13] , D. (1990). "An improved annealing scheme for the qap", *European Journal of Operational Research* 46(1): 93100.
- [14] Connolly, D. (1992). "General purpose simulated annealing", *Journal of the Operational Research Society* 43(5): 495505.
- [15] Cortez, P.; Cerdeira, A.; Almeida, F.; Matos, T. & Reis, J. (2009). "Modeling wine preferences by data mining from physicochemical properties", *Decision Support Systems* 47(1): 547533.
- [16] de los Cobos, S.; Goddard, J.; Gutiérrez, M. & Martínez, A. (2010). *Búsqueda y exploración estocástica*. México Editorial CBI, México D.F.
- [17] , A. & Aarts, E. (1991). "Global optimization and simulated annealing", *Mathematical Programming* 50(1): 367393.
- [18] Fidanova, S. (2006). "Simulated annealing for grid scheduling problem", en *International Symposium on Modern Computing*, IEEE Computer Society, Bulgaria.
- [19] Geman, S. & Geman, D. (1984). "Stochastic relaxation, gibbs distributions, and the
- [20] Hajek, B. (1988). "Cooling schedules for optimal annealing", *Mathematics of Operations Research* 13: 311329.
- [21] Johnson, D.; Aragon, C.; McGeoch, L. & Schevon, C. (1989). "Optimization by simulated annealing: an experimental evaluation; part i, graph partitioning", *Operations Research Society of America* 37(6): 865892.
- [22] Kannan, A.; Mao, G. & Vucetic, B. (2006). "Simulated annealing based wireless sensor network localization", *Journal of Computers* 1: 1522.
- [23] Kim, D.; Kim, K.; Jang, W. & Chen, F. (2002). "Unrelated parallel machine scheduling with setup times using simulated annealing", *Robotics and Computer Integrated Manufacturing* 18: 223231.
- [24] Kirkpatrick, S.; Gellat, C. & Vecchi, M. (1983). "Optimization by simulated annealing", *Science* 220(4598): 671680.
- [25] Lundy, S. & Mees, A. (1986). "Convergence of an annealing algorithm", *Mathematical programming* 34(1): 111124.

- [26] Mesa, G. (2007). "Cadenas de markov, una sencilla aplicación", *Revista Memorias* 5(9).
- [27] Osman, I. & Christofides, N. (1994). "Capacitated clustering problems by hybrid simulated annealing and tabu search", *International Transactions in Operational Research* 1(3): 317336.
- [28] Phillip, A.; Della-Marta, P.; Jacobeit, J.; Fereday, D.; Jones, P.; Moberg, A. & Wanner, H. (2007). "Long-term variability of daily north atlantic?european pressure patterns since 1850 classified by simulated annealing clustering", *Journal of Climate* 20(1): 40654095.
- [29] Rutembar, R. (1989). "Simulated annealing algorithms: an overview", *IEEE Circuits and Devices Magazine* 5: 1926.
- [30] Svergun, D. (1999). "Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing", *Biophysical Journal* 76: 28792886.
- [31] Talbi, E. (2009). *Metaheuristics: From design to implementation*. John Wiley & Sons, New Jersey.
- [32] Van Laarhoven, P.; Aarts, E. & Lenstra, J. K. (1992). "Job shop scheduling by simulated annealing", *Operations Research* 40(1): 113125.
- [33] Vořechovský, M. & Novák, D. (2009). "Correlation control in small-sample monte carlo type simulations i: A simulated annealing approach", *Probabilistic Engineering Mechanics* 24: 452462.
- [34] Wong, K. & Fung, C. (1993). "Simulated annealing based economic dispatch algorithm", *Generation Transmission and Distribution* 140: 509515.
- [35] Wong, K. & Yin-Wa-Wong, S. (1996). "Combined genetic algorithm/simulated annealing/fuzzy set approach to short-term generation scheduling with take-or-pay fuel contract", *Transactions on Power Systems* 11: 128136.
- [36] Zhu, Y.; Wen, H. & Xu, Z. (2000). "Global stability analysis and phase equilibrium calculations at high pressures using the enhanced simulated annealing algorithm", *Chemical Engineering Science* 55(1): 34513459.