

Simulated Annealing

Ing. Valerio Lacagnina

Introduzione

Il Simulated Annealing (SA) è divenuto negli ultimi anni una euristica ampiamente discussa. Autori come Johnson et al. forniscono eccellenti descrizioni del SA con analogie e paragoni rispetto alla fisica del fenomeno di formazione delle strutture cristalline. Aarts e Korst ne inquadrano la storia, dalle origini dell'idea principale negli anni '50 fino alla sua applicazione nei problemi di ottimizzazione negli anni '80. Questi ultimi presentano un modello formale del SA e provano che l'algoritmo converge alla soluzione ottima con probabilità arbitrariamente vicina ad uno ma con una velocità di convergenza che può anche essere *esponenziale*. Per contro, utilizzando il SA come tecnica euristica, il suo comportamento asintotico può essere approssimato in un tempo polinomiale. Le applicazioni più diffuse del SA sono su problemi combinatoriali, in particolare su problemi di scheduling. In particolare nell'ambito dello scheduling di processi industriali su macchine si hanno varie applicazioni da problemi con tempi di settaggio dipendenti dalla sequenza a problemi single-machine tardiness e problemi flow shop. Furono Van Laarhoven, Aarts e Lenstra nel 1988 ad applicare per la prima volta il SA per problemi generali di tipo job-shop. Ultimamente Jeffcoat e Bulfin hanno realizzato un approccio SA per problemi di scheduling a unità di processamento parallele con risorse limitate. Tutte le applicazioni SA nell'ambito di problematiche combinatoriali mostrano una notevole efficienza del metodo congiuntamente ad una robustezza rispetto il tipo di problema, confermata anche in altre applicazioni di natura combinatoriale mista-intera.

Algoritmo Simulated Annealing

È una metodologia di ricerca altamente adatta per qualunque problema di ottimizzazione non convessa, e fonda le sue basi nella statistica meccanica. È stata sviluppata originariamente da Kirkpatrick et al. per risolvere problemi di ottimizzazione combinatoriale e discreta. Il SA è nato come metodo di simulazione della tempra (*annealing*) dei solidi. L'annealing è il processo con il quale un solido, portato allo stato fluido, mediante riscaldamento ad alte temperature, viene riportato poi di nuovo allo stato solido o cristallino, a temperature basse, controllando e riducendo gradualmente la temperatura. Ad alte temperature, gli atomi nel sistema si trovano in uno stato altamente disordinato e quindi l'energia del sistema è elevata. Per portare tali atomi in una configurazione cristallina altamente ordinata (statisticamente), deve essere abbassata la temperatura del sistema. Riduzioni veloci della temperatura possono causare difettosità nel reticolo cristallino con conseguente metastabilità, con fessurazioni e fratture del reticolo stesso (*stress termico*). L'annealing evita questo fenomeno procedendo ad un graduale raffreddamento del sistema, portandolo ad una

Simulated Annealing

Ing. Valerio Lacagnina

struttura globalmente ottima stabile. Il sistema si dice essere in *equilibrio termico* alla temperatura T se la probabilità $P(E_i)$ di uno stato avente energia E_i è governata dalla *distribuzione di Boltzmann* :

$$P(E_i) = \frac{\exp\left(-\frac{E_i}{k_B T}\right)}{\sum_j \exp\left(-\frac{E_j}{k_B T}\right)},$$

dove k_B è la costante di Boltzmann.

Si noti che ad alte temperature tutti gli stati di energia sono probabilmente possibili, mentre a basse temperature il sistema si trova sicuramente in stati di minima energia. Metropolis et al. nel 1953 sviluppò un algoritmo per simulare il comportamento di una collezione di atomi in equilibrio termico ad una particolare temperatura. Questo algoritmo ha un ruolo fondamentale per l'applicazione del metodo SA a problemi di ottimizzazione. Senza perdita di generalità, si consideri un problema di minimizzazione. La caratteristica essenziale dell'algoritmo di Metropolis è che genera un insieme di configurazioni a ogni temperatura T con la proprietà che le energie delle differenti configurazioni possono essere rappresentate dalla distribuzione di Boltzmann. Il metodo comincia da una assegnata configurazione iniziale degli atomi in un sistema con energia E_0 . Vengono quindi generate successive configurazioni con piccole perturbazioni casuali della configurazione corrente. Viene deciso se accettare o rigettare la configurazione in base alla differenza fra l'energia della configurazione corrente e quella della nuova configurazione (o *configurazione candidata*). Tale decisione è influenzata dal fatto che le energie del sistema delle configurazioni accettate devono formare una distribuzione di Boltzmann se si è raggiunto l'equilibrio termico. L'algoritmo di Metropolis accetta sempre una soluzione candidato la cui energia E_j è inferiore a quella della configurazione corrente (E_i). Per contro se l'energia E_j della configurazione candidato è più grande di quella della configurazione corrente, allora il candidato è accettato con la seguente probabilità :

$$P(\Delta E) = \exp\left(-\frac{\Delta E}{k_B T}\right)$$

dove $\Delta E = E_j - E_i$.

La temperatura non ha alcun diretto analogo in ottimizzazione. Serve meramente come *parametro di controllo* che implicitamente definisce la regione dello spazio di stato esplorato dall'algoritmo in un particolare stadio. Ad alte temperature, l'algoritmo SA può attraversare quasi tutto lo spazio di stato poiché pessime soluzioni vengono facilmente accettate. Successivamente, abbassandosi il valore del parametro di controllo, l'algoritmo viene confinato in regioni sempre più ristrette dello spazio di

Simulated Annealing

Ing. Valerio Lacagnina

stato dato che la distribuzione di Boltzmann collassa a sempre più piccole o basse probabilità di accettazione. Per i problemi di ottimizzazione il SA lavora come segue:

ad alte temperature l'algoritmo si comporta più o meno come una *random search*. La ricerca salta da un punto all'altro dello spazio delle soluzioni individuandone le caratteristiche e quindi le direzioni o le aree in cui è più probabile individuare l'ottimo globale. A basse temperature l'SA è simile ai *metodi steepest descent*. Le soluzioni vengono localizzate nella zona del dominio maggiormente promettente. Da quanto detto si intuisce che vi è un certo numero di parametri che l'analista deve decidere per implementare il metodo, consentendo ampia libertà di scelta e quindi elevata applicabilità. Tuttavia c'è un prezzo da pagare: la taratura di un numero elevato di parametri causa un duro lavoro iniziale affinché il metodo possa convergere. Un vantaggio fondamentale del SA è che l'analista lo può applicare a problemi di ottimizzazione per i quali non ha una conoscenza profonda. L'algoritmo SA per un problema di ottimizzazione può essere riassunto nel seguente generico algoritmo:

- I. Sia data una configurazione iniziale o soluzione x^0 con *energia* o valore della funzione obiettivo E_0 . Selezionare un valore iniziale T_0 per la temperatura.
- II. Per ogni stadio della temperatura effettuare i seguenti passi :
 - A. Generare una configurazione candidato *ammissibile* tramite una piccola perturbazione casuale della configurazione corrente. Valutare la differenza di energia ΔE fra le due configurazioni.
 - B. Se $\Delta E \leq 0$, la configurazione candidato ha un valore della funzione obiettivo inferiore rispetto a quello della configurazione corrente. Accettare la nuova soluzione e sostituirla a quella corrente.
Se $\Delta E > 0$, la configurazione candidato ha un valore della funzione obiettivo peggiore rispetto quello della configurazione corrente. Accettare tale soluzione con una probabilità $P(\Delta E) = \exp\left(-\frac{\Delta E}{k_b T}\right)$ e aggiornare la configurazione corrente se necessario.
 - C. Se non è raggiunto l'equilibrio termico, torna a Step 2.1. Altrimenti vai a Step 3.
- III. Se il processo annealing è incompleto, ridurre la temperatura e ritornare a II.

Simulated Annealing

Ing. Valerio Lacagnina

Schema di raffreddamento

Dato che la riduzione della temperatura governa il successo dell'algoritmo nella sua ricerca delle soluzioni ottime globali, è importante selezionare un appropriato schema di riduzione degli stadi di temperatura determinando quando e quanto ridurre.

Quando accade se ad una data temperatura le energie o i valori della funzione obiettivo delle soluzioni accettate approssimano la distribuzione di Boltzmann per quello stadio di temperatura. Chiaramente più elevato è il numero di soluzioni accettate, più si approssima la distribuzione di Boltzmann. Tuttavia, generare una elevata quantità di soluzioni per ogni stadio di temperatura porta ad un aumento del tempo di calcolo. A tal fine è necessario tarare un parametro detto *transizione*. Il numero di transizioni ad ogni stadio di temperatura deve essere tale da assicurare che si passi da uno stadio termico al successivo solo quando si è raggiunto l'equilibrio ossia le configurazioni siano distribuite secondo una probabilità di Boltzmann.

Quanto ridurre è l'altra importante componente dello schema di raffreddamento. Ridurre la temperatura troppo velocemente causa la possibilità da parte dell'algoritmo di rimanere confinato in un minimo locale. D'altro canto, ridurre la temperatura molto lentamente comporterà un aumento degli stadi di temperatura con un conseguente innalzamento del tempo di calcolo. Basandosi su queste considerazioni gli analisti determinano un appropriato schema di raffreddamento in base al problema specifico. Vi sono, essenzialmente, due schemi di raffreddamento della temperatura: quello *geometrico* e quello *logaritmico*. Il primo è molto semplice, e viene utilizzato dalla maggioranza dei ricercatori: la temperatura viene ridotta, ad ogni stadio, moltiplicandola per una costante chiamata *cooling ratio*, che ha un valore compreso fra 0 ed 1. Lo schema di raffreddamento logaritmico è un metodo più complesso utilizzato in particolari casi (si veda Van Laarhoven et al.).

Criterio di interruzione

La determinazione di un appropriato criterio di interruzione è un fattore critico per il successo del SA. Il processo di annealing converge ad un ottimo globale se la temperatura è sufficientemente ridotta in modo graduale. Al fine di evitare eccessivo tempo di calcolo, ciò che viene usualmente fatto è di interrompere il processo quando il numero di soluzioni accettate ad un certo stadio è inferiore ad un valore fissato. In tal caso ci si può aspettare una soluzione prossima all'ottimo globale. Eventualmente, per affinare o verificare la bontà della soluzione ottenuta, è possibile utilizzare dei metodi di ricerca locale.

Tipicamente, quando non si voglia effettuare un criterio di interruzione, è possibile utilizzare un valore finale della temperatura. Tale valore corrisponde dal punto di vista fisico dell'annealing al

Simulated Annealing

Ing. Valerio Lacagnina

valore della temperatura ambiente quando il bagno di fusione si è solidificato in una struttura cristallina stabile e di minima energia. Tecnicamente tale valore finale della temperatura viene stabilito in base al problema che si sta affrontando.

Generazione delle soluzioni e soluzione finale

Anche la generazione delle soluzioni rappresenta un punto cardine per la velocità di calcolo dell'algoritmo SA. È necessario che la procedura di generazione, detta anche *perturbazione*, sia il più veloce possibile. Il numero elevato di transizioni per ogni stadio di temperatura, se associato ad un elevato tempo di generazione di soluzioni candidato, porterebbe sicuramente ad un aumento esponenziale del tempo di calcolo. Ecco perché solitamente con questo metodo vengono utilizzate tecniche di perturbazione locale e/o di *neighborhood search*. Infatti tali tecniche consentono la generazione di soluzioni in tempi rapidissimi, anche se il loro valore non è molto "distante" da quello della soluzione corrente. Ciò aggiunge un altro non irrilevante vantaggio: la possibilità che la configurazione candidato sia anch'essa una soluzione ammissibile. Anche se per il SA è stata dimostrata la convergenza, si ricorda che questa si ha al prezzo di un elevato tempo di calcolo; di conseguenza ridurre tale tempo implica una più elevata velocità di riduzione della temperatura e un numero inferiore di transizioni, con la probabilità di interrompere il processo di annealing prima che sia arrivato a convergenza. In tal caso, dato il fluttuare del valore della soluzione, è una buona prassi, per ottenere una maggiore bontà della soluzione finale, conservare la migliore configurazione individuata man mano che l'algoritmo lavora. Ciò assicura che tale soluzione migliore possa essere recuperata, anche se il processo termina con una soluzione di valore peggiore.

Bibliografia

- Aarts E. and Korst J.**, *Simulated Annealing and Boltzmann Machines*, John Wiley & Sons, 1990.
- Corcoran A. L., Wainwright R. L.**, *LibGA: A User-Friendly Workbench for Order-Based Genetic Algorithm Research*, Proceedings of the 1993 ACM/SIGAPP Symposium on Applied Computing (SAC 93), Indianapolis, Indiana, February 14-16, 1993
- Kirkpatrick S., Gelatt C. D., and Vecchi M. P.**, *Optimization by Simulated Annealing*, Science, vol. 1, n. 220, pp. 671-680, 1983.
- Johnson D.S., Aragon C.R., McGeoch L.A., Schevon C.**, *Optimization by Simulated Annealing : an Experimental evaluation ; Part I, Graph Partitioning*, Operations Research, 37, pp. 365-892, 1989
- Laarhoven P. J. M. van and Aarts E. H. L.**, *Simulated annealing: theory and application*, D. Reidel Publishing Company, Dordrecht, 1987
- Lundy M., Mees A.**, *Convergence of an Annealing Algorithm*, Mathematical Programming, 34, pp. 111-124, 1986
- Metropolis N., Rosenbluth A., Rosenbluth M., Teller A., and Teller E.**, *Equation of state calculations by fast computing machines*, Journal of Chemical Physics, vol. 21, pp. 1087-1092, 1953