Lecture 15 Simulated Annealing and Genetic Algorithm

Weinan E[1,2] and Tiejun Li[2]

[1]Department of Mathematics,
Princeton University,
weinan@princeton.edu

[2]School of Mathematical Sciences,
Peking University,
tieli@pku.edu.cn
No.1 Science Building, 1575

Outline

Non-convex global optimization

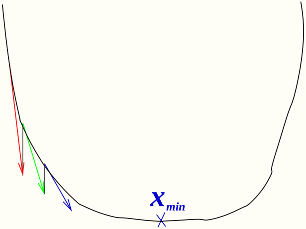
- 对于传统的凸规划问题

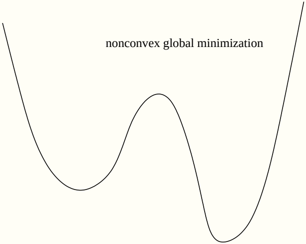
$$\min H(\boldsymbol{x})$$

最典型的可采用最速下降法

$$\frac{d\boldsymbol{x}}{dt} = -\nabla H(\boldsymbol{x})$$

则最后$\boldsymbol{x} \to \boldsymbol{x}^*$,这里$\boldsymbol{x}^*$是$V$的唯一的极小点。共扼梯度法，BFGS等都是可行的选择。

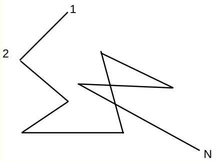




- 非凸问题全局最优化方法?

Traveling Salesman Problem

- **例** TSP问题(Traveling Salesman Problem)
  设有N个不同城市,两两之间有唯一路径连接($l_{ij}=l_{ji}$),
  试寻找一条贯穿所有城市的路径,使得每一个城市必须且只能经过一次,问最短的路径长是多少?



$$\min_{x\in X} H(x)=\sum_{i=1}^{N-1} l_{x_i x_{i+1}}$$

$X=\{(x_1,\ldots,x_N),\ x_1,\ldots,x_N\text{为}1,2,\ldots,N\text{的一个排列}\}$

- 显然所有可能的路径数目为$\frac{N!}{2}$,这是一个典型的组合爆炸问题, 当N增大时,所有可能的状态数以指数增加,而且路径函数$H(x)$ 没有任何规律可循,此时传统的算法将一筹莫展.

图像恢复问题

- **例** 设一幅图像共有J个像素点,每个像素点上有256种颜色,它能表示的所有图像的全体记为

$$X = \{(x_1, \dots, x_J) \ : \ x_i \in \{0, 1, \dots, 255\}\}.$$

一幅图像的光滑度定义为

$$H(x) = \alpha \sum_{<s,t>} (x_s - x_t)^2, \qquad \alpha > 0,$$

这里$< s, t >$指对$x = (x_1, \dots, x_J)$中相邻的像素点求和. 对一幅给定的图像$y \in X$,定义图像x与y的比较函数

$$H(x|y) = \alpha \sum_{<s,t>} (x_s - x_t)^2 + \frac{1}{2\sigma^2} \sum_s (x_s - y_s)^2.$$

在实际问题中,y通常代表一幅被噪音污染的、有缺陷的图像,为从y得到一幅更光滑的图像,可利用以下优化问题:

$$\min_{x \in X} H(x|y).$$

- 这一组合优化问题所有可能状态数为256^J! 通常的优化算法不适用!

Outline

基本框架

- 对于优化问题

$$\min_{x\in X} H(x),$$

定义$H(x)$的全局极小点集$M=\{x_0\ :\ H(x_0)=\min_{x\in X} H(x)\}$,引入参数$\beta>0$,定义

$$\Pi^{\beta}(x)=\frac{1}{Z_\beta}e^{-\beta H(x)},\qquad Z_\beta=\sum_{x\in X}\exp(-\beta H(x)).$$

显然$\Pi^{\beta}(x)$是X上的一个概率分布密度.

- 将静态的问题通过引入逆温度β动态化!

模拟退火基本手段

- **定理** $\Pi^{\beta}(x)$有性质

$$\lim_{\beta\to+\infty}\Pi^{\beta}(x)=\begin{cases}\frac{1}{|M|} & \text{如果}x\in M,\\ 0 & \text{其它}.\end{cases}$$

并且当β充分大时，对任意的$x\in M,\Pi^{\beta}(x)$作为β的函数单调增，对于$x\notin M$，$\Pi^{\beta}(x)$作为β的函数单调减.

物理直观

- $\Pi^\beta(x)$的构造给出了从随机性出发最优化$H(x)$的途径. 定理表明,如果我们能构造对于固定的$\beta > 0$时,产生随机数序列服从$\Pi^\beta(x)$分布,那么令β增加,则“最终”($\beta = +\infty$)随机数会在最小值点中跳跃. 这一过程称退火,它在物理上是有明确意义的.
- 在物理上,β对应$1/T$, T为绝对温度. 一个能量极小化意味着晶体在完美无缺陷的情形下达到自由能的极小,自然界中看到的很多晶体有缺陷是因为他们只是在能量的局部极小态而非全局极小.为生长成一个完美的晶体,可以想象下述物理过程:在温度较高时,晶体会以液态存在,此时将温度缓缓下降,晶体慢慢生成,直到最后到达零温度时,晶体完全长成,我们得到了一个理想的能量极小化过程!
- 当β固定时,关于$\Pi^\beta(x)$的构造我们正好用到标准的Metropolis算法。对于具体问题关键是预选策略要求保证对称性。

模拟退火基本定理及缺陷

- **定理** (模拟退火基本定理) 设X为一有限集,$H(x)$不恒为常数，如果退火方案选为$\beta(n) \leq \frac{1}{\tau\Delta}\ln n$,则对任意初始分布$\nu$, 有

$$\lim_{n\to+\infty} \|\nu P^{\beta(1)}\cdots P^{\beta(n)} - \Pi^{\infty}\| = 0.$$

  这里τ, Δ为相关常数。$P^{\beta(n)}$是当$\beta = \beta(n)$时对应的马氏链的转移概率矩阵。

- 定理表明对模拟退火算法的退火速度必须满足$\beta(n) \leq \frac{1}{\tau\Delta}\ln n$,注意到这是一个非常慢的速度因为$n \geq \exp(\tau\Delta\beta(n))$,为使$\beta(n)$趋于$N_0 \gg 1$, $n \sim \exp(N_0)$,即要计算指数长的时间才能在理论上达到一定的精确度,在实际计算的过程中我们不可能以这一退火速度去计算,而是采取一个更快的实际速度如$\beta(n) \sim p^{-n}$ $(p \lesssim 1)$使$\beta(n) \to +\infty$.自然,这在理论上并不严格.

Outline

遗传算法基本思想

- 遗传算法的基本思想来自于对自然界

  自然选择，　有性生殖

  的这两种基本过程进化择优的计算机模拟。自然选择决定了群体中哪些个体能够存活并繁殖，有性生殖保证了后代基因中的混合和重组。自然界就通过这种自身演化的方式使得物种不断进化，发展。
- 1970年代，John Holland提出了基于遗传进化的优化方法，称之为遗传算法。这些算法数学上基本都可归结为非时齐马氏链的遍历理论。

遗传算法的描述

- 一个简单的遗传算法由复制、杂交和变异三个遗传算子组成。
- 简单例子：假设某个问题涉及三个维度，每个维度只有两种选择，将其记为0或者1，即所有可能为

$$0\text{或}1|0\text{或}1|0\text{或}1$$

共8种。

更一般的如某个问题有n维，每个维度有2^d种选择，则所有可能为

$$\{0,\ 1\}\cdots\{0,\ 1\}|\cdots\cdots|\{0,\ 1\}\cdots\{0,\ 1\}$$

上面由||隔开的共有n个格子，每个格子中的$0,1$有d位。

遗传算法的描述

- 定义群体个数为$N=4$,即演化时每代有4个个体,假设初始个体即第0代为
$$011,\quad 001,\quad 110,\quad 010$$
这些$0,1$位称为遗传染色体. 计算其相应的环境适应值（即目标函数值,假设目标是极大化目标函数）
$$3\quad 1\quad 6\quad 2$$
其总和为12，各自占适应比例为
$$0.25\quad 0.08\quad 0.5\quad 0.17$$
- 复制算子: 即此时重新复制N个个体. 复制时按照上面各自占的适应比例在N个状态中进行随机选择.如011对应的比例为0.25,则以25%的概率选择状态011,其它类似.随机选择可以采取赌盘选择的方法,即按照上述适应比例将圆盘剖分,随机转动圆盘,最后得到的状态即为一个复制状态. 连续选取N次,即得到N个新的状态,此时我们称这N个状态之为交配池.
- 显然上述复制策略即对应强者生存,弱者淘汰的原则.

遗传算法的描述

- 假设上述复制过程之后得到4个新状态:

$$011 \quad 110 \quad 110 \quad 010$$

- 杂交算子: 定义杂交概率$p_c = 50\%$,即需要随机选取$N * p_c = 2$个个体进行杂交,产生新后代.例如单点杂交. 随机选取两个状态如011和110,在第二个位点进行杂交

$$01|1 \quad 11|0$$

得到

$$01|0 \quad 11|1$$

(一般的杂交可以是多点的).其它的直接进行原样复制. 这样我们得到4个新状态(第1代)

$$010 \quad 111 \quad 110 \quad 010$$

如此重复即可得到第2代,第3代,. . .等等.

遗传算法的描述

- 变异算子:变异算子是遗传算法中相对次要的算子.它以很小的概率p_m随机的挑选若干个个体,对于其染色体上的若干个位进行变异,从而得到新的个体.变异算子可以在每次杂交后进行应用.
- 变异算子具有增加群体多样性的效果.
- 有了复制,遗传和变异算子,就可以进行遗传算法.对于一般的离散优化问题,如TSP问题,也可以将城市编号进行排序而进行杂交和变异操作.

References

- G. Winkler, Image Analysis, Random Fields and Dynamic Monte Carlo Methods, Springer-Verlag, Berlin and New York, 1995.
- Jun S. Liu, Monte Carlo strategies in scientific computing, Springer-Verlag, 2001.
- 康立山等,非数值并行算法(第一,二册),科学出版社.