算法理论部分

> - 用粒子的排列或相应的能量表示物体所处的状态，在温度T下，物体(系统)所处的状态具有一定的随机性。**主流趋势是系统向能量较低的状态发展，但粒子的不规则热运动妨碍系统准确落入低能状态。**

简单来说就是，**在温度T下，**

有两种可能

- $f(x_i) \geq f(x_j)$， 那没得说有更好的肯定选更好的。
- $f(x_i) < f(x_j)$，这个时候我们也有一定的概率选这个。

概率为

$$P(X_{i \to j}) = e^{\frac{f(x_i)-f(x_j)}{KT}}$$

其中K为玻尔兹曼常数（在这个算法上，我们经常使用数值1代替这个）

- $K > 0$，一般设置为$K = 1$
- $T > 0$，我们在递减的过程中，终止的T，一般认为是$T = 1$

变量简单分析

不难看出，随着T下降，这个状态转移的概率是下降的。

- **原因：** $f(x_i) - f(x_j) < 0$

物理上解释：

- 因为之前的这个转移概率，认为是，在高温情况下，分子可能会产生较为不稳定的随机运动。且温度越高，这个分子不规则运动的可能是更大的。这是在模拟这个过程。
- 所以，我们称这个算法为，**模拟退火算法**

从状态转移概率到状态概率

这个分析过程，其实是类似于推理马尔可夫链的过程。

- 之前给给出的$P(X_{i \to j})$其实是条件概率。

我们假设第n个状态为X_i，那么现在就是考虑下一个状态的概率。

$$P(s_{n+1} = X_j | s_n = X_i) = P(X_{i \to j}) = e^{\frac{f(x_i)-f(x_j)}{KT}}$$

我们用s_n，表示第n个状态。

用条件概率公式得到

$$P(s_{n+1} = X_j | s_n = X_i) = \frac{P(s_{n+1} = X_j, s_n = X_i)}{P(s_n = X_i)}$$

$$P(s_{n+1} = X_j, s_n = X_i) = P(s_n = X_i) * P(s_{n+1} = X_j | s_n = X_i)$$

$$P(s_{n+1} = X_j) = \sum_i P(s_n = X_i) * P(s_{n+1} = X_j | s_n = X_i)$$

而解这个过程，就用到了以前解马尔科夫过程的方法，这里我需要假设一个均衡态，在这种状态下，经过任意次状态转移之后，整个模型的概率保持一致。

得到结果是

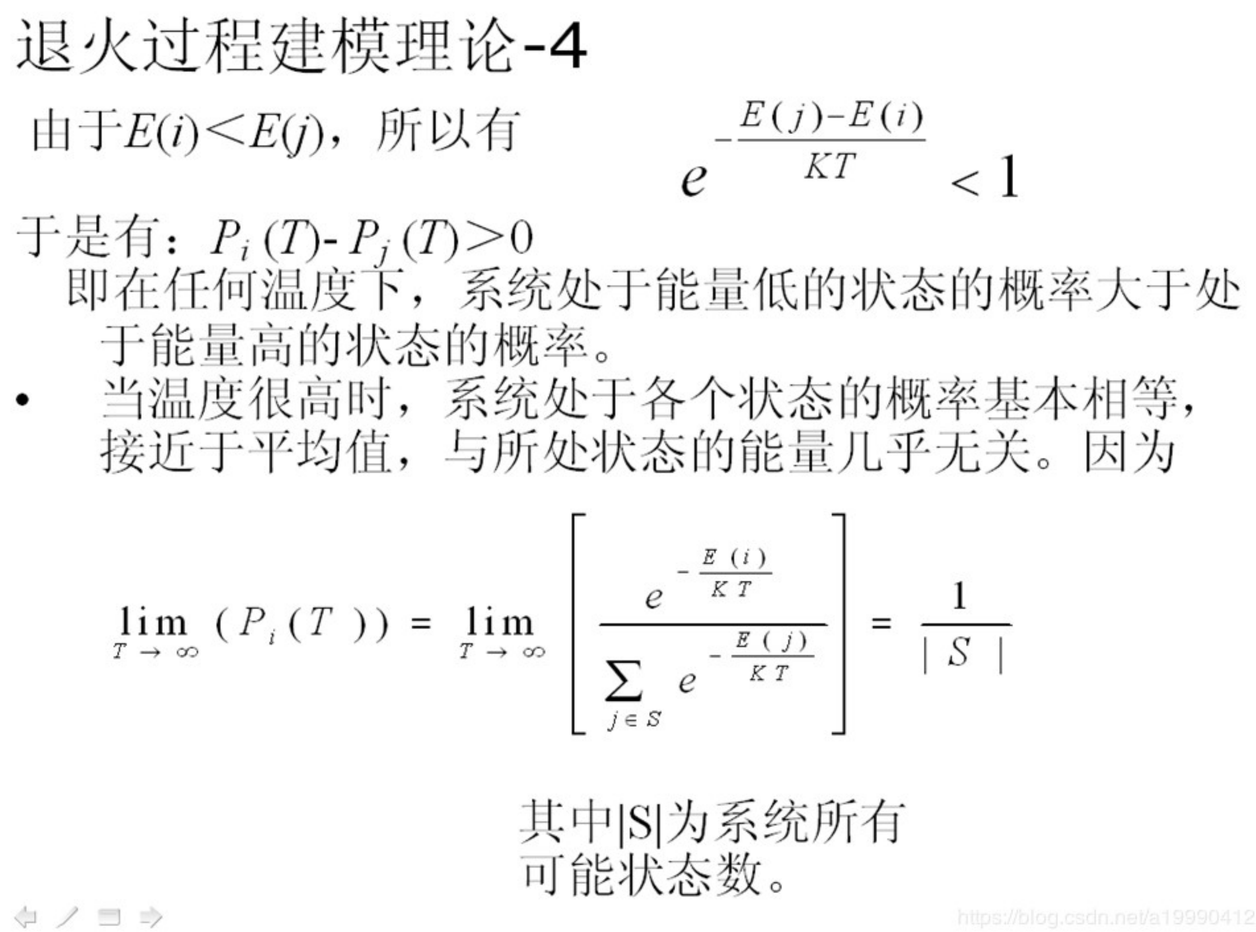
$$P_i(T) = \frac{e^{\frac{-f(x_i)}{KT}}}{Z_T}$$

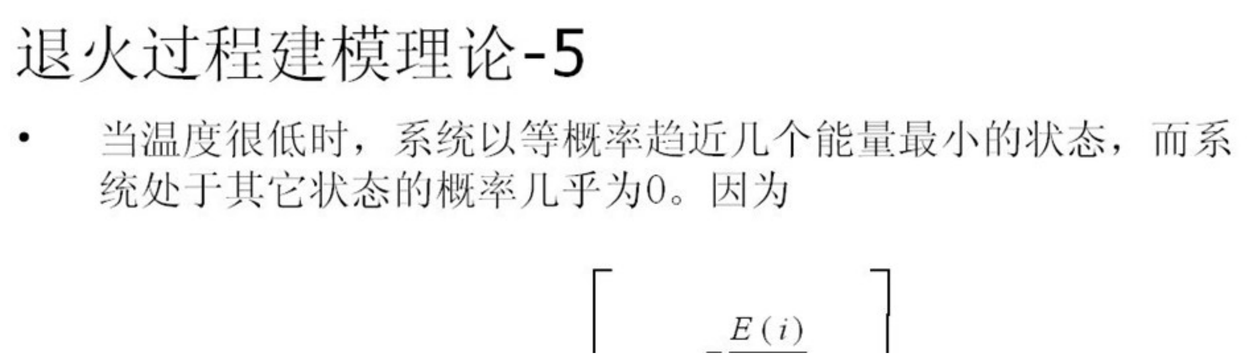
Z_T只是一个归一化的因子而已，就是把所有的这样的分子的数值求个和。使得概率和为一而已。

这个分布称之为 **Boltzmann分布**。

推导

$$P_i(T) - P_j(T) = \frac{e^{\frac{-f(i)}{KT}}}{Z_T}(1 - e^{-\frac{f(j)-f(i)}{KT}})$$





$$\lim_{T \to \infty} (P_i(T)) = \lim_{T \to \infty} \left[\frac{e^{-\frac{E(i)}{KT}}}{\sum_{j \in S} e^{-\frac{E(j)}{KT}}} \right]$$

$$= \lim_{T \to \infty} \left[\frac{e^{-\frac{E(i)}{KT}}}{\sum_{j \in S} e^{-\frac{E(j)}{KT}}} \cdot \frac{e^{-\frac{E_m}{KT}}}{e^{-\frac{E_m}{KT}}} \right]$$

理解当温度收敛到接近0的时候，收敛到结果

其实只需要理解下面这个函数的导数就可以了

$$e^{-\frac{E(i)}{KT}}$$

关于T求导。

$$e^{-\frac{E(i)}{KT}} * \frac{E(i)}{kT^2}$$

那么，这就是这个变量关于T的变化导数。再考虑关于函数值$E(i)$

这个函数是关于$E(i)$的单调减函数。$E(i) > 0$的情况下。

所以说，函数值稍微小的数值所对应的状态概率，受到T的减小而导致的减小的幅度，其实是较为小的。
那么当T趋于0的时候，就可以得到，状态概率，会集中在函数值较为小的数值点上（数值小的点的概率大）

也就是说，当模拟退火，温度越低，越有可能收敛到正确解。而且，这是收敛的。

理论部分的后记

这里，我们证明了，当温度越小，越会收敛到正确解。这只是在理论上的证明。但是我们都知道，当T趋于0，但是特别小的时候，作为分母时，这个精度就会变得非常低了（计算机上是离散的）。

- 所以，为了简单，我们一般设置T到1就截止了。