

精确算法与参数算法

肖鸣宇

电子科技大学计算机科学与工程学院

* 第一部分

精确算法和参数算法简单介绍

* 第二部分

精确算法和参数算法思想在大数据中的思考

精确算法与参数算法

- 计算机科学中“最核心”的问题是算法
- 计算机以超快速度和超强的记忆力帮人类进行计算
- 一类重要的计算方式：穷举搜索
- 很多问题容易获得简单的 $O(2^n)$ 时间的穷举搜索算法

精确算法与参数算法

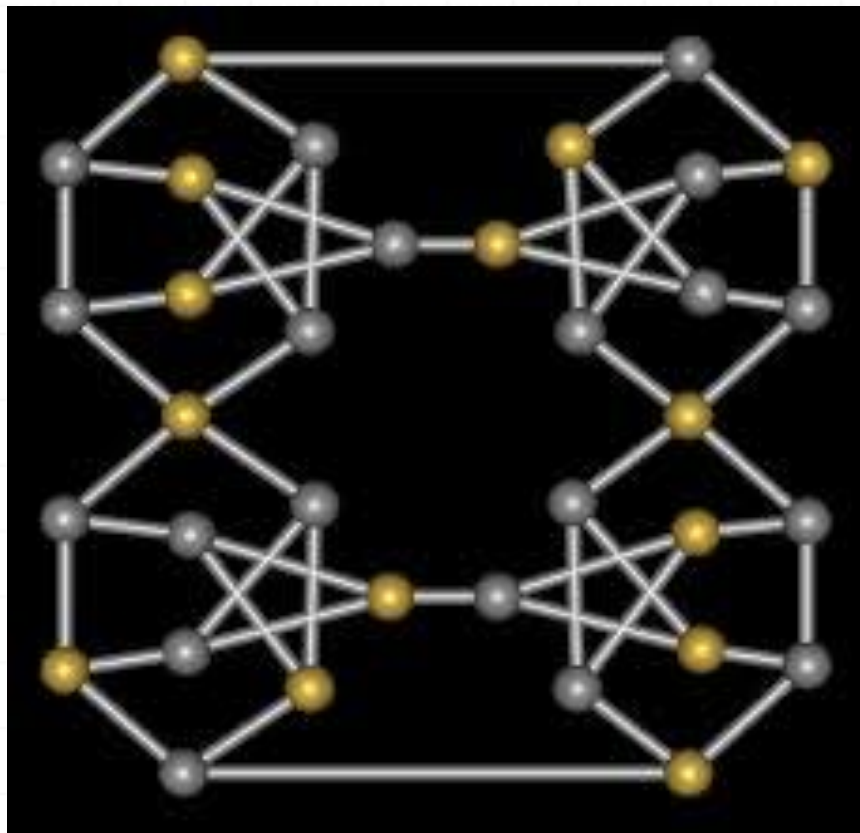
- 事实上，很多问题可能并不存在比穷举搜索明显更快的算法， $O(2^n)$ 这样的指数运行时间可能不能避免。
- 比如说NP难问题。
- 精确算法主要关注在指数运行时间下，怎样改进运算时间。
如： $2^n \rightarrow 1.2^n \rightarrow 1.0001^n$

精确算法与参数算法

图算法问题在精确算法中最为重要的问题举例：

最大独立集问题

旅行商 (TSP)



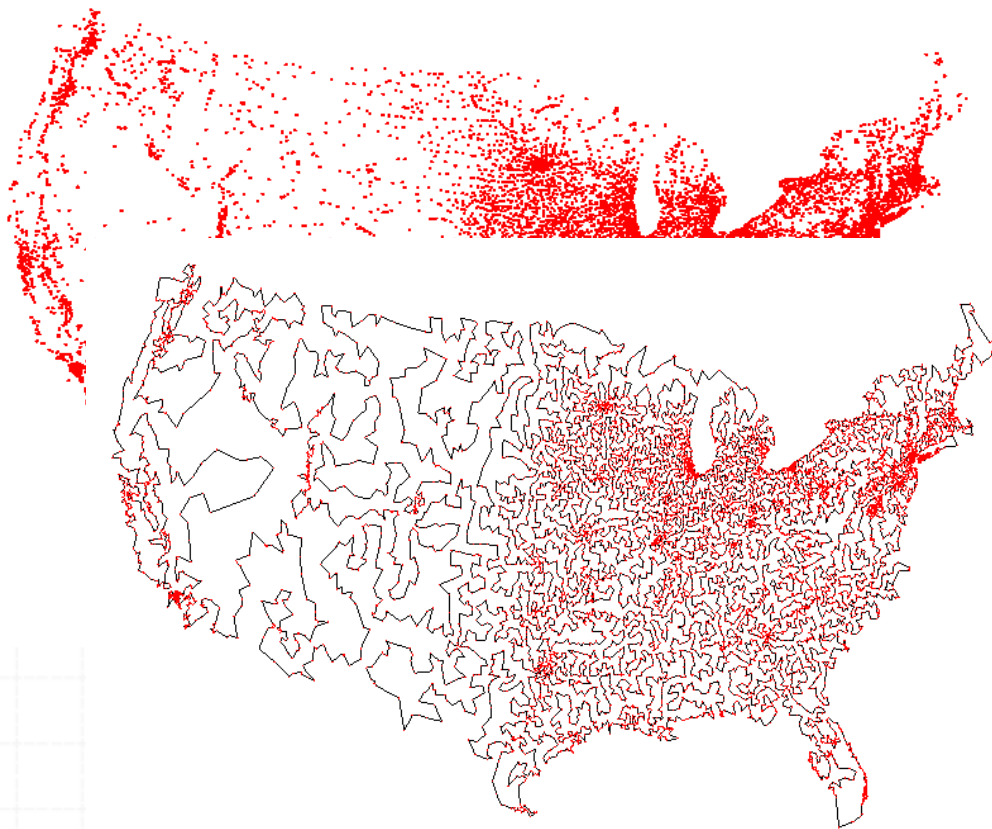
- 独立集：一组顶点两两之间都没有边

精确算法与参数算法

图算法问题在精确算法中最为重要的问题举例：

最大独立集问题

旅行商 (TSP)



- TSP: 访问每个城市一次再回到原点的最短路径

精确算法与参数算法

代表性图问题的最佳精确算法：

最大独立集问题

1. 1996^n [Xiao, Nagamochi, ISAAC 2013]

旅行商 (TSP)

$(2-a)^n$ [Bjorklund *et al.*, TALG 2012]

3度图：1. 2312^n P-space [X&N, TAMC 2013]

1. 2186^n E-space [Bodlaender *et al.*, ICALP 2013]

4度图：1. 6920^n [X&N, COCOON 2012]

精确算法与参数算法

参数算法：给出某个参数 k ，设计精确算法使得指数部分只和参数相关，和问题输入 n 无关。

传统精确算法

$$5^n n^2$$

$$2^n n$$

参数算法

$$k! n^3$$

$$2^k k^2 + kn$$

参数算法的性质：还是指数时间求得最优解，但是当参数 k 比较小的时候（哪怕输入 n 还是很大）问题可能可以很快被解决。

精确算法与参数算法

最为重要的参数问题之一：

最小点覆盖问题，参数 k 为解集的大小。

1. $2.738^k + n^{1/2}m$ [Chen et. al., TCS 2010]

3度图：1. $1.1616^k k + n^{1/2}m$ [Xiao. COCOON 2010]

现在的算法在 $n=100000$ ， $k=600$ 左右都能几个小时内完成。

精确算法与参数算法

参数算法的思想：

精确算法

$$2^n n^2$$



参数算法

$$4^k k^2 + k^2 n^2$$

$$2^n n$$



$$2^k k^2 + kn$$

参数算法里主要关注指数算法；对于多项式算法的态度呢？

$$n^4$$



$$n^2 k^2$$

大数据时代的思考

参数算法里主要关注指数算法；对于多项式算法的态度呢？

$$n^4 \longrightarrow n^2k^2$$

多项式可解的问题里参数算法真的不重要？



Fine-Grained Complexity and Algorithm Design

Aug. 19 – Dec. 18, 2015

大数据时代的思考

多项式算法中的参数算法

$$n^4 \longrightarrow n^2k^2$$

$$n^2 \longrightarrow k^3+nk$$

当参数 k 和 n 是同一个级别的时候，算法可能没有改进，甚至更差，但是当 k 很小的时候却可能非常有用。

大数据时代的思考

参数算法的思想



参数算法讲究将问题难的和容易的部分区分开来，将容易的部分用容易的算法对待，难部分用难的算法对待。

大数据时代的思考

- 大数据下，很多问题在平方时间都超时。
- 线性时间能解决的问题极少。
- 剩下难的问题又如何解决？
- 参数算法或许能派上用途？

大数据时代的思考

- 一个例子：排序问题，运行时间为 $n \log n$.
- 如果只是在 n 个数中找出最大的数，或者最大的前60个数，或者第50大的数，是否需要将 n 个数先排序？
- $n \log k$ 的算法可以找出前排名前 k 的数，当 k 为常数的时候，算法是线性的。
- 注意在最坏的情况 $k=n$ ，算法没有改进。

大数据时代的思考

- 哪些问题有小参数性质：
- 一些例子：
 - VLSI 中电路板分割，分割块数不会很大；
 - 某些社交网络中，俱乐部（club）的个数不会太多；
 - 论文合作者关系网中，一篇论文的合作者一般不会很多；等等。
- 多观察问题性质，建立参数模型。

大数据时代的思考

- 好的参数算法在参数 k 为常数时是线性时间的，是否能做到亚线性时间？
- 对一般求精确解的问题，亚线性时间算法不太可能。
- 亚线性时间算法因为没有将所有有效数据读完，因此一般只是近似算法和随机算法，不能保证最优解。

大数据时代的思考

- 在大数据里，亚线性时间的近似算法和随机算法已经有一些研究基础。
- 线性或接近线性时间的参数算法研究较少。
- 参数算法中的核心化算法等预处理方法在大数据中的应用也提得少。

大数据时代的思考

- 对大数据里的参数算法和亚线性时间算法有兴趣？
- 欢迎联系 myxiao@gmail.com