arXiv:2601.12529v1 [cs.CG] 18 Jan 2026

[1]论文于 2026 年 1 月 21 日使用 ReLaTEX 排版。

L_∞拟合的主要问题在于其对噪声和离群点的敏感性。有两种自然的补救方法。

第一种方法是改变目标函数，使其对离群点不那么敏感。例如，可以考虑平均距离，而不是最大距离。这就是 L_1-拟合问题；在这里，我们希望计算出使 $\ell_1(\mathcal{F}, P) = \min_{\gamma \in \mathcal{F}} \|\mathbf{d}(P, \gamma)\|_1 = \min_{\gamma \in \mathcal{F}} \sum_{p \in P} d(p, \gamma)$ 达到最小值的形状。类似地，在 L_2-拟合问题中，我们希望最小化各点到形状的平均平方距离；也就是说，

据我们所知，此结果唯一直接相关的既有算法 e

2. 预备知识

在整篇论文中，我们将 x_d-平行于 $\mathbb{R}^d$ 的方向称为竖直。给定 $\mathbb{R}^{d-1}$中的一个点 $x = (x_1, \ldots\ldots, x_{d-1})$，令$(x, x_d)$表示 $\mathbb{R}^d$中的点$(x_1, \ldots\ldots, x_{d-1}, x_d)$。每个点 $x \in \mathbb{R}^d$ 也都是 $\mathbb{R}^d$中的一个向量。给定一个几何对象 A，$A + x$ 表示将 A中的每个点平移 x后得到的对象。

一个曲面是 $\mathbb{R}^d$ 的子集，与任意垂直线都恰好交于一个点。一个曲面片是曲面的一部分，其到 $\mathbb{R}^{d-1}$ 中的垂直投影是一个常数复杂度的半代数集，通常为一个单形。设 A 和 B 是 $\mathbb{R}^d$ 中的点、超平面或曲面之一。若对于任意同时与 A 和 B 相交的垂直线 ℓ，都有 $x_d \geq y_d$（分别为 $x_d \leq y_d$），则称 A 位于 B上方（分别为下方），记作 $A \succeq B$（分别为 $A \preceq B$），其中 $(x_1, \ldots, x_{d-1}, x_d) = A \cap \ell$，且 $(x_1, \ldots, x_{d-1}, y_d) = B \cap \ell$。（特别地，如果 A 和 B 都是超平面，那么 $A \succeq B$ 意味着 A 和 B 是平行超平面。）

两个非负数 x 和 y 互为 $(1 \pm \varepsilon)$-近似，如果 $(1 - \varepsilon)x \leq y \leq (1 + \varepsilon)x$ 且 $(1 - \varepsilon)y \leq x \leq (1 + \varepsilon)y$。记作

这一事实为 $x \approx_\varepsilon y$。两个非负函数 $f(\cdot)$ 和 $g(\cdot)$（定义在同一定义域上）互为 $(1 \pm \varepsilon)$-近似，记作 $f \approx_\varepsilon g$，如果对于所有 x，都有 $f(x) \approx_\varepsilon g(x)$。

观察 2.1。设 x 和 y 为两个正数，且 $\varepsilon < 1/4$。我们有：

(i) 若 $x \approx_\varepsilon y$ 且 $y \approx_\varepsilon z$，则 $x \approx_{3\varepsilon} z$。*(ii)* 若 $|x - y| \leq \varepsilon x$，则 $x \approx_{2\varepsilon} y$。若 $x \leq (1+\varepsilon)y$ 且 $y \leq (1+\varepsilon)x$，则 $x \approx_\varepsilon y$。

2.1. 问题陈述

2.1.1. 圆拟合问题

为了引出本文的内容，我们首先考虑如何将圆拟合到平面上的点集这一问题。

设 $P = \{p_1, \ldots, p_n\}$ 为平面上的 n 个点组成的点集，并考虑将圆 C 拟合到 P 的代价 $\nu_P(C)$。形式上，对于一个点 $p_i \in P$，令 $f_i(C) = \left| dY p_i c - r \right|$，其中 c 是 C 的圆心，r 是 C 的半径。因此，对于圆心位于 (x, y)、半径为 r 的圆 C，总的代价为

$$\begin{aligned} \nu_P(C) = \nu_P(x, y, r) &= \sum_{i=1}^{n} f_i(C) \\ &= \sum_{i=1}^{n} \left| dY p_i c - r \right| = \sum_{i=1}^{n} \left| \sqrt{(x_i - x)^2 + (y_i - y)^2} - r \right|, \end{aligned}$$

其中 $p_i = (x_i, y_i)$，对于 $i = 1, \ldots, n$。我们要寻找使 C 最小的圆 C^{ν_P}。这是在 L_1 度量下最能拟合点集的圆。令 $\nu_{\text{opt}}(P)$ 表示最优圆 C_{opt} 的代价。

从几何上看，每个函数 f_i 都会导出一个曲面

$$\gamma_i = \left\{ (x_p, y_p, \|p - p_i\|) \;\middle|\; p \in \mathbb{R}^2 \right\}$$

在三维空间中，它是一个圆锥。这里，一个圆由三维点 $C = (x, y, r)$。编码。$f_i(C)$ 的值是点 C 与曲面 γ_i 之间的垂直距离。因此，我们有一个由 n 个三维曲面组成的集合 $\mathcal{G}$，并且我们希望找到一个点，使该点到这些 n 曲面的垂直距离之和最小。

$$\text{(i)} \qquad \min_{(x_1,\ldots,x_{d-1})\in\mathcal{D}} \sum_{i=1}^{n} w_i \cdot f_i(x_1,\ldots,x_{d-1}),$$

$$\text{(ii)} \qquad \nu_{\text{opt}}(\mathcal{G}) = \min_{x\in\mathcal{D}_d} \nu_{\mathcal{G}}(x) = \min_{(x_1,\ldots,x_d)\in\mathcal{D}_d} \sum_{i} w_i \cdot |f_i(x_1,\ldots,x_{d-1}) - x_d|,$$

$$\text{or (iii)} \quad \mu_{\text{opt}}(\mathcal{G}) = \min_{x\in\mathcal{D}_d} \mu_{\mathcal{G}}(x) = \min_{(x_1,\ldots,x_d)\in\mathcal{D}_d} \sum_{i} w_i \cdot (f_i(x_1,\ldots,x_{d-1}) - x_d)^2,$$

其中，$\mathcal{D} \subseteq \mathbb{R}^{d-1}$ 和 $\mathcal{D}_d \subseteq \mathbb{R}^{d}$ 是常数复杂度的半代数集，权重 $w_1,\ldots,w_n$ 为正整数。注意，通过设置 $\mathcal{D}_d = \mathcal{D} \times \{0\}$，**(i)** 是 **(ii)** 的特例。

为简化表述，我们假设 $\mathcal{D}_d = \mathbb{R}^d$。很容易验证，经过少量修改，我们的算法同样适用于更一般的情形。

线性化维数。下文中，一个重要参数是线性化维数$\mathbb{d}$，即我们需要将多项式 $p_1,\ldots,p_n$ 映射到的目标维数，使它们都成为线性函数。例如，若多项式的形式为 $\psi_i(x,y,z) =$

定义 3.1。对于一组加权曲面 $\mathcal{G}$ in $\mathbb{R}^d$，加权子集 $\mathcal{S} \subseteq \mathcal{G}$是 ε-核心集，如果对于任意点 $p \in \mathbb{R}^d$，都有 $\nu_{\mathcal{G}}(p) \approx_{\varepsilon} \nu_{\mathcal{S}}(p)$。

为便于阐述，以下假设 $\mathcal{G}$是无权的。加权情形也可以用类似方式处理。

第一步是对这些点进行划分。形式上，我们将$\mathbf{Z}$对称地划分为若干子集，使这些子集的大小从集合两端向中间逐渐增大。具体而言，集合 $L_i = \{z_i\}$ 包含直线上的第 i 个点，其中 $i = 1, \ldots, \mathrm{m}$，$\mathrm{m} \geq 10/\varepsilon$ 是一个稍后确定的参数。类似地，$R_i = \{z_{n-i+1}\}$，其中 $i = 1, \ldots, \mathrm{m}$。令 $\alpha_{\mathsf{m}} = \mathrm{m}$，并令 $\alpha_{i+1} = \min(\lceil (1+\varepsilon/10)\alpha_i \rceil, n/2)$，其中 $i = \mathrm{m} \ldots, \mathrm{M}$，$\alpha_{\mathsf{M}}$ 是这个序列中第一个等于 $n/2$ 的数。现在，令 $L_i = \left\{z_{\alpha_{i-1}+1}, \ldots, z_{\alpha_i}\right\}$，并令 $R_i = \left\{z_{n-\alpha_{i-1}}, \ldots, z_{n-\alpha_i+1}\right\}$，其中 $i = \mathrm{m} + 1, \ldots, \mathrm{M}$。我们将把一个

集合 L_i 或 R_i 称为一个块。考虑由这些块$L_1, L_2, \ldots, L_{\mathsf{M}}, R_{\mathsf{M}}, \ldots, R_2, R_1$构成的$\mathbf{Z}$ 的划分。这是将$\mathbf{Z}$ 划分为“指数集合”。边界两侧最前/最后的 m 个集合是单元素集合，其余所有集合的基数则呈指数增长，直至覆盖整个$\mathbf{Z}$。

接下来，我们任意选取点 $l_i \in L_i$ 和 $r_i \in R_i$，并为它们赋予权重$w_i = |R_i| = |L_i|$，其中 $i = 1$，.....，M。令 $\mathcal{S}$ 为由此得到的加权点集。我们声称这是 1- 中位数函数的一个核心集。

不过，在深入讨论之前，我们需要以下技术引理。

引理 3.2。 设 A 是一个由 n 个实数组成的集合， ψ 和 $\mathbf{z}$ 是任意两个实数。则有 $\left|\nu_{A}(\mathbf{z}) - |A| \cdot |\psi - \mathbf{z}|\right| \leq \nu_{A}(\psi)$。

证明：有

$$\begin{aligned}\left|\nu_{A}(\overline{\mathbf{z}}) - |A| \cdot |\psi - \overline{\mathbf{z}}|\right| &= \left|\sum_{p \in A} |p - \overline{\mathbf{z}}| - |A| \cdot |\psi - \overline{\mathbf{z}}|\right| \\ &\leq \sum_{p \in A} \left||\overline{\mathbf{z}} - p| - |\psi - \overline{\mathbf{z}}|\right| \leq \sum_{p \in A} |p - \psi| = \nu_{A}(\psi),\end{aligned}$$

根据三角不等式。 ∎

引理 3.3。 对于任意 $\mathbf{z} \in \mathbb{R}$，都有 $\nu_{\mathbf{Z}}(\mathbf{z}) \overset{\varepsilon/5}{\approx} \nu_{\mathcal{S}}(\overline{\mathbf{z}})$。

证明：我们断言

$$\left|\nu_{\mathbf{Z}}(\overline{\mathbf{z}}) - \nu_{\mathcal{S}}(\overline{\mathbf{z}})\right| \leq (\varepsilon/10)\nu_{\mathbf{Z}}(\overline{\mathbf{z}}),$$

对所有$\overline{\mathbf{z}} \in \mathbb{R}$均成立。事实上，令 τ 为$\mathbf{Z}$ 的中位点，并注意到$\nu_{\mathbf{Z}}(\tau)$是该函数的全局最小值。我们有

$$\begin{aligned}\mathcal{E} = \left|\nu_{\mathbf{Z}}(\overline{\mathbf{z}}) - \nu_{\mathcal{S}}(\overline{\mathbf{z}})\right| &\leq \sum_{i=1}^{\mathsf{M}} \left|\nu_{L_i}(\overline{\mathbf{z}}) - |L_i| \cdot |l_i - \overline{\mathbf{z}}|\right| + \sum_{i=1}^{\mathsf{M}} \left|\nu_{R_i}(\overline{\mathbf{z}}) - |R_i| \cdot |r_i - \overline{\mathbf{z}}|\right| \\ &= \sum_{i=\mathsf{m}+1}^{\mathsf{M}} \left|\nu_{L_i}(\overline{\mathbf{z}}) - |L_i| \cdot |l_i - \overline{\mathbf{z}}|\right| + \sum_{i=\mathsf{m}+1}^{\mathsf{M}} \left|\nu_{R_i}(\overline{\mathbf{z}}) - |R_i| \cdot |r_i - \overline{\mathbf{z}}|\right| \\ &\leq \sum_{i=\mathsf{m}+1}^{\mathsf{M}} \nu_{L_i}(l_i) + \sum_{i=\mathsf{m}+1}^{\mathsf{M}} \nu_{R_i}(r_i),\end{aligned}$$

这尤其意味着，对于任意 $\mathbf{z} \in \mathbb{R}, \mu_{\mathbf{Z}}(\mathbf{z}) = \sum_{x \in \mathbf{Z}} |\mathbf{z} - x|^2$。在这种情况下，甚至修改后的核心集 $\mathcal{R}$ 仍然是一个核心集。

引理 3.6。对于任意 $\mathbf{z} \in \mathbb{R}$，有 $\mu_{\mathbf{Z}}(\mathbf{z}) \overset{\varepsilon}{\approx} \mu_{\mathcal{R}}(\mathbf{z})$。也就是说，$\mathcal{R}$ 是 $\mathbf{Z}$ 关于 $\mu(\cdot)$ 函数的一个 ε-核心集。

证明：注意，根据上述讨论，$\mu_{\mathbf{Z}}(\mathbf{z}) \overset{\varepsilon/5}{\approx} \mu_{\mathcal{S}}(\mathbf{z})$。另一方面，固定 $\mathbf{z} \in \mathbb{R}$，并假设 $|l_i - \mathbf{z}| < |r_i - \mathbf{z}|$。这意味着

沿着这一方向推进的一大难点在于，我们选取的层具有较高的描述复杂度。我们分两个阶段来绕过这一难点。第一阶段，我们通过随机采样，将这些层替换为浅层。第二阶段，我们对这些浅层进行近似，使其引入的相对误差较小。

定义 4.1. 对于 $\mathbb{R}^d$中的一组 $\mathcal{G}$ n 曲面，点 $x \in \mathbb{R}^d$ 在排列 $\mathcal{A}(\mathcal{G})$中的层数，是在垂直方向上位于x下方的 $\mathcal{G}$ 曲面的数量。对于 $k = 0, \ldots, n-1$，令$\mathbf{L}_{\mathcal{G},k}$ 表示 $\mathcal{G}$ 中层数为 k的曲面上的所有点的闭包所构成的曲面。我们将$\mathbf{L}_{\mathcal{G},k}$ 称为

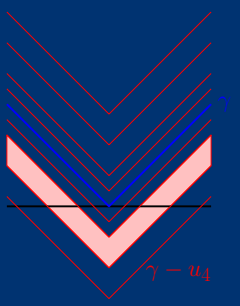
实现细节。 为了实现线性运行时间，我们需要谨慎地实现上述算法。首先，注意我们计算了 $O(\varepsilon^{-1}\log n)$个随机样本 Ψ_{m+1}、……、 Ψ_{M}。 注意，如果两个随机样本是通过以相近的概率（相差不超过两倍）对每个曲面进行采样而生成的，那么我们可以使用同一个随机样本。因此，我们只需为二的幂次概率生成随机样本（这意味着只需要 $O(\log n)$个随机样本）。具体而言，令Υ_i 为一个随机样本，其生成方式是以概率$1/2^i$对 $\mathcal{S}$ 中的每个曲面进行抽样。

为快速执行这一采样，我们通过以二分之一的概率将 Υ_i 的每个曲面选入 Υ_{i+1} ，来生成第 $(i+1)$ 个随机样本（随机样本序列 $\mathcal{S} = \Upsilon_0$、 Υ_1、…… $\Upsilon_{O(\log n)}$ 有时称为渐变序列）。也就是说，每个 Υ_i 都可替代一系列随机样本 Ψ_α、 $\ldots\Psi_\beta$；这些样本是以相似的概率生成的，其中 α和 β 是 i的函数。

接下来，我们需要近似 Ψ_i 的“浅层”，直至第

基本思路是用与 $\gamma \in \mathcal{H}$ 相关、出现在 $\nu_{\mathcal{G}}(p)$ 中的“难处理”距离函数 $\mathrm{d}(p, \gamma)$ 的等值集来替代它。具体而言，对于 $\nu_{\mathcal{G}}(p)$ 求和中的每一项，我们都将生成若干等值集，这样就无需计算 $\mathrm{d}(p, \gamma)$，而只需使用相应的等值集值。略微不严谨地说，$\mathrm{d}(p, \gamma)$ 的等值集是一个曲面，而两个相邻等值集之间区域所对应的值，是较高等值集上的 $\mathrm{d}(\cdot, \gamma)$ 值。这个过程有些类似于绘制地形图时使用的等高线。由于每个等值集都是一个曲面，因此会形成曲面排列。现在，对于任意点 $p \in \mathbb{R}^d$，只需确定 p 位于哪些等值集之间，就可以计算 $\nu_{\mathcal{G}}(p)$。因此，计算 $\nu_{\mathcal{G}}(p)$ 可归结为在曲面排列中执行点定位查询，并返回包含 p 的面所对应的值。

注意，$\|\sigma\|/2$ 是该棱柱中$\nu_{\mathcal{H}}(\cdot)$取值的下界，而 $W \cdot \|\sigma\|$ 是 $\nu_{\mathrm{opt}} = \min_{p \in} \nu_{\mathcal{H}}(p)$取值的上界，其中 $W = w_{\mathcal{H}} \geq n$，$w_{\mathcal{H}}$ 表示$\mathcal{H}$的曲面的总权重。因此，令 $u = \varepsilon \|\sigma\| / (10W^2)$。接下来，令 $u_i = iu$，对于 $i = 1, \ldots, \mathrm{m} = O(1/\varepsilon)$。令 $u_i = (1 + \varepsilon/20)u_{i-1}$，对于$i = \mathrm{m}+1, \ldots, \mathrm{M} = O(\log_{1+\varepsilon/20} W) = O(\varepsilon^{-1} \log W)$。注意，$u_{\mathrm{M}} > W^2 \|\sigma\| > 2\nu_{\mathrm{opt}}$。因此，如果点 $p \in$ 到 $\mathcal{H}$中任意曲面的距离都大于 u_{M}，那么它到这个单独曲面的距离就足以确保



th在 $\nu_{\mathcal{H}}(p) > 2\nu_{\mathrm{opt}}$。特别是，对于每个曲面 $\gamma \in \mathcal{H}$，我们创建 $O(\mathrm{M})$

个副本，如下所示： $\gamma - u_{\mathsf{M}}$， $\gamma - u_{\mathsf{M}-1}$，……， $\gamma - u_1$, γ, $\gamma + u_1$，……， $\gamma + u_{\mathsf{M}}$，其中 $\gamma + x$ 是将曲面 γ 向上平移距离 x 后得到的结果。令 $\mathcal{I}(\gamma)$ 表示由此得到的“堆叠”（即曲面集合）。

$\mathcal{I}(\gamma)$ 的曲面将棱柱（的相关定义域）划分为若干区域，在这些区域中，函数 $\mathrm{d}(p, \gamma)$ 的值在相差一个（$1+\varepsilon/20$）乘法因子的意义下相同。唯一不符合这一条件的区域是 $\gamma - u_1$ 与 $\gamma + u_1$ 之间的区域。因此，如果我们用堆叠中紧邻 p 上方曲面上的该函数值来近似 $\mathrm{d}(p, \gamma)$ 的值，那么除 $\gamma - u_1$ 与 $\gamma + u_1$ 之间的区域外，我们得到该函数的（$1 \pm \varepsilon/20$）近似。

特别地，令 $\mathcal{I} = \cup_{\gamma \in \mathcal{H}} \mathcal{I}(\gamma)$。考虑任意点 $p \in$ ，并令$a_i = \mathrm{d}(p, \gamma_i)$，其中 $\gamma_i \in \mathcal{H}$，对于 $i = 1$、……、n，满足 $|a_i| \le u_{\mathsf{M}}$。此外，令 b_i 为堆叠 $\mathcal{I}(\gamma_i)$ 中紧邻 p 上方和下方曲面所对应值的最大值。因此，我们有

$$\begin{aligned}
\nu_{\mathcal{H}}(p) \le \overline{\nu_{\mathcal{H}}}(p) &= \sum_i b_i \le \sum_i \Big(u + \Big(1 + \frac{\varepsilon}{20}\Big) a_i\Big) \\
&= n \cdot \frac{\varepsilon \, \|\sigma\|}{10 W^2} + \Big(1 + \frac{\varepsilon}{20}\Big) \sum_i a_i \\
&\le \Big(1 + \frac{\varepsilon}{5}\Big) \nu_{\mathcal{H}}(p),
\end{aligned}$$

因为 $\nu_{\mathcal{H}}(p) \ge \|\sigma\|$。

因此，令 $\mathcal{A}$ 为 $\mathcal{A}(\mathcal{I})$的排列，并计算$\mathcal{A}$ 的每个面 F 上$\nu_{\mathcal{H}}(p)$的值，其中 $p \in \mathrm{F}$。根据上述论证，实现 $\min_{x \in} \nu_{\mathcal{H}}(x)$ 的点 p_{opt} 可由 $\nu_{\mathcal{H}}(p_{\mathrm{opt}})$ 正确近似。因此， $\mathcal{A}$ 中关联值$\nu_{\mathcal{H}}$ 最低的面的任意内部点，都是所需的近似点。

很容易验证，稍作修改的同一算法也能使我们近似均值函数 $\mu_{\mathcal{G}}(\cdot)$ 的最小值。由此可得：

定理 5.1。设 $\mathcal{G}$ 是 $\mathbb{R}^d$中的一组 n 加权曲面片，其线性化维数为 d，且任意垂直线与曲面的相交部分总权重为 W。设 $\varepsilon > 0$ 为一个参数。那么，可以在$O\big(n^{3\mathrm{d}+1} \varepsilon^{-\mathrm{d}} \log^{\mathrm{d}} W\big)$ 时间内计算出一个点 $x \in \mathbb{R}^d$，使得 $\nu_{\mathcal{G}}(x) \le (1+\varepsilon)\nu_{\mathrm{opt}}(\mathcal{G})$，其中 $\nu_{\mathrm{opt}}(\mathcal{G}) = \min^{x \in \mathbb{R}^d} \nu_{\mathcal{G}}(x)$。

还可以用相同的时间复杂度计算出一个点 $y \in \mathbb{R}^d$，使得 $\mu_{\mathcal{G}}(y) \le (1+\varepsilon)\mu_{\mathrm{opt}}(\mathcal{G})$，其中 $\mu_{\mathrm{opt}}(\mathcal{G}) = \min^{x \in \mathbb{R}^d} \mu_{\mathcal{G}}(x)$。

6.1.2. 部分数据的1-中位数聚类

考虑输入为 n 个点 p_1，……，p_n，但这些点并未明确给出。取而代之的是，我们得到一组 $\mathfrak{F} = \{\mathfrak{f}_1, \ldots, \mathfrak{f}_n\}$ 个 n 仿射子空间，满足 $p_i \in \mathfrak{f}_i$；其中，仿射子空间是 $\mathbb{R}^d$的一个仿射子空间。当我们只掌握某个点的部分信息，而该点必须满足定义其所在仿射子空间的某些线性约束时，自然会出现这种情况。

现在，自然会想要对部分数据进行最佳拟合或聚类。例如，我们可能希望计算一个能包围所有部分点的最小球。这归结为计算一个与所有仿射子空间都相交的最小球$\mathbf{b}$（也就是说，我们假设真实点 p_i 位于该球的交集中的某处

定理 6.4。设 $\mathfrak{F} = \{\mathsf{f}_1, \ldots, \mathsf{f}_n\}$ 是 $\mathbb{R}^d$ 中的 n 个仿射子空间，且 $\varepsilon > 0$ 是一个参数。可以计算出一个点 $p \in \mathbb{R}^d$，使得 $\nu_{\mathfrak{F}}(p)$ 是 $(1+\varepsilon)$-近似于 $\min^q \nu_{\mathfrak{F}}(q)$。该算法的运行时间为 $O(n + \operatorname{poly}(\log n, 1/\varepsilon))$，且该算法以高概率成功。

请注意，在这种情况下，1-均值聚类是平凡的，因为它归结为最小化一个二次多项式。

7. 结论

在本文中，我们描述了一种通用的近似技术，用于在低维空间中将形状拟合到点集的 L_1 问题。新算法的运行时间为 $O(n + \operatorname{poly}(\log n, 1/\varepsilon))$，对于固定的 ε，其运行时间为线性。多对数项中隐藏的常数次幂大得令人难以启齿，无法明确给出。即便只是将圆拟合到平面中的点集这一问题，L_1 的常数也可能在 20 到 60 之间。换言之，该算法仅具有理论意义。因此，这项工作提出的首个开放问题是改进这些常数。一个有趣得多的问题是，为这类问题开发实用算法。