arXiv:2305.00337v1 [stat.AP] 29 Apr 2023

一种实用且经济的 Gas 价格预测贝叶斯方法

ChihYun Chuang[1] 和 TingFang Lee[2]

[1] AMIS，台湾台北 chihyun@maicoin.com，[2] 罗得岛大学药学院，美国罗得岛 tingfanglee@uri.edu

摘要。 在以太坊网络上，要确定合适的 Gas 价格，既能确保交易在用户要求的时限内被纳入区块，又不会支付过高费用，是一项挑战。解决这一问题的一种方法是使用 Gas 价格预言机，根据历史区块数据推荐 Gas 价格。然而，当交易量迅速增加时，这些预言机往往会低估或高估价格。本文展示了高斯过程模型如何在交易量增加时，预测即将生成的区块中最低价格的分布。这种方法之所以有效，是因为此类过程能够考虑区块之间的时间相关性。我们使用高斯过程模型对历史区块数据进行了实证分析，并将其性能与 GasStation-Express 和 Geth 的 Gas 价格预言机进行了比较。结果表明，当交易量波动很大时，高斯过程模型能够提供更准确的估计。此外，我们还证明，通过使用经过适当预处理且训练样本量较小的数据，可以改进 GasStation-Express 和 Geth。根据实证分析结果，我们推荐一种由高斯过程与 GasStation-Express 组成的混合模型 Gas 价格预言机。该预言机兼具效率、准确性和更低的成本。

关键词： 以太坊、Gas 价格预言机、区块链、高斯过程、贝叶斯

1 引言

在以太坊区块链上，[1]，*Gas* 指执行交易或智能合约所需的燃料[2]。由于链上的每个区块所能包含的 Gas 数量都有上限[3]，，矿工会优先处理 Gas 价格较高的交易，以实现利润最大化。与所有呈现供需动态的市场一样，用户若能预测并提供适当价格，便更有优势：

[3] 2021年3月4日当天，区块的平均燃料上限约为12,500,000个Gas单位。

他们可以预测并报出能够确保交易在预定时间内被纳入区块的最低 Gas 价格。

开发此类 Gas 价格预言机的难点在于交易量的高度波动。在这种情况下，一些现有的 Gas 价格预言机要么低估价格，导致这些交易必须等待很长时间才能被纳入区块；要么高估价格，使用户为交易支付过高费用。

迄今为止，Gas 价格预测的主要方法之一是分析大型内存池中待处理交易的定价结构 [3]。这种方法需要访问大量内存池，以获取足够的待处理交易数据进行分析，因此会消耗大量资源。此外，只有在假设内存池数据准确的前提下，才能准确预测 Gas 价格，而用户很难验证这一点。另一种方法是利用矿工近期纳入区块的交易来推荐价格。基于这一理念的一些算法已用于开发 Gas 价格预言机，包括 Geth、EthGasStation、GasStation-Express（简称 GS-Express），以及 Sam M. Werner 等人的研究 [4,5,6,7,8]。

其中一种 Gas 价格预言机 GS-Express 提议使用最近 200 个区块中的最低 Gas 价格集合。该模型认为，该集合的第 α 个百分位数高于下一个区块中交易的最低 Gas 价格的概率为 α%。换句话说，该集合的第 α 个百分位数有 α% 的概率会被纳入下一个区块。另一个预言机 Geth 使用最近 100 个区块中的最低 Gas 价格集合，并将该集合的第 60 个百分位数作为推荐的 Gas 价格。当待处理交易数量相对较少时，这些模型能够高效地推荐 Gas 价格。然而，当待处理交易激增时，这些模型会低估价格。

本文将重点介绍一种新颖的预测模型，该模型利用近期成功挖出的区块，估算用户为使交易以指定概率得到处理而应出价的最低价格。所提出的方法使用高斯过程（简称 GP）模型，预测即将生成的区块中的最低价格分布。包括高斯过程在内的随机过程常用于研究许多与股票市场微观结构相关的问题，包括价格发现、相关市场之间的竞争、市场参与者的策略行为，以及实时市场动态建模。这些过程可为波动率、时变相关结构、交易量、买卖价差、市场深度、交易成本和流动性风险提供潜在高效的估计器和预测器。以太坊市场中的作用力与这些市场十分相似，因此高斯过程适合用于捕捉以太坊 Gas 价格的动态。随机过程的另一个优势是其协方差函数能够让模型估算区块之间的时间相关性，也就是说，模型可以捕捉时间间隔较短的区块之间更强的相关性。即使交易量激增，我们的方法仍能提供稳定的价格预测。从长期来看，Gas

价格模型建议的价格比现有方法更经济实用[4] 。

我们的贡献包括以下几点：1）我们提出了一种将高斯过程应用于评估定价模型的新方法，并利用该方法研究 GS-Express、Geth 和 GP 的优缺点。研究结果表明，当交易量大幅波动时，GS-Express 和 Geth 会高估或低估价格，而 GP 始终保持了较高的准确性。此外，GP 在模型训练和预测方面具有时间效率。2）我们进行了敏感性分析，以研究训练数据规模对 GS-Express 模型性能的影响。结果表明，将训练数据规模减少到 50 或 30 可以有效提高预测效果。然而，这需要可靠的数据预处理流程。3）我们在算法1中提出了一种实用且经济的 Gas 价格预言机，该预言机保留了 GP 和 GS-Express 的优点，同时避免了各自的缺点。与现有方法相比，在所考虑的区块中，我们的方法在实现目标短期和长期成功率方面更具优势。值得注意的是，除 P_{50}（见表7和表8）外，我们方法的平均成本仍低于其他方法。

本文的结构如下：第 2 节介绍以太坊网络中的交易运作方式以及高斯过程。第 3 节阐述我们的方法，包括数据预处理和建模。第 4 节展示高斯过程预测模型的结果，并与 GS- Express 和 Geth 模型进行比较。最后，第 5 节提出一种结合高斯过程与 GS-Express 的 Gas 价格预言机，并利用二者各自的优势。

2 背景

本节简要概述以太坊网络中的交易以及高斯过程。

2.1 以太坊与 Gas

与其他无需许可的区块链和加密货币一样，以太坊通过一种称为“工作量证明”的密码学零知识证明机制达成共识。在此类协议中，一个名为“矿工”的角色会将交易打包成区块，并将其追加到区块链末尾。这项工作会消耗资源，因此，使用以太坊进行操作需要支付费用，矿工则因完成这项工作而获得费用。矿工根据 Gas 价格决定将哪些交易纳入区块。

交易费以Gas计算，使用$\mathbf{wei}= 10^{-18}$ETH 或$\mathbf{Gwei}= 10^{-9}$ETH 作为计量单位，其中 ETH 是以太坊中的货币。执行成本等于：

$$\text{gas cost} \times \text{gas price}.$$

Here

[4] 高斯过程模型预测新价格平均耗时 0.7 秒。

– 燃料成本的下界为 21,000，上界为**燃料上限**，即用户愿意为某项操作使用的最大 Gas 量。燃料成本的具体数额取决于执行“智能合约”的复杂程度；智能合约使用图灵完备编程语言定义一组规则。交易完成后，所有未使用的 Gas 都会退还至用户账户。如果燃料上限低于燃料成本，则该交易会被视为无效并遭到拒绝；用于执行计算的 Gas 不会退还至账户。

– Gas 价格也由用户决定，表示用户愿意为每单位 Gas 支付的价格。由于矿工的奖励很大程度上取决于 Gas 价格，Gas 价格越高，交易被矿工选中并打包进区块的概率就越大。

2.2 高斯过程

高斯过程是一种随机过程，为对函数分布进行概率推断提供了强大的工具。它提供了灵活的非参数贝叶斯框架，可根据数据估计潜在函数。简而言之，高斯过程能够进行带有不确定性的预测。例如，它会预测下一分钟的股价为 100 美元，标准差为 30 美元。了解预测不确定性对于制定定价策略十分重要。本节其余部分将遵循 [9]，介绍 GP 回归。

定义 1. 高斯过程是随机变量的集合，其中任意有限个随机变量都服从联合高斯分布。

GP 由其均值函数和协方差函数确定，二者决定函数的平滑度和变异性。给定输入向量 $\mathbf{x}$ 和 $\mathbf{x}'$，我们将实值过程$f(\mathbf{x})$的均值函数 $m(\mathbf{x})$和协方差函数 $k(\mathbf{x},\mathbf{x}')$定义为

$$
\begin{aligned}
m(\mathbf{x}) &= E[f(\mathbf{x})] \\
k(\mathbf{x},\mathbf{x}') &= E[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))]
\end{aligned}
$$

并将高斯过程写作

$$
f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x},\mathbf{x}')).
$$

给定训练数据集 $\mathcal{D} = \{(\mathbf{x}_i, y_i)|i = 1, 2, \cdots, n\}$ 其中$\mathbf{x}$ 表示输入向量，y 表示目标变量。可以考虑高斯噪声模型

$$
y_i = f(\mathbf{x}_i) + \mathcal{N}(0, \sigma_n^2).
$$

超参数为 $\theta = \{\sigma_f, l\}$ 的平方指数核，

$$
k(\mathbf{x},\mathbf{x}') = \sigma_f^2 \exp\big(\frac{-|\mathbf{x} - \mathbf{x}'|^2}{2l^2}\big),
$$

被认为是应用最广泛的协方差函数。由于近期 Gas 价格之间的相关性较强，这种协方差函数也适用于我们的模型。

观测目标值与新输入$\mathbf{x}_*$ 处函数值的联合分布为

$$\begin{bmatrix} \mathbf{y} \\ f(\mathbf{x}_*) \end{bmatrix} \sim \mathcal{N}(\mathbf{0}, \begin{bmatrix} K + \sigma_n^2 I & K_*^T \\ K_* & K_{**} \end{bmatrix}), \tag{1}$$

其中，$\mathbf{y} = \begin{bmatrix} y_1 \cdots y_n \end{bmatrix}^T$，符号 T 表示矩阵转置，$K = [k(\mathbf{x}_i, \mathbf{x}j)]_{i,j=1,\cdots,n}$ 为协方差矩阵，$K_* = [k(\mathbf{x}_*, \mathbf{x}_i)]_{i=1,\cdots,n}$，且 $K_{**} = k(\mathbf{x}_*, \mathbf{x}_*)$。因此，后验预测分布为

$$f(\mathbf{x}_*)|\mathcal{D}, \mathbf{x}_*, \theta \sim \mathcal{N}(K_*(K + \sigma_n^2 I)^{-1}\mathbf{y}, K_{**} - (K + \sigma_n^2 I)^{-1}K_*^T). \tag{2}$$

高斯过程回归模型的准确性取决于协方差函数的选择是否恰当。尤其是，估计协方差函数的超参数对预测性能至关重要。拉普拉斯近似框架常用于近似后验预测分布，其构造基于在后验分布最大值处对 $\log\ p(\mathbf{f}|\mathcal{D}, \theta)$ 进行二阶泰勒展开。研究表明，该方法能以短得多的时间提供更精确的估计 [10]。

3 方法

本节将说明数据预处理的步骤。随后，我们将使用处理后的数据拟合 GP 回归模型、Geth 和 GS-Express。第 3.3 节将介绍评估各模型性能的方法。

3.1 预处理

为保持统计显著性，我们移除了交易数量少于 7 笔的区块。此外，有些区块的交易成本异常低。例如，区块 11763787 中有三笔零手续费交易。这类交易很少见，却会给模型带来噪声。因此，我们剔除了所有手续费低于全部 Gas 价格第 2.5 百分位数的交易。处理步骤如下：

1. 选取交易数量超过六笔的区块。2. 计算每个区块的第 2.5 个百分位数，记为 $\mathcal{P}_{2.5}$。3. 删除手续费低于 $\mathcal{P}_{2.5}$的交易。4. 获取每个区块的最低 Gas 价格，记为 $y_$。

3.2 模型

取连续的 n 个区块： b_1， b_2， $\cdots$， b_n，并令训练数据集为 $\mathcal{D} = \{(i,\ y_i) | i = 1,\ 2,\ \cdots,\ n\}$，其中 $y_i :=$ 区块 b_i}. 中的最低 Gas 价格。目标是使用 GP 回归模型预测 b_{n+1}中的 y_{n+1}。

我们考虑高斯噪声模型 $y_i = f(i) + \sigma_n^2$. 使用平方指数协方差函数来估计联合分布（1）中的协方差矩阵。利用后验预测分布（2）来预测第（$n+1$）个区块的最低 Gas 价格的均值 $\hat{y}_{n+1}$ 和标准差 $\hat{s}_{n+1}$；即，

$$f(n+1) \sim \mathcal{N}(\hat{y}_{n+1}, \hat{s}_{n+1}^2).$$

更具体地说，上述估计意味着 $\hat{y}_{n+1}$ 大于第（$n+1$）个区块的最低 Gas 价格 y_{n+1} 的概率为 50%。随后，我们定义 $\hat{y}_{n+1}$ 是第（$n+1$）个区块的 P_{50}。类似地，$\hat{y}_{n+1} + 0.675 \cdot \hat{s}_{n+1}$ 是第（$n+1$）个区块的 P_{75}。（$\hat{y}_{n+1} + \hat{s}_{n+1}$ 是 P_{84}[5]；$\hat{y}_{n+1} + 1.645 \cdot \hat{s}_{n+1}$ 是 P_{95}）。

3.3 模型评估

本节将介绍一种模型比较准则——逆概率权重（IPW），用于比较模型在准确性和效率方面的表现。我们通过以下步骤比较 GP 回归模型、GS-Express 和 Geth：

I. 给定 n 个连续区块 b_1， b_2， $\cdots$， b_n 和 $0 < \alpha < 1$，我们使用 GP 回归模型、GS-Express 和 Geth 来预测 P_α。随后，将预测的 P_α 与实际的 y_{n+1}进行比较。若 $P_\alpha \geq y_{n+1}$，则认为该交易已成功纳入区块 b_{n+1}。定义

$$T_{n+1}(\alpha) = \begin{cases} 1, \text{ if } P_\alpha \geq y_{n+1}, \\ 0, \text{ otherwise.} \end{cases}$$

II. 迭代模型拟合可得到 T_{n+1}， T_{n+2}， $\cdots$，依此类推。定义区块 b_s， b_{s+1}， $\cdots$， b_{s+t-1} 中的成功率如下：

$$R_{s,t,n}(\alpha) = \frac{\sum_{j=0}^{t-1} T_{s+j}(\alpha)}{t}.$$

其中 n 为训练数据中的观测结果数量。

请注意，函数 $R_{s,\ t,\ n}(\alpha)$ 对于 $0 < \alpha < 1$ 是递增函数。我们使用 $R_{s,\ t,\ n}$，在使用包含 n 个观测结果的训练数据时，观察短期（$t \leq 3n$）和长期（$t \geq 10n$）成功率。值得注意的是

$$T_{n+1}(\alpha) \sim \text{Bernoulli}(\alpha)$$

[5] 更准确地说，应为 $P_{84.13}$

假设存在理想的燃料价格预言机。因此，从长期来看，三种方法——高斯过程模型、GS-Express 和 Geth——使用 P_α 时的成功率应约为 α%（即 $R_{s,\ t,\ n}$ 是 α 的一致估计量）。也就是说，

$$\lim_{m\to\infty} R_{s,m,n}(\alpha) = \alpha\%.$$

当 m 不大时，$R_{s,\ m,\ n}$ 可以反映各方法在短期内的预测性能。效率低下的Gas价格预言机可能导致交易待处理，例如长达50分钟。尽管用户可以从以太坊网络中撤回[6] 待处理交易，但仍需要预言机提供新的价格。因此，Gas价格预言机需要在较短时间内表现良好。

最终目标是预测最低价格，使交易能够被纳入区块。当然，较高的价格可以带来较高的成功率。因此，我们引入一种度量指标——逆概率权重，它能更好地反映定价策略。

$$\mathrm{IPW}_{s,t,n}(\alpha) = \frac{\text{average cost}}{R_{s,t,n}(\alpha)}.$$

换言之，较小的逆概率权重$_{s,\ t,\ n}(\alpha)$代表低成本和高成功率；另一方面，高成本和低成功率则会产生较大的逆概率权重$_{s,\ t,\ n}(\alpha)$。我们将在下一节使用这一度量指标评估模型性能。

4 实证分析

所有分析均使用配备英特尔® 至强® 处理器 E3-1505M v5 2.80 GHz CPU 和 32 GB 内存的笔记本电脑完成。我们使用 Mathematica 12，指定平方指数协方差函数，进行高斯过程回归模型拟合。我们使用 SageMath、GS-Express 和 Geth 执行模型拟合与预测。所有使用的数据均可在 [11] 中找到。

4.1 观测结果

分析中使用了从区块编号 11753792 到 11823790[7], 的历史区块数据。其中，有 1450 个区块没有交易，另有 4 个区块的交易少于 7 笔。我们首先移除了这些区块，然后按照第 3.1 节中的步骤 2、3 和 4 处理剩余区块。对每个区块进行预处理后，共有 68,545 个区块。我们将这些区块记为 b_1、b_2、$\cdots$、b_{68545}，例如，b_1 的区块编号为 11753792，而 b_{201} 的区块编号为 11753994。以连续的 200 个区块作为训练数据，我们用训练数据拟合高斯过程模型和 GS-Express。Geth 仅使用 100 个训练数据。每个模型都将用于预测下一个区块的最低 Gas 价格。换言之，我们将三个模型分别拟合到

[6] 使用相同的随机数并提高 Gas 价格。[7] 2021年1月29日至2021年2月9日

第 1 至第 200 个区块，然后预测第 201 个区块最低 Gas 价格的 P_{50}、 P_{75}、 P_{84} 和 P_{95} 。接下来，我们将模型拟合到第 2 至第 201 个区块，然后预测第 202 个区块最低 Gas 价格的 P_{50}、 P_{75}、 P_{84} 和 P_{95} ，依此类推。所得的 P_α 值将用于与真实最低 Gas 价格进行比较。按照第 3.3 节中的 I. 和 II.，我们将展示各模型的成功率，并比较其性能。

理想情况下，使用 P_α 与真实的最低价格进行比较时，成功率应约为 α%。我们发现，GS-Express 和 Geth 在 P_{50} 上的长期成功率$R_{201,\ 68345,\ 200}$（表 1）约为 0.5，而高斯过程为 0.36。GS-Express 和 Geth 在 P_{75}、 P_{84} 和 P_{95} 上低估了价格，而高斯过程在 P_{75}、 P_{84} 和 P_{95} 上给出的价格相对准确。各方法的平均成本，

$$\text{the average of predicted price} \cdot 10^{-9},$$

也列于表 1 中。

表 1：高斯过程和 GS-Express 的长期成功率$R_{201,\ 68545,\ 200}$（α），以及 Geth 的长期成功率$R_{201,\ 68545,\ 100}$（α）。表格右侧列出了各方法在 $\alpha = 50$，75，84，95 时的平均成本。对应的逆概率权重 IPW$_{201,68545,200}$（α）也一并列出。

	成功率				平均成本（Gwei）			
方法	P_{50}	P_{75}	P_{84}	P_{95}	P_{50}	P_{75}	P_{84}	P_{95}
GP	0.358	0.744	0.862	0.972	128.6	168.3	187.4	225.3
GS-Express (		0.696	0.784	0.914	141.6	168.0	178.2	197.1
Geth	0.500	0.712	0.798	0.922	143.1	164.7	174.1	193.3

	逆概率权重			
方法	P_{50}	P_{75}	P_{84}	P_{95}
GP	359.22	226.22	217.40	231.79
GS-Express	282.07	241.38	227.30	215.65
Geth	286.20	231.32	218.17	209.65

注：平均成本是预测价格 $\cdot\ 10^{-9}$的平均值。

我们还计算了最低短期成功率

$$\min\{R_{s,m,200}(\alpha) \mid s = 1, 2, \cdots, 68345\},$$

并在表2中报告了最低成功率。按矿工分组时，连续$m = 25$、50、100 个区块的成功率分别被视为快速、一般和缓慢。从表 1 和表2中，我们观察到，高斯过程在长期使用 P_{75}、 P_{84} 和 P_{95} 时表现更好，并且在短期内也能保持较高的成功率 $m = 25$、50、100。

现在我们关注各方法的 P_{75} 成功率。高斯过程的成功率为 0.744，而 GS-Express 和 Geth 的成功率分别为 0.696 和 0.712。我们推导出了一个 α

表2：GP 和 GS-Express 的最低成功率 $R_{s,\ m,\ 200}(\alpha)$，以及 Geth 在 $\alpha = 50$，75，84，95 条件下的最低成功率 $R_{s,\ m,\ 100}(\alpha)$。我们使用 m =25，50，100 来表示交易被纳入区块的速度分别为快、一般和慢。

	GP				GS-Express				Geth			
m	P_{50}	P_{75}	P_{84}	P_{95}	P_{50}	P_{75}	P_{84}	P_{95}	P_{50}	P_{75}	P_{84}	P_{95}
25	0	0.12	0.24	0.36	0	0	0	0.12	0	0	0	0.24
50	0.04	0.20	0.32	0.54	0.02	0.06	0.06	0.18	0.04	0.06	0.08	0.40
100	0.09	0.33	0.42	0.71	0.07	0.12	0.16	0.32	0.10	0.19	0.23	0.50

使得来自高斯过程的 P_α 具有相当的成功率，但成本低于 GS-Express 和 Geth，见表 3。因此，从长期来看，高斯过程在成本和成功率方面具有优势。不过，在使用较少训练数据时，GS-Express 的表现可能更好。

表3：GP 预测的最低短期成功率 $R_{s,\ m,\ 200}$、长期成功率$R_{201,\ 68545,\ 200}$，以及平均成本 $P_{72.24}$。

短期成功率		长期	
$m = 25$	0.12	成功率	0.712
$m = 50$	0.18	平均成本 164.16 Gwei	
$m = 100$	0.30	IPW	230.56

4.2 更多观察

在区块 11903793 至 11917694[8]期间，Gas 价格出现了大幅波动。我们也针对这些区块的交易数据提供了类似的结果。根据之前的观测结果，我们将 GS-Express 的训练数据点减少到 50 或 30。对这些近期区块数据进行预处理后，我们得到 13627 个区块，记为 $b_1, b_2, \cdots, b_{13627}$，例如，$b_1$ 的区块编号为 11903793，而 b_{201} 为 11903999。GP 的长期成功率 $R_{201,13627,200}$ 以及 GS-Express 的长期成功率 $R_{201,13627,50}$ 和$R_{201,13627,30}$ 及其平均成本见表 4。各模型的最低短期成功率（$m = 25, 50, 100$）见表 5。结果与第 4.1 节中的观测结果一致。

同样，我们发现，GP 给出的预测 $P_{69.67}$ 与其他方法的成功率相当，但成本更低，见表6。

[8] 2021年2月22日至2021年2月24日

表 4：使用 GS-Express 时，GP预测的长期成功率 $R_{201,\ 13627,\ 200}(\alpha)$、$R_{201,\ 13427,\ 30}(\alpha)$ 和 $R_{201,\ 13427,\ 50}(\alpha)$。以及每种方法的平均成本分别为 $\alpha = 50$，75，84，95。还报告了相应的逆概率权重 IPW$_{201,68545,200}(\alpha)$。

	成功率				平均成本（Gwei）			
方法	P_{50}	P_{75}	P_{84}	P_{95}	P_{50}	P_{75}	P_{84}	P_{95}
GP	0.375	0.79	0.911	0.982	248.5	314.5	346.3	409.4
GS-Express (30) (		0.70	0.786	0.882	270.9	297.0	312.4	331.8
GS-Express (50)	0.506	0.70	0.786	0.893	270.2	301.1	317.5	341.0

	逆概率权重			
方法	P_{50}	P_{75}	P_{84}	P_{95}
GP	662.67	398.10	380.13	416.60
GS-Express (30)	525.00	424.29	397.46	376.19
GS-Express (50)	533.99	430.14	403.94	381.86

表 5：各方法的最低成功率 $R_{s,\ m}(\alpha)$，其中 $\alpha = 50$，75，84，95 表示将交易纳入区块的速度分别为快、平均和慢。

	GP				GS-Express（30）				GS-Express（50）			
m	P_{50}	P_{75}	P_{84}	P_{95}	P_{50}	P_{75}	P_{84}	P_{95}	P_{50}	P_{75}	P_{84}	P_{95}
25	0	0.08	0.24	0.32	0	0	0	0.28	0	0	0	0.16
50	0.04	0.24	0.38	0.52	0.08	0.10	0.14	0.42	0.06	0.08	0.10	0.28
100	0.09	0.30	0.42	0.63	0.12	0.16	0.21	0.52	0.12	0.12	0.16	0.39

5 讨论

我们的研究旨在开发一种 Gas 价格预言机，确保交易能在用户要求的时间范围内被纳入区块，同时避免支付过高费用。所提出的 GP 回归能够高效地预测 Gas 价格，尤其是在交易量快速增加时。

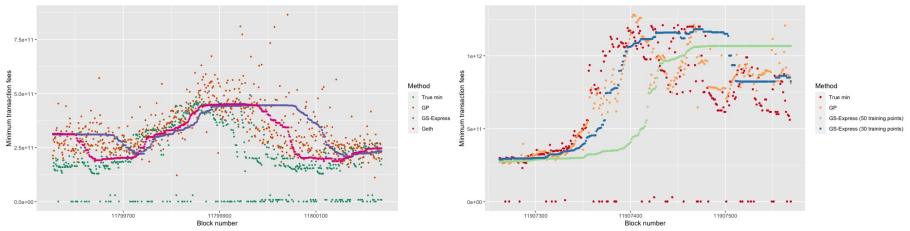
使用不同数量的数据（例如 200、100、50 或 30）时，GS-Express 模型在区块 11799554 至 11799759 以及 11907262 至 11907569 期间表现不佳。图 1 比较了各方法对每个区块的真实最低 Gas 价格的预测 P_{75} 。Gas 价格上涨时，GS-Express 往往会低估价格，导致交易一直处于待处理状态，直到价格趋于稳定或回落。此外，价格刚从峰值下跌时，GS-Express 会高估价格，导致用户多付费用。

当待处理交易量保持稳定时，如果减少训练数据观测结果的数量（从 200 个减少到 50 个或 30 个），GS-Express 的表现会更好。减少训练数据可能存在风险，因为异常交易（例如零手续费交易）可能会为模型带来更多噪声。预处理——

表6：最低短期成功率 $R_{s,\ m,\ 200}$，长期成功率R，以及使用高斯过程预测的平均成本 $P_{69.67}$。

短期成功率		长期	
$m = 25$	0.08	成功率	0.701
$m = 50$	0.20	平均成本	298.85 Gwei
$m = 100$	0.26	IPW	426.32

图1：各区块中真实最低 Gas 价格与各方法对 P_{75} 的预测结果对比。此处展示了区块编号 11799554 至 11800236（顶部）以及区块编号 11907262 至 11907569（底部）。



（a）区块编号：11799554-11800236　　（b）区块编号：11907262-11907569

数据可以有效减少这些异常数据点的影响。因此，我们从实证分析中得出的观测结果如下：

1. 当交易量快速增加时，GP 在超额支付较少的情况下保持了更高的成功率。
2. 当交易量大幅波动时，可以通过减少训练数据来改进 Geth 和 GS-Express 的预测结果。然而，异常交易可能会干扰模型。因此，数据预处理是一个重要步骤。3. 这 3 种方法的长期成功率相当。然而，当 Gas 价格快速上涨时，Geth 和 GS-Express 经常低估 Gas 价格。

除了仅使用高斯过程外，我们还提出以下 Gas 价格预言机，即算法1。它由高斯过程和 GS-Express 组成，并根据即时成功率的变化进行调整。即时成功率 R 可用于监测高斯过程和 GS-Express 估计器的偏差。当 Gas 价格稳定时，训练样本量较小的 GS-Express 表现良好。当 Gas 价格快速上涨时，成功率 R 低于期望值 α，用户可以切换到高斯过程，以维持预期的 α。如果 R 高于 α，则应将值 α 调低。该预言机兼顾效率、成功率，并能降低成本。

算法1 Gas 价格预言机

输入：期望的成功率 α、 $n_{\mathbf{GS}}$（分别为 $n_{\mathbf{GP}}$）、GS-Express（分别为高斯过程）的训练数据规模，以及允许的误差 e 。输出：对编号为 $s+n_{\mathbf{GS}}$ 的区块进行预测。1：执行预处理：i. 选取交易数量超过六笔的区块。ii. 计算每个区块的第 2.5 百分位数，记为 $P_{2.5}$ 。iii. 删除手续费低于 $P_{2.5}$ 的交易。iv. 构建集合 M，其中包含每个区块的最低 Gas 价格。2：令 P_α 为高斯过程（分别为 GS-Express）的预测值 $P_{\mathrm{GP},\alpha}$（分别为 $P_{\mathrm{GS},\alpha}$）。随着 s 递增，GS-Express 的成功率 $R_{s,n_{\mathrm{GS}},{}^{n}\mathrm{GS}(\alpha)}$ 持续更新，执行：a. 情况 $R_{s,n_{\mathrm{GS}},{}^{n}\mathrm{GS}(\alpha)} < \alpha - e$：输出为 $\max\{P_{\mathrm{GP},\alpha}, P_{\mathrm{GS},\alpha}\}$。b. 情况 $\alpha - e \le R_{s,n_{\mathrm{GS}},{}^{n}\mathrm{GS}(\alpha)} \le \alpha + e$：输出为 $P_{\mathrm{GS},\alpha}$。c. 情况 $R_{s,n_{\mathrm{GS}},{}^{n}\mathrm{GS}(\alpha)} > \alpha + e$：使用介值定理求出 α'，使得 $\alpha - e \le R_{s,n_{\mathrm{GS}},{}^{n}\mathrm{GS}(\alpha')} \le \alpha + e$。若不存在这样的 α'，则取 $\alpha' = \alpha$。随后输出 $P_{\mathrm{GS},\alpha'}$。

表7：使用区块 11753792 至 11823790 的数据，比较我们的方法、GS-Express 和 Geth 在 $\alpha = 50$，75，84，95 下的长期成功率、平均成本和逆概率权重（上方两个表格）。下方两个表格报告使用区块 11903793 至 11917694 的数据计算的长期成功率、平均成本和逆概率权重。

	成功率				平均成本（Gwei）			
方法	P_{50}	P_{75}	P_{84}	P_{95}	P_{50}	P_{75}	P_{84}	P_{95}
我们的方法	0.52	0.73	0.81	0.92	143.0	158.4	168.1	189.7
GS-Express	0.50	0.70	0.78	0.91	141.6	168.0	178.2	197.1
Geth	0.50	0.71	0.80	0.92	143.1	164.7	174.1	193.3

	逆概率权重			
方法	P_{50}	P_{75}	P_{84}	P_{95}
我们的方法	275.00	216.99	207.53	206.20
GS-Express	283.20	240.00	228.46	216.59
Geth	286.20	231.97	217.63	210.11

	成功率				平均成本（Gwei）			
方法	P_{50}	P_{75}	P_{84}	P_{95}	P_{50}	P_{75}	P_{84}	P_{95}
我们的方法	0.52	0.75	0.842	0.92	272.7	303.6	323.5	352.3
GS-Express	0.52	0.70	0.79	0.90	264.8	316.2	335.2	367.8
Geth	0.50	0.70	0.78	0.91	268.8	309.4	324.5	358.7

	逆概率权重			
方法	P_{50}	P_{75}	P_{84}	P_{95}
我们的方法	524.42	404.80	384.20	382.93
GS-Express	509.23	451.71	424.30	408.67
Geth	537.60	442.00	416.03	394.18

注：平均成本是预测价格 $\cdot\, 10^{-9}$的平均值。

为评估所提出的 Gas 价格预言机，我们使用了区块 11799554 至 11800236 以及 11903999 至 11917694 的 Gas 价格。区块 11799554 至 11800236 的 Gas 价格逐渐变化，而区块 11903999 至 11917694 的 Gas 价格则快速变化。我们在预言机中采用了 $n_{\mathrm{GS}} = 30$、$n_{\mathrm{GP}} = 200$ 和 $e = 0.1$。在大多数情况下，我们的 Gas 价格预言机具有较低的逆概率加权。结果表明，我们的预言机在保持所需短期成功率的同时，成本更低、长期成功率更高，见表7和表8。

未来的潜在工作包括在实时环境中实践所提出的燃气价格预测程序，并验证该程序的效率和准确性。这也将有助于深入研究短期成功率和实际等待时间。此外，还应仔细研究 GP 回归模型所采用的不同协方差函数，以确定哪种协方差函数最适合预测 Gas 价格。

表8：基于区块11753792至11823790和11903793至11917694的数据，各方法在 $m = 25$, 50，100下的短期成功率。

	区块11753792至11823790								
	我们的方法			GS-Express			Geth		
m	25	50	100	25	50	100	25	50	100
P_{50}	0.04	0.08	0.18	0	0.02	0.07	0	0.04	0.10
P_{75}	0.12	0.22	0.34	0	0.06	0.12	0	0.06	0.19
P_{84}	0.20	0.38	0.50	0	0.06	0.16	0	0.08	0.23
P_{95}	0.68	0.72	0.81	0.12	0.18	0.32	0.28	0.40	0.50

	区块11903793至11917694								
	我们的方法			GS-Express			Geth		
m	25	50	100	25	50	100	25	50	100
P_{50}	0	0.08	0.17	0	0.04	0.06	0	0.04	0.08
P_{75}	0.12	0.28	0.35	0	0.04	0.10	0	0.06	0.12
P_{84}	0.32	0.46	0.50	0	0.06	0.12	0	0.08	0.12
P_{95}	0.52	0.64	0.73	0	0.08	0.13	0.08	0.14	0.22

致谢

我们谨向所有为本项目作出贡献的 AMIS 数据管理团队成员和参与者表示感谢。也感谢 Yu-Te Lin 和 Gavino Puggioni 在早期草稿中提供的宝贵意见与反馈，以及所有审稿人提出的有益评论和建议。

参考文献

1. Ethereum.org。https://ethereum.org/en/，最后访问时间：2021年2月17日

2. Wood, G.：《以太坊：安全的去中心化通用交易账本》，Petersburg 版本 41c1837。以太坊黄皮书（2021年2月） 3. Gas 平台。https://www.blocknative.com/gas，最后访问时间：2021年2月17日 4. GitHub。以太坊协议的官方 Go 实现。https://github.com/ethereum/go-ethereum/，最后访问时间：2021年2月17日 5. EthGasStation。https://ethgasstation.info，最后访问时间：2021年2月17日 6. GitHub。GasStation-Express。https://github.com/ethgasstation/gasstation-express-oracle，最后访问时间：2021年2月17日 7. Antonio Pierro, G., Rocha, H., Tonelli, R., Ducasse, S.：Gas 价格预言机可靠吗？基于 EthGasStation 的案例研究。载于：2020 IEEE 国际区块链导向软件工程研讨会（IWBOSE），第 1–8 页（2020） 8. Werner, S., Pritz, P., Perez, D.：Gas 加速？一种更好的以太坊 Gas 价格推荐方法，第 161–177 页（2020年10月） 9. Rasmussen, C., Williams, C.：《机器学习中的高斯过程》（2005年1月） 10. Rue, H., Martino, S., Chopin, N.：利用积分嵌套拉普拉斯近似对潜在高斯模型进行近似贝叶斯推断。《英国皇家统计学会学报 B 辑》71，319–392（2009年4月） 11. 贝叶斯 Gas 价格预言机。
h
t
t
p
s://
g
i
t
hub.co
m/bayesian-gas-price-oracle/A-Practical-and-Economical-Bayesian-Approach-to-Gas-Price-Prediction，最后访问时间：2021年3月9日