

doi:10.3969/j.issn.1001-4616.2025.05.009

含输入噪声的多类别高斯过程 液体火箭发动机故障诊断

董宝阳¹, 解 晖², 许兆宋², 刘久富²

(1. 郑州理工职业学院信息工程学院, 河南 郑州 451100)

(2. 南京航空航天大学自动化学院, 江苏 南京 211106)

[摘要] 针对传统多类别高斯过程分类算法往往忽略数据受到的噪声污染导致预测准确性降低的问题, 提出一种基于变分推断优化算法的含输入噪声的多类别高斯过程分类算法. 以多类别高斯过程模型作为底层分类器, 在传统模型上引入加性高斯噪声, 使用变分推断方法优化改进后的模型, 近似模型隐变量的后验分布, 并据此进行新的预测. 将含输入噪声的多类别高斯过程分类方法应用到液体火箭发动机的故障分类问题中, 实验证明, 与传统多类别高斯过程分类算法相比, 提出的算法在预测精度上有一定提高, 负似然对数指标有效降低, 改进后的模型与真实后验分布更接近.

[关键词] 多类别高斯过程, 变分推断, 液体火箭发动机, 输入噪声

[中图分类号] TP181 **[文献标志码]** A **[文章编号]** 1001-4616(2025)05-0075-10

Fault Diagnosis of Multi Class Gaussian Process Liquid Rocket Engine with Input Noise

Dong Baoyang¹, Xie Hui², Xu Zhaosong², Liu Jiufu²

(1. School of Information Engineering, Zhengzhou Institute of Technology, Zhengzhou 451100, China)

(2. School of Automation, Nanjing University of Aeronautics and Astronautics, Nanjing 210096, China)

Abstract: A multi class Gaussian process classification algorithm with input noise based on variational inference optimization algorithm is proposed to address the problem that traditional multi class Gaussian process classification algorithms often ignore the noise pollution caused by data and reduce prediction accuracy. The multi class Gaussian process model is used as the underlying classifier, additive Gaussian noise is introduced into the traditional model, and the variational inference method is used to optimize the improved model, approximate the posterior distribution of the model's potential variables, and make new predictions accordingly. The multi class Gaussian process classification method containing input noise is applied to the fault classification problem of liquid rocket engines. Experiments have shown that compared with the traditional multi class Gaussian process classification algorithm, the proposed algorithm is better at predicting. There is a certain improvement in accuracy, the negative logarithmic likelihood index is effectively reduced, and the improved model is closer to the true posterior distribution.

Key words: multi class Gaussian process, variational inference, liquid rocket engine, input noise

多类别分类问题^[1-2]广泛存在于工业、科学等各个领域. 当机器学习算法处理这类问题时, 通常会假设数据集中的数据是真实且不带噪声的. 然而, 这种假设在理论上忽略了测量噪声对数据的影响. 尽管忽略噪声可以在算法设计和迭代时间上带来一些优势, 但如果未考虑的噪声对输入数据的生成过程产生重大影响, 那么多类别分类问题的预测结果可能并非最优. 相反, 将输入数据中的噪声纳入先验知识, 并结合归纳偏差, 算法的预测结果往往会更加出色. 这是因为考虑到噪声可以使算法更具鲁棒性, 能够更好地

收稿日期: 2025-01-06.

基金项目: 国家自然科学基金资助项目(61473144)

通讯作者: 刘久富, 博士, 研究方向: 故障诊断、深度学习. E-mail: liujiufu4@126.com

适应现实世界中数据的变化和不确定性. 因此, 在多类别分类问题中, 考虑噪声并将其纳入算法的设计和训练过程中通常能够取得更优秀的结果.

高斯过程(Gaussian process, GP)^[3-5]是一种在机器学习中常用的概率建模工具, 在贝叶斯框架的基础上给出概率输出, 适用于处理具有不确定性的任务. GP 算法的不确定性能够从目标变量的预测分布来体现, 可能会受到模型中噪声的影响. GP 算法通常被用来解决回归问题或简单的二分类^[6-7]问题, 为使得 GP 算法能够解决多分类问题^[8], 近年来有颇多学者对此进行了相关研究.

在二分类的问题背景下, 文献[9]考虑了为 GP 算法引入相关噪声. 这种方法的核心思想是利用训练阶段的特权信息来更好地建模分类问题中的噪声, 从而提高分类器的性能. 通过使用额外的高斯过程来建模噪声的方差, 可以更准确地捕捉数据中的噪声特性, 有助于提高分类器的性能, 同时表明在模型中引入数据噪声能够给系统带来一定的优势^[10-11].

文献[12]在解决 GP 算法向多类别问题中扩展时, 提出了基于使用伪数据点的稀疏高斯过程方法. 通过引入一组从数据空间中随机采样或其他方式获得的 $M \leq N$ 个伪数据样本, 并使用最大边际似然估计优化伪数据点的位置. GP 算法求解的复杂度为 $\mathcal{O}(N^3)$, 经过伪数据点的优化, 算法的复杂度降低为 $\mathcal{O}(NM^2)$, 因此 GP 算法得以被拓展到大型数据集中. 文献[13]向扩展的 GP 算法中引入变分推断^[14-18], 通过采用近似后验簇来绕过求解边际似然的问题, 方法的局限在于仅能求解无监督学习问题, 且要求高斯似然因子具有容易求解的形式. 文献[19]描述了 GP 隐变量模型的推广, 将近似分布限制在具有固定均值和可调整方差的高斯分布中, 但仍对高斯似然因子有相同的要求, 且无法解决多分类问题.

本文基于传统的稀疏高斯过程多分类方法, 通过添加噪声, 构建含输入噪声的分类模型. 同时, 引入变分推断(Variational Inference)用于对数据噪声和伪数据点的优化, 所提方法记为 VI-MGP. 然后生成仿真数据, 对传统多类别高斯过程分类(multi class Gaussian process, MGP)和 VI-MGP 算法以及竞争算法进行性能对比. 最后进行液体火箭发动机的故障样本实验, 验证改进后算法的性能.

1 含噪声的稀疏多类别高斯过程分类算法

1.1 多类别高斯过程

多类别分类问题的数据模型设定如下: 数据集包含 N 个实例数据 $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)^T$, 与实例数据对应的标签数据 $\mathbf{y} = (y_1, \dots, y_N)$, 其中 $y_i \in \{1, \dots, C\}$, $C > 2$. 式中 C 代表类别的数量. 类别标签根据式(1)中的规则确定:

$$y_i = \operatorname{argmax}_c f^c(\mathbf{x}_i), \tag{1}$$

式中, $f^c(\cdot)$ 是对应于每一个类别标签的隐函数, $c = 1, \dots, C$. 根据在数据点 \mathbf{x}_i 处隐函数的最大值来确定对应的类别标签.

在上述标记规则下, 在数据点 \mathbf{x}_i 处给出的类别标签 y_i 的似然函数为:

$$p(y_i | \mathbf{f}_i) = \prod_{c=y_i} \Theta(f^c(\mathbf{x}_i) - f^c(\mathbf{x}_i)),$$

式中, $\Theta(\cdot)$ 是 Heaviside 阶跃函数.

考虑系统在标记时会产生标记错误, 假定与数据点 \mathbf{x}_i 相对应的类别标签 y_i 以概率 e 被等可能地错误标记为其他的类别, 则似然函数替换为式(2)形式:

$$p(y_i | \mathbf{f}_i) = (1-e) \prod_{c=y_i} \Theta(f^c(\mathbf{x}_i) - f^c(\mathbf{x}_i)) + \frac{e}{C-1} \left[1 - \prod_{c \neq y_i} \Theta(f^c(\mathbf{x}_i) - f^c(\mathbf{x}_i)) \right], \tag{2}$$

式中, 第一项反应了类别标签识别正确的概率, 第二项表示存在 $C-1$ 种识别错误的可能.

高斯过程 \mathcal{GP} 表示不可数无限维函数上的分布, 通常由两个实值函数来定义: 均值函数和协方差函数. 假定每个隐函数 $f^c(\cdot)$ 都具有 \mathcal{GP} 先验, 即隐函数服从于均值函数为 0, 协方差函数为 $k_\theta(\cdot, \cdot)$ 的 \mathcal{GP} 分布: $p(f^c) \sim \mathcal{GP}(0, k_\theta(\cdot, \cdot))$. $k_\theta(\cdot, \cdot)$ 是一个具有超参数 θ_c 的协方差函数, 采用如下定义:

$$k_{\theta_c}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp \left\{ -\frac{1}{2} \sum_{j=1}^d \frac{(x_j - x'_j)^2}{\ell_j} \right\} + I[\mathbf{x} = \mathbf{x}'] \sigma_0^2,$$

式中, $I[\cdot]$ 是一个指示函数, σ^2 是振幅参数, ℓ_j 是长度刻度, σ_0^2 是附加在隐函数 $f^c(\cdot)$ 上的高斯噪声的水平, 三者组成超参数 θ_c .

根据贝叶斯规则, 需要对 $\mathbf{f} = \{\mathbf{f}_i\}_{i=1}^N$ 的后验分布进行计算.

$$p(\mathbf{f}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{f})p(\mathbf{f})}{p(\mathbf{y})} = \frac{\left[\prod_{i=1}^N p(y_i | \mathbf{f}_i) \right] \left[\prod_{c=1}^C p(\mathbf{f}^c) \right]}{p(\mathbf{y})}, \quad (3)$$

式中, $\mathbf{f} = (f^c(\mathbf{x}_1), \dots, f^c(\mathbf{x}_N))^T$, 且隐函数的先验分布 $p(\mathbf{f}^c) = \mathcal{N}(\mathbf{f}^c | 0, \mathbf{K}^c)$ 是一个均值为零, 协方差矩阵为 \mathbf{K}^c 的多变量高斯分布, 且 $K_{i,j}^c = k_{\theta_c}(\mathbf{x}_i, \mathbf{x}_j)$. 分母 $p(\mathbf{y}) = \int p(\mathbf{y}|\mathbf{f})p(\mathbf{f})d\mathbf{f}$, 是后验分布的归一化常数, 即边际似然. 通过最大化边际似然值, 可以获得模型的超参数 θ_c 的最优值. 假定隐函数 $f^c(\cdot)$ 之间相互独立, 因此先验因子为 $p(\mathbf{f}) = \prod_{c=1}^C p(\mathbf{f}^c)$.

1.2 稀疏高斯过程

训练集大小为 N 的高斯过程的后验分布的计算需要计算协方差矩阵的逆, 计算成本为 $\mathcal{O}(N^3)$. 因此, 传统多分类高斯过程只适用于数量级不高的数据集, 而稀疏高斯过程可以降低计算成本. 引入 M 个伪数据点, 记为: $\mathbf{Z}^c = (z_1^c, \dots, z_M^c)$. 伪数据点 \mathbf{Z}^c 与训练数据具有相同的维度空间 \mathbf{R}^d , 通过最大化边际似然估计能够确定其在训练期间的位置. 同时考虑伪数据点 \mathbf{Z}^c 通过相对应的隐函数计算获得的伪数据输出 \mathbf{u}^c , 且 $u_j^c = f^c(z_j^c)$. 在稀疏近似中, 每个 \mathbf{u}^c 服从高斯过程先验, 即 $p(\mathbf{u}^c) = \mathcal{N}(\mathbf{u}^c | 0, \mathbf{K}_{\mathbf{Z}^c, \mathbf{Z}^c}^c)$.

通过 \mathbf{Z}^c 和 \mathbf{u}^c 与对应类别的隐函数 $f^c(\cdot)$ 构成高斯条件概率分布, 可以推出:

$$p(\mathbf{f}^c | \mathbf{u}^c) = \mathcal{N}(\mathbf{f}^c | \mathbf{K}_{\mathbf{X}, \mathbf{Z}^c}^c (\mathbf{K}_{\mathbf{Z}^c, \mathbf{Z}^c}^c)^{-1} \mathbf{u}^c, \mathbf{K}_{\mathbf{X}, \mathbf{X}}^c - \mathbf{K}_{\mathbf{X}, \mathbf{Z}^c}^c (\mathbf{K}_{\mathbf{Z}^c, \mathbf{Z}^c}^c)^{-1} \mathbf{K}_{\mathbf{Z}^c, \mathbf{X}}^c), \quad (4)$$

式中, $\mathbf{K}_{\mathbf{X}, \mathbf{Z}^c}^c$ 是在观测数据点 \mathbf{X} 和伪数据点 \mathbf{Z}^c 处 $N \times M$ 的协方差矩阵, $\mathbf{K}_{\mathbf{Z}^c, \mathbf{Z}^c}^c$ 是 $M \times M$ 的协方差矩阵.

根据式(4)中的描述, 新的似然函数只需要求解 $\mathbf{K}_{\mathbf{Z}^c, \mathbf{Z}^c}^c$ 的逆以及计算乘积 $\mathbf{K}_{\mathbf{X}, \mathbf{Z}^c}^c (\mathbf{K}_{\mathbf{Z}^c, \mathbf{Z}^c}^c)^{-1}$ 两项, 其计算复杂度转变为 $\mathcal{O}(M^2N)$. 如果 M 的数值远小于数据量 N , 则训练的代价将大幅度降低.

1.3 含噪声的多分类高斯过程

传统的多类别稀疏高斯过程分类假定数据不含噪声, 在决策边界的分类预测会过于自信. 因此在模型中引入噪声并建立具有噪声的多分类高斯过程模型.

设定 $\tilde{\mathbf{X}}$ 是噪声观测的数据矩阵, \mathbf{X} 为无噪声输入的数据矩阵, 两者关系如(5)所示:

$$\tilde{\mathbf{x}}_i = \mathbf{x}_i + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \mathbf{V}_i), \quad (5)$$

式中, $\tilde{\mathbf{x}}_i \in \mathbf{R}^d$ 是观测到的具有噪声的数据, \mathbf{x}_i 是不含噪声的真实数据, $\boldsymbol{\varepsilon}_i$ 是附加在真实数据上的高斯噪声. 假定噪声具有独立性, 则 \mathbf{V}_i 是维度 $d \times d$ 的对角噪声矩阵. $\tilde{\mathbf{x}}_i$ 是观测数据, $i = 1, \dots, N$ 其中噪声未知, 实际无噪声输入 $\mathbf{x}_i, i = 1, \dots, N$ 是模型隐变量.

由式(5)可得如下似然函数:

$$p(\tilde{\mathbf{x}}_i | \mathbf{x}_i) = \mathcal{N}(\tilde{\mathbf{x}}_i | \mathbf{x}_i, \mathbf{V}_i).$$

观测到的数据 $\tilde{\mathbf{x}}_i$ 视为均值是实际无噪声输入数据 \mathbf{x}_i , 协方差矩阵为 \mathbf{V}_i 的正态分布. 指定无噪声观测数据 \mathbf{x}_i 具有如下先验分布. 即:

$$p(\mathbf{x}_i) = \mathcal{N}(\mathbf{x}_i | 0, \mathbf{I}_s),$$

式中, \mathbf{I} 是单位矩阵, 当 s 为较大常数时, 该先验分布与无信息均匀分布类似, 通过多次试验验证, 本文采用 $s = 1000$.

模型所有观测变量及隐变量的联合分布如下:

$$p(\mathbf{X}, \tilde{\mathbf{X}}, \mathbf{y}, \mathbf{F}, \mathbf{U}) = \left[\prod_{i=1}^N p(y_i | \mathbf{f}_i) \right] \left[\prod_{c=1}^C p(\mathbf{f}^c | \mathbf{u}^c) p(\mathbf{u}^c) \right] \left[\prod_{i=1}^N p(\tilde{\mathbf{x}}_i | \mathbf{x}_i) p(\mathbf{x}_i) \right], \quad (6)$$

式中, $\tilde{\mathbf{X}} = (\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_N)^T$ 是观测数据矩阵, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)^T$ 是实际无噪声输入的矩阵, \mathbf{y} 是观测标签向量, $\mathbf{F} = (\mathbf{f}_1, \dots, \mathbf{f}_N)^T$ 是实际无噪声输入计算矩阵, $\mathbf{U} = (\mathbf{u}^1, \dots, \mathbf{u}^C)^T$ 是伪数据点的计算矩阵.

2 优化方案

2.1 参数后验分布

为求解隐变量的最优参数,写出式(6)在观测变量已知时的后验分布公式:

$$p(\mathbf{X}, \mathbf{F}, \mathbf{U} | \mathbf{y}, \tilde{\mathbf{X}}) = \frac{p(\tilde{\mathbf{X}}, \mathbf{X}, \mathbf{y}, \mathbf{F}, \mathbf{U})}{p(\mathbf{y}, \tilde{\mathbf{X}})}, \quad (7)$$

后验分布的求解需要计算边际似然 $p(\mathbf{y}, \tilde{\mathbf{X}})$ 的高维积分,在实际情况中,边际似然的值难以求解,并且其计算代价极高,本文采用变分推断通过将推断问题转化为优化问题,以避免边际似然的求解.

2.2 变分推断近似

实际工作中,伪数据点的值 \mathbf{u}^c 和无噪声真实输入数据 \mathbf{x}_i 是未知的,需要为这些变量指定一个近似高斯后验分布. 变分推断通过最小化变分近似 q 与真实后验分布之间的距离来优化参数,距离采用 KL 散度进行计算.

设定变分近似簇为 $q(\mathbf{X}, \mathbf{F}, \mathbf{U})$, 则其与真实后验之间的 KL 散度可以写成:

$$KL = \mathbb{E} \left[\log \frac{q(\mathbf{X}, \mathbf{F}, \mathbf{U})}{p(\mathbf{X}, \mathbf{F}, \mathbf{U} | \mathbf{y}, \tilde{\mathbf{X}})} \right] = \mathbb{E}[\log q(\mathbf{X}, \mathbf{F}, \mathbf{U})] - \mathbb{E}[\log p(\mathbf{X}, \mathbf{F}, \mathbf{U} | \mathbf{y}, \tilde{\mathbf{X}})] = \mathbb{E}[\log q(\mathbf{X}, \mathbf{F}, \mathbf{U})] - \mathbb{E}[\log p(\mathbf{X}, \tilde{\mathbf{X}}, \mathbf{y}, \mathbf{F}, \mathbf{U})] + \log p(\mathbf{y}, \tilde{\mathbf{X}}), \quad (8)$$

式中, $\log p(\mathbf{y}, \tilde{\mathbf{X}})$ 是常数项,要求 KL 散度的值最小等价于等式右侧前两项值最小,其负值称为变分下界,记作 \mathcal{L} :

$$\mathcal{L} = \mathbb{E}_q \left[\log \frac{p(\tilde{\mathbf{X}}, \mathbf{X}, \mathbf{y}, \mathbf{F}, \mathbf{U})}{q(\mathbf{X}, \mathbf{F}, \mathbf{U})} \right]. \quad (9)$$

变分近似 $q(\mathbf{X}, \mathbf{F}, \mathbf{U})$ 包含隐变量 $\mathbf{X}, \mathbf{F}, \mathbf{U}$, 其中 \mathbf{X} 是实际无噪声输入矩阵,与隐变量 \mathbf{F}, \mathbf{U} 相互独立. $q(\mathbf{X}, \mathbf{F}, \mathbf{U})$ 可以分解成 $q(\mathbf{F}, \mathbf{U})q(\mathbf{X})$ 两项. 第一项 $q(\mathbf{F}, \mathbf{U})$ 可以利用式(4)中后验分布 $p(\mathbf{f}^c | \mathbf{u}^c)$, 采用变分推断重新近似为:

$$q(\mathbf{F}, \mathbf{U}) = q(\mathbf{F} | \mathbf{U})q(\mathbf{U}) = p(\mathbf{f}^c | \mathbf{u}^c)q(\mathbf{U}),$$

隐变量 \mathbf{U} 的变分分布选定为高斯后验分布, $q(\mathbf{u}^c) = \mathcal{N}(\mathbf{u}^c | \mathbf{n}_c, \mathbf{S}_c)$, 其中 $c = 1, \dots, C$. 考虑第二项的变分近似 $q(\mathbf{X})$, 其变分函数服从于高斯后验分布 $q(\mathbf{x}_i) = \mathcal{N}(\mathbf{x}_i | \boldsymbol{\mu}_i^x, \mathbf{V}_i^x)$.

$$q(\mathbf{X}, \mathbf{F}, \mathbf{U}) = \left[\prod_{c=1}^C p(\mathbf{f}^c | \mathbf{u}^c) q(\mathbf{u}^c) \right] \left[\prod_{i=1}^N q(\mathbf{x}_i) \right]. \quad (10)$$

结合式(6)、(9)、(10),化简后的损失函数如下:

$$\mathcal{L} = \sum_{i=1}^N \mathbb{E}_q[\log p(y_i | \mathbf{f}_i)] + \sum_{i=1}^N \mathbb{E}_q[\log p(\tilde{\mathbf{x}}_i | \mathbf{x}_i)] - \sum_{c=1}^C KL(q(\mathbf{u}^c) \| p(\mathbf{u}^c)) - \sum_{i=1}^N KL(q(\mathbf{x}_i) \| p(\mathbf{x}_i)), \quad (11)$$

由式(11)可知,损失函数分解为 4 项,前两项分别为两项对数似然函数关于变分分布的期望,后两项是 KL 散度.

式(11)中的第一项期望无法求得解析解,考虑小批量数据近似方法,对原本的第一项期望进行无偏估计,得到如下公式:

$$\sum_{i=1}^N \mathbb{E}_q[\log p(y_i | \mathbf{f}_i)] \approx \frac{\mathcal{B}}{N} \sum_{i \in \mathcal{B}} \mathbb{E}_q[\log p(y_i | \mathbf{f}_i)]. \quad (12)$$

q 分布关于随机变量 \mathbf{U} 的期望可以解析计算:

$$\mathbb{E}_{q(\mathbf{f}_i)}[\log p(y_i | \mathbf{f}_i)] = \frac{e}{C-1} + (1-e) \frac{1}{S} \sum_{i=1}^S \int \mathcal{N}(f^i(\mathbf{x}_i) | m_{y_i}, v_{y_i}) \prod_{c \neq y_i} \Phi\left(\frac{f^i(\mathbf{x}_i) - m_c}{\sqrt{v_c}}\right) df^i(\mathbf{x}_i). \quad (13)$$

$\Phi(\cdot)$ 是标准高斯分布的累计概率,式(13)中

$$m_c = \mathbf{k}_{\mathbf{x}_i, \mathbf{z}^c}^c (\mathbf{K}_{\mathbf{z}^c, \mathbf{z}^c}^c)^{-1} \mathbf{n}_c,$$

$$v_c = k_{\mathbf{x}_i, \mathbf{x}_i}^c - \mathbf{k}_{\mathbf{x}_i, \mathbf{z}^c}^c (\mathbf{K}_{\mathbf{z}^c, \mathbf{z}^c}^c)^{-1} \mathbf{K}_{\mathbf{z}^c, \mathbf{x}_i}^c + \mathbf{k}_{\mathbf{x}_i, \mathbf{z}^c}^c (\mathbf{K}_{\mathbf{z}^c, \mathbf{z}^c}^c)^{-1} \mathbf{S}_c (\mathbf{K}_{\mathbf{z}^c, \mathbf{z}^c}^c)^{-1} \mathbf{k}_{\mathbf{z}^c, \mathbf{x}_i}^c,$$

式中, $k_{\mathbf{x}_i, \mathbf{x}_i}^c$ 是 $f^c(\mathbf{x}_i)$ 的方差, $\mathbf{k}_{\mathbf{x}_i, \mathbf{z}^c}^c$ 是 $f^c(\mathbf{x}_i)$ 和 $f^c(\cdot)$ 在伪数据点 \mathbf{z}^c 的协方差矩阵, $\mathbf{K}_{\mathbf{z}^c, \mathbf{z}^c}^c$ 是 $f^c(\cdot)$ 在点 \mathbf{z}^c 处的

协方差矩阵. 然后使用 Gauss-Hermite 求积法^[20]计算与 f_i 相关的期望.

计算 $\mathbb{E}_q[\log p(y_i | f_i)]$ 关于 $q(\mathbf{x}_i)$ 的期望采用蒙特卡洛近似和重参数化技巧来实现. 使用重参数化技巧分离 $q(\mathbf{x}_i)$ 的随机性和参数的独立性, 将 \mathbf{x}_i 的高斯噪声转移到变量 $\boldsymbol{\varepsilon}$ 中, 得到下式:

$$\begin{aligned}\hat{\mathbf{x}}_i &= \mathbf{L}_i \boldsymbol{\varepsilon} + \boldsymbol{\mu}_i^x, \\ \boldsymbol{\varepsilon} &\sim \mathcal{N}(0, \mathbf{I}),\end{aligned}\quad (14)$$

式中, \mathbf{L}_i 是一个对角阵, 其中每一项包含了 \mathbf{V}_i^x 对角线平方根, $\boldsymbol{\mu}_i^x$ 是 \mathbf{x}_i 的均值.

重新定义 $\hat{f}_i = (f^1(\hat{\mathbf{x}}_i), \dots, f^c(\hat{\mathbf{x}}_i))^T$, 代入上式可以得到:

$$\mathbb{E}_{q(\mathbf{x}_i)}[\mathbb{E}_{q(f_i)}[\log p(y_i | f_i)]] \approx \mathbb{E}_{q(\hat{f}_i)}[\log p(y_i | \hat{f}_i)] \quad (15)$$

式中, 等式右侧是 $\mathbb{E}_q[\log p(y_i | f_i)]$ 关于 $q(\mathbf{x}_i)$ 的期望的无偏估计.

损失函数的第二项通过小批量估计进行计算:

$$\begin{aligned}\sum_{i=1}^N \mathbb{E}_q[\log p(\bar{\mathbf{x}}_i | \mathbf{x}_i)] &\approx \frac{|\mathcal{B}|}{N} \sum_{i \in \mathcal{B}} \mathbb{E}_q[\log p(\bar{\mathbf{x}}_i | \mathbf{x}_i)] = \\ &\frac{|\mathcal{B}|}{N} \sum_{i \in \mathcal{B}} \left[-\frac{d}{2} \log 2\pi - \frac{1}{2} \log |\mathbf{V}_i| - \frac{1}{2} \text{trace}(\mathbf{V}_i(\mathbf{V}_i^x + \boldsymbol{\mu}_i^x (\boldsymbol{\mu}_i^x)^T)) + \bar{\mathbf{x}}_i^T \mathbf{V}_i \boldsymbol{\mu}_i^x - \frac{1}{2} \bar{\mathbf{x}}_i^T \mathbf{V}_i \bar{\mathbf{x}}_i \right],\end{aligned}\quad (16)$$

损失函数的第三项与第四项是两个高斯分布之间的 KL 散度^[21], 公式如下:

$$\sum_{c=1}^C \text{KL}(q(\mathbf{u}^c) | p(\mathbf{u}^c)) = \sum_{c=1}^C \frac{1}{2} \left[\text{trace}((\mathbf{K}_{z^c, z^c}^c)^{-1} \mathbf{S}_c) + \mathbf{n}_c^T (\mathbf{K}_{z^c, z^c}^c)^{-1} \mathbf{n}_c - M + \log \frac{|\mathbf{K}_{z^c, z^c}^c|}{|\mathbf{S}_c|} \right]. \quad (17)$$

第四项计算同式(17):

$$\sum_{i=1}^N \text{KL}(q(\mathbf{x}_i) | p(\mathbf{x}_i)) \approx \frac{|\mathcal{B}|}{N} \sum_{i \in \mathcal{B}} \text{KL}(q(\mathbf{x}_i) | p(\mathbf{x}_i)) = \frac{|\mathcal{B}|}{N} \sum_{i \in \mathcal{B}} \frac{1}{2} \left[\text{trace}(\mathbf{V}_i^x \mathbf{I}_s^{-1}) + (\boldsymbol{\mu}_i^x)^T \boldsymbol{\mu}_i^x s^{-1} - d + \log \frac{s^d}{|\mathbf{V}_i^x|} \right], \quad (18)$$

最后对式(11)损失函数 \mathcal{L} 采用 ADAM 优化器进行迭代计算, 获得变分参数的最优解.

2.3 分类预测

通过使用变分推断方法对含噪声的多分类高斯过程模型进行优化, 可以获得变分近似 q 的参数, 其中包括 $q(\mathbf{u}^c)$ 的参数 $\mathbf{m}_c, \mathbf{S}_c$ 以及 $q(\mathbf{x}_i)$ 的参数 $\boldsymbol{\mu}_i^x, \mathbf{V}_i^x$. 优化后的模型根据新的实例数据 \mathbf{x}_* 预测其结果 y_* , 使用近似后验代替实际后验得到如下预测公式:

$$p(y_* | \mathbf{x}_*) \approx \int p(y_* | f_*) \left[\prod_{c=1}^C p(f_*^c | \mathbf{u}^c) q(\mathbf{u}^c) d\mathbf{u}^c \right] p(\mathbf{x}_* | \bar{\mathbf{x}}_*) d\mathbf{x}_* df_*, \quad (19)$$

式中, $p(\mathbf{x}_* | \bar{\mathbf{x}}_*)$ 是给定含噪声数据 $\bar{\mathbf{x}}_*$ 时真实数据 \mathbf{x}_* (不含噪声) 的后验分布, 该后验是先验与似然的归一化乘积, 因此可以闭式求解且为高斯分布, 即:

$$p(\mathbf{x}_* | \bar{\mathbf{x}}_*) = \frac{p(\bar{\mathbf{x}}_* | \mathbf{x}_*) p(\mathbf{x}_*)}{p(\bar{\mathbf{x}}_*)} = \mathcal{N}(\mathbf{x}_* | \boldsymbol{\mu}_*^x, \mathbf{V}_*^x), \quad (20)$$

式中, $\mathbf{V}_*^x = (\mathbf{V}_*^{-1} + \mathbf{I}_s^{-1})^{-1}$, $\boldsymbol{\mu}_*^x = \mathbf{V}_*^x (\mathbf{V}_*^{-1} \bar{\mathbf{x}}_*)$. \mathbf{V}_* 是在数据点 \mathbf{x}_* 处的具有高斯噪声的方差的角矩阵, 若方差 s 较大, 则实际上 $\mathbf{V}_*^x \approx \mathbf{V}_*$.

式(19)中的积分无法以封闭形式进行求解, 因此借助蒙特卡洛方法, 从 $p(\mathbf{x}_* | \bar{\mathbf{x}}_*)$ 抽取样本, 计算相应的近似积分:

$$p(y_* | \mathbf{x}_*) \approx \frac{1}{S} \sum_{s=1}^S \int p(y_* | f_*^s) \left[\prod_{c=1}^C p(f_*^c | \mathbf{u}^c) q(\mathbf{u}^c) d\mathbf{u}^c \right] df_*^s, \quad (21)$$

式中, S 是抽取样本的数量, $f_*^s = (f^1(\mathbf{x}_*^s), \dots, f^c(\mathbf{x}_*^s))^T$, $f_*^s = f^c(\mathbf{x}_*^s)$, \mathbf{x}_*^s 是抽样生成的第 s 个 \mathbf{x}_* 样本.

式右侧对 f 和 \mathbf{u} 进行积分计算, 与 \mathbf{u} 相关的积分可以进行高斯累积分布进行解析计算, 与 f 有关的积分采用 Gauss-Hermite 求积法进行计算, 预测公式如下:

$$p(y_* | \mathbf{x}_*) \approx \frac{e}{C-1} + \frac{1-e}{S} \sum_{s=1}^S \int \mathcal{N}(f_*^s | m_{y_*}^s, v_{y_*}^s) \prod_{c \neq y_*} \Phi\left(\frac{f_*^c - m_c^s}{\sqrt{v_c^s}}\right) df_*^s, \quad (22)$$

式中, $f_*^s = f^*(\mathbf{x}_*^s)$, $f_*^s = f^c(\mathbf{x}_*^s)$, $\Phi(\cdot)$ 是标准高斯累积概率分布, 且

$$m_c^s = k_{x_{s^*}, z^c}^c (\mathbf{K}_{Z^c, Z^c}^c)^{-1} \mathbf{n}_c,$$

$$v_c^s = k_{x_{s^*}, x_{s^*}}^c - k_{x_{s^*}, z^c}^c (\mathbf{K}_{Z^c, Z^c}^c)^{-1} \mathbf{K}_{Z^c, x_{s^*}}^c + k_{x_{s^*}, z^c}^c (\mathbf{K}_{Z^c, Z^c}^c)^{-1} \mathbf{S}_c (\mathbf{K}_{Z^c, Z^c}^c)^{-1} k_{Z^c, x_{s^*}}^c.$$

使用变分推断方法对含噪声的多分类高斯过程模型进行优化,并根据新的实例数据进行标签预测的伪代码如下:

算法 含输入噪声的多类别高斯过程变分优化算法

输入:含噪声的训练数据 $\tilde{\mathbf{X}} \in \mathbf{R}^{N \times d}$, 类别标签 $y \in \{1, \dots, C\}$, 类别数 C , 噪声协方差矩阵 $\{V_i\}_{i=1}^N$, 伪数据点数量 M , 协方差函数超参数 θ_c , 标签错误率 e , 批量大小 \mathcal{B} , 学习率 η , 迭代次数 T .

输出:测试数据的预测类别 y_* .

1. 初始化伪数据点 Z^c 、变分分布参数和隐变量后验分布;
2. 定义协方差函数 $k_{\theta_c}(\mathbf{x}, \mathbf{x}')$;
3. for t in 1 to T :
 4. 重参数化生成隐变量样本;
 5. 计算隐函数 f 的均值和方差;
 6. 使用 Gauss-Hermite 积分法近似每个数据点的 $\log p(y_i | f_i)$, 根据式计算期望对数似然项;
 7. 根据式和式计算 KL 散度项;
 8. 计算损失函数 \mathcal{L} ;
 9. 使用 ADAM 优化器更新参数 $\theta_c, \mathbf{S}_c, Z^c, \mu_i^x, V_i^x$;
 10. 计算隐变量后验 $p(\mathbf{x}_* | \tilde{\mathbf{x}}_*)$;
 11. 借助蒙特卡洛方法采样生产 S 个样本, 计算每个样本对应各类别 c 的隐函数 f_c^s 的均值和方差;
 12. 采用 Gauss-Hermite 求积计算预测概率 $p(y_* | \mathbf{x}_*)$;
 13. 输出测试数据的预测类别 $y_* = \operatorname{argmax}_c p(y_* | \mathbf{x}_*)$.

3 实例分析与验证

3.1 仿真数据对比实验

通过生成仿真实验数据,对传统 MGP 分类算法和 VI-MGP 算法以及竞争算法进行对比,竞争算法如下:

NN-MGP:该方法使用一个神经网络来计算每个数据实例无噪声属性的后验近似参数,减少近似分布的参数数量,克服了需要为每个数据实例存储参数的局限性,并通过对模型进行正则化提高泛化效果.

FO-MGP:该方法将输入中的噪声传播到高斯过程预测分布的方差中,使用一阶近似来处理输入噪声,在每个输入位置使用高斯过程的局部线性近似,使得输入噪声可以被重新表示为与高斯过程预测均值的梯度平方成正比的输出噪声.

实验主要通过两个方面对结果进行评估:平均误差与平均负对数似然(NLL).平均误差能够直观的评估模型的性能,NLL 作为多类别分类问题中用于衡量预测分布质量的指标,能够反应预测分布与实际分布的接近程度,NLL 值越低,模型更接近真实分布,模型对不同类别的概率建模更准确.

从多类别高斯过程分类方法隐函数中采样生成数据,给采样数据添加高斯噪声获得仿真数据,生成的实验数据类别数量为 3,每类数据样本量为 1 000. 算法采用的高斯过程核函数超参数为 $\sigma^2 = 0.5, \sigma_0^2 = 0, \ell_j = 2, \forall j$, 伪数据点数量设置为 $M = 100$, 训练批量大小设置为 200. 在 NN-MGP 中,神经网络设有 2 个隐藏层,每个隐藏层有 50 个单元.

考虑观测数据在给定噪声的情况下传统 MGP 算法和 VI-MGP 算法以及竞争算法进行对比. 输入数据的给定噪声方差为 $V_i = \{0.1I, 0.25I, 0.5I\}$.

将噪声方差作为给定输入参数的情况下,实验结果如表 1 所示.

当噪声方差作为给定输入参数时,随着方差的数值上升,所有分类算法对应的平均误差与平均 NLL 都有所上升. 在同一噪声水平下,由于决策边界相似,每种方法的平均误差相近,但 VI-MGP 分类算法的平均误差比传统 MGP 算法平均下降了约 11.2%,平均 NLL 只有原 MGP 方法的 1/3,平均下降了约 65.5%,且低于竞争算法 NN-MGP 和 FO-MGP,凸显了准确建立输入噪声模型以获得更好预测分布的优势. 平均 NLL 指标次优的是 NN-MGP,该实验表明使用神经网络计算后验近似参数具有一定优势. 基于一阶近似的

FO-MGP 算法也比传统 MGP 算法有所改进,但改进差异较小。

保持上述实验条件不变,通过最大化式给出的下界来估计输入噪声方差,将噪声方差分别作为学习参数进行实验对比,结果如表 2 所示。

表 1 噪声方差作为给定输入参数时每种方法的平均误差和平均 NLL 实验结果对比

Table 1 Comparison of average error and average NLL experimental results of each method when noise variance is used as a given input parameter

衡量指标	噪声方差	MGP	VI-MGP	NN-MGP	FO-MGP
平均误差	0.1	0.124	0.108	0.109	0.121
	0.25	0.173	0.154	0.171	0.168
	0.5	0.218	0.197	0.215	0.226
平均 NLL	0.1	0.665	0.249	0.273	0.327
	0.25	1.131	0.375	0.380	0.565
	0.5	1.528	0.502	0.526	0.747

表 2 噪声方差作为学习参数时每种方法的平均误差和平均 NLL 实验结果对比

Table 2 Comparison of average error and average NLL experimental results of each method when noise variance is used as a learned parameter

衡量指标	噪声方差	MGP	VI-MGP	NN-MGP	FO-MGP
平均误差	0.1	0.124	0.110	0.112	0.129
	0.25	0.173	0.162	0.173	0.174
	0.5	0.218	0.203	0.221	0.237
平均 NLL	0.1	0.665	0.411	0.443	0.452
	0.25	1.131	0.870	0.896	0.947
	0.5	1.528	0.931	0.927	1.373

引入输入噪声后,将输入噪声的方差作为学习参数进行推断和预测时,所有分类算法对应的平均误差和平均 NLL 都在表 1 实验基础上有所提升。VI-MGP 分类算法的平均误差仍比传统 MGP 算法平均降低约 8.2%,平均负对数似然值仅为传统 MGP 算法的 2/3,平均降低约 33.4%,对比 NN-MGP 和 FO-MGP,VI-MGP 仍是最优方法,仅在给定噪声方差为 $V_i = 0.5I$ 时,NN-MGP 的平均 NLL 低于 VI-MGP,推测是由于神经网络引入了正则化特性,间接提高了模型对噪声的鲁棒性。

仿真实验结果表明,输入数据的噪声方差 $V_i = \{0.1I, 0.25I, 0.5I\}$ 分别作为给定输入参数和学习参数时,相比传统 MGP 算法、NN-MGP 和 FO-MGP 算法,VI-MGP 算法在多类别分类任务中具有更高的分类精确度和预测分布质量。

图 1 和图 2 展示了在噪声方差作为输入参数给定的情况下,平均误差和平均负对数似然的随着训练数据量增大的变化。随着训练数据量的增加,传统的 MGP 算法和 FO-MGP 算法预测模型的负对数似然反而有所上升,而 VI-MGP 算法和 NN-MGP 持续下降,表明训练数据越多,可被算法利用的解释属性越多,模型后验分布拟合效果越好。在平均误差方面,含输入噪声的 VI-MGP 模型与传统 MGP 方法具有类似的下降曲线,在误差范围整体略优于其他模型。

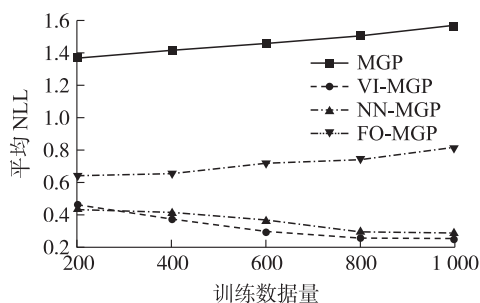


图 1 不同训练数据量对 NLL 的影响
Fig. 1 Influence of different training data sizes on NLL

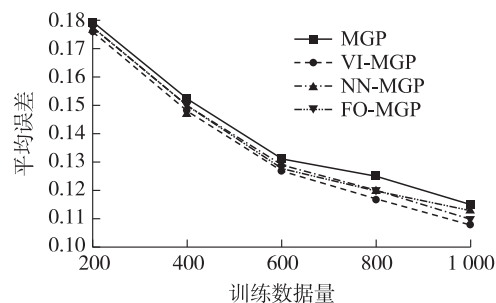


图 2 不同训练数据量对平均误差的影响
Fig. 2 Influence of different training data sizes on average error

3.2 实例分析与验证

本节使用多类别高斯过程分类预测模型对液体火箭发动机故障进行诊断。

3.2.1 数据类别与特征

液体火箭发动机的故障分类按故障发生的阶段可以分为启动过程的故障、稳态过程的故障和关机过程的故障。在稳态工作阶段主要受供应系统的流体及机械运动、燃气发生器与燃烧室能量转换过程的控制,该状态的主要故障包括管路泄露或阻塞、涡轮泵系统故障等。

实验数据取自于液体火箭发动机稳态工作过程试车数据,包括 5 类易发生的管路故障类型以及 21 类与之相关的特征属性。故障类型和特征属性见表 3、4 所示。

表 3 液体火箭发动机稳态故障类型

Table 3 Steady-state fault types of liquid rocket engines

故障类型	标记
氧稳压阀出口泄露	Fault0
氧涡轮入口燃气泄露	Fault1
氧泵后泄露	Fault2
发生器氢副控阀泄露	Fault3
发生器氧副控阀泄露	Fault4

表 4 特征属性与对应符号

Table 4 Feature attributes and corresponding symbols

特征属性	符号	特征属性	符号	特征属性	符号	特征属性	符号
氧泵入口温度	TOHY	发生器氧喷前压力	PFY	燃烧室压力	PK	氢泵出口温度	TER
氢泵入口温度	TOHR	发生器氢喷前压力	PFR	氧泵流量	GY	氧泵出口温度	TEY
氧泵入口压力	POY	氧涡轮入口压力	POWR	氢泵流量	GR	氧贮箱压力	Pg
氢泵入口压力	POR	氢涡轮入口压力	POWF	氧泵转速	NWY		
氧泵出口压力	PEY	燃气发生器压力	PF	氢泵转速	NWR		
氢泵出口压力	PER	燃烧室氧喷前压力	PY	冷却套压力	PEL		

数据集包含 5 种不同管路故障类型下不同时间段的监测数据,故障数据数量为 5 400,训练集取其中的 90%,测试机选取剩下的 10%。对故障数据进行随机排列,重复生成 100 次进行实验。考虑真实数据取值范围较大影响算法收敛,因此观测到的液体火箭发动机的特征属性均进行标准化操作,将数据转化为均值为 0,方差为 1 的标准值。在实验中向原始数据分别注入均值为 0,方差为 0.1、0.25 和 0.5 的噪声,与原始无噪声数据一起形成 4 组对比实验。训练采用的批量大小为 50,迭代上限为 1000 次。伪数据点数量选择总数据量的 5%。

3.2.2 分类测试结果

根据 3.1 节仿真实验的结论可知,本文方法在给定噪声参数的情况下更具有优势。输入噪声的方差为 0.1 时,测得 VI-MGP 算法在液体火箭发动机故障诊断应用中的分类结果如图 3 混淆矩阵所示,相应评估指标如表 5 所示。

表 5 故障分类评价指标

Table 5 Fault classification evaluation indicators

故障类型	TP Rate	FP Rate	Precision	Recall	F1 score
Fault0	0.943	0.007	0.965	0.943	0.954
Fault1	0.937	0.041	0.875	0.937	0.905
Fault2	0.898	0.015	0.883	0.898	0.890
Fault3	0.929	0.010	0.886	0.929	0.907
Fault4	0.924	0.022	0.967	0.924	0.945

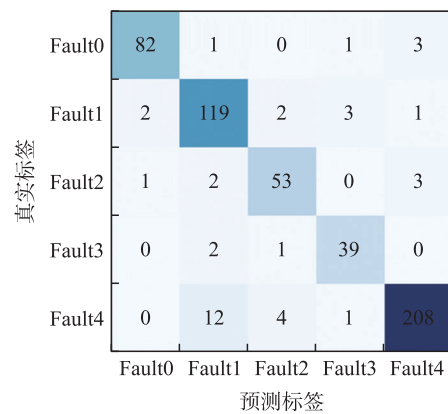


图 3 液体火箭发动机的故障数据混淆矩阵

Fig. 3 Confusion matrix of liquid rocket engine fault data

从混淆矩阵和故障分类评价指标来看,混淆矩阵对角线元素之和为 501,总样本数为 540,计算得整体准确率为 92.8%,表明 VI-MGP 算法在全局范围内具有较高的分类能力。Fault4 被误判为 Fault1 的样本数为 12,误判次数最多,可能是因为两类特征存在局部相似。五类故障的 F1 分数均值为 0.920,总体标准差为 0.025,各故障类型的综合分类性能较为均衡,未出现显著短板。

对比不同噪声方差下传统 MGP 算法和 VI-MGP 算法在液体火箭发动机故障诊断应用中的平均误差和平均 NLL,其箱线图如图 4、图 5 所示。

随着向真实数据中添加的噪声方差的增大,两种方法的平均分类误差和平均负对数似然在同步递

增. 噪声方差为 0.5 时, 噪声对输入数据的影响显著, 导致两种方法的误差率都接近 15%. 引入过高的噪声方差参数使得模型预测的结果不佳. 在未添加输入噪声, 即方差为 0 时, 平均误差仍比传统 MGP 方法要低, 对此合理的解释是液体火箭发动机故障真实数据本就包含了一定的噪声, 因此使用本文算法时更具有优势. 真实数据中引入合适的方差大小时, 含输入噪声的算法模型的平均误差比传统算法略有降低, 平均负对数似然比传统算法平均降低 32.3%, 且分布更加集中, 这表明添加噪声后的模型对真实后验具有更好的拟合.

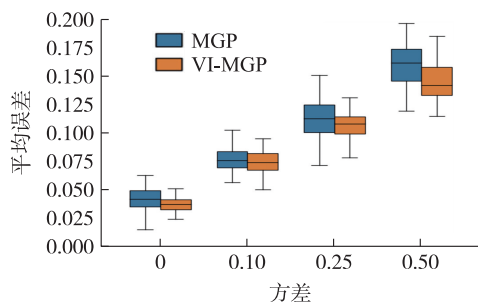


图 4 不同噪声方差下传统 MGP 算法与 VI-MGP 算法的平均误差对比

Fig. 4 Comparison of average error between traditional MGP algorithm and VI-MGP algorithm under different noise variances

表 6 展示了传统算法和添加输入噪声后算法的平均迭代时长. 引入噪声后, 算法需要迭代优化伪数据点和噪声输入上的两组变分参数, 相比较传统方法的计算量有所上升, 计算复杂度仍保持在 $\mathcal{O}(M^2N)$ 级别.

4 结论

本文在多类别高斯过程分类算法中引入数据噪声, 采用变分推断作为优化方法, 逼近模型的后验分布. 并推导引入输入噪声的多类别高斯过程分类算法模型. 提出的引入噪声的多类别高斯过程分类算法与传统无噪声多类别高斯过程分类算法对比, 能够降低一定的误差, 并且负对数似然要远小于传统算法, 对真实后验的拟合程度更好.

在提出含有输入噪声的多类别高斯过程分类算法中, 对不同的噪声方差数值进行对比. 实验证明数据含有的噪声的方差越大, 预测的结果误差越大; 数据样本量越大, 含输入噪声的多类别高斯过程分类算法的负似然对数与传统 MGP 方法相比优势越大.

该算法在液体火箭发动机故障诊断应用中表现良好, 整体准确率达 92.8%, 且各故障类型的 F1 分数均高于 0.89, 具备高鲁棒性和均衡性. 尽管存在局部误判, 但模型未因样本数量不均衡严重失准, 在小样本场景下仍具备较强泛化性, 能够满足实际工程应用的可靠性需求.

从含噪声数据中学习推断噪声方差参数时, 虽然误差与负对数似然均有下降, 但迭代时间略有提升, 且对噪声方差的估计效果不佳. 未来的研究方向将着重探索学习噪声的方差的方法优化, 以进一步提高引入噪声后的多分类高斯过程分类算法的性能.

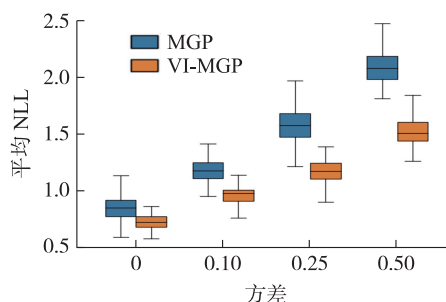


图 5 不同噪声方差下传统 MGP 算法与 VI-MGP 算法的平均 NLL 对比

Fig. 5 Comparison of average NLL between traditional MGP algorithm and VI-MGP algorithm under different noise variances

表 6 算法平均迭代时长

Table 6 Algorithm average iteration time

算法	传统方法	噪声方法
平均迭代时间/s	4.152	4.838

[参考文献]

- [1] GRANDINI M, BAGLI E, VISANI G. Metrics for multi-class classification; an overview[J]. arXiv.2020. 2008.05756.
- [2] BISHOP C M, NASRABADI N M. Pattern recognition and machine learning[M]. New York: Springer, 2006.
- [3] 何志昆, 刘光斌, 赵曦晶, 等. 高斯过程回归方法综述[J]. 控制与决策, 2013, 28(8): 1121-1129, 1137.
- [4] SCHULZ E, SPEEKENBRINK M, KRAUSE A. A tutorial on Gaussian process regression: modelling, exploring, and exploiting functions[J]. Journal of mathematical psychology, 2018, 85: 1-16.

- [5] DERINGER V L, BART6K A P, BERNSTEIN N, et al. Gaussian process regression for materials and molecules[J]. *Chemical reviews*, 2021, 121(16):10073–10141.
- [6] KUSS M, RASMUSSEN C E, HERBRICH R. Assessing approximate inference for binary gaussian process classification[J]. *Journal of machine learning research*, 2005, 6(10):1679–1704.
- [7] 张炳晖, 张研, 王伟, 等. 基于高斯过程二元分类模型的溶洞规模预测方法[J]. *中国岩溶*, 2020, 39(02):259–263.
- [8] 黄维, 童莹, 曹雪虹. 基于高斯过程的多分类器[J]. *计算机工程与设计*, 2016, 37(04):1021–1026.
- [9] MCHUTCHON A, RASMUSSEN C. Gaussian process training with input noise[C]//*Proceedings of the 25th International Conference on Neural Information Proceeding Systems*, Granada Spain, Curran Associates Inc, 2011:1341–1349.
- [10] XUE Y, LIU Y, JI C, et al. System identification of ship dynamic model based on Gaussian process regression with input noise[J]. *Ocean engineering*, 2020, 216:107862.
- [11] VILLACAMPA-CALVO C, ZALDÍVAR B, GARRIDO-MERCHÁN E C, et al. Multi-class gaussian process classification with noisy inputs[J]. *The journal of machine learning research*, 2021, 22(1):1696–1747.
- [12] SNELSON E, GHAHRAMANI Z. Sparse Gaussian processes using pseudo-inputs[C]//*Proceedings of the 19th International Conference on Neural Information Proceeding Systems*. Vancouver British Columbia Canada, MIT Press, 2005:1257–1264.
- [13] TITSIAS M, LAWRENCE N D. Bayesian Gaussian process latent variable model [C]//*Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Sardinia Italy, PMLR, 2010:844–851.
- [14] BLEI D M, KUCUKELBIR A, MCAULIFFE J D. Variational inference; A review for statisticians[J]. *Journal of the American statistical association*, 2017, 112(518):859–877.
- [15] ZHANG C, BÛTEPAGE J, KJELLSTRÖM H, et al. Advances in variational inference [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2018, 41(8):2008–2026.
- [16] TITSIAS M. Variational learning of inducing variables in sparse Gaussian processes[C]//*The 12th International Conference on Artificial intelligence and statistics*. Florida USA, PMLR, 2009:567–574.
- [17] LEIBFRIED F, DUTORDOIR V, JOHN S T, et al. A tutorial on sparse Gaussian processes and variational inference[J]. *arXiv*, 2020, 2012.13962.
- [18] WYNNE G, WILD V. Variational gaussian processes: A functional analysis view[C]//*Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*. Virtual Conference, PMLR, 2022:4955–4971.
- [19] DAMIANOU A, LAWRENCE N D. Semi-described and semi-supervised learning with Gaussian processes[J]. *arXiv*, 2015:1509.01168.
- [20] ZHANG Y, LIU W, CHEN Z, et al. On the properties of kullback-leibler divergence between multivariate Gaussian distributions [J]. *arXiv*, 2021, 2102.05485.
- [21] Trefethen L N. Exactness of quadrature formulas[J]. *Siam review*, 2022, 64(1):132–150.

[责任编辑:陆炳新]