QUBO决策树：退火机扩展决策树划分

Koichiro Yawata日立有限公司研发组
日本国分寺市
koichiro.yawata.rt@hitachi.com

Yoshihiro Osakabe日立有限公司研发组日本国分寺市
yoshihiro.osakabe.fj@hitachi.com

Takuya Okuyama日立有限公司研发组日本国分寺市
takuya.okuyama.mn@hitachi.com

Akinori Asahara日立有限公司研发组
日本国分寺市
akinori.asahara.bq@hitachi.com

arXiv:2303.09772v1 [cs.LG] 17 Mar 2023

摘要—本文提出了一种基于二次无约束二元优化（QUBO）的回归树扩展方法。回归树是非常流行的、可利用表格数据集训练的预测模型，但由于决策规则过于简单，其准确率不足。所提方法将决策树中的决策规则扩展为多维边界。由于计算能力有限，这种扩展通常无法实现；然而，所提方法将训练过程转换为QUBO，使退火机能够解决这一问题。

索引词—**退火机，二次无约束二元优化（QUBO），决策树，回归树，马尔可夫链蒙特卡洛**

I. 引言

退火机能够同时处理多种可能状态，因此在求解组合优化问题方面具有优势 [1]。近年来，退火机的性能和易用性均有所提升 [2], [3]。作为量子计算机的一种，量子退火引起了广泛关注 [4], [5]。因此，除了旅行商问题和最大割问题等典型组合优化问题的应用之外，人们还期待开发出其他使用退火机的算法 [6]。

回归树是用于回归任务的决策树；决策树是一种机器学习算法，可根据解释变量估计目标变量[7]。XGBoost [8] 和 LightGBM [9]等基于决策树的方法因性能出色 [10], [11]且具有可解释性 [12]而得到广泛应用。在决策树中，会迭代地按照各种特征值对样本进行划分，以使用贪心算法降低预测误差。

然而，受计算能力限制，回归树在每次划分中只能使用一个特征，不支持使用多个特征的复杂条件进行划分 [13], [14]。这是因为搜索空间的大小会随着所用特征数量的增加而呈指数级增长，从而使优化更加困难。这一限制可能导致划分效果不佳，进而对模型的准确率和可解释性产生负面影响。

为解决这一问题，我们考虑使用模拟退火 [15],来进行回归树划分优化；模拟退火是退火机的一种。退火技术是一种随机逼近算法，用于寻找使给定函数最小化的状态。在模拟退火中，通过基于马尔可夫链蒙特卡洛（MCMC）算法进行随机状态转移来实现收敛。

为了使用模拟退火优化回归树，我们提出了两种技术。首先，针对寻找能够降低回归树预测误差的划分这一搜索问题，我们将其表述为二次无约束二元优化（QUBO）问题，该问题可由退火机求解。其次，我们扩展回归树，使其能够搜索更复杂的条件。在传统方法中，样本划分只能使用一个特征值。相比之下，如图1所示，所提方法支持包含多个特征的决策规则。预计使用多个特征进行划分能够减少决策树中的划分次数，这也可能有助于提高决策树的可解释性 [16]。

在本研究中，假设输入特征在转换为 QUBO问题时为二元变量。对于所提方法，原本就是二元变量的特征（例如经典自然语言处理中词语的出现情况 [17]，或化学信息学中使用的分子描述符 [18], [19] ）无需任何预处理即可使用。即使变量不是二元变量，也可以通过独热编码等通用二值化技术 [20] 来使用。先前的研究中已经进行过此类预处理 [21], [22]。

II. 相关研究

决策树。决策树凭借出色的性能和可解释性 [23]，在许多场景中被用于表格数据分析。近年来，基于决策树的方法（如

IEEE版权声明 版权所有 (c) 2023年 IEEE 允许将本材料用于个人用途。本文已被2022年IEEE知识图谱国际会议（ICKG）接收并将发表。数字对象标识符：10.1109/ICKG55886.2022.00052

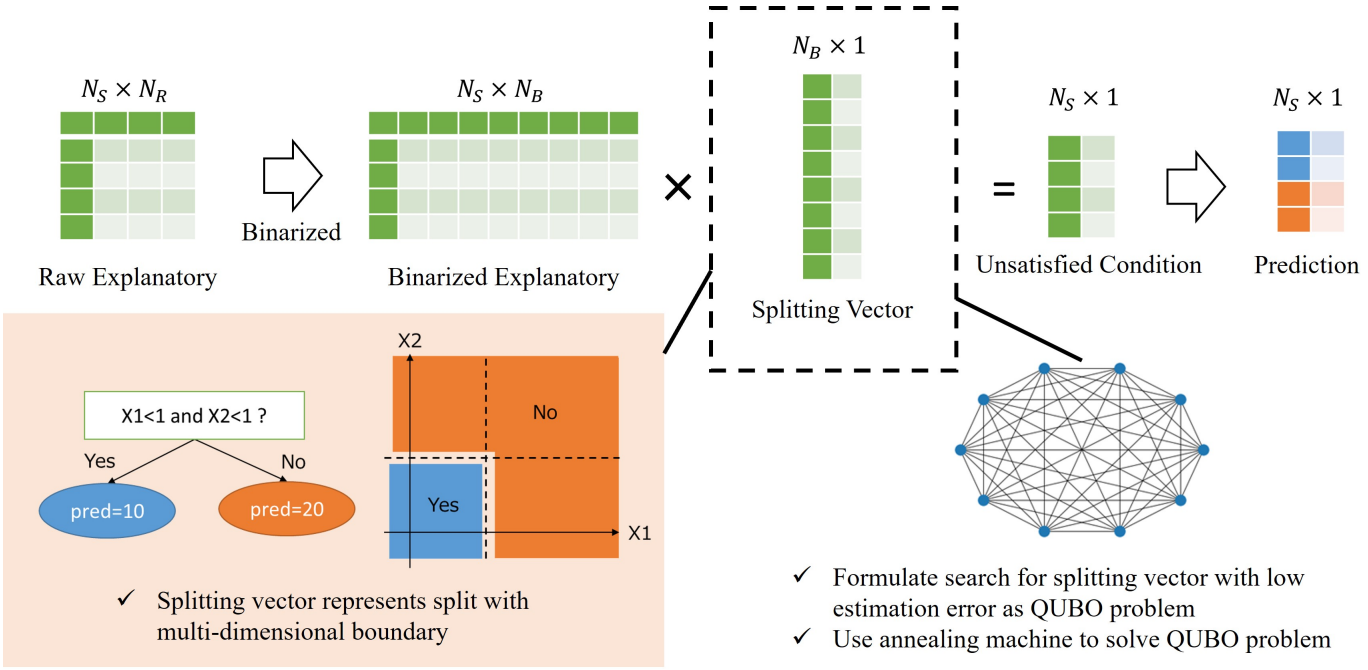


图1. 概述。我们将寻找回归树分裂条件的搜索问题表述为一个二次无约束二元优化问题，可由退火机求解。此外，我们扩展了回归树，使其能够搜索更复杂的条件，这些条件由多维边界定义。

随机森林 [12]，XGBoost [8] 和 LightGBM [9] ）的性能通过决策树集成和提升技术得到了改善。然而，构成决策树的样本划分方法并未发生显著变化[13]，[14]。在这些决策树中，每次划分只使用一个特征，并采用贪心算法。另一方面，多变量决策树会使用多个特征进行划分[24]–[29]。例如， [25],使用基于多个特征的线性判别函数进行划分。这些研究探讨的不仅是表达能力更强的分裂方法，也包括高效的搜索方法。此外，许多研究试图优化整棵决策树 [30]–[33]。这些研究旨在减少整棵决策树中的分割数量，这与本研究的目标相似。

使用马尔可夫链蒙特卡洛的机器学习。马尔可夫链蒙特卡洛是一种从概率分布中抽样的算法 [35]，[36]。在贝叶斯估计中，通常很难通过计算推导出后验分布。因此，马尔可夫链蒙特卡洛常用于机器学习 [37]–[39]。这些研究侧重于计算得到的分布，而我们关注退火机抽样得到的结果。动量退火 [40]通过使用 Metropolis算法 [35]，（一种马尔可夫链蒙特卡洛方法）来探索基态中的自旋构型。

退火机。关于退火机的硬件和软件，已有大量研究。它们在解决组合优化问题方面具有优势，且已应用于现实世界 [41]。量子退火[4] 和绝热量子演化算法 [42] 推动了退火机的发展。D-Wave 提供云端量子退火机 [2]。采用图形处理器并使用一种称为动量退火的技术的退火机也已问世 [40]，[43]。由于不同硬件和算法所能处理的 QUBO问题规模各不相同，因此有必要根据所要处理的问题选择合适的方法 [44]。在软件方面，PyQUBO [45]和 Ocean [46]等库也已发布。

由退火机优化的机器学习。人们尝试通过将问题表述为 QUBO问题，利用退火技术解决机器学习算法中的问题。在 [47],中，退火机已用于聚类。在 [6]，中，线性回归、支持向量机（SVM）和平衡K均值聚类这三种方法被转化为 QUBO问题。还有许多研究将深度学习任务表述为 QUBO问题 [48]–[50]。退火机也已用于图像处理 [51]–[53]、点云处理 [54]，[55]以及贝叶斯网络 [56]，[57]。

III. 问题设定

设 x_{r,s_i,r_i} 为解释性原始特征的值r_i， t_{s_i} 为样本 s_i 的目标变量， y_{s_i} 为样本 s_i的预测值。此外，设 S 为所有样本的集合， S_1 为满足分割样本条件的样本集合， S_0 为不满足条件的样本集合。 S 中的样本数量记为 N_S。

特征二值化。为了将回归树学习轻松地视为一个 QUBO问题，我们将解释变量二值化，作为预处理步骤。二值化特征记为 x_{b,s_i,b_i} ，用于表示样本 s_i 是否满足基本条件 b_i，如图 2 所示。基本条件由单个原始特征生成，例如 $x_{r,s_i,0} > 0$。

IEEE 版权声明 版权所有 (c) 2023年 IEEE。允许将本材料用于个人用途。本文已被 2022年IEEE知识图谱国际会议（ICKG）接收发表。数字对象标识符：10.1109/ICKG55886.2022.00052

id	*score*
0	10
1	14
2	12
3	18
4	20
5	16

Binarizing Feature

id	*score*<13	*score*<17	13<*score*	17<*score*
0	1	1	0	0
1	0	1	1	0
2	1	1	0	0
3	0	0	1	1
4	0	0	1	1
5	0	1	1	0

图 2. 特征二值化。为了将回归树学习作为 QUBO问题轻松处理，根据基本条件对原始解释变量进行二值化。二值化方法不受任何限制。

Condition	Use
class ≠ "1"	0
class ≠ "2"	0
class ≠ "3"	1
class ≠ "4"	1
score<13	0
score<17	1
13<*score*	1
17<*score*	0

Equivalent

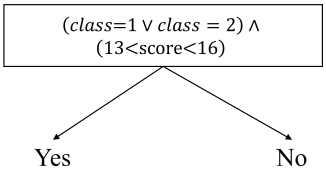


图 3. 分割表示。用于划分的条件以二进制向量表示。每个元素表示是否使用一个基本条件：值为 1 时使用该条件，值为 0 时则不使用。

如果样本 s_i 满足条件 b_i，则 $x_{b,s_i,b_i} = 1$；否则，$x_{b,s_i,b_i} = 0$。我们使用逻辑乘积条件来生成用于划分的复杂条件。逻辑乘积条件是由多个基本条件的逻辑乘积表示的条件，例如 $x_{r,s_i,0} > 0 \wedge x_{r,s_i,1} > 0$。设 N_F 和 N_B 分别为原始特征和基本条件的数量。

基本条件。连续特征和类别特征的基本条件构造方法有所不同。对于连续特征，我们构造条件，以判断特征是否大于或小于某个阈值。通过这种方式，我们可以构造条件来指定特征的取值范围，例如 $0 < x_{r,\ s_i,\ 0} < 1$。对于类别特征，我们构造条件，以判断特征是否等于某个类别变量，例如 $x_{r,\ s_i,\ 0} \neq$"1"。通过这种方式，我们可以表示类别变量的逻辑和。例如，如果类别变量的唯一取值为"1"、"2"、"3"和"4"，那么 $x_{r,\ s_i,\ 0} \neq$ "1" $\wedge\ x_{r,\ s_i,\ 0} \neq$ "2" 与 $x_{r,\ s_i,\ 0} =$ "3" $\vee\ x_{r,\ s_i,\ 0} =$ "4"表示相同的划分。虽然我们介绍了两种典型的基本条件，但其他条件也可以接受，例如判断特征是否缺失，以及多个特征之间关系的条件，如 $x_0 > x_1$。请注意，唯一可用于样本划分的条件，是使用预先构造的基本条件形成的逻辑乘积条件。相反，不使用逻辑和条件。

划分向量。本文将用于划分的逻辑乘积条件表示为长度为 N_B 的二进制向量，并称之为划分向量。每个元素表示是否采用相应的基本条件：值为 1 时采用该条件，值为 0 时则不采用。如图 3 所示，划分向量可以通过逻辑乘积条件表示复杂条件。由于条件有 2^{N_B} 种不同的组合，通常很难搜索到最优解。因此，本文使用擅长求解组合优化问题的退火机。请注意，如果划分时仅使用一个条件，并且准备了足够多的基本条件，那么这几乎与传统的决策树划分条件搜索相同。具体而言，可以根据排序后的特征值，通过不等式确定基本条件的阈值，从而得到这些条件。然而，实际上，考虑到计算成本和准确率，没有必要使用所有阈值 [9]

[34].

*A. QUBO*问题

在退火机中，待优化的参数由取二进制值 $\sigma_i \in \{-1, 1\}$ 的自旋变量表示，目标变量则由伊辛模型的哈密顿量表示。伊辛模型由以下方程给出：

$$H = -\sum_{i<j} J_{ij}\sigma_i\sigma_j - \sum_i h_i\sigma_i, \tag{1}$$

其中，J_{ij} 称为自旋相互作用系数，h_i 为磁场系数。通过 $\sigma_i = 2\theta_i - 1$，可以将公式 1 中的伊辛模型转换为QUBO形式：

$$H = \sum_{i<j} Q_{ij}\theta_i\theta_j + \sum_i b_i\theta_i,$$

其中，$\theta_i \in \{0, 1\}$ 是二元变量，称为QUBO变量；Q_{ij} 和 b_i 是用于表征QUBO问题的系数。由于QUBO问题的表达形式比伊辛问题更简单，本文采用QUBO形式。

*B. QUBO*问题的形式化

通过QUBO格式将哈密顿量表示为损失函数 L 与约束函数C 之和，从而构建QUBO问题，如下所示：

$$H = L\left(\boldsymbol{X}, \boldsymbol{t}|\boldsymbol{\theta}\right) + C(\boldsymbol{\theta})$$

这里，$\boldsymbol{\theta}$ 是表示模型参数的二进制向量。$\boldsymbol{X}$ 是样本 i的解释变量，$\boldsymbol{t}$ 是表示目标变量的向量。此外，L 是优化目标，$C(\boldsymbol{\theta})$是用于获得有效解的集合，通常要求为 $C(\boldsymbol{\theta}) = 0$。使用退火机来寻找使哈密顿量最小化的 $\boldsymbol{\theta}$。

已有许多约束函数被提出 [60], ，本文使用线性不等式约束。线性不等

IEEE版权声明 版权所有 (c) 2023年 IEEE。允许个人使用本材料。本文已获准在2022年IEEE知识图谱国际会议（ICKG）上发表。数字对象标识符：10.1109/ICKG55886.2022.00052

式约束是指可表示为公式2的约束，而约束函数则表示为公式3：

$$\alpha \leq \sum_i a_i \theta_i \leq \beta \tag{2}$$

$$C(\boldsymbol{\theta}) = (\sum_i a_i \theta_i - \sum_{\alpha \leq j \leq \beta} j\theta_{s,j})^2 + (\sum_{\alpha \leq j \leq \beta} \theta_{s,j} - 1)^2, \tag{3}$$

其中，$\theta_{s,\ j}$ 是一个辅助变量，也是优化的目标。当$C(\boldsymbol{\theta})$ $= 0$时，公式2的约束得到满足。当 α 和 β 相等时，即对应于等式约束，公式3中的第二项是不必要的。

C. 贡献

本文的贡献如下。

1）我们提出了一种方法，将为回归树寻找分裂条件这一搜索问题表述为二次无约束二元优化问题，并可由退火机求解。2）我们扩展了回归树，使其能够搜索更复杂的条件，即由多维边界定义的分裂条件。3）为提高解的收敛性，我们提出了一种分割约束，要求同时存在满足分裂条件的样本和不满足分裂条件的样本。4）我们在合成数据和真实数据上评估了所提方法。结果表明，尽管该方法可处理的问题规模有限，但仍能在合成数据上随机找到复杂条件，并找到能够降低估计误差的条件，其效果优于传统划分方法。

总的来说，本文展示了如何利用退火机，使广泛应用于表格数据分析的回归树变得更加复杂和完善。这项技术为机器学习提供了一种新选择。

IV. 二次无约束二元优化决策树

我们首先将误差函数——均方误差（MSE）——转换为QUBO格式。回归树中的MSE可写为：

$$\text{MSE} = \frac{1}{N_S}(\sum_{s_i \in S_1} (\text{pred}_1 - t_{s_i})^2 + \sum_{s_i \in S_0} (\text{pred}_0 - t_{s_i})^2),$$

其中，pred_1 是满足条件时的估计值，pred_0 是不满足条件时的估计值。回归树的学习涉及寻找划分条件以及 pred_1,pred_0 ，以降低均方误差。显然，pred_1 和 pred_0 是按条件划分的样本的均值。因此，均方误差等于各划分后样本组的方差乘以各样本组所占比例后求和。因此，均方误差可以写成以下公式：

$$\begin{aligned}\text{MSE} &= \sum_{b=0,1} \text{Var}\left(\{t_{s_i} | s_i \in S_b\}\right) \frac{N_{S_b}}{N_S} \\ &= \sum_{b=0,1} \left(\frac{1}{N_S} \sum_{s_i \in S_b} t_{s_i}^2 - \frac{1}{N_S N_{S_b}} (\sum_{s_i \in S_b} t_{s_i})^2\right).\end{aligned}$$

然而，由于MSE涉及使用变量 N_{S_b}进行除法运算，而 N_{S_b} 是QUBO变量之和，因此很难将其表述为QUBO问题。在这里，由于 N_S 是样本总数，因而为常数，所以不存在这个问题。因此，我们提出了平方加权均方误差（SWMSE）。在SWMSE中，计算方差之和时，不再使用样本组所占比例作为权重，而是使用该值的平方：

$$\begin{aligned}\text{SWMSE} &= \sum_{b=0,1} \text{Var}\left(\{t_{s_i} | s_i \in S_b\}\right) \left(\frac{N_{S_b}}{N_S}\right)^2 \\ &= \sum_{b=0,1} \left(\frac{N_{S_b}}{N_S} \sum_{s_i \in S_b} t_{s_i}^2 - \frac{1}{N_S} (\sum_{s_i \in S_b} t_{s_i})^2\right).\end{aligned} \tag{4}$$

通过这一转换，MSE最小化问题变成了QUBO问题。MSE与SWMSE之间的关系将在第VI节中讨论。

A. QUBO问题的形式化

待优化的QUBO变量有两类： $\boldsymbol{\theta}_B$和 $\boldsymbol{\theta}_X$，其元素分别表示为$\theta_{B、\ b_i}$ 和$\theta_{X、\ s_i、\ c_i}$。设 $\boldsymbol{\Theta}_B$ 为 $\boldsymbol{\theta}_B$可能取值向量的集合。首先，$\theta_{B、\ b_i}$ 是表示划分逻辑乘积条件的二元变量。根据样本是否满足由b_i 构成的条件的逻辑乘积来划分样本，其中这些条件满足$\theta_{B、\ b_i} = 1$。$\theta_{X、\ s_i、\ c_i}$ 是用于计算SWMSE的辅助二元变量，其中$\theta_{X、\ s_i、\ c_i} = 1$意味着样本 s_i 不满足 c_i 个条件。请注意，$\theta_{X、\ s_i、\ 0} = 1$ 表示样本 s_i 满足所有划分条件，即$\theta_{X、\ s_i、}$0表示分割结果。此处， c_i 是一个从0到 M的整数，其中 M 是逻辑乘积条件中可使用的二元特征的最大数量，也是一个学习参数。不满足这些条件的样本数量记为 s_i ，表示为 $unsatisfied\ \ feature_{s_i}$，其表达式如下：

$$unsatisfied_feature_{s_i} = \sum_{b_i} (1 - x_{b,s_i,b_i})\, \theta_{B,b_i}.$$

如果 $unsatisfied\ \ feature_{s_i} = 0$，那么样本 s_i满足所有划分条件。

损失函数。方程（4）中 t_{s_i} 出现的时刻可通过以下方程使用QUBO变量表示：

$$\begin{aligned}\sum_{s_i \in S_1} t_{s_i}^n &= \sum_{s_i} \theta_{X,s_i,0} t_{s_i}^n \qquad (n = 1, 2, ..), \\ \sum_{s_i \in S_0} t_{s_i}^n &= \sum_{s_i} (1 - \theta_{X,s_i,0}) t_{s_i}^n \quad (n = 1, 2, ..).\end{aligned}$$

IEEE版权声明 版权所有 (c) 2023年 IEEE。允许个人使用此材料。已获准在2022年IEEE知识图谱国际会议（ICKG）上发表。数字对象标识符：10.1109/ICKG55886.2022.00052

同样，按某个条件划分的样本组数量 N_{S_1} 和 N_{S_0}可表示如下：

$$N_{S_1} = \sum_{s_i} \theta_{X,s_i,0}$$
$$N_{S_0} = \sum_{s_i} (1 - \theta_{X,s_i,0}).$$

因此，SWMSE可通过以下公式表示为QUBO形式的问题：

$$\begin{aligned}\text{SWMSE} = &(\sum_{s_i} \theta_{X,s_i,0} t_{s_i}^2)(\sum_{s_i} \theta_{X,s_i,0}) \\ &- (\sum_{s_i} \theta_{X,s_i,0} t_{s_i})^2 \\ &+ (\sum_{s_i} (1 - \theta_{X,s_i,0}) t_{s_i}^2)(\sum_{s_i} (1 - \theta_{X,s_i,0})) \\ &- (\sum_{s_i} (1 - \theta_{X,s_i,0} t_{s_i}))^2.\end{aligned}$$

约束函数。如下方公式（5）至（7）所示，QUBO变量有三类约束。这里，公式（5）约束每个样本中θ_{B,b_i} 与θ_{X,s_i,c_i} 之间的关系。满足此约束意味着θ_{X,s_i,c_i} 可以表示每个样本中构成划分条件的已满足基本条件的数量。公式（6）约束每个样本中θ_{X,s_i,c_i} 的有效性。公式（7）是一个可选约束，用于缩小搜索空间。尽管 M 的取值最高可达 N_B，但出于实际考虑，我们可以对其进行限制。

$$\forall s_i \sum_{b_i} (1 - x_{b,s_i,b_i}) \theta_{B,b_i} - \sum_{c_i} c_i \theta_{X,s_i,c_i} = 0 \quad (5)$$
$$\forall s_i \sum_{c_i} \theta_{X,s_i,c_i} = 1 \quad (6)$$
$$1 \le \sum_{b_i} \theta_{B,b_i} \le M \quad (7)$$

当所有这些约束都满足时，实际的 SWMSE 等于退火机计算出的 SWMSE。反之，由于违反少数约束时两个 SWMSE 之间的差异很小，对学习的影响也很小，因此不必满足所有约束。

分割约束。除了上述三个约束外，我们还提出了一项用于实现有效划分的约束。通过添加这项约束，我们可以避免划分后的样本数量为零，从而使学习过程更加稳定。这里， a 是划分后的最小样本比例。

$$aN_S \le \sum_{s_i} \theta_{X,s_i,0} \le (1 - a) N_S \quad (8)$$

哈密顿量。通过最小化下方公式（9）中的哈密顿量来探索划分方法。在该公式中，C_1，C_2，C_3 和 C_{add} 是公式（5）至（8）中所述约束对应的哈密顿量，而 w 是用于调整各哈密顿量权重的参数。请注意，这些约束并非都必须满足。换言之，即使存在违反的约束，仍可使用获得的 $\theta_{B,\ b_i}$ 创建估计器。

$$\begin{aligned}H = &\frac{w_q}{N_S} \text{SWMSE}(\boldsymbol{X}, \boldsymbol{t} | \boldsymbol{\theta}_B, \boldsymbol{\theta}_X) + \frac{w_{c1}}{N_S} C_1(\boldsymbol{\theta}_B, \boldsymbol{\theta}_X) \\ &+ \frac{w_{c2}}{N_S} C_2(\boldsymbol{\theta}_X) + C_3(\boldsymbol{\theta}_X) + C_{add}(\boldsymbol{\theta}_X)\end{aligned} \quad (9)$$

预测。根据从退火机获得的 θ_{B,b_i}创建估计器，如下列公式所示。

$$S_1 = \{s_i | s_i \in S \wedge \sum_{b_i} (1 - x_{b,s_i,b_i}) \theta_{B,b_i} = 0\}$$
$$S_0 = \{s_i | s_i \in S \wedge \sum_{b_i} (1 - x_{b,s_i,b_i}) \theta_{B,b_i} > 0\}$$
$$y_{s_i} = \begin{cases} \frac{1}{N_{S_1}} \sum_{s_i \in S_1} t_{s_i}, & s_i \in S_1 \\ \frac{1}{N_{S_0}} \sum_{s_i \in S_0} t_{s_i}, & s_i \in S_0 \end{cases}$$

V. 实验

由于模拟退火具有随机性，每次试验都会返回不同的解。因此，我们使用不同的随机种子多次执行优化，并通过比较所得结果来确定最终结果。本研究通过观察多次试验中获得的理想结果数量，对所提方法进行了评估。

合成数据集。为了评估所提方法寻找最优条件的能力，我们生成了合成数据集。这些数据集使用三个参数生成：K、N_S和 N_F，如图4所示。这里，K 是构成逻辑乘积条件的特征数量，而目标变量t_{s_i} 则通过以下公式由$x_{b、s_i、0}$、 $x_{b、s_i、1}$、...、$x_{b、s_i、K-1}$确定：

$$t_{s_i} = \begin{cases} 1 & \text{if } x_{b,s_i,0} = x_{b,s_i,1} = .. = x_{b,s_i,K-1} = 1 \\ 0 & \text{else} \end{cases}.$$

由于$x_{b、s_i、b_i}$是二元变量，因此也可以写作$t_{s_i} = \prod_{b_i=0}^{K-1} x_{s_i、b_i}$。显然，满足$\theta_{B、0}$ =1、...、$\theta_{B、K-1}$ = 1、 $\theta_{B、K}$ = 0、..$\theta_{B、N_B-1}$ = 0的 $\boldsymbol{\theta_B}$ 是能使均方误差达到最小值的最优划分向量 $\boldsymbol{\theta_B^*}$ （图4）。对于使用词语出现向量进行的垃圾邮件分类，可以将包含所有选定词语的邮件与不包含这些词语的邮件区分开来。具体而言，数据集生成基于算法1，其参数经过调整，使得 $\boldsymbol{t}$ 的平均值为0.5。

真实数据集。为了评估所提方法在真实数据上训练回归树的能力，我们使用了艾姆斯房价数据集 [58]。该数据集包含 1460 套房屋的数据，涉及 79 个解释变量以及目标变量房价。为了将特征二值化，我们按 q 分位数（$q = 0.33$，0。66）对连续变量进行划分，并移除了唯一值超过 3 个的类别特征。最终，二元特征的数量为 136。

IEEE版权声明 版权所有 (c) 2023年 IEEE。允许个人使用本材料。已获准发表于2022年IEEE知识图谱国际会议（ICKG）。数字对象标识符：10.1109/ICKG55886.2022.00052

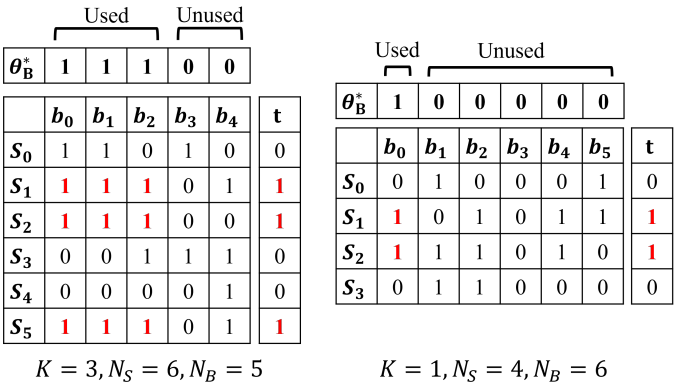


图4. 合成数据集。数据集使用三个参数生成：K，N_S和 N_F。如果 $x_{b,\ s_i,\ 0} = x_{b,\ s_i,\ 1} = \cdots = x_{b,\ s_i,\ K-1}$为1，则目标变量为 $t_{s_i} = 1$。$\boldsymbol{\theta^*_B}$是最优划分向量。

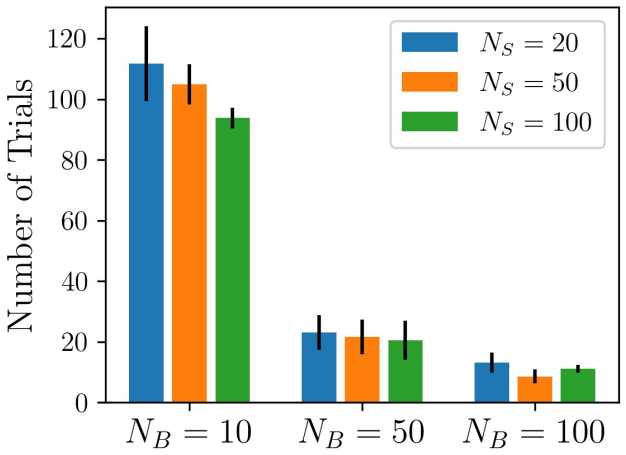


图 5. 找到最优划分向量（$K = 1$）所需的试验次数。黑线表示标准差。

算法1 生成合成数据

输入： N_S，N_B，K

输出： $x_{s_i,\ b_i}, t_{s_i}$

对于 $s_i = 0$ 到 $N_B - 1$ **执行**
  对于 $b_i = 0$ 至 N_B **执行**
    $rand \leftarrow Uniform(0,\ 1)$
    如果 $b_i < K$ **则**
      $th \leftarrow (1 - 0.5^{1/K})$
    else
      $th \leftarrow 0.5$
    结束如果
    如果 $rand > th$ **则**
      $x_{s_i,b_i} = 1$
    else
      $x_{s_i,b_i} = 0$
    结束条件判断
  结束循环
  $t_{s_i} = \prod_{b_i=0}^{K-1} x_{s_i,b_i}$
结束循环

A. 评估

合成数据集。使用合成数据集，我们评估了所提方法寻找满足 $\theta^*_{B,\ 0} = 1$，..，$\theta^*_{B,\ K-1} = 1$，$\theta^*_{B,\ K} = 0$，..$\theta^*_{B,\ N_B-1} = 0$ 的最优划分向量 $\boldsymbol{\theta^*_B}$ 的能力。这些实验使用 PyQUBO 库在中央处理器上进行。参数设置为试验次数 1000，每次试验的迭代次数为 10000，其他参数采用默认值。我们评估了找到最优划分向量的试验次数。共进行了五次实验，数据集均为随机生成。我们计算了找到最优划分向量所需试验次数的均值和标准差。

图 5 展示了 $K = 1$、$M = 1$ 以及 N_S 的不同取值（= 20、50、100）下的实验结果，以及 N_B（= 10、50、100）。结果发现，随着问题规模增大，找到最优划分向量所需的试验次数减少。影响条件数量的特征数 N_B 对结果的影响尤为显著。这些结果表明，使用所提方法可以随机地获得与传统方法中的决策树相同的分割结果。

表 I M 与找到最优划分向量所需试验次数之间的关系。

	M=1	M=2	M=3	M=4
N_S=20	111.8±12.3	69.4±20	2.8±2.5	0.0±0.0
N_S=50	105.0±6.6	5.8±5.2	0.0±0.0	0.0±0.0
N_S=100	93.8±3.3	0.8±0.8	0.0±0.0	0.0±0.0

表 I 列出了使用 $K = 1$、$N_S = 20$、$N_B = 10$以及不同 M 值（= 1、2、3、4）进行实验的结果。随着 M 增加，找到最优划分向量所需的试验次数减少；当 M 为 4 时，所有 N_S 的试验次数均为 0。我们预计，随着满足约束的 $\boldsymbol{\theta}_B$ 增多，找到最优划分所需的试验次数会减少。

接下来，我们评估了所提方法能否找到表示逻辑乘积条件的最优划分向量。我们使用带有 $K = 2$ 的数据生成算法对此进行了验证；在这种情况下，$\theta_{B,\ b_i} = \{1, 1, 0, ..., 0\}$ 是最优划分向量。表 II 列出了针对 $M = 2$ 和 $K = 1$、2 找到最优划分向量的试验次数。即使条件变得更加复杂，也未观察到明显的精度下降，这证实了 $\theta_{B,\ b_i}$ 的搜索空间具有显著影响。

真实数据集。使用真实数据集，我们评估了所提方法寻找能够实现较低均方误差的划分向量的能力。我们将使用 $\boldsymbol{\theta}_B$ 个划分条件的划分向量所得到的均方误差记为均方误差($\boldsymbol{\theta}_B$)。我们评估了所提方法能否找到一个划分向量，使其均方误差

IEEE版权声明 版权所有 (c) 2023年 IEEE 允许将此材料用于个人用途。已接受，待发表于2022年IEEE知识图谱国际会议（ICKG）。数字对象标识符：10.1109/ICKG55886.2022.00052

表 II K 与找到最优划分向量所需试验次数之间的关系。

	K=1	K=2
N_S=20	69.4±20	61.0±14.0
N_S=50	5.8±5.2	8.2±3.6
N_S=100	0.8±0.8	2.4±1.3

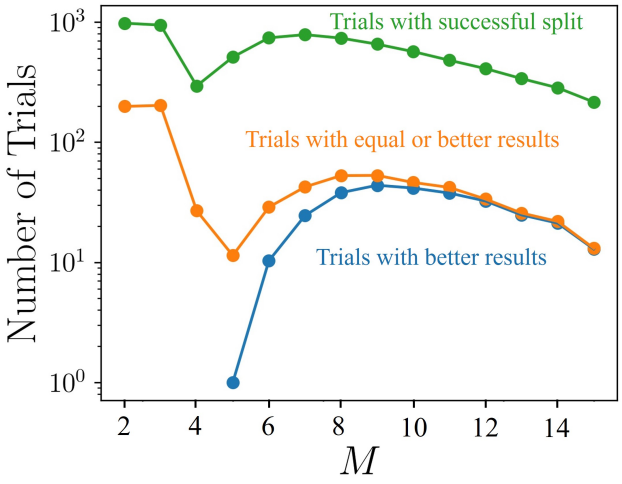


图 6. 不同 M 值下划分结果的评估结果。请注意， M = 2,3,4均未出现结果更好的试验。

低于仅使用一个基本条件时的最佳均方误差。仅使用一个基本条件时的最佳均方误差记为 cMSE，其计算公式如下：

$$\boldsymbol{\Theta}_B^1 = \{\boldsymbol{\theta}_B | \boldsymbol{\theta}_B \in \boldsymbol{\Theta}_B \wedge \sum_{b_i} \theta_{B,b_i} = 1\},$$

$$\text{cMSE} = \min_{\boldsymbol{\theta}_B \in \boldsymbol{\Theta}_B^1} \text{MSE}(\boldsymbol{\theta}_B),$$

其中， $\boldsymbol{\Theta}_B^1$ 是仅包含一个条件的划分向量集合。如果基本条件准备充分，则使用传统回归树时，cMSE 与均方误差一致。真实数据集上的实验使用图形处理器和动量退火进行。试验次数参数设为 1000。共进行了十次实验，并从数据集中随机选取样本。再次计算试验次数的均值和标准差。我们将最小分割比设为 $a = 0.2$。

图 6 从三个方面展示了使用不同 M 值对每次试验（N_S =20）进行评估的结果：是否生成了可分割条件、是否取得了等于或优于对比方法的均方误差（MSE($\boldsymbol{\theta}_B$) ≤ cMSE），以及是否取得了优于对比方法的均方误差（MSE($\boldsymbol{\theta}_B$) <cMSE）。这里，可分割条件是指既存在满足该条件的样本，也存在不满足该条件的样本的条件。令 N_{gs}、 N_{eq} 和 N_{su} 分别表示满足上述各项条件的试验次数。有趣的是， N_{gs} 和 N_{eq}

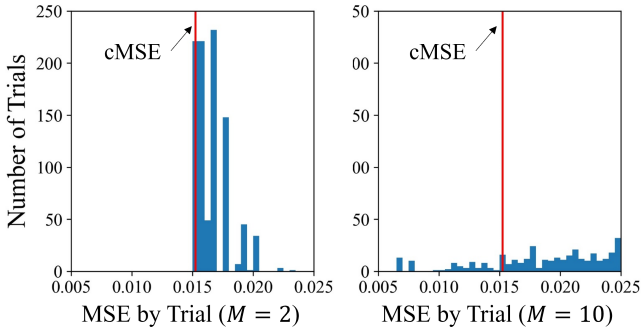


图7. 均方误差值的直方图。红线表示 cMSE 值。

表IIIN_{su} 不同 N_S 和 M 值下。

N_S	3	5	8	10
20	0.0±0.0	1.0±0.9	38.0±19.8	41.5±29.1
30	0.0±0.0	3.4±1.7	2.2±1.8	1.3±1.6
40	0.0±0.0	0.4±0.7	0.0±0.0	0.0±0.0
50	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0

随着 M 增大而减少。然而， N_{su} 在 $M = 9$ 时达到最大值。这可能是因为在对 $\theta_{B,\ b_i}$施加较强约束时，算法收敛到了局部解。

图7显示了 $M = 2$，10时均方误差值的直方图。当 M 为2时，许多试验的均方误差值集中在传统方法的均方误差值附近。另一方面，当 M 为10时，均方误差值分布较为分散，且在若干次试验中，均方误差低于传统方法。表III列出了 N_{su} 在不同 N_S 和M取值下的结果。即使样本数量增加，所提方法在较小的 M下仍无法找到更优的划分。表 IV总结了分割约束的影响。可以看出，加入分割约束使平均N_{su} 有所增加。

条件冗余。构成所得逻辑乘积条件的部分分裂条件并未影响分割结果。这些冗余条件可能在评估过程中产生了负面影响。因此，我们从逻辑乘积条件中逐一移除不会改变分割结果的条件，直至不再有此类条件。请注意，这一过程并不能保证达到该分割所需的条件数量最少。图 8 显示了冗余去除过程前后使用的特征数量（M = 10, N_S = 20）。为了更清楚地显示数据，我们在每个点的坐标位置上添加了噪声。如图所示，只有大约一半的条件

表 IV 分割约束影响的消融结果。

	3	5	8	10
w/o	0.0±0.0	0.8±1.1	22.7±12.7	22.7±17.1
w/	0.0±0.0	1.0±0.9	38.0±19.8	41.5±29.1

IEEE版权声明 版权所有 (c) 2023 IEEE。允许将本材料用于个人用途。已被 2022年IEEE知识图谱国际会议 (ICKG) 接收并将发表。数字对象标识符：10.1109/ICKG55886.2022.00052

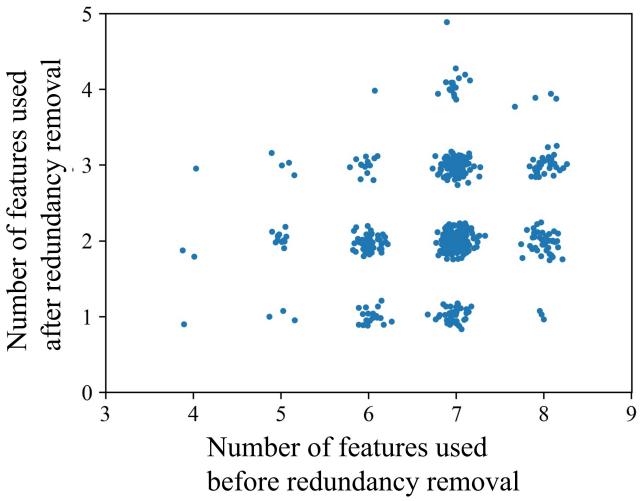


图 8. 冗余消除前后使用的特征数量（$M = 10$，$N_S = 20$），图中添加了噪声以提高可视性。

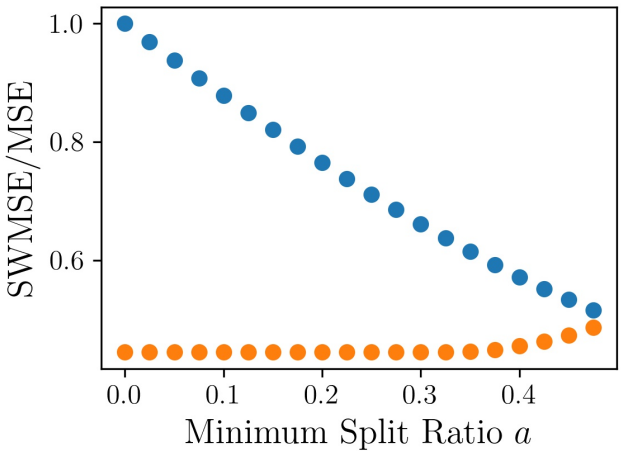


图 9. 最小划分比例与 SWMSE/均方误差比之间的关系。蓝色和橙色点分别表示 SWMSE/均方误差的最大值和最小值。

实际上有效。我们认为，这些冗余分割条件导致算法无法收敛到局部解。更重要的是，我们了解到，使用它时需要仔细调整参数 M 。

VI. 讨论

均方误差与 SWMSE 之间的关系。我们首先讨论均方误差和 SWMSE 之间的差异。设 ρ 和 V 分别为满足条件的样本组比例和目标变量的方差，设 γV 为不满足条件的样本组比例。那么，这两个指标可表示如下。

$$\begin{aligned}\text{SWMSE} &= V\rho^2 + \gamma^2 V(1-\rho)^2 \\ &= V(\rho^2 + \gamma^2(1-\rho)^2)) \\ &= (\frac{\text{MSE}}{\rho + \gamma^2(1-\rho)})(\rho^2 + \gamma^2(1-\rho)^2) \\ &= \frac{\rho^2 + \gamma^2(1-\rho)^2}{\rho + \gamma^2(1-\rho)} \cdot \text{MSE}\end{aligned}$$

图 9 显示了在假设 $0.5 < \gamma < 2.0$ 的情况下，采用分割约束时最小分割比 a 变化所对应的两个指标比值的最大值和最小值。随着 a 增大，该比值的取值范围缩小，这表明分割约束有助于减小优化中 SWMSE 与均方误差之间的差异。

此外，假设划分后样本组的 y 值方差相等，则以下等式成立。

$$\text{SWMSE} = (2(\rho_1 - 0.5)^2 + 0.5) \cdot \text{MSE}$$

如果均方误差相同，SWMSE 会因为样本被均等划分而发挥作用。这可能有助于防止过拟合，而且这两个指标之间的差异似乎不会对学习产生极其负面的影响。

与深度回归树的区别：使用浅层逻辑乘积条件有两个优点。首先，它不会对不满足条件的样本进行额外划分。在深度回归树中，不满足条件的样本会被划分到不同组中。因此，可用于预测的样本会减少，从而更容易受到噪声的影响。其次，大量分割会降低模型的可解释性。事实上，很少有人能逐一查看回归树的分割情况。

实际应用。如实验所示，寻找良好划分的过程具有随机性，在某些情况下，其准确度低于传统方法。为解决这一问题，最好同时采用所提方法和确定性的传统方法，然后针对每次划分选择准确度更高的方法。此外，由于所提方法关注的是划分，因此似乎不难应用与决策树相关的技术，例如集成方法、提升和早停。

VII. 结论

本文将回归树中寻找能够降低预测误差的划分这一搜索问题表述为二次无约束二元优化问题。QUBO形式使得我们能够搜索基于多个特征的复杂分裂条件。我们展示了所提划分方法的有效性。在合成数据上，我们展示了所提方法发现可表示为逻辑乘积的条件的能力。在真实数据上，我们展示了所提方法发现优于传统方法的划分条件的能力。然而，在实际的决策树中，需要重复进行多次分割，才能输出最终估计。未来，需要将迄今开发的回归树技术应用于所提方法并进行验证。我们相信，这些技术将为机器学习提供一种新的选择。

IEEE 版权声明 版权所有 (c) 2023年 IEEE。允许个人使用本材料。已获准发表于2022年IEEE知识图谱国际会议 (ICKG)。数字对象标识符：10.1109/ICKG55886.2022.00052

参考文献

[1] R. Martok、G. E. Santoro 和 E. Tosatti，“旅行商问题的量子退火”，《物理评论E：统计物理、等离子体、流体及相关跨学科主题》，第70卷，第5期，第057701页，2004年11月。[2] S. W. Shin、G. Smith、J. A. Smolin 和 U. Vazirani，“D-Wave机器有多‘量子’？”，arXiv预印本 arXiv:1401.7087，2014年。[3] “IBM Quanum”2022，[在线]。网址：https://quantum- computing.ibm.com/。[访问日期：08.06.2022]。[4] T. Kadowaki 和 H. Nishimori，“横场伊辛模型中的量子退火”，《物理评论E》，第58卷，第5355–5363页，1998年11月。[5] J. J. Brooke、D. Bitko、T. F. Rosenbaum 和 G. Aeppli，《无序磁体的量子退火》。《科学》，284(5415)：779–781，1999年。[6] P. Date、D. Arthur 和 P. Lauren，“用于训练机器学习模型的QUBO形式”，《科学报告》，第11卷，10，2021 1，1第。页码 -。[7] S. R. Safavian 和 D. Landgrebe，“决策树分类器方法综述”，载于《IEEE系统、人类与控制论汇刊》，第21卷，第3期，第660–674页，1991年5月至6月，数字对象标识符：10.1109/21.97458。[8] T. Chen 和 C. Guestrin，“XGBoost：一种可扩展的树提升系统”，SIGKDD会议论文集，第785–794页，2016年。[9] G. Ke 等，“LightGBM：一种高效的梯度提升决策树”，第31届神经信息处理系统会议（NIPS）论文集，第3146–3154页，2017年。[10] J. Bennett 和 S. Lanning，《Netflix奖》。载于《2007年KDD杯研讨会论文集》，第3–6页，纽约，2007年8月。[11] X. He、J. Pan、O. Jin、T. Xu、B. Liu、T. Xu、Y. Shi、A. Atallah、R. Herbrich、S. Bowers 和 J. Q. n. Candela，《Facebook广告点击预测的实践经验》。载于《第八届在线广告数据挖掘国际研讨会论文集》，ADKDD’14，2014年。[12] 2001 L. Breiman，《随机森林》。技术报告；W. Y. Loh，“分类与回归树”，《Wiley跨学科数据挖掘与知识发现评论》，第1卷，第14–23页，2011年。[13] [14] K. Grabczewski 和 N. Jankowski，“基于决策树准则的特征选择”，《第五届混合智能系统国际会议》（HIS’05），2005年，第6页，数字对象标识符：10.1109/ICHIS.2005.43。[15] S. A. Kirkpatrick、C. D. Gellatt Jr. 和 M. P. Vecchi，“模拟退火优化”，《科学》，第220卷，第671–680页，1983年。[16] A. Ignateiv 等，《计算最优决策集的可扩展两阶段方法》。arXiv预印本 arXiv:2102.01904，2021年。[17] A. McCallum 和 K. Nigam，“朴素贝叶斯文本分类事件模型的比较”，人工智能促进协会文本分类学习研讨会，1998年。[18] A. Mauri，“alvaDesc：用于计算和分析分子描述符及指纹的工具”，载于《生态毒理学定量构效关系》。药理学与毒理学方法，纽约：Humana，2020年。[19] C. W. Yap，“PaDel-descriptor：用于计算分子描述符和指纹的开源软件”，《计算化学杂志》，第32卷，第7期，第1466–1474页，2011年。[20] H. T. Cheng、L. Koc、J. Harmsen、T. Shaked、T. Chandra、H. Aradhye、G. Anderson、G. Corrado、W. Chai、M. Ispir 等，“面向推荐系统的宽深度学习”，《第一届推荐系统深度学习研讨会论文集》，第7–10页，2016年。[21] H. Verhaeghe、S. Nijssen、G. Pesant、C.-G. Quimper 和 P. Schaus，“使用约束规划学习最优决策树”，CP2019，2019年。[22] S. Verwer 和 Y. Zhang，“使用二元线性规划表述学习最优分类树”，第33届AAAI人工智能会议，2019年。[23] Scott M. Lundberg、Gabriel G. Erion 和 Su-In Lee，“树集成方法的一致个体化特征归因”，arXiv预印本 arXiv:1802.03888，2018年。[24] C. E. Brodley 和 P. E. Utgoff，“多变量决策树”，《机器学习》，第19卷，第45–77页，1995年。[25] X.B. Li、J.R. Sweigar、J.T.C. Teng、J.M. Donohue、L.A. Thombs 和 S.M. Wang，“使用线性判别式和禁忌搜索的多变量决策树”，《IEEE系统、人类与控制论汇刊A辑》，第33卷，第2期，第194–205页，2003年3月。[26] S. K. Murthy、S. Kasif 和 S. Salzberg，“一种用于归纳斜决策树的系统”，《人工智能研究杂志》，第2卷，第1–32页，1994年。[27] H. Kim 和 W.-Y. Loh，“具有无偏多路分裂的分类树”，《美国统计学会杂志》，第96卷，第589–604页，2001年。[28] W.-Y. Loh 和 Y.-S. Shih，“分类树的分裂选择方法”，《统计学报》，第7卷，第815–840页，1997年。[29] W.-Y. Loh 和 N. Vanichsetakul，“基于广义判别分析的树结构分类”，《美国统计学会期刊》，第83卷，第403期，第715-728页，1988年。[30] N. Narodytska、A. Ignatiev、F. Pereira和J. Marques-Silva，“利用可满足性问题学习最优决策树”，第二十七届国际人工智能联合会议论文集（IJCAI 2018），第1362-1368页，2018年7月13日至19日，2018年。[31] A. Schidler、S. Szeider. 基于可满足性问题的大型数据集决策树学习。载于：人工智能促进协会会议论文集。2021年。[32] C. Bessiere、E. Hebrard和B. O’Sullivan，“将决策树规模最小化作为组合优化问题”，第15届国际约束编程原理与实践会议论文集，第173-187页，2009年。[33] F. Avellaneda，“高效推断最优决策树”，人工智能促进协会，第3195-3202页，2020年。[34] P. Li、Q. Wu和C. J. Burges，“Mcrank：使用多重分类和梯度提升进行排序学习”，神经信息处理系统进展会议论文集，第897-904页，2007年。[35] N. Metropolis、A. W. Rosenbluth、M. N. Rosenbluth、A. H. Teller和E. Teller，“利用高速计算机计算状态方程”，《化学物理杂志》，第21卷，第6期，第1087-1092页，1953年。[36] W. K. Hastings，“使用马尔可夫链的蒙特卡洛抽样方法及其应用”，《生物统计学》，第57卷，第97-109页，1970年。[37] J. C. Pinheiro和D. M. Bates，《S和S-Plus中的混合效应模型》，纽约：施普林格，2000年。[38] A. D. Cobb和B. Jalaian，“通过对称分裂扩展贝叶斯神经网络的哈密顿蒙特卡洛推断”，载于 arXiv:2010.06772，2020年。[39] J. Viinikka和M. Koivisto，“用于贝叶斯网络结构学习的分层马尔可夫链蒙特卡洛”，第36届人工智能不确定性会议（UAI）论文集，机器学习研究论文集124：839-848，2020年。[40] T. Okuyama、T. Sonobe、K.-I. Kawarabayashi和M. Yamaoka，“通过动量退火进行二元优化”，《物理评论E：统计物理、等离子体、流体及相关跨学科主题》，第100卷，第1期，2019年7月。[41] T. Stollenwerk等，“将量子退火应用于空中交通管理中的最优轨迹冲突消解”，IEEE智能交通系统汇刊，第21卷，第1期，第285-297页，2020年1月。[42] E. Farhi、J. Goldstone、S. Gutmann、J. Lapan、A. Lundgren和D. Preda，“应用于随机NP完全问题实例的量子绝热演化算法”，《科学》，第292卷，475，2001年5516，472 no. pp. -。[43] M. Tao、K. Nakano、Y. Ito、R. Yasudo、M. Tatekawa、R. Katsuki等，“适用于二次无约束二元优化和伊辛模型的工作时间最优并行穷举搜索算法及其图形处理器实现”，国际并行与分布式处理研讨会专题会议论文集，第557-566页，2020年。[44] M. J¨unger等。量子退火与数字计算：实验比较。《实验算法学杂志》（JEA）26，1–30（2021年）。[45] “PyQUBO” 2022，[在线]。网址：https://pyqubo.readthedocs. io/en/latest/。[访问日期：08.06.2022]。[46] “D-Wave Ocean软件文档” 2022，[在线]。网址：https://docs.ocean.dwavesys.com/en/stable/。[访问日期：08.06.2022]。[47] K. Kurihara、S. Tanaka和S. Miyashita，“用于聚类的量子退火”，载于人工智能不确定性会议（UAI），2009年。[48] M. Sasdelli和T. -J. Chin，“二值神经网络的量子退火建模”，2021年数字图像计算：技术与应用（DICTA），2021年，第1-10页，数字对象标识符：10.1109/DICTA52665.2021.9647321。[49] L. M. R. Rere、M. I. Fanany和A. M. Arymurthy，“用于深度学习的模拟退火算法”，《计算机科学程序汇刊》，第72卷，第137-144页，2015年。[50] G. Sato、M. Konoshima、T. Ohwa、H. Tamura和J. Ohkubo，“用于整流线性单元类型函数的二次无约束二元优化建模”，《物理评论E：统计物理、等离子体、流体及相关跨学科主题》，第99卷，第4期，2019年4月。[51] T. Birdal、V. Golyanik、C. Theobalt和L. Guibas，“量子置换同步”，IEEE计算机视觉与模式识别会议（CVPR），2021年。[52] J. Li和S. Ghosh，“用于精确目标检测的量子软QUBO抑制”，载于欧洲计算机视觉会议（ECCV），2020年。[53] W. Cruz-Santos、S. E. Venegas-Andraca和M. Lanzagorta，“针对D-Wave量子退火机的立体匹配问题QUBO建模”，《熵》，第2卷，2018年。

IEEE 版权声明 版权所有 (c) 2023年 IEEE。允许将此材料用于个人用途。本文已被接收，将发表于2022年IEEE知识图谱国际会议（ICKG）。数字对象标识符：10.1109/ICKG55886.2022.00052

[54] M. Noormandipour和H. Wang，“基于量子电路学习的点集匹配”，ICASSP 2022——2022年IEEE国际声学、语音与信号处理会议（ICASSP），2022年，第8607-8611页，数字对象标识符：10.1109/ICASSP43922.2022.9746800。[55] V. Golyanik和C. Theobalt，“点集对应问题的量子计算方法”，IEEE计算机视觉与模式识别会议（CVPR），2020年。[56] B. O'Gorman、R. Babbush、A. Perdomo-Ortiz、A. Aspuru-Guzik和V. Smelyanskiy，“利用量子退火进行贝叶斯网络结构学习”，《欧洲物理学杂志专题》，第224卷，188，2015年1，163 no. pp. - 。[57]Y. Shikuri，“将贝叶斯网络学习高效转换为二次无约束二元优化”，arXiv预印本 arXiv:2006.06926，2020年。[58] D. D. Cock，“爱荷华州艾姆斯市：作为学期末回归项目的波士顿房价数据替代数据集”，《统计学教育杂志》，第19卷，第3期，2011年。[59] D. Venturclli、S. Mandr`a、S. Knysh、B. OGorman、R. Biswas和V. Smelyanskiy，“全连接自旋玻璃的量子优化”，《物理评论X》，第5卷，第3期，第031040页，2015年。[60] F. Glover、G. Kochenberger和Y. Du，“QUBO模型的建模与应用教程”，arXiv:1811.11538，2018年11月。

附录

A. 可复现性

本文描述了研究中所用合成数据的完整生成过程；除概率性元素外，合成数据均可复现。真实数据可通过 [58] 公开获取。关于源代码，使用人工数据的代码提供了充分的参数说明，可通过 PyQUBO [45] 进行计算。

B. 二值化特征的数量

本节具体介绍特征二值化方法，并说明最终二值化特征的数量。首先，假设所有特征均为连续变量或类别变量。如果未指定类型，则将所有特征视为连续变量。对于类别特征，我们设置条件，以判断某个特征是否等于某个类别变量。换言之，独特特征值有多少个，就会创建多少个二元特征。对于连续特征，我们设置条件，以判断某个特征是否大于或小于某个阈值。分位数是一种确定这些阈值的方法。设置$1/Q$，$2/Q\ldots$，$(Q-1)/Q$并取参数 Q，将生成总计 $2(Q-1)$ 个二元特征。如果所有特征都是连续特征，则可写成如下形式。

$$N_B = 2N_R(Q-1)$$

Q 的最大可能值为 N_S-1, N_B ，可表示如下。

$$N_B \leq 2N_R(N_S-2)$$

这种二值化方式会产生大量 QUBO变量，因此需要适当设置 Q 。如第 III 节所述，适当减小 Q 不仅有利于降低计算成本，也有助于提高准确率。

QUBO变量数量。

	QUBO变量数量
θ_{B,b_i}	$N_B(\leq 2N_R(N_S-2))$
θ_{X,s_i,c_i}	$N_S(M+1)$
$\theta_{s,c3,j}$	M
$\theta_{s,add,j}$	$N_S(1-2a)$

C. 最终哈密顿量与 QUBO变量数量

所提方法中的最终哈密顿量如下。

$$\begin{aligned}
H =& \frac{w_q}{N_S}\mathrm{SWMSE}(\boldsymbol{X},\boldsymbol{t}|\boldsymbol{\theta}_B,\boldsymbol{\theta}_X) \\
&+ \frac{w_{c1}}{N_S}C_1(\boldsymbol{\theta}_B,\boldsymbol{\theta}_X) + \frac{w_{c2}}{N_S}C_2(\boldsymbol{\theta}_X) + C_3(\boldsymbol{\theta}_X) + C_{add}(\boldsymbol{\theta}_X) \\
=& \frac{w_q}{N_S}\Bigg((\sum_{s_i}\theta_{X,s_i,0}t_{s_i}^2)(\sum_{s_i}\theta_{X,s_i,0}) - (\sum_{s_i}\theta_{X,s_i,0}t_{s_i})^2 \\
&+ (\sum_{s_i}(1-\theta_{X,s_i,0})t_{s_i}^2)(\sum_{s_i}(1-\theta_{X,s_i,0})) \\
&-(\sum_{s_i}(1-\theta_{X,s_i,0}t_{s_i}))^2\Bigg) \\
&+ \frac{w_{c1}}{N_S}\left(\sum_{s_i}\left(\sum_{b_i}(1-x_{b,s_i,b_i})\theta_{B,b_i} - \sum_{c_i}c_i\theta_{X,s_i,c_i}\right)^2\right) \\
&+ \frac{w_{c2}}{N_S}\left(\sum_{s_i}\left(\sum_{c_i}\theta_{X,s_i,c_i}-1\right)^2\right) \\
&+ \left((\sum_{b_i}\theta_{B,b_i} - \sum_{1\leq j\leq M}j\theta_{s,c3,j})^2 + (\sum_{1\leq j\leq M}\theta_{s,c3,j}-1)^2\right) \\
&+ \Bigg(\sum_{s_i}\theta_{X,s_i,0} - \sum_{aN_S\leq j\leq(1-a)N_S}j\theta_{s,add,j})^2 \\
&+ (\sum_{aN_S\leq j\leq(1-a)N_S}\theta_{s,add,j}-1)^2\Bigg)
\end{aligned}$$

一般而言，退火机可处理的 QUBO变量数量存在上限。所提方法创建的 QUBO问题中的 QUBO变量数量为 $N_B+N_S(M+1)+M+N_S(1-2a)$，如表 V 所示。由于 N_S、N_B 和 M 对 QUBO 变量数量有很大影响，因此必须根据任务和硬件调整这些值。在此公式中，许多 QUBO变量对之间存在相互作用。因此，最好使用能够轻松处理全连接 QUBO问题的退火机。如果难以直接处理该问题，则需要使用图小嵌入技术对问题进行转换，并确保 QUBO变量不会增加太多 [59]。