

- 1 引言
- 2 问题重述与分析
- 3 数据预处理与探索性分析
- 4 模型的建立与求解

① 引言

研究背景

研究目的与意义

② 问题重述与分析

③ 数据预处理与探索性分析

④ 模型的建立与求解

① 引言

研究背景

研究目的与意义

② 问题重述与分析

③ 数据预处理与探索性分析

④ 模型的建立与求解

研究背景

- 全球能源需求的增长与环境问题的日益严重，开发清洁、高效的可再生能源成为关键。
- 生物质是一种丰富的可再生资源，与煤共热解可利用两者的协同效应，提高产物质量和能源转化效率。

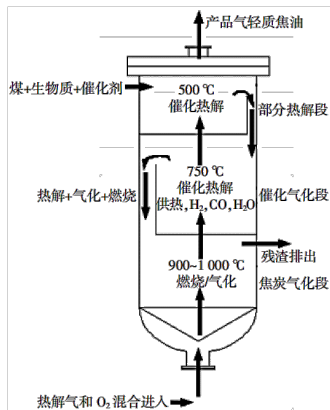


图 1: 生物质与煤共热解工艺示意图^[1]

研究背景

研究目的与意义

② 问题重述与分析

③ 数据预处理与探索性分析

④ 模型的建立与求解

- ① 问题一：分析正己烷不溶物 (INS) 对各热解产物产率的影响。

问题分析

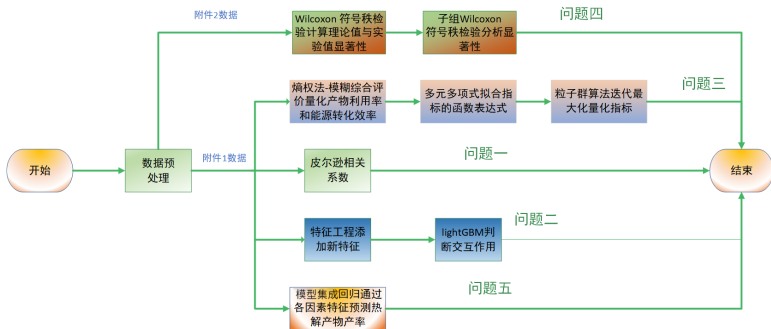


图 2: 问题分析流程图

数据可视化分析

通过可视化分析,数据的分布和变量间关系得到了更为直观的展示。图 4 显示了焦油、水、INS 等关键变量的箱线图,从中可以观察到变量的集中趋势及其潜在的异常点,各指标均不能通过正态性检验,所以在相关性分析前进行了 Z-score 标准化使数据具备标准正态分布的特征(均值为 0,标准差为 1)。

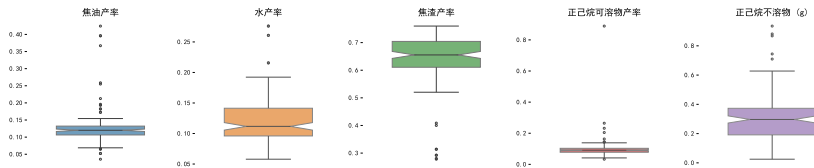


图 4: 关键变量箱线图

问题一：结果分析与结论

结果解读

- **INS 与焦油产率：**INS 含量增加有助于提高焦油产率。
- **INS 与焦渣产率：**INS 含量增加会降低焦渣产率。
- **INS 与水产率：**相关性较弱，影响不明显。

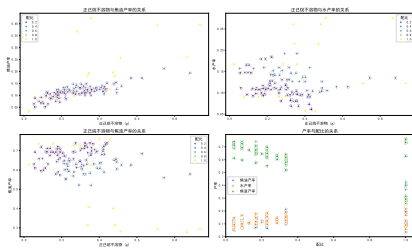


图 5: INS 与产物产率的关系

结论

INS 在热解过程中促进了焦油的生成,抑制了焦渣的形成,可能影响了热解反应的化学路径。

在散点图（图 5）中呈现了生物质混合比例与焦油产率、焦渣产率、水产率的关系。随着生物质比例的增加，焦油和水产率呈现上升趋势，而焦渣产率则下降。

问题三：熵权法-模糊综合评价模型

熵权法核心推导

给定 m 个评价对象, n 个指标, 原始数据矩阵 $X = (x_{ij})_{m \times n}$:

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^m x_{ij}}, \quad E_j = -\frac{1}{\ln m} \sum_{i=1}^m p_{ij} \ln p_{ij}, \quad w_j = \frac{1 - E_j}{\sum_{k=1}^n (1 - E_k)}$$

计算得到的评价指标权重

指标	符号	权重 w_j
焦油产率	Y_{tar}	0.3572
焦渣产率	Y_{char}	0.2980
水产率	Y_{water}	0.1981
正己烷可溶物产率	Y_{INS}	0.1467

问题三：熵权法-模糊综合评价模型

模糊综合评价实现目标融合

权重确定后，模糊综合评价将多个目标融合为单一目标。其步骤如下：

- 通过隶属函数将每个目标值标准化到区间 $[0, 1]$ 。
- 对每个目标的隶属度进行加权求和:

$$S = W \cdot R = \sum_{j=1}^m w_j \cdot r_j$$

其中 w_j 为权重, r_j 为目标的隶属度, S 为综合得分。

综合得分 S 即为单一优化目标, 从而实现将多目标问题简化为单目标优化。

问题三：熵权法-模糊综合评价模型

表 3: 评价得分情况（部分展示）

配比	样品 g	焦油 (Char) g	水 (Water) mL	正己烷不溶物 (INS) g	量化指标
1	10.5737	1.3179	0.58	0.400732	0.040207
1	10.2179	1.2832	0.59	0.4566	0.0326
1	10.3176	1.008171	0.943029	0.264736	0.003489
1	10.1371	1.075369	0.926531	0.296578	0.009528
1	9.299	0.9803	0.85	0.261862	0.007144

问题三：多元多项式拟合模型的构建与分析

模型构建理论基础

基于 Taylor 展开理论，二阶多项式可以较好地逼近局部非线性关系：

$$f(\mathbf{x}) \approx f(\mathbf{a}) + \nabla f(\mathbf{a})^T (\mathbf{x} - \mathbf{a}) + \frac{1}{2} (\mathbf{x} - \mathbf{a})^T H(\mathbf{a}) (\mathbf{x} - \mathbf{a})$$

其中， $H(\mathbf{a})$ 为 Hessian 矩阵。

拟合模型及其分析

$$f(x_1, \dots, x_5) = 0.119 + 0.176x_1 - 0.012x_2 - 0.120x_3 + 0.070x_4 - 0.071x_5 + (\text{交互项})$$

模型解释：

- x_1 （配比）和 x_4 （水）对目标函数有正面影响
- x_2 （样品）、 x_3 （焦油）和 x_5 （INS）呈负面影响
- R-squared 为 0.861528，表明模型解释了 86.15% 的数据变异性

问题三：粒子群优化算法的理论基础与应用

PSO 算法的理论基础

基于群体智能理论，粒子的运动受三个因素影响：

- 惯性：保持当前运动趋势
- 认知：个体历史最优位置的吸引
- 社会：群体最优位置的吸引

算法核心公式及其解释

速度更新：

$$v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (p_{id}^k - x_{id}^k) + c_2 r_2 (p_{gd}^k - x_{id}^k)$$

位置更新：

$$x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1}$$

其中， ω 为惯性权重， c_1 和 c_2 分别为个体和群体学习因子。

问题三：结果分析与理论验证

优化结果及其物理意义

- 最佳混合比例：28.44%
- 最大量化指标：0.3511
- 物理解释：在此比例下，生物质和煤的协同效应最强

与理论模型的一致性分析

- 结果落在 30% 临界值附近，符合文献^[2] 的预测
- 验证了低掺比下的正协同作用假说，提供了定量优化结果

第四章 结论与展望

4.1 全文结论

- 1)随着混合样中生物质掺混比例的加大,半焦产率不断降低,由 66.23%下降到 34.65%,而气体产率由 7.78%上升到 19.52%,焦油产率变化幅度同气体产率变化比较一致,也呈现上升趋势,由 8.74%上升到 18.58%,增大幅度较大,同时焦油中直链烷烃,以及一些高附加值的化合物如:萘、酚、苯、萘等含量不断增大,实现了低温煤焦油中高附加值化工产品的富集。
- 2)从 TG-DTG 曲线可以看出,随着温度升高,样品失重量逐渐增大;与煤单独热解相比,混合热解失重量及热解速率也随之提高,混合物最终剩余物质分数逐渐减小;随着生物质掺混比例增大, TG 曲线向低温区移动,煤热解起始温度逐渐降低,热解速率明显增大,达到相同失重量所需的热解时间也缩短。
- 3)生物质与煤共热解时,生物质的掺混比例低于 30%时,样品失重量比理论值大,两者之间存在有利的协同作用,促使了样品向热解方向进行,且这些协同作用均在高温处较为明显,当生物质掺混比例高于 30%,样品失重量比理论值小,两者之间的协同作用越来越弱,甚至可能抑制样品向热解方向的进行。
- 4)长焰煤、生物质及其不同配比的混合物热解动力学模型均符合一级动力学方程,煤的平均活化能 E_a 为46.43kJ/mol,指前因子为102.34 min⁻¹; 生物质的平均活化能 E_a =13.96 kJ/mol,指前因子为0.07min⁻¹;随着生物质掺混比例的增大,混合后热解表现活化能和指前因子明显减小,热解所需能量降低,系统反应速率加快,说明生物质的加入使混合热解反应更加容易的进行。

图 7: 文献验证^[2]

1 引言

2 问题重述与分析

3 数据预处理与探索性分析

4 模型的建立与求解

问题一：INS 对产率的影响分析

问题二：INS 与混合比例的交互效应

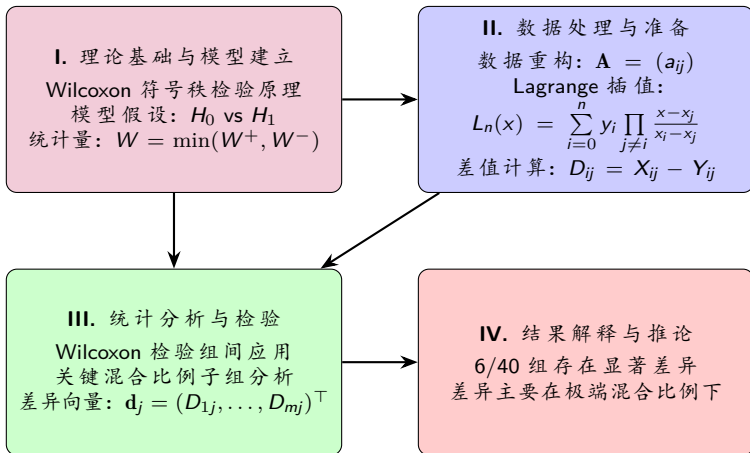
问题三：最佳混合比例的确定

问题四：差异显著的热解组合

问题五：热解产物产率预测模型

研究展望

© 2004 Blackwell Publishing Ltd, *Journal of Internal Medicine* 255: 101–108



方法选择理由:

采用 **Wilcoxon** 符号秩检验, 适用于**小样本**和**非正态分布**的数据, 能够考虑差值的**符号和大小**, 适合**配对样本**的比较, 对**异常值**不敏感, 提高结果的稳健性。

通过 Wilcoxon 符号秩检验，针对不同组别进行了组间和子组分析。以下展示了组间分析和各子组的检验结果。

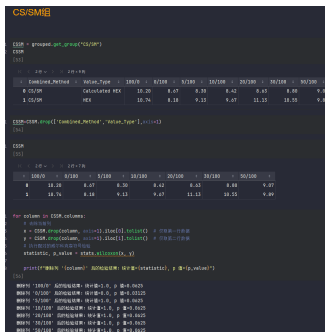


图 8: 组间分析

图 9: CS/SM 组分析

[illegible]

```

SDHS组

1 SDHS = grouped.get_group("SD/HS")
2 SDHS=SDHS.drop(['Combined_Method','Value_Type'],axis=1)
3 for column in SDHS.columns:
4     # 对每列进行
5     x = SDHS.drop(column,axis=1).iloc[0].tolist() # 获取第一行数据
6     y = SDHS.drop(column,axis=1).iloc[1].tolist() # 获取第二行数据
7     # 执行配对的威尔逊符号检验
8     statistic,p_value = stats.wilcoxon(x,y)
9
10 print("“{}列” 后的检验结果：统计量={statistic}，p值={p_value}")
11
12 [59]
13
14 删除列 '100/0' 后的检验结果：统计量=1.0，p值=0.0625
15
16 删除列 '0/100' 后的检验结果：统计量=0.0，p值=0.03125
17
18 删除列 '5/100' 后的检验结果：统计量=2.0，p值=0.09375
19
20 删除列 '10/100' 后的检验结果：统计量=2.0，p值=0.09375
21
22 删除列 '20/100' 后的检验结果：统计量=2.0，p值=0.09375
23
24 删除列 '30/100' 后的检验结果：统计量=2.0，p值=0.09375
25
26 删除列 '40/100' 后的检验结果：统计量=2.0，p值=0.09375
27
28 删除列 '50/100' 后的检验结果：统计量=2.0，p值=0.09375
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
95
```

39 / 57

Wilcoxon 符号秩检验运行结果

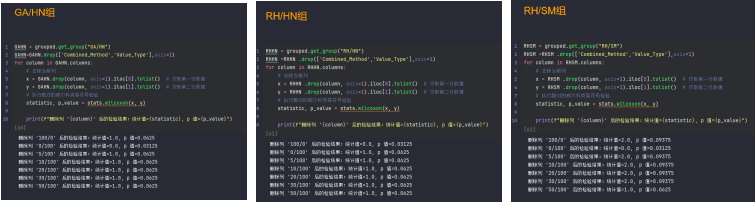


图 12: GA/HN 组分析

图 13: RH/HN 组分析

图 14: RH/SM 组分析

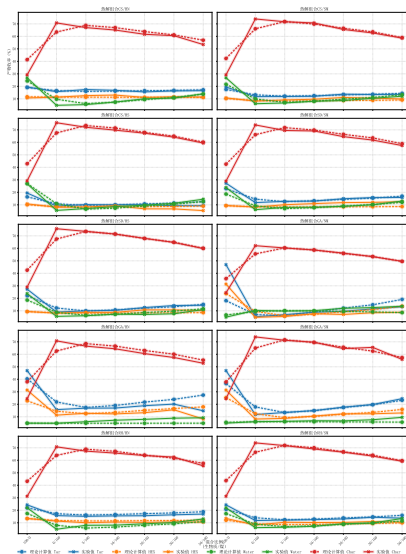
问题四：差异显著的热解组合

差异显著的热解组合		共热解组合子组分析结果		
组合	物质类型	组合	关键混合比例	剔除后 p 值
CS/SM	HEX	CS/SM	0/100	0.03125
SD/SM	HEX	SD/SM	0/100	0.03125
SD/HS	HEX	SD/HS	0/100	0.03125
GA/HN	Water	GA/HN	0/100	0.03125
RH/HN	HEX	RH/HN	100/0	0.03125
RH/SM	HEX	RH/SM	0/100	0.03125

表 4: 差异显著的热解组合及子组分析结果

主要发现：在40 组实验中，有6 组存在显著差异，主要涉及HEX（正己烷）产物。关键混合比例集中在纯煤（0/100）和高比例生物质（100/0），表明在极端混合比例下，理论模型的预测准确性降低。非极性产物（HEX）的产率预测偏差较大，需进一步研究生物质与煤在极端比例下的相互作用机制，考虑改进理论模型，纳入更多影响因素。

不同生物质煤共热解与实验产物收率比较可视化



答辩人：陆子凯、狄诗琪

生物质和煤共热解技术中特征选择和优化研究

Team # 2024050727407, Problem B

42 / 57

1 引言

2 问题重述与分析

3 数据预处理与探索性分析

4 模型的建立与求解

- 问题一：INS 对产率的影响分析
 - 问题二：INS 与混合比例的交互效应
 - 问题三：最佳混合比例的确定
 - 问题四：差异显著的热解组合
 - 问题五：热解产物产率预测模型
- 研究展望

问题五：热解产物产率预测模型 - 理论基础与方法选择

问题分析

生物质-煤共热解过程是一个复杂的热化学转化过程, 其产物产率预测涉及多个挑战:

- 多因素影响：配比、样品质量、焦油量、水量等因素共同作用。
- 非线性关系：因素间可能存在复杂的非线性交互作用。
- 数据特征：实验数据可能存在噪声和不确定性。
- 预测精度要求：需要在有限样本条件下实现高精度预测。

这些特点决定了我们需要选择能够适应复杂系统、捕捉非线性关系、并具有良好的泛化能力的预测模型。

理论依据

基于问题特性, 我们选择了三种不同类型的模型:

- ① 多元线性回归：适合捕捉线性关系，提供基础因素分析。
- ② 随机森林回归：非参数模型，适合处理非线性关系与复杂特征交互。
- ③ 高斯过程回归：非参数回归，在贝叶斯框架下优化选用，尝试量化不确定性并通过核函数捕捉复杂模式。

这种多模型策略允许我们从不同角度分析问题，比较模型性能，并获得更全面的预测结果。

问题五：多元线性回归模型

模型原理

多元线性回归模型假设因变量（热解产物产率）与自变量（影响因素）之间存在线性关系：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon$$

其中， y 为焦油产率， x_i 为影响因素， β_i 为回归系数， ε 为随机误差项。
模型估计采用最小二乘法，最小化残差平方和：

$$\min_{\beta} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}))^2$$

通过求解正规方程 $(X^T X)\beta = X^T y$ ，我们可以得到参数估计 $\hat{\beta} = (X^T X)^{-1} X^T y$ 。

问题五：多元线性回归模型

参数估计结果与解释

拟合得到的模型为:

$$\hat{y} = 0.0000 - 0.3108x_1 - 0.0431x_2 + 0.0194x_3 + 0.7672x_4$$

其中, x_1 : 配比, x_2 : 样品质量, x_3 : 焦油量, x_4 : 水量。

解释：

- 配比 (x_1) 和样品质量 (x_2) 对产率有负面影响, 可能是由于高配比和大样品量导致热解不充分。
- 焦油量 (x_3) 和水量 (x_4) 对产率有正面影响, 特别是水量的影响最为显著 (系数最大)。

问题五：多元线性回归模型结果可视化

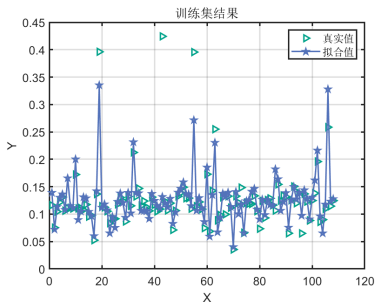


图 16: 训练集拟合结果

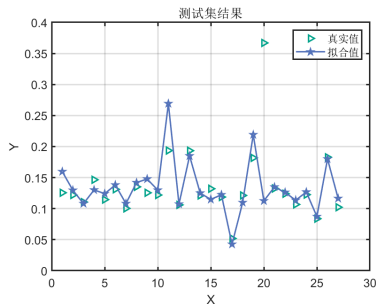


图 17: 测试集预测结果

从图中可以看出，模型在训练集上具有较好的拟合度，但在测试集上仍然存在一定的偏差，表明线性模型可能无法完全捕捉系统中的非线性特征。

问题五：随机森林回归模型

模型原理

随机森林是一种集成学习方法，通过构建多个决策树并结合它们的预测来进行回归：

- 基于 Bootstrap 采样构建多个决策树，每棵树使用随机选择的特征子集。
- 决策树生长过程中，节点分裂基于最小化不纯度（如 Gini 指数或均方误差）。
- 最终预测通过平均所有决策树的预测结果得到。

模型优势与局限性

优势：

- 随机森林能够有效处理非线性关系和特征交互。
- 不受多重共线性影响，适用于高维数据。
- 可以评估特征重要性，具有良好的解释性。

局限性：

- 模型复杂度高，训练和预测时间可能较长。
- 难以捕捉超出训练数据范围的模式（外推能力有限）。
- 可能出现过拟合，尤其是当树的数量不足或深度过大时。

问题五：随机森林模型结果可视化

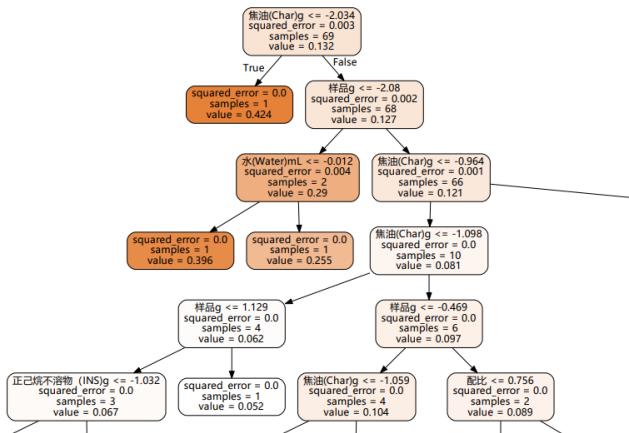


图 18: 随机森林中的一棵决策树部分可视化

问题五：贝叶斯优化-高斯过程回归模型

模型原理

高斯过程回归（GPR）是一种基于贝叶斯框架的非参数回归方法：

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

其中， $m(x)$ 是均值函数， $k(x, x')$ 是协方差函数（即核函数），通常采用 RBF 核：

$$k(x, x') = \sigma_f^2 \exp(-\frac{1}{2l^2} \|x - x'\|^2)$$

该模型能够量化预测的不确定性，并通过核函数捕捉复杂的非线性关系。

贝叶斯优化与超参数选择

- 贝叶斯优化可以帮助我们自动调整 GPR 的超参数（如 σ_f^2 、 l 、 σ_n^2 ），以提高模型性能。
- 通过最大化边际似然（或最小化负对数边际似然）来选择最优的超参数组合。

问题五：贝叶斯优化-高斯过程回归模型结果可视化

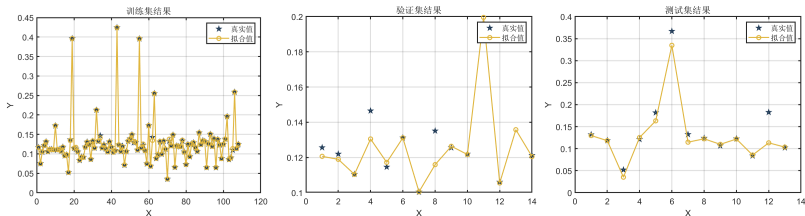


图 19: 高斯回归-贝叶斯优化模型结果

模型性能比较与结果分析

评估指标对比

模型	MAE	MSE	RMSE
线性回归	0.0215	0.0028	0.0527
随机森林	0.0201	0.0031	0.0561
高斯过程	0.0130	0.0005	0.0229

结果分析

- 高斯过程回归在所有指标上都优于其他两种方法，特别是在 MAE 和 RMSE 方面有显著改善。
- 随机森林在捕捉非线性特征方面表现不错，但在 MSE 和 RMSE 上的表现稍逊。
- 线性回归由于其线性假设，在捕捉复杂特征方面表现不佳。

① 引言

② 问题重述与分析

③ 数据预处理与探索性分析

④ 模型的建立与求解

问题一：INS 对产率的影响分析

问题二：INS 与混合比例的交互效应

问题三：最佳混合比例的确定

问题四：差异显著的热解组合

问题五：热解产物产率预测模型

研究展望

- ◀ ◻ ▶ ◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ↺ 🔍 ↻

感谢您的聆听！