可解释数据挖掘实践

数据挖掘课程大作业

TA：李振宇 孙宇涛 潘藤予 刘方舟

2025 秋

目录

1. 任务背景

- 现有的机器学习分类模型从性能和可解释性两个维度大致分为两类：
  - 以深度学习和集成学习（如 **随机森林、XGBoost** ）为代表的分类模型具有良好的性能，但模型复杂度高、可解释性差
  - 以 **决策树和逻辑回归** 为代表的模型可解释性强，但性能不理想
- 随着可解释性越发受到人们重视，近期不少工作提出了保持可解释性同时具有不错性能的模型（如 **RRL** ）
- 本次作业，我们将通过探索可解释模型在现实任务中的应用，让大家对先进的数据挖掘模型的特点有初步的了解和认识，若有兴趣可以对其可解释性做出进一步分析和改进
- **本次作业为组队作业，与专题展示共用一组**

2. 任务描述

（点击数据集名称即可进入对应链接）

- **Bank Marketing**
  - 经典 **分类任务** ，预测人群是否会在银行进行存款，具有混合的数据特征
  - 下载数据集后，统一使用 bank.zip 中的 bank-full.csv
- **Boston Housing**
  - 经典 **回归任务** ，根据特征预测房价

3.1 模型选择：可解释模型

- 我们建议使用规则表征学习器 **RRL** [1]，可直接基于 **开源代码** 进行修改
  - RRL 能够通过自动学习可解释的非模糊规则进行数据表征和分类，同时在效果上，RRL 显著优于其他可解释模型，与 **LightGBM** 和 **XGBoost** 等复杂模型有着相当的结果
  - 为了适配回归任务，需要对 loss 函数进行简单修改
  - 如果有熟悉的其他可解释模型也可以使用

Dataset	**RRL**	C4.5	CART	SBRL	CORESL	CRS	LR	SVM	PLNN	BNN	RF	LGBM	XGB	FT	SAINT
adni	84.68	79.71	81.02	80.20	81.19	83.10	83.83	83.93	84.65	83.49	83.82	**84.76**	84.65	82.21	84.33
adult	80.42	77.77	77.06	79.88	70.56	**80.95**	78.43	79.01	79.60	77.26	79.22	80.36	80.64	79.01	79.31
bank-marketing	**77.18***	71.24	71.38	72.67	66.86	73.34	69.81	72.99	72.40	72.49	72.67	75.28	74.71	77.04	75.60
banknote	**100.0***	98.45	97.85	94.44	98.49	94.93	98.82	**100.0**	**100.0**	99.64	99.40	99.48	99.55	99.93	99.04
chess	89.73*	79.90	79.15	26.44	24.86	80.21	33.06	87.04	77.85	55.03	75.00	88.73	**90.04**	87.88	86.73
connect-4	72.01*	61.66	61.24	48.54	51.72	65.88	49.87	69.85	70.64	61.94	62.72	70.53	70.65	72.45	**72.85**
letRecog	96.14*	88.20	87.62	64.32	61.13	84.96	72.05	95.57	92.34	81.06	96.59	96.51	96.38	**97.17**	96.72
magic04	86.29*	82.44	81.20	82.52	77.37	80.87	75.72	85.35	85.50	79.50	86.48	86.67	**86.69**	85.95	85.46
tic-tac-toe	**100.0**	91.70	94.21	98.39	98.92	99.77	98.12	98.07	98.26	98.92	98.37	99.89	99.89	97.84	96.95
wine	98.37	95.48	94.39	95.84	97.43	97.78	95.16	96.05	76.07	95.77	98.31	**98.44**	97.78	94.63	95.50
activity	98.96	94.24	93.35	11.34	51.61	5.05	98.47	98.67	98.27	97.86	97.80	**99.41**	99.38	98.56	98.94
dota2	**60.08***	52.08	51.91	34.83	46.21	56.31	59.34	59.25	59.46	54.76	57.39	58.81	58.53	59.70	59.58
facebook	**90.11***	80.76	81.50	31.16	34.93	11.38	88.62	87.20	89.43	85.94	87.49	85.87	88.90	86.52	88.79
fashion	89.64*	80.49	79.61	47.38	38.06	66.92	84.53	**90.09**	89.36	85.33	88.35	89.91	89.82	89.23	89.69
AvgRank	**2.50**	11.57	12.14	11.86	12.64	9.29	10.43	6.21	6.79	9.71	7.29	3.86	3.50	6.14	5.57

3.2 模型选择：其他模型

- 除可解释模型之外，每个任务选择至少 2 个模型或算法作为对比
- 其中每个任务都需要自己 **手动实现** 至少 1 个模型
  - 可以手动实现 1 个模型，同时适配分类和回归任务
  - 也可以手动实现 2 个模型，分别适配两种任务
  - 不可直接调包或使用开源代码

4.1 实验内容：数据分析及预处理

- 对数据进行分析，以发现数据存在的可能影响后续模型训练的问题
- 根据数据分析的结论，对数据集进行预处理
  - 包括数据清洗、数据归一化、数据缺失处理等等
- RRL 需要制备 data 和 info 文件
  - 可参考开源仓库中 dataset 文件夹下的 README.md

4.2 实验内容：模型训练与评估

(1) 数据集划分

- 训练集、验证集和测试集如何划分
  - RRL 中默认 5 折交叉验证，通过 `-ki` 参数选择折数；也可以设计其他划分方式

(2) 模型选择

- 为什么选择这些模型（每个任务需手动实现至少一种）

(3) 评价指标选择

- 结合之前预处理的分析以及可能的实际应用需求、不同的任务类型，选择合适的评价指标对各个模型进行评价

(4) 模型调参

- 通过适当的调参方法，尽量避免过拟合或欠拟合等情况

4.2 实验内容：模型训练与评估

(5) 模型间的对比与分析

- 不同模型的优劣及背后的原理，适用于什么样的任务类型
- 可以从以下角度进行分析：
  - 性能：结合之前选择的评价指标进行分析
  - 可解释性：模型解释的可理解性，以及这些解释对模型真实决策过程的忠实度
  - 复杂度：RRL 以及基于决策树的模型可以计算连边数量之和的自然对数
    - RRL 代码中已实现 $\log(\#\text{edges})$ 的计算，见 experiment.py 中的 `test_model()` 函数
    - 也可以选择其他复杂度指标
  - 还可以选择其他角度进行分析

4.3 实验内容：选做（Bonus）

- 可解释特性的进一步探索
  - 产生的规则有多大的实际意义?
  - 模型效果对规则条数是否敏感？可观察 RRL 保留不同规则条数对性能的影响
- 模型结构优化
  - 通过调整模型结构提升模型效果
  - 如更改 RRL 的逻辑激活函数、合取与析取层的结构，在计算过程中添加扰动等等

5. 报告提交

- **时间：第 16 周周四（2026 年 1 月 1 日）23:59 前** 提交到网络学堂
  - 网络学堂会设置为分组作业，每组提交一次即可
- **格式**：要求 PDF 格式，文件名不限，与代码打包为 zip 压缩包提交
- 只需提交报告和完整代码，无需提交中间结果和数据文件
- 报告需包含以下内容：
  - 所有实验内容的过程和结果
  - 必要的分析解释和图表
  - 小组各成员分工情况

6. 课堂展示

- **时间：第 16 周周五（2026 年 1 月 2 日）课上**
- **PPT 提交：第 16 周周四（2026 年 1 月 1 日）23:59 前** 提交到网络学堂
  - **PPT 命名格式**：“组号-组长姓名-组长学号”
  - 网络学堂会设置为分组作业，每组提交一次即可
- **时长**：每组 5 分钟展示 + 1 分钟互动，**请严格控制时间**
- **评分**：助教打分
- 每组上台展示的人数不限
- 教学团队后续将在网络学堂上发布展示顺序

参考文献

[1] Z. Wang, W. Zhang, N. Liu, 和 J. Wang, 《Learning Interpretable Rules for Scalable Data Representation and Classification》, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 卷 46, 期 2, 页 1121–1133, 2024, doi: 10.1109/TPAMI.2023.3328881.

Thanks

Q & A