医渡云
YIDUCLOUD

知识图谱与机器学习的融合
——可解释的模型构建

医渡云

CCKS 2020

综述

医渡云
YIDUCLOUD

医疗健康领域的可解释性需求

- 临床辅助决策支持系统
- 疫情防控中的风险评估模型

模型可解释性问题

- 问题定义
- 意义和价值

面向模型可解释性的相关工作

- 相关领域研究分类逻辑
- 相关工作的系统介绍

应用案例

- 案例一：规则对知识的逼近
- 案例二：知识图谱结构下的深度网络桥接

总结与展望

医渡云对AI模型可解释性的诉求

医渡云
YIDUCLOUD

临床辅助决策系统

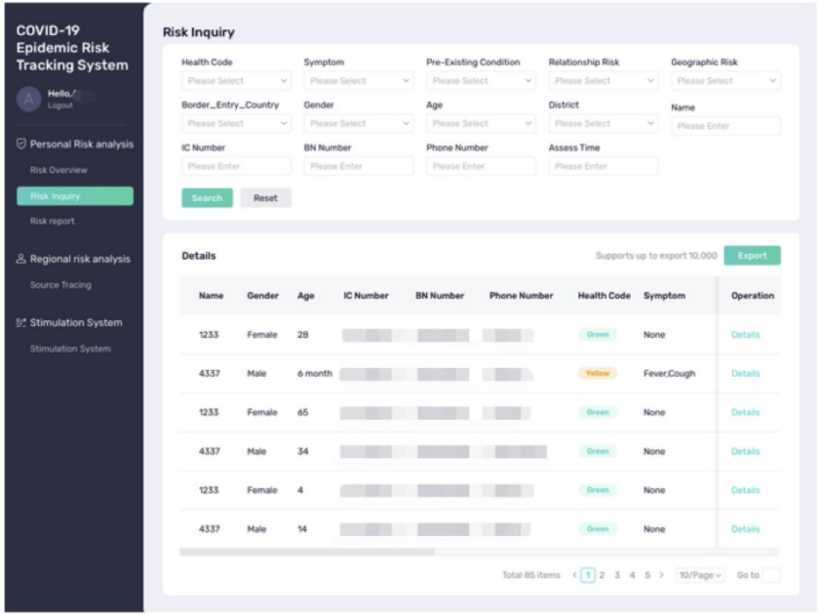
- 以肿瘤辅助治疗决策为例，既有知识无法应对所有场景
- 相似病例是否足以支撑治疗方案推荐？
- 如何像资深专家一样解释诊断决策的来龙去脉？

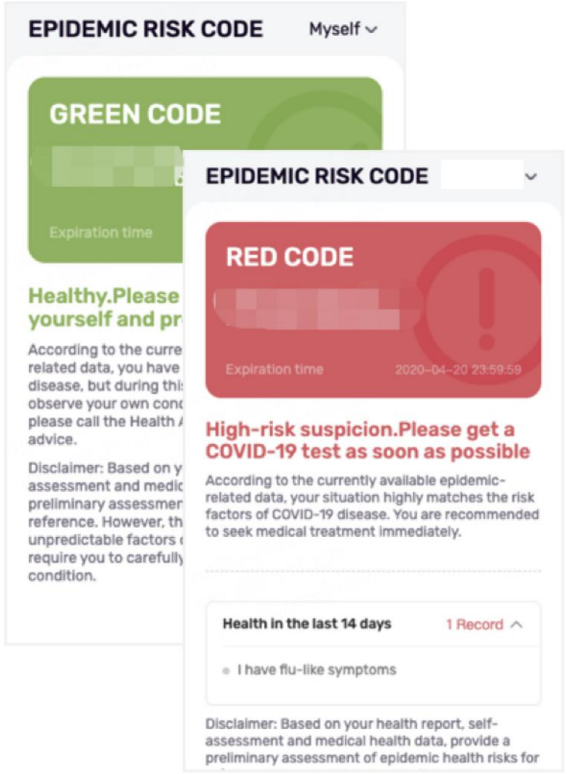
医疗健康领域的可解释性需求

医渡云 YIDUCLOUD

防疫健康码

- 如何评价标准和最终发码的决策过程？
- 给红码的原因？给绿码的原因？





模型的可解释性问题

医渡云
YIDUCLOUD

深度学习中的可解释性

- 深度学习极高的特征抽象和非线性拟合能力
- 结果以外的决策过程：智能 or 曲线拟合?

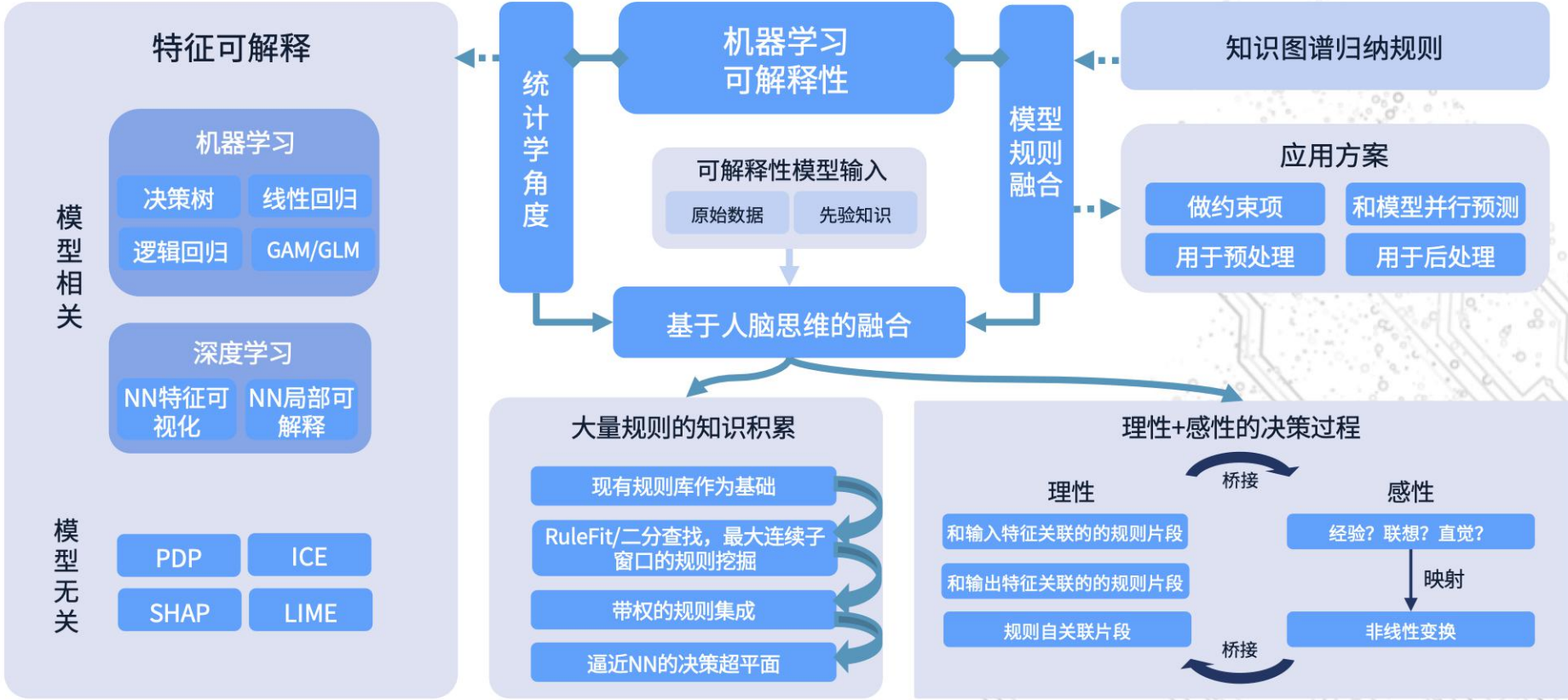
“黑盒”带来的阻碍

- 严谨场景下对统计机器学习的偏向
- 示例：各类医学决策，金融量化投资，个人征信评估等

更多应用角度

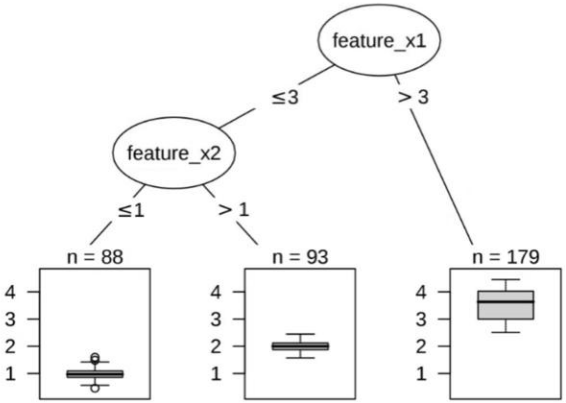
- 更充分理解需要解决的问题
- 模型的定向优化

机器学习可解释性-逻辑框架



模型相关类方法

- **典型代表：决策树 【Wenjing，2018】**
- **通过决策路径的回溯直观地实现可解释性**
  - 从根节点开始，根据边的判断转到下一个子集，直到叶节点。
  - 当树的节点和层级变大时，解释将变得复杂和困难
  - 同类其他模型：逻辑回归，线性回归，GAM
  - 示例：常规决策路径（左），基于样本分析特征贡献（右）



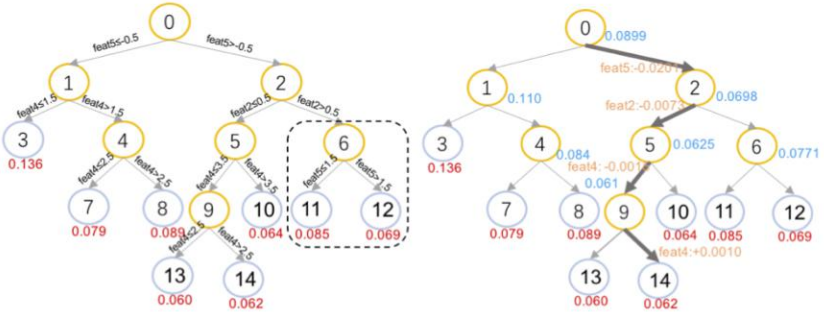


Fig. 1. Feature Contribution Example for GBDT

模型相关类方法

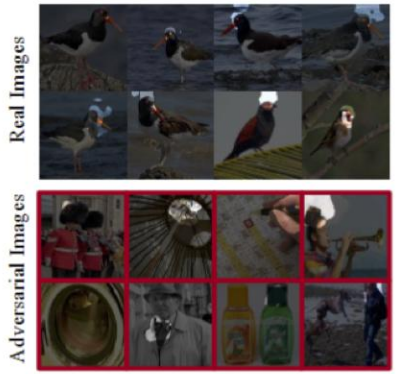
- **典型代表：卷积神经网络 【Christoph，2019】**
- **通过可视化学习到特征的最大化激活值来实现可解释性**
  - 卷积神经网络从原始图像像素中学习抽象特征和概念
  - 构建神经网络单元的(平均)激活最大化的新图像
  - 存在问题：激活的新图像依赖于人类的识别能力
  - 示例： 神经元被轮子正向激活的图像（左图）；神经元被眼睛负向激活后的图像（右图）

$$img^* = \arg\max_{img} \sum_{x,y} h_{n,x,y,z}(img)$$

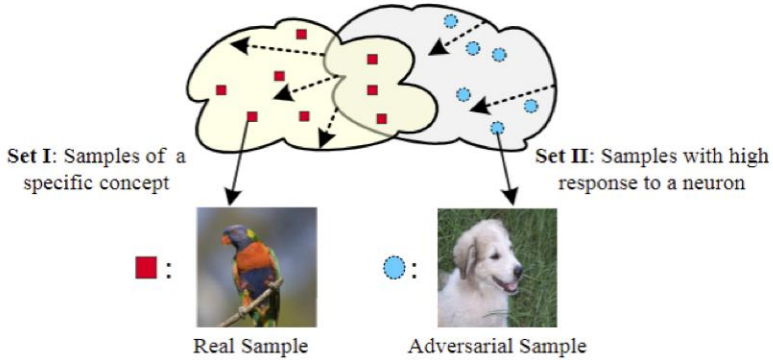
模型相关类方法

- 典型代表：卷积神经网络 【Yingpeng，2019】
- 通过挖掘决策性子网络实现可解释性
  - 网络各层对特征的抽象具有不同意义
  - 挖掘对最终决策影响最大的局部子网络/神经元
  - 存在问题：在测试集分布发生较大偏移后失效

示例： VGG-16对鸟进行分类，pool5层17号神经元对鸟头部某特征响应强烈，造成对抗样本的错误识别



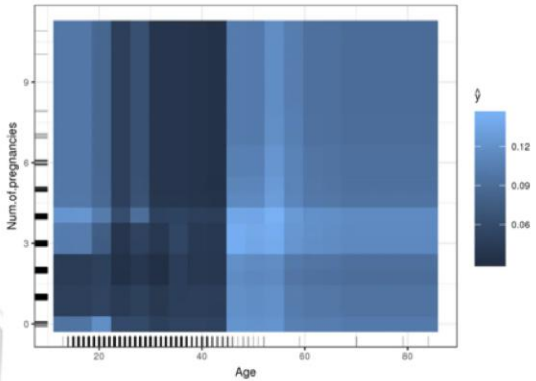
(a) Neuron 147: Bird



模型无关类方法

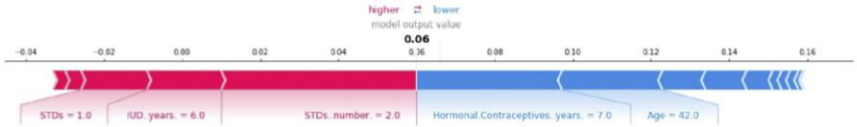
- **典型代表：部分依赖图 (PDP) 【Christoph，2019】**
- **通过特征和预测结果的关系可视化实现可解释性**
  - 显示特征对先前模型预测结果的边际效应(J. H. Friedman 200126)
  - 通过对其他特性的边缘化，得到仅依赖于目标特征的函数
  - 存在问题：最大特征数量为2；存在独立性假设
  - 类似方法：ICE图,ALE图
  - 示例：乳腺癌和怀孕次数，年龄相关性

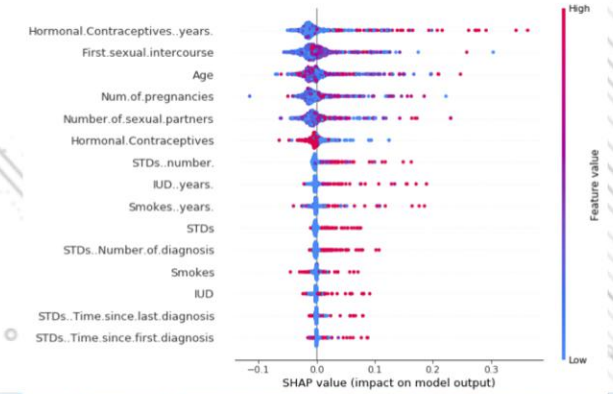
$$\hat{f}_{xS}(x_S) = \frac{1}{n}\sum_{i=1}^{n}\hat{f}\left(x_S, x_C^{(i)}\right)$$



模型无关类方法

- 典型代表：SHAP【Scott，2017】
- 通过特征归因方法解释模型输出
  - 通过计算在合作中个体的贡献来确定该个体的重要程度
  - 最大的优势是SHAP能对于反映出每一个样本中的特征的影响力，而且还表现出影响的正负性。
  - 存在问题：计算效率低。
  - 示例：对样本的解释（左），对特征的解释（右）

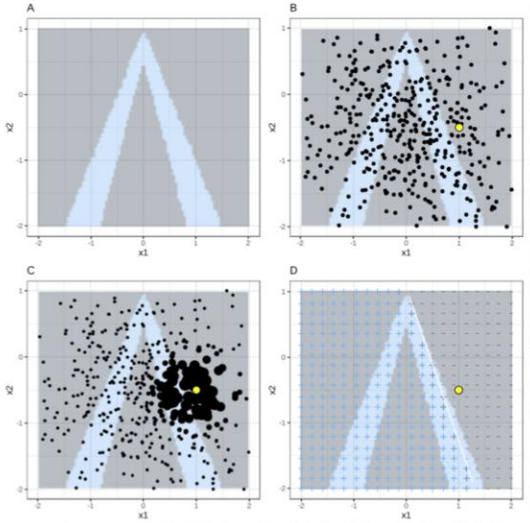




模型无关类方法

- 典型代表：LIME【Thomas，2017】
- 使用简单模型对复杂模型的局部进行解释
  - 构建新的数据集，并训练简单模型g
  - 希望原始模型f与新模型g预测值之间的误差尽可能小
  - 存在问题： 定义需要解释的点的周围边界
  - 示例：对于非线性分类器的局部可解释性

$$\mathcal{L}(f, g, \Pi_x) = \sum_{z, z' \in \mathcal{Z}} \Pi_x(z) \left(f(z) - g(z')\right)^2$$

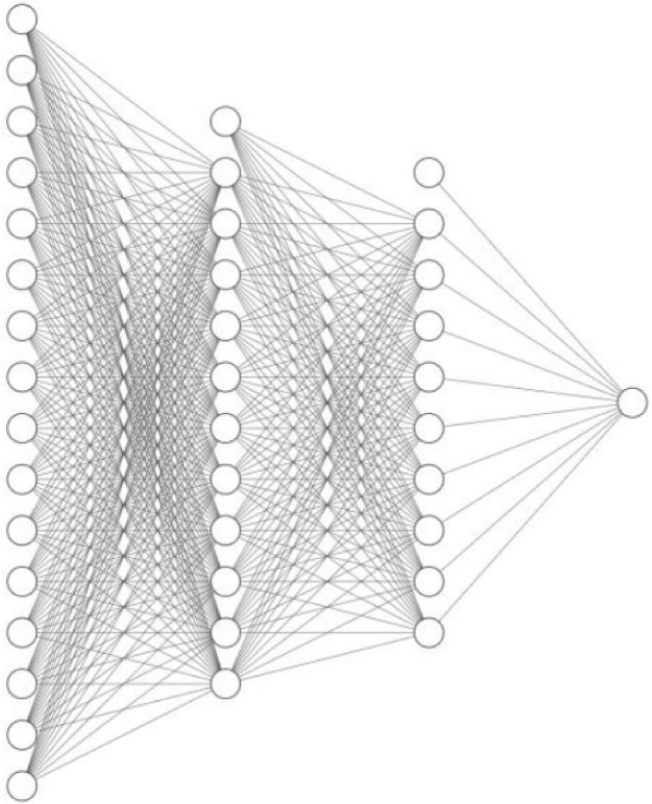


挑战

医渡云 YIDUCLOUD

- **融合符号逻辑和机器学习模型融合的难点**
  - 符号逻辑如何引入机器学习模型?
  - 复杂网络如何做到可解释?
  - 人脑的思维过程是什么?

- $\forall x\ ((even(x) \vee odd(x)) \wedge \neg(even(x) \wedge odd(x)))$
- $\forall x\ (even(x) \leftrightarrow 2 \mid x)$
- $\forall x\ (even(x) \rightarrow even(x^2))$
- $\forall x\ (even(x) \leftrightarrow odd(x+1))$
- $\forall x\ (prime(x) \wedge x > 2 \rightarrow odd(x))$
- $\forall x\ \forall y\ \forall z(x \mid y \wedge y \mid z \rightarrow x \mid z)$



符号逻辑融合机器学习

医渡云 YIDUCLOUD

- 典型代表：规则引擎作为约束项【Nicholas，2017】
- 将规则融入到神经网络模型的损失函数中，从而定向对预测方向进行优化

  - 当我们需要寻找目标函数最小的θ值，有如下公式：

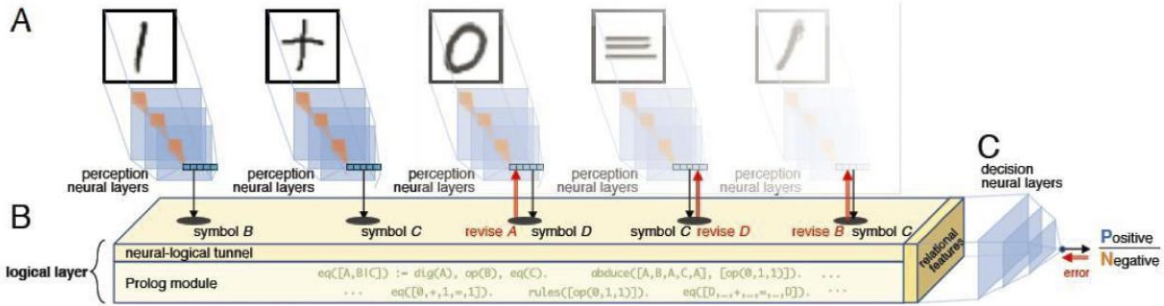
$$\theta^* = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} L(y_i, f(x_i; \theta)) + \lambda \phi(\theta)$$

  - 其中 $\phi(\theta)$即正则化项，L为损失函数
  - 设定 $\phi(\theta)$为L1范数，视为规则引擎，则模型优化方向将会使参数稀疏化
  - 缺点：当规则增加后，难以确定模型优化方向

符号逻辑融合机器学习

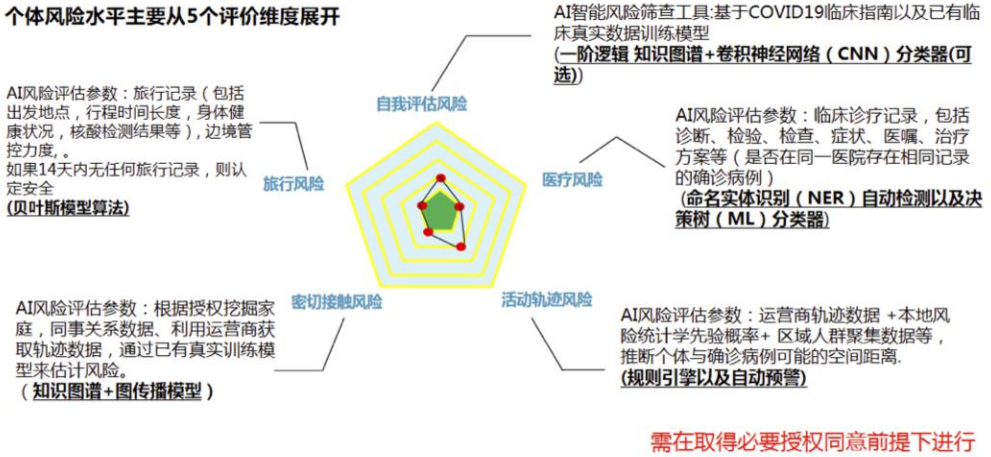
- 典型代表：规则引擎和模型并行优化【Dai，2018】
- 通过逻辑溯因和神经感知同时训练模型并做出决策

- Perception neural layer：用于区分手写图片代表的符号
- Logical layer：使用已知KB计算具体式子形式，训练perception neural layer，计算得到relation feature
- Decision neural layer：对relation feature进行分类



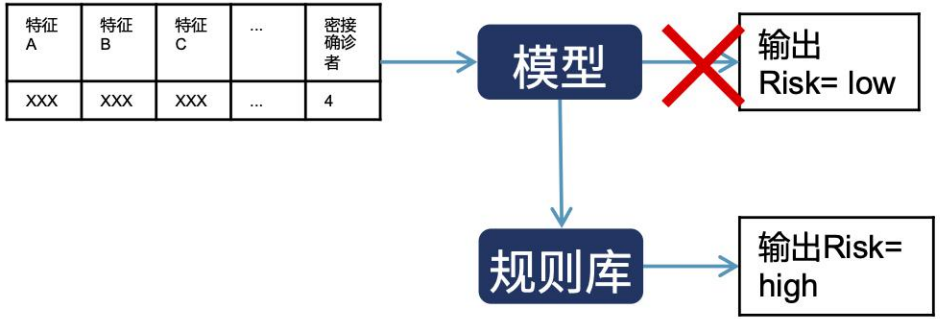
符号逻辑融合机器学习

- **典型代表：规则引擎作为预处理**
- **规则引擎可以对模型的输入数据进行筛选或者转化**
  - 利用已有知识对输入数据进行转化，减少模型计算负担
  - 以疫情的个人风险评估为例，活动轨迹风险不直接采用运营商轨迹数据
  - 通过规则引擎进行参数评估，再和其他风险结合做后续计算



符号逻辑融合机器学习

- **典型代表：规则引擎作为后处理**
- **规则引擎作为各类模型决策前的最终步骤**
  - 神经网络模型作为黑盒模型，很难对决策进行定向优化
  - 后处理引入规则引擎，使模型最终决策服从已知规则
  - 例如在评估新冠密切接触风险时模型的评估修正

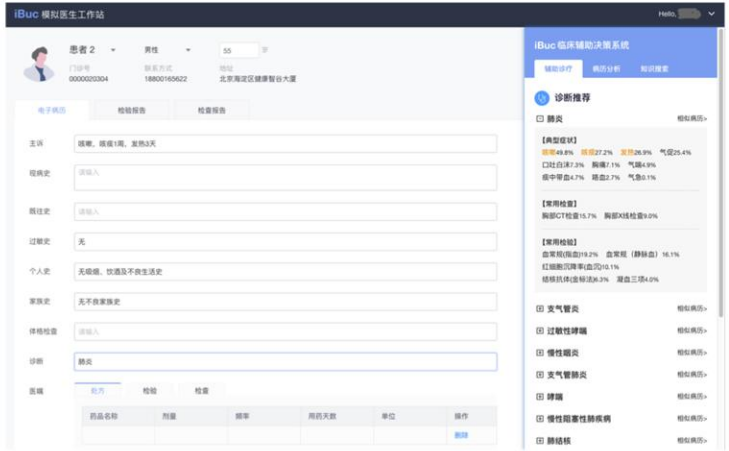


医渡云应用案例一

医渡云
YIDUCLOUD

临床辅助决策支持系统-决策的解释

- 对于各类诊断决策，既有知识仅可实现部分路径的解释
- 通过规则积累和深度学习，可解释更多诊断的决策
- 输入：患者电子病历，检查检验结果
- 输出：推荐的诊断



诊断推荐

根据输入的患者信息和主诉现病史等信息，当医生下诊断时，自动推荐疑似诊断，并提供这些诊断的典型症状、常用检查、常用检验以及医院内相似病历。

1）疑似疾病典型症状展示，匹配的症状飘红展示；

2）疑似疾病常用检查推荐，以及在医院相似病历中开立占比。帮助医生进一步确诊；

3）疑似疾病常用检验推荐，以及在医院相似病历中开立占比。帮助医生进一步确诊；

4）相似病历推荐。根据患者特征、主诉/现病史、当次疑似诊断，在本院召回相似病历数据；

5）推荐诊断、推荐检查/检验，支持HIS监听用户点击事件，以完成回填功能。

医渡云应用案例一

医渡云 YIDUCLOUD

- **融合方案：规则对知识逼近的方法**
  - 在足够知识的累积下，规则引擎可以无限逼近非线性分类器
  - 以二维图做示例，通过rule的集成，逼近曲线的决策
  - 现有规则基础上，规则挖掘方案：

Algorithm 1: Rule-Mining

Data: Based on current rule, finding the key features and store to S
Result: Rules mined from S

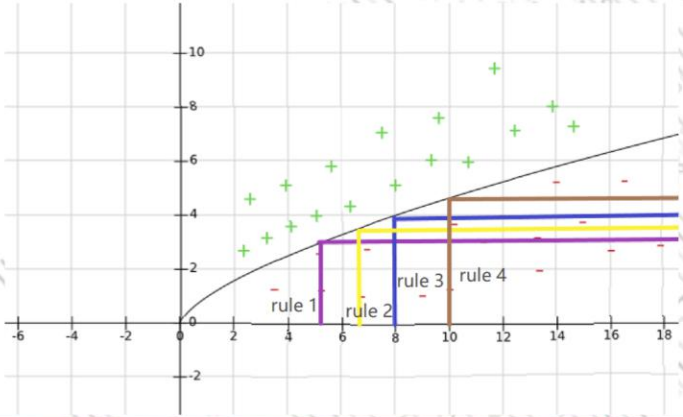
```
Rule_Set=[]
for Fea_Pair in S:
    range1=Sub_Window_List(Fea_pair[0])
    range2=Sub_Window_List(Fea_pair[1])
    loss= inf
    while(loss >torlerance):
        for sub_l1 in range1:
            for sub_l2 in range2:
                loss=cal_loss_w_range(Rule_Set,sub_l1,sub_l2)
    Rule_Set.append(Fea_Pair:[sub_l1,sub_l2])

def Sub_Window_List(Fea):
    using sliding window to generate the range for candidate features.
    Size of sliding window= [(max(size)*i/100) for i in range[100,0]]
```

$$ML \rightarrow g(z)$$

$$rules \rightarrow f_1, f_2, \ldots$$

$$f = \lim_{i \to \infty} f_i \approx g(z)$$

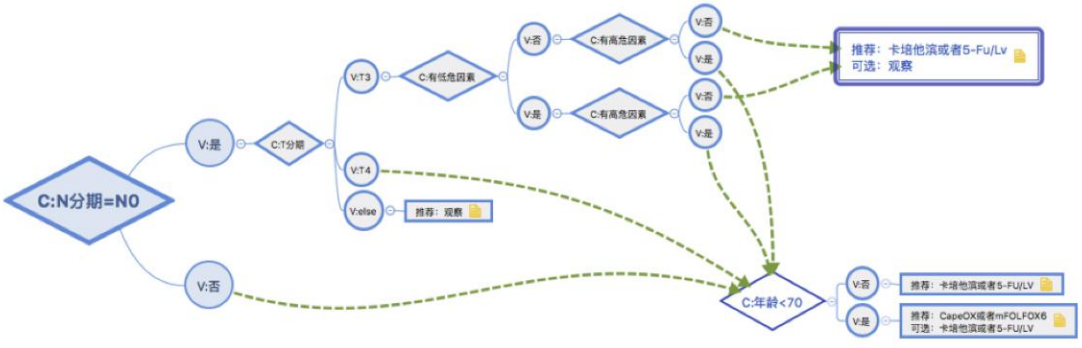


医渡云应用案例一

医渡云 YIDUCLOUD

- **实例：CDSS中根据肿瘤分期做治疗方案的推荐**
  - M1: 根据现有医学和专家指南构建规则库，无法涵盖全局
  - M2: 根据规则库+DNN并行决策模型
  - M3: 现有知识基础上，多轮动态阈值下的规则挖掘扩充方案
  - 示例: 某类肿瘤治疗方案决策路径

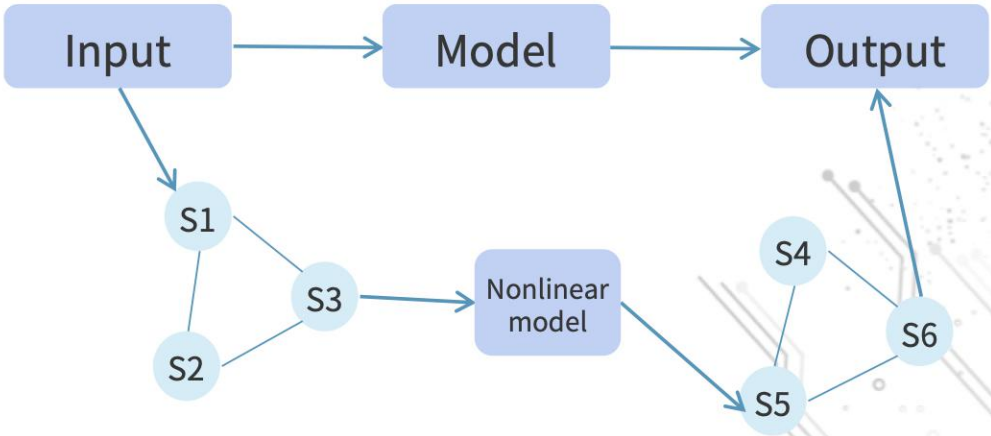
模型	F1 score
M1	0.772
M2	0.833
M3（目前最多支持5特征组合）	0.817



医渡云应用案例二

医渡云
YIDUCLOUD

- **融合方案：知识图谱结构下的深度网络桥接**
  - 感性决策→缺乏规则基础，不可控
  - 理性决策→难以应对所有场景，规则集有限
  - 人脑的思维过程→理性（规则）+感性（非线性变换）
  - 如图所示，S为知识片段，模型只学习S3到S5的变换过程

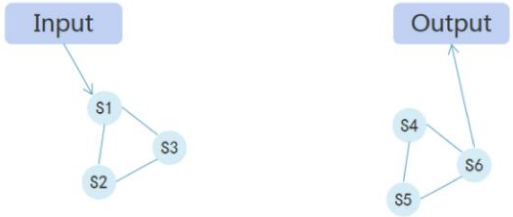


医渡云应用案例二

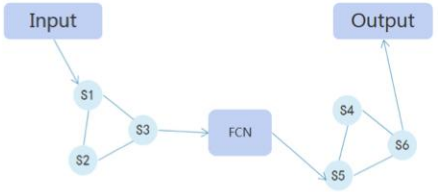
医渡云 YIDUCLOUD

- **具体实现逻辑- 知识片段构建与FCN的桥接**
  - A.和输入输出关联的知识片段集合构建
  - B.基于核心节点的片段非线性（FCN）关联
  - C.基于graph embedding（Deepwalk）的FCN优化
  - D.根据FCN被优化的频度进行剪枝，完成混合决策路径图

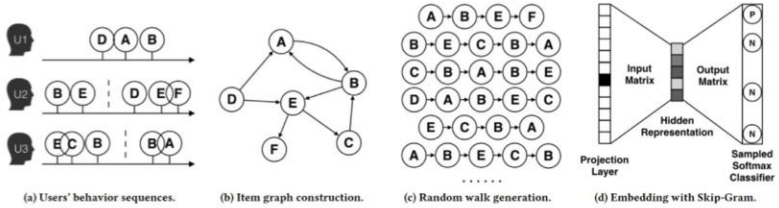
A



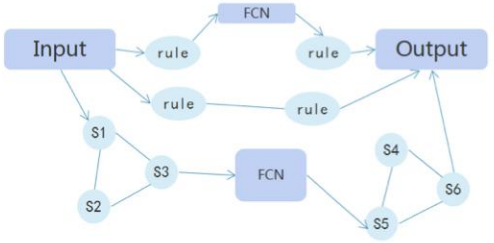
B



C



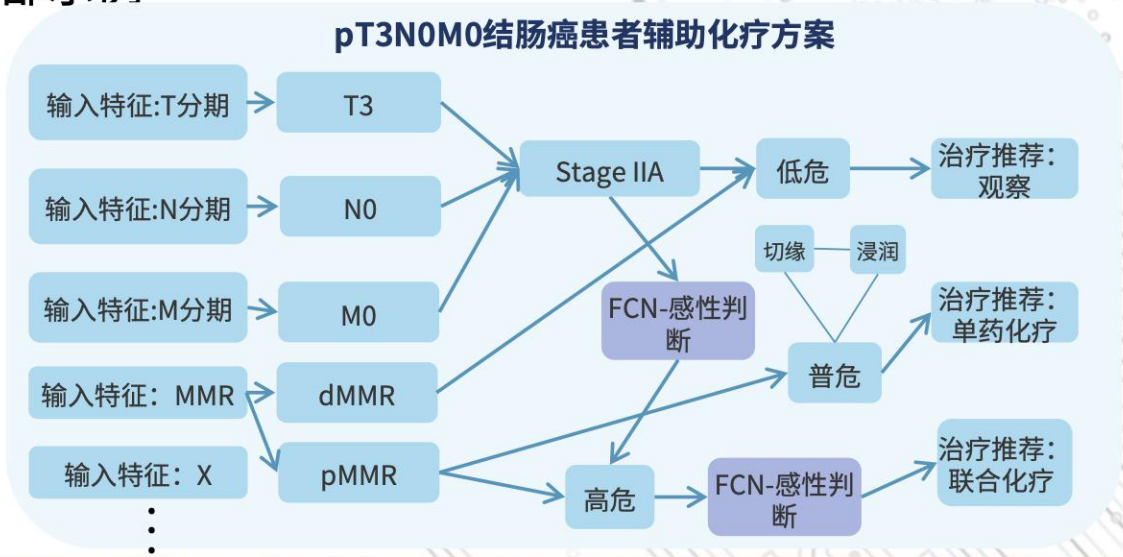
D



医渡云应用案例二

- **实例：**
  - M1：根据指南等权威医学知识构建规则库的决策
  - M2：构建图谱片段的非线性关联后进行决策
  - 示例：M2方案决策的局部示例

模型	覆盖率
M1	0.907
M2	0.962



总结与展望

模型可解释是AI在各领域发展的必经之路

数据不是模型的唯一输入，知识的嵌入同样重要

人与机器间知识表示的交互过程是可解释性学习的关键

We Are Hiring!

医渡云
YIDUCLOUD

可解释性学习交流群

医渡云XAI交流群

算法工程师 全职/实习

岗位职责
1. 研究自然语言处理、数据挖掘或统计学习领域的前沿技术,并用于医学数据问题的解决和优化
2. 和医学深度配合，将大量非结构化或半结构化数据，进行结构化信息处理
3. 通过医疗人工智能技术进行敏锐洞察和深入治理

简历请投至：
recruiting@yiducloud.cn

相关文献

1. Guangyou Cai, et al. Artificial Intelligence: Principles and Applications (Third Edition). Tsinghua University Press. 2003

2. Bo Zhang. 迈向第三代人工智能 . SCIENTIA SINICA Informationis. 2020. (1281–1302)

3. Christoph Molnar, et al. Interpretable Machine Learning: A Guide for Making Black Box Models Explainable，Lulu, 1st edition, March 24, 2019; eBook (GitHub, 2020-04-27)

4. Wang-Zhou Dai, et al. Tunneling Neural Perception and Logic Reasoning through Abductive Learning. 2018

5. Wenjing Fang, et al. Unpack Local Model Interpretation for GBDT. 2020

6. 人工智能发展史总结. (n.d.). Retrieved October 25, 2020, from https://blog.csdn.net/york1996/article/details/98845544

7. Perozzi. DeepWalk: Online Learning of Social Representations. KDD 2014

相关文献

8. Scott M. Lundberg, et al. A Unified Approach to Interpreting Model Predictions. NIPS 2017

9. Thomas Lin Pedersen, et al. LIME: Local Interpretable Model-Agnostic Explanations

10. Nicholas Tsagkarakis, et al. L1-norm Principal-Component Analysis of Complex Data. 2017 IEEE Transactions on Signal Processing PP(99)

11. First Order Logic. (n.d.). Retrieved October 25, 2020, from https://leanprover.github.io/logic_and_proof/first_order_logic.html

12. Dong Y P, Su H, Zhu J, et al. Towards interpretable deep neural networks by leveraging adversarial examples.In: Proceedings of the IJCAI workshop on AISC, Sydney, 2019. 1–61298

13. Friedman, Jerome H. "Greedy function approximation: A gradient boosting machine." Annals of statistics (2001): 1189-1232.