软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2020,31(4):1063−1078 [doi: 10.13328/j.cnki.jos.005920]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于规则推理网络的分类模型*

黄德根[1], 张云霞[1], 林红梅[2], 邹 丽[2], 刘 壮[1]

[1](大连理工大学 计算机科学与技术学院,辽宁 大连 116024)

[2](辽宁师范大学 计算机与信息技术学院,辽宁 大连 116081)

通讯作者: 邹丽, E-mail: zoulicn@163.com

摘 要: 为了缓解神经网络的“黑盒子”机制引起的算法可解释性低的问题,基于使用证据推理算法的置信规则库推理方法(以下简称RIMER)提出了一个规则推理网络模型.该模型通过RIMER中的置信规则和推理机制提高网络的可解释性.首先证明了基于证据推理的推理函数是可偏导的,保证了算法的可行性;然后,给出了规则推理网络的网络框架和学习算法,利用 RIMER 中的推理过程作为规则推理网络的前馈过程,以保证网络的可解释性;使用梯度下降法调整规则库中的参数以建立更合理的置信规则库,为了降低学习复杂度,提出了“伪梯度”的概念;最后,通过分类对比实验,分析了所提算法在精确度和可解释性上的优势.实验结果表明,当训练数据集规模较小时,规则推理网络的表现良好,当训练数据规模扩大时,规则推理网络也能达到令人满意的结果.

关键词: 规则推理;RIMER;可解释性网络;机器学习;不确定性分类

中图法分类号: TP181

中文引用格式: 黄德根,张云霞,林红梅,邹丽,刘壮.基于规则推理网络的分类模型.软件学报,2020,31(4):1063−1078. http://www.jos.org.cn/1000-9825/5920.htm

英文引用格式: Huang DG, Zhang YX, Lin HM, Zou L, Liu Z. Rule inference network model for classification. Ruan Jian Xue Bao/Journal of Software, 2020,31(4):1063−1078 (in Chinese). http://www.jos.org.cn/1000-9825/5920.htm

Rule Inference Network Model for Classification

HUANG De-Gen[1], ZHANG Yun-Xia[1], LIN Hong-Mei[2], ZOU Li[2], LIU Zhuang[1]

[1](School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China)

[2](School of Computer and Information Technology, Liaoning Normal University, Dalian 116081, China)

Abstract: The black-box working mechanism of artificial neutral network brings the confusion of interpretability. Therefore, a rule inference network is proposed based on rule-base inference methodology using the evidential reasoning approach (RIMER). It is interpretable by the rules and the inference engine in RIMER. In the present work, the partial derivatives of inference functions are proved as the theoretical fundamental of the proposed model. The framework and the learning algorithm of rule inference network for classification are presented. The feed forward of rule inference network using the inference process in RIMER contributes for the interpretability. Meanwhile, parameters in belief rule base such as attribute weights, rule weights and belief degree of consequents are trained by gradient descent as in neural network for belief rule base establishment. Moreover, the gradient is simplified by proposing the “pseudo gradient” to reduce the learning complex during the training process. The experimental results indicate the advantages of proposed rule inference network on both interpretability and learning capability. It shows that the rule inference network performs well when the scale of the training dataset is small, and when the training data scale increases, it also achieves comforting results.

Key words: rule inference network; RIMER; interpretable network; machine learning; uncertainty classification

* 基金项目: 国家自然科学基金(61772250, U1936109, 61672127)

Foundation item: National Natural Science Foundation of China (61772250, U1936109, 61672127)

本文由“非经典条件下的机器学习方法”专题特约编辑高新波教授、黎铭教授、李天瑞教授推荐.

收稿时间: 2019-03-10; 修改时间: 2019-07-11; 采用时间: 2019-09-20; jos 在线出版时间: 2020-01-10

CNKI 网络优先出版: 2020-01-14 09:53:15, http://kns.cnki.net/kcms/detail/11.2560.TP.20200114.0953.005.html

分类问题是机器学习领域最常见的任务之一.很多分类方法,如 *K*-邻近算法、随机森林、支持向量机、人工神经网络等在现实应用和学术研究中都取得了优异的成绩.尤其是随着大数据的发展,神经网络以其黑盒表示的低可解释性为代价取得了优越的性能.然而,近年来,社会对算法可解释性的关注越来越多.欧盟 2018 年 5 月 25 日提出的《通用数据保护条例》规定所有算法解释其输出原理,保证数据隐私和算法公平性(欧洲联盟.通用数据保护条例.2018.5.25);2017 年中国国务院在《新一代人工智能发展规划》中提出"实现具备高可解释性,强泛化能力的人工智能".政府对人工智能算法可解释性的关注对工业界产生了巨大的影响.为应对欧盟《通用数据保护条例》的规定,Facebook 不得不将 15 亿用户数据从爱尔兰转移到非欧盟地区的美国;Google 公司发布《The Building Blocks of Interpretability》,探讨了如何结合特征可视化和其他可解释性技术来理解神经网络如何决策.Google 公司还发布了一个建立在 DeepDream 上的神经网络可视化库 Lucid,作为研究神经网络可解释性的一套基础架构和工具.随着政府和工业界需求的增加,可解释性人工智能正逐渐成为研究热点.

神经网络一直以来被广泛质疑的"黑盒子"机制在这样的环境下受到了巨大的挑战.神经网络参数学习的不透明性,学习模型的不可解释性产生了很多的问题,例如,人们对机器学习模型的结果信心不足;对直观样本学习能力较强,但对推理性知识的学习能力不足;对数据量的要求较高,不易纠错,会产生"机器偏见"等等.文献[1]指出,模型的可解释性可以帮助人们突破深度学习的几个瓶颈,例如,从少量注释中学习、在语义级别上通过人机通信学习以及在语义上调试网络表示.

因此,关于神经网络解释性的研究日渐兴起.其中包括对神经网络结构本身进行分析解释的研究,文献[2]介绍了一种新的多层神经网络的通用解释方法:分解网络分类决策到其输入元素所做贡献,该方法适用于广泛的输入数据、学习任务和网络结构,称为深度泰勒分解,通过将解释从输出反向传播到输入层,有效地利用了网络的结构.另外,还有结合其他算法对神经网络进行改进的研究,以缓解神经网络低解释性带来的问题.文献[3]为了降低模型复杂度,实现模型的可解释性,采用分类规则提取方法对神经网络模型串联进行网格划分;层级相关度传播(layer-wise relevance propagation,简称 LRP)技术作为可解释性的一种补救方法被引进到神经网络算法中,在解释性和准确性上都取得了良好的表现,并在单次电脑信号分类领域得到了验证[4,5];文献[6]为了将表示空间与人类知识对齐,提出了一种 DNN,称为概念对齐深层神经网络(conceptual alignment deep neural network,简称 CADNN),它能够在隐含层中产生可解释的表示;文献[7]提出了一种基于本体的深度学习模型(ontology-based deep learning model,简称 ORBM$^+$),用于无向图和节点属性图的人类行为预测,不仅准确地预测了人类行为,而且还对每个预测行为产生解释.

这些方法都从一定程度上缓解了神经网络的低解释性问题,但往往存在解释性或者准确性上的不足.因此,对高性能的可解释性神经网络的探索尚未停止.鉴于不确定性推理具有较高可解释性的特点,不确定性推理和神经网络相融合的方法是一个可行的、有效的探索方向.

在不确定环境中,信息具有多种类型的不确定性知识表示形式,如数值表示、语言值表示、偏好关系表示等.传统的处理不确定知识的方法,例如,贝叶斯概率理论、证据理论和模糊集合理论等都只针对某一种特定类型的不确定性环境.置信规则库(belief rule base,简称 BRB)系统是常见的知识表示框架,可以处理多种类型的不确定性[8].对规则库系统的研究有很多[9−11],Yang 等人提出的基于证据推理方法的置信规则库推理方法(rule-base inference methodology using the evidential reasoning,简称 RIMER)是一种基于置信规则库的近似推理方法,通过置信度分布表示建立规则,处理不同类型的不确定性知识[12].在 RIMER 中,知识和规则表示是通过 BRB 实现的,其中有大量的专家给定参数,如前件属性权重、规则权重、后件置信度等,近似推理过程是基于证据推理理论实现的.文献[13]证明 RIMER 不但可以以足够高的精度拟合确定性和随机性系统,而且可以处理确定性和随机性数学模型不能表达的复杂不确定问题.因其能够处理多种类型的不确定性知识和非线性因果关系,具有很大的应用潜力.RIMER 的相关研究已在创伤分诊预判[14]、研究开发项目风险评估[15]、加热器的热效率优化系统[16]、故障诊断[17]等多个领域取得了丰硕的成果[18].

对 RIMER 的研究主要包括对基础理论的扩展研究[19]和对规则库的自动生成的研究[20,21].为了将 RIMER 推广到更一般的应用情况,文献[22]将置信度结构扩展到前件所有可能的先行词,提出了扩展的 RIMER

(extended belief rule based inference methodology based on evidential reasoning,简称 ERIMER),使得 RIMER 的应用更加灵活和广泛,且实现了简单的样本规则迭代;文献[23]将 RIMER 扩展到区间值上,提出了区间 RIMER (interval-valued belief rule inference methodology based on evidential reasoning,简称 IRIMER),研究了区间规则表示形式和推理方法,并通过对比实例说明 IRIMER 比 RIMER 和 ERIMER 更加灵活、有效.

为了突破 RIMER 方法置信规则库中专家给定参数的限制,优化置信规则库,置信规则库的自动生成研究吸引了学者们的注意[24].文献[25]建立了一个完备的调节结构,以解决 RIMER 中不可再现、不完备、实际效用值的限制、过多规则导致过完备、过少的规则导致不完备等问题;为了降低 BRB 参数学习的复杂性,同时保持 BRB 系统的逼近精度,文献[26]提出了一种基于扩展的因果强度逻辑的 BRB 参数学习方法;为了平衡建模精度与建模复杂度之间的相关性,文献[27]推导出基于 Akaike 信息准则目标来表示建模精度和建模复杂性,并建立了双层优化模型和相应的双层优化算法;文献[28]基于贝叶斯估计提出了一种不需要求单个参数的最优值的 BRB 参数在线更新方法,该方法通过考虑所有可能的参数来估计 BRB 参数的后验分布并产生预测输出.

尽管已有很多更新 RIMER 中参数的方法,RIMER 系统的自动生成和优化性能仍有待提升,RIMER 与神经网络相融合不仅能够提高神经网络的可解释性,同时利用神经网络学习 RIMER 中参数,实现置信规则库的自动生成.因此,本文基于 RIMER 提出了一种可解释的分类模型,称为规则推理网络,既保持神经网络优异的学习性能,又保留了 RIMER 的高可解释性,为分类系统提供必要的解释并便于人工干预.

本文第 1 节简单回顾 RIMER 的理论知识.第 2 节讨论规则推理网络的理论基础,证明推理函数的可导性.第 3 节提出规则推理网络的框架和学习算法.第 4 节通过分类实验来说明规则推理网络的优越性.第 5 节总结全文,并对未来研究方向进行探讨.

1 RIMER 基础理论

本节介绍 RIMER 中的置信规则库及其推理机制,是规则推理网络生成和置信规则库参数优化的重要理论基础.RIMER 方法通过将置信度嵌入到规则后件的所有可能中,将传统的 IF-THEN 规则扩展成置信规则.知识表示的参数体现在置信规则中,包括规则权重、前件属性权重和后件置信度.置信规则的推理通过证据推理方法实现[12].

1.1 不确定知识表示形式

假设给定可信度规则库 $\boldsymbol{R}=\{R_1,R_2,\ldots,R_L\}$,第 k 条规则表示为[12]

$$R_k:\text{如果 } U \text{ 是 } A^k,\text{则 } V \text{ 是 } \{(D_n,\beta_n^k)\},\text{带有规则权重 } \theta_k \text{ 和属性权重}\{\delta_1,\ldots,\delta_M\} \tag{1}$$

其中,U 表示前件属性向量$\{U_1,U_2,...,U_L\}$,V 是后件属性,A^k 是前件的指标值集合 $\{A_1^k,A_2^k,...,A_T^k\}$, 这里, $A_i^k(i=1,2,...,T)$ 是前件属性 U_T 在第 k 条规则的指标值,T 是规则中用到的前件属性的总数. $\{(D_n,\beta_n^k)\}$ 表示带有可信度值β^k 的 D,即$\{(D_1,\beta_{1k}),(D_2,\beta_{2k}),...,(D_N,\beta_{Nk})\}$,这里,$D$ 是后件向量$\{D_1,D_2,...,D_T\}$,β^k 是可信度向量$(\beta_{1k},\beta_{2k},...,\beta_{Nk})$且$k\in\{1,2,...,L\}$,$\beta_{sk}(s\in\{1,\ldots,N\})$表示输入满足前件组 A^k 时后件是 D_s 的可信度.$\theta_k(\in\Re^+, k=1,2,\ldots,L)$是第 k 条规则的权重,δ是后件的属性权重向量,$\delta_i(\in\Re^+, i=1,2,\ldots,T)$.$L$ 是规则库中所有规则组的数量.并且, $\sum_{s=1}^{N}\beta_{sk}\leqslant 1$, 若 $\sum_{s=1}^{N}\beta_{sk}=1$, 称第 k 组规则是完全的,否则称为不完全的.

在规则库中,指标集是用来描述属性的有实际意义和特定评价标准的集合,一般使用语言值描述来反映和模拟概念中的模糊性或不准确性,通过不同的前件属性参考值组合生成置信规则库,见式(1),可以等价表示为

$$S(\boldsymbol{A}^k)=\{(D_i,\beta_{ik});i=1,...,J_D\} \tag{2}$$

其中,$\boldsymbol{A}^k$ 是一组前件,即样本输入,J_D 是后件的总数.

1.2 不确定知识推理机制

置信规则库中的规则推理过程使用证据推理方法完成.证据推理方法的核心是证据合成算法,它基于置信决策矩阵、决策理论和 D-S 证据理论的证据合成规则提出的多属性聚合方法[12].规则推理过程包括激活权重的

计算和激活规则合成.

首先,由置信度分布表示的输入样本计算第 k 条规则的激活权重.

$$w_k=\frac{\theta_k\cdot\alpha_k}{\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)} \tag{3}$$

其中,θ_i 表示规则权重,L 表示规则的数量,且

$$\alpha_k=\prod_{i=1}^{T_k}(\alpha_i^k)^{\bar{\delta}_i} \tag{4}$$

其中,$\bar{\delta}_i=\frac{\delta_i}{\underset{i=1,2,\ldots,T}{\mathrm{Max}}\{\delta_i\}}$ 为属性权重标准化函数,使得 $0\leqslant\bar{\delta}_i\leqslant 1$, δ_i 表示属性权重,$\bar{\delta}_i$ 表示 δ_i 标准化后的属性权重,后文中对标准化后的属性权重训练学习,亦简称为属性权重.本文中假设不同规则中的属性权重是相同的.

α_i^k 表示第 k 条规则中第 i 个前件属性 U_i 的输入属于指标值 A_i^k 的可信度.

然后,综合置信度可以利用解析证据推理算法合成激活规则,得到

$$y_j=\frac{\mu\cdot\left[\prod_{k=1}^{L}\left(w_k\beta_{j,k}+1-w_k\sum_{i=1}^{J_D}\beta_{i,k}\right)-\prod_{k=1}^{L}\left(1-w_k\sum_{i=1}^{J_D}\beta_{i,k}\right)\right]}{1-\mu\cdot\left[\prod_{k=1}^{L}(1-w_k)\right]} \tag{5}$$

其中,

$$\mu=\left[\sum_{j=1}^{J_D}\prod_{k=1}^{L}\left(w_k\beta_{j,k}+1-w_k\sum_{j=1}^{J_D}\beta_{j,k}\right)-(J_D-1)\prod_{k=1}^{L}\left(1-w_k\sum_{j=1}^{J_D}\beta_{j,k}\right)\right]^{-1} \tag{6}$$

证据推理递归算法是本文提出的规则推理网络的基础.y_j 是指样本属于第 j 类的程度.例如,包含 3 个类的分类系统,后件的置信度(y_1,y_2,y_3)表示输入属于第 1 类、第 2 类、第 3 类的程度分别为 y_1、y_2、y_3.通常,认为样本属于第 i 类,其中,i 是 $\mathrm{Max}(y_1,y_2,y_3)$的下标.

2 规则推理网络可导性

本节对规则推理网络算法的理论基础进行了证明以说明建立规则推理网络的可行性.在规则推理网络的训练算法中,使用梯度下降对算法的参数进行调整.因此,有必要对算法的可偏导性进行证明.

定理 1. 若 $\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)\neq 0$, 则激活权重函数,如式(3),对规则权重 $\theta_i(i=1,\ldots,L)$和属性权重 $\bar{\delta}_i\,(i=1,\ldots,T)$是可偏导的.

证明:(1) 首先证明激活权重对规则权重$\theta_i(i=1,\ldots,L)$和$\alpha_j(i=1,\ldots,L)$存在偏导数.

显然,$\theta_k.\alpha_k$ 和 $\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)$ 是可导的,且 $\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)\neq 0$. 根据函数可导定理,$w_k$ 对 $\theta_i(i=1,\ldots,L)$和$\alpha_j(i=1,\ldots,L)$可导.

(2) 接下来证明激活权重对属性权重 $\bar{\delta}_i\,(i=1,\ldots,T)$可偏导.

易知, $\alpha_k=\prod_{i=1}^{T_k}(\alpha_i^k)^{\bar{\delta}_i}$ 是 $\bar{\delta}_i$ 的指数函数,因此,对 $\bar{\delta}_i\,(i=1,\ldots,T)$是可导的.又因为 w_k 对$\alpha_j(i=1,\ldots,L)$可导.根据函数可导定理,w_k 对 $\bar{\delta}_i\,(i=1,\ldots,T)$是可偏导的. □

定理 2. 当 $\prod_{j=1}^{J_D}\prod_{k=1}^{L}\beta_{j,k}\neq 0$ 且 $0<w_k<1$ 时,综合置信度函数,如式(5),对置信度$\beta_{j,k}$和激活函数 w_i 是可偏导的.

证明:首先,令

$$\mu_0=\sum_{j=1}^{J_D}\prod_{k=1}^{L}(w_k\beta_{j,k}+1-w_k)-(J_D-1)\prod_{k=1}^{L}(1-w_k),$$

则有

$$\mu=\frac{1}{\mu_0}.$$

易知,$f=(w_k\beta_{j,k}+1-w_k)$对$\beta_{j,k},w_i$偏导数存在,且$g=\prod_{k=1}^{L}(1-w_k)$对w_i偏导数存在.根据函数可导定理可知,μ_0对$\beta_{j,k},w_i$是可偏导的.并且,

$$\mu_0=\sum_{j=1}^{J_D}\prod_{k=1}^{L}(w_k\beta_{j,k}+1-w_k)-(J_D-1)\prod_{k=1}^{L}(1-w_k)>\sum_{j=1}^{J_D}\prod_{k=1}^{L}(1-w_k)-(J_D-1)\prod_{k=1}^{L}(1-w_k)$$

$$=J_D\prod_{k=1}^{L}(1-w_k)-(J_D-1)\prod_{k=1}^{L}(1-w_k)=\prod_{k=1}^{L}(1-w_k).$$

可知$\mu_0>0$,因为$0<w_k<1$.因此,$\mu=\frac{1}{\mu_0}$对$\beta_{j,k},w_i$是可偏导的.

已知f对$\beta_{j,k},w_i$存在偏导数,g对w_i存在偏导数,μ对$\beta_{j,k},w_i$存在偏导数,且$1-\mu\cdot\left[\prod_{k=1}^{L}(1-w_k)\right]>0$,因此,$y_j=\frac{\mu\cdot[f-g]}{1-\mu\cdot g}$对$\beta_{j,k},w_i$是可偏导的. □

从定理 1 和定理 2 可知,y_j对w_k是可偏导且连续的,w_k对$\overline{\delta}_i$、θ_l是可偏导的,由函数可导定理可知,y_j对$\overline{\delta}_i$、θ_l是可偏导的.

函数的可导性是使用梯度下降法建立网络的充分条件,从而对规则置信度和属性权重、规则权重进行调节,得到自动生成的置信规则库.

具体求导公式见附录 1.

3 规则推理网络

在上述理论的基础上,我们通过融合 RIMER 和 BP 神经网络的思想建立了一个规则推理网络(rule inference network,简称 RIN)模型.该模型中前馈过程是 RIMER 中的规则推理方法,置信规则库中参数训练类似 BP 神经网络,通过梯度下降完成.RIN 的推理机制使得它在可解释性方面优于传统的神经网络.置信规则库中的信息通过带有推理机制的神经网络传播.其中,自动生成置信规则库需要更新的参数包括:属性权重$\overline{\delta}_i$(i=1,…,T),规则权重θ_k(k=1,…,L)和后件置信度$\beta_{j,k}$(j=1,…,T;k=1,…,L)(如图 1 所示).接下来,我们将介绍 RIN 的网络框架和训练算法.

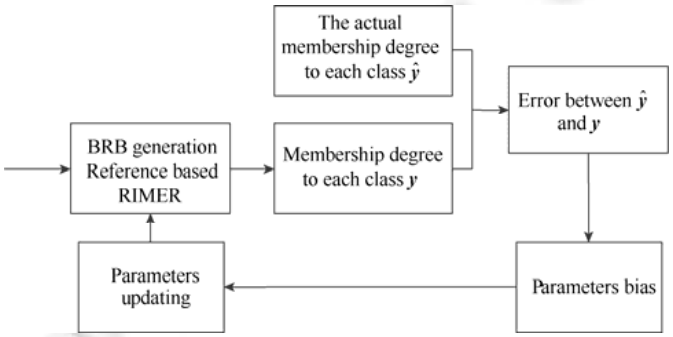

Fig.1 BRB generation process

图 1 置信规则库生成过程

3.1 RIN框架结构

RIN 由 4 层组成,包括输入层、聚合层、规则激活层和结论层(输出层)(如图 2 所示).与传统的神经网络不同,在 RIN 中,神经元之间的信息传递不需要任何激活函数,而是基于 RIMER 的置信规则库推理方法,能够形成有意义的规则,而不是传统的“黑盒子”.这种机制提高了 RIN 逻辑上的可解释性.

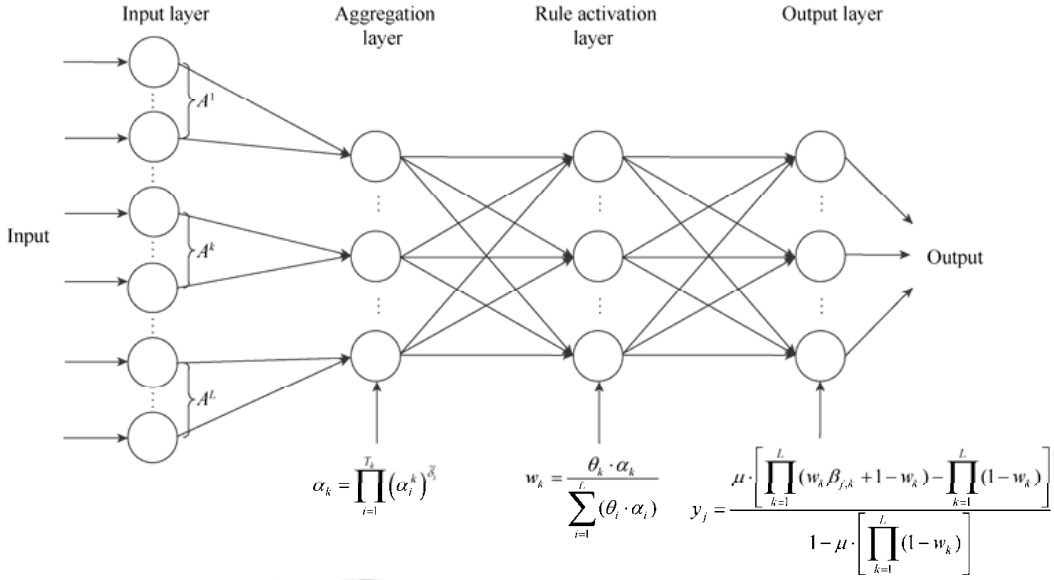

Fig.2 The framework of the inference mechanism

图 2 推理机制的网络结构

RIN 输入层的输入是样本所对应于每一条规则的参考值组,因此,如果样本是数值,则需要转换为置信度分布以便产生参考值组;在聚合层,每条规则的前件属性参考值组根据式(4)进行合并;规则激活层的输入为聚合层合并的结果,由式(3)得到规则激活权重,规则激活权重表示该条规则被激活的程度;在结论层(输出层),规则激活权重和规则的置信度由式(5)聚合可得到结论.在分类问题里,结论是一个置信度分布,表示样本属于每个类的程度,而样本所属类别为置信度最大的类.

根据推理函数,每个输入对输出的影响程度和方式都是不同的,这是由 RIMER 的推理机制引起的,有利于可解释性.例如,在聚合层,α_k 只与 α_i^k (i=1,2,…,T)有关,而与 α_i^l (i=1,2,…,T,$l\neq k$)无关;在规则激活层,α_k(k=1,…,L)和 α_i(i=1,2,…,k–1,k+1,…,L)对 w_k 的影响方式不同.图 2 用加粗的直线标记连接对输出影响更多的输入.

3.2 RIN训练算法

为了建立一个可行的 RIN,本文利用梯度下降作为 RIN 的学习算法.训练算法流程图如后文图 3 所示,主要包含 4 个部分:数据准备、置信规则库建立、规则推理和参数更新.

3.2.1 数据准备

对于一般的数值数据集,使用最大最小算法归一化到[0,1],如式(7)所示.

$$\boldsymbol{Nor}(\boldsymbol{A}) = 1 - \frac{\text{Max} - \boldsymbol{A}}{\text{Max} - \text{Min}} \tag{7}$$

其中,$\boldsymbol{Nor}(\boldsymbol{A})$是样本 $\boldsymbol{A}$ 的归一化,Max 和 Min 分别是数据集上的最大值和最小值.

在 RIMER 中,置信规则库中的规则和输入都使用置信度分布.然而,我们并不能总是得到标准的置信规则库的数据,因此需要将一般的数值数据转换成置信度分布表示.数值转换成置信度分布的方法有很多,本文采用文献[12]中的方法.

输入$\gamma_i \in x_j(i=1,\ldots,M,j=1,\ldots,SN)$转换为前件属性 U_i 指标集的置信度分布如下:

$$E(\gamma_i)=\{(A_{ij},\alpha_{ij})\big|j=1,...,J_i\},\ i=1,...,M \tag{8}$$

其中,α_{ij} 表示对应前件属性 U_i 的指标值 A_{ij} 的置信度.

$$\left.\begin{aligned}&\alpha_{ij}=\frac{A_{i(j+1)}-\gamma_i}{A_{i(j+1)}-A_{ij}},\ \alpha_{i(j+1)}=1-\alpha_{ij},\ \text{if}\ A_{ij}\leqslant\gamma_i\leqslant\alpha_{i(j+1)}\\&\alpha_{it}=0,\ t=1,...,J_i,\ \text{and}\ t\neq j,j+1\end{aligned}\right\} \tag{9}$$

本文中若无特别说明,$\boldsymbol{A}$={(*L*,0),(*M*,0.5),(*H*,1)},其中,*H* 表示“high”,*M* 表示“medium”,*L* 表示“low”,且{A_{i1}=0, A_{i2}=0.5,A_{i3}=1},*i*={1,…,*T*}.

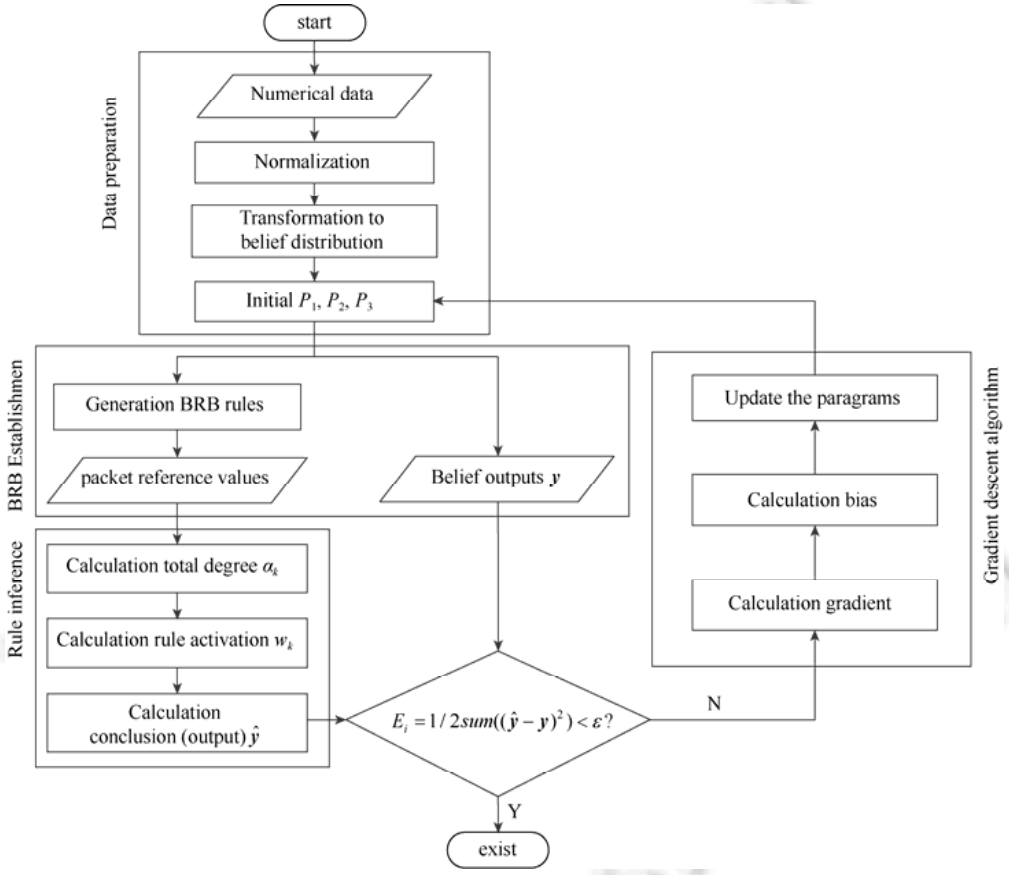

Fig.3 Flow chart of the learning algorithm for RIN

图 3 RIN 算法流程图

所有样本的属性值都可以基于转换技术表示为置信度分布.

训练样本的类别也表示为置信度分布.假设样本属于第 i^{th},$i\in\{1,2,3\}$类,则可表示为

$$\{(C_1,b_1),(C_2,b_2),...,(C_{J_D},b_{J_D})\} \tag{10}$$

其中,$b_j=\begin{cases}1,\ j=i,\\0,\ j\neq i\end{cases}$.

为了简化,在不致混淆的情况下,置信度分布表示{(A_{ij},α_{ij})|*j*=1,…,J_i}简略为{α_{ij}|*j*=1,…,J_i}.此外,每条规则后件的置信度分布β^k、规则权重θ_k以及属性权重$\bar{\delta}$都是随机初始化的,并在训练过程中更新优化.

3.2.2 建立置信规则库

由上文置信度分布表示的样本输入和规则后件可以得到如式(1)所示形式的置信规则库.

置信规则库中规则见表 1.

Table 1 Rule generation in BRB

表 1 BRB 中的生成规则

Rule	Rule weight	Antecedent	Consequent			
		Packet reference values (attributes weight δ)	C_1	C_2	…	C_{J_D}
R_1	θ_1	A^1	$\beta_{1,1}$	$\beta_{1,2}$	…	β_{1,J_D}
R_2	θ_2	A^2	$\beta_{2,1}$	$\beta_{2,2}$	…	β_{2,J_D}
⋮	⋮	⋮	⋮	⋮	…	⋮
R_L	θ_L	A^L	$\beta_{L,1}$	$\beta_{L,2}$	…	β_{L,J_D}

其中, $A^k=\left\{A_1^k,A_2^k,...,A_J^k\right\}$, $A_i^k\in\{L,M,H\}(k=1,\ldots,L;i=1,2,\ldots,J)$,$\beta_{sk}$ 分别表示第 k 条规则中后件属于类 C_s 的置信度, J 表示指标值的个数,J_D 是类的数量,L 是规则的个数.

对于一个规则推理模型,如果指标值的个数是 J,属性的数量为 T,那么置信规则库里规则的个数是

$$L=J^T \tag{11}$$

当置信度分布表示的训练样本输入 BRB 中时,根据每条规则中的前件指标值进行组合,用于不确定推理.同时,样本的类标签被保存下来作为训练 RIN 模型的实际输出.

3.2.3 推理过程

基于 RIMER 对样本输入进行推理,是一个前馈过程.样本输入是置信度分布表示,推理过程与 RIMER 一致.

首先,由式(4)得到前件的综合置信度,其中,数值输入 α_i^k 由式(9)可得;

其次,由式(3)计算每条规则的激活权重;

最后,由式(5)产生最终的结论输出.

经过推理过程可以得到一个输出,在训练过程中需要将它和实际输出的差距优化到足够小.

3.2.4 RIN 网络训练

训练一个 RIN 也就是对置信规则库中的参数进行优化,包括属性权重、规则权重和后件置信度.优化过程使用梯度下降算法最小化系统输出和实际输出的差异.

(1) 计算差异

训练样本集记作 $\{(\boldsymbol{x}_i,\boldsymbol{y}_i)\}$,其中,$\boldsymbol{x}_i$ 是输入, $\boldsymbol{y}_i=\{y_1^{(i)},...,y_J^{(i)}\}$ 是输入对应的实际输出,即样本所属类别的置信度分布.置信规则库推理后的系统输出记作 $\hat{\boldsymbol{y}}_i=\{\hat{y}_1^{(i)},...,\hat{y}_J^{(i)}\}$. $\boldsymbol{y}_i$ 和 $\hat{\boldsymbol{y}}_i$ 之间的差值使用均方差函数计算.

$$E_i=\frac{1}{2}\sum_{j=1}^{J}\left(\hat{y}_j^{(i)}-y_j^{(i)}\right)^2 \tag{12}$$

(2) 参数更新

由式(3)~式(6)可知规则的激活权重是关于所有属性权重的函数,并且输出函数是关于所有激活权重的复杂函数,这使得 $\frac{\partial E_i}{\partial\bar{\delta}_t}$、$\frac{\partial E_i}{\partial\theta_l}$ 的计算复杂度非常高,要付出很大的代价.为了简化算法,考虑将推理层的激活权重偏差作为激活层输出的差值,即分别对规则推理网络中的规则激活层和输入层进行梯度下降更新.由定理 1 和定理 2 可知, $\frac{\partial E_i}{\partial w_k}$、$\frac{\partial w_k}{\partial\bar{\delta}_t}$、$\frac{\partial w_k}{\partial\theta_k}$ 存在,用偏导 $\frac{\partial\Delta w_k}{\partial\bar{\delta}_t}$、$\frac{\partial\Delta w_k}{\partial\theta_k}$ 替代 $\frac{\partial E_i}{\partial\bar{\delta}_t}$、$\frac{\partial E_i}{\partial\theta_k}$ 作为梯度,称为 $\bar{\delta}_t$ 和 θ_k 的“伪梯度”.

使用“伪梯度”代替复杂的“真梯度”降低了复杂度,参数更新算法的主要步骤如下.

步骤 1.参数偏差计算.

计算推理层梯度如下:

$$\frac{\partial E_i}{\partial \hat{y}_j^{(i)}} = \hat{y}_j^{(i)} - y_j^{(i)}.$$

$$\frac{\partial E_i}{\partial \beta_{j,k}} = \sum_{l=1}^{J} \frac{\partial E_i}{\partial \hat{y}_l^{(i)}} \cdot \frac{\partial \hat{y}_l^{(i)}}{\partial \beta_{j,k}}, \frac{\partial E_i}{\partial w_k} = \sum_{j=1}^{J} \frac{\partial E_i}{\partial \hat{y}_j^{(i)}} \cdot \frac{\partial \hat{y}_j^{(i)}}{\partial w_k}.$$

推理层参数偏差如下:

$$\Delta \beta_{j,k} = -\eta_1 \frac{\partial E_i}{\partial \beta_{j,k}} \tag{13}$$

$$\Delta w_k = -\eta_1 \frac{\partial E_i}{\partial w_k} \tag{14}$$

其中,η_1是学习率.

特别需要注意的是,计算$\bar{\delta}_t$和θ_k的偏差值时,使用$\bar{\delta}_t$和θ_k的"伪梯度",即由$\frac{\partial \Delta w_k}{\partial \bar{\delta}_t}$、$\frac{\partial \Delta w_k}{\partial \theta_k}$代替$\frac{\partial E_i}{\partial \bar{\delta}_t}$、$\frac{\partial E_i}{\partial \theta_k}$来降低算法复杂度.$\bar{\delta}_t$和$\theta_k$的偏差值如下:

$$\Delta \theta_l = \eta_2 \frac{\partial \Delta w_k}{\partial \theta_l} = -\eta_1 \eta_2 \frac{\partial E_i}{\partial w_k} \cdot \frac{\partial w_k}{\partial \theta_l} \tag{15}$$

$$\Delta \bar{\delta}_t = \eta_2 \frac{\partial \Delta w_k}{\partial \bar{\delta}_t} = -\eta_1 \eta_2 \frac{\partial E_i}{\partial w_k} \cdot \frac{\partial w_k}{\partial \bar{\delta}_t} \tag{16}$$

其中,η_1和η_2是学习率.

偏导数$\frac{\partial \hat{y}_l^{(i)}}{\partial \beta_{j,k}}$、$\frac{\partial \hat{y}_l^{(i)}}{\partial w_k}$、$\frac{\partial \Delta w_k}{\partial \bar{\delta}_t}$、$\frac{\partial \Delta w_k}{\partial \theta_k}$可根据附录 1 求得.

步骤 2.更新参数.

基于梯度下降算法调整参数,对于任意参数 v,参数更新估计函数如下:

$$v \leftarrow v + \Delta v \tag{17}$$

由此,属性权重$\bar{\delta}_t$、规则权重θ_k和后件置信度$\beta_{j,k}$更新如下:

$$\bar{\delta}_t = \bar{\delta}_t + \Delta \bar{\delta}_t,\ t = 1,...,T \tag{18}$$

$$\theta_l = \theta_l + \Delta \theta_l,\ l = 1,...,L \tag{19}$$

$$\beta_{j,k} = \beta_{j,k} + \Delta \beta_{j,k},\ j = 1,...,J; k = 1,...,L \tag{20}$$

步骤 3.重复步骤 1~步骤 3 更新参数,直到$\boldsymbol{y}_i$和$\hat{\boldsymbol{y}}_i$的差值满足设定条件,即$\sum_{i=1}^{J} E_i$小于足够小的ε.

3.3 算法描述

下面给出 3 种算法,即 BRB-BP1、BRB-BP2 和 RIN 的算法描述.其中,BRB-BP1 和 BRB-BP2 融合了置信规则库和 BP 神经网络,即通过输入置信度分布的样本使用 BP 神经网络更新参数,不涉及 RIMER 中的证据推理过程.BRB-BP1 横向拼接输入样本的置信度分布,BRB-BP2 纵向拼接输入样本的置信度分布.它们用来对比说明推理机制在所提 RIN 算法中所起的重要作用.

算法 1. BRB-BP1.

输入:训练样本集$\{(A_i,y_i)\}$,A_i是输入样本的属性集,y_i是输入对应的样本类别;

输出:BP 神经网络模型 BRB-BP1.

置信度分布表示 1 (step 1~step 8).

1. 归一化,根据最大最小法,由式(7)归一化样本集到[0,1]区间
2. **While** $A_i \in \{(A_i,y_i)\}$, **Do**
3. **For** j=1 to T, **Do**
4. 由式(8)将样本$\alpha_{ij} \in A_i$表示为置信度分布$(\alpha_1^{(ij)},...,\alpha_J^{(ij)})$;

5. **End For**
6. 由式(9)将样本类别表示为置信度分布 $\boldsymbol{y}_i=(y_1^{(i)},...,y_J^{(i)})$;
7. **End While**
8. 置信度分布的样本矩阵横向拼接,即 $\boldsymbol{x}_i=\{x_{ik}\}=(\alpha_1^{(i1)},...,\alpha_J^{(j1)},...,\alpha_J^{(ij)},...)$;
9. 将$\{(\boldsymbol{x}_i,\boldsymbol{y}_i)\}$作为训练样本,利用 BP 神经网络训练得到 BP 神经网络模型 BRB-BP1;

算法 2. BRB-BP2.

输入:训练样本集$\{(A_i,y_i)\}$,A_i是输入样本的属性集,y_i是输入对应的样本类别;

输出:BP 神经网络模型 BRB-BP2.

置信度分布表示 2 (step 1~step 9).

1. 归一化,根据最大最小法,由式(7)归一化样本集到[0,1]区间
2. **While** $A_i\in\{(A_i,y_i)\}$, **Do**
3. **For** j=1 to T, **Do**
4. 由式(8)将样本$\alpha_{ij}\in A_i$表示为置信度分布$(\alpha_1^{(ij)},...,\alpha)$;
5. **End For**
6. 由式(9)将样本类别表示为置信度分布 $\boldsymbol{y}_i=(y_1^{(i)},...,y_J^{(i)})$;
7. **End While**
8. 置信度分布的样本矩阵纵向拼接,即 $\boldsymbol{x}_i=\{x_{ik}\}=\begin{pmatrix}\alpha_1^{(i1)} & \cdots & \alpha_J^{(i1)}\\ & \cdots & \\ \alpha_1^{(iT)} & \cdots & \alpha_J^{(iT)}\end{pmatrix}$;
9. 样本类别置信度分布转置 $\boldsymbol{y}_i=\boldsymbol{y}_i^T$;
10. 将$\{(\boldsymbol{x}_i,\boldsymbol{y}_i)\}$作为训练样本,训练得到 BP 神经网络模型 BRB-BP2;

算法 3. RIN.

输入:训练样本集$\{(A_i,y_i)\}$,A_i是输入样本的属性,y_i是输入对应的样本类别.给定期望误差$\varepsilon_{\min}$,最大训练次数$\lambda_{\max}$;

输出:置信规则库,包括属性权重$\overline{\delta}_i$(i=1,…,T)、规则权重θ_k(k=1,…,L)和后件置信度$\beta_{j,k}$(j=1,…,J_D;k=1,…,L).

1. 执行置信度分布表示 2;
2. 随机初始化 $P_1=\{\overline{\delta}_i|i=1,...,T\}$, $P_2=\{\theta_k|k=1,\ldots,L\}$,$P_3=\{\beta_{j,k}|j=1,\ldots,J_D;k=1,\ldots,L\}$;
3. **While** $\varepsilon>\varepsilon_{\min}\&\&\lambda>\lambda_{\min}$, **Do**
4. **For** $\boldsymbol{x}_i$, **Do**
5. 由式(2)生成置信规则库,$S(\boldsymbol{A}^k,\boldsymbol{\delta},\theta_k)=\{(D_i,\beta_{j,k});i=1,\ldots,J_D\}$; // $A^k=\{\alpha_1^k,...,\alpha_J^k\}$ (k=1,…,L)是 $\boldsymbol{x}_i$ 置信度分布参考值的组合作为前件,$\{\beta_{j,k}|j=1,\ldots,J_D\}$($k$=1,…,$L$)作为后件
6. 由式(4)计算$\boldsymbol{\alpha}=\{\alpha_k\}$;
7. 由式(3)计算 $\boldsymbol{w}=\{w_k\}$;
8. 由式(5)和式(6)计算 $\boldsymbol{y}_i=\{y_1^{(i)},...,y_J^{(i)}\}$;
9. 由式(12)计算均方差 e_i;
10. 由式(13)、式(15)、式(16)计算偏差值;
11. 由式(18)~式(20)更新参数;
12. **End For**
13. $\varepsilon=sum(\mathrm{e}_i)/M$; //$M$ 是样本个数
14. $\lambda=\lambda+1$;
15. **End While**
16. 输出置信规则库;

4 实验及分析

本节通过与传统的 BP 神经网络、自适应神经模糊系统(adaptive network-based fuzzy inference system,简称 ANFIS)[29]及随机配置网络(stochastic configuration networks,简称 SCN)[30]比较说明所提方法在精确度和可解释性上的优势.

4.1 实验结果

本文使用的数据集全部来自著名的 UC Irvine Machine Learning Repository 数据库,选取其中部分具有代表性的分类数据集.其中,Data1 是 Iris 数据集;Data2 来自 Wine 数据集,选取其中 6 个属性:Malic acid、Alcalinity of ash、Total phenols、Nonflavanoid phenols、Color intensity、OD280/OD315 of diluted wines;Data3 来自 Wine 数据集,选取其中 6 个属性:Ash、Magnesium、Flavanoids、Proanthocyanins、Hue、Proline;Data4 来自 Seeds 数据集,选取其中 5 个属性:compactness $C=4\times pi\times A/P^2$、length of kernel、width of kernel、asymmetry coefficient、length of kernel groove;Data5 来自 Banknote Authentication 数据集,包含 4 个属性,1 372 样本;Data6 来自 Wireless Indoor Localization 数据集,包含 7 个属性,1 500 个样本.其中,Data2、Data3、Data4 只包含部分属性值,是不完整的数据.

在 k 折交叉验证法中,初始样本被随机分成 k 个几乎规模相等的子样本,一个单独的子样本被保留作为验证模型的数据,其他 k–1 个样本用来训练.交叉验证重复 k 次,每个子样本验证 1 次(https://en.wikipedia.org/wiki/Cross-validation_(statistics)#k-fold_cross-validation).为了说明 RIN 在训练数据远远少于测试数据的情况下的效果,一个单独的子样本被作为训练集,其他 k–1 个样本作为测试集,与传统的 k 折交叉验证法相反,称为 k 折逆交叉验证.本文采用 5 折逆交叉验证法来测试分类的精确度,以 BP 神经网络作为基线.BP、ANFIS、SCN、RBR-BP1、RBR-BP2 与 RIN 分类实验结果如表 2 和图 4 所示.

Table 2 Classification precision of BP, ANFIS, SCN, RBR-BP1, RBR-BP2 and RIN

表 2 BP、ANFIS、SCN、RBR-BP1、RBR-BP2 与 RIN 分类准确率比较

	Data1 (%)	Data2 (%)	Data3 (%)	Data4 (%)	Data5 (%)	Data6 (%)
BP	87.00	85.67	82.33	83.93	89.3	83.20
ANFIS	89.00	56.17	58.00	74.05	**94.8**	74.55
SCN	78.50	81.67	66.17	**98.33**	**99.73**	**97.73**
RBR-BP1	87.50	80.17	79.00	83.45	88.26	80.81
RBR-BP2	89.17	59.67	42.00	68.45	69.72	66.78
RIN	**91.83**	**89.33**	**85.50**	**89.17**	94.53	**90.63**

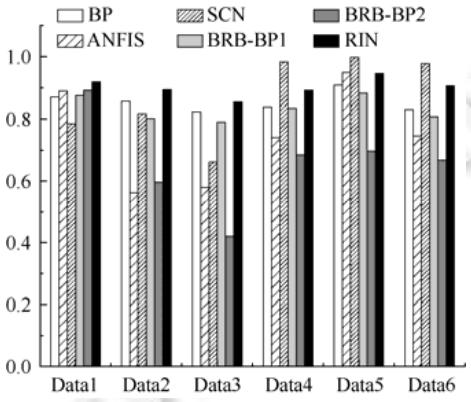

Fig.4 Histogram of classification precision

图 4 分类准确率对比图

从表 2 和图 4 可以看出,RIN 的性能总是优于基线 BP 神经网络,准确率高出 3.17%~7.43%.在 Data1、Data2、Data3 上 RIN 的表现优于其他所有方法,准确率分别为 91.83%、89.33%、85.50%.在 Data4、Data5、Data6 上 SCN 的表现最好,但 RIN 在 Data4、Data6 上仅次于 SCN,准确率分别为 89.17%、90.63%.在所有数据集上 RIN

的表现优于 BP 和 RBR-BP1、RBR-BP2,除了 Data5 之外,RIN 的表现在其他数据集上优于 ANFIS.其中,本节所提方法 RIN 和 ANFIS 具有一定的可解释性,但 ANFIS 在大多数数据集上的性能都差强人意,表现不够稳定.

为了研究数据规模对分类结果的影响,图 5 展示了不同训练数据规模下的分类准确率变化.横坐标表示数据的规模,即训练所用数据(即训练数据和验证数据)占总数据的比例.例如,*scales*=0.4 表示随机选取 40%的数据作为 5 折逆交叉训练的数据,余下数据作为测试集.纵坐标表示分类准确率.

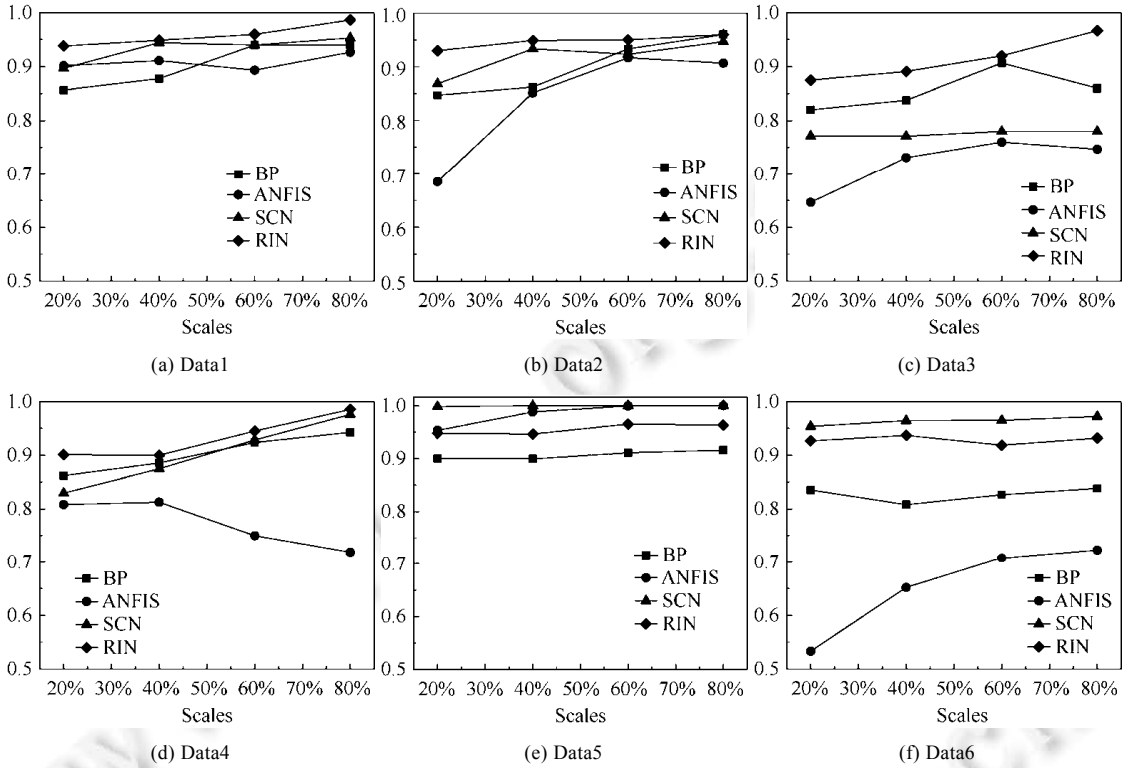

Fig.5 Classification precision trend chart in different scales

图 5 不同数据规模下分类准确率变化趋势

由图 5 可以看到,随着训练数据集规模的扩大,曲线都呈整体上升的趋势,说明所有分类方法的准确性都呈整体提高的趋势,且结果逐渐趋于一致.当数据规模较小时,RIN 曲线保持在较高位置(如图 5(a)~图 5(d)所示),说明方法优势更加明显.ANFIS 作为另一可解释性神经网络,曲线位置较低,且波动大,说明该方法表现性能普遍不佳且性能不够稳定.RIN 和 SCN 曲线波动较小,说明它们的性能更加稳定,但是 SCN 方法可解释性较低.

4.2 结果分析

本文所提方法结合了 RIMER 和神经网络的思想,构建了一个可解释的网络——RIN,利用神经网络的学习机制优化置信规则库中的参数,自动生成置信规则库,反过来利用置信规则库中的规则缓解神经网络可解释性低的问题.通过训练学习到 RIN 系统的同时,建立了一个置信规则库,规则库中的知识通过梯度下降训练获得,推理过程由 RIMER 中的不确定推理完成,由此来实现 RIN 的解释性.专家或其他人可根据专业知识或者经验,通过调节规则库中的知识规则来实现对分类系统的修正.例如,在 iris 5 折逆交叉验证实验室中,训练生成 RIMER 的置信规则库中部分规则在表 3 中列出.其中每行表示一条规则,例如第 1 行表示规则:

R_1:若$\{U_1,U_2,U_3,U_4\}$是$\{L,L,L,L\}$,则 Iris 分类为$\{(C_1,0.3529),(C_2,0.5152),(C_3,0.1397)\}$,带有规则权重 0.125 4,属性权重分别为{0.2863,0.4176,0.7994,0.6349}.

其中,U_1、U_2、U_3、U_4分别表示鸢尾花的花萼长度、花萼宽度、花瓣长度、花瓣宽度.L、M、H则可表示

语气词“短”“一般”“长”.因此,该规则表示:

如果花萼长度短且花萼宽度短且花瓣长度短且花瓣宽度短,则该鸢尾花是山鸢尾的置信度是 0.352 9,该鸢尾花是杂色鸢尾的置信度是 0.515 2,该鸢尾花是维吉尼亚鸢尾的置信度是 0.139 7,此条规则的权重为 0.125 4,以上属性的权重分别是 0.286 3、0.417 6、0.799 4、0.634 9.

Table 3 Generation BRB of Iris date

表 3 Iris 数据生成置信规则库

规则	规则权重	前件				后件		
		A_1(0.2863)	A_2(0.4176)	A_3(0.7994)	A_4(0.6349)	C_1	C_2	C_3
R_1	0.125 4	*L*	*L*	*L*	*L*	0.352 9	0.515 2	0.139 7
R_2	0.947 9	*L*	*L*	*L*	*M*	0.637 2	0.052 3	0.319 1
R_3	0.903 2	*L*	*L*	*L*	*H*	0.534 1	0.328 9	0.147 7
R_4	0.997 0	*L*	*L*	*M*	*L*	0.414 8	0.091 5	0.493 7
R_5	0.152 2	*L*	*L*	*M*	*M*	0.034 7	0.815 4	0.163 6
R_6	0.600 1	*L*	*L*	*M*	*H*	0.412 2	0.522 5	0.075 9
R_7	0.079 3	*L*	*L*	*H*	*L*	0.199 7	0.543 1	0.259 1
R_8	0.597 3	*L*	*L*	*H*	*M*	0.282 9	0.493 2	0.493 2
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
R_{75}	0.483 4	*H*	*H*	*L*	*H*	0.325 9	0.488 2	0.185 9
R_{76}	0.447 5	*H*	*H*	*M*	*L*	0.111 4	0.398 1	0.490 5
R_{77}	0.810 7	*H*	*H*	*M*	*M*	0.404 4	0.244 1	0.352 9
R_{78}	0.063 4	*H*	*H*	*M*	*H*	0.506 1	0.352 3	0.142 4
R_{79}	0.616 8	*H*	*H*	*H*	*L*	0.375 7	0.451 2	0.173 1
R_{80}	0.732 0	*H*	*H*	*H*	*M*	0.245 3	0.146 9	0.609 4
R_{81}	0.822 2	*H*	*H*	*H*	*H*	0.185 5	0.312 3	0.555 1

本文所提方法产生了有可解释性的规则库,专家可根据已有经验修改置信规则库中的规则,方便人工干预.ANFIS 也具有一定的可解释性,可以通过调节系统曲线来实现系统的修正,SCN 和 BP 神经网络解释性较弱.但是,ANFIS 除了在 Data5 中表现良好以外,在其他数据集上表现都很差,对数据要求比较高,不够稳定.从实验结果可以看出,RIN 在大多数数据集上表现很好,尤其是在训练数据规模较小时能够取得优异的性能.这可能得益于置信规则库的置信度分布表示和基于证据理论的近似推理方法,提高了网络的学习能力.然而,BRB-BP1 和 BRB-BP2 的表现并不理想,可见单纯地使用置信度分布表示的训练数据并不会使网络性能提高,甚至会降低网络的性能,因此,基于 RIMER 的不确定推理机制在学习过程中起着至关重要的作用.

RIN 不仅在小数据上表现良好,从实验结果可以看到,当数据量增大时,RIN 依然能够取得令人满意的结果.这可能是因为在置信规则库中的有意义的规则和可解释的推理机制,使得 RIN 对数据量的依赖没有那么强.当数据量足够大时,SCN 的表现最好,SCN 在节点设置上的改进在后续更加完善 RIN 的研究上具有一定的指导意义.

综上所述,RIN 在数据量较小时表现明显优于其他方法,随着数据量的增大,RIN 依旧能够取得令人满意的结果.SCN 在数据量较大时表现优异,但其可解释性较低,RIN 的性能仍有提升空间,在将来的研究中可以借鉴 SCN 的改进方法.

5 结 论

在本文中,基于 RIMER 提出了一个规则推理网络(RIN),利用机器学习思想建立自动生成置信规则库系统,规则库中的规则和知识都采用置信度分布表示.该系统弥补了神经网络“黑盒子”机制导致的可解释性低的问题.实验结果表明,当训练数据规模很小时,RIN 性能最佳,这得益于 RIMER 中的置信度分布表示形式和使用证据理论的推理机制带来的可解释性,尤其是 RIMER 中不确定推理在可解释性和系统性能上都至关重要.另外,当规模扩大时,RIN 依旧能够取得令人满意的分类准确性.

然而,知识表示形式的复杂性也增加了训练系统的复杂度,随着属性和参考值的增加,规则数的指数增长引起的维数爆炸问题是基于规则的机器学习的一个重要问题,本文利用简化的“伪梯度”代替梯度的方法降低系

统的复杂度,但依旧不能完全避免此类问题,因此算法对高维数据不友好.在未来的工作中,我们将探究降低系统复杂度的方法,例如,利用规则约简算法、属性约简算法、学习率优化、归结原理等等.作为一个新的可解释网络算法,RIN 还并非十分完善,我们也将从多方面对其进行改善,提高它的性能,充分发挥其可解释性的能力.

References:

[1] Zhang QS, Zhu SC. Visual interpretability for deep learning: A survey. Frontiers of Information Technology & Electronic Engineering, 2018,19(1):27–39.

[2] Grégoire M, Sebastian L, Alexander B, *et al*. Explaining nonlinear classification decisions with deep Taylor decomposition. Pattern Recognition, 2017,65:211–222.

[3] Petru M, Young JL, Charles C, *et al*. Accurate and interpretable classification of microspectroscopy pixels using artificial neural networks. Medical Image Analysis, 2017,37:37–45.

[4] Irene S, Sebastian L, Wojciech S, *et al*. Interpretable deep neural networks for single-trial EEG classification. Journal of Neuroscience Methods, 2016,274:141–145.

[5] Grégoire M, Wojciech S, Klaus-Robert M. Methods for interpreting and understanding deep neural networks. Digital Signal Processing, 2018,73:1–15.

[6] Dai YL, Wang GJ, Li KC. Conceptual alignment deep neural networks. Journal of Intelligent & Fuzzy Systems, 2018,34:1631–1642.

[7] Nhathai P, Dou DJ, Wang H, *et al*. Ontology-based deep learning for human behavior prediction with explanations in health social networks. Information Sciences, 2017,384:298–313.

[8] Sun R. Robust reasoning: Integrating rule-based and similarity-based reasoning. Articial Intelligence, 1995,75(2):241–295.

[9] Yang ZL, Wang J. Use of fuzzy risk assessment in FMEA of offshore engineering systems. Ocean Engineering, 2015,95:195–204.

[10] Zhou ZJ, Hua CH, Yang JB, *et al*. Online updating belief rule based system for pipeline leak detection under expert intervention. Expert Systems with Applications, 2009,36(4):7700–7709.

[11] Chen SW, Liu J, Wang H, *et al*. A group decision making model for partially ordered preference under uncertainty. Information Fusion, 2015,25:32–41.

[12] Yang JB, Liu J, Wang J, *et al*. Belief rule-base inference methodology using the evidential reasoning approach-RIMER. IEEE Trans. on Systems Man and Cybernetics Part A-Systems and Humans, 2006,36(2):266–285.

[13] Guo M. A belief-rule-based inference method for modeling systems under uncertainties. Systems Engineering —Theory & Practice, 2016,36(8):1975–1982 (in Chinese with English abstract).

[14] Kong GL, Xu DL, Yang JB. Belief rule-based inference for predicting trauma outcome. Knowledge-based Systems, 2016,95:35–44.

[15] Yang Y, Wang J, Wang G, *et al*. Research and development project risk assessment using a belief rule-based system with random subspaces. Knowledge-based Systems, 2019,178:51–60.

[16] Gao X, Lyu W, Qi L, *et al*. RIMER and SA based thermal efficiency optimization for fired heaters. FUEL, 2017,205:272–285.

[17] Cheng C, Wang JH, Teng WX, *et al*. Health status prediction based on belief rule base for high-speed train running gear system. IEEE Access, 2019,7:4145–4159.

[18] Wei H, Hu GY, Zhou, ZJ, *et al*. A new BRB model for security-state assessment of cloud computing based on the impact of external and internal environments. Computers & Security, 2018,73:207–218.

[19] Jin LQ, Liu J, Xu Y, *et al*. A novel rule base representation and its inference method using the evidential reasoning approach. Konwledge-based Systems, 2015,87:80–91.

[20] Yang JB, Liu J, Xu DL, *et al*. Optimization models for training belief-rule-based systems. IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2007,37(4):569–585.

[21] Clazada A, Liu J, Wang H, *et al*. A new dynamic rule activation method for extended belief rule-based systems. IEEE Trans. on Knowledge & Data Engineering 2015,27(4):880–894.

[22] Liu J, Luis M, Calzada A. A novel belief rule base representation, generation and its inference methodology. Knowledge-based Systems, 2013,53:129–141.

[23] Zhu H, Zhao J, Xu Y. Interval-valued belief rule inference methodology based on evidential reasoning-IRIMER. Int'l Journal of Information Technology & Decision Making, 2016,15(6):1345–1366.

[24] Chang LL, Zhou Y, Jiang J. Structure learning for belief rule base expert system—a comparative study. Knowledge-based Systems, 2013,39:159–172.

[25] Wang YM, Yang LH, Fu YG. Dynamic rule adjustment approach for optimizing belief rule-base expert system. Knowledge-based Systems, 2016,96:40–60.

[26] Sun JB, Huang J X, Chang LL, *et al*. BRBcast: A new approach to belief rule-based system parameter learning via extended causal strength logic. Information Sciences, 2018,444:51–71.

[27] Chang LL, Zhou ZJ, Chen YW, *et al*. Akaike information criterion-based conjunctive belief rule base learning for complex system modeling. Knowledge-based Systems, 2018,161:47–64.

[28] Tang XL, Xiao MQ, Liang YJ, *et al*. Online updating belief-rule-base using Bayesian estimation. Knowledge-based Systems, 2019,171:93–105.

[29] Jang J. Anfis-adaptive-network-based fuzzy inference system. IEEE Trans. on Systems Man and Cybernetics, 1993,23(3):665–685.

[30] Wang DH, Li M. Stochastic configuration networks: fundamentals and algorithms. IEEE Trans. on Cybernetics, 2017,47(10): 3466–3479.

附中文参考文献:

[13] 郭敏.基于置信规则库推理的不确定性建模研究.系统工程理论与实践,2016,36(8):1975–1982.

附录 1. $\frac{\partial w_k}{\partial \theta_i}$、$\frac{\partial w_k}{\partial \delta_i}$、$\frac{\partial y_j}{\partial \beta_{j,k}}$、$\frac{\partial y_j}{\partial w_k}$ 求导过程

(1) $\frac{\partial w_k}{\partial \theta_i}$ 求导过程

当 $i=k$ 时,

$$\frac{\partial w_k}{\partial \theta_k}=\frac{(\theta_k\cdot\alpha_k)'\cdot\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)-\left(\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)\right)'\cdot(\theta_k\cdot\alpha_k)}{\left(\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)\right)^2}=\frac{\alpha_k\cdot\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)-\alpha_k\cdot(\theta_k\cdot\alpha_k)}{\left(\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)\right)^2}=\frac{\alpha_k\cdot\sum\limits_{i=1,i\neq k}^{L}(\theta_i\cdot\alpha_i)}{\left(\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)\right)^2}.$$

当 $j\neq k$ 时,

$$\frac{\partial w_k}{\partial \theta_i}=-\theta_k\cdot\alpha_k\frac{\left(\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)\right)'}{\left(\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)\right)^2}=-\theta_k\cdot\alpha_k\frac{\alpha_i}{\left(\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)\right)^2}.$$

(2) $\frac{\partial w_k}{\partial \delta_i}$ 求导过程

$$\frac{\partial w_k}{\partial \delta_i}=\frac{\partial w_k}{\partial \alpha_j}\cdot\frac{\partial \alpha_j}{\partial \delta_i}.\ \frac{\partial \alpha_j}{\partial \delta_i}=((\alpha_i^j)^{\delta_i})'\prod_{l=1,l\neq i}^{T}(\alpha_l^k)^{\delta_l}=\ln\alpha_i^j\prod_{l=1}^{T}(\alpha_l^k)^{\delta_l}.$$

当 $j=k$ 时,

$$\frac{\partial w_k}{\partial \alpha_k}=\frac{\theta_k\cdot\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)-\theta_k\cdot(\theta_k\cdot\alpha_k)}{\left(\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)\right)^2}=\frac{\theta_k\cdot\sum\limits_{i=1,i\neq k}^{L}(\theta_i\cdot\alpha_i)}{\left(\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)\right)^2}.$$

当 $j\neq k$ 时,

$$\frac{\partial w_k}{\partial \alpha_j}=-\theta_k\cdot\alpha_k\frac{\left(\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)\right)'}{\left(\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)\right)^2}=-\theta_k\cdot\alpha_k\frac{\theta_j}{\left(\sum_{i=1}^{L}(\theta_i\cdot\alpha_i)\right)^2}.$$

(3) $\frac{\partial y_j}{\partial \beta_{j,k}}$ 求导过程

令 $f_j=\prod_{k=1}^{L}(w_k\beta_{j,k}+1-w_k)-\prod_{k=1}^{L}(1-w_k)$,

$$\mu_0=\sum_{j=1}^{J_D}\prod_{k=1}^{L}(w_k\beta_{j,k}+1-w_k)-(J_D-1)\prod_{k=1}^{L}(1-w_k),\ \frac{\partial y_j}{\partial \beta_{j,k}}=\frac{(\mu_\beta' f_j+\mu f_{j\beta}')\left(1-\mu\cdot\left[\prod_{k=1}^{L}(1-w_k)\right]\right)+\prod_{k=1}^{L}(1-w_k)\cdot\mu_\beta'\mu f_j}{\left\{1-\mu\cdot\left[\prod_{k=1}^{L}(1-w_k)\right]\right\}^2},$$

其中, $\mu_\beta'=\frac{\partial\mu}{\partial\beta_{j,k}}, f_{j\beta}'=\frac{\partial f_j}{\partial\beta_{j,k}}$.

$$\frac{\partial\mu}{\partial\beta_{j,k}}=\frac{-\mu_0'}{\mu_0^2}=\frac{-w_k\prod_{l=1,l\neq k}^{L}(w_k\beta_{j,k}+1-w_k)}{\mu_0^2},\frac{\partial f_j}{\partial\beta_{j,k}}=w_k\prod_{l=1,l\neq k}^{L}(w_k\beta_{j,k}+1-w_k).$$

(4) $\frac{\partial y_j}{\partial w_k}$ 求导过程

令 $g=1-\mu\cdot\left[\prod_{k=1}^{L}(1-w_k)\right]$.

$$\frac{\partial y_j}{\partial w_k}=\frac{(\mu_w' f_j+\mu f_{jw}')\cdot g-g'\mu f_j}{g^2},$$

其中, $\mu_w'=\frac{\partial\mu}{\partial w}, f_{jw}'=\frac{\partial f_j}{\partial w}, g'=\frac{\partial g}{\partial w}$.

$$\frac{\partial\mu}{\partial w}=\frac{-\mu_0'}{\mu_0^2}=\frac{\sum_{j=1}^{J_D}(\beta_{j,k}-1)\cdot\prod_{l=1,l\neq k}^{L}(w_l\beta_{j,k}+1-w_l)+(J_D-1)\prod_{l=1,l\neq k}^{L}(1-w_l)}{\mu_0^2},$$

$$\frac{\partial f_j}{\partial w}=(\beta_{j,k}-1)\cdot\prod_{l=1,l\neq k}^{L}(w_l\beta_{j,k}+1-w_l)+\prod_{l=1,l\neq k}^{L}(1-w_l),\frac{\partial g}{\partial w}=-\mu_w'\cdot\prod_{l=1}^{L}(1-w_l)+\mu\cdot\prod_{l=1,l\neq k}^{L}(1-w_l).$$

黄德根(1965－),男,福建南平人,博士,教授,博士生导师,CCF 高级会员,主要研究领域为机器学习,自然语言处理,机器翻译.

张云霞(1987－),女,博士,主要研究领域为计智能信息处理,不确定性推理.

林红梅(1996－),女,学士,主要研究领域为智能信息处理.

邹丽(1971－),女,博士,教授,CCF 专业会员,主要研究领域为智能信息处理.

刘壮(1982－),男,博士生,CCF 学生会员,主要研究领域为自然语言处理,机器阅读理解、问答.