

# 动量梯度法优化的置信规则库推理方法

作者名<sup>1)</sup> 作者名<sup>2),3)</sup> 作者<sup>3)</sup>

<sup>1)</sup>(单位全名 部门(系) 全名, 市(或直辖市) 国家名 邮政编码)

<sup>2)</sup>(单位全名 部门(系) 全名, 市(或直辖市) 国家名 邮政编码)

<sup>3)</sup>(单位全名 部门(系) 全名, 市(或直辖市) 国家名 邮政编码)

**摘 要** 扩展置信规则库在传统置信规则库基础上引入规则前件属性置信分布结构, 使用训练数据构建规则库的规则, 该方法简单高效。但在训练数据数目较多时需要进行规则约简或使用数据结构优化规则的存储与激活过程。为了解决这些问题, 本文对传统扩展置信规则库的结构进行简化, 取消规则前件属性置信分布形式, 引入高斯函数优化规则激活权重计算过程, 最终使用动量优化的随机梯度下降法训练置信规则库推理系统的参数。在实验分析中, 通过拟合多极值非线性函数验证该方法的有效性, 并在多个公共数据集比较该方法与其他置信规则库参数训练方法, 实验结果表明该方法在简化置信规则库结构、优化激活过程的同时提升了推理的精度。

**关键词** 置信规则库; 结构优化; 参数训练; 梯度法; 动量优化

中图法分类号 TP391 DOI号 10.11897/SP.J.1016.01.2020.00001

## Belief rule base inference method based on gradient descent with momentum

NAME Name-Name<sup>1)</sup> NAME Name<sup>2),3)</sup> NAME Name-Name<sup>3)</sup>

<sup>1)</sup>(Department of \*\*\*\*, University, City ZipCode, Country)

<sup>2)</sup>(Department of \*\*\*\*, University, City ZipCode)

<sup>3)</sup>(Department of \*\*\*\*, University, City ZipCode, Country)

**Abstract** Extended belief rule base introduce the belief structure of rule antecedent attributes based on the traditional belief rule base, and use the training data to construct the rules of the rule base. This method is simple and efficient. However, when the number of training data is large, it is necessary to perform rule reduction or use data structure to optimize the storage and activation process of rules. In order to solve these problems, this paper simplifies the structure of the traditional extended belief rule base, cancels the pre-attribute attribute belief distribution form, introduces the Gaussian function to optimize rule activation weight calculation method, and finally uses the momentum-optimized stochastic gradient descent method to train the belief rule base parameters. In the experimental analysis, the method is validated by fitting a multi-extremal nonlinear function, and the method is compared with other belief rule base parameter training methods in multiple public data sets. Experimental results show that the method simplifies the structure of the belief rule base and optimizes the activation process while improving the accuracy of inference.

**Key words** belief rule base; structure optimization; Parameter training; Gradient method; Momentum optimization

## 1 引言

置信规则库<sup>[1]</sup>及其推理方法是 Yang 等在传统 IF-THEN<sup>[2]</sup>生成式规则的基础上通过引入置信分

布结构解决不确定信息的专家系统。Yang 使用 D-S 证据理论<sup>[3,4]</sup>、决策理论<sup>[5]</sup>、模糊理论<sup>[6]</sup>建模对该专家系统进行建模, 使其可以有效的综合不同定量、定性信息与模糊不确定信息, 广泛的应用于不同领域的各种难题包括输油管道泄漏检测<sup>[7]</sup>、军事能力预估<sup>[8]</sup>、消费者行为预测<sup>[9]</sup>等。

在置信规则库推理过程中, 系统中属性权重、规则权重、置信度分布等参数直接影响了最终推理预测结果的精度, 为了避免人为设给定参数影响推理精度, Yang 使用 Fmincon<sup>[10]</sup>函数优化模型参数以提升系统的推理精度。Chang<sup>[11]</sup>首先使用了优化步长的梯度法直接寻找梯度下降方向进行参数训

收稿日期: 2020-07-24; 修改日期: 2020-07-24 本课题得到……基金中文完整名称 (No. 项目号)、……基金中文完整名称 (No. 项目号)、……基金中文完整名称 (No. 项目号) 资助。作者名, 性别, xxxx 年生, 学位(或目前学历), 职称, 是/否计算机学会 (CCF) 会员 (提供会员号), 主要研究领域为 \*\*\*\*、\*\*\*\*. E-mail: \*\*\*\*. 作者名, 性别, xxxx 年生, 学位(或目前学历), 职称, 是/否计算机学会 (CCF) 会员 (提供会员号), 主要研究领域为 \*\*\*\*、\*\*\*\*. E-mail: \*\*\*\*. 作者(通信作者), 性别, xxxx 年生, 学位(或目前学历), 职称, 是/否计算机学会 (CCF) 会员 (提供会员号), 主要研究领域为 \*\*\*\*、\*\*\*\*. E-mail: \*\*\*\*. 第 1 作者手机号码: ……., E-mail: \*\*\*\*.

练优化步长信息, 随后 Chang<sup>[12]</sup> 使用梯度法与二分法结合分别求解模型参数梯度下降方向并根据约束条件二分寻找最大步长实现了对模型参数的快速优化, Wu<sup>[13]</sup> 提出更全面的加速梯度法优化模型参数, 提升了梯度法训练下推理模型的精度。同时还有 Su<sup>[14]</sup> 提出的粒子群算法, Wang<sup>[15]</sup> 提出的差分进化算法等一系列智能算法对置信规则库推理系统进行参数训练。在传统置信规则库推理系统的基础上, Liu<sup>[16]</sup> 引入了置信分布结构的前件属性形式并使用训练数据构建规则库的扩展置信规则推理系统, 简化了规则库的构建过程并提高了推理速度。

目前置信规则库系统的参数优化模型多基于各种智能算法, 其过程较为复杂且中间训练参数多, 传统的梯度法对梯度下降方向步长的选取受到规则库参数约束条件限制。而扩展置信规则库由于未引入参数训练过程使得系统对选取构建规则库的数据代表性要求较高且在规则数据较多的情况下需要进行规则约简或使用数据结构优化规则的存储与激活过程。针对这些问题本文在扩展置信规则库的基础上通过简化规则前件置信结构, 并针对新的前件结构提出了一种基于高斯函数的规则权重激活方法, 避免了规则零激活问题, 并使用线性整流函和指数归一化函数数避免了参数训练过程中权重与置信分布数值受到约束条件限制, 最后通过求导证明了推理系统中参数都可以使用梯度法进行训练, 并应用了动量优化的随机梯度下降法训练整个推理系统的参数, 提高了推理精度与参数训练速度。在实验中通过对多极值非线性函数的拟合与在多个公共分类数据集上同以往方法的对比验证了本文方法的有效性。

## 2 改进的置信规则库推理系统

### 2.1 简化前件置信分布结构的置信规则库构建

在传统规则产生式的基础上, Yang<sup>[1]</sup> 通过引入置信分布结构、规则属性与权重参数提出了置信规则的表示形式, 其具体表示如下:

$$R_k : if \{X_1 is A_1^k \wedge \dots \wedge X_{T_k} is A_{T_k}^k\} \\ then \{(D_1, \beta_1^k), \dots, (D_N, \beta_N^k)\}, \sum_{i=1}^N \beta_i^k \leq 1$$

等号在规则信息完整时取得, 同时每条规则有其规则权重  $\theta_k$ , 前件属性权重  $\delta_1, \delta_2, \dots, \delta_{T_k}$ , 扩展置信

规则库推理系统在此基础上通过对前件属性引入置信分布结构, 其规则形式表示如下:

$$R_k : if \{[(A_{11}^k, \alpha_{11}^k), \dots, (A_{1J_1}^k, \alpha_{1J_1}^k)] \wedge \\ \dots \wedge [(A_{T_k 1}^k, \alpha_{T_k 1}^k), \dots, (A_{T_k J_{T_k}}^k, \alpha_{T_k J_{T_k}}^k)]\} \\ then \{(D_1, \beta_1^k), \dots, (D_N, \beta_N^k)\}, \sum_{i=1}^N \beta_i^k \leq 1$$

使用原始数据构建的扩展置信规则库将输入数据转换为具有置信分布形式的规则前件属性, 对于输入数据

$$X^k = (x_1^k, \dots, x_T^k)$$

考虑对第  $i$  个属性参数进行转换构建对应规则第  $i$  个具有置信分布形式的前件属性

$$E(x_i^k) = [(A_{i1}, \alpha_{i1}^k), \dots, (A_{iJ_i}, \alpha_{iJ_i}^k)]$$

设第  $i$  属性的对应参考值由  $\{\gamma_{ij} | j = 1, \dots, J_i, \gamma_{ij} < \gamma_{i(j+1)}\}$  给出, 则对应的转换公式为:

$$\alpha_{ij}^k = \frac{\gamma_{i(j+1)} - x_i^k}{\gamma_{i(j+1)} - \gamma_{ij}}, \gamma_{ij} \leq x_i^k \leq \gamma_{i(j+1)}$$

$$\alpha_{i(j+1)}^k = 1 - \alpha_{ij}^k, \gamma_{ij} \leq x_i^k \leq \gamma_{i(j+1)}$$

$$\alpha_{it}^k = 0, t = 1, \dots, (j-1), (j+2), \dots, J_i$$

根据相同的转换方法可以将数据输出转换成规则的结果属性置信分布。

由输入数据转化生成规则的过程需要人为设定前件属性信息的参考值并消耗计算资源, 同时需要额外空间存储具有置信分布结构的规则前件属性信息。为了解决上述问题, 本文提出了改进的置信规则形式, 通过取消前件属性置信分布结构, 简化构建置信规则形式如下:

$$R_k : if (x_1^k, \dots, x_{T_k}^k) then \{(D_1, \beta_1^k), \dots, (D_N, \beta_N^k)\}, \sum_{i=1}^N \beta_i^k \leq 1$$

简化后使用训练数据生成规则前件属性时仅需存储原始信息而无需转换为置信分布结构, 避免人为设定前件属性参考等级, 节省了计算资源与存储空间。

### 2.2 高斯函数优化的证据推理合成方法

对规则库的规则进行激活合成是置信规则库推理系统的核心内容。整个过程主要包括两个步骤: 计算激活权重、激活规则合成, 整个流程如图 1 所示。

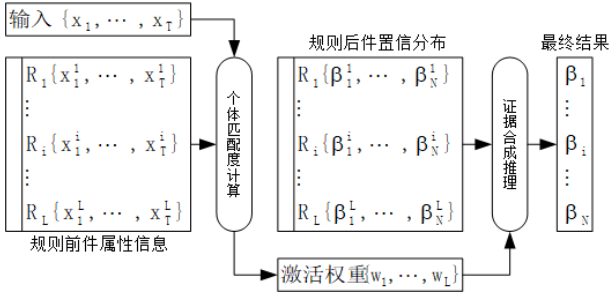


图1 置信规则库证据合成推理流程

传统扩展置信规则库中每条规则激活权重的计算可看作两个置信分布之间相似度的计算，使用欧式距离对于第  $i$  个属性的个体匹配度进行计算，将输入数据对应属性信息转换为置信分布形式后该属性个体匹配度为：

$$S_i^k = 1 - d_i^k = 1 - \sqrt{\sum_{j=1}^{J_i} (\alpha_{i,j} - \alpha_{i,j}^k)^2}$$

每个属性的个体匹配度计算完成后，对所有属性进行聚合，合取规则形式下聚合函数为：

$$\alpha_k = \prod_{i=1}^{T_k} (S_i^k)^{\delta_i}, \bar{\delta}_i = \frac{\delta_i}{\max_{j=1, \dots, T_k} \delta_j}$$

该条规则的激活权重由以下公式计算得到：

$$w_k = \frac{\theta_k \alpha_k}{\sum_{l=1}^L \theta_l \alpha_l}$$

规则权重归一化操作使得所有权重满足  $0 \leq w_k \leq 1, \sum w_k = 1$ 。

使用置信分布相似度作为激活权重计算的方法已经不再适用于简化后的置信规则形式，为了进行有效的权重激活，本文使用高斯函数计算个体匹配度进行激活权重计算。对于第  $i$  个属性信息，输入数据  $X = (x_1, \dots, x_{T_k})$  与规则库中第  $k$  条规则  $R_k : if(x_1^k, \dots, x_{T_k}^k) then \{(D_1, \beta_1^k), \dots, (D_N, \beta_N^k)\}$  在该属性上个体匹配度使用高斯函数计算为：

$$S_i^k = e^{-\frac{a(x_i - x_i^k)^2}{b}}, a > 0, b > 0$$

其中正数参数  $a$  代表该属性对距离的敏感程度，距离不变时取值越大越敏感，属性匹配度越接近 0。正数参数  $b$  代表该属性的权重，距离不变时取值越大属性重要度越低，属性匹配度越接近 1。为了减少参数训练的数量，将参数  $a$  与  $b$  组合形成该属性对距离的敏感程度与属性权重相融合的属性辅助参数  $\delta = \frac{a}{b}$ ，该属性个体匹配度可记为：

$$S_i^k = e^{-\delta(x_i - x_i^k)^2}, \delta > 0$$

对于合取规则置信规则库中单条规则激活权重由下式计算：

$$w_k = \frac{\theta_k \alpha_k}{\sum_{i=1}^L \theta_i \alpha_i}, \alpha_k = \prod_{i=1}^{T_k} S_i^k = e^{-\sum_{i=1}^{T_k} \delta_i (x_i - x_i^k)^2}, \delta_i > 0$$

激活过程中参数  $\delta$  的取值可通过优化算法进行计算，也可以根据经验人为设定。对属性参数进行标准化去量纲处理后可直接设置参数  $\delta = 1$ ，即初始情况下属性不受距离敏感程度参数与属性权重参数影响。

传统置信规则库推理系统计算规则前件置信分布结构个体匹配度过程中，由于原始数据转换生成的规则在单个属性上至多仅有 2 个相邻的非零评价等级，导致不在这 2 个评价等级激活区域内的属性参数激活值为零。使用了改进的置信规则结构与基于高斯函数的激活权重计算方法后，任意规则的属性个体匹配度聚合后都是非零正值，避免了规则零激活问题影响推理系统性能。

规则权重计算完成后对所有规则进行合成并获得推理结果，首先将规则的置信分布转化为对应的可信度信息：

$$m_{j,k} = w_k \beta_j^k, j = 1, \dots, N$$

$$m_{D,k} = 1 - \sum_{j=1}^N m_{j,k} = 1 - w_k \sum_{j=1}^N \beta_j^k$$

$$\bar{m}_{D,k} = 1 - w_k, \quad \tilde{m}_{D,k} = w_k (1 - \sum_{j=1}^N \beta_j^k)$$

其中  $m_{j,k}$  代表第  $k$  条规则在第  $j$  个结果属性上的可信度， $\bar{m}_{D,k}$  代表第  $k$  条规则未分配给任何结果属性的可信度， $\tilde{m}_{D,k}$  代表第  $k$  条规则结果参考属性不完备导致缺失的可信度，不确定性可信度总和由  $m_{D,k} = \bar{m}_{D,k} + \tilde{m}_{D,k}$  给出。根据每条规则的可信度信息进行合成，得到第  $j$  个结果属性最终置信度结果：

$$m_j = k[\prod_{i=1}^L (m_{j,i} + m_{D,i}) - \prod_{i=1}^L m_{D,i}]$$

$$\bar{m}_D = n[\prod_{i=1}^L \bar{m}_{D,i}], \quad \tilde{m}_D = k[\prod_{i=1}^L m_{D,i} - \prod_{i=1}^L \bar{m}_{D,i}]$$

$$n = [\sum_{j=1}^N \prod_{i=1}^L (m_{j,i} + m_{D,i}) - (N-1) \prod_{i=1}^L m_{D,i}]^{-1}$$

$$\beta_j = \frac{m_j}{1 - \bar{m}_D}, \quad \beta_D = \frac{\tilde{m}_D}{1 - \bar{m}_D}$$

### 2.3 改进的置信规则库推理系统复杂度分析

本文提出的简化前件属性置信分布结构与使用高斯函数优化激活权重计算方法对置信规则库推理系统复杂度优化分为构建与推理两个部分。假设前件属性有  $T$  个, 每个属性的候选数目  $J$  个, 规则总数  $L$  条, 一共  $N$  个结果评价等级。传统扩展置信规则库推理系统在构建过程中, 生成每条规则的时间复杂度为  $O(TJ)$ , 所有规则生成的时间复杂度为  $O(LTJ)$ 。在推理过程中, 对输入数据的转换与规则匹配度计算的时间复杂度为  $O(LTJ)$ , 对规则进行证据合成推理的时间复杂度为  $O(NL)$ , 推理过程的总复杂度为  $O((N + TJ) \times L)$ 。

优化后的置信规则库推理系统由于简化了规则前件属性置信度分布结构, 生成每条规则的时间复杂度降为  $O(T)$ , 对输入数据的转换与规则匹配度计算的时间复杂度降为  $O(LT)$ , 使得规则库构建与推理过程的总复杂度分别下降  $O(LT)$  和  $O((N + T) \times L)$ , 同时避免了零激活问题。

## 3 动量优化随机梯度下降法参数训练

传统梯度法应用于置信规则库推理系统参数训练过程中受到规则属性偏导公式构造困难、训练步长受参数约束条件限制等问题。上文改进的置信规则库推理系统避免了传统置信规则系统求导困难的问题, 同时通过引入指数归一化函数与线性整流函数避免参数训练过程中受到特定约束。最终采用动量优化随机梯度下降法进行参数训练, 在降低计算资源消耗的同时提高了训练速度。

### 3.1 改进的置信规则库推理系统梯度求解与可微分证明

置信规则库推理的最终置信分布形式由公式给出, 展开得到第  $j$  个结果属性数值:

$$\beta_j = \frac{m_j}{1 - \bar{m}_D}$$

$$= \frac{\prod_{i=1}^L (m_{j,i} + m_{D,i}) - \prod_{i=1}^L m_{D,i}}{\sum_{t=1}^N \prod_{i=1}^L (m_{t,i} + m_{D,i}) - (N-1) \prod_{i=1}^L m_{D,i} - \prod_{i=1}^L \bar{m}_{D,i}}$$

本文推理系统的实现及相关实验均是在信息完整情况下进行, 规则结果属性的置信分布中不包括不确定信息:

$$\sum_{i=1}^N \bar{\beta}_{ik} = 1 (k = 1, \dots, L), m_{D,i} = \bar{m}_{D,i}, \tilde{m}_{D,i} = 0$$

完备情况下第  $j$  个结果属性表示为:

$$\begin{aligned} \beta_j &= \frac{\prod_{i=1}^L (m_{j,i} + m_{D,i}) - \prod_{i=1}^L m_{D,i}}{\sum_{t=1}^N \prod_{i=1}^L (m_{t,i} + m_{D,i}) - N \times \prod_{i=1}^L m_{D,i}} \\ &= \frac{\prod_{i=1}^L (\frac{m_{j,i}}{m_{D,i}} + 1) - 1}{\sum_{k=1}^N [\prod_{i=1}^L (\frac{m_{k,i}}{m_{D,i}} + 1) - 1]} \end{aligned}$$

第  $j$  个结果属性未进行归一化的置信度表示为:

$$\bar{\beta}_j = \prod_{i=1}^L (\frac{m_{j,i}}{m_{D,i}} + 1) - 1, \beta_j = \frac{\bar{\beta}_j}{\sum_{k=1}^N \bar{\beta}_k}$$

将  $m_{j,i} = w_i \beta_{j,i}, m_{D,i} = 1 - w_i, w_i = \frac{\theta_i \alpha_i}{\sum_{k=1}^L \theta_k \alpha_k}$  代入第  $j$  个未归一化置信度表达式中有:

$$\bar{\beta}_j = \prod_{i=1}^L (\frac{\theta_i \alpha_i \beta_{j,i}}{\sum_{k=1}^L \theta_k \alpha_k} + 1) - 1$$

根据高斯函数优化的规则权重激活过程, 第  $t$  规则个体匹配度聚合值  $\alpha_t$  由该条规则前件属性信息与属性辅助参数的计算得到, 设置最终置信度输出结果的损失函数表示为  $loss = Loss(\beta_1, \dots, \beta_N)$ , 根据多元复合函数求导链式法则, 最终结果损失对第  $t$  条规则第  $n$  个前件属性求导为:

$$\frac{dloss}{dx_n^t} = \sum_{i=1}^N \sum_{j=1}^N \frac{loss}{d\beta_i} \frac{d\beta_i}{d\bar{\beta}_j} \frac{d\bar{\beta}_j}{d\alpha_t} \frac{d\alpha_t}{dx_n^t}$$

最终结果损失对第  $n$  个规则属性辅助参数求导为:

$$\frac{dloss}{d\delta_n} = \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^L \frac{loss}{d\beta_i} \frac{d\beta_i}{d\bar{\beta}_j} \frac{d\bar{\beta}_j}{d\alpha_k} \frac{d\alpha_k}{d\delta_n}$$

最终结果损失对第  $t$  条规则的规则权重求导为:

$$\frac{dloss}{d\theta_t} = \sum_{i=1}^N \sum_{j=1}^N \frac{dloss}{d\beta_i} \frac{d\beta_i}{d\bar{\beta}_j} \frac{d\bar{\beta}_j}{d\theta_t}$$

注意到任意规则的第  $j$  个结束属性置信度仅与第  $j$  个未归一化的结果属性置信度相关, 最终结果损失对第  $t$  条规则第  $j$  个结果置信分布求导为:

$$\frac{dloss}{d\beta_{j,t}} = \sum_{i=1}^N \sum_{j=1}^N \frac{dloss}{d\beta_i} \frac{d\beta_i}{d\bar{\beta}_j} \frac{d\bar{\beta}_j}{d\beta_{j,t}}$$

由于上述导数中除损失函数偏导外其他部分的偏导均存在 <sup>A</sup> 且连续 <sup>B</sup>, 当选定适当的损失函数后, 最终结果损失对推理模型的所有参数都是可微分的, 这为使用梯度法优化模型参数提供了条件。



根据上述导数与选定的损失函数,可以得到损失结果在模型各部分参数上的梯度。损失函数在所有规则前件属性参数上的梯度是:

$$\nabla_x loss = \begin{bmatrix} \frac{dloss}{dx_1^1} & \cdots & \frac{dloss}{dx_{T_k}^1} \\ \vdots & \ddots & \vdots \\ \frac{dloss}{dx_1^L} & \cdots & \frac{dloss}{dx_{T_k}^L} \end{bmatrix}$$

损失函数在所有规则属性辅助参数上的梯度是:

$$\nabla_\delta loss = \begin{bmatrix} \frac{dloss}{d\delta_1} & \cdots & \frac{dloss}{d\delta_{T_k}} \end{bmatrix}$$

损失函数在所有规则权重上的梯度是:

$$\nabla_\theta loss = \begin{bmatrix} \frac{dloss}{d\theta_1} & \cdots & \frac{dloss}{d\theta_L} \end{bmatrix}$$

损失函数在所有规则结果置信度分布上的梯度是:

$$\nabla_\beta loss = \begin{bmatrix} \frac{dloss}{d\beta_{1,1}} & \cdots & \frac{dloss}{d\beta_{N,1}} \\ \vdots & \ddots & \vdots \\ \frac{dloss}{d\beta_{1,L}} & \cdots & \frac{dloss}{d\beta_{N,L}} \end{bmatrix}$$

根据推理系统输出的置信分布与结果的损失函数对各部分参数的梯度,通过沿着负梯度方向更新参数进行模型优化。由于推理模型的参数有各自的约束条件,使用概率归一化函数和线性整流函数对受约束参数进行预处理避免训练过程中破坏特定的参数约束条件。为了满足结果置信度和为 1 且各个置信度非负,使用指数归一化函数:

$$\beta_j^k = \frac{e^{\bar{\beta}_j^k}}{\sum_{i=1}^N e^{\bar{\beta}_i^k}} (\sum_{i=1}^N \beta_j^k = 1, \beta_j^k > 0)$$

为了满足规则权重和属性辅助参数的非负约束,使用激活函数处理可能出现的负数情况,可使用线性整流函数 SoftPlus:

$$\theta_n = f(\bar{\theta}_n), \delta_n = f(\bar{\delta}_n), f(x) = \ln(1 + e^x)$$

### 3.2 动量优化随机梯度下降法训练参数

使用梯度下降法更新推理模型参数的优化过程由下面式子给出:

$$M_{new}(x, \beta, \theta, \delta) = M_{old} - \mu \nabla_{M_{old}} loss$$

其中梯度信息根据最终置信度分布的损失函数给出,学习率  $\mu$  即更新的步长信息需要设定。Chang<sup>[11]</sup>与 Wu<sup>[13]</sup>给出的梯度训练过程中使用二分法在约束空间中迭代寻找最优步长,在梯度为 0 时还需添

加摄动参数避免训练过程停滞。上文提到新的梯度参数模型使用了概率归一化函数和线性整流函数预处理受约束参数,在满足约束条件的同时对步长无限制要求。为了提高训练的速度,本文使用了动量优化随机梯度下降的方法。

随机梯度下降法<sup>[18]</sup>是一种在目标函数可微分条件下的迭代优化方法,使用数据的随机子集计算得到的梯度作为整个数据集梯度的估算值,在高维优化问题中降低了计算负担,以较低的收敛速度实现快速迭代训练。对于所有数据样本,其损失函数输出由所有样本的决定并根据其梯度更新模型参数

$$loss = \frac{1}{n} \sum_i^n Loss(\beta_i), M_{new} = M_{old} - \mu \nabla_{M_{old}} loss$$

随机梯度下降法随机选取了单个样本作为所有输出样本损失函数的预估值进行参数更新

$$M_{new} = M_{old} - \mu \nabla_{Loss(\beta_i)} M_{old}$$

避免训练过程中对每个样本的梯度进行计算,随机近似过程和凸优化理论<sup>[19]</sup>证明了选取适当学习率  $\mu$  时其收敛性,且在目标函数是凸函数或伪凸时随机梯度下降法几乎可以肯定的收敛到全局最小值。

随机梯度下降法的明显缺点是其更新方向完全依赖当前样本的梯度因而十分不稳定,本文采用了动量法<sup>[20]</sup>去优化随机梯度下降过程。动量法借用物理学中的动量概念,模拟运动时物体的惯性,即在更新过程中保留一定程度的历史梯度信息,与当前样本梯度结合,可以同时增加更新过程的稳定性与速度,并增强了摆脱局部最优解的能力。动量法使用加权融合历史更新参数与当前梯度信息作为本次更新参数:

$$v_t = \beta v_{t-1} + \alpha \nabla_H M, M = M - v_t$$

将初始  $v_0$  设置为 0,  $\beta$  一般设置为 0.9。若当前梯度与之前更新方向相同,则动量法会使得模型在该梯度方向上加速参数更新过程,提高训练的速度。若当前梯度方向与之前更新方向相反,则动量法可以抑制本次梯度信息造成的模型参数震荡。

## 4 实验与结果

本文引入非线性多极值函数及多个公共数据分类集,研究验证梯度训练下改进的置信规则库推

理模型的训练效率与精确度。实验中模型使用 TensorFlow 2 搭建, 运行于 Intel® Core™ i5 8500@3GHz, 16GB 内存, Nvidia 1060 6GB 显卡, Ubuntu 20.04 操作系统。

#### 4.1 非线性函数拟合

Liu<sup>[16]</sup> 证明了置信规则库可以逼近任意函数, 本节通过引入一个非线性多极值函数进行训练验证模型的推理性能。非线性多极值函数如下所示:

$$f(x) = e^{-(x-2)^2} + 0.5e^{-(x+2)^2}, -5 \leq x \leq 5$$

在其定义域上均匀的选取 1000 组数据作为训练数据, 使用均方误差作为损失函数, 由函数曲线图所显示的各个极值点可设置规则库结果属性评价等级与相应的效用值:

$$\{D_1, D_2, D_3, D_4, D_5\} = \{-0.5, 0, 0.5, 1, 1.5\}$$

在规则构建过程中, 通过随机生成定义域范围内的点作为规则前件属性信息, 同时使用随机正态分布生成规则的置信分布并使用指数归一化函数进行预处理, 设定规则数目为 16, 初始属性辅助参数与初始规则权重均设置为 1。进行梯度训练时学习率  $\alpha = 0.01$ , 动量优化参数  $\beta = 0.9$ , 随机梯度下降训练过程中每批次样本数量为 64, 对数据集训练 1000 批次后终止。

对比图 1 初始随机生成规则的模型推理效果与模型训练完成后的推理效果, 证明了梯度下降法能有效的提高推理系统的精确度。

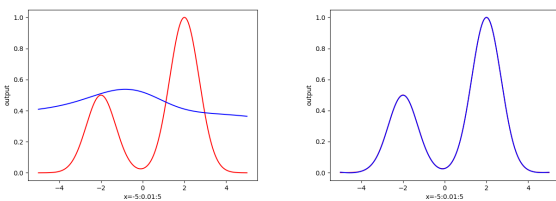


图 2 模型训练前后的输出分别与函数图像比较

图 2 为每批次训练样本上的损失函数值, 可以发现动量优化的随机梯度下降法训练速度极快, 在第 10 秒均方差损失就下降到  $2 \times 10^{-5}$  水平。

表 1 中对比了本文方法与 Fmincon 方法、Chen 的局部优化参数训练方法、Wang 的专家干预差分进化法及 Li 的布谷鸟搜索法, 可以发现新方法无论在精度还是速度上都较以往方法有极大提升。

#### 4.2 公共分类数据集测试

本文从 UCI 上选择了 9 种较为常见的公共分类数据集进行测试, 数据集详细信息见表 2。使

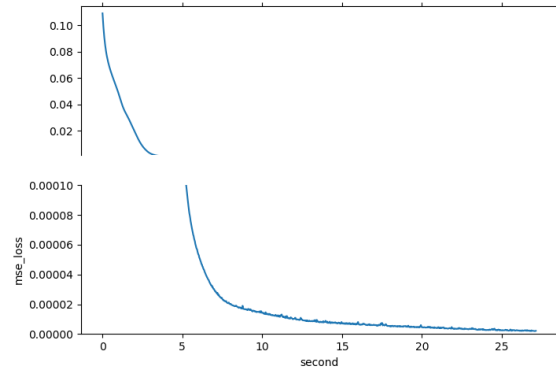


图 3 训练过程中每批次样本上的损失函数值

表 1 不同参数训练方法训练结果

	MSE 均方误差	运行时间
Fmincon	$5.16 \times 10^{-5}$	496.71
Chen-BRB	$6.3228 \times 10^{-5}$	-
Wang-BRB	$3.9284 \times 10^{-5}$	386.63
Li-BRB	$3.3322 \times 10^{-5}$	357
本文方法	$1.3691 \times 10^{-6}$	30.60

用 10 折交叉验证实验方法生成训练数据与测试数据, 并重复实验 10 次求平均实验结果减少误差波动。模型使 32 条规则, 为了简化前件属性参数生成过程, 对训练数据所有属性信息进行标准化去量纲预处理, 并使用标准正态分布随机生成所有前件属性参数值, 同时使用标准正态分布随机生成结果置信分布并进行指数归一化预处理, 初始化所有属性辅助参数与规则权重为 1。梯度训练过程中批次大小设置为 128, 训练 1000 批次后终止, 学习率与动量优化参数分别设置为 0.01 与 0.9, 使用交叉熵作为分类结果的损失函数。表 3 中使用准确率作为衡量分类结果评价指标, 统计对比本文方法与 Liu-BRB<sup>[16]</sup>、DRA-BRB<sup>[21]</sup>、SRA-BRB<sup>[22]</sup>、VP-BRB、MVP-BRB<sup>[23]</sup> 这几种置信规则库系统的分类效果。

从表 3 中可以发现本文方法在超过一半的数据集上超越了其他置信规则库推理方法, 在 ecoli、glass、thyroid、transfusion 四个数据集上未能超越现有方法, 且仅在 glass 上落后其他方法较多, 其余数据集上均超过或达到其他推理方法的水平。平均排名结果最高。验证了本文方法在推理精度与效率上均超越了以往方法。

表 2 分类数据集信息

数据集	类别数目	属性数目	样本数目
iris	3	4	150
wine	3	13	178
diabetes	2	8	768
ecoli	8	7	336
glass	6	9	214
seeds	3	7	210
yeast	10	8	1484
thyroid	3	5	215
transfusion	2	4	748

表 3 分类数据集实验结果

数据集	Liu-BRB	DRA-BRB	SRA-BRB	VP-BRB	MVP-BRB	本文方法
iris	95.33(4)	95.50(3)	94.80(6)	95.13(5)	95.87(2)	96.50(1)
wine	96.32(5)	96.46(4)	96.85(3)	93.01(6)	97.02(2)	97.44(1)
diabetes	73.39(2)	71.44(6)	71.71(5)	71.89(4)	72.59(3)	73.64(1)
ecoli	81.61(6)	83.76(5)	84.85(4)	84.87(3)	85.61(1)	85.43(2)
glass	67.85(6)	69.65(5)	73.08(1)	71.75(3)	72.06(2)	69.91(4)
seeds	91.33(5)	92.02(4)	91.24(6)	92.57(2)	92.38(3)	94.02(1)
yeast	45.61(6)	54.13(5)	56.85(4)	58.15(2)	57.49(3)	59.49(1)
thyroid	81.47(3)	97.19(1)				95.95(2)
transfusion	76.14(6)	76.57(4)		77.33(3)	80.36(1)	79.45(2)
average rank	4.77	4.11	4.14	3.5	2.12	1.66

5 结束语

针对以往扩展置信库推理系统中构建规则消耗计算资源，规则权重零激活问题，本文提出通过简化前件属性置信分布结构，使用高斯函数优化激活权重计算过程，同时使用动量优化的随机梯度下降方法训练模型参数解决对应问题并提高推理性能。该方法结合结构优化与参数训练两个方面，通过优化结构使快速的梯度训练方法能够应用，最终提升了整个系统的推理效率。该方法具有良好的扩展性能，在实验分析中，本文方法在多个数据集上都较以往方法具有更高的推理性能。在接下来的工作中，以这种简化模型与快速参数训练的方法将使得多个推理模型进行集成学习成为可能的研究方向。

参考文献

[1] YANG J B, LIU J, WANG J, et al. Belief rule-base inference methodology using the evidential reasoning approach-rimer[J]. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 2006, 2(36):266-285.

[2] SUN R. Robust reasoning: integrating rule-based and similarity-based reasoning[J]. Artificial Intelligence, 1995, 2(75):241-295.

[3] DEMPSTER A P. A generalization of bayesian inference[J]. Journal of the Royal Statistical Society: Series B (Methodological), 1968, 2(30): 205-232.

[4] SHAFER G, SMITH A F M. A mathematical theory of evidence[J]. Biometrics, 1976, 3(32):703.

[5] YOON K, HWANG C L. Multiple attribute decision making[M]. Thousand Oaks, CA: Sage Publications, 1995.

[6] ZADEH L. Fuzzy sets[J]. Information and Control, 1965, 3(8):338-353.

[7] ZHOU Z J, HU C H, YANG J B, et al. Online updating belief rule based system for pipeline leak detection under expert intervention[J]. Expert Systems with Applications, 2009, 4(36):7700-7709.

[8] JIANG J, LI X, JIE ZHOU Z, et al. Weapon system capability assessment under uncertainty based on the evidential reasoning approach[J]. Expert Systems with Applications, 2011.

[9] YANG Y, FU C, CHEN Y W, et al. A belief rule based expert system for predicting consumer preference in new product development[J]. Knowledge-Based Systems, 2016(94):105-113.

[10] YANG J B, LIU J, XU D L, et al. Optimization models for training belief-rule-based systems[J]. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 2007, 4(37):569-585.

[11] 常瑞, 张速. 基于优化步长和梯度法的置信规则库参数学习方法[J]. 华北水利水电学院学报, 2011, 1(32):154-157.

[12] CHANG R Y, WANG H, YANG J B. An algorithm for training parameters in belief rule-bases based on the gradient and dichotomy methods [J]. Systems Engineering, 2007.

[13] 吴伟昆, 杨隆浩, 傅仰耿, 等. 基于加速梯度求法的置信规则库参数训练方法[J]. 计算机科学与探索, 2014, 8(8):989-1001.

[14] 苏群, 杨隆浩, 傅仰耿, 等. 基于变速粒子群优化的置信规则库参数训练方法[J]. 计算机应用, 2014, 34(8):2161-2165.

[15] 王韩杰, 杨隆浩, 傅仰耿, 等. 专家干预下置信规则库参数训练的差分进化算法[J]. 计算机科学, 2015, 42(5):88-93.

[16] LIU J, MARTINEZ L, CALZADA A C, et al. A novel belief rule base representation, generation and its inference methodology[J]. Knowledge-Based Systems, 2013(53):129-141.

[17] SHANNON C E. A mathematical theory of communication[J]. Bell System Technical Journal, 1948, 4(27):623-656.

[18] ROBBINS H, MONRO S. A stochastic approximation method[J]. The Annals of Mathematical Statistics, 1951, 3(22):400-407.

[19] KIWIEL K C. Convergence and efficiency of subgradient methods for quasiconvex minimization[J]. Mathematical Programming, 2001, 1(90):1-25.

[20] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986, 6088(323): 533-536.

[21] CALZADA A, LIU J, WANG H, et al. A new dynamic rule activation method for extended belief rule-based systems[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 4(27):880-894.

[22] 林燕清, 傅仰耿. 基于改进相似性度量的扩展置信规则库规则激活方法[J]. 中国科学技术大学学报, 2018, 48(1):21-27.

- [23] LIN Y Q, FU Y G, SU Q, et al. A rule activation method for extended belief rule base with vp-tree and mvp-tree[J]. Journal of Intelligent and Fuzzy Systems, 2017, 33(6):3695-3705.

## 附录 A 置信规则库推理模型偏导求解

第  $i$  个结果属性置信度  $\beta_i$  对第  $j$  个未归一化结果属性置信度  $\bar{\beta}_j$  求导为:

$$\frac{d\beta_i}{d\bar{\beta}_j} = \begin{cases} \frac{\sum_{k \neq j}^N \bar{\beta}_k}{(\sum_{k=1}^N \bar{\beta}_k)^2}, j = i \\ -\frac{\beta_i}{(\sum_{k=1}^N \bar{\beta}_k)^2}, j \neq i \end{cases}$$

第  $j$  个未归一化结果属性置信度对第  $t$  条规则的个体匹配度聚合值求导为:

$$\begin{aligned} \frac{d\bar{\beta}_j}{d\alpha_t} &= \frac{\theta_t \beta_{j,t}}{\sum_{k \neq t}^L \theta_k \alpha_k} \prod_{i=1, i \neq t}^L \left( \frac{\theta_i \alpha_i \beta_{j,i}}{\sum_{k \neq i}^L \theta_k \alpha_k} + 1 \right) \\ &+ \sum_{l=1, l \neq t}^L \left( -\frac{\theta_l \alpha_l \beta_{j,l} \theta_t}{(\sum_{k \neq l}^L \theta_k \alpha_k)^2} \prod_{i=1, i \neq l}^L \left( \frac{\theta_i \alpha_i \beta_{j,i}}{\sum_{k \neq i}^L \theta_k \alpha_k} + 1 \right) \right) \end{aligned}$$

第  $t$  条规则个体匹配度聚合值对相应规则的第  $n$  个前件属性参数求导为:

$$\frac{d\alpha_t}{dx_n^t} = -2\delta_n(x_n - x_n^t) e^{-\sum_{i=1}^{T_k} \delta_i(x_i - x_i^t)^2}$$

第  $t$  条规则个体匹配度聚合值对第  $n$  个规则属性辅助参数求导为:

$$\frac{d\alpha_t}{d\delta_n} = -(x_n - x_n^t)^2 e^{-\sum_{i=1}^{T_k} \delta_i(x_i - x_i^t)^2}$$

第  $j$  个未归一化结果属性置信度对第  $t$  条规则的规则权重求导为:

$$\begin{aligned} \frac{d\bar{\beta}_j}{d\theta_t} &= \frac{\alpha_t \beta_{j,t}}{\sum_{k \neq t}^L \theta_k \alpha_k} \prod_{i=1, i \neq t}^L \left( \frac{\theta_i \alpha_i \beta_{j,i}}{\sum_{k \neq i}^L \theta_k \alpha_k} + 1 \right) \\ &+ \sum_{l=1, l \neq t}^L \left( -\frac{\theta_l \alpha_l \beta_{j,l} \alpha_t}{(\sum_{k \neq l}^L \theta_k \alpha_k)^2} \prod_{i=1, i \neq l}^L \left( \frac{\theta_i \alpha_i \beta_{j,i}}{\sum_{k \neq i}^L \theta_k \alpha_k} + 1 \right) \right) \end{aligned}$$

第  $j$  个未归一化结果属性置信度对第  $t$  条规则对应第  $j$  个结果属性置信度求导为:

$$\frac{d\bar{\beta}_j}{d\beta_{j,t}} = \frac{\theta_t \alpha_t}{\sum_{k \neq j}^L \theta_k \alpha_k} \prod_{i=1, i \neq t}^L \left( \frac{\theta_i \alpha_i \beta_{j,i}}{\sum_{k \neq i}^L \theta_k \alpha_k} + 1 \right)$$

## 附录 B 置信规则库推理模型可微分证明

置信规则库推理过程是一个多元复合函数过程, 根据多元复合函数可微分条件, 各中间函数皆需满足微分条件, 同时对各个变量的偏导要求存在且连续。根据损失结果对第  $t$  条规则第  $n$  个前件属性的导数表达式:

$$\frac{dloss}{dx_n^t} = \sum_{i=1}^N \sum_{j=1}^N \frac{loss}{d\beta_i} \frac{d\beta_i}{d\bar{\beta}_j} \frac{d\bar{\beta}_j}{d\alpha_t} \frac{d\alpha_t}{dx_n^t}$$

当选择损失函数的导数存在并连续时, 任选一条求导路径  $\frac{dloss}{d\beta_i} \frac{d\beta_i}{d\bar{\beta}_j} \frac{d\bar{\beta}_j}{d\alpha_t} \frac{d\alpha_t}{dx_n^t}$  进行连续证明。



对过程  $\beta_i(\bar{\beta}_1, \dots, \bar{\beta}_N)$  中对  $\bar{\beta}_j$  的偏导  $\frac{d\beta_i}{d\bar{\beta}_j}$  进行连续性证明:

明:

$$\frac{d\beta_i}{d\bar{\beta}_j}(\bar{\beta}_j) = \begin{cases} \frac{\sum_{k \neq j}^N \bar{\beta}_k}{(\sum_{k=1}^N \bar{\beta}_k)^2}, j = i \\ -\frac{\beta_i}{(\sum_{k=1}^N \bar{\beta}_k)^2}, j \neq i \end{cases}$$

由于上述偏导具有初等函数形式, 且  $\bar{\beta}_j \in (0, +\infty)$ , 该偏导在其定义域上是连续的。

对过程  $\bar{\beta}_j(\alpha_1, \dots, \alpha_L, \theta_1, \dots, \theta_L, \beta_{j,1}, \dots, \beta_{j,L})$  中对  $\alpha_i$  的偏导为  $\frac{d\bar{\beta}_j}{d\alpha_i}$  进行连续性证明:

$$\begin{aligned} \frac{d\bar{\beta}_j}{d\alpha_i}(\alpha_i) &= \frac{\theta_i \beta_{j,i}}{\sum_{k \neq i}^L \theta_k \alpha_k} \prod_{i=1, i \neq t}^L \left( \frac{\theta_i \alpha_i \beta_{j,i}}{\sum_{k \neq i}^L \theta_k \alpha_k} + 1 \right) \\ &+ \sum_{l=1, l \neq i}^L \left( -\frac{\theta_l \alpha_l \beta_{j,l} \theta_i}{(\sum_{k \neq l}^L \theta_k \alpha_k)^2} \prod_{i=1, i \neq l}^L \left( \frac{\theta_i \alpha_i \beta_{j,i}}{\sum_{k \neq i}^L \theta_k \alpha_k} + 1 \right) \right) \end{aligned}$$

同样由于上述偏导具有初等函数形式, 且  $\alpha_i \in (0, 1]$ , 该偏导在其定义域上是连续的。

对过程  $\alpha_i(x_1^t, \dots, x_{T_k}^t, \delta_1, \dots, \delta_{T_k})$  中对  $x_n^t$  的偏导为

$$\frac{d\alpha_i}{d\delta_n} = -(x_n - x_n^t)^2 e^{-\sum_{i=1}^{T_k} \delta_i (x_i - x_i^t)^2}$$

同样具有初等函数形式且  $x_n^t \in (-\infty, +\infty)$ , 该偏导在其定义域上是连续的。

由于路径  $\frac{d\text{loss}}{d\beta_i} \frac{d\beta_i}{d\bar{\beta}_j} \frac{d\bar{\beta}_j}{d\alpha_i} \frac{d\alpha_i}{dx_n^t}$  为上各部分中间偏导均存在且在各自定义域上连续, 根据连续函数四则运算法则, 其连乘得到路径偏导同样连续。由于任意路径偏导连续, 最终求和得到偏导表达式仍具有连续性质。由于置信规则库推理系统中激活与推理部分均由初等函数构成, 各参数不同路径的偏导都具有相似的连续性质, 可证明当选取合适损失函数时损失结果对推理系统各个参数是可微分的。

**NAME Name-Name**, Ph.D., asociate profesor. His/her research interests include \*\*\*, \*\*\*, and \*\*\*.

**NAME Name**, 英文作者介绍内容包括: 出生年, 学位 (或目前学历), 职称, 主要研究领域 (与中文作者介绍中的研究方向一致)。

**NAME Name-Name**, 英文作者介绍内容包括: 出生年, 学位 (或目前学历), 职称, 主要研究领域 (与中文作者介绍中的研究方向一致)。