

深圳大学学报(理工版)

*Journal of Shenzhen University(Science and Engineering)*

ISSN 1000-2618,CN 44-1401/N

## 《深圳大学学报(理工版)》网络首发论文

题目: 基于相关性学习的多任务模型及其应用  
作者: 徐薇, 骆剑平, 李霞, 曹文明  
收稿日期: 2022-01-04  
网络首发日期: 2023-06-07  
引用格式: 徐薇, 骆剑平, 李霞, 曹文明. 基于相关性学习的多任务模型及其应用[J/OL]. 深圳大学学报(理工版).  
<https://kns.cnki.net/kcms2/detail/44.1401.N.20230606.0910.004.html>



**网络首发:** 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

**出版确认:** 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

【电子与信息科学 / Electronics and Information Science】

# 基于相关性学习的多任务模型及其应用

徐薇, 骆剑平, 李霞, 曹文明

深圳大学电子与信息工程学院, 广东深圳 518060

**摘要：**多任务学习是一种联合多个任务同时学习来增强模型表示和泛化能力的手段，任务之间的相关性是多任务学习的基本因素。为解决任务差异的内在冲突会损害部分任务的预测的问题，提出一种基于相关性学习层(correlation learning layer, CLL)的多任务学习模型，并将该模型作为新的代理模型应用于贝叶斯优化算法中，以期解决昂贵的优化问题。在传统的多任务学习网络后面增加相关性学习层，使得已经完成初步共享学习的任务在该层进行优化共享，令多个任务所学习到的知识充分交互起来。根据不同的基于参数的共享机制，构建带有相关性层的LeNet和径向基(radial basis function, RBF)多任务学习模型，在多任务版本的美国国家标准与技术研究所数据集(Mixed National Institute of Standards and Technology database, MNIST)数据集和任务相关性可控制的综合数据集上进行实验，验证了所提出的基于相关性层的多任务学习模型的有效性。将所提多任务学习网络作为代理模型应用于贝叶斯优化算法中，不仅能减少模型对目标问题的评价次数，还能成倍地扩充训练数据数量，进而提升模型的性能。

**关键词：**神经网络；多任务学习；贝叶斯优化；深度学习；径向基函数；相关性学习层

**中图分类号：**TP181；TP193

**文献标志码：**A

## Correlation learning based multi-task model and its application

XU Wei, LUO Jianping, LI Xia, and CAO Wenming

College of Electronic and Information Engineering, Shenzhen University, Shenzhen 518060, Guangdong Province, P. R. China

**Abstract:** Multi-task learning is a means of combining multiple tasks simultaneously to enhance model representation and generalization ability. The correlation between tasks is the basic factor of multi-task learning. In order to solve the problem that the inherent conflict of task differences will damage the prediction of some tasks, a multi-task learning model based on the correlation learning layer (CLL) is proposed. Meanwhile, to solve the expensive optimization problem, the proposed multi-task learning model is applied as a new agent model to the Bayesian optimization algorithm. A correlation learning layer is added behind the traditional multi-task learning network to make the tasks that have completed the preliminary shared learning carry out advanced sharing in this layer, so that the knowledge learned by multiple tasks can fully interact with each other. According to different parameter-based sharing mechanisms, LeNet and radial basis function (RBF) multi-task learning models with correlation layers are constructed. According to different parameter-based sharing mechanisms, LeNet and RBF multi-task learning models with correlation layers are constructed. They are experimented on the multi-task version of the Mixed National Institute of Standards and Technology database (MNIST) data set and the comprehensive data

**Received:** 2022-01-04; **Revised:** 2023-02-13; **Accepted:** 2023-02-14

**Foundation:** National Natural Science Foundation of China (62176161); Shenzhen Science and Technology Program (JCYJ20220818100005011, JCYJ20200813144831001)

**Corresponding author:** Associate professor LUO Jianping. E-mail: ljp@szu.edu.cn

**Citation:** XU Wei, LUO Jianping, LI Xia, et al. Correlation learning based multi-task model and its application [J]. Journal of Shenzhen University Science and Engineering. (in Chinese)

<http://journal.szu.edu.cn>

set with controllable task correlation. The experimental results verify the effectiveness of the proposed multi-task learning model based on the correlation layer. Meanwhile, the proposed multi-task learning network as a proxy model is applied to the Bayesian optimization algorithm, which not only reduces the evaluation times of the model to the target problem, but also enlarges the number of training data exponentially, thus improving the accuracy of the model.

**Key words:** neural network; multi-task learning; Bayesian optimization; deep learning; radial basis function; correlation learning layer

多任务学习旨在联合多个任务同时学习,使任务之间彼此能相互帮助提升,是机器学习领域一个很有前景的研究方向.当多个任务之间具有相似特征时,每个任务之间能够共享有用信息,所以联合学习的性能比每个任务独立学习更好.人们根据不同的场景提出了不同的多任务学习算法,并在计算机视觉<sup>[1-3]</sup>、自然语言处理<sup>[4-6]</sup>、生物信息学<sup>[7-8]</sup>和语音<sup>[9-10]</sup>等领域都已取得了巨大成功.最近,有学者将目光汇集于基于深度学习的多任务学习上,他们中的一部分人研究如何通过使用更深入的学习体系结构来增强表示能力,从而提升预测精度<sup>[11]</sup>,另一部分人探索如何通过建模来确定任务关系<sup>[2, 12-13]</sup>,如十字绣结构通过学习输入激活映射的线性组合来建模共享表示,在网络的每一层,从两个任务中学习激活映射的线性组合,下一层的过滤器对这个共享表示进行操作.任务的相关性对多任务学习来说至关重要,直接影响模型性能.多任务学习中的各个任务之间往往既相关又独立,因此多任务学习的目的是利用多个任务中共同的有效信息来提升单个任务的学习.然而在现实中,许多基于深度神经网络的多任务学习模型对数据分布差异和任务之间的相关性程度等因素很敏感<sup>[13]</sup>,导致多任务学习模型并不是在所有任务上都能优于单任务模型<sup>[5, 14]</sup>.当模型参数在所有任务之间广泛共享时,任务差异的内在冲突会损害部分任务的预测.因此,在任务相关性程度低时,多任务模型的质量反而会受到消极影响.为解决此问题,本研究提出基于相关性学习层(correlation learning layer, CLL)的多任务模型,在传统多任务模型的隐藏层和输出层之间增加一个相关性层.此层将多个任务的输出通过一定形式进行组合,将多个任务所对应的网络之间学习到的知识更充分地交互起来,从而提升多任务模型的性能.本研究采用有监督学习任务,在分类实验和回归实验上对提出的算法做了有效性证明.

贝叶斯优化具有优化昂贵的黑盒无导数函数的能力, SNOEK 等<sup>[15]</sup>观察到贝叶斯优化对训练深度神经网络很有用,这引起了机器学习领域的兴趣.

贝叶斯优化算法基于目标函数的过去评估结果建立替代函数(概率模型),来找到最小化目标函数值.该算法与网格搜索<sup>[16]</sup>和随机搜索<sup>[17]</sup>的不同之处在于,它在尝试下一组超参数时,会参考之前的评估结果,不需评价真实函数,因此可省去很多无用功,性能更好.多任务贝叶斯优化采用带有任务间内核的多任务高斯过程来捕获辅助函数和目标函数之间的相关性<sup>[18-20]</sup>.然而,类似于标准的高斯过程,多任务高斯过程无法拓展到大型离线数据集.最近的工作提出将神经网络与概率模型相结合以提高可扩展性<sup>[21-24]</sup>.例如, MT-BOHAMIANN<sup>[21]</sup>使用贝叶斯神经网络<sup>[22]</sup>作为的多任务贝叶斯优化代理模型,解决了有限表达能力的问题.由于待优化函数的评估代价高昂,一个自然的想法就是使用概率模型近似替代原始函数的模型.概率代理模型根据模型参数数量是否固定可分为参数模型和非参数模型<sup>[23]</sup>.与参数模型相比较,非参数模型更灵活,可扩展性更高,且使用贝叶斯方法不易发生“过拟合”<sup>[24]</sup>,但是模型的参数随着数据量的增加而增加.高斯过程是一种常用的非参数模型,但高斯过程中协方差矩阵的计算量会随着数据量的增加变得非常大,因此,寻找新的深度神经网络来取代高斯过程模型也是目前非常热门的研究.本研究提出采用基于相关性学习层的多任务模型作为新的代理模型来解决昂贵优化问题.

## 1 基于相关性层的多任务学习模型

基于相关性层的多任务学习模型通过在传统的多任务学习网络框架引入一组相关性学习层实现在任务的相关性程度降低时减少任务对模型的影响.在初阶段的训练中,多个任务已经学习到了部分有效特征,而通过相关性学习层的再次训练,可让多个任务在优选出来的信息中学习到有价值的部分.

### 1.1 单任务学习和传统的多任务学习模型

现有的机器学习和深度学习大都围绕单个任务



学习进行, 因此每训练一个新任务, 就需更换一个新的网络. 但在自然界中, 人类的大脑往往是同时处理多个任务, 这千差万别的任务全靠一套神经系统, 因此多任务学习应运而生. 单任务学习即每个任务独自学习, 任务与任务之间、模型与模型之间无任何交互与联系. 单任务学习模型结构如图 1(a). 当有多个任务需要学习时, 需单独为每个任务设计新的网络模型来进行训练. 多任务学习则是联合多个任务一起学习, 所有任务在网络中信息共享并进行并行学习. 图 1(b)给出了一种基于硬共享的多任务学习模型结构

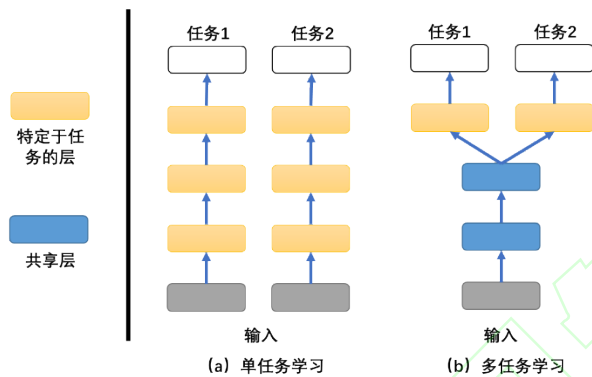


图 1 (a)单任务学习与(b)传统多任务学习模型

Fig. 1 (a) Single-task learning model and (b) traditional multi-task learning model.

## 1.2 基于相关性学习层(CLL)的多任务模型

为将多个任务所学习到的知识充分交互起来, 并提升多任务模型的鲁棒性和稳定性, 本研究在传统多任务学习模型的隐藏层和输出层之间增加了一组全连接层(可带激活层的多层网络), 如图 2 中红色部分所示, 此即相关性学习层(CLL). 多个任务在底层已经共享过一遍信息, 并学习到了部分有价值的特征, 此时再将多个任务在相关性层上进行进一步优化和共享. 与传统多任务学习模型结构相比, 此层将多个原本已经初步训练好的任务输出通过全连接的形式连接起来. 在这一层中, 本研究使用权值参数表示各任务之间的相关性, 即每个任务的信息共享程度. 通过此相关性层, 一个任务的最终输出可以表达为各个任务之间的相关信息加权和. 当单层全连接层不足以学到更复杂的任务相关性的任务时, 可适当增加全连接层的层数, 组成一个相关性模块, 学习到更多的特征信息.

本研究根据不同的基于参数的共享机制, 提出两个不同的基于相关性层的多任务学习模型: 一是基于 LeNet 的多任务学习模型, 记为 MultiLeNet+;

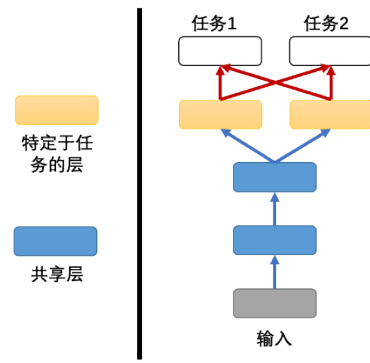


图 2 带相关性层的多任务学习模型

Fig. 2 Multi-task learning model with correlation layers.

二是基于径向基函数<sup>[26]</sup>(radial basis function, RBF)的多任务学习模型, 记为 MultiRBF.

### 1.2.1 MultiLeNet+模型

近几年来, 卷积神经网络在图像识别中取得了非常成功的应用, 而 LeNet5 在卷积神经网络的发展历程中有着里程碑的意义<sup>[25]</sup>. 本研究提出基于 LeNet 的带有相关性层的多任务学习模型 MultiLeNet+, 其基础的硬共享模型的体系结构如图 3. 该模型将卷积层和池化层(蓝色部分)作为硬共享模型的底部共享层, 后端将 3 层全连接层(黄色部分)作为任务对应层, 分别对应 2 个具体任务. 最后嵌入相关性层, 以便对已完成初步学习的任务中的有效信息进行再共享.

保持其他部分不变(即共享底层保留原来的网络结构), 对虚线方框内的部分进行修改, 提出特定于任务的层带有十字绣单元的多任务学习模型 MultiLeNetCrossStitch+ (记为 MultiLeNetCS+). 该模型的共享底层由两层卷积层和两层池化层组成. 在原始特定于任务的层中嵌入十字绣结<sup>[2]</sup>构, 将两组特定于任务的层通过十字绣单元缝合起来, 让机器自动学习并判断 3 层特定于任务的层之间是否需要信息共享, 然后将这部分的输出通过相关性层进行再学习. 十字绣单元由固定大小的可训练矩阵组成, 在两个任务之间共享信息, 其收集每个任务的输入, 通过线性变换生成各自任务的输出. 此模型为基于参数的硬共享结构和软共享结构相结合的一个特殊模型, 如图 4 所示. 为方便比较, 还构建了其他 3 个模型: 基于 LeNet 的单任务学习模型 SingleLeNet、不带相关性层的多任务学习模型 MultiLeNet 以及 MultiLeNetCS. 在后续实验章节, 将两个新提出的带相关性层的多任务学习模型与单任务学习模型及其不带相关性层的模型进行对比.

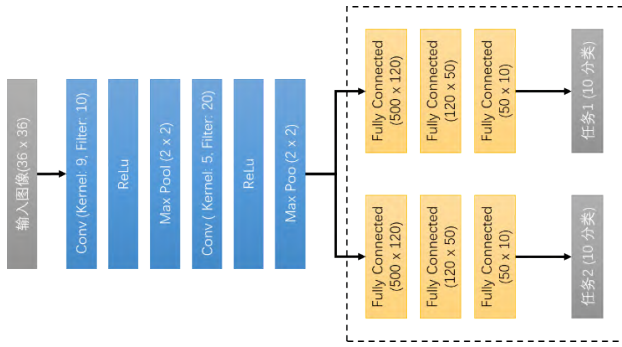


图3 MultiLeNet+模型的体系结构

Fig. 3 Multi-task learning model based on LeNet (MultiLeNet+).

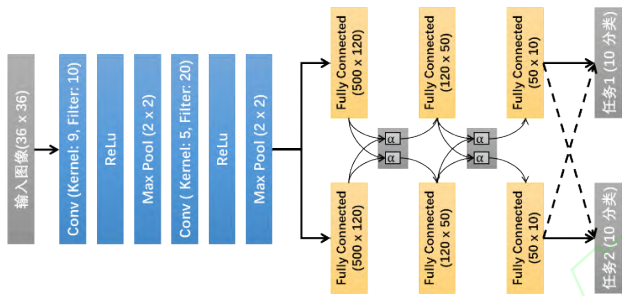


图4 带十字绣单元的多任务学习模型(MultiLeNetCS+)

Fig. 4 Multi-task learning model with cross-stitch units (MultiLeNetCS+).

### 1.2.2 MultiRBF 模型

为在多元的多任务模型中验证相关性层的有效性, 采用RBF<sup>[26]</sup>单独对应一个任务来构建多任务模型, 提出基于RBF的多任务学习网络(MultiRBF+), 如图5所示. 由图5可见, MultiRBF+模型是一个由输入层、隐藏层、相关性层和输出层构成的4层神经网络. 该模型的重点在于其末端增加了一层相关性层, 相关性层在该模型中为单层全连接层, 它将原本两个任务所对应的径向基神经网络的输出通过全连接层联系起来, 让这两个任务进行再学习. 由于整个MultiRBF+模型层数少, 且RBF具有良好的

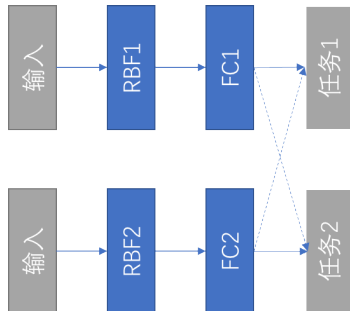


图5 基于径向基函数的多任务学习网络(MultiRBF+)

Fig. 5 Multi-task learning network based on RBF.

局部逼近特性, 因此该模型具有训练速度快和逼近性能好的优点. 在进行多任务联合学习时, 多个任务的分布大多数是非线性且复杂的, 为进一步增强该模型的非线性映射能力和表达能力, 在相关性层增加激活函数. 称此带有激活层的MultiRBF+模型为MultiRBF+ReLU模型. 此外, 本文在实验部分还对基于RBF的传统多任务学习模型MultiRBF进行比较.

## 2 基于多任务学习的贝叶斯优化算法

贝叶斯优化算法的核心由对目标函数进行建模和构造采集函数两部分组成. 前者用于计算每一个预测点的函数均值和方差, 后者用于决定本次迭代是在哪个最有潜力的点处进行采样. 由于径向基神经网络具有良好的局部逼近特性, 可逼近任意的非线性函数, 同时还具备计算能力高、泛化能力强的优点, 本研究提出一种带相关性层的基于RBF的多任务学习模型, 并替代高斯过程模型, 当作新代理模型嵌入到贝叶斯优化算法框架中, 最终构建一个基于RBF的多任务学习模型的贝叶斯优化框架Bay\_MRBFCor. 图6为任务数为2的Bay\_MRBFCor框架结构. 其中,  $x$ 为输入数据向量;  $\varphi_j(x)$ 为径向基函数输出;  $h_{ij}$ 为相关性层节点;  $y_1$ 和 $y_2$ 为2个任务的具体输出.

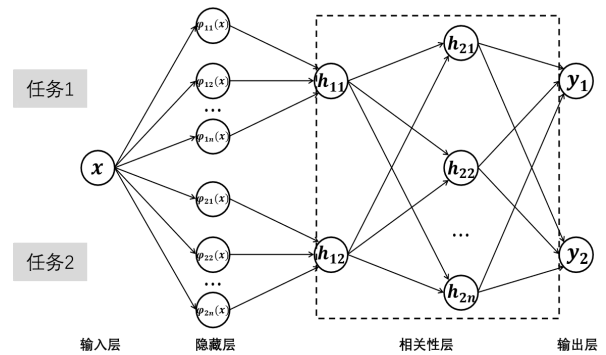


图6 任务数为2的基于RBF的多任务学习模型Bay\_MRBFCor

Fig. 6 A RBF-based multi-task model (Bay\_MRBFCor) when task number is 2.

由图6可见, Bay\_MRBFCor模型框架由多个径向基神经网络组合而成, 其网络结构有输入层、隐藏层、2层相关性层和输出层共5层. 相比传统的多任务学习模型将特定于任务层的输出直接作为任务的最终输出, Bay\_MRBFCor模型增加了两层相

关性层. 这两层将多个 RBF 网络的输出通过全连接形式进行再连接, 并且带有激活函数, 多个任务之间在相关性层共享优化信息, 让模型在相关性程度较低的任务之间也能学到更多有效信息, 从而帮助每个任务的学习都得到提升. 隐藏层中每个径向基神经网络对应一个单独的任务, 根据训练数据数量的改变, 可适当调整径向基神经网络中心点的数量. 每个径向基神经网络的输入是相同的, 由于每个任务都有自己需要学习的东西, 在网络中训练后, 每个任务都有特定的输出.

选择高斯核函数作为 RBF, 采用非线性优化的学习策略, 核函数的中心点的选择采用  $k$ -means 算法. 高斯核函数为

$$K(x - x_c) = e^{-\frac{\|x - x_c\|^2}{2\sigma^2}} \quad (1)$$

其中,  $\| \cdot \|$  为求模运算;  $x$  为空间中的任意一点;  $x_c$  为核函数的中心点;  $\sigma$  为核函数的宽度参数.

代理模型确定后, 候选点的选择采用标准粒子群算法与采集函数 EI (expectation of the improvement)<sup>[22]</sup> 策略相结合的方法, 即将采集函数 EI 策略作为粒子群算法的适应度函数, 而 EI 策略通过模型的预测值计算得到. 模型根据已有的训练数据, 在搜索空间中寻找下一个有潜力的候选点, 再不断地将候选点添加到初始训练集中, 从而实现模型更新. 可见, 使用粒子群算法寻找候选点时, 省去了对目标函数的真实评价步骤, 从而提升了模型的运行效率, 避免了在昂贵优化问题中因进行真实评价而产生的开销.

### 3 实验和结果分析

为证明基于相关性学习层 (CLL) 的多任务模型的有效性, 在由美国国家标准与技术研究所数据集 (Mixed National Institute of Standards and Technology database, MNIST) 数据集和 Fashion 数据集改编的多任务版本数据集上处理多标签分类问题, 然后在可控制任务之间相关性的合成数据集上进行误差分析实验, 这两个问题的任务数均设置为 2, 实验类型包含分类和回归. 又为证明多任务学习网络作为代理模型应用到贝叶斯优化模型中对解决优化问题具有有效性, 采用 5 组测试函数进行真实评价.

#### 3.1 基于多任务版本 MNIST 数据集的分类实验

MNIST 数据集是一个大型的手写体数字识别数据库, 常在机器学习的多个领域中被用作基准数据

集. FashionMNIST 数据集是在 MNIST 数据集上的改进, 整体数据结构上跟 MNIST 一致. 这两个数据集中均包含了 10 种类别的物品, 训练集有  $6 \times 10^4$  幅图像, 测试集有  $1 \times 10^4$  幅图片, 图像尺寸均为  $28 \times 28$  像素. 通常这两个数据集只能代表某个单独的具体任务, 于是本研究参考文献[27]方法, 采用图像叠加方式构建多任务版本的 MNIST 数据集, 将单任务分类转化为多任务的分类问题. 本研究从两个原始数据集中任意选取 2 张图像进行叠放, 则有 3 种组合, 因此可构建 3 个不同的多任务数据集 Multi-MNIST (原始数据皆来自 MNIST 数据集)、Multi-FashionMNIST (原始数据皆来自 FashionMNIST 数据集) 和 Multi-(Fashion+MNIST) (原始数据分别来自 MNIST 和 FashionMNIST 数据集) 数据集. 从原始数据集中随机地选取 2 幅不同的图像, 然后将这两幅图像一个放在左上角, 一个放在右下角, 合并成一个新的图像. 3 个数据集的构建方式区别在于原始数据集, 分别为 MNIST 数据集和 MNIST 数据集, MNIST 数据集和 FashionMNIST 数据集, 以及 MNIST 数据集和 Fashion+MNIST 数据集. 图 7 分别展示了 3 个多任务数据集的图像的示例. 由于原始数据集的两两组合很多, 在此次实验中, 每个多任务版本的数据集只选取  $120 \times 10^3$  张训练集图像,  $20 \times 10^3$  张测试集图像. 根据原始数据集及其组合方式, 多任务版本数据集中两个任务的相似性程度由高到低排列为: Multi-MNIST > Multi-FashionMNIST > Multi-(Fashion+MNIST). 对于每个数据集, 左上角的图像为任务 1, 右下角的图像为任务 2.

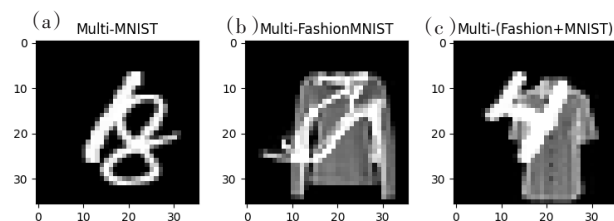


图 7 多任务版本 MNIST 数据集示例 (a) Multi-MNIST 数据集; (b) Multi-FashionMNIST 数据集; (c) Multi-(Fashion+MNIST) 数据集

Fig. 7 Multi-tasking version MNIST dataset. (a) Multi-MNIST, (b) Multi-FashionMNIST, and (c) Multi-(Fashion+MNIST).

实验使用 Pytorch 深度学习平台, 所有多任务学习模型在训练过程中使用的超参数设置一致. 基于 LeNet 的多任务学习模型均采用综合性能良好的 Adam 优化器来训练网络. 设置初始学习率 0.001, 在训练过程中, 模型的学习率会随着迭代次数的增



加进行衰减, 迭代次数每增加 15 次, 学习率减小 1/2. 使用衰减学习率的目的是使训练到达一定程度后, 减小步长以提高模型训练的精度.

使用 MultiLeNet、MultiLeNet+、MultiLeNetCS 和 MultiLeNetCS+4 种基于 LeNet 的多任务模型和基于 LeNet 的单任务模型对构建的多任务版本数据集

Multi-MNIST、Multi-FashionMNIST 和 Multi-(Fashion+MNIST)进行训练, 实验采集测试集的分类准确率, 实验结果如表 1. 表中的结果都是经过 10 次独立实验取的平均值, 括号里的数据是 10 次实验中取到的最大值, 粗体是最优值, 粗体加下划线为次优值, 下划线为次次优值.

表 1 不同多任务数据集中不同单任务和多任务分类模型的分类准确率<sup>1)</sup>

Table 1 Classification accuracy comparison

数据集	任务类别	分类模型				
		单任务 LeNet	MultiLeNet	MultiLeNet+	MultiLeNetCS	MultiLeNetCS+
Multi-MNIST	task1	0.913 2(0.915 9)	0.934 5(0.937 6)	0.938 6(0.942 7)	0.936 5(0.939 0)	0.937 3(0.942 3)
	task2	0.893 1(0.896 6)	0.919 2(0.921 7)	0.923 7(0.927 2)	0.920 4(0.923 3)	0.921 3(0.924 1)
Multi-FashionMNIST	task1	0.8154(0.818 1)	0.838 5(0.844 1)	0.842 3(0.846 1)	0.838 2(0.841 4)	0.840 0(0.844 9)
	task2	0.807 7(0.810 5)	0.830 4(0.835 7)	0.836 1(0.841 0)	0.834 6(0.838 0)	0.834 6(0.839 7)
Multi-(Fashion + MNIST)	task1	0.936 9(0.940 1)	0.959 0(0.962 3)	0.960 3(0.963 7)	0.957 7(0.960 8)	0.957 7(0.961 3)
	task2	0.844 6(0.848 5)	0.858 2(0.863 6)	0.859 8(0.870 6)	0.856 5(0.862 9)	0.856 6(0.859 7)

<sup>1)</sup> 灰底数据为该项目下的最优值

由表 1 可见: 基于 LeNet 的多任务学习模型在 3 个数据集上的分类准确率都大于单任务学习模型 SingleLeNet, 验证了多任务学习模型能让两个任务共享任务之间的相关性, 帮助网络模型学习对两个任务都有帮助的特征. 3 个数据库中, MultiLeNet+ 模型的准确率均高于 MultiLeNet 模型, 表明在传统的多任务学习模型中增加一层相关性层, 有助于模型性能的提升. 为进一步验证相关性层的有效性, 对 MultiLeNetCS+ 模型与 MultiLeNetCS 模型进行对比实验. 从表 1 可见, MultiLeNetCS 模型在 Multi-MNIST 数据集上的分类准确率高于 MultiLeNet 模型, 但在 Multi-FashionMNIST 数据集上, 仅任务 2 的分类准确率有所提升. 出现上述情况的原因可能是 MultiLeNetCS 模型在任务相关性低时自动学习能力低, 也可能是本研究将共享底层进行了固定, 未将十字绣单元的学习能力发挥到极致. 但实验仍可证明, 带有十字绣结构的 MultiLeNetCS 模型的性能优于单任务学习模型的性能. 而且增加相关性层后, MultiLeNetCS+ 模型分类准确率比 MultiLeNetCS 模型均有所提升, 表明模型性能均得到改善. 综上, 基于相关性学习层的多任务模型有效, 且在任务相关性程度较高时, 使用十字绣单元自动选择学习任务之间的有效信息还能进一步提升传统多任务学习模型的性能.

### 3.2 任务相关性的影响——综合数据实验

多任务学习模型的性能与任务之间的相关性信息

息相关, 为验证不同任务关系对模型质量的影响程度, 故在生成的任务相关性可控制的综合数据<sup>[28]</sup>上进行实验. 采用 MultiRBF+、MultiRBF 和 SingleRBF 模型进行对比实验. 此外, 由于相关性层可根据任务之间的复杂程度进行修改, 在 MultiRBF+ 模型的相关性层加上激活函数 LeakyReLU, 以提升模型的非线性映射能力, 将 MultiRBF+ReLU 模型纳入对比实验中进行性能比较.

实验分为 3 步进行: ① 设定任务维数, 给定任务相关性  $p$  分别为 1.0、0.9、0.5, 并分别生成一个多任务数据集; ② 在每个数据集上分别对不同模型进行训练, 保持所有模型的超参数一致; ③ 分别计算在各生成数据集上各模型的预测误差表现.

基于 RBF 神经网络的多任务学习模型均采用综合性能良好的 Adam 优化器来训练网络, RBF 中心点数设置为 50. 初始学习率均设置为 0.000 1, 并随着迭代次数的增加而进行衰减, 迭代次数达到  $1 \times 10^3$ 、 $3 \times 10^3$ 、 $7 \times 10^3$  次时, 学习率衰减 1/2. 为避免实验结果的随机性, 所有模型在相同实验设置下独立运行 20 次. 模型输出均为 20 次独立实验取平均值后的损失函数值, 实验结果如图 8.

由图 8 中可见, 在 3 种任务相关性程度下, MultiRBF+ 和 MultiRBF+ReLU 模型的损失均小于未加相关性学习层的 MultiRBF 模型, 说明多个具有一定相关性的任务联合学习时, 基于相关性学习层

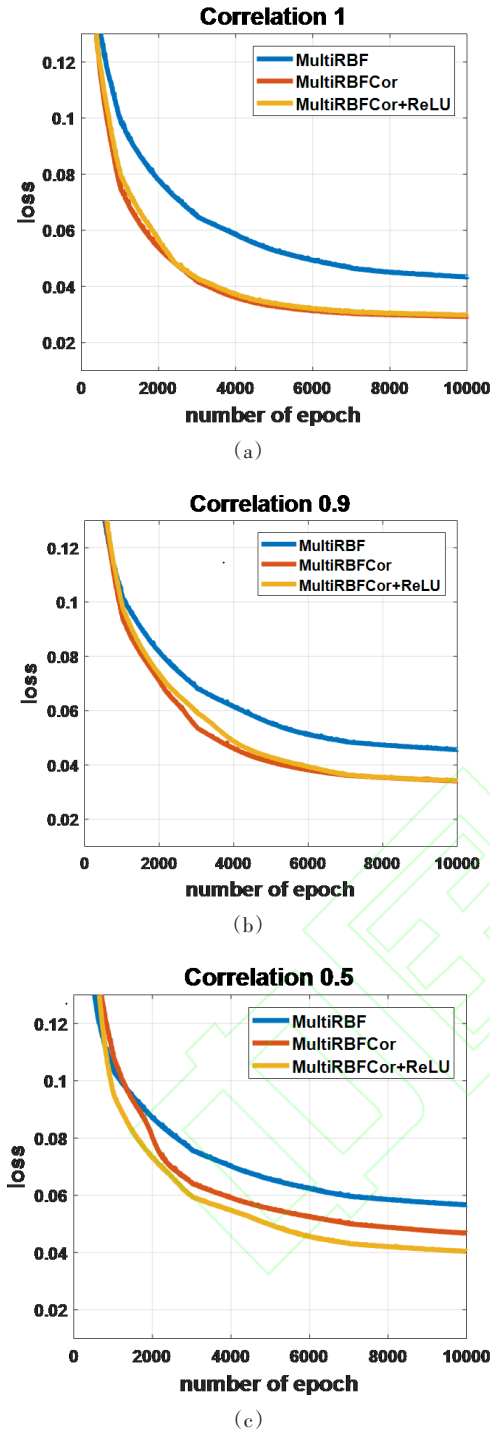


图8 不同模型在任务相关性分别为(a)1.0、(b)0.9和(c)0.5时的误差曲线

Fig. 8 Error curves of different models under task correlations of (a) 1.0, (b) 0.9, and (c) 0.5, respectively.

的多任务学习模型的学习能力高于传统的多任务学习模型. 在 $p$ 分别为1.0和0.9时, MultiRBF+模型和MultiRBF+ReLU模型的误差曲线接近, 但是当 $p$ 降为0.5时, MultiRBF+ReLU模型的损失比MultiRBF+模型更小, 性能更优. 可见, 多任务模型的

性能与任务之间的相关性呈正相关. 但是基于相关性学习层的多任务模型在 $p$ 降低至0.5时的性能还是优于普通多任务模型, 且激活函数对基于相关性层的多任务学习模型也有积极影响. 综上, 本研究提出的基于相关性学习层的多任务模型有效, 且在任务之间相关性程度降低时, 能减少无关离群任务信息对模型的影响.

### 3.3 基于多任务学习的贝叶斯优化算法实验

为探究贝叶斯优化框架 Bay\_MRBFCor 的多任务优化性能, 本研究对 Single-RBF、RMLA 和 Bay\_MRBFCor 模型在 $p$ 分别为1.0、0.9、0.5的情况下进行了多任务数值优化(求最优质)实验. 通俗来说, 贝叶斯优化实际上就是一个求极值的过程, 其目的是找到函数的最优值. 由于现实生活中, 待优化的问题往往很昂贵, 本研究选取 Ackley、Rastrigin、Schwefel 和 Griewangk 等4个广泛被使用的测试函数作为待优化的目标函数, 来模拟昂贵的优化问题. 为生成相关性可控制的两组任务, 从5种测试函数中选择两个测试函数进行组合, 形成新的5种测试函数. 例如, 将测试函数 Ackley 和 Rastrigin 组合的具体步骤为: 首先根据测试函数 Ackley 和 Rastrigin 的搜索空间, 分别生成20个二维的向量 $\mathbf{x}_0$ 和 $\mathbf{x}_1$ , 然后将 $\mathbf{x}_0$ 和 $\mathbf{x}_1$ 组合, 得到四维输入数据 $\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1)$ . 将 $\mathbf{x}_0$ 和 $\mathbf{x}_1$ 分别代入对应的测试函数, 计算得到 $f_A(\mathbf{x}_0)$ 和 $f_R(\mathbf{x}_1)$ 的值并归一化到 $[-1, 1]$ , 再组合为二维输入 $\mathbf{x}' = (f_A(\mathbf{x}_0), f_R(\mathbf{x}_1))$ , 最后将这组二维输入代入任务相关性公式中<sup>[28]</sup>, 生成如式(2)的任务相关性可控制的任务 $y_1$ 和任务 $y_2$ ,

$$\begin{cases} y_1 = \mathbf{w}_1^T \mathbf{f}(\mathbf{x}) + \sum_{i=1}^S \sin(\alpha_i \mathbf{w}_1^T \mathbf{f}(\mathbf{x}) + \beta_i) + \varepsilon_1 \\ y_2 = \mathbf{w}_2^T \mathbf{f}(\mathbf{x}) + \sum_{i=1}^S \sin(\alpha_i \mathbf{w}_2^T \mathbf{f}(\mathbf{x}) + \beta_i) + \varepsilon_2 \end{cases} \quad (2)$$

$$\mathbf{w}_1 = c \mathbf{u}_1, \mathbf{w}_2 = c(p \mathbf{u}_1 + \sqrt{(1-p)^2} \mathbf{u}_2) \quad (3)$$

式(3)中,  $c$ 为尺度常数;  $\mathbf{u}_1$ 和 $\mathbf{u}_2$ 为两个正交单位向量;  $p$ 为相关性系数. 式(2)中,  $\alpha_i$ 和 $\beta_i$ 是给定的控制正弦函数形状的参数;  $\varepsilon$ 表示均值为0、标准差为0.1的高斯随机噪声.

由于式(2)存在随机数, 故除了噪声外, 本研究在算法训练和预测的过程中将其他随机参数对应固定. 本研究随机生成了5种组合, 分别为 Ackley+Rastrigin 组合、Rastrigin+Griewangk 组合、Rastrigin+Schwefel 组合、Ackley+Griewangk 组合和 Ackley+Schwefel 组合. 本研究将分别设置 $p$ 为1.0、



0.9 和 0.5.

为评估贝叶斯优化算法,将改进的正则化多任务学习模型(regularized multi-task learning model, RMLA)<sup>[29]</sup>作为对比模型. RMLA 的参数被设置同[29],同时假设所有任务的模型参数彼此接近,因此去除相关层,且在损失函数中加入了正则化的惩罚项. 表 2 列出了在输入数据维度为 4, Single-RBF、RMLA 和 Bay\_MRBFCor 模型在 5 个合成后的测试函数上的实验结果. 本次实验中值得注意的问题是,在模型的迭代次数相同时,采用单径向基函数作为代理模型能寻找到与迭代次数相同的候选点

个数. 而采用本研究提出的多任务模型作为代理模型时,两个任务寻找到的候选点个数是迭代次数的 2 倍,故而其进行真实评价的次数为单任务模型作为代理模型时的 2 倍. 若要保证多任务学习模型作为代理模型时,其评价次数与单任务模型作为代理模型时相同,该模型在进行迭代时,迭代次数需减少 1/2. 为增加所提贝叶斯优化算法的可靠性验证,考察算法在不同维度和不同任务关系设定的对比实验,详情请扫描论文末页右下角二维码见补充材料.

为避免实验的随机性,实验的真实评价次数设

表2 各方法在 5 个测试问题上的优化结果(求每个任务最小值)

Table 2 Optimization results of each method on five test settings

测试函数	$p$	任务	Single-RBF	RMLA	Bay_MRBFCor
Ackley+Rastrigin	1.0	1	-7.327 5(1.518 6)	-7.8025(2.001 6)	-8.204 3(1.559 6)
		2	-7.181 5(0.888 7)	-7.7684(2.074 3)	-8.191 6(1.503 7)
	0.9	1	-7.132 6(1.155 9)	-7.2200(1.268 8)	-7.601 6(1.629 4)
		2	-18.6561(0.462 1)	-18.8831(0.559 6)	-19.017 6(0.702 0)
	0.5	1	-7.2751(1.624 3)	-6.1069(0.936 4)	-7.999 0(2.003 8)
		2	-20.5251(1.151 5)	-20.404 8(0.867 6)	-21.577 7(1.230 8)
Rastrigin+Griewangk	1.0	1	-27.6798(6.310 7)	-35.874 0(2.311 5)	-36.130 3(5.284 1)
		2	-27.926 4(6.300 0)	-35.820 9(2.329 3)	-36.150 4(5.356 5)
	0.9	1	-28.704 5(7.199 9)	-38.153 6(1.525 2)	-38.684 5(1.430 2)
		2	-44.867 2(8.038 9)	-50.384 4(0.575 0)	-50.890 8(0.356 2)
	0.5	1	-27.582 5(4.903 1)	-38.198 6(0.935 5)	-38.366 3(1.486 9)
		2	-57.051 5(4.718 6)	-58.737 0(0.573 3)	-59.402 4(0.194 7)
Rastrigin+Schwefel	1.0	1	-288.782 4(45.152 9)	-289.640 7(37.289 4)	-294.085 2(32.350 2)
		2	-273.381 5(34.381 5)	-289.649 9(37.280 1)	-294.167 4(32.298 4)
	0.9	1	-287.149 2(39.339 1)	-289.516 0(35.523 9)	-293.165 1(33.475 2)
		2	-376.286 6(40.221 4)	-386.546 2(49.072 2)	-397.3423(45.0698)
	0.5	1	-249.137 2(13.585 8)	-270.475 1(40.151 2)	-271.3577(32.9476)
		2	-410.581 8(28.943 2)	-383.282 7(57.256 6)	-381.409 1(46.784 0)
ackley+griewangk	1.0	1	-36.053 1(3.115 1)	-36.234 1(4.427 6)	-37.869 0(1.912 0)
		2	-34.614 3(3.914 8)	-36.261 0(4.436 7)	-37.813 5(1.881 7)
	0.9	1	-34.227 7(5.383 7)	-36.478 5(6.203 2)	-37.709 5(1.329 0)
		2	-50.476 9(0.335 3)	-45.658 0(7.288 4)	-50.740 6(0.234 1)
	0.5	1	-33.694 7(6.155 6)	-35.477 3(4.524 5)	-37.209 9(1.832 2)
		2	-53.876 1(2.848 8)	-54.310 7(3.951 8)	-55.826 8(0.187 2)
Ackley+Schwefel	1.0	1	-304.505 0(49.997 4)	-308.309 9(17.625 6)	-310.325 2(28.089 7)
		2	-307.612 6(41.224 6)	-307.215 2(17.647 7)	-310.272 6(28.085 9)
	0.9	1	-306.764 2(29.458 8)	-305.508 8(9.157 2)	-320.756 0(26.545 9)
		2	-392.976 1(59.446 5)	-393.908 7(13.091 2)	-430.434 3(32.919 0)
	0.5	1	-295.482 1(27.234 9)	-293.353 8(15.637 8)	-304.910 9(27.923 0)
		2	-378.029 4(40.519 6)	-409.486 0(20.265 0)	-424.796 0(36.945 9)

<sup>1)</sup> 灰底数据为该项目下的最优值

置为100次,所有算法的迭代次数都设置为100次,表格中的结果都是根据10次实验的结果统计出的均值和标准差,以方便性能比较.括号中的数据为标准差,粗体数据为最优值,带有下划线的数据为次优值.

由表2可见,本研究提出的贝叶斯优化算法在29组对比实验中找到的最小值均优于其他算法找到的最小值,且实验结果的标准差远小于平均值.只有在Rastrigin+Schwefel组合的测试函数上,当 $p=0.5$ 时,基于SingleRBF的贝叶斯优化算法在任务2上找到的最小值更小,原因可能是该函数组合更复杂,全局最优与局部的最优值之间存在很多干扰项,任务之间共享的知识有限,当优化陷入局部最小值时难度更大.这表明所提出贝叶斯优化算法在5组复杂的测试函数中都有效.单独观察每个待优化的测试函数,当 $p$ 从1.0降至0.5时,本研究提出的贝叶斯优化算法依旧能保持良好的性能,表明所提贝叶斯优化算法在任务之间的相关性程度降低时具有一定的稳定性.然而,基于RMLA模型的贝叶斯优化算法只有24组实验结果优于基于SingleRBF模型的贝叶斯优化算法,当 $p$ 降低时,算法性能亦有所下降.

因此,本研究提出的基于多Bay\_MRBFCor模型的贝叶斯优化算法在5组复杂的且相关性程度可控制的测试问题中具有有效性和稳定性,在真实评价次数一致的情况下具有一定的优势.当任务之间的相关性程度降低时,还能减缓模型性能的降低,这表明Bay\_MRBFCor算法框架是有效的.本研究算法能够同时对多个任务进行优化,任务之间能够共享有效的优化信息,从而为每个任务找到更好的优化方案.此模型可应用在多个相关任务现有的数据集有相同的输入变量以及不同的目标函数时,此外该模型还能应用在初始训练数据少的多任务场景中,它通过每次迭代能够快速扩充数据集.

## 结 语

提出一种基于相关性学习层(CLL)的多任务网络,即在传统的多任务学习网络后嵌入一组相关性学习层.让已经完成初步学习的多个任务在相关性层进行再次训练,使得多任务学习网络更有效地利用多个任务之间的相关性.这在一定程度上解决了多任务学习网络在任务之间的相关性程度降低时,

网络性能也有所降低的问题.此外,将提出的多任务学习网络作为代理模型应用于贝叶斯优化算法中,不仅减少了模型对目标问题的评价次数,还能成倍地扩充训练数据数量,从而提升模型的准确度,并将针对单任务的优化问题拓展至针对多任务的优化问题.

**基金项目:** 国家自然科学基金资助项目(62176161); 深圳市基础研究资助项目(JCYJ20220818100005011, JCYJ20200813144831001)

**作者简介:** 徐薇(1997—), 深圳大学硕士研究生. 研究方向: 机器学习. E-mail: 745436009@qq.com

**引 文:** 徐薇, 骆剑平, 李霞, 等. 基于相关性学习的多任务模型及其应用[J]. 深圳大学学报理工版.

## 参考文献 / References:

- [1] ZHANG Zhanpeng, LUO Ping, LOY C C, et al. Facial landmark detection by deep multi-task learning [C]// Computer Vision-ECCV 2014. Cham: Springer International Publishing, 2014: 94-108.
- [2] MISRA I, SHRIVASTAVA A, GUPTA A, et al. Cross-stitch networks for multi-task learning [C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, USA: IEEE, 2016: 3994-4003.
- [3] ZAMIR A R, SAX A, SHEN W, et al. Taskonomy: disentangling task transfer learning [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018: 3712-3722.
- [4] COLLOBERT R, WESTON J, BOTTOU L, et al. Natural language processing (almost) from scratch [J]. Journal of Machine Learning Research, 2011, 12: 2493-2537.
- [5] LUONG M T, LE Q V, SUTSKEVER I, et al. Multi-task Sequence to Sequence Learning [J]. Computer Science, 2015.
- [6] LIU Pengfei, QIU Xipeng, HUANG Xuanjing. Recurrent neural network for text classification with multi-task learning [EB/OL]. (2016-05-17) <https://arxiv.org/abs/1605.05101>.
- [7] ZHOU Jiyun, LU Qin, GUI Lin, et al. MTFsite: cross-cell type TF binding site prediction by using multi-task learning [J]. Bioinformatics, 2019, 35(24): 5067-5077.
- [8] HAN Lei, LI Lei, WEN Feng, et al. Graph-guided multi-task sparse learning model: a method for identifying antigenic variants of influenza A (H3N2) virus [J]. Bioinformatics, 2019, 35(1): 77-87.
- [9] BELL P, RENALS S. Regularization of context-dependent deep neural networks with context-independent multi-task training [C]// IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway,

- USA: IEEE, 2015: 4290-4294.
- [10] CHEN Zhuo, WATANABE S, ERDOGAN H, et al. Speech enhancement and recognition using multi-task learning of long short-term memory recurrent neural networks [C]// *Proceeding in Interspeech 2015*. Baixas, France: ISCA, 2015: 3274-3278.
- [11] CAO Jiajiong, LI Yingming, ZHANG Zhongfei. Partially shared multi-task convolutional neural network with local constraint for face attribute learning [C]// *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, USA: IEEE, 2018: 4290-4299.
- [12] DUONG L, COHN T, BIRD S, et al. Low resource dependency parsing: cross-lingual parameter sharing in a neural network parser [C]// *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. Stroudsburg, USA: ACL, 2015: 845-850.
- [13] ZHAO Jiejie, DU Bowen, SUN Leilei, et al. Multiple relational attention network for multi-task learning [C]// *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, USA: Association for Computing Machinery, 2019: 1123-1131.
- [14] KAISER L, GOMEZ A N, SHAZEER N, et al. One model to learn them all [EB/OL]. (2017-06-16). <https://arxiv.org/abs/1706.05137>.
- [15] SNOEK J, LAROCHELLE H, ADAMS R P. Practical bayesian optimization of machine learning algorithms [C]// *Proceedings of the 25th International Conference on Neural Information Processing Systems-Volume 2*. Red Hook, USA: Curran Associates Inc., 2012: 2951-2959.
- [16] BARDENET R, KÉGL B. Surrogating the surrogate: accelerating gaussian-process-based global optimization with a mixture cross-entropy algorithm [C]// *Proceedings of the 27th International Conference on International Conference on Machine Learning*. Madison, USA: Omnipress, 2010: 55-62.
- [17] BERGSTRA J, BENGIO Y. Random search for hyperparameter optimization [J]. *Journal of Machine Learning Research*, 2012, 13(1): 281-305.
- [18] SWERSKY K, SNOEK J, ADAMS R P. Multi-task bayesian optimization [C]// *Proceedings of the 26th International Conference on Neural Information Processing Systems-Volume 2*. Red Hook, USA: Curran Associates Inc., 2013: 2004-2012.
- [19] BONILLA E V, CHAI K M A, WILLIAMS C K I. Multi-task gaussian process prediction [C]// *Proceedings of the 20th International Conference on Neural Information Processing Systems*. Red Hook USA: Curran Associates Inc., 2007: 153-160.
- [20] POLOCZEK M, WANG Jialei, FRAZIER P I. Warm starting Bayesian optimization [C]// *Winter Simulation Conference (WSC)*. Piscataway, USA: IEEE, 2016: 770-781.
- [21] SPRINGENBERG J T, KLEIN A, FALKNER S, et al. Bayesian optimization with robust Bayesian neural networks [C]// *Proceedings of the 30th International Conference on Neural Information Processing Systems*. Red, USA: Curran Associates Inc., 2016: 4141-4149.
- [22] NEAL R M. Bayesian learning for neural networks [M]. [S. l.]: Springer Science & Business Media, 2012.
- [23] 崔佳旭, 杨博. 贝叶斯优化方法和应用综述[J]. *软件学报*, 2018, 29(10): 3068-3090.  
CUI Jiaxu, YANG Bo. Survey on Bayesian optimization methodology and applications [J]. *Journal of Software*, 2018, 29(10): 3068-3090. (in Chinese)
- [24] GHAHRAMANI Z. Probabilistic machine learning and artificial intelligence [J]. *Nature*, 2015, 521(7553): 452-459.
- [25] EL-SAWY A, EL-BAKRY H, LOEY M. CNN for handwritten Arabic digits recognition based on LeNet-5 [C]// *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016*. Cham: Springer International Publishing, 2017: 566-575.
- [26] BROOMHEAD D S, LOWE D. Radial basis functions, multi-variable functional interpolation and adaptive networks [EB/OL]. <https://www.researchgate.net/publication/233783084>.
- [27] LIN Xi, ZHEN Huiling, LI Zhenhua, et al. Pareto multi-task learning [C]// *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Red Hook, USA: Curran Associates Inc., 2019: 12060-12070.
- [28] MA Jiaqi, ZHAO Zhe, YI Xinyang, et al. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts [C]// *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York, USA: Association for Computing Machinery, 2018: 1930-1939.
- [29] EVGENIOU T, PONTIL M. Regularized multi-task learning [C]// *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, USA: Association for Computing Machinery, 2004: 109-117.

【中文责编：英子】