

基于元学习思想的算法选择问题综述

曾子林^{1,2}, 张宏军¹, 张睿¹, 王之腾¹

(1. 解放军理工大学指挥信息系统学院, 南京 210007; 2. 南昌陆军学院, 南昌 330103)

摘要: 算法选择实际上可视为一种学习任务. 鉴于此, 首先分析基于元学习思想的算法选择框架; 然后从数据集特征和元算法两个角度对基于元学习思想的算法选择方法进行归纳总结; 最后分析基于元学习思想的算法选择存在的问题, 并指出未来发展方向.

关键词: 算法选择; 元学习; 数据集特征

中图分类号: TP181

文献标志码: A

Summary of algorithm selection problem based on meta-learning

ZENG Zi-lin^{1,2}, ZHANG Hong-jun¹, ZHANG Rui¹, WANG Zhi-teng¹

(1. Institute of Command and Information System, PLA University of Science Technology, Nanjing 210007, China; 2. Nanchang Military Academy, Nanchang 330103, China. Correspondent: ZENG Zi-lin, E-mail: zzljxnu@163.com)

Abstract: The algorithm selection problem can be considered as a learning task. Therefore, the framework of algorithm selection based on meta-learning is analyzed firstly. Then the algorithm selection based on meta-learning is classified and summarized from the viewpoint of characteristics of data set and meta-algorithm. Finally, the problems of algorithm selection based on meta-learning are analyzed, and the develop directions are proposed in the future.

Key words: algorithm selection; meta-learning; characteristics of data set

0 引言

近30年来, 随着机器学习的迅猛发展, 人们提出了许多不同类型的机器学习算法, 主要包括决策树^[1]、神经网络^[2]、基于实例的学习^[3]、基于规则的学习^[4]、基于统计理论的学习^[5-6]等算法. 由 Wolpert 等^[7]提出的“没有免费的午餐”(NFL)定理, 如果对具体学习任务的本质不作任何先验假设, 则没有任何理由认定某种学习算法比另外一种更好; 如果某种算法对某个特定的问题看上去比另外一种算法更好, 则仅仅是它更适合这一特定的学习任务, 而并非泛泛地说该算法是“优越”的^[8]. 由于不同的学习算法都有不同的归纳偏置^[9], 在处理不同特征问题时也有不同的学习性能, 如何从大量的学习算法中选择适合目前学习任务的算法是各行业、各领域迫切需要解决的问题.

本文首先分析了基于元学习思想的算法选择框架; 然后从数据集特征和元算法两个角度对基于元学习思想的算法选择方法进行归纳总结; 最后指出目前基于元学习思想的算法选择存在的问题及未来发展

方向.

1 算法选择

算法选择主要解决这样一类问题: 对于现存的一些可用算法, 哪一个算法在特定的问题领域中执行得最好? Rice^[10]于1976年提出一个正式的算法选择抽象模型, 如图1所示. 该抽象模型由4部分构成: 1) 问题空间 P 表示问题实例的集合; 2) 特征空间 F 表示对问题实例 P 进行特征提取得到的可测特征; 3) 算法空间 A 表示适用于问题实例 P 的算法集合; 4) 性能空间 Y 表示每个算法到其性能测度集的映射.

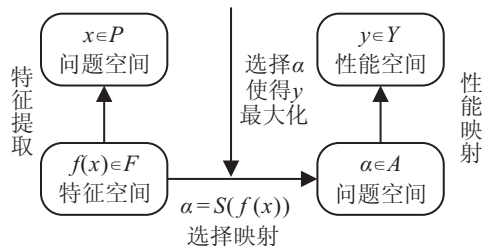


图1 Rice的算法选择抽象模型

收稿日期: 2013-09-18; 修回日期: 2014-02-05.

基金项目: 国家自然科学基金项目(70791137).

作者简介: 曾子林(1981—), 女, 讲师, 博士生, 从事元学习、特征选择的研究; 张宏军(1963—), 男, 教授, 博士生导师, 从事机器学习、效能评估、系统仿真等研究.

Rice 提出的算法选择问题抽象模型考虑了用一个通用的符号和模型来描述不同领域中不同任务的最优算法选择问题, 算法选择问题也可形式化表述如下: 对于给定问题实例 $x \in P$, 其特征为 $f(x) \in F$, 找到选择映射 $S: f(x) \rightarrow A$, 使得所选择的算法 $\alpha \in A$ 最大化性能映射。

算法选择最直观的方法是通过反复实验选取性能最优的算法, 如 Schaffer^[11] 提出在每个学习任务上依次执行所有可供选择的算法, 然后使用交叉验证的方法对准确率进行估计, 选取得分最高的算法。该方法虽然简单、直观, 但弊端是对计算资源要求过高, 而且相当耗时。另一种方法是通过专家经验^[12] 对算法选择进行指导, 如 Brodley^[13] 提出的算法选择系统 (MCS)。该系统以已经存在的专家知识为基础, 通过规则的形式表示出专家知识, 然后根据规则从一些候选模型和算法中选择适当的模型和算法。基于专家经验的算法选择方法存在两方面缺陷: 一是获取专家经验代价较高, 且专家经验不一定可靠; 二是开发的专家系统缺乏可扩展性, 即它不能对新出现的模型或算法进行处理, 从而降低了系统的灵活性。

从算法选择问题的形式化到 NFL 定理, 人们一直承认不存在某种最好的算法能够处理所有问题。为了选择适合给定问题的算法, 往往需要理解更多所研究问题的特征, 同时也要考虑对于某个特定问题领域, 什么特征可能会与算法性能相关、能不能对这些特征与算法性能的关系进行建模等一系列问题。在 20 世纪 90 年代初期, 有学者意识到算法选择实际上是一种学习任务, 因此算法选择在机器学习学科中逐渐发展, 形成了元学习领域^[14-15]。关于元学习的定义有很多种, 但简而言之, 元学习就是关于学习的学习, 其最终目的是寻求挖掘关于学习算法性能的元知识以提高学习算法的性能。元学习思想的雏形最早出现在文献 [16] 中, 该文献提出一种刻画分类问题的特征, 并验证了这些特征对算法行为产生的影响。Aha^[17] 对该思想进行了扩展, 用基于规则的学习算法产生如下形式的规则: 若给定的数据集具有特征 C_1, C_2, \dots, C_n , 则采用算法 A_1 , 而不采用算法 A_2 。

1991~1994 年, 欧洲开发了大型分类算法比较项目——STATLOG (comparative testing of statistical and logical learning)^[18-19], 元学习思想作为 STATLOG 项目的副产品开始引起研究者的注意。STATLOG 项目的主要目标是研究不同类型分类算法在不同问题上的性能, 并通过研究算法性能与数据集特征的联系获取关于算法选择的知识。STATLOG 的研究结果表明, 不存在任何一个算法能够解决所有分类问题, 这

与 NFL 定理不谋而合。在 STATLOG 项目的研究基础上, 欧洲一些国家先后开展了其他基于元学习思想的算法选择研究项目, 如 METAL 项目和 MiningMart 项目。元学习思想在算法选择问题上的应用已经引起越来越多机器学习领域学者的关注, 2001 年的知识发现会议上一个重要研究问题就是“依赖数据自动地选择数据挖掘参数和算法”, 该议题在 2006 年的知识发现会议上又被 Fogelman 重新提起。

2 基于元学习思想的算法选择框架

“元学习”最早出现在心理学学科中, 1992 年, Aha 首次将“元学习”思想用于机器学习领域。元学习是指应用机器学习的方法 (元算法) 寻求问题的特征 (元特征) 与算法相对性能测度间的映射, 从而形成元知识的学习过程^[20]。事实上, 通过对现有关于算法选择和元学习的研究, 可以提出一个统一的基于元学习思想的算法选择框架如图 2 所示, 该框架也可以应用于多个学科领域, 包括计算机科学、人工智能、运筹学、统计学、机器学习、生物信息学等。

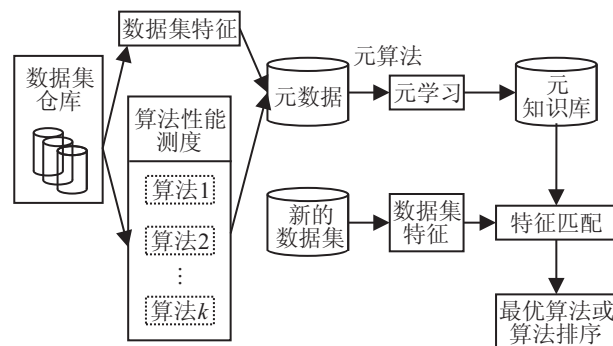


图 2 基于元学习思想的算法选择框架

算法选择任务可以视为一种学习任务。在初始阶段, 每个候选算法应用到各个数据集, 评估其性能, 并提取这些数据集的形态特征 (元特征), 形成元数据; 然后采用适当的元算法对元数据进行元学习, 以获得数据集特征与算法性能间的映射关系, 形成元知识库。算法的选择需要依靠元知识的指导, 当新的数据集加入时, 可先对该数据集进行特征提取, 然后根据与元知识库的特征匹配程度, 选择最优的算法或对算法进行排序。

由图 2 可见, 基于元学习思想的算法选择框架有以下关键技术需要研究: 数据集的选取、数据集特征的提取、算法性能测度的选取和元算法的确定。目前基于元学习思想的算法选择研究主要集中于对数据集特征的刻画和元算法的确定, 因此, 本文分别从数据集特征和元算法两个角度对基于元学习思想的算法选择方法进行分类。

3 基于元特征类型的分类

基于元思想的算法选择问题的一个重要任务是提取数据集的特征, 高质量的数据集特征(也称元特征)^[21-22]能为算法性能的差别提供合理解释. 具有代表性的元特征可以分为 3 类: 基于统计和信息论的元特征、基于基准分类器的元特征、基于模型的元特征.

3.1 基于统计和信息论的元特征

基于统计和信息论的元特征主要有 3 类: 数据集的简单特征、统计特征和信息特征. 数据集的简单特征描述数据集的简单结构, 包括数据集大小、属性类型、属性个数等; 数据集的统计特征主要反映数据的中心趋势和属性的离散程度, 如几何平均、调和平均、方差、偏斜度、峰度等; 信息特征反映不同属性间的关联程度和数据的一致性, 如属性变量的信息熵、平均互信息、噪声率等.

早期, Rendell 等提出利用与类的大小和集中程度相关的特征刻画分类问题, Aha 对 Rendell 等提出的数据集特征进行了扩展, 考虑的数据集特征包括实例的个数、类的种类、每类的个数、相关和不相关属性的个数、实例的分布范围等. 随后, STATLOG 项目又将描述数据集的特征扩充到 16 个, 如今这些数据集特征仍被元学习的研究者们广泛使用^[23]. 1999 年, Lindner 等^[24]将元特征的范围扩展到与区分能力相关的统计分布检验和异常点检验, 并研制了计算数据集特征的工具——DCT(data characterization tool).

上述元特征都只是采用各属性特征的均值来刻画整个数据集的特征, 容易造成数据集特征信息的丢失, 为了克服这个缺陷, Kalousis 等^[25-26]提出一种新方法——NOEMON, NOEMON 的数据集特征侧重于描述属性间的关联性, 主要通过直方图的值代替均值刻画数据集的形态特征, 直方图形式的元特征更能反映元特征的分布情况, 但数目比均值形式多, 文

献[25]使用的元特征为 45 个. 另外, Song 等^[27-28]基于频率对数据集特征进行提取, 从而得到数据集的特征向量. 一些常见的基于统计和信息论的元特征如表 1 所示.

目前, 大多数文献^[29-30]对数据集特征进行提取时, 采用的均是基于统计^[31-32]和信息论类型的元特征, 而且使用的多是元特征的平均值. 该类型的元特征有两点不足之处: 一是有的特征测度计算复杂度较高; 二是各属性特征的平均值不足以反映数据集的整体特征.

3.2 基于基准思想的元特征

在研究 STATLOG 项目的过程中, 研究者发现计算数据集特征所耗费的时间甚至比直接运行某个简单算法耗费的时间还多. 这个问题引起人们的思考, 能否仅用简单算法的性能来学习预测不同算法的性能呢? 该问题的提出逐渐形成了后来的基准思想, 这也是元学习领域中另一个重要的发展方向. Landmarker 方法^[33]的主要思想是: 给定两个简单的分类器 A 和 B, 看作是基准分类器, 若分类器 A 在某项给定的学习任务中执行得比 B 好, 则在执行类似挖掘任务时, 选择与 A 有相同归纳偏置的分类器.

基于基准思想的元特征是指通过基准算法的性能(如准确率)或候选算法在基准数据集上的性能刻画数据集特征. 基准算法^[34]是指算法的简单版本形式, 如决策树桩在机理上与决策树算法十分接近, 但在结构上却比决策树算法简单, 因此可看作是决策树算法的基准. 文献[33-34]即是以基准算法的性能作为元特征来估计算法的相对性能. 但该方法必须保证所选的基准算法有完全不同的归纳偏置, 因此基准算法的选择是使用 Landmarker 方法的难点. 常用的基准算法包括单层决策树桩、朴素贝叶斯、1-近邻和线性判别分类器等.

表 1 基于统计和信息论的元特征

简单特征	统计特征		信息特征
样本的大小	几何均值	变异系数	类的熵
属性的个数	调和均值	典型相关	属性的平均熵
类的数目	算术平均	偏斜度	属性与类的平均互信息
正例百分比	中位数	峰度	类的熵与平均互信息之比
负例百分比	众数	重心	噪声率
离散变量百分比	截尾均值	卡方检验值	
连续变量百分比	四分位数间距	离差指数	
缺失值百分比	最小绝对离差	相关系数	
	极值	正态分布检验值	
	标准差	异常点检验	
	方差		

文献[35-36]沿着基准思想的思路,提出了以算法在数据集的简单版本上的性能作为元特征来预测算法的性能,数据集的简单版本也称为采样基准,但实验结果并不理想.通过对实验结果的进一步分析,Leite等^[37]提出了利用学习曲线的信息^[38]刻画数据集特征.学习曲线是算法的性能测度关于采样基准数据集大小的函数,可由向量 $\langle a_{A_k,D,1}, a_{A_k,D,2}, \dots, a_{A_k,D,\#S} \rangle$ 表示.其中: D 为一个数据集, A_k 为第 k 个算法, $\#S$ 为在数据集 D 上的采样个数(采样大小每次以几何级数增长), $a_{A_k,D,r}$ 为算法 A_k 在数据集 D 的第 r 个采样上的准确率.各个数据集的学习曲线信息可用来识别与当前数据集最近的学习曲线,但直接以学习曲线作为元特征的效果并不好.为了提高预测准确率,文献[37]采用了基于实例推理中的适应性概念,即为样本容量越大的采样基准赋予更大权重的方法以调整学习曲线,使其适应新的环境.实验表明,适应性学习曲线能获得更高的预测准确率,且预测准确率随着采样个数的增加而提高.

3.3 基于模型的元特征

基于模型的元特征是指通过数据集训练出一个模型,然后将该模型的一些结构特性作为数据集的特征测度.Bensusan等^[39]提出了从得到的决策树模型中获取信息来刻画学习的复杂度,基于决策树列出10个特征测度,如决策树节点数与属性个数的比值、节点数与训练实例的比值等.Peng等^[40]在此基础上作了一些改进,重新分析了决策树的特性,提出15个新的测度,这些测度着重于刻画决策树的结构性质,如节点和叶子的数目、关于决策树每一层和每一分枝节点分布的统计测度、决策树的深度和宽度、决策树中属性的分布等,Peng等根据这些特征测度对10个学习算法进行了排序.Sun等^[41]在前两种类型元特征集合的基础上加入了反映目标算法间逻辑联系的信息,通过RIPPER算法生成规则集来描述在哪种情形下一种算法优于另一种算法,然后将该规则集或规则集中的每条规则看作一个布尔变量,得到新的元特征.

可以看出,基于统计和信息论的元特征以直接方式获得数据集的特征,但提取的特征是否与算法性能关联紧密却不得而知.基于模型和基于基准思想的元特征间接通过模型的某些特性提取数据集特征,两者的区别在于基于模型的特征测度是根据单个模型(如决策树)的结构特性刻画数据集特征;基于基准思想的特征测度通过多个模型(基准分类器)的性能刻画数据集特征.无论哪种类型的元特征,均必须满足两个最基本的条件:一是所提取的元特征要与学习算法

的性能高度相关;二是与学习过程相比,元特征的计算复杂度要尽可能的小.

4 基于元算法类型的分类

算法选择结果(元目标)最后以何种形式提供给用户取决于用户的需求.例如,在某些情况下,若用户只想简单地知道哪种算法最优,则算法选择系统只需直接提供单个最优算法结果即可;若用户希望了解关于算法性能更详细的信息,则算法选择系统需返回算法性能的估计值;若用户只对各个算法的相对性能感兴趣,则算法选择系统应给用户提供算法的排序.因此,根据用户的不同需求,元目标形式主要包括单个最优算法、算法的排序、算法的性能估计3种形式.元目标的形式决定了元算法的类型,按照元目标的形式,元算法分为分类算法、回归算法和排序算法3种类型.

4.1 分类算法

当用户只想知道哪种算法最优时,算法选择结果一般以单个最优算法形式提供给用户,这时元目标为单个分类变量,在这种情况下元算法采用分类算法类型.基于规则的分类算法以其具备较强解释性的优点,经常用来挖掘元知识,STATLOG项目便是通过决策树算法(C4.5)挖掘元知识.例如,STATLOG中对分类算法CART的学习结果可表示如下:若实例个数 <6435 ,且偏斜度 >0.57 ,则CART可用;若实例个数 >6435 ,或实例个数 <6435 且偏斜度 <0.57 ,则CART不可用.

NOEMON方法通过算法的两两比较生成若干个元学习模型,即若有 n 个候选算法,则可建立 C_n^2 个元学习模型,分别对这些元学习模型采用决策树算法生成规则形式的元知识.

在机器学习领域中,分类算法研究较多,因此将其作为元算法的优点是具有较大的选择余地,但分类算法的返回结果为单个最优算法.大量实验表明,算法选择系统提供给用户的单个算法不一定是最优的,因此,分类算法作为元算法的缺点是当选择的算法不是最优时,不能给用户提供任何备选算法信息.

4.2 回归算法

当用户对所选择算法的性能感兴趣时,元目标为连续变量,元算法一般采用回归算法类型.对于分类算法而言,可用的回归算法较少,Gama等^[42]首次提出了利用回归算法预测学习算法性能,分别使用简单线性回归、基于实例的回归和双线性回归3种不同的回归方法对算法误差进行估计.当要预测的目标为多个算法性能时,该问题又转化为多目标回归问题,可

以通过建立多个回归模型分别对多个算法性能进行预测, 如文献[43]通过建立7个回归支持向量机模型, 对7种聚类算法性能进行比较. 多目标回归也可以通过建立单个回归模型同时对多个算法性能进行预测, 如文献[44]通过聚类树方法建立单个模型, 实现了同时对多种算法性能的预测. 虽然单个回归模型比多个回归模型具有更好的可读性, 但是该模型并不能提供关于算法相对性能的元知识. 回归类型算法^[45]主要应用于对算法具体性能值的预测, 该预测值的误差容易影响算法的相对性能排序.

4.3 排序算法

在实际应用中, 用户有时不需要准确地预测算法在某个数据集上的真实性能, 只需预测出算法的相对性能即可. 在这种情况下, 算法选择的任务便是对算法进行排序, 这也是近几年算法选择研究的热点. k -NN算法经常作为元算法用来对候选算法进行排序. 其基本思想是: 若一个算法在数据集 A 上的分类性能较好, 则该算法同样会在与 A 相似的数据集上得到好的分类性能. Bradzil 等^[46]首次利用 k -NN 算法生成机器学习算法排序, 并提出了一套评估算法排序的方法. 文献[47]通过标签排序技术解决了算法排序问题. 文献[48]将算法排序转化为算法的两两比较, 并提出了利用主动学习技术减少算法两两比较的次数. 文献[49]使用统计关系学习理论直接预测算法的排序.

由上述分析可见, 算法排序更符合用户实际需求. 事实上, 单纯从算法角度看, 算法排序问题与搜索引擎中的排序问题在本质上是相同的, 因此, 信息检索中的排序学习算法均可以应用到算法排序中. 除了排序学习算法技术, 信息检索领域还为算法排序评估提供了多个评估测度, 如 SRCC (斯皮尔曼秩相关系数)、WRC (加权秩相关)、LA (松散准确性)、NDCG (归一化打折累计收益) 等.

5 元学习算法选择存在的问题和发展方向

目前, 基于元学习思想的算法选择问题虽然取得了一定的研究成果, 也研发了若干个帮助和引导用户进行算法选择的系统, 但算法选择涉及到的元学习关键技术仍需要不断改进, 基于元学习思想的算法选择问题仍有较大的发展空间.

5.1 元特征的有效性和完备性

元特征的有效性是指所提取的元特征要与算法性能高度相关; 元特征的完备性是指所列举的元特征能够完全、充分地反映算法性能. 不同研究者提出了不同数目的特征测度, 如 STATLOG 使用的特征测度

为 16 个, Ali 等提出了 31 个特征测度, 而 DCT 工具箱中涉及的特征测度有 25 个. 在这些数据集特征中, 哪些特征测度真正对算法性能产生影响? 哪些与算法性能不相关? 哪些特征之间是冗余的? 所列举的特征测度是否能完全描述数据集? 这些问题都值得进一步研究. Kalousis 等^[50]在 NOEMON 的基础上对数据集的特征进行了约简, 实验表明, 对数据集特征进行约简后, NOEMON 的系统效能有显著提高. 另外, Rice 的算法选择模型只考虑了问题特征, 忽略了各种算法的特征, 每种算法都有其相应的适用范围和局限性, 因此不但要提取问题集合的特征, 还要描述算法集合的特征. 如 Hilario 等^[51]提出了利用算法能处理的数据类型、学习策略 (是消极学习还是积极学习)、算法的适应性 (对不相关属性是否敏感)、算法的可操作性 (参数设置是否复杂) 等刻画算法特征. 但目前对元特征的研究更侧重于对数据集特征的刻画, 将算法特征作为元特征的文献非常少. 因此, 本文认为对元特征的研究可以朝两个方向发展: 一方面, 可在现有特征测度基础上对数据集特征进行约简, 去除冗余的、不相关的特征测度; 另一方面, 也可尝试结合多种类型的特征测度 (包括算法的特征) 提高算法推荐的准确率.

5.2 元学习思想在算法选择问题上应用的狭隘性

目前, 将元学习思想用于解决分类算法选择问题的相关文献较多, 而元学习思想在其他模型和算法选择问题上的应用偏少. 少数研究者将元学习思想用于解决除分类外的算法选择问题, 这些应用主要集中于对回归算法、时间序列模型和支持向量机参数的选择. 如: Ricardo 等^[52]利用 NOEMON 方法对时间序列模型进行选择和排序; Soares 等^[53]提出了利用元学习方法选择支持向量机的核宽度; Ali 等^[54]将元学习思想用于解决支持向量机核函数的自动选择问题; 文献[55-56]采用元学习得到的 SVM 参数作为搜索算法的初始值; METAL 项目的研究目的也是为了将分类算法的选择扩展到回归算法的选择. 实际上, 元学习思想在其他领域的算法选择问题上还有较大的发展空间, 还可将其用于优化问题, 即在满足一系列约束条件的前提下, 花最小的代价寻找问题的解. 然而, 元学习思想在其他领域的拓宽应用给研究者带来很大的挑战, 各种实际问题的特征难以提取是元学习研究者面临的最大障碍.

5.3 算法性能测度的确定

机器学习领域有多个评价算法性能的测度^[57], 较常用的性能测度有准确率、真正例比、真负例比、查准率、F-测度、ROC 曲线面积、均值绝对误差、训

练时间、测试时间、可解释性等. 大多数研究者在使用元学习方法进行算法选择时考虑的算法性能测度更侧重于准确率和训练时间, 而忽略了其他的性能测度. 文献[46]考虑的算法性能测度有算法的准确率和训练时间, 并通过 ARR(adjust ratio of ratios)测度综合二者. NOEMON分类器的性能测度由准确率、执行时间、训练时间和资源需求(内存空间)组成, 为了将多个性能评价准则的结果映射为一个可比较的标量, NOEMON采取数据包络分析(DEA)方法^[58-59]解决该问题. 但DEA的缺陷是算法性能测度的权重只能由DEA方法自身生成, 不能由用户指定, 缺乏一定的灵活性. Ali等^[60]考虑了多种分类性能测度, 包括TPR、TNR、正确分类率、加权F-测度和计算复杂度(包括训练时间和测试时间). 文献[25]将算法的可解释性纳入算法性能测度. 选择适当的算法性能测度、将多个性能测度映射成一个标量和多个性能测度权重的确定均需进一步深入研究.

5.4 元知识的可扩展性和可解释性

每种类型的算法都有其相应的适用范围和局限性, 如: 决策树算法最适合由“属性-值”对表示、目标函数具有离散的输出值、可能需要析取等特征描述的学习问题; 支持向量机中的序贯最小优化算法(SMO)^[61]在高度稀疏的数据集上有较好的性能; k -近邻算法适合训练实例个数多的数据集等. 以往的实验经验^[62]均可以作为元知识指导用户选择适当的算法, 但是仅依靠这些专家经验还远远不够, 且获取的元知识也缺乏可扩展性. 另一方面, 基于元学习思想的算法选择方法大多是黑盒形式, 获取的元知识缺乏可解释性^[63], 因此可以考虑结合专家经验和元学习方法, 通过元知识的合成进行算法选择.

6 结 论

算法选择问题是各研究领域普遍面临的一项挑战任务, 利用元学习思想能够有效地挖掘元知识, 辅助用户选择适当算法. 在挖掘元知识的过程中, 数据集特征的提取和元算法的确定是元学习方法的关键技术. 本文首先提出了基于元思想的算法选择框架, 按照数据集特征和元算法类型对基于元学习思想的算法选择问题进行了分类, 并指出了目前研究存在的不足和下一步的研究方向, 为完善算法选择问题研究提供了理论基础和技术参考.

参考文献(References)

[1] Quinlan J R. Introduction of decision trees[J]. Machine Learning, 1986, 1(1): 81-106.
[2] Bishop C M. Neural networks for pattern recognition[M]. Oxford: Oxford University Press, 1996: 164-191.

[3] Aha D, Kibler D, Albert M. Instance-based learning algorithm[J]. Machine Learning, 1991, 6(1): 37-66.
[4] Cohen W W. Fast effective rule induction[C]. Proc of the 12th Int Conf on Machine Learning. California: Morgan Kaufmann, 1995: 115-123.
[5] Vapnik V N. The nature of statistical learning[M]. New York: Springer, 1995: 123-167.
[6] Vapnik V N. An overview of statistical learning theory[J]. IEEE Trans on Neural Network, 1999, 10(5): 988-999.
[7] Wolpert D H, Macready W G. No free lunch theorems for search[J]. IEEE Trans on Evolutionary Computation, 1997, 1(1): 67-82.
[8] Mitchell T. Machine learning[M]. New York: McGraw-Hill, 1997: 201-225.
[9] Gordon D F, Desjardins M. Evaluation and selection of biases in machine learning[J]. Machine Learning, 1995, 20(1/2): 5-22.
[10] Rice J R. The algorithm selection problem[J]. Advances in Computers, 1976, 15(1): 65-118.
[11] Schaffer C. Selecting a classification method by cross validation[J]. Machine Learning, 1993, 13(1): 135-143.
[12] Sleeman D, Rissakis M, Craw S, et al. Consultant-2: Pre and post-processing of machine learning applications[J]. Int J of Human Computer Studies, 1995, 43(1): 43-63.
[13] Brodley C E. Recursive automatic bias selection for classifier construction[J]. Machine Learning, 1995, 20(1/2): 63-94.
[14] Carrier C G, Vilalta R, Brazdil P. Introduction to the special issue on meta-learning[J]. Machine Learning, 2004, 54(3): 187-193.
[15] Vilalta R, Drissi Y. A perspective view and survey of meta-learning[J]. Artificial Intelligence Review, 2002, 18(2): 77-95.
[16] Rendell L, Cho H. Empirical learning as a function of concept character[J]. Machine Learning, 1990, 5(3): 267-298.
[17] Aha D. Generalizing from case studies: A case study[C]. Proc of the 9th Int Conf on Machine Learning. Aberdeen: Morgan Kaufmann, 1992: 1-10.
[18] Michie D, Spiegelhalter D J, Taylor C C. Machine learning, neural and statistical classification[M]. New York: Ellis Horwood, 1994: 4-5.
[19] King R D, Feng C, Sutherland A. STATLOG: Comparison of classification algorithms on large real-world problems[J]. Applied Artificial Intelligence, 1995, 9(3): 289-333.
[20] Brazdil P, Giraud-Carrier C, Soares C, et al. Metalearning: Applications to data mining[M]. Berlin: Springer, 2009: 31-60.

- [21] Engels R, Theusinger C. Using a data metric for offering preprocessing advice in data-mining applications[C]. Proc of the 13th European Conf on Artificial Intelligence. Chichester: John Wiley and Sons, 1998: 430-434.
- [22] Ho T K, Basu M. Complexity measures of supervised classification problems[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2002, 24(3): 289-300.
- [23] Rossi A L D, Carvalho A C P L F, Soares C, et al. MetaStream: A meta-learning based method for periodic algorithm selection in time-changing data[J]. Neurocomputing, 2014, 127: 52-64.
- [24] Linder C, Studer R. AST: Support for algorithm selection with a CBR approach[C]. Lecture Notes in Computer Science—Principles of Data Mining and Knowledge Discovery. Heidelberg: Springer, 1999, 1704: 418-423.
- [25] Kalousis A, Theoharis T. NOEMON: Design, implementation and performance results of an intelligent assistant for classifier selection[J]. Intelligent Data Analysis, 1999, 3(5): 319-337.
- [26] Kalousis A, Hilario M. Model selection via meta-learning: A comparative study[J]. Int J on Artificial Intelligence Tools, 2001, 10(4): 525-554.
- [27] Song Q B, Wang G T, Wang C. Automatic recommendation of classification algorithms based on data set characteristics[J]. Pattern Recognition, 2012, 45(7): 2672-2689.
- [28] Tatti N. Distances between data sets based on summary statistics[J]. J of Machine Learning Research, 2007, 8(1): 131-154.
- [29] Smith K A, Woo F, Ciesielski V, et al. Matching data mining algorithm suitability to data characteristics using a self-organising map[C]. Advances in Soft Computing—Hybrid Information Systems. Heidelberg: Springer, 2002, 14: 169-179.
- [30] Soares C, Brazdil P. Zoomed ranking: Selection of classification algorithms based on relevant performance information[C]. Proc of Principles of Data Mining and Knowledge 4th European Conf. Berlin: Springer-Verlag, 2000: 160-181.
- [31] Gnanadesikan R. Methods for statistical data analysis of multivariate observations[M]. New York: Wiley, 1997: 139-220.
- [32] Tamhane A C, Dunlop D D. Statistics and data analysis[M]. Upper Saddle River: Prentice-Hall, 2000: 20-52.
- [33] Pfahringer B, Bensusan H, Garrier C G. Meta-learning by landmarking various learning algorithms[C]. Proc of the 17th Int Conf on Machine Learning. San Francisco: Morgan Kaufmann, 2000: 743-750.
- [34] Bensusan H, Giraud-Carrier C. Discovering task neighbourhoods through landmark learning performances[C]. Proc of the 4th European Conf on Principles and Practice of Knowledge Discovery in Databases. Berlin: Springer-Verlag, 2000: 325-330.
- [35] Furnkranz J, Petrak J. An evaluation of landmarking variants[C]. Working Notes of the ECML/PKDD2000 Workshop on Integrating Aspects of Data Mining, Decision Support and Meta-Learning. Heidelberg: Springer, 2001: 57-68.
- [36] Soares C, Petrak J, Brazdil P. Sampling-based relative landmarks: Systematically test-driving algorithms before choosing[C]. Lecture Notes in Computer Science—Progress in Artificial Intelligence. Heidelberg: Springer, 2001, 2258: 88-94.
- [37] Leite R, Brazdil P. Predicting relative performance of classifiers from samples[C]. Proc of the 22nd Int Conf on Machine Learning. Bonn, 2005: 497-503.
- [38] Leite R, Brazdil P. Improving progressive sampling via meta-learning on learning curves[J]. Lecture Notes in Computer Science—Machine Learning. Heidelberg: Springer, 2004, 3201: 250-261.
- [39] Bensusan H, Giraud C, Kennedy C. Higher-order approach to meta-learning[C]. The ECML 2000 workshop on Meta-Learning: Building Automatic Advice Strategies for Model Selection and Method Combination. Heidelberg: Springer, 2000: 109-117.
- [40] Peng Y H, Flach P A, Soares C, et al. Improved dataset characterization for meta-learning[C]. Proc of Discovery Science 5th Int Conf. Lubeck Germany, 2002: 141-152.
- [41] Sun Q, Pfahringer B. Pairwise Meta-Rules for better meta-learning-based algorithm ranking[J]. Machine Learning, 2013, 93(1): 141-161.
- [42] Gama J, Brazdil P. Characterization of classification algorithms[C]. Proc of the 7th Portugese Conf in AI. Madeira Island, 1995: 83-102.
- [43] Marcilio C P de Souto, Ricardo B C Prudencio, Rodrigo G F Soares, et al. Ranking and selecting clustering algorithms using a meta-learning approach[C]. Proc of Int Joint Conf on Neural Networks. Hong Kong, 2008: 3729-3735.
- [44] Todorovski L, Blockeel H, Dzeroski S. Ranking with predictive clustering trees[C]. Lecture Notes in Computer Science—Machine Learning: ECML 2002. Heidelberg: Springer, 2002, 2430: 444-455.
- [45] Bensusan H, Alexandros K. Estimating the predictive accuracy of a classifier[C]. Lecture Notes in Computer Science—Machine Learning: ECML 2001. Heidelberg: Springer, 2001, 2167: 25-36.

- [46] Brazdil P, Soares C, Costa J. Ranking learning algorithms: Using IBL and meta-learning on accuracy and time results[J]. *Machine Learning*, 2003, 50(3): 251-277.
- [47] Kanda J, Soares C, Hruschka E, et al. A meta-learning approach to select meta-heuristics for the traveling salesman problem using MLP-Based label ranking[C]. *Lecture Notes in Computer Science—Neural Information Processing*. Heidelberg: Springer, 2012, 7665: 488-495.
- [48] Leite R, Brazdil P. Active testing strategy to predict the best classification algorithm via sampling and metalearning[C]. *Proc of the 2010 Conf on ECAI 2010: 19th European Conference on Artificial Intelligence*. Lisbon, 2010: 309-314.
- [49] Kotthoff L, Gent I P, Miguel I. An evaluation of machine learning in algorithm selection for search problems[J]. *AI Communications*, 2012, 25(3): 257-270.
- [50] Kalousis A, Hilario M. Feature selection for meta-learning[C]. *Lecture Notes in Computer Science—Advances in Knowledge Discovery and Data Mining*. Heidelberg: Springer, 2001, 2035: 222-233.
- [51] Hilario M, Kalousis A. Fusion of meta-knowledge and meta-data for case based model selection[C]. *Lecture Notes in Computer Science—Principles of Data Mining and Knowledge Discovery*. Heidelberg: Springer, 2001, 2168: 180-191.
- [52] Ricardo B C P, Teresa B L. Selecting and ranking time series models using the NOEMON approach[C]. *Lecture Notes in Computer Science—Artificial Neural Networks and Neural Information Processing*. Heidelberg: Springer, 2003, 2714: 654-661.
- [53] Soares C, Brazdil P, Kuba P. A meta-learning method to select the kernel width in support vector regression[J]. *Machine Learning*, 2004, 54(3): 195-209.
- [54] Ali S, Smith K A. A meta-learning approach to automatic kernel selection for support vector machines[J]. *Neurocomputing*, 2006, 70(1-3): 173-186.
- [55] Taciana A F G, Ricardo B C P, Soares C, et al. Combining meta-learning and search techniques to select parameters for support vector machines[J]. *Neurocomputing*, 2012, 75(1): 3-13.
- [56] Reif M, Shafait F, Dengel A. Meta-learning for evolutionary parameter optimization of classifiers[J]. *Machine Learning*, 2012, 87(3): 357-380.
- [57] Lim T S, Loh W E, Shih Y S. A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms[J]. *Machine Learning*, 2000, 40(3): 203-229.
- [58] Nakhaeizadeh G, Schnabl A. Development of multi-criteria metrics for the evaluation of data mining algorithms[C]. *Proc of the 3th Int Conf on Knowledge Discovery in Databases and Data Mining*. California, 1997: 37-42.
- [59] Charnes A, Cooper W W, Rhodes E. Measuring the efficiency of decision making units[J]. *European J of Operational Research*, 1978, 2(6): 429-444.
- [60] Ali S, Smith K A. On learning algorithm selection for classification[J]. *Applied Soft Computing*, 2006, 6(2): 119-138.
- [61] Platt J C. Sequential minimal optimization: A fast algorithm for training support vector machines[R]. Microsoft Research, 1998.
- [62] Holte R C. Very simple classification rules perform well on most commonly used datasets[J]. *Machine Learning*, 1993, 11(1): 63-90.
- [63] Kalousis A, Gama J, Hilario M. On data and algorithms: Understanding inductive performance[J]. *Machine Learning*, 2004, 54(3): 275-312.

(责任编辑: 郑晓蕾)