

基于多分类支持向量机的优化算法智能推荐系统与实证分析^{*}

崔建双, 车梦然

(北京科技大学东凌经济管理学院, 北京 100083)

摘 要: 算法智能推荐是超启发式算法研究领域一个重要分支, 其目标是从众多“在线”算法中自动选择出最适于当前问题的算法, 从而大大提升解决问题的效率。基于此提出并验证了一种优化算法智能推荐系统, 理论依据是无免费午餐定理和 Rice 算法选择框架, 并假设问题特征与算法性能表现之间存在潜在关联关系, 从而可以把算法推荐问题转换为一个多分类问题。为了验证假设的成立, 以多模式资源约束项目调度问题为测试样本数据集, 以粒子群、模拟退火、禁忌搜索和人工蜂群等元启发式优化算法为推荐对象, 以支持向量机多分类策略实现算法的分类推荐。交叉验证结果表明, 推荐准确率均在 90% 以上, 各项评价指标表现优秀。

关键词: 算法推荐; 问题特征; 多分类支持向量机; 多模式资源约束项目调度问题

中图分类号: TP181

文献标志码: A

doi: 10.3969/j.issn.1007-130X.2019.01.020

An intelligent recommendation system for optimization algorithms based on multi-classification support vector machine and its empirical analysis

CUI Jian-shuang, CHE Meng-ran

(Dolinks School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, China)

Abstract: Intelligent algorithm recommendation is an important branch of the research field of hyper-heuristic algorithms. Its goal is to automatically select the most suitable algorithm for the problem to be solved from many "online" algorithms, thereby greatly improving the efficiency of problem solving. We propose and validate an intelligent optimization algorithm recommendation system, whose theoretical basis is the No Free Lunch theorem and Rice's algorithm selection framework. It assumes that there is a potential correlation between problem features and algorithm performance, thus the algorithm recommendation problem can be converted into a multi-classification problem. In order to verify the assumption, the multi-mode resource constrained project scheduling problem is chosen as the test sample data, a number of meta-heuristic optimization algorithms such as the particle swarm optimization, simulated annealing, tabu search, and artificial bee swarm, are used as the recommended algorithms, and the multi-classification strategy of support vector machine is applied to achieve algorithm classification recommendation. Cross-validation results show that the recommendation accuracy exceeds 90% and the evaluation indicators perform well.

Key words: algorithm recommendation; problem feature; multi-classification support vector machine; multi-mode resource constrained project scheduling problem

^{*} 收稿日期: 2018-04-23; 修回日期: 2018-06-07

基金项目: 国家自然科学基金(71472013); 中央高校基本科研业务费专项资金(06106175)

通信地址: 100083 北京市北京科技大学东凌经济管理学院

Address: Donlinks School of Economics and Management, University of Science and Technology Beijing, Beijing 100083, P. R. China

1 引言

在解决大规模组合优化问题时,各种元启发式优化算法以其高效率、高时效性的特点获得了广泛的应用,但一个不可忽略的事实是人们对这类算法的适用条件和适用场合仍然缺乏普遍认可的规则和依据。为了达到理想的优化效果,多凭个人经验具体问题具体设计,并试图通过个别算例结果的改进来证明其算法的有效性和先进性^[1]。No free lunch 定理^[2]指出:世间不存在一个与具体应用无关的普遍适用的“通用算法”能够高效地解决所有问题。换句话说,与其费尽心思地通过不断地改进算法来适应问题,或许主动地让问题来选择算法效果会更好。多年来,人们已经注意到在算法的性能表现与问题自身固有属性之间可能隐藏着某种未知的内在关系,这种关系使得一种算法具备更高效地解决一个(或一类)问题的能力^[3]。所谓算法推荐就是要依据“问题-算法”之间这种潜在的内在关系,从众多“在线”的可用算法中自发地寻找出最适于解决当前问题的算法。但是迄今为止,研究者们对这种内在关系的揭示仍然处于探索阶段,特别是在优化算法智能推荐领域,相关研究文献乏善可陈。

事实上,算法推荐并非是一个全新的概念,在不同的研究领域有着不同的表述和侧重点,如超启发式算法^[4]、算法自动配置^[5]、元学习^[6]、适值景观分析^[7]、算法组合方法^[8]等。Rice^[3]最早给出了算法推荐的概念框架,并将其形式化地表述为:对于给定问题实例 $x \in P$ (问题空间),其特征为 $f(x) \in F$ (可测特征空间),找到算法选择的映射 $\alpha = S(f(x)) \rightarrow A$ (算法空间),使得所选算法 $\alpha \in A$ 最大化性能 $y(\alpha(x)) \in Y$ (算法性能测度空间)。这一表述虽然简单但实现起来却有很大的难度。如何提取反映问题真实特征的特征?其有效性如何度量?使用算法的哪些性能测度指标?如何实现特征到算法的映射?这些问题为研究者提供了许多待解课题。

近年来,随着人工智能和大数据研究的兴起,有学者把算法推荐看成是一种学习过程,借鉴机器学习领域元学习的研究成果、方式和方法来实现算法自动推荐。Smith-Miles^[9]曾提出基于元学习思想的不同学科领域实现算法推荐的统一框架,认为在包括统计回归、排序、时间序列预测、约束满足和最优化等领域,算法智能推荐的实现均有潜在的发展空间。20 世纪 90 年代早期,Rendell 等^[10]首次

把算法推荐看成是一种机器学习过程,Aha^[11]更进一步明确地提出了基于规则的算法推荐规则。此后,基于机器学习的算法智能推荐逐渐获得重视,形成了元学习领域^[12]。元学习强调对算法在分类问题上的表现进行学习,基于元学习的算法选择就是应用元学习系统寻求问题的元特征与算法相对性能测度之间的映射模型。欧洲的 STATLOG^[13]项目比较了机器学习、神经网络和统计学分类方法在不同问题上的性能,研究数据集特征与算法性能表现之间的关系以获取关于算法推荐的知识。Cui 等^[14]提出了基于元学习框架的大厦能源模型推荐系统,并进一步研究了使用通用元建模的元学习自动机。Messelis 等^[15]研究了基于实证难度模型的多模式资源受限项目调度问题的智能算法推荐。Miranda 等^[16]提出了基于元学习框架的支持向量机参数改进模型。Smith-Miles 等^[17]量化了组合优化问题一般特征和部分组合优化问题的特定特征。崔建双等^[18]证明了组合优化问题算法选择定理,指出在多个优化算法中必定有一个最适用于求解该问题的算法。Bhatt 等^[19]对当前基于数据集特征的学习方法所面临的挑战做出了综述。

综上所述,算法推荐问题一直是众多学者们关注的目标,然而这些研究成果大多是聚焦在机器学习领域针对机器学习算法的推荐,而涉及到优化算法推荐的研究成果则十分匮乏。

本文将在 Rice 概念框架的基础上,把机器学习领域基于元学习的算法推荐过程拓展至基于元启发式优化算法解决大规模组合优化问题的算法推荐领域。以多模式资源约束项目调度问题 MRCPS(P(Multi-mode Resource Constrained Project Scheduling Problem))为数据集^[20],以粒子群优化 PSO(Particle Swarm Optimization)、模拟退火 SA(Simulated Annealing)、禁忌搜索 TB(TaBu search)和人工蜂群 ABC(Artificial Bee Colony)4 种元启发式优化算法为推荐对象,利用多分类支持向量机 MCSVM (Multi-Classification Support Vector Machine)^[21]来构建元启发式优化算法智能推荐系统。

本文的贡献:在对 MRCPS 深入分析的基础上,提出了 47 项基本静态特征代替了文献^[15]提出的 686 项特征,大大简化了特征提取的难度;其次,采用了不同于实证难度模型的优化算法智能推荐方法,该方法完全可以推广至解决其他大规模组合优化问题,如背包问题、装箱问题、旅行商问题、图着色问题等。

实证过程分别选取了 540 个具有相同活动规模和 378 个具有不同活动规模的样本数据进行验证。从实验结果来看,基于机器学习的方法不但获得了很高的推荐正确率,而且计算速度快、鲁棒性强,充分验证了基于机器学习的方法在算法推荐领域的可行性,为提升优化算法的使用效率开辟了一条新的途径。

2 MRCPSP 的样本特征

MRCPSP 是一种典型的组合优化问题,其目标是在满足活动顺序和资源约束的前提下求得项目工期最小化。目前,已经有多种求解算法,国际

流行的标杆算例库提供的算例^[21]可用于比较这些算法的优劣。但是,本文并非研究不同优化算法之间的优劣比较,而是要通过提取 MRCPSP 算例特征来推荐最佳算法,其目的是提升解决组合优化问题的效率。特征是反映和区别算例内在属性的重要方法,本着特征有效的原则,本文仅提取了 47 个有关 MRCPSP 的基本静态特征(见表 1),以及在这些特征基础上得到的若干统计特征,大大简化了此类问题的特征空间,降低了特征维数。这些特征多基于网络结构性质、网络复杂度、资源系数、资源强度与可用性等。从后面的使用结果来看,这些特征足够准确地区分问题,使用这些特征得到的优化算法推荐准确率都能够超过 90%。

Table 1 Basic static characteristics

表 1 基本静态特征

序号	特征	变量	特征描述	序号	特征	变量	特征描述
(1)	Anum	<i>NumActivity</i>	活动数量	(25)	AvgRRuse	<i>SumRR_Anum</i>	可更新资源平均可用量
(2)	Mnum	<i>NumMode</i>	模式数量	(26)	AvgNRuse	<i>SumNRR_Anum</i>	不可更新资源平均可用量
(3)	Enum	<i>SumEdges</i>	边的数量	(27)	TotalDur	<i>Sumall actsDur_Mnum</i>	全体活动持续时间和
(4)	RRTnum	<i>typesRR</i>	可更新资源类型数	(28)	AvDur	<i>TotalDur_Anum</i>	活动持续时间均值
(5)	NRRnum	<i>typesNRR</i>	不可更新资源类型数	(29)	Maxdur	<i>Maxall actsDur</i>	活动持续时间最大值
(6)	ARTnum	<i>typesRR_NRR</i>	所有资源类型数	(30)	Mindur	<i>MinallactsDur</i>	活动持续时间最小值
(7)	SumRR	<i>SumRR</i>	可更新资源总量	(31)	CNC	<i>NumEdges_Anum</i>	网络复杂度系数
(8)	SumNRR	<i>SumNRR</i>	不可更新资源总量	(32)	RF	$RF = \frac{1}{N} \frac{1}{K} \sum_{j=1}^N \sum_{k=1}^K \text{sgn}(r_{jk})$ $\text{sgn}(r_{jk}) = \begin{cases} 1, & r_{jk} > 0 \\ 0, & \text{otherwise} \end{cases}$	可更新资源系数,反映活动对不同类型可更新资源的依赖程度, N 是活动数量, K 是可更新资源数量
(9)	Maxsuc	<i>Maxsuccessor</i>	前向最大后继数量	(33)	OS	<i>PR_Anum</i>	次序强度
(10)	MaxsucPos	<i>Maxsuc position</i>	反向最大后继数量	(34)	CI	<i>MinNode</i>	网络复杂度指数
(11)	Minsuc	<i>Minsuccessor</i>	前向最小后继数量	(35)	RS _k	$RS_k = \frac{R_k - R_k^{\min}}{R_k^{\max} - R_k^{\min}}$	可更新资源强度,反映第 k 种可更新资源的稀缺状况
(12)	MinsucPos	<i>Minsuc position</i>	反向最小后继数量	(37)	NRS _k	$NRS_k = (NR_k - NR_k^{\min}) / (NR_k^{\max} - NR_k^{\min})$	不可更新资源强度,反映第 k 种不可更新资源的稀缺状况
(13)	Maxpre	<i>Maxpredecessor</i>	前向最大前驱数量	(39)	RC _k	$RC_k = \frac{\bar{R}_k}{R_k}$	可更新资源受限程度,反映资源稀缺程度
(14)	MaxprePos	<i>Maxpre position</i>	反向最大前驱数量	(41)	NRC _k	$NRC_k = \frac{\overline{NR_k}}{NR_k}$	不可更新资源受限程度,反映了资源的稀缺程度
(15)	Minpre	<i>Minpredecessor</i>	前向最小前驱数量	(43)	MD	<i>medianx</i>	特征中位值
(16)	MinprePos	<i>Minpre position</i>	反向最小前驱数量	(44)	AV	<i>meanx</i>	特征均值
(17)	UBRRusek	<i>UBRR</i>	可更新资源上限	(45)	GM	<i>prodx</i>	特征几何均值
(18)	UBNRusek	<i>UBNR</i>	不可更新资源上限	(46)	VR	<i>varx</i>	特征方差
(19)	MaxNRusek	<i>MaxNRu</i>	不可更新资源最大使用量	(47)	ED	<i>pdistx</i>	以上样本值之间的欧氏距离
(20)	MinNRusek	<i>MinNRu</i>	不可更新资源最小使用量				

3 优化算法的相对性能排序

算法 PSO、SA、TB、ABC 虽然在实现机理上有较大区别,但在现实应用中都能很好地解决类似 MRCPSP 的大规模组合优化问题,在不同算例上的表现也不尽相同。为了能够实现准确的算法推荐,这 4 种算法均使用相同的调度编码解和串行进度生成机制^[20],通过快速计算来获取算法针对每一算例的相对性能表现。为了尽可能地做到公平,避免因个人编程技巧、编程语言或硬件设备的不同对算法性能排序造成的影响,以每个算例执行 2 000 次调度邻域解得到的目标值与标准值的偏差以及相应的运行时间作为算法优劣的评价标准,按照 5 次独立运行结果的均值给出排位标号值(1、2、3、4)。排位规则是:距目标值偏差越小排位越靠前,若偏差相同,则计算时间越短越靠前。此外,在调试每一个元启发式算法时,以对大多数算例表现良好的版本作为最终版本,一旦参数调试完成即不再做出修改。

4 优化算法智能推荐系统的实现

4.1 系统实现框架

图 1 是本文提出的优化算法智能推荐系统框架。具体实现步骤如下:(1)从 MRCPSP 标杆算例中筛选整理问题样本数据集,根据表 1 统一提取样本特征;(2)依据第 3 节取得各算例的相对性能排位标号;(3)基于多分类支持向量机训练推荐模型;(4)对未见算例推荐优化算法,对推荐结果做出评价。

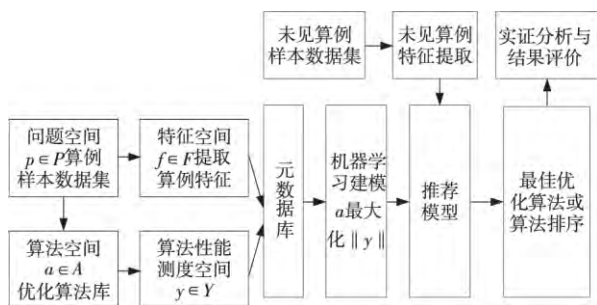


Figure 1 Framework of the optimization algorithm intelligent recommendation system

图 1 优化算法智能推荐系统框架

4.2 多分类支持向量机

SVM 基本模型^[22]定义为特征空间上间隔最大的线性分类器。为了充分利用 SVM 的二分类

泛化能力,本文将利用一对一和一对多的分类策略把一个多分类问题转化为多个二分类问题实现分类推荐。

LIBSVM(LIBrary for Support Vector Machines)是一个知名的 SVM 分类回归软件包。该软件包提供了两个关键函数:分类模型训练函数 *svmtrain* 和预测函数 *svmpredict*,其优点在于参数调节相对较少,具备交互检验功能,用于分类或回归时还可以实现大范围组合参数寻优。

4.3 推荐效果评价指标

有多项指标可用于衡量优化算法的推荐效果,主要包括正确率、准确率、灵敏度以及综合评价指标等。可视化分类推荐效果通过混淆矩阵图、类别比例对照图、预测类别与实际类别对比图等做出评价。(1)正确率(Accuracy):样本分类正确的个数与样本总个数之比。(2)准确率(Precision):实际为正样本且分类为正样本占有所有标记为正样本的比例。(3)灵敏度(Sensitive):实际为正样本中被预测为正样本的比例。(4)综合评价指标(F-Score):准确率和灵敏度的加权调和值, $F-Score = 2 \times Precision \times Sensitive / (Precision + Sensitive)$ 。以上评价指标均在[0,1],值越大越好。

4.4 一对多分类推荐结果

一对多分类 OvR(One versus Rest)^[23]推荐策略是在训练时依次把一个类别的样本归为一类,其余的样本归为另一类,这样 k 个类别的样本可构造出 k 个二分类器。然后使用这 k 个二分类器对未知样本做出分类预测,得到 k 个分类函数值,分类结果确定为具有最大分类函数值的那一类。

4.4.1 540 个相同规模的样本集 OvR 实验结果

从 PSPLIB 算例库^[21]选用 540 个活动数量为 20、活动模式为 3、使用 2 种可更新资源、2 种不可更新资源的一组标杆样本算例。这些算例的特征值由表 1 给定的特征定义提取,即每个算例对应 47 项特征值。排位标号按照第 3 节给定方法获得,即每一算例对应唯一的一个最佳算法标号。这样,优化算法推荐问题被转换为一个多分类问题。样本集数据是一个 540×47 的特征向量矩阵 *featureMatrix540* 和 540×1 的分类标号向量 *rank540*。

依据 OvR 分类策略,对 540 个算例进行分类推荐。实现步骤如下:(1)读入 *featureMatrix540* 和 *rank540*;(2)随机分组测试样本和训练样本;(3)分别对两组样本数据做均值为 0、方差为 1 的

归一化预处理;(4)使用交叉验证搜索最佳惩罚系数 c 和参数 g ;(5)把参数 c 、 g 以及径向基核函数 RBF(Radical Basis Function)代入函数 $svmtrain()$,训练获得 k 个二分类器;(5)利用 k 个二分类器和 $svmpredict()$ 函数对测试样本做出预测;(6)对推荐结果按照 4.3 节给出的指标做出评价。

典型实例:540 个总体样本随机划分为 450 个训练样本和 90 个测试样本,然后以训练样本做最佳参数 c 和 g 三重交叉验证搜索,每次取 1/3 样本做测试,2/3 样本做训练,仅以正确率指标判断是否得到最佳参数。无论是训练还是测试过程均采用 OvR 分类准则,即类别为 i 的样本作为正集($i=1,2,3,4$),其余作为负集。参数 $c=32767$,参数 $g=0.002$ 。虽然有不同的 c 和 g 都对应最高预测准确率,但我们选用具有最小 c 值的那一组 c 和 g 。 c 是惩罚系数,较大的 c 意味着误差宽容度较小。 g 是选择径向基核函数后自带的一个参数,决定了数据映射到新的特征空间后的分布。默认核函数是 RBF, $b=1$ 表明采用输出概率估计模型(Probability Estimation)来决定分类结果。表 2 给出了 10 个包含特征值、实际类别和预测类别的测试实例。

执行上述步骤后,推荐正确率最大值为 98.82%,均值为 92.31%。依据分类结果,图 2a 给出了 540 个样本一对多可视化混淆矩阵图(左上角)、预测类别与实际类别散点对比图(右上角)、实际类别与预测类别比例对照图(下半部)。混淆矩阵(Confusion Matrix)的每一列代表预测类别(Predicted class),每一行代表数据的实际归属类别(True

Class)。预测类别与实际类别散点对比图(Predicted Class vs True Class)中的符号‘o’代表预测类别,‘+’代表实际类别。若‘o’与‘+’基本上重叠,说明预测正确率高。图中箭头所指是此例中唯一出现的一个预测错误。实际类别(True Class)与预测类别(Predicted Class)比例对照图以条形图(Bar Chart)的形式显示出二者比例基本一致,条形图上的数字标出了各类样本的数量。由混淆矩阵可计算得到 PSO、SA、TB、ABC 4 个优化算法的分类准确率分别是 1.00,1.00,1.00,1.00;灵敏度分别是 1.00,1.00,1.00,1.00;综合评价指标分别是 1.00,1.00,1.00,1.00。推荐结果的准确率达到 100%,可见 OvR 预测方法取得了十分理想的推荐结果。

4.4.2 378 个不同规模的样本集 OvR 实验结果

从 PSPLIB 算例库中选择 J10,J12,J14,J16,J18,J20 和 J30 共 7 组算例子集,每个子集随机选择 54 个算例,共 378 个样本算例。这组样本的活动数量分别为 10,12,14,16,18,20,30,故规模不同,活动模式数量均为 3,使用 2 种可更新资源和 2 种不可更新资源。

同样,依据 OvR 分类策略,对这 378 个算例进行分类推荐。首先得到 378×47 的特征向量矩阵 $featureMatrix378$ 和分类标号向量 $rank378$;然后使用 LIBSVM 预测函数完成一对多分类预测,实现过程和步骤与 4.4.1 节完全一样,改变的仅是数据集。典型实例:参数 $c=32767$,参数 $g=0.003$ 。随机选取 314 个训练样本、64 个测试样本,平均推荐正确率为 95.31%。图 2b 是 378 个样本一对多

Table 2 A section of 10 test instances with features,true and predicted classes

表 2 10 个测试实例的特征值、实际类别和预测类别

Ins	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	实际类别	预测类别
1	0.17	-1.04	-0.44	1.28	0.44	-0.13	-1.25	-0.77	-1.13	0.44	-0.83	1	1
2	1.14	-1.19	0.81	-0.12	1.10	1.15	-1.33	-0.98	-1.48	-0.64	-0.86	2	2
3	-1.91	-0.29	1.19	-0.89	-1.91	-1.83	0.15	-0.70	0.20	-1.18	-0.52	2	2
4	-0.38	1.62	0.44	-0.12	-0.73	0.01	1.34	1.76	0.83	1.30	1.06	2	2
5	1.56	0.14	-0.30	-0.50	1.49	1.57	0.23	0.04	1.80	0.49	1.11	3	3
6	0.17	0.29	-0.67	-0.50	0.05	0.30	0.19	0.36	-0.06	0.22	0.87	3	3
7	-0.24	1.36	0.81	0.27	-0.08	-0.41	1.00	1.59	0.65	1.63	1.48	3	3
8	-1.07	-1.25	0.44	2.21	-0.99	-1.12	-1.33	-1.08	-1.04	-0.32	-0.96	4	4
9	0.45	-0.13	-1.41	-0.12	0.31	0.58	0.00	-0.24	0.20	-0.54	-0.96	4	4
10	0.10	0.51	-1.78	1.04	0.31	-0.13	1.00	0.01	0.03	-1.40	-0.39	4	4

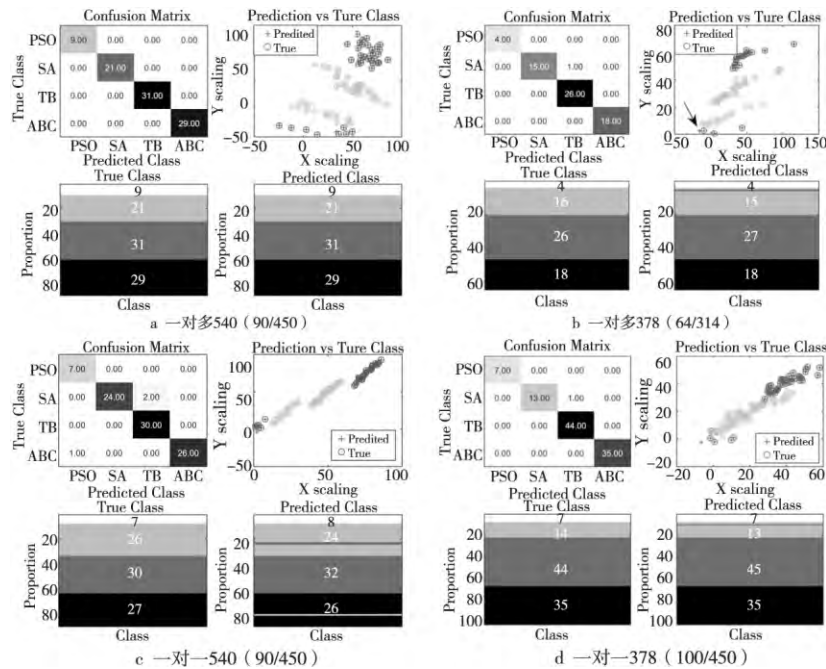


Figure 2 Visualized classification effect diagram

图2 可视化分类效果图

分类混淆矩阵等可视化效果图。算法 PSO、SA、TB、ABC 的分类准确率分别是 1.00, 1.00, 0.96, 1.00; 灵敏度分别是 1.00, 0.94, 1.00, 1.00; 综合评价指标分别是 1.00, 0.97, 0.98, 1.00。

4.5 一对一分类推荐结果

一对一分类 OvO(One versus One)^[23] 推荐首先在每两个类别之间构造分类器, k 个类别需要构造 $k(k-1)/2$ 个分类器。对测试样本进行分类时, 依次使用这些分类器, 每次分类等效于一次投票, 每一类别选择经 $k-1$ 次投票后得票最多的那个类作为该样本的最终类。

4.5.1 540 个相同规模的样本 OvO 实验结果

与 4.4.1 节实现路线基本一致, 但过程有所区别。把样本集分为训练集和测试集, 从训练集训练获得 6 个二分类模型, 分别是 1vs2(含 2vs1), 1vs3(含 3vs1), 1vs4(含 4vs1), 2vs3(含 3vs2), 2vs4(含 4vs2), 3vs4(含 4vs3)。用这 6 个模型依次对测试样本进行二分类, 每一类别分别获得 3 组分类结果。例如, 对第 3 类测试样本进行分类时, 分别利用 1vs3, 2vs3 和 3vs4 3 个二分类模型, 可以得到 3vs1, 3vs2 和 3vs4 的 3 组分类结果, 用这 3 组结果投票决定其最终类别。实验中随机分出 450 个训练样本和 90 个测试样本, 并做 3 重交叉验证搜索最佳参数 c 和 g 。使用最佳参数和 LIBSVM 提供的训练函数和预测函数, 获得最高分类正确率为 96.67%, 均值为 90.56%, 最差为 82.22%。图 2c

给出了 540 个样本的分类效果图, 不难计算出 PSO、SA、TB、ABC 4 个算法的分类推荐准确率分别是 0.88, 1.00, 0.94, 1.00 以及其它评价指标。表 3 是对第 2 类预测投票的一个示例。

Table 3 Prediction vote results for class 2

表3 第2类预测投票结果示例

Num	2vs1	2vs3	2vs4	Vote
1	0(=2)=1	0(=3)=0	1(=2)=1	1
2	0(=2)=1	0(=3)=0	1(=2)=1	1
3	0(=2)=1	1(=2)=1	1(=2)=1	1
4	0(=2)=1	1(=2)=1	1(=2)=1	1
5	0(=2)=1	1(=2)=1	1(=2)=1	1
6	0(=2)=1	0(=3)=0	1(=2)=1	1
7	0(=2)=1	1(=2)=1	1(=2)=1	1
8	0(=2)=1	0(=3)=1	1(=2)=1	1
9	0(=2)=1	0(=3)=0	1(=2)=1	1
10	0(=2)=1	0(=3)=0	1(=2)=1	1
11	0(=2)=1	1(=2)=1	1(=2)=1	1
12	0(=2)=1	1(=2)=1	1(=2)=1	1
13	0(=2)=1	1(=2)=1	1(=2)=1	1
14	0(=2)=1	0(=3)=0	1(=2)=1	1
15	0(=2)=1	1(=2)=1	1(=2)=1	1
16	0(=2)=1	0(=3)=0	1(=2)=1	1
17	0(=2)=1	1(=2)=1	1(=2)=1	1

4.5.2 378 个不同规模的样本 OvO 实验结果

使用 LIBSVM 预测函数和一对一多分类预测

方法,取 278 个训练样本算例,100 个测试样本算例,得到的最高正确率为 94.00%,平均正确率为 89.80%。图 2d 给出了可视化分类效果图。

5 结束语

随着更多优化算法的不断“涌现”,算法的优化特性、适用性和适用范围成为迫切需要解决的问题。首先,许多优化算法表现出了相近的实现机制,但对它们在实现细节上的异同则缺乏深层次的理论分析与对比,多通过大批算例的实测,采用描述统计学的方法做出评价;其次,算法的成功应用多依靠使用者的经验和偏好,启发式痕迹明显,应用效果起伏不定;再者,人们更注重通过算法的改进来解决问题,许多研究仅关注对特定问题特定条件下算法的性能改进,缺乏对算法适用领域规律性的探索,很少关注问题的特征与算法性能表现之间的内在联系。对于众多可用算法中到底哪一种算法更适于所要解决的问题,至今尚无系统性的研究成果。

针对算法研究领域存在的以上问题,本文重点研究了优化算法智能化推荐的可能。这一概念的提出可以认为是比改进算法、拓展算法的应用,甚至是发明新的算法更有意义的一项新课题,也是增强已有算法使用效果、降低算法使用成本的一条新途径。

本文所做的工作主要有如下几个贡献点:(1)尝试回答“一个给定的算法更适宜解决什么样的问题以及为什么的问题”;(2)实现了从多种可用优化算法中准确地选择最适用算法的目标;(3)进一步证实了问题特征与算法性能表现之间可以通过建立关联模型达到高精度预测。

为了进一步巩固研究成果,未来仍需要从以下几个方面展开深入的研究:(1)特征选择问题,如何保证所提取的问题特征与优化算法的性能测度强相关?(2)动态(景观)特征的选择和使用问题,动态特征反映了算法执行期间搜索空间的结构。(3)把本文研究路线推广至其他组合优化问题,如旅行商问题、背包问题等。

参考文献:

- [1] Chen Wang, Shi Yan-jun, Teng Hong-fei, et al. An efficient hybrid algorithm for resource-constrained project scheduling [J]. Information Sciences, 2010, 180(6): 1031-1039.
- [2] Wolpert D H, Macready W G. No free lunch theorems for optimization [J]. IEEE Transactions on Evolutionary Computa-

tion, 1997, 1(1): 67-82.

- [3] Rice J R. The algorithm selection problem [J]. Advances in Computers, 1976, 15(2): 65-118.
- [4] Burke E K, Gendreau M, Hyde M, et al. Hyper-heuristics: A survey of the state of the art [J]. Journal of the Operational Research Society, 2013, 64(12): 1695-1724.
- [5] Lindauer M, Hoos H H, Hutter F, et al. AutoFolio: An automatically configured algorithm selector [J]. Journal of Artificial Intelligence Research, 2015, 53(1): 745-778.
- [6] Cui C, Hu M Q, Weir J D, et al. A recommendation system for meta-modeling: A meta-learning based approach [J]. Expert Systems with Applications, 2016, 46(C): 33-44.
- [7] Morgan R, Gallagher M. Fitness landscape analysis of circles in a square packing problems[C]//Proc of the 10th International Conference on Simulated Evolution and Learning, 2014: 455-466.
- [8] Silverthorn B, Mikkilainen R. Latent class models for algorithm portfolio methods [C]//Proc of the 24th AAAI Conference on Artificial Intelligence, 2010: 167-172.
- [9] Smith-Miles K A. Cross-disciplinary perspectives on meta-learning for algorithm selection [J]. ACM Computing Surveys, 2008, 41(1): Article No. 6.
- [10] Rendell L, Cho H. Empirical learning as a function of concept character [J]. Machine Learning, 1990, 5(3): 267-298.
- [11] Aha D W. Generalizing from case studies: A case study[C]//Proc of the 9th International Workshop on Machine Learning, 1992: 1-10.
- [12] Carrier C G, Vilalta R, Brazdil P. Introduction to the special issue on meta-learning [J]. Machine Learning, 2004, 54(3): 187-193.
- [13] 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.
- [14] Cui C, Wu T, Hu M, et al. Short-term building energy model recommendation system: A meta-learning approach [J]. Applied Energy, 2016, 172(15): 251-263.
- [15] Messelis T, de Causmaecker P. An automatic algorithm selection approach for the multi-mode resource-constrained project scheduling problem [J]. European Journal of Operational Research, 2014, 233(3): 511-528.
- [16] Miranda P B C, Prudencio R B C, de Carvalho A P L F, et al. A hybrid meta-learning architecture for multi-objective optimization of SVM parameters [J]. Neurocomputing, 2014, 143: 27-43.
- [17] Smith-Miles K, Lopes L. Measuring instance difficulty for combinatorial optimization problems [J]. Computers & Operations Research, 2012, 39(5): 875-889.
- [18] Cui Jian-shuang, Liu Xiao-chan, Yang Mei-hua. A meta-learning based intelligent recommendation framework for optimization algorithm automatic selection and its empirical study [J]. Journal of Computer Applications, 2017, 37(4): 1105-1110. (in Chinese)

- [19] Bhatt N, Thakkar A, Ganatra A. A survey & current research challenges in meta-learning approaches based on dataset characteristics [J]. International Journal of Soft Computing and Engineering, 2012, 2(1): 239-247.
- [20] Cui Jian-shuang, Yang Jian-hua. Multi-resource constrained project scheduling problems with discrete particle swarm optimization [J]. Computer Engineering and Applications, 2015, 51(14): 253-257. (in Chinese)
- [21] Chang C C, Lin C J. LIBSVM: A library for support vector machines [J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3): Article No. 27.
- [22] Cortes C, Vapnik V. Support-vector networks [J]. Machine Learning, 1995, 20(7): 273-297.
- [23] Duan K B, Rajapaks J C, Nguyen M N. One-versus-one and one-versus-all multiclass SVM-RFE for gene selection in cancer classification [M] // Lecture Notes in Computer Science, Berlin Heidelberg: Springer-Verlag, 2007.

附中文参考文献:

- [18] 崔建双, 刘晓婵, 杨美华. 基于元学习推荐的优化算法自动选择框架与实证分析[J]. 计算机应用, 2017, 37(4): 1105-1110.

- [20] 崔建双, 杨建华. 多资源约束的项目调度问题离散粒子群算法[J]. 计算机工程与应用, 2015, 51(14): 253-257.

作者简介:



崔建双(1961-), 男, 河北衡水人, 博士, 副教授, 研究方向为机器学习、智能优化和项目调度。E-mail: cuijs@manage.ustb.edu.cn

CUI Jian-shuang, born in 1961, PhD, associate professor, his research interests include machine learning, intelligent optimization, and project scheduling.



车梦然(1994-), 女, 河北保定人, 硕士生, 研究方向为机器学习、智能优化和项目调度。E-mail: 17801235119@163.com

CHE Meng-ran, born in 1994, MS candidate, her research interests include machine learning, intelligent optimization, and project scheduling.

《计算机工程与科学》征文通知

《计算机工程与科学》是由国防科技大学计算机学院主办的中国计算机学会会刊, 是国内外公开发行的计算机类综合性学术刊物, 现为月刊。本刊欢迎关于计算机科学理论、计算机组织与系统结构、计算机软件、计算机应用、计算机器件设备与工艺等学科领域方面的来稿。本刊常年设有高性能计算专栏。

来稿论文必须未发表、未投到其他会议或期刊。

来稿要求和注意事项:

- (1) 主题明确、文字精练、语句通顺、数据可靠。
- (2) 标题、作者单位、摘要、关键词采用中英文间隔行文; 请注明是否基金资助项目论文(注明项目名称和编号), 并注明文章中图法分类号。务必附上所有作者中英文简历(姓名、性别、出生年月、籍贯、学位、职称、研究方向)、1寸证件照片(军人请用便服照)、中英文通信地址、联系电话和 Email。
- (3) 作者在投稿时须注明是否是 CCF 会员(高级会员、普通会员、学生会会员), 若是会员, 请注明会员号。
- (4) 来稿请用 WORD 软件编辑, 格式为 A4, 40 行×40 列, 通栏排版, 正文为 5 号宋体, 论文长度不得低于 5 个标准版面, 并请自留底稿。
- (5) 来稿中图形绘制要求工整、清晰、紧凑, 尺寸要适当, 图中文字用 6 号宋体, 线为 0.5 磅。
- (6) 每篇论文格式要求: 1 引言; ……; 最后是结束语。引言和结束语中尽量不用图和表。附录应放参考文献之后。参考文献限已公开发表的。

(7) 来稿文责自负, 要遵守职业道德, 如摘引他人作品, 务请在参考文献中予以著录。署名的作者应为参与创作, 对内容负责的人。文章发表后, 如不同意其他报、刊、数据库等转载、摘编其作品, 请在来稿时声明。

(8) 本刊不收取作者任何费用(免审稿费、版面费等所有费用)。

联系地址: 410073 湖南省长沙市国防科技大学《计算机工程与科学》编辑部

联系电话: 0731-87002567

电子邮件: jsjgcykx@vip.163.com

投稿主页: <http://joces.nudt.edu.cn>