中国科学: 技术科学 20xx 年 第 xx 卷 第 x 期: 79 ~ 89
SCIENTIA SINICA Technologica tech.scichina.com
《中国科学》杂志社
SCIENCE CHINA PRESS

论 文

近似最优并行算法组智能汇聚构造

刘晟材[①②], 杨鹏[①②], 唐珂[①②*]

① 南方科技大学计算机科学与工程系, 深圳 518055;
② 广东省类脑智能计算重点实验室, 深圳 518055;
* E-mail: tangk3@sustech.edu.cn

收稿日期: 20xx-xx-xx; 接受日期: 20xx-xx-xx; 网络出版日期: 20xx-xx-xx

广东省重点实验室(批准号：2020B121201001)和国家自然科学基金青年科学基金项目(批准号：61806090)资助

摘要 作为一种高性能通用并行求解器，近年来并行算法组(Parallel Algorithm Portfolios，简称 PAPs)在判定、计数以及连续、离散优化等问题上取得了突出的求解效果。传统人工构造 PAP 的方式依赖于大量领域知识，门槛极高。为了解决这一问题，本文提出了一种基于演化优化的 PAP 智能汇聚自动构造方法 AutoPAP。整体上，AutoPAP 遵循$(n+1)$演化优化框架，即在每一代生成n个候选算法，并保留最优算法加入到 PAP 中。考虑到算法配置空间往往非常巨大且涉及混合变量，本文设计了专用变异算子以提升 AutoPAP 的实际性能，并证明了 AutoPAP 在理论上可以达到$(1-1/e)$近似最优构造效果。最后，本文以旅行商问题(Traveling Salesman Problems，简称 TSP)为例，使用 AutoPAP 构造得到 TSP_PAP。实验结果表明，在主流 TSP 测试集上，TSP_PAP 的求解效率和效果均显著好于当前 TSP 上公认性能最佳的求解器 EAX 和 LKH。在 128 个规模 1000 至 30000 的 TSP 测试样例上，相比于 EAX 和 LKH，TSP_PAP 可以将平均求解时间缩短至少 45.71%，并将平均误差率降低至少 87.50%，体现出 AutoPAP 在算法自动设计与演化方面的巨大潜力。

关键词 求解器，并行算法组，演化计算，自动算法设计，旅行商问题，组合优化

1 引言

过去二十年间，并行计算架构得到了巨大发展[1]。现如今多核中央处理器(Central Processing Unit，简称 CPU)已然成为个人计算机的标准配置，而在那些专为科学计算而搭建的计算平台如工作站和服务器上，CPU 核数动辄几十上百。如此丰富的计算资源也给算法/求解器设计者们提出了新的挑战：如何有效利用并行计算平台以更好地解决实际应用中的复杂问题？事实上，无论是学术界还是工业界，对于并行求解器的研究已持续多年。例如，在一些重要的基本计算问题如布尔可满足性问题(Boolean Satisfiability Problem，简称 SAT)、约束满足问题(Constraint Satisfaction Problem，简称 CSP)、回答集编程问题(Answer Set Programming，简称 ASP)，混合线性整数规划问题(Mixed Integer Linear Programming，简称 MILP)和黑箱连续优化问题(Black-box Continuous Optimization，简称 BO)上，并行求解器[2-6] 已经成为

引用格式: xx

主流。此外，在很多基础工业软件例如数学规划求解器 CPLEX[1]和 EDA 软件 Synopsys[2]中，并行求解器已经成为其核心模块。

虽然并行求解器已取得一定程度的发展，其仍面临着研发难度高、周期长的困境。一般而言，并行求解器的研发始于改造已有串行求解器，并需进一步集成信息交互机制以实现问题分解和不同求解线程之间的协作[9]。这一流程要求研发人员具备大量相关领域知识，并需要其耗费大量时间对求解器进行迭代改进，耗时耗力。

近年来，一种新的并行求解器形式—并行算法组(Parallel Algorithm Portfolios，简称 PAP)，在判定[8]、计数[8]以及连续[6]、离散[8-10]优化等问题上取得了突出的求解效果，逐渐成为了研究热点。本质上 PAP 是一个包含若干算法的集合，其中每个算法都被称作该 PAP 的成员算法。当 PAP 被用于求解某个问题样例时，其所有成员算法都将在该问题样例上相互独立地并行运行，且在达到停止条件时同时终止。可以看到，在 PAP 运行过程中，各个成员算法之间不涉及任何信息交互，因而 PAP 的结构复杂度相比于传统并行求解器要低得多。

另一方面，构造高性能 PAP 并不容易。假设一个 PAP 的某个成员算法θ在任何问题样例上的性能都超过了其它成员算法，那么该 PAP 的性能实际上仅等价于算法θ的性能。换言之，除θ之外的成员算法虽然使用了和θ相同的计算资源，却并未给 PAP 带来任何性能增益。因此，直观上而言，一个高性能 PAP 所包含的成员算法在性能上应具备差异性，即各个成员算法都应该有自身最为擅长解决的问题样例类型。可以说，认识到这种差异性是构造高性能 PAP 的前提条件。然而，在实际应用中找到符合以上条件的算法并不容易，这要求研发人员对各种算法的优势以及短板有充分认识，且需耗费大量时间进行后续调整。这不仅带来了高昂的人力成本，也在无形中提高了 PAP 的研发门槛，阻碍其进一步发展。

为了解决以上问题，本文提出了一种基于$(n+1)$演化优化框架的 PAP 智能汇聚构造方法 AutoPAP，其接受一个算法配置空间和一个问题样例集合作为输入，从前者中自动为 PAP 选择成员算法，以使得 PAP 在后者上的性能达到最优。为了提升 AutoPAP 的实际应用效果，本文结合自动算法配置技术和分而治之策略设计了可以高效搜索算法配置空间的变异算子。进一步地，本文从理论上证明了 AutoPAP 可以达到$(1-1/e)$的近似最优比。

最后，本文以旅行商问题(Traveling Salesman Problems，简称 TSP)为例，验证 AutoPAP 的有效性。具体而言，本文使用 AutoPAP 构造得到 TSP_PAP，并将其和当前 TSP 上性能最佳的求解器 LKH[11]以及 EAX[12]作对比。实验结果表明，在主流 TSP 测试集上，TSP_PAP 无论是在求解效率还是在求解效果上均显著好于 LKH 和 EAX。在 128 个规模 1000 至 30000 的 TSP 测试样例上，相比于 LKH 和 EAX，TSP_PAP 可以将平均求解时间缩短至少 45.71%，并将平均误差率降低至少 87.50%。可以说，TSP_PAP 是当前综合性能最强的 TSP 求解器，这体现出 AutoPAP 在算法自动设计与演化方面的巨大潜力。

本文的后续章节安排如下。第二节分别给出 PAP 以及 PAP 自动构建问题的形式化定义。第三节首先描述 AutoPAP 框架以及变异算子，然后证明 AutoPAP 的近似最优性。第四节以 TSP 问题为例，验证 AutoPAP 的有效性。第五节对全文进行总结。

2 问题定义

2.1 PAP 形式化定义

令P表示 PAP，令k表示P的规模，即成员算法的数量，P的形式化定义如下：

$$P := \{\theta_1, \theta_2, \dots, \theta_k\}, \tag{1}$$

其中θ_i表示P中第i个成员算法。

为了避免引入过多的数学符号，本文以 m(solver, instances)表示在性能指标m下，求解“solver”在问题样例集合“instances”上的性能。注意“solver”可以是单个算法θ，也可以是一个 PAP，而“instances”既可以是单个问题样例(此时可以看作只含有一个样例的集合)，也可以是包含多个问题样例的集合。

当使用P来求解给定问题样例z时，P中所有成员算法，即$\theta_1, \theta_2, \dots, \theta_k$，都将在$z$上相互独立地并行运行，

[1] https://www.ibm.com/analytics/cplex-optimizer

[2] https://www.synopsys.com

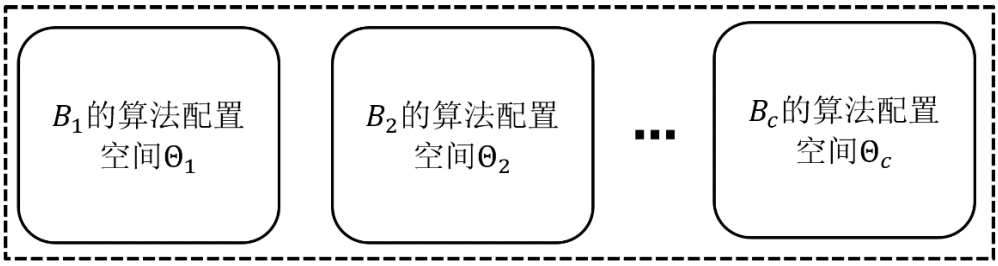

图 1　基础算法$B_1, B_2, \ldots, B_c$ 所定义的算法配置空间Θ

直到达到终止条件。这里的终止条件依赖于待求解的问题和感兴趣的性能指标m。具体而言，假设待求解的问题为判定类(decision)问题(例如 SAT 问题)，那么在 PAP 运行过程中只要任何一个成员算法率先输出了对z的判定结果，即“是”或“否”，所有成员算法都会立即被终止。因此求解z所需的运行时间就是其最佳成员算法求解z所需的运行时间。此外，在这种情况下，通常会引入一个最长运行时间(又叫截止时间) T_{max}以防止 PAP 运行时间过长。如果在T_{max}时间内，没有任何成员算法输出判定答案，那么所有成员算法也会被立即终止，且此次求解被判定为超时(Timeout)或失败(Failure)。

另一方面，假设待求解的问题为优化问题(例如 TSP 问题)，终止条件则依赖于感兴趣的性能指标m。如果m是“找到一个近似最优解(例如，与最优解的差距不超过预先给定的阈值)所需的运行时间”，那么只要任何一个成员算法率先找到了这类解，所有成员算法都会立即被终止。与考虑决策问题时一样，在这种情况下可以引入最长运行时间T_{max}以防止 PAP 运行时间过长。如果m是“在给定时间T_{max}内找到的解的质量”，那么每个成员算法都将在运行T_{max}时间后终止，最后P将成员算法找到的所有的解中的最好的那个作为输出。

可以看到，无论考虑何种问题类型和何种性能指标，P在问题样例z上的性能$\mathrm{m}(P,z)$总是其成员算法$\theta_1, \ldots, \theta_k$在$z$上取得的最好性能，即：

$$\mathrm{m}(P,z) = \max_{\theta \in P} \mathrm{m}(\theta, z)。 \quad (2)$$

不失一般性，假设对性能指标m而言，值越大越好。值得注意的是，实际中考虑优化问题且m是“找到一个近似最优解所需的运行时间”时，我们往往并不知道待求解问题的真正最优解，也就无法计算成员算法找到的解与最优解的距离，从而无法判断是否是近似最优解，最终导致无法测量算法的运行时间。但是，这也不会影响等式(2)的成立。进一步地，P在问题样例集合I上的性能是P在I中各个问题样例上的性能的平均值：

$$\mathrm{m}(P,I) = \frac{1}{|I|}\sum_{z \in I} \mathrm{m}(P,z)\,, \quad (3)$$

其中$|I|$表示集合I的势。

2.2　PAP 自动构建问题

为了将 PAP 构建过程自动化，本文考虑如下的优化问题：从给定算法配置空间Θ中选择k个成员算法组成P，以使得后者在给定问题样例集合I上的性能达到最优。

其中，I被称作训练集，其应当充分代表P的目标使用场景。换言之，如果P在I上性能良好，那么可以推断P在实际使用时也具有良好的性能。算法配置空间Θ由一组参数化算法$B_1, B_2, \ldots, B_c$所定义，方便起见，我们称这类算法为基础算法。其中每个基础算法B_i都有若干参数，参数的取值控制着B_i方方面面的行为，进而对B_i的性能有着巨大的影响。因此，当B_i的参数取不同值时，可以认为得到了不同的算法；在此基础上，对B_i的参数取遍所有可能的值，最终得到的所有独一无二的算法所组成的集合，就是B_i的算法配置空间，记为Θ_i。例如，假设B_1有两个参数α和β，其各有两种取值，$\alpha \in \{0,1\}$，$\beta \in \{0,1\}$；那么Θ_1一共包含 4 个不同的算法，即$\Theta_1 = \{(\alpha=0,\beta=0),(\alpha=0,\beta=1),(\alpha=1,\beta=0),(\alpha=1,\beta=1)\}$。如图 1 所示，多个基础算法$B_1, \ldots, B_c$所定义的算法配置空间$\Theta = \Theta_1 \cup \Theta_2 \cup \ldots \cup \Theta_c$。例如，假设现在除了上面例子中的$B_1$，还考虑另一个基础算法$B_2$，其有一个参数$\gamma \in \{0,1\}$，因此$\Theta_2 = \{(\gamma=0),(\gamma=1)\}$，最终$\Theta = \Theta_1 \cup \Theta_2 = \{(\alpha=0,\beta=0),(\alpha=0,\beta=1),(\alpha=1,\beta=0),(\alpha=1,\beta=1),(\gamma=0),(\gamma=1)\}$。

如前所述，P中每个成员算法θ_i都是从Θ中选择

得到，因此θ_i的搜索空间大小为$|\Theta|$。此外，由等式(2)可知，假设P已经包含成员算法θ_i，那么再将同样的算法θ_i加入到P则不会带来任何性能增益。因此，在为P寻找新成员算法时，可以简单地先将P已包含的成员算法排除掉，这可以保证最终得到的P中各个成员算法互不相同。实际上，P可以看作Θ的子集，那么P的整体搜索空间大小为$\prod_{i=1}^{k}(|\Theta|-i+1)=O(|\Theta|^k)$。综上所述，PAP 自动构建问题的形式化定义如下：

$$P^* = \operatorname{argmax}_{P\subseteq\Theta \text{ 且 } |P|=k} \mathrm{m}(P, I)\,, \tag{4}$$

其中P^*表示最优 PAP。

3 PAP 智能汇聚构造方法

3.1 AutoPAP 框架

本文提出了一种基于演化优化的方法 AutoPAP 来解决上文中的 PAP 自动构建问题。演化算法是一类具有广泛适用性的全局优化方法[13,24]，其通过将种群进行交叉(crossover)、变异(mutation)等操作，并在一定的控制参数(如演化代数)作用下，于解空间中进行搜索。值得注意的是，传统演化算法演化的每个个体通常都是待求解问题的完整解；而在 AutoPAP 中，种群(而非个体)才是完整解，对应着 PAP，种群中的个体则对应着 PAP 的成员算法。也正是基于此，AutoPAP 和传统演化算法有一个明显区别：对于后者，在搜索过程中种群规模通常保持不变；而在 AutoPAP 中，种群(PAP)规模是逐渐增大的，这可以保证 PAP 的性能不会出现退化(详细分析请见 3.3 节引理 1)。AutoPAP 的具体步骤如下：

输入：算法配置空间Θ，训练集I，性能指标m，PAP 规模k，每一代生成的候选算法数量n

输出：最终种群$P=\{\theta_1,\theta_2,\dots,\theta_k\}$

步骤 1 初始化当前迭代计数变量$i\leftarrow 1$；初始化种群$P\leftarrow\emptyset$。

步骤 2 使用变异算子在Θ中搜索生成n个候选算法$\theta_1',\dots,\theta_n'$。

步骤 3 将$\theta_1',\dots,\theta_n'$在训练集$I$上进行测试，找到对当前$P$具有最大性能增益的算法：

$$\theta_i = \operatorname*{argmax}_{j\in\{1,\dots,n\}} \mathrm{m}\left(P\cup\theta_j', I\right)。 \tag{5}$$

步骤 4 将θ_i加入到 PAP 中：$P\leftarrow P\cup\theta_i$。

步骤 5 如果$i=k$，算法结束并返回P，否则$i\leftarrow i+1$，并重复步骤 2-5。

可以看到，AutoPAP 一共有k次迭代，遵循$(n+1)$演化优化框架。在每一次迭代中，AutoPAP 使用变异算子从算法配置空间Θ中搜索得到n个候选算法，然后保留其中对当前 PAP 性能提升最大的那个算法加入到P中。值得注意的是，该算法有可能无法给P带来性能增益，即$\mathrm{m}(P\cup\theta_i)=\mathrm{m}(P)$，这种情况下$\theta_i$仍然会被加入到$P$中。这样做的原因主要有两点。首先，加入$\theta_i$绝对不会降低$P$的性能(请见 3.3 节引理 1)；其次，虽然在训练集上θ_i没有给P带来性能增益，但在训练集没有覆盖到的问题样例上(例如实际使用场景中出现的问题样例)，θ_i有可能会给P带来性能增益。

随着 AutoPAP 迭代次数增加，P的规模逐步增大。显然，变异算子在Θ中的搜索效率将在很大程度上决定P的最终性能。下一节将会详细介绍 AutoPAP 中的变异算子。AutoPAP 的计算代价主要由两部分组成。第一部分是在步骤 2 中使用变异算子生成候选算法的计算代价。令t_c表示生成一个算法的时间，那么这一部分的 CPU 时间为nt_c。第二部分则是在步骤 3 中对候选算法进行测试的计算代价。令测试一个算法的时间为t_v，那么这一部分的 CPU 时间为nt_v。由此，AutoPAP 一次迭代所消耗的 CPU 时间为$n(t_c+t_v)$，总 CPU 时间则为$kn(t_c+t_v)$。

3.2 变异算子

如 2.2 节所述，Θ是一组参数化算法(称作基础算法)$B_1,B_2,\dots,B_c$的联合配置空间，即$\Theta=\Theta_1\cup\Theta_2\cup\dots\cup\Theta_c$且$|\Theta|=\sum_{i=1}^{c}|\Theta_i|$。换言之，$|\Theta|$将会随着基础算法的数量线性增长。当基础算法本身的配置空间较大，或者基础算法数量较多时，$|\Theta|$会变得非常巨大，在这种情况下直接对Θ进行搜索效率很低。其次，Θ可能会涉及不同类型的参数，包括浮点型(例如演化算法的变异速率)、整数型(例如演化算法的种群规模)以及范畴型(例如演化算法的种群初始化方式)，这要求变异算子能同时处理各种类型的决策变量。

本文使用了基于模型的顺序采样技术(Sequential Model-based Algorithm Configuration，简称 SMAC[16-17])来实现变异算子。具体而言，变异算子调用 SMAC 从Θ中寻找一个能够最大程度提高当前 PAP 性能的算法，作为最终生成的候选算法。在具体搜索过程中，SMAC 会建立回归模型(Regression Model)以预测给定算法配置的性能，然后在该模型基础之上最大化特定的获取函数(Acquisition Function)来决定下一个算

法配置空间中的采样点。因为使用了随机森林来建立回归模型，SMAC 可以处理浮点型、整数型、范畴性甚至是序数型(Ordinal)变量。感兴趣的读者可以查阅文献[16]以进一步了解 SMAC。本文使用了 SMAC 的开源版本 v3，可在 https://github.com/automl/SMAC3 获取到。

为了解决$|\Theta|$过大而难以高效搜索的问题，本文采用了分而治之的策略。具体而言，考虑到不同基础算法$B_1,\dots,B_c$的配置空间互相没有交集，因此可以将Θ自然地分解为c个子空间$\Theta_1,\dots,\Theta_c$，并在每个子空间内部使用 SMAC 进行搜索。当$c\geq 2$时，子空间远小于原空间，这可以大幅降低搜索难度。进一步地，在c个子空间内部分别搜索最终将会得到c个新算法，而 AutoPAP 在每一代都要生成n个新算法。因此，可以将这n个新算法均匀地分配到不同子空间内部的 SMAC 搜索进程上。例如，假设$n=10$，$c=2$，那么$\theta_1',\theta_3',\theta_5',\theta_7',\theta_9'$由 SMAC 在$\Theta_2$中搜索得到，$\theta_2',\theta_4',\theta_6',\theta_8',\theta_{10}'$则由 SMAC 在$\Theta_1$中搜索得到。

综上所述，AutoPAP 中变异算子的具体步骤如下：

输入： 算法配置空间$\Theta=\Theta_1\cup\dots\cup\Theta_c$，训练集$I$，性能指标$m$，每一代生成的候选算法数量$n$，当前$P$

输出： n个候选算法$\theta_1',\dots,\theta_n'$

步骤 1 初始化算法标记变量$i\leftarrow 1$。

步骤 2 $j\leftarrow(i \bmod c)+1$。

步骤 3 使用 SMAC 在Θ_j中搜索$\theta_i'=\underset{\theta\in\Theta_j}{\operatorname{argmax}}\,\mathrm{m}(P\cup\theta,I)$。

步骤 4 如果$i=n$，算法结束并返回$\theta_1',\dots,\theta_n'$，否则$i\leftarrow i+1$，并重复步骤 2-4。

在变异算子的步骤 2 中，mod表示取模运算。在步骤 3 中，SMAC 的优化目标是$\max\limits_{\theta\in\Theta_j}\mathrm{m}(P\cup\theta,I)$。当运行时间$t_c$趋向于无穷时，SMAC 以概率 1 找到最优算法；在实际中，t_c不可能为无穷大，因此 SMAC 最终返回的往往是近似最优算法。值得注意的是，只有当n不小于c时才能保证每个子空间都会被变异算子搜索至少一次，因此在使用 AutoPAP 时需设置$n\geq c$。

3.3　理论分析

本节从理论上分析 AutoPAP 在式(4)所定义的 PAP 自动构造问题上所能达到的近似比。简便起见，以$f(P)$表示$\mathrm{m}(P,I)$，首先证明如下引理 1 和引理 2。

引理 1. 式(4)中的目标函数f具有单调性，即对于任意$A\subseteq B\subseteq\Theta$，满足$f(A)\leq f(B)$。

证明： 根据等式(2)，对于任意问题样例z，以下不等式成立。

$$\mathrm{m}(B,z)=\max\{\mathrm{m}(A,z),\mathrm{m}(B\backslash A,z)\}\geq\mathrm{m}(A,z)$$

进一步地，结合等式(3)，以下不等式成立。

$$\begin{aligned}f(B)-f(A)&=\mathrm{m}(B,I)-\mathrm{m}(A,I)\\&=\frac{1}{|I|}\sum_{z\in I}[\mathrm{m}(B,z)-\mathrm{m}(A,z)]\\&\geq 0\end{aligned}$$

证明完毕。 □

引理 2. 式(4)中的目标函数f具有子模性，即对于任意$A\subseteq B\subseteq\Theta$和任意$\theta\in\Theta\setminus B$，满足

$$f(A\cup\{\theta\})-f(A)\geq f(B\cup\{\theta\})-f(B)。\quad(6)$$

证明： 根据等式(2)，对于任意问题样例z，以下等式成立。

$$\mathrm{m}(A\cup\{\theta\},z)=\max\{\mathrm{m}(A,z),\mathrm{m}(\theta,z)\}$$
$$\mathrm{m}(B\cup\{\theta\},z)=\max\{\mathrm{m}(A,z),\mathrm{m}(B\backslash A),\mathrm{m}(\theta,z)\}$$

考虑以下三种不同情况。

情况(1)：$\mathrm{m}(\theta,z)<\mathrm{m}(A,z)$，那么如下等式成立。

$$\mathrm{m}(A\cup\{\theta\},z)=\mathrm{m}(A,z)$$
$$\mathrm{m}(B\cup\{\theta\},z)=\mathrm{m}(B,z)$$

由此得到$f(A\cup\{\theta\})-f(A)=f(B\cup\{\theta\})-f(B)=0$。

情况(2)：$\mathrm{m}(B,z)\geq\mathrm{m}(\theta,z)\geq\mathrm{m}(A,z)$，那么如下等式成立。

$$\mathrm{m}(A\cup\{\theta\},z)=\mathrm{m}(\theta,z)$$
$$\mathrm{m}(B\cup\{\theta\},z)=\mathrm{m}(B,z)$$

由此得到$f(A\cup\{\theta\})-f(A)\geq 0=f(B\cup\{\theta\})-f(B)$。

情况(3)：$\mathrm{m}(B,z)<\mathrm{m}(\theta,z)$，那么如下等式成立。

$$\mathrm{m}(A\cup\{\theta\},z)=\mathrm{m}(\theta,z)$$
$$\mathrm{m}(B\cup\{\theta\},z)=\mathrm{m}(\theta,z)$$

进一步地，可以得到如下不等式。

$$\begin{aligned}\mathrm{m}(A\cup\{\theta\},z)-\mathrm{m}(A,z)&=\mathrm{m}(\theta,z)-\mathrm{m}(A,z)\\&\geq\mathrm{m}(\theta,z)-\mathrm{m}(B,z)\\&=\mathrm{m}(B\cup\{\theta\},z)-\mathrm{m}(B,z)\end{aligned}$$

这表明$f(A\cup\{\theta\})-f(A)\geq f(B\cup\{\theta\})-f(B)$。

证明完毕。 □

直观上，引理 1 意味着随着 PAP 的规模增大，其性能单调非递减。引理 2 则意味着对于 PAP 来说，添加成员算法所带来的性能增益具有如下的边际递减效应：将成员算法θ添加到A的性能收益总是大于或等于将θ添加到A的超集B的性能收益。简便起见，将所谓的性能收益记为：

$$\Delta(\theta|A):=f(A\cup\{\theta\})-f(A)。\quad(7)$$

表 2　实验所采用的 TSP 问题样例在各问题规模区间上的分布情况

规模区间 / 类型	[1000,5000]	[5001,10000]	[10001,15000]	[15001,20000]	[20001,25000]	[25000,30000]	总计
TSPlib	23	3	3	2	0	0	31
National	4	11	2	0	2	0	19
VLSI	48	7	3	4	4	6	72
DIMACS	0	0	6	0	0	0	6
总计	75	21	14	6	6	6	128

下面的定理 1 表明了 AutoPAP 在理论上可以达到$(1-1/e)$近似最优构造效果。为了证明定理 1，定义P_i为 AutoPAP 运行过程中第i次迭代结束时的 PAP，即$P_i=\{\theta_1,\dots,\theta_i\}$，并令$P^*=\{\theta_1^*,\theta_2^*,\dots,\theta_k^*\}$表示最优 PAP，即$P^*=\operatorname{argmax}_{P\subseteq\Theta\ 且|P|=k} f(P)$。

定理 1. 当$t_c\to\infty$，且$n\geq c$时，对于任意正整数i和k，满足$f(P_i)\geq(1-e^{-i/k})f(P^*)$。特别地，当$i=k$时，$f(P_k)\geq(1-1/e)f(P^*)$。

证明： 如前所述，在 AutoPAP 的每次迭代中，$t_c\to\infty$可以保证每个子配置空间中对当前 PAP 性能提升最大的算法被找到，$n\geq c$则可以保证每个子配置空间都会被变异算子搜索至少一次。综上，AutoPAP 在第i次迭代最终找到的算法θ_i满足$\theta_i=\operatorname{argmax}_{\theta\in\Theta}\Delta(\theta|P_{i-1})$。

因此，如下不等式成立。

$$f(P^*)\leq f(P^*\cup P_i)\qquad\text{引理 1}$$
$$=f(P_i)+\sum_{j=1}^{k}\Delta\left(\theta_j^*\mid P_i\cup\{\theta_1^*,\theta_2^*,\dots,\theta_{j-1}^*\}\right)$$
$$\leq f(P_i)+\sum_{\theta\in P^*}\Delta(\theta\mid P_i)\qquad\text{引理 2}$$
$$\leq f(P_i)+\sum_{\theta\in P^*}\Delta(\theta_{i+1}\mid P_i)$$
$$=f(P_i)+k\Delta(\theta_{i+1}\mid P_i)$$

重新整理上式，得到以下不等式。

$$\Delta(\theta_{i+1}\mid P_i)\geq\frac{1}{k}\left(f(P^*)-f(P_i)\right)\qquad(8)$$

定义$\delta_i:=f(P^*)-f(P_i)$，这意味着如下等式成立。

$$\delta_i-\delta_{i+1}=f(P_{i+1})-f(P_i)=\Delta(\theta_{i+1}|P_i)\qquad(9)$$

将式(9)代入前面不等式(8)，得到以下不等式。

$$\delta_i-\delta_{i+1}\geq\frac{\delta_i}{k}\qquad(10)$$

将不等式(10)重新整理，并递归应用，得到下式。

$$\delta_i\leq\left(1-\frac{1}{k}\right)^i\delta_0\leq\left(1-\frac{1}{k}\right)^i[f(P^*)-f(\emptyset)]$$
$$\leq\left(1-\frac{1}{k}\right)^i f(P^*)$$

基于$\left(1-\frac{1}{k}\right)\leq e^{-1/k}$，得到$\delta_i\leq e^{-i/k}f(P^*)$。因为$\delta_i=f(P^*)-f(P_i)$，那么下式成立。

$$f(P_i)\geq\left(1-e^{-i/k}\right)f(P^*)。$$

证明完毕。 □

4　实验验证

本文选择以 TSP 问题为例，验证 AutoPAP 的有效性。TSP[16]是著名的 NP 难问题，其描述如下：给定若干城市，寻找一条遍历所有城市的最短路径。TSP 在物流、通讯、电路设计等领域有着广泛应用[17]，多年来一直吸引着众多研究人员对其求解器进行持续改进。目前，学术界公认的最佳 TSP 求解器是采用 Lin-Kernighan[25]启发式的 LKH[11,26]以及使用边组装交叉算子的 EAX[12,27]。本文将使用 AutoPAP 针对 TSP 问题构造 PAP(记为 TSP_PAP)，然后将 TSP_PAP 和这些求解器进行比较。

4.1　问题样例集

为了尽可能全面地测试求解器的性能，本文从 TSP 基准集网站 http://www.math.uwaterloo.ca/tsp 收集了 4 种类型共 128 个规模 1000 至 30000 的 TSP 问题样例。这四种类型分别是:

a) TSPlib[17]，来自实际应用，共 31 个样例；
b) National TSP，从各国地图抽象得到，共 19 个样例；
c) VLSI TSP，来自超大规模集成电路设计场景，共 72 个样例；
d) DIMACS Test Set[16]，曾在 DIMACS TSP 求解挑战中作为基准测试集，共 6 个样例。

表 1 总结了这些样问题例在各规模区间和各问题类型上的分布情况。一般而言，当 TSP 问题规模

超过 1000 时，就被认为是中等规模 TSP，而超过 5000 时则被认为是大规模 TSP。可以看到本文所采用的 TSP 样例全面覆盖了小、中、大三种规模。此外，这些问题样例也覆盖了多种 TSP 距离类型，包括定义在二维、三维、四维欧式距离上的 EUC_2D、EUC_3D、EUC_4D 类型，直接以距离矩阵给出的 MATRIX 类型，以及地理距离类型 GEO，感兴趣的读者可以查询 TSPlib 文献[17]获取这些距离的详细定义。值得注意的是，TSP_PAP 适用的 TSP 距离类型是由其基础算法 LKH 和 EAX 决定的。因此，和 LKH 以及 EAX 一样，TSP_PAP 对于以上所有距离类型都适用。进一步地，本文收集了这些 TSP 样例的最优解作为评判标准以衡量求解器找到的解的好坏。由于部分 TSP 样例的最优解目前仍然未知，对于这些样例本文使用当前已知最好的解(Best Known Solution)作为最优解。

4.2　性能指标

本文使用了两种性能指标来衡量求解器的性能。其中，带惩罚的平均运行时间(Penalized Average Runtime-10，简称 PAR10)衡量了求解器找到 TSP 样例最优解所需要的平均时间(即求解效率)，值越小越好。在实际操作中，被测试的求解器将会被限定最长运行时间$T_{max}=3600$秒。如果在运行了$t\le T_{max}$时求解器成功找到了问题样例的最优解，那么此次运行视为“成功”，且时间记录为t；否则此次运行视为“超时”，时间记录为$10T_{max}$。最终，将求解器在所有样例上所需运行时间的平均值记为 PAR10。

此外，本文使用了平均误差率(Average Deviation Ratio，简称 ADR)来衡量求解器找到的解的质量(即求解效果)，值越小越好。在实际操作中，被测试的求解器将会被限定最长运行时间$T_{max}=3600$秒，当运行T_{max}后，假定求解器返回的解的代价为$cost$，而问题样例的最优解的代价为$cost^*$，则计算

$$DR=\frac{cost-cost^*}{cost^*}。\tag{11}$$

最终，将求解器在所有样例上获得的 DR 的平均值记为 ADR。

4.3　构造 TSP_PAP

如前所述，AutoPAP 的输入为训练集I和算法配置空间Θ，后者由若干基础算法定义。本文从 128 个 TSP 样例中随机抽取了 50 个样例构成训练集I，并选择了 LKH[11] (版本 2.0.9)和 EAX[12] (版本 1.0)作为 AutoPAP 的基础算法。LKH 一共有 29 个参数，EAX 一共有 8 个参数。

考虑到如今 4 核 CPU 已经相当常见，本文将 AutoPAP 的参数k(成员算法数量)设置为 4，n(每一代生成候选算法的数量)设置为 10。AutoPAP 优化的性能指标m为 PAR10，生成单个算法的时间t_c为 24 小时，测试单个算法的时间t_v为 6 小时。如 3.1 节所述，AutoPAP 消耗的总 CPU 时间为$kn(t_c+t_v)$，因此本实验中 AutoPAP 一共使用了 1200CPU 小时(共 50 CPU 天)构造 TSP_PAP。另一方面，由于 AutoPAP 的每一次迭代中变异算子生成 10 个新算法并测试的过程是相互独立的，因此可以并行运行，那么在 CPU 核数大于等于 10 的计算平台上，TSP_PAP 的构造时长可以缩短为$k(t_c+t_v)=5$ 天。实验中所用计算平台为 Intel Xeon CPU E5-2699A v4 @ 2.40GHz，22 核心，55MB 高速缓存。

4.4　基线求解器

本文考虑了如下 TSP 上公认性能最好的两个求解器作为基线求解器：

a) LKH[11]，本文所考虑的基础求解器之一，使用默认参数配置；

b) EAX[12]，本文所考虑的基础求解器之一，使用默认参数配置。

考虑到 LKH 和 EAX 的性能会受其参数配置影响，本文进一步使用 SMAC 对 LKH 和 EAX 分别进行自动算法配置，其中训练集和总 CPU 时间与构造 TSP_PAP 时保持一致：

c) LKH-TUNED，自动算法配置得到的 LKH；

d) EAX-TUNED，自动算法配置得到的 EAX。

考虑到 TSP_PAP 包含了 4 个算法，而以上各求解器实际上仅为一个算法，因此我们人为地将以上 4 个算法进一步组成 PAP，得到：

e) EAX_LKH，由以上 4 个算法组成。

此外，为了验证 AutoPAP 中分而治之策略的有效性，还得到了如下 PAP：

f) TSP_PAP∗，在构造该 PAP 过程中使用变异算子直接搜索整个算法配置空间，而非子配置空间。

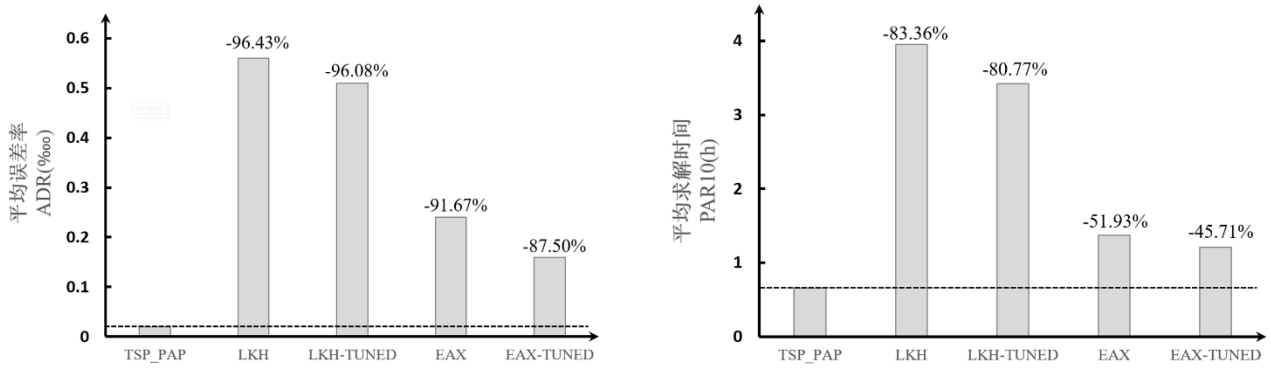

图 3　TSP_PAP 以及 LKH、EAX 的平均求解时间和平均误差率，百分数表示 TSP_EAX 相比于对应求解器的提升幅度

表 2　各求解器在 TSP 样例集上的超时数量(#Timeous)，PAR10 以及 ADR

	TSP_PAP	LKH	EAX	LKH-TUNED	EAX-TUNED	EAX_LKH	TSP_PAP*
#Timeouts	**7**	50	16	43	14	9	9
PAR10(s)	**2369.38**	14242.85	4928.88	12319.73	4364.25	2921.51	2980.09
ADR(‰)	**0.02**	0.56	0.24	0.51	0.16	0.12	0.39

4.5　实验结果与分析

在每个 TSP 样例上，每个求解器被重复运行 3 次(每次使用不同的随机种子)，3 次运行结果的中位数作为该求解器在这个样例上的结果，在此基础上计算出每个求解器在整个 TSP 样例集上的超时数(超时意味着运行 3600 秒仍未找到问题样例的最优解)，PAR10 和 ADR。表 2 展示了各求解器的测试结果，其中每个性能指标上的最佳结果以“_”标记。

首先，TSP_PAP 在所有求解器中获得了最少的超时数，最短的运行时间和最小的平均误差率，这意味着无论是在求解效率还是在求解效果上，TSP_PAP 都显著好于其余基线求解器。图 2 直观对比了 TSP_PAP 和 LKH 以及 EAX。可以看到，相对于 LKH 和 EAX，TSP_PAP 可以将平均求解时间大幅缩短至少 45.71%，而在平均误差率上的降低幅度更为巨大，达到了至少 87.50%。注意以上结果是在已经对 LKH 和 EAX 进行了算法配置的基础之上达成的，这说明 TSP_PAP 已经成为了 TSP 上综合性能最强的求解器，也验证了 AutoPAP 作为 PAP 构造方法的有效性。

相比于 TSP_PAP*，TSP_PAP 将超时数减少了 28.57%，将 PAR10 缩短了 20.49%，将 ADR 降低了 94.87%，这说明 AutoPAP 采取的分而治之策略确实能够让变异算子更加高效地搜索算法配置空间，从而构造出性能更强的 PAP。此外，TSP_PAP 相比于 EAX_LKH 也取得了类似幅度的性能提升，考虑到后者是人为将已有 TSP 求解器组合而成，这表明了 AutoPAP 所遵循的“PAP 自动构造”方法论相对于“手工构造 PAP”有性能优势。

虽然 EAX 和 LKH 都是目前公认最佳的 TSP 求解器，但在表 2 中前者的表现比后者要好得多，这与近期文献[18-19]中的结果相符。此外，经过自动算法配置之后，EAX 和 LKH 的性能都有了小幅提升，这表明参数取值确实会影响求解器的性能。

我们进一步分析了在不同问题规模区间上 TSP_PAP 的性能。对于规模不大于 5000 的中小 TSP 样例，TSP_PAP 的 PAR10 为 40.75 秒，且 ADR 为 0(即全部都找到了最优解)。相比较而言，EAX_LKH 在中小规模 TSP 样例上的 ADR 为 0.03‰，PAR10 则为 109.57 秒，后者是 TSP_PAP 所需时间的 2.7 倍。随着 TSP 规模变大，问题解空间指数级增长，TSP_PAP 所需求解时间也显著增长。在规模大于 5000 的 TSP 样例上，TSP_PAP 的 PAR10 为 5664.61 秒，而 EAX_LKH 则为 6900.67 秒。后者为前者的 1.2 倍。

最后，可以看到表 2 中所有 PAP 型求解器

(TSP_PAP，EAX_LKH 以及 TSP_PAP*)的表现都远好于非 PAP 型求解器，这表明了 PAP 这种求解器形式所具有的巨大优越性。

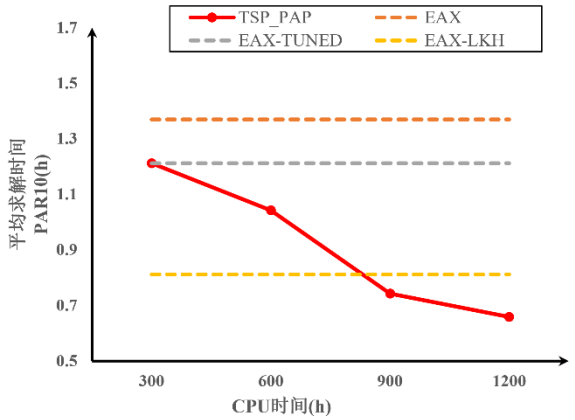

图 3　随着 AutoPAP 消耗 CPU 时间增多，TSP_PAP 的性能(PAR10)变化趋势

4.6　TSP_PAP 性能变化趋势

图 3 展示了在 AutoPAP 运行过程中每一代结束时的 PAP(P_1, P_2, P_3, P_4)在 TSP 样例集上的 PAR10 测试结果。注意图中的横坐标以 AutoPAP 消耗的 CPU 时间给出。可以看到，随着迭代次数增多，PAP 的性能单调提升，这符合引理 1。此外，AutoPAP 消耗 300 个 CPU 小时生成的P_1(仅包含一个成员算法)和 EAX-TUNED 性能相近，而消耗 900 个 CPU 小时生成的P_3(包含 3 个成员算法)的性能则超过了同为 PAP 的 EAX_LKH。AutoPAP 消耗了 1200 个 CPU 小时生成了P_4，从图 3 中可以看出此时 PAR10 下降趋势仍未停止。可以预想的是，随着 AutoPAP 迭代次数的进一步增加，TSP_PAP 的性能将进一步提升。

5　总结

本文提出了一种基于演化优化的 PAP 智能汇聚构造方法 AutoPAP，其可以基于给定算法配置空间和训练集自动构造出高性能 PAP。AutoPAP 在理论上可以达到$(1-1/e)$近似最优构造效果，其针对 TSP 问题所构造的 TSP_PAP 的性能大幅超过了 EAX 以及 LKH，成为了综合性能最强的 TSP 求解器。

TSP_PAP 的强悍性能充分展示了 AutoPAP 的巨大应用价值。事实上，只需提供相应算法配置空间和训练集，AutoPAP 可以被用于针对几乎任何问题类构建 PAP。本文下一步工作便是将 AutoPAP 应用到其它问题类例如 XX 问题[]，车辆路径规划问题[20]和交通拥堵控制[23]。

另一个重要的研究方向是对 AutoPAP 进行改进。当前 AutoPAP 的局限性主要体现在两方面。首先，由引理 2 可知，随着 PAP 成员算法增多，对 PAP 添加更多成员算法所带来的性能增益呈现边际递减效应，这表明当前 AutoPAP 有造成“计算资源浪费”的风险：即对 PAP 性能增益有限的成员算法和对 PAP 性能增益明显的成员算法使用了同样的计算资源。造成这一现象的根本原因在于 AutoPAP 当前仅能添加成员算法，而不涉及删除、替换等操作，未来可将这些操作引入 AutoPAP 以提高其灵活性。AutoPAP 的第二个局限性在于当前变异算子的分而治之策略仅对基础算法数量大于 2 的情况有效。当基础算法数量为 1 时，可尝试将该算法的配置空间进行分解。具体而言，可首先分析找出对该算法行为有决定性影响的参数，然后将该参数的取值范围均分为不相交的若干集合，从而得到分解方案。在此基础之上，即可按照当前策略在各子空间中并行搜索以达到加速效果。最后，还可以借助演化计算中较为成熟的多种群技术[21]、负相关搜索技术[22]等，进一步提高 AutoPAP 中变异算子对算法配置空间的搜索效率。

参考文献

1　Asanovic K, Bodík R, Demmel J, et al. A view of the parallel computing landscape. Commun ACM, 2009, 52: 56-67.

2 Biere A, Fazekas K, Fleury M, et al. CaDiCaL, Kissat, Paracooba, Plingeling and Treengeling entering the SAT Competition 2020. In: Proceedings of SAT Competition 2020: Solver and Benchmark Descriptions. University of Helsinki, 2020. 50-53.
3 Gottlob G, Okulmus C, Pichler R. Fast and parallel decomposition of constraint satisfaction problems. In: Proceedings of the 29th International Joint Conference on Artificial Intelligence. AAAI Press, 2020. 1155-1162.
4 Gebser M, Kaufmann B, Neumann A, et al. clasp : A conflict-driven answer set solver. In: Proceedings of the 9th International Conference on Logic Programming and Nonmonotonic Reasoning. Springer, 2007. 260-265.
5 Ralphs T K, Shinano Y, Berthold T, et al. Parallel solvers for mixed integer linear optimization. Handbook of Parallel Constraint Reasoning. Springer, 2018: 283-336.
6 Tang K, Peng F, Chen G, et al. Population-based algorithm portfolios with automated constituent algorithms selection. Inf Sci, 2014, 279:94-104.
7 Hamadi Y, Wintersteiger C M. Seven challenges in parallel SAT solving. AI Mag, 2013, 34: 99-106.
8 Liu S, Tang K, Yao X. Automatic construction of parallel portfolios via explicit instance grouping. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence. AAAI Press, 2019. 1560–1567.
9 Liu S, Tang K, Yao X. Generative adversarial construction of parallel portfolios. IEEE Trans Cybern, 2021, to be published, DOI:10.1109/TCYB.2020.2984546.
10 Tang K, Liu S, Yang P, et al. Few-shots parallel algorithm portfolio construction via co-evolution. IEEE Trans Evol Comput, 2021, 25: 595-607.
11 Helsgaun K. General k-opt submoves for the Lin-Kernighan TSP heuristic. Math Program Comput, 2009, 1: 119-163.
12 Nagata Y, Kobayashi S. A powerful genetic algorithm using edge assembly crossover for the traveling salesman problem. INFORMS J Comput, 2013, 25:346-363.
13 Kenneth D J. Evolutionary computation: a unified approach. MIT Press, 2006.
14 Hutter F, Hoos H H, Leyton-Brown K. Sequential model-based optimization for general algorithm configuration. In: Proceedings of the 5th International Conference on Learning and Intelligent Optimization. Springer, 2011. 507-523.
15 Liu S, Tang K, Yao X. On performance estimation in automatic algorithm configuration. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence. AAAI Press, 2020. 2384–2391.
16 Gutin G, Punnen A P. The traveling salesman problem and its variations. Springer, 2006.
17 Gerhard R. TSPLIB—A traveling salesman problem library. INFORMS J Comput, 1991, 3: 376-384.
18 Taillard É D, Helsgaun K. POPMUSIC for the travelling salesman problem. Eur J Oper Res, 2019, 272: 420-429.
19 Yuichi N. High-order entropy-based population diversity measures in the traveling salesman problem. Evol Comput, 2020, 28: 595-619.
20 Liu S, Tang K, Yao Xin. Memetic search for vehicle routing with simultaneous pickup-delivery and time windows. Swarm Evol Comput, 2021, to be published, DOI: 10.1016/j.swevo.2021.100927.
21 Ma H, Shen S, Yu M, et al. Multi-population techniques in nature inspired optimization algorithms: A comprehensive survey. Swarm Evol Comput, 2019, 44: 365-387.
22 Tang K, Yang P, Yao X. Negatively correlated search. IEEE J Sel Areas Commun, 2016, 34: 542-550.
23 张敖木翰, 高自友, 任华玲. 突发事故下交通拥堵控制策略. 中国科学:技术科学, 2011, 41: 955-961.
24 何振亚, 刘琚, 杨绿溪,等. 盲均衡和信道参数估计的一种 ICA 和进化计算方法. 中国科学 E 辑:技术科学, 2000, 2: 47-54.
25 Lin S, Kernighan B W. An effective heuristic algorithm for the traveling-salesman problem. Oper Res, 1973, 21: 498-516.
26 Nagata Y, Kobayashi S. Edge assembly crossover: A high-power genetic algorithm for the travelling salesman problem. In: Proceedings of the 7th International Conference on Genetic Algorithms. Morgan Kaufmann, 1997. 450-457.
27 Helsgaun K. An effective implementation of the Lin-Kernighan traveling salesman heuristic. Eur J Oper Res, 2000, 126: 106-130.

Approximately Optimal Construction of Parallel Algorithm Portfolios by Evolutionary Intelligence

LIU ShengCai[1,2], Yang Peng[1,2], Tang Ke[1,2*]

[1] *Department of Computer Science and Technology, Southern University of Science and Technology, Shenzhen 518055, China;*
[2] *Guangdong Key Laboratory of Brain-Inspired Intelligent Computation, Shenzhen 518055, China*
[*] *E-mail: tangk3@sustech.edu.cn*

As a high-performance general-purpose form of parallel solvers, Parallel Algorithm Portfolios (PAPs) have shown great performance when solving decision, counting, continuous and discrete optimization problems. However, the manual construction of PAPs still remains a laborious work, which typically requires plenty of domain knowledge and human effort. To address this issue, this paper proposes AutoPAP, an evolutionary intelligence-based method for PAP construction. Overall, AutoPAP follows the $(n+1)$ evolutionary framework, i.e., in each generation n candidate algorithms are generated and the best one among them is retained to be included into the PAP. Considering that the algorithm configuration space is often very large, this paper designs a highly-effective mutation operator to improve the practical performance of AutoPAP and further theoretically proves that AutoPAP can achieve $(1-1/e)$-approximation. Finally, to validate the effectiveness of AutoPAP, this paper uses it to build a PAP, namely TSP_PAP, for the Traveling Salesman Problem (TSP). The testing results show that TSP_PAP significantly outperforms the state-of-the-art TSP solvers EAX and LKH in both efficiency (runtime) and effectiveness (solution quality). On 128 TSP instances of sizes ranging from 1000 to 30000, compared to LKH and EAX, TSP_PAP can reduce the average runtime by at least 45.71% and lower the average deviation ratio by at least 87.50%, indicating the huge potential of AutoPAP in automatic algorithm design and evolution.

Problem Solver, Parallel Algorithm Portfolios, Evolutionary Computation, Automatic Algorithm Design, Traveling Salesman Problem, Combinatorial Optimization

doi: xx.xx