

启发式算法设计中的骨架分析与应用

江贺¹ 邱铁¹ 胡燕¹ 李明楚¹ 罗钟铤^{1,2}

摘要 骨架是指一个 NP-难解问题实例的所有全局最优解的相同部分, 因其在启发式算法设计中的重要作用而成为该领域的研究热点. 本文对目前骨架及相关概念的研究成果进行了全面综述, 将骨架本身的研究工作归纳为三个层面: 理论基础层面主要考虑骨架与计算复杂性的关系问题; 应用基础层面主要考虑如何高效地获取骨架; 应用层面主要考虑如何利用骨架进行高效启发式算法设计. 在此基础上, 本文详细讨论了骨架研究亟待解决的难题, 并指出了解决这些问题的努力方向.

关键词 NP-难解, 骨架, 启发式算法, 计算复杂性

DOI 10.3724/SP.J.1004.2011.00257

Backbone Analysis and Applications in Heuristic Algorithm Design

JIANG He¹ QIU Tie¹ HU Yan¹ LI Ming-Chu¹ LUO Zhong-Xuan^{1,2}

Abstract As the shared common parts of all global optimal solutions to an NP-hard problem instance, the backbone has become a hot research line due to its essential impacts on the heuristic algorithm design. In this paper, research results on the backbone and related conceptions were thoroughly surveyed. The existing research of backbone consists of three areas, including the basic theoretical research focusing on investigation of the relationship between the backbone and computational complexity, the application research on how to efficiently approximate the backbone, and the applications on how to employ the backbone within the heuristic algorithm design. Furthermore, we discussed in detail the existing problems to be solved in this topic, and provided some guidelines on how to tackle them.

Key words NP-hard, backbone, heuristic algorithm, computational complexity

NP-难解问题普遍存在于道路交通、物流管理和物联网等众多领域中. 鉴于获取该类问题精确解的算法在最坏情况下所需时间为指数级^[1-2], 许多研究者探索如何在合理时间(多项式级别)内获取近似最优解, 并因此提出了大量的启发式算法(Heuristic algorithm)(包括蚁群算法^[3-4]、遗传算法^[5]、模拟退火^[6]、人工免疫^[7]等). 针对应用领域中出现的 NP-难解问题的规模的迅速扩张, 现有的启发式算法的性能难以进一步提升, 限制了本领域的发展. 研究者逐渐意识到: 解决这类组合爆炸问题的根本出路在于深入理解所求解问题的本身特征. 骨架(Backbone)作为描述 NP-难解问题内在特征的新途径, 开始被广泛运用于启发式算法设计, 形成了本领域的研究热点.

给定一个 NP-难解问题的实例, 骨架是指它的

所有全局最优解的共同部分. 研究骨架对于启发式算法的意义包括: 1) 基于骨架的启发式算法容易实现: 骨架可以提供一种新的框架, 其十分容易与现有启发式算法相互结合, 从而给出新的启发式算法; 2) 基于骨架的启发式算法性能提升显著: 基于骨架的算法可以对实例的搜索空间进行显著收缩, 从而明显提高搜索的效率, 例如: 基于骨架可以将旅行商问题(Traveling salesman problem, TSP)^[1]的实例转化成规模为原实例 30%~40% 的新实例, 所获得的多级归约算法比原最好算法 ILK (Iterative Lin-Kernighan) 在求解质量(即解与全局最优解的差别)上提升了 30%~60%^[8].

自从 Parkes^[9] 在研究相变(Phase transition)的时候最早地提出了骨架概念以后, 众多研究者对骨架进行了深入研究, 并取得了许多重要成果. 本文全面地综述了目前骨架及相关概念的研究现状, 并对骨架研究存在的问题进行了归纳分析. 本文首先将骨架本身的研究工作分为三个层面: 1) 理论基础研究层面主要考虑骨架与计算复杂性(Computational complexity)的关系问题, 具体包括骨架与相变、后门(Backdoor)的相关性以及获取骨架的

收稿日期 2009-09-01 录用日期 2010-12-02
Manuscript received September 1, 2009; accepted December 2, 2010

国家自然科学基金(60805024, 61033012), 大连市自然科学基金(201000117)资助

Supported by National Natural Science Foundation of China (60805024, 61033012) and Natural Science Foundation of Dalian (201000117)

1. 大连理工大学软件学院 大连 116621 2. 大连理工大学数学学院 大连 116621

1. School of Software, Dalian University of Technology, Dalian 116621 2. School of Mathematics, Dalian University of Technology, Dalian 116621

¹TSP 问题是指给定若干个顶点和它们之间的边, 寻求从某个顶点出发, 经过所有顶点一次且仅经过一次的最短环路. 根据顶点间往返的边长是否对称, 可以分为对称旅行商问题(Symmetric TSP, STSP)和非对称旅行商问题(Asymmetric TSP, ATSP). 一般在没有特别指明的情况下, TSP 问题是指 STSP.

计算复杂性等; 2) 应用基础研究层面主要考虑如何高效地获取骨架, 包括局部最优解近似法和限界交叉 (Limit-crossing) 方法; 3) 应用研究层面主要考虑如何利用骨架进行高效启发式算法设计. 在此基础上, 本文重点归纳了骨架研究存在的一系列问题, 包括获取骨架的多尺度计算复杂性问题、骨架的高效逼近问题、面向小规模骨架实例的启发式算法设计问题、骨架与适应度地貌分析 (Fitness landscape analysis) 等. 针对主要问题, 本文进行了深入分析并给出了解决这些问题的努力方向.

1 骨架的理论基础研究

本节主要分析骨架与计算复杂性的关系问题, 包括骨架与相变、后门的相关性, 骨架的计算复杂性等. 此外, 本节简要介绍一些与骨架相关的概念和研究现状.

1.1 骨架与相变的相关性

相变的概念源自于统计物理学, 它是指当复杂系统的控制参数在某个特定点 (相变点) 附近微小变化时, 系统的物理性质产生急剧变化的现象. 比如, 当温度 (控制参数) 低于零度 (相变点) 时, 水由液态凝结为固态 (冰). 有趣的是, 研究人员发现 NP- 难解问题的求解难度也普遍存在极为相似的相变现象, 并且与骨架关系密切.

由于可满足性问题 (Satisfiability problem, SAT)² 是最著名的 NP- 难解问题之一, 很多研究人员纷纷针对 SAT 问题及其变种开展了有关相变与骨架关系的研究. Monasson 等^[10] 研究了 $(2+p)$ -SAT 问题的相变与骨架规模的关系. 他们发现, 当子句数/变元数比值低于相变点 (4.0 ~ 4.2 左右) 时几乎 100% 的 $(2+p)$ -SAT 实例均可解, 而子句数/变元数比值超过相变点时又几乎均不可解, 在相变点附近时实例是最难求解的. 而在相变点附近, 骨架规模恰好是实例规模的 50% 左右, 在远离相变点的两端, 骨架规模分别趋近 0 和实例规模. Dimitriou^[11] 在加权 Max k -SAT 问题上观察到类似的现象. 通过一种实例难度可控的实例生成器, Dimitriou 发现该问题的求解难度表现出明显的相变现象, 并且相变区恰好是骨架规模急剧变化的区间. Zhang^[12] 则研究了 Max 3-SAT 问题的骨架自身的相变现象, 发现随着子句数/变元数比值增长, 骨架规模从几乎为 0 突然增加并迅速逼近实例规模, 并且 Max 3-SAT 的

骨架相变与 3-SAT 问题的 (可满足性) 相变几乎线性相关.

除 SAT 问题及其变种外, 其他 NP- 难解问题的相变与骨架关系也引起了研究人员的关注. Zhang^[13-14] 发现 ATSP 问题的求解难度也表现出显著的相变现象, 在相变点附近恰好是骨架规模为实例规模的 50% 左右. Boettcher 等^[15] 研究了随机图上的 3-着色 (3-coloring) 问题, 发现该问题的求解难度表现出了相变现象, 图上顶点的平均度数 (即每个顶点平均与多少边相关联) 为 4.7 左右为相变点, 而骨架的规模在相变点 4.7 附近也恰好表现出了相变现象. Slaney 等^[16] 发现图着色问题 (Graph coloring problem, GCP) 的难度与骨架规模正相关, 而模块世界规划 (Blocks world planning, BWP) 问题和 TSP 问题的难度与骨架规模负相关. Watson 等^[17] 研究车间调度问题 (Job-shop scheduling problem, JSP) 时发现, 随机 JSP 实例的全局最优解个数与骨架规模相关, 而局部最优解与全局最优解之间的差别则和骨架规模无明显关联.

一些研究人员还探索了骨架规模与启发式算法搜索代价之间的关系. Singer 等^[18] 发现随机 3-SAT 问题处于相变区附近时, 若骨架规模较小, 则全局最优解的个数对于启发式算法的搜索代价有直接影响, 全局最优解越多搜索代价越小; 而当骨架规模较大时, 全局最优解的个数对于启发式算法的搜索代价无显著影响. Singer 等^[19] 还研究了 $(2+p)$ -SAT 问题上骨架对于子句的依赖性, 他们发现 p 值越大, 实例的骨架越受到删除子句的影响. 而骨架对于删除子句的敏感性越高, 启发式算法求解所需的搜索代价就越高.

1.2 骨架与后门的相关性

后门是 Williams 等^[20-21] 最早在 SAT 问题的研究中提出的. 虽然随机 SAT 实例很难求解, 但是高效求解工具已经可以求解包含上百万个变元和几百万个子句的 (来自于实际应用中的) 实例. 因此, Williams 等认为现实世界中遇到的 SAT 实例实际上是包含了某种特殊的结构 (后门), 它由一组变元构成, 并存在一种赋值方式, 使得原 SAT 实例可以被启发式算法在多项式时间内判定是否可满足. 对于一个 SAT 实例而言, 它的骨架是唯一的, 而后门可能有很多. Williams 等从理论上分析了后门规模对于 SAT 启发式算法运行时间的影响, 后门规模越大则 SAT 实例求解时间越长. Williams 等通过实验发现: 实际应用中的大规模 SAT 实例容易求解的原因是包含了很小规模的后门. Kilby 等^[22] 专门针对 SAT 问题研究了骨架与后门的关系. 他们提出了一系列算法用于获取 SAT 问题的后门. 利用这些算

²SAT 问题是指给定一组布尔变元以及一组子句 (由若干布尔变元和它们的非构成的“或”运算), 寻找一种对布尔变元的赋值方式, 使所有的子句均为真. 2-SAT、3-SAT 是指子句长度分别固定为 2 和 3. $(2+p)$ -SAT 是 2-SAT 和 3-SAT 的混合形式, 其中长度为 3 的子句比例为 p , 而长度为 2 的子句比例为 $(1-p)$. Max k -SAT 是指子句长度固定为 k , 而寻求一种赋值方式使得最多的子句能被满足.

法, Kilby 等发现 SAT 问题的骨架与后门存在弱相关性.

1.3 骨架的计算复杂性分析

骨架的计算复杂性分析主要是从理论上判断: 是否存在多项式时间的精确算法可以获得 NP- 难解问题的骨架? 若存在这种多项式时间的精确算法, 则研究人员可以直接利用骨架本身来进行启发式算法设计. 否则, 则应该考虑如何有效地获取近似骨架, 用于启发式算法设计.

在骨架计算复杂性分析中, 研究人员主要采用了由 3 个步骤构成的证明方法.

首先, 构造 NP- 难解问题任意实例的偏移实例. 向 NP- 难解问题实例的约束条件中加入数值很小的微偏移量, 使之满足: 1) 在添加微偏移量前后, 原实例任意两个解的目标函数值大小关系在新实例上保持不变; 2) 新实例任意两个解的目标函数值均不相同. 这种新实例即是偏移实例.

其次, 分析偏移实例与原实例关系. 验证偏移实例是否满足: 只有唯一的全局最优解, 且该解亦是原实例的一个全局最优解.

最后, 分析完整骨架的计算复杂性. 若构造偏移实例成功, 则可以推断获取该 NP- 难解问题的完整骨架是 NP- 难解的, 原因在于: 反之, 则存在多项式时间算法可以获得偏移实例 (偏移实例也属于原 NP- 难解问题的一个实例) 的完整骨架, 由此也就得到原实例的全局最优解 (偏移实例的完整骨架就是它本身的唯一全局最优解, 也是原实例的全局最优解), 该结果与 NP- 难解问题的定义相矛盾.

按照上述思路, Kilby 等^[23] 证明了获取 TSP 问题的完整骨架和部分 (固定比例或者固定规模) 骨架是 NP- 难解的. 同时他们还给出了一种更强的计算复杂性结果, 即使是获取纯度大于 50% 的近似骨架也是 NP- 难解的. 在 SAT 问题上, Kilby 等^[22] 也给出了类似的结果. Jiang 等^[24] 则证明了获取图的二划分 (Graph bi-partitioning problem, GBP) 的完整骨架和部分 (固定比例或者固定规模) 骨架是 NP- 难解的. 采用类似的技巧, 他们^[25] 还证明了获取二次分配问题 (Quadratic assignment problem, QAP) 的骨架是 NP- 难解的.

1.4 骨架的相关概念研究

与骨架相对应, 还存在着脂肪 (Fat) 和肌肉 (Muscle) 的概念. 脂肪指一个 NP- 难解问题的实例中不在任何全局最优解出现的成分, 其研究重点在于如何剔除实例中的脂肪, 以缩小实例的搜索空间. 肌肉则定义为一个 NP- 难解问题的实例中所有全局最优解的并集, 其研究重点在于如何获取近似肌肉, 并在近似肌肉中进行高效搜索.

1) 脂肪

Climer 等^[26] 在 TSP 问题研究中最明确地提出了脂肪的概念, 并给出了获取脂肪的精确算法. 在 TSP 问题的实例中, 那些不在任何最短环路中出现的边就属于脂肪. 根据脂肪的定义, 从待求解的实例中将脂肪剔除后不影响实例的求解, 且剔除后的搜索空间将更加紧致, 因此对空间的搜索将更加高效. Climer 给出了获取 TSP 问题脂肪的精确算法——限界交叉方法. 该方法的核心思想是, 若将某条边固定后, TSP 实例的全局最优解会变差, 则说明该条边属于脂肪. 由于限界交叉方法的时间复杂度相对较高 (判定一条边属于脂肪的时间复杂度为 $O(n^4)$), 其主要用于 TSP 问题的精确求解. 以 ATSP 实例为例, 采用限界交叉方法后, 规模超过 500 个顶点的实例所需求解时间不到传统方法的 $1/3$ ^[26]. 另一方面, 随着实例规模的增加, 限界交叉所能获得的脂肪比例有所下降, 因此其不适合用于启发式算法设计.

江贺等^[27] 根据边在局部最优解中出现的频率与脂肪关系, 给出了 TSP 问题的动态候选集搜索算法. TSP 问题的局部搜索算法为了提高运行效率, 通常将与顶点关联的 (可能包含在合法环路上的) 边限制为有限的若干条, 这些边的集合称为顶点候选集. 江贺等通过实验发现: 随着边在局部最优解中出现频率的增加, 属于脂肪的边比例迅速降低. 根据上述现象, 他们设计了动态候选集算法. 该算法包括两个阶段: 前一阶段是在初始候选集上, 利用现有的局部搜索算子求得 TSP 实例的多个局部最优解; 后一阶段将那些在局部最优解中频繁出现的边调整到候选集的前部, 然后在新候选集上继续调用局部搜索算子求解. 他们利用新算法对目前 TSP 最好的算法之一的 LKH (Lin-Kernighan Helsgaun) 进行了改进, 并在旅行商问题库 (TSP library, TSPLIB) 的多组不同规模的典型实例上进行实验比较. 实验结果表明, 改进后的算法在全局最优解数量, 解的平均质量等方面较原 LKH 算法有较为明显的提高. 该算法的不足在于, 其有效性依赖于具体实例的结构特征, 难以从理论上保证始终能获得高质量的解.

Gutin 等^[28] 则研究了如何剔除推广的旅行商问题 (Generalized TSP, GTSP) 的脂肪, 以实现实例的归约. 给定一些顶点集合, 不同集合中的顶点间存在加权边, GTSP 问题的目标是如何从每个集合挑选一个顶点, 使得由这些顶点形成的环路最短. Gutin 等分别考虑了如何删除不属于任何全局最优解的顶点和边 (这些顶点和边即是脂肪). 以判定顶点为例, 给定顶点集合 X 的一个顶点 x , 若对于任意分属于其他顶点集合 Y, Z 的两个顶点 y, z , 若能找到 X 中的另一顶点 x' , 使得 x' 到 y, z 的距离之和小于 x 到 y, z 的距离之和, 则可以判定 x 属于

脂肪而可以从实例中删除. 大量的测试结果表明, 利用该方法可以将实例规模平均缩小 15% ~ 20%.

2) 肌肉

Jiang 等^[29] 在 3 次分配问题 (3-index assignment problem, AP3) 研究中引入了肌肉的概念. 江贺等分析了该问题上肌肉的计算复杂性, 并给出了基于近似肌肉的全局搜索算法. 该算法的基本做法是利用现有的启发式算法求得 AP3 实例的多个局部最优解, 将这些局部最优解合并作为近似肌肉, 然后在其中实施全局优化. 为了提高全局优化的效率, 使用了多种剪枝加速策略. 实验结果表明, 新的算法在困难实例上比已有的最好结果^[30-31] 有显著改进.

实际上, 有些研究者虽然没有明确提出肌肉的概念, 但也采用了类似的算法设计技巧来求解很多 NP- 难解问题.

在 TSP 问题的求解中, Applegate 等^[32] 提出了环路合并方法: 首先重复调用 ILK 算法^[33] 求得多个高质量环路 (局部最优解), 然后由这些环路构成一个更加稀疏的图, 最后在这个稀疏图上使用精确算法^[34] 求解. 他们将环路合并方法用于求解大规模 TSP 实例 (实验中最大实例包含 85 900 个顶点), 发现其在平均解质量和获得全局最优解的次数上, 比 ILK 算法有明显优势. 以 pcb3038 为例, 新算法在 10 次运行中能获得 8 次全局最优解, 平均解的长度仅比全局最优解超出 0.001%. 而对比算法 ILK 在 10 次运行中未能获得全局最优解, 平均解长度比全局最优解超出 0.19%. Cook 等^[35] 提出了一种新的环路合并算法, 主要改进是采用性能更高的 LKH 算法^[36] 替换 ILK 算法. 新的环路合并算法是目前求解中等规模和大规模 TSP 实例最好的算法之一.

在 p -中心 (p -median) 问题的求解中, Rosing 等^[37-39] 提出了一种启发式浓缩策略 (Heuristic concentration). 该策略分为三个阶段, 首先调用 Teitz 等^[40] 提出的启发式算法获得若干个局部最优解, 然后从这些解中选出一部分构成一个规模更小的 p -median 实例, 最后使用一个数学规划工具 CPLEX^[41] 精确求解. 作为对比, Rosing 等在 21 个典型 p -median 实例上测试了启发式浓缩策略和之前最好算法之一的禁忌搜索算法^[42], 发现新算法可以获得其中 17 个实例的全局最优解或者已知最优解, 而禁忌搜索算法仅能获得两个 2 个实例的全局最优解或已知最优解. 而在运行时间方面, 启发式浓缩策略在某些实例上时间长于禁忌搜索算法, 在另外一些实例上又优于禁忌搜索算法.

在集合覆盖问题 (Set covering problem, SCP) 的求解中, Finger 等^[43] 采用了与 Rosing 等^[37-39] 类似的技巧. 所不同的是在最后一个阶段采用了一种模拟退火方法^[44] 来求解转化后的实例. 改用模拟

退火方法原因在于, 转化后的实例的搜索空间依然非常巨大, 若采用精确算法则时间开销过大.

2 骨架的应用基础研究

在运用骨架进行启发式算法设计前必须有效获取骨架, 现有的方法包括局部最优解近似法和限界交叉两种.

2.1 局部最优解近似法

当前很多学者采取局部最优解近似法来获得近似骨架, 也就是提取多个局部最优解中的共同部分作为近似骨架. 该方法的依据是所谓的“大坑” (Big valley) 猜想, 即许多 NP- 难解问题的局部最优解的结构与全局最优解有很高的相似性. 形象地描述, 形成了以全局最优解为底而周围遍布局部最优解的一个“大坑”. 该现象最早是由 Boese^[45] 在 TSP 问题上发现的, 他通过实验证实很多 TSP 实例的局部最优解与全局最优解有 80% 左右的边是相同的. 后来, Merz 等^[46]、Reeves^[47] 和 Zhang^[48] 也在其他问题 (如 GBP 问题、SAT 和 Max-SAT 问题) 上分别观察到类似的现象.

针对一个具体的 NP- 难解问题, 局部最优解近似法的基本想法是首先通过小实例进行采样分析, 判断是否存在“大坑”, 其次再设法获取近似骨架. 具体步骤包括: 首先构造或者寻找该问题的若干个小实例, 并在其上重复运行某现有启发式算法来获得多个局部最优解, 再采用穷尽搜索或分支限界等方法来求出小实例的全局最优解, 然后通过比较判断全局最优解和局部最优解是否相似 (即是否有“大坑”), 若有“大坑”现象, 则可以提取多个局部最优解的共同部分来充当近似骨架.

局部最优解近似法的优点是实现简单、时间复杂度小且获得骨架规模较大, 现有的各种启发式算法均可以用于获得局部最优解. 该方法的缺陷是所获取的骨架为近似骨架, 并且纯度变化很大 (比如 TSP 问题的近似骨架纯度在 85% 以上^[8], 而 QAP 问题的近似骨架纯度仅有 50% ~ 60%^[25, 49]). 针对局部最优解近似法的缺陷, Zhang 等^[50] 提出: 为了保证近似骨架有较高的纯度, 用于获取近似骨架的局部最优解应具备良好的无偏性和多样性, 这就要求初始解的选取应该尽量随机化.

2.2 限界交叉方法

限界交叉方法是 Climer 等^[26] 提出的一种确定性方法. 它的基本原理是依次删除 NP- 难解问题的实例中的一部分结构 (比如 TSP 问题的某些边), 然后判定新实例的全局最优解的目标函数值是否变差, 若是则说明被删除部分属于骨架 (该判定规则可以

由骨架定义推导出来). 由于直接求解原实例和新实例均很困难, 所以实际应用中采用松弛技巧: 以最小化问题为例, 用新实例的解下界与原实例的解上界进行比较, 若新实例的解下界超过原实例的解上界, 则判定被删除部分属于骨架. 一种计算下界的方法是利用拉格朗日松弛等手段求解 NP- 难解问题对应的整数规划形式. 上界可以使用原实例上局部最优解的目标函数值, 而现有的启发式算法均可以用于获取局部最优解.

限界交叉方法的优点是能从理论上保证所能判定的部分必定属于骨架; 缺点是存在漏判现象: 下界和上界并非足够紧致 (Tight), 导致很多本属于骨架的部分不能被发现. 另外, 该方法的计算代价很高. 在 TSP 问题中, 判定一条边是否属于骨架的时间复杂度高达 $O(n^4)$.

3 骨架的应用研究

骨架应用层面的任务是如何设计基于骨架的高效启发式算法. 根据骨架使用方式的不同, 这类算法可以分为确定型和概率型两种.

3.1 确定型骨架算法

确定型骨架算法的特征是通过固定 (近似) 骨架实现实例的归约或者搜索空间的限制. 根据骨架的作用, 确定型骨架算法可以进一步分为空间受限型和实例归约型.

1) 实例归约型骨架算法

实例归约型骨架算法通过固定 (近似) 骨架, 把 NP- 难解问题的实例转变成规模更小的新实例, 以便进行高效求解. 实例归约型算法由三个阶段构成: 第一阶段是归约, 固定 (近似) 骨架来归约 NP- 难解实例, 获得更小规模的新实例, 而这个新实例可以类

似归约, 直到得到新实例的规模足够小为止; 第二阶段是求解, 运用现有某高效算法求解最终的小实例; 第三阶段是反向优化, 其可以视为第一阶段的逆向过程, 从最小规模实例开始, 逐渐把 (近似) 骨架插入恢复到规模更大的实例中, 并进一步优化, 如此迭代直至恢复到原实例.

图 1 给出了 TSP 问题的实例归约的例子. 其中, 图 1(a), 1(b) 和 1(c) 代表了归约阶段, 虚线表示近似骨架, 它们将在实例中被删除 (近似骨架的两个端点将用长度为 0 的边连接起来), 以获得规模更小的实例; 图 1(d) 代表求解阶段, 调用某种启发式算法求解经过多次归约后的最小实例; 图 1(e), 1(f) 和 1(g) 代表了反向优化阶段, 从最小实例开始, 反向将近似骨架加入到实例中以得到更大规模实例, 并将最小实例的解和近似骨架拼接成更大规模实例的初始解, 然后调用现有算法来进一步优化此解, 如此操作, 直至恢复到原始实例.

按照实例归约的思路, 很多研究人员给出了针对 TSP 问题的新启发式算法. 邹鹏等^[8] 给出了求解 TSP 问题的多级归约算法, 通过采用多次运行 LK 算法求得若干局部最优解, 再提取局部最优解的共同部分当作近似骨架, 最后通过固定近似骨架来对实例进行多级归约及反向优化. 实验表明, 在典型小规模 TSP 实例上, 多级归约算法比 ILK 算法^[33] 更加容易获得全局最优解. 对于大规模实例, 多级归约算法所获得解质量比 ILK 算法也有较明显提高. Schneider^[51-52] 运用类似想法, 使用并行化策略来提取近似骨架, 由此给出了用于 TSP 问题求解的并行多级归约算法. 并行多级归约算法的求解质量与邹鹏等的多级归约算法类似. 由于其中获取局部最优解步骤容易并行化, 该并行多级归约算法可以提供接近线性的加速比, 这为将来各种基于骨架的算法

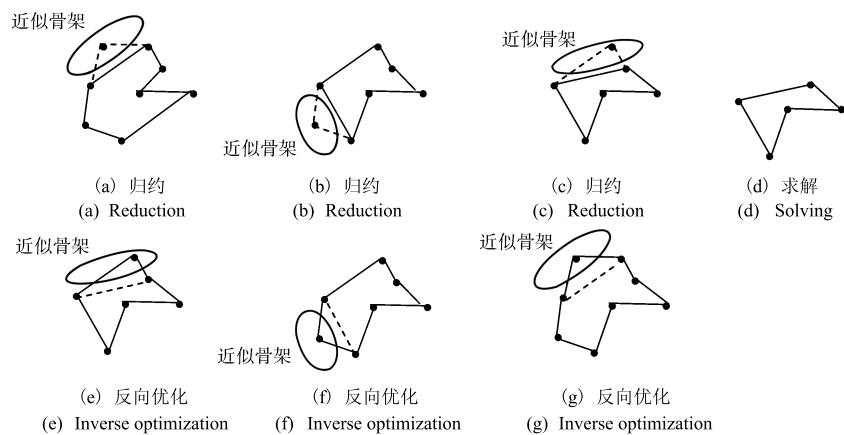


图 1 实例归约的示意图

Fig. 1 Illustration of instance reduction

的加速提供了一条可行的思路. 威玉涛等^[53]从理论的角度分析了决定 TSP 问题多级算法^[8]性能的原因, 据此提出了面向 TSP 问题的自适应归约免疫算法框架, 该框架通过近似骨架的进化, 使其规模自适应地增加, 同时不断提高近似骨架的纯度, 以达到提高归约后算法找到全局最优解的概率值. 在 6 个典型 TSP 实例上的结果表明, 自适应归约免疫算法运行时间比邹鹏等的归约算法稍长, 但解的质量有较明显地提高. Fischer 等^[54]提出了 4 种获取 TSP 问题近似骨架的方法, 并根据实例归约的思路设计了新启发式算法, 实验结果表明, 该算法在大规模 TSP 实例上的性能 (在解的质量、运行时间指标上) 超过目前最好算法中的 Memetic 算法^[55-56].

Dong 等^[57]主要针对来自于超大规模集成电路 (Very large scale integrated, VLSI) 领域的 TSP 实例, 设计了实例归约算法, 该算法的特点是在实例归约中采用了不同版本的 LKH 算法: 首先利用 LKH 的低版本软件工具 LKH-1.3 获得多个局部最优解, 然后利用这些解获取近似骨架来归约实例, 最后则采用 LKH 的最新版本 LKH-2.0 来求解这个规模更小的归约后实例. 利用这种归约算法, Dong 等刷新了 VLSI 领域中 6 个大规模 TSP 实例的最好解记录.

除了 TSP 问题之外, 一些学者也将骨架的概念运用到了聚类分析中. 宗瑜等^[58]设计了一种基于近似骨架的聚类算法. 他们利用现有的启发式聚类算法得到同一聚类实例的多个局部最优解, 通过对局部最优解求交得到近似骨架, 将近似骨架固定得到规模更小的聚类实例, 然后在新实例上求解. 在 26 个仿真数据集和 3 个实际数据集上的实验结果表明, 骨架理论对提高聚类质量、降低初始解影响及加快算法收敛速度等方面均十分有效. 与宗瑜等的工作类似, 李雪^[59]将近似骨架运用到不确定数据的聚类中, 通过提取多个聚类结果的共同部分, 实现了对于聚类实例的归约. 在大量不确定数据集上的实验结果表明, 该方法比现有算法在聚类准确度上有较显著提高.

2) 空间受限型骨架算法

空间受限型骨架算法是指通过固定 (近似) 骨架, 使得启发式算法在搜索时仅考虑搜索空间的一个子集. 与实例归约型骨架算法不同的是, 空间受限型骨架算法一直是在同一个实例上进行求解, 而不是多个规模不同实例. 以包含 3 个布尔变元 (X, Y, Z) 和 3 个子句 ($X \vee Y \vee Z, X \vee \neg Y \vee Z, \neg X \vee Y \vee \neg Z$) 的 3-SAT 实例为例, 若近似骨架是变元 X 为真, 则对于解 (X 为真, Y 为假, Z 为真) 进行局部搜索时, 仅考虑改变 Y 或 Z 的赋值, 而不改变 X 的赋值. 因此, 通过固定骨架, 解空间仅有 4 个需要尝

试的解, 而原始解空间有 8 个解.

在 3-SAT 问题上, Dubois 等^[60]将近似骨架运用于求解困难实例 (即子句数/变元数在 4.25 左右相变区的随机 3-SAT 实例). 通过利用近似骨架对 DPLL (Davis-Putnam-Logemann-Loveland) 类的启发式算法的搜索树剪枝, 在 200~600 个变元的局部随机 3-SAT 实例上比此前的最好算法 (比如 satz214^[61]) 快 1.2~3 倍, 同时解决了包含多达 700 个变元的随机 3-SAT 实例.

在 QAP 问题上, 邹鹏等^[49]利用近似骨架对搜索空间进行收缩, 提出了近似骨架导向的蚁群算法. 该算法首先通过局部最优解近似法获取近似骨架, 然后将这一部分固定, 最后利用蚁群在受限制的搜索空间上进一步求解. 在 17 个典型 QAP 实例上的结果表明, 近似骨架导向的蚁群算法在运行时间相近的情况下, 有 15 个实例均取得比传统的蚁群算法更好的结果. Jiang 等^[25]改进了邹鹏等的工作, 通过变换 QAP 实例, 提高了实例的骨架规模, 给出了基于偏移实例的近似骨架导向的蚁群算法. 实验结果表明, 对于二次分配问题库 (QAP library, QAPLIB) 中的典型实例, 该算法在解的质量上较近似骨架导向的蚁群算法、随机重启蚁群算法等有较为显著的提高.

Zou 等^[62]归纳了利用近似骨架对搜索空间进行约束的技巧, 给出了基于交叉的缩放算法 (Intersection based scaling, IBS), 并在 TSP 及图划分问题 (Graph partitioning problem, GPP) 上进行了测试. IBS 算法在 TSP 问题上的效果与多级归约算法^[8]类似, 而在 GPP 问题上比原有的 KL (Kernighan-Lin) 算法^[63]在解的质量上有所提高.

3.2 概率型骨架算法

概率型骨架算法是根据当前解中基本单元 (比如 TSP 问题中的边) 在近似骨架中出现的概率, 来引导启发式算法的求解过程. 概率型骨架算法大致可分为三个阶段: 第一阶段是局部最优解求解阶段, 采用现有的启发式算法获得若干个局部最优解; 第二阶段是近似骨架概率计算阶段, 计算解中各个基本单元在这些局部最优解中出现的频率, 并将其作为近似骨架概率; 第三阶段是概率导向求解阶段, 将近似骨架概率用于初始解的生成、实例变换或局部搜索中邻域的确等.

图 2 以 TSP 问题为例, 给出了概率型骨架算法的执行示意图. 首先, 通过现有的启发式算法 (比如 LKH) 获得的 2 个局部最优解 (图 2(a), 2(b)). 其次, 计算出该实例中各条边在这些局部最优解中出现的频率, 并将此频率作为近似骨架概率 (图 2(c)). 比如, 边 (2, 4) 在 1 个局部最优解中出现, 则它的近

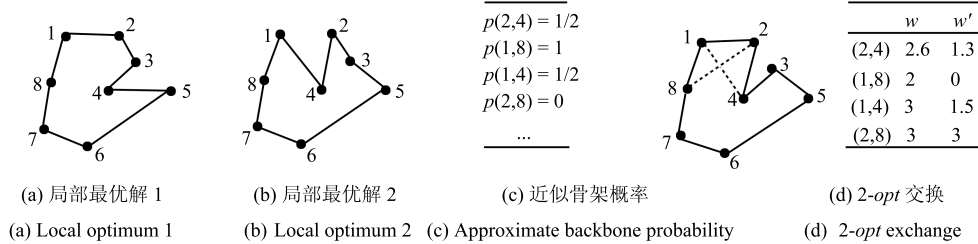


图 2 概率型骨架算法的示意图

Fig. 2 Illustration of probabilistic backbone based algorithm

似骨架概率 $p(2, 4) = 1/2$; 而边 $(1, 8)$ 在 2 个局部最优解中均出现, 故它的近似骨架概率 $p(1, 8) = 2/2 = 1$. 最后, 利用近似骨架概率来引导后续的求解过程. 在本例子中, 近似骨架概率被用来重新调整各边的权值, 以生成新的实例. 给定边 (i, j) 及其权值 w , 一种新的权值计算公式^[50] 是 $w' = w(1 - p(i, j))$. 图 2(d) 给出了 4 条边的原始权值和调整后权值. 以边 $(2, 4)$ 为例, 原始的权值为 2.6, 而根据近似骨架概率用现有启发式算法在新实例上求解. 根据权值计算公式可以发现, 那些频繁在局部最优解中出现的边权值被缩小, 从而在后续的求解中更加容易出现环路中.

根据概率型骨架算法的思路, 很多研究者针对 TSP 问题设计了多种新算法. Zhang 等^[50] 给出了一种骨架导向的 LK 算法. 该算法利用局部最优解来逼近骨架, 并根据边在近似骨架中出现概率来选择初始解中的边; 在搜索过程中根据边在近似骨架中出现的概率来决定是否替换它, 并在搜索过程中根据概率不断调整优化方向. Richter 等^[64] 利用 Zhang 等^[50] 提供的近似骨架作为候选集来引导搜索过程, 改进了目前 TSP 问题的最好算法之一的 LKH 算法^[36]. 2009 年, LKH 算法的作者 Helsgaun^[65] 将 Zhang 等^[50] 的算法设计思想吸收到 LKH 软件包的升级中, 所获得的新版本 LKH 2.0 在求解大规模 TSP 实例 ($10^4 \sim 10^7$ 个顶点) 上取得了目前启发式算法最好的结果.

在 SAT 问题研究中, Zhang^[48] 通过实验发现 SAT 问题和 Max-SAT 问题存在“大坑”结构(局部最优解与全局最优解相似), 基于这一发现, Zhang 利用变元和子句在近似骨架出现的频率对 WalkSat 中的 5 个算子(包括初始解生成、子句选择等)进行了修改, 实验结果表明这种骨架导向的 WalkSat 算法与原算法相比: 更容易获得 SAT 问题的解, 而在 Max-SAT 问题上解的质量平均提高了 20%. Yoon^[66] 对于 Zhang 提出的骨架导向的 WalkSat 算法进行了分析, 发现算法有效的关键在于近似骨架的纯度足够高, 而在很多困难的 SAT 实例上,

近似骨架纯度都能达到 70% 以上, 故此骨架导向的 WalkSat 算法性能优于原 WalkSat 算法. Hsu 等^[67-68] 针对 SAT 问题, 分析了多种概率消息传递机制(概况传播 (Survey propagation, SP)、信念传播 (Belief propagation, BP)) 等对于近似骨架的影响, 并将近似骨架与 SAT 求解工具 MiniSat 相结合, 显著提高了算法在某些困难的随机实例上的搜索效率. Valnir^[69] 等讨论了骨架对于 SAT 问题的典型动态局部搜索算法——纯附加权重机制 (Pure additive weighting scheme, PAWS) 及其变种^[70-72] 的影响, 实验结果表明近似骨架可以加快大规模 SAT 实例的求解, 而小规模实例则不明显.

在 SAT 的变种研究中, Telelis 等^[73] 针对 Max-SAT 问题给出了基于骨架采样的启发式算法. 该算法首先随机生成多个解保存在一个集合中, 然后对解集进行多次迭代, 并输出解集中最好解. 在每次迭代中, 根据解集中变元的赋值与解质量的关系, 计算出变元属于近似骨架的概率. 根据这一概率再生成大量解, 并从中挑选出一批解进行局部搜索得到局部最优解, 最后从这些局部最优解中挑选出最好的一个将解集中最差的一个替换掉. Menai^[74] 给出了求解部分最大可满足性问题 (Partial Max-SAT, PMSAT) 的基于骨架的两段式启发式算法. PMSAT 问题是 Max-SAT 问题的变种, 它由两组子句构成, 要求在完全满足第一组子句的前提下, 满足近量多的第二组子句. Menai 的新算法分为两个阶段: 在第一阶段, 首先将两组子句合并在一起作为一般的 Max-SAT 问题, 并调用 Zhang 等^[48] 的基于骨架的算法来求解; 在第二阶段, 则选出前一阶段中最好的解作为初始解, 然后再次调用 Zhang 等^[48] 的算法来确保第一组子句可以完全被满足. Menai^[75] 还利用文献 [74] 的算法高效地求解了另一个著名的 NP-难解问题——Steiner 树问题 (Steiner tree problem, STP) 问题, 其基本做法是首先将 STP 问题转化为 PMSAT 问题, 然后利用基于骨架的两段式启发式算法进行求解, 最后再将解转化为原始 STP 问题的解. 这种做法所获得的解比以前的启发式算法有了

较显著地提高. Jiang 等^[76] 针对加权 Max-SAT 给出了基于近似骨架概率的随机游走算法. 该算法首先利用随机游走算子获取一系列解, 然后根据这些解计算出近似骨架概率, 再利用 Zhang 等^[48] 的算法在近似骨架概率的引导下进行求解. 在 44 个典型 Max-SAT 实例上的实验结果表明, 基于近似骨架概率的随机游走算法的运行时间明显小于此前最好算法之一的 GRASP (Greedy randomized adaptive search procedure) 算法^[77], 并且解的质量有普遍提高. 在全部实例中, GRASP 算法仅能获得 4 个实例的全局最优解, 而基于近似骨架概率的随机游走算法能获得 19 个实例的全局最优解.

4 下一步研究的方向

虽然骨架研究已经取得了令人振奋的成果, 但围绕着骨架在启发式算法研究中的不同层面还存在一系列重要的问题亟待解决.

4.1 骨架的计算复杂性

当前, 骨架的理论研究部分主要是考虑优化类问题 (包括最小化问题和最大化问题), 而缺少组合类问题的成果 (仅有的成果是 SAT 问题的骨架分析^[22]). 对于大量存在的组合类问题, 骨架理论分析包括两个尺度:

1) 获取完整骨架的计算复杂性

是否存在多项式时间的算法可以获得完整骨架? 如何来分析获取完整骨架的计算复杂性? 这还有待更深入研究.

2) 获取部分骨架的计算复杂性

相对于获取完整骨架, 一个更强化的问题是获取部分 (固定比例或者固定规模) 骨架的计算复杂性如何? 这是一个很有意义的问题: 若能得到固定比例或固定规模的骨架, 也可以利用它对 NP- 难解问题的实例进行归约或者搜索空间的收缩, 从而提高求解效率.

4.2 骨架的高效逼近问题

事先获取骨架是利用它来指导启发式算法设计的前提. 目前, 获取骨架的两种手段均存在较大的缺陷: 限界交叉方法^[26] 的计算代价太高且实现复杂, 很难运用于启发式算法设计中; 局部最优解近似法虽然实现简单, 却存在纯度偏低的局限性. 因此, 如何高效逼近骨架是当前提高基于骨架的启发式算法性能的关键. 新的骨架逼近方法应该满足: 速度快、实现简单、纯度高要求. 这可以从以下两个方面入手.

1) 将局部最优解与下界近似手段相结合, 利用共同部分作为近似骨架

与启发式算法类似, 用于估计 NP- 难解问题的全局最优解下界 (最小优化问题为下界, 最大优化问题则为上界) 方法本身也是一种对于全局最优解的逼近. 因此, 下界近似手段得到的解 (或者部分解) 同样包含了骨架的一些信息. 基于这种考虑, 采用下界近似手段和局部最优解相结合的办法有望克服传统局部最优解近似法缺乏无偏性和多样性的不足.

2) 不同启发式算法获得局部最优解, 利用它们的共同部分作为近似骨架

针对传统的局部最优解近似法存在纯度偏低的局限性, Zhang 等^[50] 提出利用初始解的随机性来加以改善. 然而, 采用同一种启发式算法来获取多个局部最优解的做法本身就缺乏足够的无偏性和多样性. 解决这一问题的一种可行途径是用多种不同的启发式算法来获得局部最优解, 然后用这些局部最优解的共同部分来模拟骨架.

4.3 面向小规模骨架实例的启发式算法设计问题

目前各种基于骨架的算法存在共同的缺陷: 一个具体实例的骨架规模可能非常小 (比如多数典型 QAP 实例的骨架规模不到解规模的 10%^[25, 49]), 此时基于骨架的算法难以体现其优势. 一种极端的情况是, 当骨架规模为 0 时, 利用骨架进行算法设计就没有意义了.

目前存在的困难有: 1) 具体实例的骨架规模很难事先预知, 而判定一个实例的骨架规模本身可能就是 NP- 难解的. Beacham^[78] 证明了判定 SAT 问题和 Hamiltonian path 问题的骨架是否为空是 NP- 难解的; 2) 即使骨架规模可知, 基于骨架的启发式算法对于小规模骨架的实例也难以发挥骨架的优势. 因此, 如何提高待求解实例的骨架规模, 对于基于骨架的启发式算法设计具有重要意义. 在文献 [79] 中, 江贺等实现了 GPP 问题的骨架规模提升, 并给出了基于偏移实例的近似骨架导向的归约算法. 在该文工作中, 他们通过扰动待求解实例得到偏移实例, 而偏移实例可以证明仅包含一个全局最优解 (即骨架), 从而有效提升了骨架规模. 这种技巧, 对于其他的 NP- 难解问题也具有一定的参考价值.

4.4 骨架与适应度地貌分析

骨架和适应度地貌均是刻画 NP- 难解问题特征的方式. 骨架是从单个解的构成来刻画 NP- 难解问题, 而适应度地貌则是从解空间的整体来刻画 NP- 难解问题. 适应度地貌分析的一些成果有力地推动了骨架的研究工作.

适应度地貌 (Fitness landscape) 的概念来源于理论生物学中^[80]. 随后, 这一概念广泛应用于组合优化^[1]、复杂系统^[81-82] 等领域中. 对于 NP- 难解问题 (如 TSP 问题) 的实例 I , 其适应度地貌可以定

义成一个三元组 $L = (S, f, d)$, 其中 S 为实例 I 的解空间, $f: S \rightarrow R$ 为适应度函数 (常取为目标函数), $d: S \times S \rightarrow R$ 为满足以下条件的距离函数: 1) $d(s, t) \geq 0$; 2) $d(s, t) = 0 \Leftrightarrow s = t$; 3) $d(s, t) \leq d(s, u) + d(u, t)$, 其中 $\forall s, t, u \in S$. 对于给定问题的一个实例而言, 适应度地貌是与具体的距离函数相关联的. 最简单且常见的距离函数可以定义为: 对求解该问题的某种启发式算法, 定义距离函数 $d: S \times S \rightarrow R$, 使得 $d(s, t)$ 为在该算法下从 s 搜索到 t 的最少步数.

适应度地貌存在多种特征, 近似骨架所依据的“大坑”结构仅是其中一种特征 (适应度距离相关性) 的研究成果. 适应度距离相关系数 (Fitness distance correlation, FDC) 是 Jones 等^[83] 最早提出的概念. 通过 FDC 的图形化表示, 可以直观分析局部最优解到最近的全局最优解的距离和其适应度之间的统计关系. FDC 图是通过大量运行启发式算法来获取局部最优解, 对于每一个局部最优解 loc 在图上用点表示 $(d(loc, opt), f(loc))$. 对 TSP 问题, 典型实例上的 FDC 图如图 3^[8] 所示, 从中可以看到: 对不同的实例, $d(loc, opt) \approx 0.2n$, 其中 n 为实例规模 (顶点数). 这就是所谓的“大坑”结构.

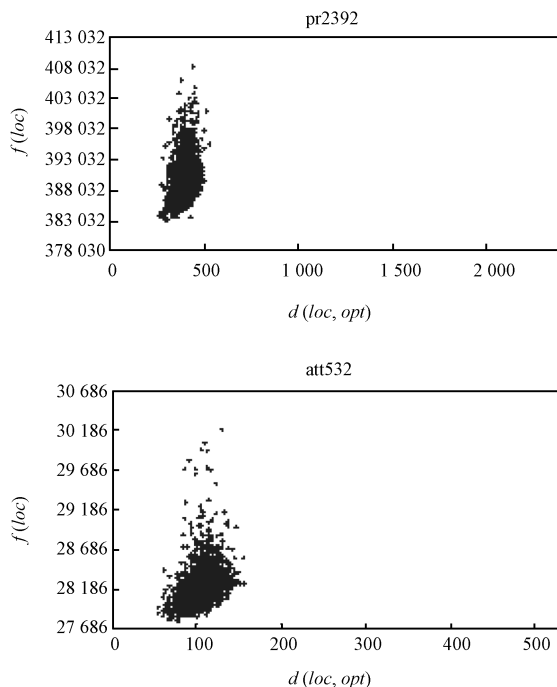


图 3 TSPLIB 中的典型实例 pr2392 和 att532 的 FDC 分析图

Fig. 3 FDC analysis figures of typical samples, pr2392 and att532, in TSPLIB

除了 FDC 外, 适应度地貌存在自相关性、分形性等特征. 自相关性 (Autocorrelation function)

是 Weinberger^[84] 为了衡量适应度地貌的平滑性而提出的. 相关性长度越大, 适应度地貌越光滑, 局部搜索就越容易获得全局最优解. 近年来, Angel 等^[85-88] 采用概率统计的方法给出了 QAP 和 GBP 等问题自相关性等参数的解析公式, 并由此分析了适应度地貌与启发式算法性能的关系. 适应度地貌的分形性是 Sorkin^[89] 提出的. 对于适应度地貌上的任意两点对 $\langle x, y \rangle$, 若它们的适应度方差与它们之间的距离正相关, 则称该地貌为分形的. Weinberger^[90] 和 Reeves 等^[91] 对地貌的分形类型进行了深入研究.

对于适应度地貌的自相关性和分形性与骨架的关系还缺乏相关的研究成果, 存在大量的有待研究课题, 包括: 适应度地貌的光滑性与骨架规模是否存在关联; 适应度地貌的分形特征对于获取骨架是否有帮助; 特别是分析适应度地貌所采用的解析方法, 对于获取骨架是否有参考意义等. 这些课题都有待进一步的深入研究.

4.5 肌肉和脂肪的进一步研究

脂肪和肌肉是与骨架相对应的概念, 目前启发式算法设计中有关脂肪和肌肉的研究成果还不丰富. 有关脂肪的研究仅包含了获取脂肪的限界交叉方法和基于脂肪的候选集搜索算法. 而肌肉的研究主要还是利用它来进行全局优化. 由于脂肪和肌肉的概念与骨架相似, 因此有关骨架研究的很多热点也适用于脂肪和肌肉, 包括脂肪和肌肉的计算复杂性分析方法、近似逼近方法及各类启发式算法设计等.

4.6 讨论

根据以上小节的分析, 给定一个 NP- 难解问题, 其骨架研究可以按照图 4 的路线展开. 首先分析其骨架的计算复杂性, 若获取骨架不是 NP- 难解的, 则存在多项式时间算法可以获取精确骨架, 然后就可根据骨架进行确定型骨架算法设计 (包括所求解实例进行归约或者对搜索空间进行限制). 反之, 若获取骨架是 NP- 难解的, 则进行骨架的规模提升, 再通过适应度地貌分析来获得高效逼近骨架的手段, 最后利用近似骨架进行启发式算法设计.

5 结论

骨架是近年来启发式算法设计中的重要工具. 本文对于骨架的研究现状进行了分析, 将现有的骨架研究按照理论基础、应用基础和应用分为三个层次, 分别进行了归纳总结. 在此基础上, 分析了骨架研究存在的一系列问题: 在理论上缺乏骨架的多尺度计算复杂性分析; 在应用基础上缺乏骨架的高效逼近手段; 在骨架应用上难以应对小规模骨架

实例等. 针对这些问题, 给出了下一步工作的思路. 这些重要问题的最终解决有望显著提高基于骨架的启发式算法的性能, 拓展计算复杂性理论的研究范畴.

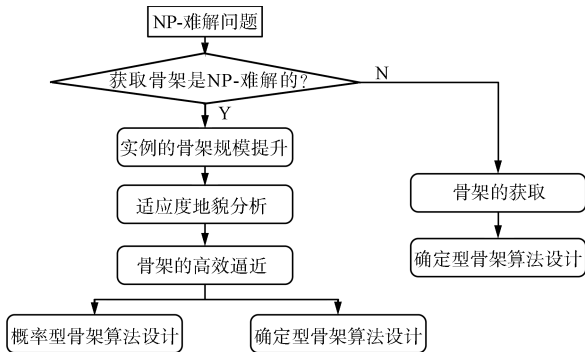


图 4 骨架的主要研究路线

Fig. 4 Main research steps of the backbone

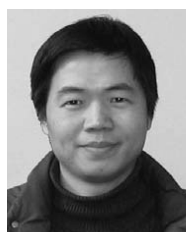
References

- Garey M R, Johnson D S [Author], Zhang Li-Ang, Shen-Hong, Bi Yuan-Zhang [Translator]. *Computers and Intractability: A Guide to the Theory of NP-completeness*. Beijing: Science Press, 1987
(Garey M R, Johnson D S [著], 张立昂, 沈泓, 毕源章 [译]. 计算机和难解性: NP 完全性理论引导. 北京: 科学出版社, 1987)
- Huang Wen-Qi, Xu Ru-Chu. *Introduction to Modern Computational Theory: Background, Perspective, and Algorithms Research of NP-hard Problems*. Beijing: Science Press, 2004
(黄文奇, 许如初. 近世计算理论导引: NP 难度问题的背景、前景及其求解算法研究. 北京: 科学出版社, 2004)
- Dorigo M, Blum C. Ant colony optimization theory: a survey. *Theoretical Computer Science*, 2005, **344**(2–3): 243–278
- Dorigo M, Maniezzo V, Colnari A. Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 1996, **26**(1): 29–41
- Holland J H. *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press, 1975
- Kirkpatrick S, Gelatt C D J, Vecchi M P. Optimization by simulated annealing. *Science*, 1983, **220**(4598): 671–680
- Farmer J D, Packard N H, Perelson A S. The immune system, adaptation and machine learning. *Physica D*, 1986, **2**(1–3): 187–204
- Zou Peng, Zhou Zhi, Chen Guo-Liang, Gu Jun. A multi-level reduction algorithm to TSP. *Journal of Software*, 2003, **14**(1): 35–42
(邹鹏, 周智, 陈国良, 顾钧. 求解 TSP 问题的多级归约算法. 软件学报, 2003, **14**(1): 35–42)
- Parkes A J. Clustering at the phase transition. In: Proceedings of the 14th National Conference on Artificial Intelligence and the 9th Conference on Innovative Applications of Artificial Intelligence. Rhode Island, USA: AAAI Press, 1997. 340–345
- Monasson R, Zecchina R, Kirkpatrick S, Selman B, Troyansky L. Determining computational complexity for characteristic “phase transition”. *Nature*, 1999, **400**(6740): 133–137
- Dimitriou T. SAT distributions with planted assignment and phase transitions between decision and optimization problems. *Discrete Applied Mathematics*, 2005, **153**(1): 58–72
- Zhang W X. Phase transitions and backbones of 3-SAT and maximum 3-SAT. In: Proceedings of the 7th International Conference on Principles and Practice of Constraint Programming. London, UK: Springer-Verlag, 2001. 153–167
- Zhang W X. Phase transitions, backbones, measurement accuracy, and phase-aware approximation: the ATSP as a case study. In: Proceedings of the 4th International Workshop on Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems. Le Croisic, France: Springer, 2002. 345–358
- Zhang W X. Phase transition and backbones of the asymmetric traveling salesman problem. *Journal of Artificial Intelligence Research*, 2004, **21**(1): 471–497
- Boettcher S, Percus A G. Extremal optimization at the phase transition of the three-coloring problem. *Physical Review E*, 2004, **69**(6): 1–8
- Slaney J, Walsh T. Backbones in optimization and approximation. In: Proceedings of the 17th International Joint Conference on Artificial Intelligence. San Francisco, USA: Morgan Kaufmann Publishers, 2001. 254–259
- Watson J P, Beck J C, Howe A E, Whitley L D. Problem difficulty for tabu search in job-shop scheduling. *Artificial Intelligence*, 2003, **143**(2): 189–217
- Singer J, Gent I P, Smaill A. Backbone fragility and the local search cost peak. *Journal of Artificial Intelligence Research*, 2000, **12**(1): 235–270
- Singer J, Gent I P, Smaill A. Local search on random 2 + p-SAT. In: Proceedings of the 14th European Conference on Artificial Intelligence. Amsterdam, Holland: IOS Press, 2000. 113–117
- Williams R, Gomes C P, Selman B. Backdoors to typical case complexity. In: Proceedings of the 18th International Joint Conference on Artificial Intelligence. San Francisco, USA: Morgan Kaufmann Publishers, 2003. 1173–1178
- Williams R. On computing k-CNF formula properties. In: Proceedings of the 6th International Conference on Theory and Applications of Satisfiability Testing. Heidelberg, Germany: Springer, 2003. 330–340

- 22 Kilby P, Slaney J, Thiebaut S, Walsh T. Backbones and backdoors in satisfiability. In: Proceedings of the 20th National Conference on Artificial Intelligence. Pittsburgh, USA: AAAI, 2005. 1368–1373
- 23 Kilby P, Slaney J, Walsh T. The backbone of the traveling salesperson. In: Proceedings of the 19th International Joint Conference on Artificial Intelligence. San Francisco, USA: Morgan Kaufmann Publishers, 2005. 175–181
- 24 Jiang H, Zhang X C, Chen G L. Unique optimal solution instance and computational complexity of backbone in the graph bi-partitioning problem. *Chinese Science Bulletin*, 2007, **52**(20): 2871–2875
- 25 Jiang H, Zhang X C, Chen G L, Li M C. Backbone analysis and algorithm design for the quadratic assignment problem. *Science in China Series F: Information Sciences*, 2008, **51**(5): 476–488
- 26 Climer S, Zhang W X. Searching for backbone and fat: a limit-crossing approach with applications. In: Proceedings of the 18th National Conference on Artificial Intelligence. California, USA: AAAI, 2002. 707–712
- 27 Jiang He, Hu Yan, Li Qiang, Yu Hong. Fat computational complexity and heuristic design for the TSP. *Journal of Software*, 2009, **20**(9): 2344–2351
(江贺, 胡燕, 李强, 于红. TSP 问题的脂肪计算复杂性与启发式算法设计. *软件学报*, 2009, **20**(9): 2344–2351)
- 28 Gutin G, Karapetyan D. Generalized traveling salesman problem reduction algorithms. *Algorithmic Operations Research*, 2009, **4**(2): 144–154
- 29 Jiang H, Xuan J F, Zhang X C. An approximate muscle guided global optimization algorithm for the three-index assignment problem. In: Proceedings of the IEEE Congress on Evolutionary Computation. Hong Kong, China: IEEE, 2008. 2404–2410
- 30 Huang G F, Lim A. A hybrid genetic algorithm for the three-index assignment problem. *European Journal of Operational Research*, 2006, **172**(1): 249–257
- 31 Aiex R M, Resende M G C, Pardalos P M, Toraldo G. GRASP with path relinking for three-index assignment. *INFORMS Journal on Computing*, 2005, **17**(2): 224–247
- 32 Applegate D, Bixby R, Chvatal V, Cook W. Finding Tours in the TSP, Technical Report 99885, University of Bonn, Germany, 1999
- 33 Glover F, Kochenberger G A. *Handbook of Metaheuristics*. Norwell: Kluwer Academic Publishers, 2002. 321–353
- 34 Applegate D, Bixby R, Chvatal V, Cook W. Concorde: a computer code for traveling salesman problems [Online], available: <http://www.tsp.gatech.edu/concorde/downloads/codes/src/970827/README>, August 18, 1999
- 35 Cook W, Seymour P. Tour merging via branch-decomposition. *INFORMS Journal on Computing*, 2003, **15**(3): 233–248
- 36 Helsingaun K. An effective implementation of the Lin-Kernighan traveling salesman heuristic. *European Journal of Operational Research*, 2000, **126**(1): 106–130
- 37 Rosing K E. Heuristic concentration: a study of stage one. *Environment and Planning B: Planning and Design*, 2000, **27**(1): 137–150
- 38 Rosing K E, ReVelle C S. Heuristic concentration: two stage solution construction. *European Journal of Operational Research*, 1997, **97**(1): 75–86
- 39 Rosing K E, ReVelle C S. Heuristic concentration and tabu search: a head to head comparison. *European Journal of Operational Research*, 1998, **104**(1): 93–99
- 40 Teitz M B, Bart P. Heuristic methods for estimating the generalized vertex median of a weighted graph. *Operations Research*, 1968, **16**(5): 955–961
- 41 IBM ILOG CPLEX Optimizer [Online], available: <http://www-01.ibm.com/software/integration/optimization/cplex-optimizer/>, July 1, 2009
- 42 Rolland E, Schilling D A, Current J R. An efficient tabu search procedure for the p -median problem. *European Journal of Operational Research*, 1997, **96**(2): 329–342
- 43 Finger M, Stutzle T, Lourenco H. Exploiting fitness distance correlation of set covering problems. In: Proceedings of the Applications of Evolutionary Computing on EvoWorkshops. London, UK: Springer-Verlag, 2002. 61–71
- 44 Brusco M J, Jacobs L W, Thompson G M. A morphing procedure to supplement a simulated annealing heuristic for cost- and coverage-correlated set-covering problems. *Annals of Operations Research*, 1999, **86**(1): 611–627
- 45 Boese K D. Cost Versus Distance in the Traveling Salesman Problem, Technical Report CSD-950018, University of California, USA, 1995
- 46 Merz P, Freisleben B. Fitness landscapes, memetic algorithms, and greedy operators for graph bipartitioning. *Evolutionary Computation*, 2000, **8**(1): 61–91
- 47 Reeves C R. Landscapes, operators and heuristic search. *Annals of Operations Research*, 1999, **86**(1): 473–490
- 48 Zhang W X. Configuration landscape analysis and backbone guided local search: Part I: satisfiability and maximum satisfiability. *Artificial Intelligence*, 2004, **158**(1): 1–26
- 49 Zou Peng, Zhou Zhi, Chen Guo-Liang, Jiang He, Gu Jun. Approximate-backbone guided fast ant algorithms to QAP. *Journal of Software*, 2005, **16**(10): 1691–1698
(邹鹏, 周智, 陈国良, 江贺, 顾钧. 求解 QAP 问题的近似骨架导向快速蚁群算法. *软件学报*, 2005, **16**(10): 1691–1698)

- 50 Zhang W X, Looks M. A novel local search algorithm for the traveling salesman problem that exploits backbones. In: Proceedings of the 19th International Joint Conference on Artificial Intelligence. San Francisco, USA: Morgan Kaufmann Publishers, 2005. 343–348
- 51 Schneider J, Froschhammer C, Morgenstern I, Husslein T, Singer J M. Searching for backbones — an efficient parallel algorithm for the traveling salesman problem. *Computer Physics Communications*, 1996, **96**(2–3): 173–188
- 52 Schneider J. Searching for backbones — a high-performance parallel algorithm for solving combinatorial optimization problems. *Future Generation Computer Systems*, 2003, **19**(1): 121–131
- 53 Qi Yu-Tao, Liu Fang, Jiao Li-Cheng. Immune algorithm with selfadaptive reduction for large-scale TSP. *Journal of Software*, 2008, **19**(6): 1265–1273
(戚玉涛, 刘芳, 焦李成. 求解大规模 TSP 问题的自适应归约免疫算法. 软件学报, 2008, **19**(6): 1265–1273)
- 54 Fischer T, Merz P. Reducing the size of traveling salesman problem instances by fixing edges. In: Proceedings of the 17th European Conference on Evolutionary Computation in Combinatorial Optimization. Heidelberg, Germany: Springer-Verlag, 2007. 72–83
- 55 Moscato P, Norman M G. A “memetic” approach for the traveling salesman problem implementation of a computational ecology for combinatorial optimization on message-passing systems. In: Proceedings of the International Conference on Parallel Computing and Transputer Applications. Amsterdam, Holland: IOS Press, 1992. 177–186
- 56 Merz P, Freisleben B. Memetic algorithms for the traveling salesman problem. *Complex Systems*, 2001, **13**: 297–345
- 57 Dong C X, Jäger G, Richter D, Molitor P. Effective tour searching for TSP by contraction of pseudo backbone edges. In: Proceedings of the 5th International Conference on Algorithmic Aspects in Information and Management. San Francisco, USA: Springer, 2009. 175–187
- 58 Zong Yu, Li Ming-Chu, Jiang He. Approximate backbone guided reduction algorithm for clustering. *Journal of Electronics and Information Technology*, 2009, **31**(12): 2953–2957
(宗瑜, 李明楚, 江贺. 近似骨架导向的归约聚类算法. 电子与信息学报, 2009, **31**(12): 2953–2957)
- 59 Li Xue. Research on Clustering of Uncertain Data [Master dissertation], Dalian University of Technology, China, 2009
(李雪. 不确定数据聚类研究 [硕士学位论文], 大连理工大学, 中国, 2009)
- 60 Dubois O, Dequen G. A backbone-search heuristic for efficient solving of hard 3-SAT formulae. In: Proceedings of the 17th International Joint Conference on Artificial Intelligence. San Francisco, USA: Morgan Kaufmann Publishers, 2001. 248–253
- 61 Li C M. A constraint-based approach to narrow search trees for satisfiability. *Information Processing Letters*, 1999, **71**(2): 75–80
- 62 Zou P, Zhou Z, Wan Y Y, Chen G L, Gu J. New metaheuristic for combinatorial optimization problems: intersection based scaling. *Journal of Computer Science and Technology*, 2004, **19**(6): 740–751
- 63 Kernighan B W, Lin S. An efficient heuristic procedure for partitioning graphs. *Bell Systems Technical Journal*, 1970, **49**(2): 291–307
- 64 Richter D, Goldengorin B, Jager P, Molitor P. Improving the efficiency of Helsgaun’s Lin-Kernighan heuristic for the symmetric TSP. In: Proceedings of the 4th Conference on Combinatorial and Algorithmic Aspects of Networking. Berlin, Germany: Springer-Verlag, 2007. 99–111
- 65 Helsgaun K. General *k-opt* submoves for the Lin-Kernighan TSP heuristic. *Mathematical Programming Computation*, 2009, **1**(2–3): 119–163
- 66 Yoon D T S. Empirical Analysis of Local Search Algorithms and Problem Difficulty in Satisfiability [Master dissertation], University of Toronto, Canada, 2006
- 67 Hsu E I, Muise C J, Beck J C, McIlraith S A. Probabilistically estimating backbones and variable bias: experimental overview. In: Proceedings of the 14th International Conference on Principles and Practice of Constraint Programming. Berlin, Germany: Springer-Verlag, 2008. 613–617
- 68 Hsu E I, Muise C J, McIlraith S A, Beck J C. Applying Probabilistic Inference to Heuristic Search by Estimating Variable Bias, Technical Report CSRG-577, Department of Mechanical and Industrial Engineering, University of Toronto, Canada, 2008
- 69 Valnir F. Improvements to Clause Weighting Local Search for Propositional Satisfiability [Ph. D. dissertation], Griffith University, Australia, 2006
- 70 Ferreira V, Thornton J. Tie breaking in clause weighting local search for SAT. In: Proceedings of the 18th Australian Joint Conference on Artificial Intelligence. Sydney, Australia: Springer, 2005. 70–81
- 71 Thornton J, Pham D N, Bain S, Ferreira V. Additive versus multiplicative clause weighting for SAT. In: Proceedings of the 19th National Conference on Artificial Intelligence. San Jose, USA: AAAI, 2004. 191–196
- 72 Pham D N, Thornton J, Sattar A, Ishtaiwi A. SAT-based versus CSP-based constraint weighting for satisfiability. In: Proceedings of the 20th National Conference on Artificial Intelligence. California, USA: AAAI, 2005. 455–460
- 73 Telelis O, Stamatopoulos P. Heuristic backbone sampling for maximum satisfiability. In: Proceedings of the 2nd Hellenic Conference on Artificial Intelligence. Thessaloniki, Greece: Springer, 2002. 129–139

- 74 Menai M E B. A two-phase backbone-based search heuristic for partial MAX-SAT: an initial investigation. In: Proceedings of the 18th International Conference on Innovations in Applied Artificial Intelligence. London, UK: Springer-Verlag, 2005. 681–684
- 75 Menai M E. A logic-based approach to solve the steiner tree problem. In: Proceedings of the 5th IFIP Conference on Artificial Intelligence Applications and Innovations. Thessaloniki, Greece: Springer, 2009. 73–79
- 76 Jiang H, Xuan J F. Backbone guided local search for the weighted maximum satisfiability problem. *Evolutionary Computation*. Victoria: InTech Press, 2009. 261–272
- 77 Festa P, Pardalos P M, Pitsoulis L S, Resende M G C. GRASP with path relinking for the weighted MAXSAT problem. *ACM Journal of Experimental Algorithms*, 2006, **11**: 1–16
- 78 Beacham A J. The Complexity of Problems without Backbones [Master dissertation], University of Alberta, Canada, 2000
- 79 Jiang He, Qiu Tie. Backbone analysis and heuristic design for the graph partitioning problem. *Chinese Journal of Computers*, 2009, **32**(8): 1662–1667
(江贺, 邱铁. GPP 问题的骨架分析与启发式算法设计. 计算机学报, 2009, **32**(8): 1662–1667)
- 80 Wright S. Evolution in populations in approximate equilibrium. *Journal of Genetics*, 1935, **30**(2): 257–266
- 81 Binder K, Young A P. Spin glasses: experimental facts, theoretical concepts and open questions. *Reviews of Modern Physics*, 1986, **58**(4): 801–976
- 82 Mezard M, Parisi G, Virasoro M. *Spin Glass Theory and Beyond*. Singapore: World Scientific, 1987
- 83 Jones T, Forrest S. Fitness distance correlation as a measure of problem difficulty for genetic algorithms. In: Proceedings of the 6th International Conference on Genetic Algorithms. San Francisco, USA: Morgan Kaufmann Publishers, 1995. 184–192
- 84 Weinberger E. Correlated and uncorrelated fitness landscapes and how to tell the difference. *Biological Cybernetics*, 1990, **63**(9): 325–336
- 85 Angel E, Zissimopoulos V. On the hardness of the quadratic assignment problem with metaheuristics. *Journal of Heuristics*, 2002, **8**(4): 399–414
- 86 Angel E, Zissimopoulos V. On the landscape ruggedness of the quadratic assignment problem. *Theoretical Computer Science*, 2001, **263**(1–2): 159–172
- 87 Angel E, Zissimopoulos V. On the quality of local search for the quadratic assignment problem. *Discrete Applied Mathematics*, 1998, **82**(1–3): 15–25
- 88 Angel E, Zissimopoulos V. Autocorrelation coefficient for the graph bipartitioning problem. *Theoretical Computer Science*, 1998, **191**(1–2): 229–243
- 89 Sorkin G B. Efficient simulated annealing on fractal energy landscapes. *Algorithmica*, 1991, **6**(3): 367–418
- 90 Weinberger E D, Stadler P F. Why some fitness landscapes are fractal. *Journal of Theoretical Biology*, 1993, **163**(2): 255–275
- 91 Reeves C R. Fitness landscapes and evolutionary algorithms. In: Proceedings of the 4th European Conference on Artificial Evolution. London, UK: Springer-Verlag, 2000. 3–20



江贺 博士, 大连理工大学副教授. 主要研究方向为智能计算和数据挖掘. 本文通信作者.

E-mail: jianghe@dlut.edu.cn

(JIANG He Ph.D., associate professor at Dalian University of Technology. His research interest covers intelligent computation and data mining.

Corresponding author of this paper.)



邱铁 博士研究生, 大连理工大学讲师. 主要研究方向为嵌入式高性能计算和无线传感器网络系统建模.

E-mail: qiutie@dlut.edu.cn

(QIU Tie Ph.D. candidate, lecturer at Dalian University of Technology. His research interest covers embedded high performance computing, and wireless sensor networks and systems modeling.)



胡燕 博士, 大连理工大学讲师. 主要研究方向为智能计算和数据挖掘.

E-mail: wohuyan@gmail.com

(HU Yan Ph.D., lecturer at Dalian University of Technology. His research interest covers intelligent computation and data mining.)



李明楚 博士, 大连理工大学教授. 主要研究方向为网络安全, 网络计算和智能计算.

E-mail: mingchul@dlut.edu.cn

(LI Ming-Chu Ph.D., professor at Dalian University of Technology. His research interest covers network security, grid computing, and intelligent computation.)



罗钟铉 博士, 大连理工大学教授. 主要研究方向为智能计算和计算几何.

E-mail: zxluo@dlut.edu.cn

(LUO Zhong-Xuan Ph.D., professor at Dalian University of Technology. His research interest covers intelligent computation and computational geometry.)