

CVRPTW 可以用数学公式表示如下 [33]:K 个车队用于服务 n 个客户。服务开始时间bi必须落在客户vi的时间窗口 [ei , fi] 内,其中ei和fi分别代表服务qi 的最早和最晚时间 (即 vi 的需求)。

如果车辆在时间ai < ei到达vi ,则等待时间wi = max {0, ei – ai}发生。因此,服务开始时间bi = max {ei , ai}。每辆载客量为 Q 的车辆行驶在一条连接客户子集的路线上,从v0开始并在计划范围[e0, f0] 内结束。 dij表示从客户vi到客户vj 的距离。

决策变量: Xk = 1,如果从vi到vj的边被分配在车辆k的路线上,否则Xk = 0。 ij目标函数:

最小化

钾

(5)

最小化

$$\sum_{k \in K} \sum_{vi \in V} \sum_{vj \in V} Xk \ ij \ dij$$

(6)

约束:

$$\sum_{k \in K} \sum_{vi \in V} Xk \ ij \ = 1, \ \forall vi \in V \setminus \{vo\}$$

(7)

$$\sum_{k \in K} \sum_{vj \in V} Xk \ ij \ = 1, \ \forall vj \in V \setminus \{vo\}$$

(8)

$$\sum_{k \in K} \sum_{vi \in V} \sum_{vj \in V \setminus \{vo\}} Xk \ ij \ = n$$

(9)

$$\sum_{vj \in V} Xk \ oj \ = 1, \ \forall k \in K$$

(10)

$$\sum_{vi \in V} Xk \ ij \ - \sum_{vj \in V} Xk \ ji \ = 0, \ \forall k \in K, vj \in V \setminus \{vo\}$$

(11)

$$\sum_{i=0}^6 Xk \ i0 \ = 1, \ \forall k \in K$$

(12)

$$ei \ bi \ fi, \ \forall vi \in V$$

(13)

$$\sum_{vi \in V} \sum_{vj \in V} Xk \ ij \ qi \leq Q, \ \forall k \in K$$

(14)

$$Xk \ ij \ \in \{0, 1\}, \ \forall vi, vj \in V, k \in K$$

(15)

第一个目标是最小化车辆数量 (等式 (5)),而第二个目标是最小化总行驶距离 (等式 (6))。约束条件 (7-9) 限制每个客户只能访问一次,同时确保为所有客户提供服务。约束 (10-12) 定义车辆 k 的路线。约束 (13)和 (14)分别定义了客户时间窗口约束和车辆容量约束。约束 (15) 定义了决策变量Xk ij的域。

我们在文献中采用相同的评估函数,其中两个目标被转换为具有权重 [34] 的单个目标,如等式 (16)所示。

钾

$$f = \sum_{k \in K} \sum_{vi \in V} \sum_{vj \in V} Xk \ ij \ dij + 1000 \times K \ (16)$$

Solomon 基准数据集 [35] 由六组不同特征的实例 (C1.C2.R1.R2.RC1 和 RC2)组成。这些实例在客户的地理位置、车辆容量、密度和时间紧迫性方面有所不同 windows. 实例集 C1 和 C2 中的客户在地理上是聚类的;而实例集 R1 和 R2 中的客户是随机分布的。实例集 RC1 和 RC2 包含随机客户和集群客户的混合。

对于相同类型的问题实例,客户坐标是相同的。一种类型中的实例在时间窗的密度和紧密度方面有所不同,即时间受限客户的百分比和时间窗的宽度。

B. 学习模式的有效性

本节从两个方面研究了所提出的 RL-GSF 模型的有效性:Q 值函数逼近器的影响;政策更新机制的影响。

1)Q值函数逼近器对学习模型的影响:QL代表表格RL方法,DQN代表函数逼近RL方法,在本节建立的GSF中应用。选择随机算法作为基线算法来证明 RL 方法的性能。 Random-GSF方法在优化过程的不同阶段在已建立的GSF内随机选择算法组件,无需任何学习,即每个算法组件被选中的概率相同。

状态空间由相对于初始种群适应度的总适应度改进定义,即表 IX 中的f1。由于 CVRPTW 是一个具有有限适应度搜索空间的 NP-hard 问题,QL-GSF 方法需要处理大量的状态。在 QL-GSF 中使用基于状态聚合概念的近似技术 [36]、[37]、[38] 将状态空间聚合为几个不相交类别。

从初步的实验观察来看,在 R1 型和 RC1 型实例中,获得的值落在 [0.4,0.6] 范围内。该范围在 type-C、type-R2 和 type-RC2 实例中略有不同,从实验中观察到 [0.3,0.5]。因此,type-R1和type-RC1实例的状态空间划分为: f1 ∈ [0, 0.4), [0.41, 0.42), ..., [0.59, 0.6), [0.6, 0.7), ... , [0.9, 1];对于type-C、type-R2和type-RC2实例,状态空间划分为: f1 ∈ [0, 0.3), [0.31, 0.32), ..., [0.49, 0.5), [0.5, 0.7), [0.7, 0.8], ..., [0.9, 1]。在带有神经网络函数逼近器的DQN-GSF中,不需要聚合状态;相反, f1可以简单地用作神经网络的唯一输入

网络。

这三种算法除了 Q 值函数逼近器外,实验环境和参数设置完全相同。在所有算法中,种群大小、时间步数 NoT 和预定义的最大值

一集的运行时间tmax分别设置为100、50和600s。为了训练策略, NoE 的集数设置为 500。出于测试目的, 如表 X XII 所示, 通过运行每个学习算法 10 次, 我们收集了平均最佳适应度 (AVG)、标准差 (SD)、最佳适应度 (BEST) 和 BEST 与文献中最知名的解决方案之间的 GAP [39]。两个最先进的自动设计算法 (即 RT 和 HG) 的已发布结果 (BEST) 列在所有比较表中。

表 X所选

选择这两种人工设计的算法作为比较,因为它们具有已发表论文中的大部分结果和信息。

应该注意的是,通常不可能比较自动方法和手动设计方法的设计时间,因为这些信息通常不会在已发表的论文中报告。其中一些只公布了他们的结果并没有提供申请时间。由于不同的计算平台和实现语言,直接比较所提出的自动化方法和产生最知名解决方案的手动设计方法之间的计算费用是不公平的。此外,在大多数已发布的算法中,终止条件和独立运行的次数因方法而异。与其他方法相比,所提出的方法在训练和测试过程中需要额外的计算时间。然而,我们的目标不是开发一种快速的方法,而是自动开发可以产生具有更高通用性的最先进结果的搜索算法。额外的时间可以通过解决不同的问题实例来补偿,而无需长期重新设计或微调算法。

表 XI 所选

在 type-C 实例上,表 XI 中的 BEST 和 GAP 的结果表明这三种方法可以产生当前最知名的解决方案 [39],这种类型的实例可以通过进化搜索来解决,无需任何学习技术。不同的 AVG 和 SD 表明所提出的 RL-GSF 方法,尤其是 DQN-GSF,更稳定地自动设计用于解决具有统计显著性的 C 类实例的搜索算法(通过 Wilcoxon 秩和检验测量, $p < 0.05$),并在所有结果表中用 * 表示。

表 XII 所选

在 type-R 和 type-RC 实例上,如表 XI 和表 XII 所示,DQN-GSF 在大多数实例中取得了这三种算法中最好的结果。QL-GSF 是第二好的,在大多数情况下比 Random-GSF 具有更高的 AVG 和更小的 GAP。这表明基于学习的模型比非学习搜索过程更有效。

总之,神经网络函数逼近器执行简单的 Q 表。有了更多的特征来定义状态空间,学习方法的有效性可能会进一步提高。然而,一个简单的 Q 表处理多个特征所需的内存将会增加,探索每个状态以创建所需 Q 表所需的时间变得不现实。与 Q 表相比,神经网络能够处理多个特征。

还可以观察到,Random-GSF 在 type-C 实例上显示出与其他两种方法相当的性能

但在 type-R 和 type-RC 实例上的性能较差,这表明学习机制可以帮助找到算法组件的更好组合,从而获得更好的解决方案。在下一节中,将进一步研究两种基于神经网络的 RL GSF 方法 DQN-GSF 和 PPO-GSF。

[illegible]

TYPE-R实例的比较 (Q 值函数逼近器的影响)。并且*表明DQN-GSF分别与RANDOM-GSF和QL-GSF显著不同,即, $p < 0.05$

TYPE-RC实例的比较 (Q 值函数逼近器的影响)。并且*表明DQN-GSF分别与RANDOM-GSF和QL-GSF显著不同,即, $p < 0.05$

2)策略更新机制对学习模型的影响:在基于价值的方法DQN-GSF和基于策略的方法PPO-GSF中,除了策略更新机制外,其他参数,如人口规模和最大运行时间都相同,以进行公平比较。

PPO-GSF和DQN-GSF的策略在训练过程中逐渐完善。然而,必须保持一定程度的随机性以避免陷入局部最优。奖励曲线似乎

乳胶类文件杂志,卷。 , 不。 , 2022 年 8 月

因此,随着一些波动而上升。为了更清楚地呈现训练效果,使用滑动窗口滤波方法 (移动平均)平滑奖励曲线,如等式 (17)所示。

$$\text{平滑} = \frac{\text{convolve}(\text{xbuf f}, \text{ybuf f})}{\text{convolve}(\text{zbuf f}, \text{ybuf f})} \quad (17)$$

xbuf f是每集原始奖励, ybuf f[1, ···, 1]q是一个向量,其长度为平滑因子 = [实验中设置为5。 zbuf f是一个长度为q的向量, zbuf f是整个训练数据。q在实验中设置为5。

在 C 类实例中,如图 3 所示,在大多数情况下,PPO-GSF 的性能优于 DQN-GSF。如表 XIII 所示,平均最佳适应度和标准差也证明了 PPO-GSF 优于 DQN-GSF。

同样,这两种学习方法都可以产生当前最著名的解决方案 [39]。

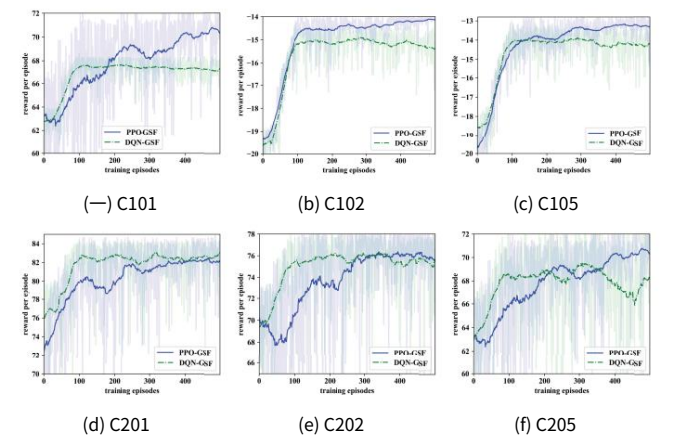


图 3. 策略更新机制对学习模型的影响 (C 类问题实例)

表 XII 对选定的 C 类实例的比较 (策略更新机制的影响)。表示 PPO-GSF 与 DQN-GSF 显著不同, IE, p < 0.05

实例	C101	C202	10828.94 [40]	10828.94 [40]	10828.94 [40]	3591.56 [40]	3591.56 [40]	3588.88 [40]	C205
最知名的解决方案	10828.94	10828.94	3591.56	10828.94	10828.94	3591.56	11148.36	10900.96	3612.52
转交 [40]	10828.94	10828.94	3591.56	10828.94	10828.94	3591.56	10900.96	3612.52	3591.56
转交 [41]	10828.94	10828.94	3591.56	10828.94	10828.94	3591.56	10900.96	3612.52	3591.56
DQN-GSF	10828.94	10828.94	3591.56	10828.94	10828.94	3591.56	10900.96	3612.52	3591.56
PPO-GSF	10828.94	10828.94	3591.56	10828.94	10828.94	3591.56	10900.96	3612.52	3591.56

在 type-R 实例上,如图 4 所示,PPO-GSF 在算法收敛性和解决方案质量方面优于 DQN-GSF。此外,表 XIV 显示,在大多数情况下,PPO-GSF 在所有四个指标 AVG、SD、BEST 和 GAP 方面取得了更好的结果。除 R202 实例外,大多数情况下 PPO-GSF 的 GAP 小于 3%。

在图 5 所示的 type-RC 实例中,PPO-GSF 在所有实例中明显优于 DQN-GSF。在表 XV 中,在大多数 type-RC 实例中,PPO-GSF 和 DQN-GSF 获得的解是非支配解

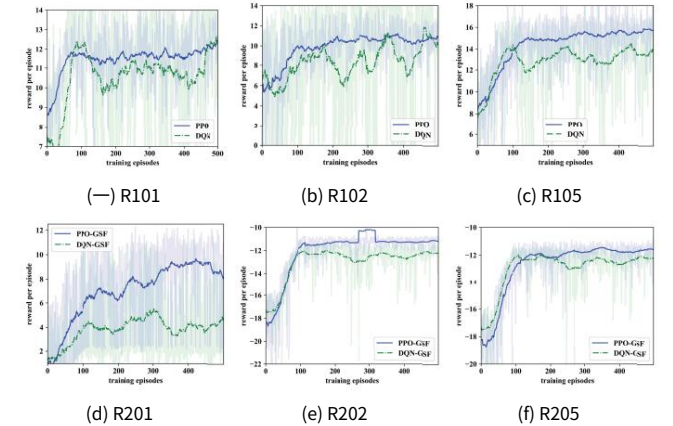


图 4. 策略更新机制对学习模型的影响 (R 类问题实例)

表 XIV 对所选 TYPE-R 实例的比较 (策略更新机制的影响)。表示 PPO-GSF 与 DQN-GSF 显著不同, IE, p < 0.05

实例	R101	R102	R202	20645.79 [42]	18486.12 [40]	18486.12 [40]	5252.37 [40]	4191.7 [43]	3994.42 [43]	R205
最知名的解决方案	20650.80	18486.12	5088.07	20656.79	18490.39	4198.45	AVG 20081.19	19910.38	5201.13	SD 454.50
转交 [40]	20656.49	19495.43	5152.05	20656.79	18490.39	4198.45	AVG 20081.19	19910.38	5201.13	SD 454.50
转交 [41]	20656.49	19495.43	5152.05	20656.79	18490.39	4198.45	AVG 20081.19	19910.38	5201.13	SD 454.50
DQN-GSF	15377.11	5252.37	16434.78	5378.77	4280.21	4162.06	5378.77	4280.21	4162.06	4162.06
PPO-GSF	15377.11	5252.37	16434.78	5378.77	4280.21	4162.06	5378.77	4280.21	4162.06	4162.06

由文献 [39] 中所有其他元启发式方法确定的解决方案。

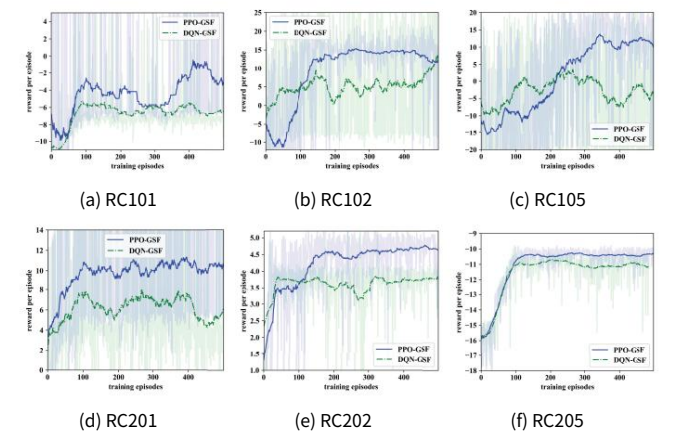


图 5. 策略更新机制对学习模型的影响 (类型 RC 问题实例)

总之,实验结果表明,PPO-GSF 和 DQN-GSF 方法都可以支持 GSF 中的有效学习,自动生成用于解决不同类型 CVRPTW 实例的进化算法。特别是,对于神经网络逼近器,基于策略的模型 PPO-GSF 比基于值的模型 DQN-GSF 更有效。主要有两个原因。首先,基于策略的方法可以学习随机策略,而基于价值的方法只能学习确定性策略。基于策略的方法更有能力进行更好的环境探索。

表 XV对所选
TYPE-RC实例的比较（策略更新机制的影响）。表示PPO-GSF与DQN-GSF显著不同, IE, p < 0.05

实例	RC101	RC102	RC202	15696.94 [44]	13554.72 [40]	26628.44 [45]	5406390 [6]	4367.09 [47]	RC205				
最知名的解决方案	5297.19 [46]	16623.58	14477.54	5165.57	15697.43	13558.07	4665.56	AVG 17586.44	16146.86	5351.60	SD 728.65	486.43	
转发 [40]	45.33	BEST 16570.05	15586.85	5736.64	GAP 25.63	56.22%	0.4	QAPWLB 81892.4	1820.18	6005.04 *	5359.51	SD 8183.8	525.97
乘 [41]					14637.15		5418.86				5302.42		
					17041.4		5661.65				5535.85		
DQN-GSF					499.10		269.55				284.21		
					16603.84		5507.77				5361.69		
					13.5%		1.87%				1.2%		
					16543.15 *		5567.00 *				5450.50 *		
PPO-GSF					309.87		78.01				36.60		
					15617.44	6.8%	5467.35				5359.96	1.2%	
							1.12%						

其次,由于其有效的价值函数优化方法,PPO-GSF 可以确保学习到的策略单调递增,从而更好地利用。

C. 学习模型的推广

RL-GSF 模型的训练过程非常耗时。本节调查由所提出的 RL-GSF 模型训练的策略的普遍性,可能会减少时间并重用在自动算法设计中学习的策略来解决新问题实例。从两个方面进行了分析:跨相同类型实例的泛化和跨不同类型实例的泛化。

1)跨同类型实例的泛化:DQN-GSF 和 PPO-GSF 在实例 R101 上训练的策略用于验证它们对其他类型 R 实例的泛化。

表 XVI 中将这些策略应用于其他五个实例的结果证明了良好的泛化程度。NV 表示车辆数量,TD 表示总距离。由 DQN-GSF 训练的策略导致与实例 R202 的差距小于 2%。使用 PPO-GSF,GAP 在所有情况下都小于 3%,获得的结果与文献中最著名的结果相当 [39]。

2)跨不同类型实例的泛化:通过将 DQN-GSF 和 PPO-GSF 在实例 R101 上训练的策略的通用性直接应用于具有不同特征的 type-C 和 type-RC 实例来验证。表 XVII 和其他 12 个实例中的结果再次证明了训练策略的泛化。对于 type-C 实例,所有 GAP 值都等于 0,这意味着 DQN-GSF 和 PPO-GSF 的训练策略可以产生当前最好的已知解决方案。在 type-RC 实例上,在大多数情况下,经过训练的策略可以获得最佳已知解决方案的非支配解决方案。例如,在 RC101 上,经过训练的 PPO-GSF 策略可以获得 15 辆车总行驶距离为 1671.46 的解,而最著名的解是 14 辆车总行驶距离为 1696.94。

总之,实验结果表明,由 DQN-GSF/PPO-GSF 自动设计的算法能够针对相同和不同类型的不同问题实例生成高质量的解决方案。这表明所提出的框架对于不同的场景是可靠的,这是自动化算法设计的目标。

五.结论

在这项研究中,首先建立了一个通用搜索框架 (GSF)来制定不同的元启发式算法,包括基于单一解决方案的算法和基于群体的算法。在已建立的统一 GSF 中设计了强化学习方法、深度 Q 网络 (DQN) 和近端策略优化 (PPO),通过在不同的过程中智能选择算法组件 (即进化算子)的适当组合来自动设计基于种群的算法优化过程的各个阶段。通过从与环境的交互 (优化过程)中学习,所提出的模型表明能够在 GSF 内有效地设计算法。

所提出的两种强化学习模型的性能已经在具有时间窗口的受限车辆路径问题的不同基准上进行了评估,以研究它们的有效性和通用性。关于学习模型的有效性,对 Q 值函数逼近器和策略更新机制的研究表明,具有神经网络函数逼近器 (即 PPO) 的基于策略的模型更适合自动设计搜索算法。关于通用性,在一个实例上学习的策略适用于相同类型和不同类型的实例。结果验证了 DQN-GSF 和 PPO GSF 模型训练策略的普遍性。这为基于统一通用搜索框架内的基本算法组件设计算法学习可重用的新知识提供了有希望的证据。

对于未来的工作,可以扩展所提出的通用搜索框架来制定多目标进化算法,以支持多目标算法的自动化设计。解决方案空间和目标空间中种群多样性的精确测量,以及算法组合搜索空间的适应度景观分析,可以进一步识别依赖于搜索的特征以更好地表示状态,增强基于强化学习的方法朝着有效方向发展学习算法设计。

附录

DQN-GSF 中使用的神经网络的细节和 PPO-GSF 分别如图 6 和图 7 所示。在图 6 中,状态和动作分别作为 Q 网络、评估网络和目标网络的输入(1, 3)。目标网络的参数每 n 集被评估网络替换(11)。Q 网络的输出是所有动作(2, 4)的一组 Q 值。动作由最大 Q 值(5)决定。所选动作 a 在环境中执行(6),称为一步。在每一步生成的 (st, rt, at, st+1)被存储在回放缓冲器(7)中。计算评估网络参数的损失(8)。Q 个网络(10, 11)更新后,数据流回1在图7中,状态作为actor神经网络的输入(1),其输出为概率分布所有动作(2)。动作由获得的概率分布 (5)决定。所选动作 a 被执行

实例的概括。表示PPO-GSF与DON-GSF显著不同, $IE, p < 0.05$

文献中最著名的解决方案			DQN-GSF					PPO-GSF										
			转发[40]	汞[41]	(训练有素的政策)			(训练有素的政策)										
实例	NV	TD	AVG		标清	最好的	平均差距	标清	最好的	差距								
R102	17	1486.12 [40]	18486.12	18490.39	19198.08	455.54	18499.58	0.07%	18918.03	486.78	18507.80	0.12%	15377.11	15377.11	16446.11	438.62	15475.15	0.64%
R105	14	1377.11 [40]	5325.66	1.40%	5131.13	22.41%	5103.13	4069.41					16343.12	* 297.87	15451.57	0.48%	5344.44	24.74
R201	4	3	1252.37 [41]	5281.58	1191.70	5252.37	5369.89	39.99					5311.65	1.13%	273.80	4286.01	2.25%	26.08
R202		3	[43] 5088.07	994.42	[43]	4198.45	5180.19	31.71										
R205			4063.24			4162.06	4165.78	58.63			1.88%	4135.23				4096.00		2.54%

表 XVI 对不同类型

实例的概括。表示PPO-GSF与DON-GSF显著不同, IE, $p < 0.05$

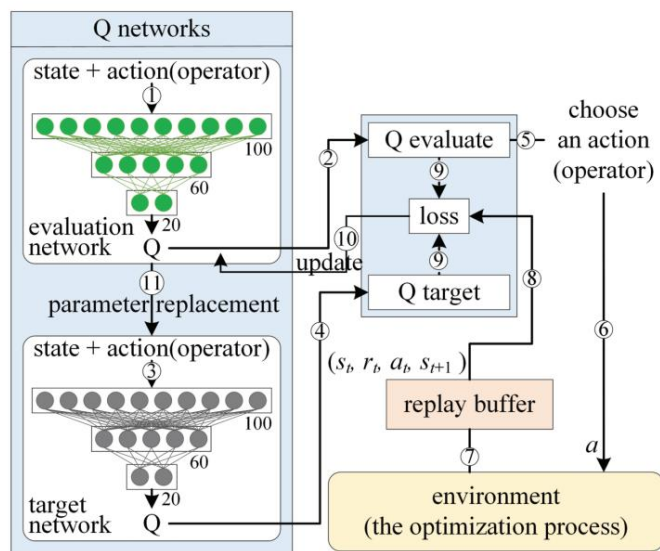
[illegible]

图 6. DON-GSF 方法的神经网络细节

(6) 在环境中,称为一步。在每一步生成的(st, rt, at, st+1)被存储在回放缓冲器(7)中。通过最小化优势函数值来更新(8)评价神经网络的参数。另一方面,状态和奖励被作为critic神经网络的输入(3),当前状态(4)的输出优势函数值,指导actor神经网络(9)的更新方向。然后,数据流回1

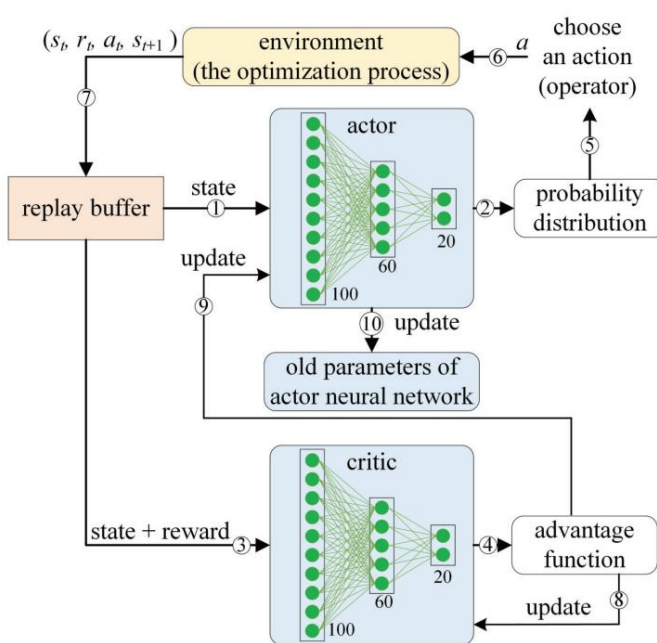


图 7. PPO-GSF 方法的神经网络细节

神经网络是算法中的关键超参数。
具体来说,自适应调整学习率,开始时设置为0.002,输出稳定后减半。

贴现率设置为 0.99,因此学习的策略更侧重于顺序决策。网络的拓扑结构是根据问题的复杂程度和数量来设置的

在 RL 中,学习率、折扣率和

层和神经元的结构如图 6 和图 7 所示。

参考

[1] N. Pillay,R. Qu,D. Srinivasan,B. Hammer 和 K. Sorensen,“机器学习 and 搜索算法的自动设计 [客座编辑]”,IEEE 计算智能杂志,卷。 13,没有。 2,第 16–17 页,2018 年。

[2] HJ Escalante, Q. Yao, W.-W. Tu, N. Pillay, R. Qu, Y. Yu 和 N. Houlsby,“客座社论:自动机器学习”,IEEE 模式分析和机器智能交易,卷。 43,没有。 9,第 2887–2890 页,2021 年。

[3] R. Qu,G. Kendall 和 N. Pillay,“一般组合优化问题:走向自动化算法设计”,IEEE 计算智能杂志,卷。 15,没有。 2,第 14–23 页,2020 年。

[4] RS Sutton 和 AG Barto,强化学习:介绍。麻省理工学院出版社,2018 年。

[5] SS Choong, L.-P. Wong 和 CP Lim,“基于强化学习的超启发式自动设计”,信息科学,卷。 436,第 89–107 页,2018 年。

[6] CJ Watkins 和 P. Dayan,“Q-learning”,机器学习,卷。 8,没有。 3–4,第 279–292 页,1992 年。

[7] JE Pettinger 和 RM Everson,“通过强化学习控制遗传算法”,载于第四届遗传和进化计算年会论文集,2002 年,第 692–692 页。

[8] F. Chen, Y. Gao, Z.-q. Chen 和 S.-f. Chen,“SCGA:使用 SARSA (0) 控制遗传算法”,国际建模、控制和自动化计算智能会议和智能代理、网络技术和互联网商务国际会议 (CIMCA-IAWTIC 06),卷。 1. IEEE, 2005 年,第 1177–1183 页。

[9] V. Mnih,K. Kavukcuoglu,D. Silver,AA Rusu,J. Veness,MG Bellemare, A. Graves, M. Riedmiller, AK Fiedelnd, G. Ostrovski 等人,“通过深度强化学习实现人类水平的控制”,自然,卷。 518,没有。 7540,第 529–533 页,2015 年。

[10] J. Schulman,F. Wolski,P. Dhariwal,A. Radford 和 O. Klimov,“Proximal policy optimization algorithms”,arXiv 预印本 arXiv:1707.06347,2017 年。

[11] F. Hutter,HH Hoos 和 T. Stutzle,“自动算法配置基于本地搜索”,在 Aaai,卷。 7, 2007, pp. 1152–1157。

[12] M. Birattari,Z. Yuan,P. Balaprakash 和 T. Stutzle,“F-Race 和迭代 F-Race:概述”,优化算法分析的实验方法,第 311–336 页,2010 年。

[13] M. Lopez-Ibanez,J. Dubois-Lacoste,L.P. Caceres,M. Birattari 和 T. Stutzle,“irace 包:用于自动算法配置的迭代竞赛”,运筹学展望,卷。 3,第 43–58 页,2016 年。

[14] T. Bartz-Beielstein,CW Lasarczyk 和 M. Preuß,“顺序参数优化”,2005 年 IEEE 进化计算大会,卷。 1. IEEE,2005 年,第 773–780 页。

[15] F. Hutter,HH Hoos 和 K. Leyton-Brown,“基于序列模型的通用算法配置优化”,国际学习和智能优化会议,斯普林格,2011 年,第 507–523 页。

[16] B. van Stein,H. Wang 和 T. Back,“Automatic configuration of deep neural networks with EGO”,arXiv 预印本 arXiv:1810.05526,2018 年。

[17] J. Bergstra,D. Yamins,DD Cox 等人,“Hyperopt:用于优化机器学习算法超参数的 Python 库”,第 12 届 Python 科学会议论文集,卷。 13. Citeseer, 2013, p. 20。

[18] K. Tang,F. Peng,G. Chen 和 X. Yao,“具有自动成分算法选择的基于人口的算法组合”,信息科学,卷。 279,第 94–104 页,2014 年。

[19] L. Xu,H. Hoos 和 K. Leyton-Brown,“Hydra:为基于投资组合的选择自动配置算法”,第 20 届 AAAI 人工智能会议,2010 年。

[20] JC Ortiz-Bayliss,J. Amaya,JM Cruz-Duarte,AE Gutierrez Rodriguez,SE Conant-Pablos 和 H. Terashima-Marín,“基于机器学习的通用框架,用于约束满足中的算法选择问题”,应用科学,卷。 11,没有。第 6 页2749 年,2021 年。

[21] P. Cowling,G. Kendall 和 E. Soubeiga,“安排销售峰会的超启发式方法”,在国际自动时间表实践和理论会议上,斯普林格,2000 年,第 176–190 页。

[22] EK Burke,M. Gendreau,M. Hyde,G. Kendall,B. McCollum,G. Ochoa,AJ Parkes 和 S. Petrovic,“跨域启发式搜索挑战——一项国际研究竞赛”,载于学习与智能优化国际会议,斯普林格,2011 年,第 631–634 页。

[23] N. Pillay 和 D. Beckedahl,“EvoHyp-A 用于进化算法超启发式的 Java 工具包”,2017 年 IEEE 进化计算大会 (CEC)。 IEEE,2017 年,第 2706–2713 页。

[24] SL Tilahun 和 MA Tawhid,“Swarm 超启发式框架”,启发式杂志,卷。 25,没有。 4,第 809–836 页,2019 年。

[25] JM Cruz-Duarte,J. Amaya-JC Ortiz-Bayliss,SE Conant Pablos,H. Terashima-Marín 和 Y. Shi,“为持续优化定制元启发式的超启发式”,Swarm 和进化计算,卷。 66,页。 100935,2021 年。

[26] S. v. Rijn,C. Doerr 和 T. Back,“走向自适应 CMA-ES 配置器”,来自自然并行问题解决国际会议,斯普林格,2018 年,第 54–65 页。

[27] R. Boks,H. Wang 和 T. Back,“粒子群优化和微分进化的模块化混合”,载于 2020 年遗传和进化计算会议论文集,2020 年,第 1418–1425 页。

[28] A. Buzdalova,V. Kononov 和 M. Buzdalov,“使用强化学习选择进化算子:初步探索”,载于 2014 年遗传和进化计算年会配套出版物的论文集,2014 年,pp. 1033–1036。

[29] Y. Sakurai,K. Takada,T. Kawabe 和 S. Tsuruta,“使用强化学习控制进化算法参数的方法”,2010 年第六届信号图像技术和基于互联网的系统国际会议。 IEEE,2010 年,第 74–79 页。

[30] J. Schuchardt,V. Golkov 和 D. Cremers,“学习进化”,arXiv 预印本 arXiv:1905.03389,2019 年。

[31] G. Duflo, G. Danoy, E.-G. Talbi 和 P. Bouvry,“高效蜂群行为的自动设计:一种 Q 学习超启发式方法”,载于 2020 年遗传和进化计算会议论文集,2020 年,第 227–228 页。

[32] O. Brayssy 和 M. Gendreau,“带时间窗的车辆路径问题,第二部分:元启发式”,交通科学,卷。 39,没有。 1,第 119–139 页,2005 年。

[33] B. Chen,R. Qu,R. Bai 和 H. Ishibuchi,“VRPTW 复合社区调查”,运筹学和企业系统国际会议,斯普林格,2016 年,第 3–19 页。

[34] S. Kirkpatrick,“模拟退火优化:定量研究”,统计物理学杂志,卷。 34,没有。 5,第 975–986 页,1984 年。

[35] MM Solomon,“具有时间窗约束的车辆路径和调度问题的算法”,运筹学,卷。 35,没有。 2,第 254–265 页,1987 年。

[36] J. Baras 和 V. Borkar,“具有自适应状态聚合的马尔可夫决策过程的学习算法”,第 39 届 IEEE 决策与控制会议论文集 (目录号 00CH37187),卷。 4. IEEE,2000 年,第 3351–3356 页。

[37] T. Mori 和 S. Ishii,“强化学习中价值函数估计的增量状态聚合”,IEEE Transactions on Systems, Man, and Cybernetics,B 部分 (控制论),卷。 41,没有。 5,第 1407–1416 页,2011 年。

[38] SP Singh,T. Jaakkola 和 MI Jordan,“使用软状态聚合进行强化学习”,神经信息处理系统进展,第 361–368 页,1995 年。

[39] SINTEF,“VRPTW 基准问题,在 sintef 传输优化门户上”,https://www.sintef.no/projectweb/top/vrptw/solomon_benchmark/100-customers/,2008 年。

[40] Y. Rochat 和 ED Taillard,“车辆路径局部搜索中的概率多样化和强化”,启发式杂志,卷。 1,没有。 1,第 147–167 页,1995 年。

[41] J. Homberger 和 H. Gehring,“具有时间窗的车辆路径问题的两种进化元启发式”,INFOR:信息系统和运筹学,卷。 37,没有。 3,第 297–318 页,1999 年。

[42] J. Homberger,“Eine verteilt-parallele metaheuristik”,在 Verteilt parallele Metaheuristiken zur Tourenplanung 中。 Springer,2000 年,第 139–165 页。

[43] L.-M. Rousseau,M. Gendreau 和 G. Pesant,“使用基于约束的算子解决带有时间窗的车辆路径问题”,启发式杂志,卷。 8,没有。 1,第 43–58 页,2002 年。

[44] E. Taillard,P. Badeau,M. Gendreau,F. Guertin 和 J.-Y. Potvin,“具有软时间窗的车辆路径问题的禁忌搜索启发式”,交通科学,卷。 31,没有。 2,第 170–186 页,1997 年。

© 2022 IEEE.允许个人使用,但再版/再分发需要 IEEE 许可.有关详细信息,请参阅 https://www.ieee.org/publications/rights/index.html。 授权许可使用限于:河北工业大学.于 2022 年 8 月 16 日 09:52:15 UTC 从 IEEE Xplore 下载.限制适用。

[45] J. Berger 和 M. Barkaoui, “具有时间窗的车辆路径问题的并行混合遗传算法”,计算机与运筹学,卷。 31,没有。 12,第 2037–2053 页,2004 年。

..

[46] D. Mester,O. Braysy 和 W. Dullaert, “车辆路径问题的多参数进化策略算法”,专家系统与应用,卷。 32,没有。 2,第 508-517 页,2007 年。

[47] ZJ Czech 和 P. Czarnas, “带时间窗的车辆路径问题的并行模拟退火”,第 10 届 Euromicro 并行、分布式和基于网络处理的研讨会论文集。 IEEE,2002 年,第 376–383 页。



牛奔 (会员,IEEE)获得博士学位。 2008 年毕业于中国科学院沈阳自动化研究所,中国沈阳。

现任深圳大学管理科学系广东省珠江学者特聘教授。他在国际期刊和国际会议上发表了 150 多篇论文。主要研究领域为演化计算和运筹学。



Wenjie Yi获得理学学士学位。和理学硕士分别于 2016 年和 2019 年在中国深圳的深圳大学获得学位。她目前正在攻读博士学位。英国诺丁汉诺丁汉大学计算机科学学院计算机科学学士学位

她目前的研究兴趣包括进化计算、机器学习、自动化算法设计和车辆路径问题。



Rong Qu (高级会员,IEEE)获得理学学士学位。 1996 年在中国西安的西安电子科技大学获得计算机科学及其应用学士学位,并于 1996 年获得博士学位。 2003 年获得英国诺丁汉大学计算机科学学士学位。

她是诺丁汉大学计算机科学学院的副教授。她的研究兴趣包括通过使用 EA、数学规划、运筹学和人工智能中的约束规划对物流运输调度、人员调度、网络路由、投资组合优化和时间表问题进行建模和优化。这些计算技术与知识发现、机器学习和数据挖掘相结合,为中小企业的物流车队运营、医院的劳动力调度、教育决策以及联网和自动驾驶汽车的网络安全提供智能决策支持。

曲博士是 IEEE Computational Intelligence Magazine,IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION,Journal of Operational Research Society 和 PeerJ Computer Science 的副主编。

自 2019 年以来,她一直担任进化计算任务委员会副主席。她是 2015 年至 2018 年 IEEE 计算智能学会智能系统应用技术委员会副主席。



Licheng Jiao (Fellow, IEEE) 于 1982 年在中国上海的上海交通大学获得学士学位,并获得硕士和博士学位。分别于 1984 年和 1990 年在中国西安的西安交通大学获得学位。

1992年起任西安电子科技大学电子工程学院教授,现任智能感知与图像理解教育部重点实验室主任。

他的研究兴趣包括图像处理、自然计算、机器学习和智能信息处理。

焦博士是中国电子学会理事;中国神经网络委员会委员;国务院学位委员会专家,奖励与表彰委员会主任委员;中国人工智能学会副理事长;欧洲科学院和俄罗斯自然科学院外籍院士; IET、CAAI、CIE、CCF 和 CAA 院士。