

# 基于近似边界和聚类的昂贵多目标优化算法

张 峰

(中国电子科技集团公司第二十八研究所,江苏 南京 210007)

**摘 要:**不少工程优化问题涉及同时优化超过2个冲突的目标,并且目标函数的评估比较耗时,这类问题就属于昂贵的超多目标优化问题。由于目标空间过于庞大,并且只能使用较少的目标函数评估次数进行求解,这使得算法的收敛速度比较缓慢以及难以保持较好多样性。此外,许多算法往往忽略使用极值点的有效信息来加速算法收敛。为了解决上述问题,该文在一种新颖的超多目标进化算法的基础上,提出了一种基于近似边界和聚类的昂贵多目标优化算法。通过使用一组高斯过程近似目标函数来辅助算法进行评估,算法还使用极值点来加速收敛并优化出一个较好的候选种群,然后进一步提出使用一种评价指标来批量挑选出一些最有价值的候选解,借此使得算法能够保持较好的收敛性和多样性。最后通过与多个流行的求解昂贵超多目标优化问题算法进行对比实验,证明了算法的有效性。

**关键词:**昂贵的超多目标优化问题;极值点;超多目标进化算法;昂贵多目标优化算法;高斯过程

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2022)0021-05

## Expensive Multi-objective Optimization Algorithm Based on Approximate Boundary and Clustering

ZHANG Feng

(28th Research Institute of China Electronics Technology Group Corporation, Nanjing 210007, China)

**Abstract:** Many engineering optimization problems involve optimizing more than two conflicting objectives at the same time, and the evaluation of the objective function is time-consuming. This type of problem is the expensive many-objective optimization problem. Because the objective space is too large, and only a small number of objective function evaluation times can be used for solving, this makes the convergence speed of the algorithm slower and it is difficult to maintain good diversity. In addition, many algorithms often neglect to use the effective information of extreme points to accelerate the algorithm's convergence. In order to solve the above problems, we propose an expensive multi-objective optimization algorithm based on approximate boundaries and clustering on the basis of a novel many-objective evolutionary algorithm. By using a set of Gaussian processes to approximate the objective function to assist the algorithm for evaluation, the algorithm also uses extreme points to accelerate convergence and optimize a better candidate population. Then we further propose to use an evaluation indicator to select some of the most valuable candidate solutions in batches, so that the algorithm can maintain better convergence and diversity. Finally, through comparative experiments with a number of popular algorithms for solving expensive many-objective optimization problems, the effectiveness of the algorithm is proved.

**Key words:** expensive many-objective optimization problem; nadir point; many-objective evolutionary algorithm; expensive multi-objective optimization algorithm; Gaussian process

## 0 引 言

许多实际工程优化问题通常都面临同时优化多个相互冲突的目标,并且由于一些特殊的原因,导致目标函数的评估需要进行十分耗时的仿真实验,这会造成目标函数的评估变得十分昂贵,这一类问题就统称为昂贵的多目标优化问题(expensive multiobjective optimization problem, EMOP)<sup>[1]</sup>。当冲突目标数大于2时,这类问题可以称为昂贵的超多目标优化问题。昂贵的超多目标优化问题广泛存在于工业生产和科学研

究中,因此,高效地求解昂贵的超多目标优化问题拥有十分重要的理论价值。

## 1 概 述

### 1.1 昂贵的目标优化问题定义

一个最小化的连续昂贵多目标优化问题(EMOP)的数学表达式如下:

$$\begin{aligned} & \text{minimize } F(x) = (f_1(x), \dots, f_m(x)) \\ & \text{subject to } x \in \Omega \end{aligned} \quad (1)$$

收稿日期:2021-06-02

作者简介:张 峰(1993-),男,研究方向为智能计算与机器学习。

其中,决策空间用  $\Omega \in R^n$  来表示,  $F: \Omega \rightarrow R^m$  是由  $m$  个实值目标函数组成的目标空间,  $x = (x_1, \dots, x_n)^T \in \Omega$  则用来表示 EMOP 中的一个解。

## 1.2 相关定义简介

假设存在两个解  $a, b \in \Omega$ , 做出如下定义, 当且仅当对  $\forall i \in \{1, \dots, m\}$  都有  $f_i(a) \leq f_i(b)$ , 并且至少存在  $\exists j \in \{1, \dots, m\}$  使得  $f_j(a) < f_j(b)$  时, 则称  $a$  支配  $b$  并记为  $a < b$ 。再假设存在一个解  $x^* \in \Omega$ , 若不存在一个解  $x \in \Omega$  使得  $x < x^*$ , 则称  $x^*$  为 Pareto 最优解,  $F(x^*)$  则为 Pareto 最优目标向量。所有  $x^*$  形成的集合称为 Pareto Set (PS), 其对应的目标向量则称为 Pareto Front (PF) [2]。

边界一般包括理想点  $z^*$  (Ideal Point) 和极值点  $z^{\text{nad}}$  (Nadir Point)。这两个点的定义如下所示:

$$z_j^* = \min_{x \in \Omega} f_j(x), j \in \{1, \dots, m\} \quad (2)$$

$$z_j^{\text{nad}} = \max_{x \in PS} f_j(x), j \in \{1, \dots, m\} \quad (3)$$

昂贵的多目标优化算法最终会优化出一组近似 PF 的解集, 决策者可以根据需求来挑选出一个合适的解进行决策。

昂贵的多目标优化算法性能则是通过衡量其近似 PF 的收敛性和多样性来评价的。收敛性和多样性分别是指算法近似出来的 PF 距离真实 PF 的远近, 以及能否较好地表示整个 PF。

## 1.3 昂贵多目标优化算法研究综述

多目标进化算法被认为是一种高效求解多目标优化问题 [3-5] 的有效方法, 其主要通过进行大量的目标函数评估来有效地求解问题。然而, 由于 EMOP 往往只能进行少量的目标函数评估, 多目标进化算法难以有效地求解问题。在此背景下, 许多相应算法被设计出来求解 EMOP。

多目标进化算法 NSGA-II [6] 主要通过快速非支配排序和计算拥挤距离, 在求解低目标的 EMOP 上仍有较为理想效果。随着机器学习领域的发展, 许多学者开始使用机器学习的方法来构建代理模型进行辅助算法评估, Knowles 提出了一个经典的昂贵多目标进化算法 ParEGO [7], 其主要思想是随机挑选一个子问题并使用高斯过程 [8] 建立代理模型进行辅助选解, 最后根据贝叶斯优化 [9] 的相关思想, 挑选出最有价值的一个解进行评估, 通过优化多个单目标子问题的最优解来近似出一组多目标优化问题的 PF。随后, Zhang 等人提出的 MOEA/D-EGO [10] 则使用多个高斯过程近似目标函数来辅助选解, 并在 MOEA/D-DE [11] 的基础上进行协同优化出一组解, 并挑选出一部分最有前途的解进行真实目标函数评估, 与 ParEGO 不同的是, MOEA/D-EGO 可以一次批量优化出多个解, 方便结

合并行技术来同时评估多个解, 以提升算法运行效率。在针对目标数大于 3 的 EMOP 上, Chugh 等人提出的 K-RVEA [12], 虽然同样使用高斯过程来近似目标函数, 但在代理模型管理上, 则根据高斯过程预测的不确定性、参考向量分布的情况还有解的位置来权衡算法是趋向收敛性还是多样性, 此外还提出一种挑选训练样本的方法来节省高斯过程训练模型的时间, 通过以上方法使得 K-RVEA 能够有效地求解昂贵的超多目标优化问题。尽管上述算法能够有效地求解一些 EMOP, 但缺乏有效地利用边界点的有用信息来提升进化效率, 此外, 随着目标数的上升, 导致目标空间过于庞大, 算法难以较好地保持收敛性和多样性, 如何使得算法能够高效地求解 EMOP 问题, 依然是一个值得探讨的问题。

## 1.4 基于边界和聚类的超多目标进化算法

在 Singh 等人提出的一种角点解 [13] 的基础上, 有人提出了一种基于近似边界和层次聚类的超多目标进化算法 (many-objective evolutionary algorithm based on approximate boundary and hierarchical clustering, MaOEA-ABHC) [14], MaOEA-ABHC 首先通过近似极值点来加速算法收敛, 并通过层次聚类的方法自适应地产生一组权重向量来确保算法的收敛性和多样性。通过以上方法结合, 该算法能够有效求解一些目标数大于 3 的多目标优化问题。下面简单介绍下 MaOEA-ABHC 的算法流程, 更多详细的内容可以参考这篇论文。

算法 1: MaOEA-ABHC。

输入: 初始种群  $C$

输出: 优化后的种群  $C$

1. while 没到达终止条件 do
2. 对  $C$  使用 SBX 算子产生子代  $C_o$ ;
3.  $C_u = C \cup C_o$ ;
4. 使用  $C_u$  近似  $z^*$ ;
5. 使用  $C_u$  求角点解集  $S_c$  并近似  $z^{\text{nad}}$ ;
6. 对  $C_u$  进行层次聚类, 并根据聚类中心产生若干权重向量, 根据聚合函数值挑选出新种群  $C$ ;
7. end
8. return  $C$ 。

## 1.5 研究动机

在求解目标数大于 3 的 EMOP 时, 目标空间随着目标数的上升, 而变得十分庞大, 但算法往往只能使用有限的评估次数来求解 EMOP, 这使得算法的收敛性能受到严重的影响并且难以保持较好的多样性。此外, 许多算法缺乏有效利用边界 (极值点) 的有用信息来加速算法收敛过程。该文在 MaOEA-ABHC 选解方法的基础上, 使用多个高斯过程来近似目标函数, 并使用贝叶斯优化中的效用准则来考虑代理模型预测的不

确定性,将效用准则值取代预测目标函数值进行选择,借此,将一些有用信息通过各个目标进行传递来引导搜索,最后批量挑选出对当前种群有价值的一部分解进行真实的目标函数评估,从而保证算法的收敛性和多样性。

## 2 算法设计

### 2.1 算法框架

这章节主要介绍提出的基于近似边界和聚类的昂贵多目标优化算法(EMOA-ABC)的实现细节,EMOA-ABC 的算法框架如下所示:

算法 2:EMOA-ABC 算法框架。

输入:决策变量个数  $n$ ; 目标变量个数  $m$ ; 候选种群数  $N$ ; 批量挑选候选解数  $K_E$ ; 算法的终止条件

输出:所有评估过的解集  $P$

1. 使用 Latin hypercube routine 生成初始种群  $P = \{(x_i, F(x_i)) \mid i = 1, \dots, 11n - 1\}$ ;
2. while 没到达终止条件 do
3.   使用  $P$  训练  $m$  个高斯过程 GP 来近似目标函数;
4.   从  $P$  随机挑选  $N$  个解作为候选种群  $C$ ;
5.   使用 MaOEA-ABHC 来优化  $C$ ;
6.    $C_B = \text{GetBestSols}(P, C, K_E)$ ;
7.   评估  $C_B$ , 且  $P = P \cup C_B$ ;
8. end
9. return  $P$ 。

EMOA-ABC 主要分为 3 个阶段:

(1) 初始化:主要使用 latin hypercube routine 来生成一个规模为  $11n - 1$  的初始种群  $P$ , 并使用真实的目标函数评估来  $P$ 。

(2) 优化候选种群:首先使用  $P$  来训练  $m$  个高斯过程 GP 来近似目标函数,然后从  $P$  中随机挑选  $N$  个解作为初始的候选种群  $C$ , 再使用 MaOEA-ABHC 来优化  $C$ 。

(3) 挑选最佳候选解:从优化出来的候选种群  $C$  中批量挑选出  $K_E$  个最有价值的候选解集  $C_B$  进行真实的目标函数评估。

### 2.2 高斯过程

高斯过程是一种能够有效解决小规模昂贵多目标优化问题的代理模型,假设  $P = \{x^1, \dots, x^{T_n}\}$  是训练样本,  $T_n$  是所有评估过的解的个数,高斯过程表达式如下所示:

$$y = \mu(x) + \varepsilon(x) \quad (4)$$

其中,  $\varepsilon(x) \sim N(0, \sigma^2)$ ,  $\mu(x)$  表示  $p$  个选择函数和  $\beta$  个系数的回归模型  $F(\beta, x) = \beta_1 c_1(x) + \dots + \beta_p c_p(x)$ 。

$\forall x^1, x^2 \in P$ ,  $x^1$  和  $x^2$  的协方差由下式计算得出:

$$\text{cov}[\varepsilon(x^1), \varepsilon(x^2)] = \sigma^2 C_T([d(x^1, x^2)]) \quad (5)$$

其中:

$$d(x^1, x^2) = \exp\left(\sum_{i=1}^n -\theta_i |x_i^1 - x_i^2|^2\right) \quad (6)$$

$n$  是决策变量数,  $C_T$  是  $T_R \times T_R$  的协方差矩阵,  $\theta_1, \dots, \theta_n$  是一组超参数。

一个待测点  $x$  的近似值可由下式计算:

$$\tilde{f}(x) = \beta + c^T(x) C_T^{-1}(y - F\beta) \quad (7)$$

$c(x)$  是  $x$  和  $P$  的协方差向量, 根据上式具有广义最小二乘解, 可得  $\beta = (F^T C_T^{-1} F)^{-1} F^T C_T^{-1} y$ 。

此外, 预测方差  $\sigma^2$  则为:

$$\sigma^2(x) = \frac{(y - F\beta)^T C_T^{-1}(y - F\beta)}{T_R} \quad (8)$$

可以通过最大化下式来求出超参数  $\theta$ 。

$$\Psi(\theta) = -\frac{N_R \ln \sigma^2 + \ln |C_T|}{2} \quad (9)$$

该文使用  $P$  做训练样本, 训练出  $m$  个高斯过程来近似  $m$  个目标函数。

### 2.3 优化候选种群

该文使用在处理超多目标优化问题上性能良好的 MaOEA-ABHC 优化候选种群  $C$ 。此外, MaOEA-ABHC 迭代优化过程中, 使用 LCB 效用准则<sup>[9]</sup>代替近似的目标函数值进行挑选种群, LCB 效用准则的表达式如下所示:

$$\text{LCB}(x) = \mu(x) - \alpha \varepsilon(x) \quad (10)$$

其中,  $\alpha$  是参数, 用来权衡开发与探索, 文中  $\alpha = 1$ 。

### 2.4 批量挑选最佳候选解

在获得优化后的候选种群  $C$  后, 使用 Hypervolume Indicator<sup>[15]</sup> (HV) 来批量挑选最佳候选解进行真实的目标函数评估。

算法 3: GetBestSols。

输入: 已评估解集  $P$ ; 优化后的候选种群  $C$ ; 批量挑选最佳候选解个数  $K_E$

输出: 最佳候选解集  $C_B$

1.  $D_P = \text{nondominated-sorting}(P)$ ;
2.  $H = \emptyset$ ;
3. For each  $x \in C$  do
4.   使用 LCB 效用准则取代高斯过程预测  $x$  的目标函数值;
5.    $D_T = D_P \cup x$ ;
6.    $H_i = \text{HV}(D_T, D_P)$ ;
7. end
8. 根据  $H$  挑选出最佳的  $K_E$  个解, 记为  $C_B$ ;
9. return  $C_B$ 。

正如算法 2 所示, 首先对所有评估过的解集  $P$  取非支配操作, 获得非支配解集  $D_P$ , 对候选种群  $C$  中的每一个解  $x$ , 使用 LCB 效用准则取代高斯过程近似的目标函数值进行计算, 将  $x$  和  $D_P$  进行合并, 组成一个临时种群  $D_T$ , 通过计算  $D_T$  的 HV 指标值来度量的  $x$

潜在价值,所有解的 HV 指标值保存在  $H$ ,并挑选出最佳的  $K_e$  个对应的候选解保存到  $C_B$  中。最后使用真实的目标函数进行评估。

### 3 实验设计

#### 3.1 实验参数设置

选择 DTLZ 系列测试问题<sup>[16]</sup>来检验 EMOA-ABC 的算法性能。此外,选择 4 个求解昂贵超多目标优化问题的流行算法,来和 EMOA-ABC 进行对比,这些对比算法主要包括 NSGA-II、MOEA/D-EGO、K-RVEA 和 ParEGO。每个算法都独立地运行目标变量数  $m$  为 3 和 5 的 DTLZ 测试问题 10 次。文中的决策变量数  $n$  均为 8。所有算法的种群数都设为 150,并且每个测试问题的最大目标函数评估次数都设为 300 次。每个算法都把所有评估过的解取非支配操作后再进行计算指标。

除了上文提到的 HV 指标外,还有 Inverted

表 1 所有算法在全部测试问题上的 IGD 均值、显著性符号和排名(每个问题的最佳指标值加粗显示)

测试问题	$m$	EMOA-ABC	MOEA/D-EGO	K-RVEA	ParEGO	NSGA-II
DTLZ1	3	<b>36.481 6(1)</b>	57.823 9 +(4)	43.745 9 =(2)	51.949 1 +(3)	62.674 7 +(5)
DTLZ2	3	<b>0.058 0 (1)</b>	0.270 8 +(5)	0.105 7 +(2)	0.259 5 +(4)	0.233 1 +(3)
DTLZ3	3	<b>96.106 5(1)</b>	128.403 3 +(3)	139.186 5 +(4)	127.641 8 +(2)	191.932 9 +(5)
DTLZ4	3	0.710 1(5)	0.530 2 -(3)	<b>0.206 3 -(1)</b>	0.395 2 -(2)	0.609 2 -(4)
DTLZ5	3	<b>0.006 0(1)</b>	0.188 6 +(5)	0.038 8 +(2)	0.151 8 +(3)	0.164 9 +(4)
DTLZ6	3	1.418 2(3)	<b>1.279 1 =(1)</b>	2.120 1 +(4)	1.280 6 =(2)	4.586 8 +(5)
DTLZ7	3	0.241 8(4)	<b>0.136 4 =(1)</b>	0.163 9 =(3)	0.154 4 =(2)	4.136 9 +(5)
DTLZ1	5	<b>14.254 9(1)</b>	22.576 1 +(3)	15.034 0 =(2)	22.880 7 +(4)	24.347 6 +(5)
DTLZ2	5	<b>0.177 3(1)</b>	0.268 9 +(4)	0.187 9 +(2)	0.294 8 +(5)	0.241 5 +(3)
DTLZ3	5	<b>43.105 1(1)</b>	72.008 0 +(5)	49.039 7 =(2)	71.761 6 +(4)	67.510 6 =(3)
DTLZ4	5	0.682 8(4)	0.578 1 -(3)	<b>0.339 6 -(1)</b>	0.554 5 -(2)	0.704 1 =(5)
DTLZ5	5	<b>0.014 6(1)</b>	0.123 4 +(5)	0.030 5 +(2)	0.122 4 +(4)	0.081 6 +(3)
DTLZ6	5	0.686 4(2)	0.876 1 =(3)	0.911 3 +(4)	<b>0.660 3 =(1)</b>	3.014 4 +(5)
DTLZ7	5	<b>0.436 5(1)</b>	0.645 9 +(4)	0.515 5 =(2)	0.529 3 +(3)	5.546 7 +(5)
平均排名		<b>1.93</b>	3.5	2.34	2.93	4.23

### 4 结束语

不少工程优化问题都和昂贵的超多目标优化问题相关。由于目标空间过于庞大,并且算法往往只能使用较少目标函数评估次数,这使得多数算法收敛速度比较缓慢。此外,许多算法往往忽视使用极值点的有用信息来帮助算法加速收敛并提升优化性能。在一个最新求解超多目标优化问题算法的基础上,使用高斯过程进行辅助选解,并使用效用准则来引导搜索,然后进一步提出用评价指标来挑选最佳候选解,保证了算法的收敛性和多样性。在未来,打算使用这个算法来

Generational Distance<sup>[17]</sup>(IGD)等评价指标,该文主要使用 IGD 来衡量所有算法的性能,上述算法均在 PlatEMO<sup>[18]</sup>上实现,并使用 PlatEMO 自带的 HV 计算方法来计算所有算法的 HV 值。此外,还使用置信度为 0.05 的 Wilcoxon 秩和函数来衡量 EMOA-ABC 与其他对比算法的显著性,其中符号 =,表示 EMOA-ABC 和其他对比算法没有显著性区别,而符号+和-分别表示 EMOA-ABC 显著地比其他对比算法好和差。对每个测试问题,根据算法效果进行了排名,并计算相应的平均排名来对比算法性能。

#### 3.2 实验结果

实验结果如表 1 所示,提出的 EMOA-ABC 在大多数测试问题上都取得了最佳的 IGD 值。在大多数测试问题上,EMOA-ABC 的性能都显著地优于其他对比算法,并获取了最佳平均排名。这表明 EMOA-ABC 能够有效求解一些昂贵的超多目标优化问题。

求解一些已经建模完毕的实际工程优化问题,并使用一些近似 HV 指标方法来提升算法运行效率和性能。

#### 参考文献:

- [1] ABATABAEI M, HAKANEN J, HARTIKAINEN M, et al. A survey on handling computationally expensive multiobjective optimization problems using surrogates: non-nature inspired methods[J]. Structural & Multidisciplinary Optimization, 2015, 52(1): 1-25.
- [2] MIETTINEN K. Nonlinear multiobjective optimization[M]. Springer Science & Business Media, 2012.
- [3] DEB K. Multi-objective optimisation using evolutionary al-



- gorithms; an introduction [M]//Multi-objective evolutionary optimisation for product design and manufacturing. London: Springer, 2011: 3–34.
- [4] 康 锰, 许 峰. 多进化策略自适应免疫多目标进化算法 [J]. 安徽理工大学学报: 自然科学版, 2019, 39(5): 43–47.
- [5] 施 展, 赵宗贵, 许 腾. 基于模糊偏好的海军多兵种合同作战资源规划技术 [J]. 指挥信息系统与技术, 2015, 6(5): 68–73.
- [6] DEB K, PRATAP A, AGARWAL S, et al. A fast and elitist multiobjective genetic algorithm: NSGA-II [J]. IEEE Transactions on Evolutionary Computation, 2002, 6(2): 182–197.
- [7] KNOWLES J. ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems [J]. IEEE Transactions on Evolutionary Computation, 2006, 10(1): 50–66.
- [8] STEIN M L. Interpolation of spatial data: some theory for kriging [M]. Springer Science & Business Media, 2012.
- [9] BROCHU E, CORA V M, DE FREITAS N. A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning [J]. arXiv:1012.2599, 2010.
- [10] ZHANG Q, LIU W, TSANG E, et al. Expensive multiobjective optimization by MOEA/D with Gaussian process model [J]. IEEE Transactions on Evolutionary Computation, 2010, 14(3): 456–474.
- [11] LI H, ZHANG Q. Multiobjective optimization problems with complicated pareto sets, MOEA/D and NSGA-II [J]. IEEE Transactions on Evolutionary Computation, 2009, 13(2): 284–302.
- [12] CHUGH T, JIN Y, MIETTINEN K, et al. A surrogate-assisted reference vector guided evolutionary algorithm for computationally expensive many-objective optimization [J]. IEEE Transactions on Evolutionary Computation, 2018, 22(1): 129–142.
- [13] SINGH H K, ISAACS A, RAY T. A Pareto corner search evolutionary algorithm and dimensionality reduction in many-objective optimization problems [J]. IEEE Transactions on Evolutionary Computation, 2011, 15(4): 539–556.
- [14] 张 峰, 顾一凡. 基于近似边界和层次聚类的超多目标进化算法 [J]. 计算机技术与发展, 2020, 30(12): 61–65.
- [15] ZITZLER E, THIELE L. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach [J]. IEEE Transactions on Evolutionary Computation, 1999, 3(4): 257–271.
- [16] DEB K, THIELE L, LAUMANN S M, et al. Scalable test problems for evolutionary multiobjective optimization [M]//Evolutionary multiobjective optimization. London: Springer, 2005: 105–145.
- [17] BOSMAN P A N, THIERENS D. The balance between proximity and diversity in multiobjective evolutionary algorithms [J]. IEEE Transactions on Evolutionary Computation, 2003, 7(2): 174–188.
- [18] TIAN Y, CHENG R, ZHANG X, et al. PlatEMO: a MATLAB platform for evolutionary multi-objective optimization [educational forum] [J]. IEEE Computational Intelligence Magazine, 2017, 12(4): 73–87.
- +++++
- (上接第 20 页)
- text with recurrent neural networks [C]//International conference on machine learning. Bellevue, Washington, USA: The International Machine Learning Society. 2011.
- [7] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks [C]//Advances in neural information processing systems. Montreal, Quebec, Canada: MIT Press, 2014: 3104–3112.
- [8] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks [J]. IEEE Transactions on Signal Processing, 1997, 45(11): 2673–2681.
- [9] GRAVES A, SCHMIDHUBER J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures [J]. Neural Networks, 2005, 18(5–6): 602–610.
- [10] CHO K, VAN MERRIËNBOER B, BAHDANAU D, et al. On the properties of neural machine translation: encoder-decoder approaches [J]. arXiv:1409.1259, 2014.
- [11] MA C, YANG C, YANG F, et al. Trajectory factory: tracklet cleaving and re-connection by deep siamese bi-gru for multiple object tracking [C]//2018 IEEE international conference on multimedia and expo (ICME). San Diego, CA, USA: IEEE, 2018: 1–6.
- [12] TAI K S, SOCHER R, MANNING C D. Improved semantic representations from tree-structured long short-term memory networks [J]. arXiv:1503.00075, 2015.
- [13] YANG Z, YANG D, DYER C, et al. Hierarchical attention networks for document classification [C]//Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. San Diego, California: Association for Computational Linguistics, 2016: 1480–1489.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//Advances in neural information processing systems. Long Beach, CA, USA: MIT Press, 2017: 5998–6008.
- [15] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding [J]. arXiv:1810.04805, 2018.