如何证明粒子群优化算法超参数的最佳值?

Abel C. H. Chen 中华电信股份有限公司
chchen.scholar@gmail.com；ORCID
0000-0003-3628-3033

摘要

近年来，许多群智能优化算法陆续被提出，并应用于解决各种优化问题。然而，群智能优化算法可能有许多超参数值需要设定。例如，粒子群优化算法虽然应用广泛，且具有较高的优化效能，但需要设定惯性速度权重、个体最优解权重、群体最优解权重等多个超参数值。因此，本研究提出一个分析框架，运用数学模型分析粒子群优化算法在不同适应函数下的平均适应函数值，并通过最小化平均适应函数值来确定最合适的超参数值。实验结果显示，采用本研究提出的超参数值，可以获得较高的收敛速度和较小的适应函数值。

关键词：粒子群优化算法、超参数、最佳值。

1. 前言

为解决各种各样的优化问题，许多群智能优化算法应运而生，包括粒子群优化算法(Particle Swarm Optimization, PSO)[1]、海豚群算法(Dolphin Swarm Algorithm, DSA)[2]、灰狼优化算法(Grey Wolf Optimizer, GWO)[3]、遗传算法(Genetic Algorithm, GA)[4] 、人工蜂群算法(Artificial Bee Colony, ABC)[5]、鼠群优化器(Rat Swarm Optimizer, RSO)[6]等。然而，不同的群智能优化算法可能需要设置不同的超参数，这些参数可能会影响收敛速度和优化结果。因此，如何设置群智能优化算法的超参数是一个重要的研究议题。

其中，粒子群优化算法是一种应用广泛且具有较高优化性能的群智能优化算法。在粒子群优化算法中，需要设置惯性速度权重、个体最优解权重、群体最优解权重等多个超参数值。然而，目前粒子群优化算法相关研究中的超参数值大多是通过经验或试错的方式来确定合适的超参数值。但这种方式不一定能找到超参数最优解，而且可能需要较长的计算时间。

有鉴于此，本研究在既有的研究基础[7]上提出一个分析框架，可以分析粒子群优化算法在各个不同适应函数中的平均适应函数值，并且从优化平均适应度函数值来决定最合适的超参数值。同时深入探讨超参数最优解与其他变量和其他超参数之间的关联，以及归纳出粒子群优化算法超参数设置原则。

本论文共分为五个章节。第2 节中介绍粒子群优化算法，以及说明粒子群优化算法的超参数。第 3节提出本研究的分析框架，并且运用分析框架深入讨论粒子群优化算法在三个不同适应函数中的超参数最优解，以及在 3.5节中归纳本研究在粒子群优化算法超参数设置方面的发现和建议。第 4节将提供实例验证本研究的超参数最优解在各个适应函数中的收敛表现。第 5节总结本研究的贡献并讨论未来研究方向。

2. 粒子群优化算法

粒子群优化算法由学者 James Kennedy和 Russell Eberhart 于 1995 年[8]提出，后由学者 YuhuiShi 和Russell Eberhart 于 1998 年[9]修改并完善。该优化算法已广泛应用于各个领域。其中，粒子群优化算法在优化过程中，主要依据适应函数值作为优化目标，并分析惯性速度为v_w、粒子变量 w值与当前个体最优解之间的距离为$o_{p,w}$、粒子变量 w值与当前群体最优解之间的距离为$o_{g,w}$，以优化

变量 w值，如公式(1)所示[8]-[9]。因此，需要设置超参数：惯性速度权重α、个体最优解权重c_1、群体最优解权重c_2，用以调整变量 w值，如公式(2)所示[8]-[9]。其中，$r_{1,w}$和$r_{2,w}$为随机数。

$$v_{w,new} = \alpha v_w + r_{1,w} c_1 o_{p,w} + r_{2,w} c_2 o_{g,w} \quad (1)$$

$$w_{new} = w + v_{w,new} = w + \alpha v_w + r_{1,w} c_1 o_{p,w} + r_{2,w} c_2 o_{g,w} \quad (2)$$

3. 研究方法

本研究旨在证明粒子群优化算法的超参数最优解，提出一个分析框架，通过该分析框架计算平均适应函数值，并根据优化平均适应函数值推导超参数的最优解。此方法可应用于最大化或最小化平均适应函数值，本研究以最小化为例进行说明。在 3.1节中说明分析框架的具体步骤和流程，并在 3.2节和 3.3节分别以单变量适应函数和多变量适应函数说明如何优化超参数。在 3.4节中，以线性回归适应函数分析超参数最优解。最后，在3.5节中总结各个适应函数的推导过程、发现及其差异。

3.1 分析框架

图 1是本研究提出的分析框架及其具体流程。首先定义适应函数，确定求取适应函数的最大值或最小值，并假设适应函数中的变量呈均匀分布时推导平均适应函数值。例如，以粒子群优化算法对神经网络模型进行优化时，神经网络模型中的变量w 和变量 b等在初始值时主要采用随机数，所以在多次实验后将呈现均匀分布。因此，在上述假设下，根据粒子群优化算法推导进行优化后的平均适应函数值。最后，再根据此平均适应函数值进行优化，推导平均适应函数值优化时的超参数最优解。

3.2 适应函数1—单变量分析

为演示如何运用分析框架证明粒子群优化算法超参数最优解，本节以单变量适应函数为例进行超参数优化。

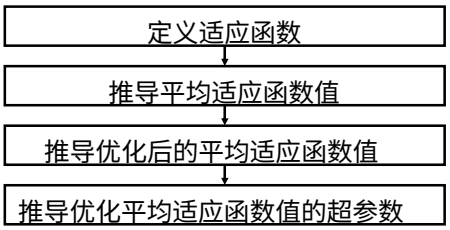

图 1 分析框架

3.2.1函数 1 定义

函数 **1**是单变量适应函数$F_1(w)$如公式(3)所示，包含单变量 w。在此例中，以$F_1(w)$值作为适应函数值，并且优化目标是最小化适应函数值(即$F_1(w)$值)。假设变量 w服从均匀分布，并且介于 0 到1之间，则可以通过公式(4)估计平均适应函数值。

$$F_1(w) = (w - 0.5)^2 \quad (3)$$

$$\int F_1(w)\, dw = \int (w - 0.5)^2\, dw \quad (4)$$

3.2.2函数 1 的超参数优化

在本节中将采用粒子群优化算法对函数**1** 进行优化，并讨论粒子群优化算法超参数的最优解。其中，变量 w的优化主要通过超参数惯性速度权重α、个体最优解权重c_1、群体最优解权重c_2来修正。因此，将修正后的值代入函数**1** 中，计算粒子群优化算法优化后的平均适应函数值，其中惯性速度为v_w、与个体最优解的距离为$o_{p,\ w}$、与群体最优解的距离为$o_{g,\ w}$、$r_{1,\ w}$和$r_{2,\ w}$为随机数，如公式(5)所示。

为优化超参数(即惯性速度权重α、个体最优解权重c_1、群体最优解权重c_2)并最小化平均适应函数值，定义函数$G_1(\alpha, c_1, c_2)$，如公式(6)所示。并对公式(6)中的α求微分，当一阶导函数为0时，求得α的值(即惯性速度权重最优解α^*)，如公式(7)所示；同理，可以推导个体最优解权重c_1和群体最优解权重，以进行修

$$\int F_1\left(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}\right)dw$$
$$=\int\left(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}-0.5\right)^2dw \tag{5}$$
$$=\frac{\left(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}-0.5\right)^3}{3}$$

$$G_1(\alpha,c_1,c_2)=\int F_1\left(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}\right)dw \tag{6}$$

$$\alpha^*=\underset{0\le\alpha\le1}{arg\,min}\,G_1(\alpha,c_1,c_2)=\underset{0\le\alpha\le1}{\arg\min}\frac{\left(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}-0.5\right)^3}{3}$$
$$\Longrightarrow\alpha^*=\frac{(w-0.5)+\left(r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}\right)}{-v_w}=\frac{\delta_1+\left(r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}\right)}{-v_w} \tag{7}$$

$${c_1}^*=\underset{0\le c_1\le1}{arg\,min}\,G_1(\alpha,c_1,c_2)=\underset{0\le c_1\le1}{\arg\min}\frac{\left(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}-0.5\right)^3}{3}$$
$$\Longrightarrow {c_1}^*=\frac{(w-0.5)+\left(\alpha v_w+r_{2,w}c_2o_{g,w}\right)}{-r_{1,w}o_{p,w}}=\frac{\delta_1+\left(\alpha v_w+r_{2,w}c_2o_{g,w}\right)}{-r_{1,w}o_{p,w}} \tag{8}$$

$${c_2}^*=\underset{0\le c_2\le1}{arg\,min}\,G_1(\alpha,c_1,c_2)=\underset{0\le c_2\le1}{\arg\min}\frac{\left(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}-0.5\right)^3}{3}$$
$$\Longrightarrow {c_2}^*=\frac{(w-0.5)+\left(\alpha v_w+r_{1,w}c_1o_{p,w}\right)}{-r_{2,w}o_{g,w}}=\frac{\delta_1+\left(\alpha v_w+r_{1,w}c_1o_{p,w}\right)}{-r_{2,w}o_{g,w}} \tag{9}$$

正c_2，如公式(8)和公式(9)所示。其中，如果将$w-0.5$定义为函数1的误差δ_1，则可以观察到，通过将超参数作为变量进行加权调整，可以使误差最小化。

3.3 适应函数2—多变量分析

为演示如何运用分析框架证明粒子群优化算法在多变量适应函数中的超参数最优解，本节以双变量适应函数为例进行超参数优化。

3.3.1函数 2 定义

函数 2是双变量适应函数$F_2(w,b)$，如公式(10)所示，包含变量 w 和变量 b。在此例中，以$F_2(w,b)$值作为适应函数值，并且优化目标是最小化适应函数值（即$F_2(w,b)$值）。假设变量 w 和变量 b 服从均匀分布，且介于 0 到 1 之间，则可以通过公式(11)估计平均适应函数值。

$$F_2(w,b)=(w+b)^2 \tag{10}$$

$$\int\int F_2(w,b)\,dw\,db=\int\int(w+b)^2\,dw\,db \tag{11}$$

3.3.2函数 2 的超参数优化

本节将采用粒子群优化算法对函数2 进行优化，并讨论粒子群优化算法超参数的最优解。其中，变量 w和变量 b 的优化主要通过超参数惯性速度权重α、个体最优解权重c_1、群体最优解权重c_2来修正。因此，将修正后的值代入函数 2中，计算经过粒子群优化算法优化后的平均适应函数值，其中，变量 w的惯性速度

为v_w、变量 b的惯性速度为v_b、变量 w 与个体最优解的距离为$o_{p,\ w}$、变量 w与群体最优解的距离为$o_{g,\ w}$、变量b 与个体最优解的距离为$o_{p,\ b}$、变量b 与群体最优解的距离为$o_{g,\ b}$、以及$r_{1,\ w}$、$r_{2,\ w}$、$r_{1,\ b}$、$r_{2,\ b}$为随机数，如公式(12)所示。

为优化超参数(即惯性速度权重α、个体最佳解权重c_1、群体最优解权重c_2)和最小化平均适应函 数值，定义函数$G_2(\alpha, c_1, c_2)$，如公式(13)所示。并 且对公式(13)的α求微分，得到一阶导函数为0时的α值(即惯性速度权重的最优解α^*)，如公式(14)所示；同 理，可以推导个体最优解权重c_1和群体最优解权重来修正c_2，如公式(15)和公式(16)所示。其中，如果

$$\begin{aligned}&\int\int F_2\big(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w},b+\alpha v_b+r_{1,b}c_1o_{p,b}+r_{2,b}c_2o_{g,b}\big)\,dw\,db\\&=\int\int\Big(\big(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}\big)+\big(b+\alpha v_b+r_{1,b}c_1o_{p,b}+r_{2,b}c_2o_{g,b}\big)\Big)^2\,dw\,db\\&=\frac{\Big(\big(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}\big)+\big(b+\alpha v_b+r_{1,b}c_1o_{p,b}+r_{2,b}c_2o_{g,b}\big)\Big)^{12}}{12}\end{aligned}\tag{12}$$

$$\begin{aligned}G_2(\alpha,c_1,c_2)=\int\int F_2\big(&w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w},b+\alpha v_b+r_{1,b}c_1o_{p,b}\\&+r_{2,b}c_2o_{g,b}\big)\,dw\,db\end{aligned}\tag{13}$$

$$\begin{aligned}\alpha^*&=\underset{0\le\alpha\le1}{arg\,min}\,G_2(\alpha,c_1,c_2)\\&=\underset{0\le\alpha\le1}{\arg\min}\frac{\Big(\big(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}\big)+\big(b+\alpha v_b+r_{1,b}c_1o_{p,b}+r_{2,b}c_2o_{g,b}\big)\Big)^{12}}{12}\\\Longrightarrow\alpha^*&=\frac{(w+b)+\big[c_1\big(r_{1,w}o_{p,w}+r_{1,b}o_{p,b}\big)+c_2\big(r_{2,w}o_{p,w}+r_{2,b}o_{g,b}\big)\big]}{-(v_w+v_b)}\\&=\frac{\delta_2+\big[c_1\big(r_{1,w}o_{p,w}+r_{1,b}o_{p,b}\big)+c_2\big(r_{2,w}o_{p,w}+r_{2,b}o_{g,b}\big)\big]}{-(v_w+v_b)}\end{aligned}\tag{14}$$

$$\begin{aligned}{c_1}^*&=\underset{0\le c_1\le1}{arg\,min}\,G_2(\alpha,c_1,c_2)\\&=\underset{0\le c_1\le1}{\arg\min}\frac{\Big(\big(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}\big)+\big(b+\alpha v_b+r_{1,b}c_1o_{p,b}+r_{2,b}c_2o_{g,b}\big)\Big)^{12}}{12}\\\Longrightarrow{c_1}^*&=\frac{(w+b)+\big[\alpha(v_w+v_b)+c_2\big(r_{2,w}o_{g,w}+r_{2,b}o_{g,b}\big)\big]}{-\big(r_{1,w}o_{p,w}+r_{1,b}o_{p,b}\big)}\\&=\frac{\delta_2+\big[\alpha(v_w+v_b)+c_2\big(r_{2,w}o_{g,w}+r_{2,b}o_{g,b}\big)\big]}{-\big(r_{1,w}o_{p,w}+r_{1,b}o_{p,b}\big)}\end{aligned}\tag{15}$$

$$\begin{aligned}{c_2}^*&=\underset{0\le c_2\le1}{arg\,min}\,G_2(\alpha,c_1,c_2)\\&=\underset{0\le c_2\le1}{\arg\min}\frac{\Big(\big(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}\big)+\big(b+\alpha v_b+r_{1,b}c_1o_{p,b}+r_{2,b}c_2o_{g,b}\big)\Big)^{12}}{12}\\\Longrightarrow{c_2}^*&=\frac{(w+b)+\big[\alpha(v_w+v_b)+c_1\big(r_{1,w}o_{p,w}+r_{1,b}o_{p,b}\big)\big]}{-\big(r_{2,w}o_{g,w}+r_{2,b}o_{g,b}\big)}\\&=\frac{\delta_2+\big[\alpha(v_w+v_b)+c_1\big(r_{1,w}o_{p,w}+r_{1,b}o_{p,b}\big)\big]}{-\big(r_{2,w}o_{g,w}+r_{2,b}o_{g,b}\big)}\end{aligned}\tag{16}$$

把$w+b$定义为函数 2 的误差δ_2，则可以观察到超参数作为变量进行加权调整，使误差最小化。

3.4 适应函数3—回归应用

为演示如何运用分析框架证明粒子群优化算法在回归及神经网络模型适应函数的超参数最优解，本节中以线性回归的适应函数为例进行超参数优化。

3.4.1函数 3 定义

函数 3是双变量的适应函数$F_3(w,b)$，如公式(17)所示，包含变量 w 和变量 b。在此例中，以$F_3(w,b)$值作为适应函数值，并且优化目标是最小化适应函数值（即$F_3(w,b)$值）。假设变量 w 和变量 b 服从均匀分布，并且介于 0 到 1 之间，则可以通过公式(18)估计平均适应函数值。

$$F_3(w,b)=(wx+b-y)^2 \tag{17}$$

$$\int\int F_3(w,b)\,dw\,db = \int\int (wx+b-y)^2\,dw\,db \tag{18}$$

3.4.2函数 3 的超参数优化

本节将采用粒子群优化算法对函数3 进行优化，并讨论粒子群优化算法超参数的最优解。其中，变量 w和变量 b 的优化主要通过超参数惯性速度权重α、个体最优解权重c_1、群体最优解权重c_2来修正。因此，将修正后的值代入函数 3中，计算粒子群优化算法优化计算后的平均适应函数值。其中，变量 w的惯性速度为v_w、变量 b的惯性速度为v_b、变量 w 与个体最优解的距离为$o_{p,w}$、变量 w与群体最优解的距离为$o_{g,w}$、变量b 与个体最优解的距离为$o_{p,b}$、变量b 与群体最优解的距离为$o_{g,b}$，以及$r_{1,w}$、$r_{2,w}$、$r_{1,b}$、$r_{2,b}$为随机数，如公式(19)所示。

为了优化超参数(即惯性速度权重α、个体最优解权重c_1、群体最优解权重c_2)并最小化平均适应函数值，定义函数$G_3(\alpha,c_1,c_2)$，如公式(20)所示。并对公式(20)中的α求微分，得到一阶导函数为0时的α值(即惯性速度权重最优解α^*)，如公式(21)所示；同理，可以推导个体最优解权重c_1和群体最优解权重，以修正c_2，如公式(22)和公式(23)所示。其中，如果将$wx+b-y$定义为函数3 的误差δ_3，则可以观察到，通过超参数作为变量进行加权调整，可使误差最小化。

3.5 小结与讨论

本文定义了3个适应函数，分别从单变量、多变量以及回归应用出发，采用最小平方误差作为目标来观察超参数最优解。通过公式推导结果，可以归纳出下面几条设置原则，未来可推广到神经网络模型优化等应用。

1. 由公式(7)、公式(8)、公式(9)、公式(14)、

$$\int\int F_3\left(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w},b+\alpha v_b+r_{1,b}c_1o_{p,b}+r_{2,b}c_2o_{g,b}\right)dw\,db$$
$$=\int\int\left(\left(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}\right)x+\left(b+\alpha v_b+r_{1,b}c_1o_{p,b}+r_{2,b}c_2o_{g,b}\right)-y\right)^2 dw\,db \tag{19}$$
$$=\frac{\left(\left(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w}\right)x+\left(b+\alpha v_b+r_{1,b}c_1o_{p,b}+r_{2,b}c_2o_{g,b}\right)-y\right)^{12}}{12}$$

$$G_3(\alpha,c_1,c_2)=\int\int F_3\left(w+\alpha v_w+r_{1,w}c_1o_{p,w}+r_{2,w}c_2o_{g,w},b+\alpha v_b+r_{1,b}c_1o_{p,b}+r_{2,b}c_2o_{g,b}\right)dw\,db \tag{20}$$

$$\alpha^* = \underset{0\le\alpha\le1}{arg\ min}\, G_3(\alpha, c_1, c_2)$$
$$= \underset{0\le\alpha\le1}{\arg\min} \frac{\left(\left(w + \alpha v_w + r_{1,w}c_1 o_{p,w} + r_{2,w}c_2 o_{g,w}\right)x + \left(b + \alpha v_b + r_{1,b}c_1 o_{p,b} + r_{2,b}c_2 o_{g,b}\right) - y\right)^{12}}{12}$$
$$\Longrightarrow \alpha^* = \frac{(wx + b - y) + \left(r_{1,w}c_1 o_{p,w}x + r_{2,w}c_2 o_{g,w}x + r_{1,b}c_1 o_{p,b} + r_{2,b}c_2 o_{g,b}\right)}{-(v_w x + v_b)} = \frac{\delta_3 + \left[c_1\left(r_{1,w}o_{p,w}x + r_{1,b}o_{p,b}\right) + c_2\left(r_{2,w}o_{g,w}x + r_{2,b}o_{g,b}\right)\right]}{-(v_w x + v_b)} \quad (21)$$

$${c_1}^* = \underset{0\le c_1\le1}{arg\ min}\, G_3(\alpha, c_1, c_2)$$
$$= \underset{0\le c_1\le1}{\arg\min} \frac{\left(\left(w + \alpha v_w + r_{1,w}c_1 o_{p,w} + r_{2,w}c_2 o_{g,w}\right)x + \left(b + \alpha v_b + r_{1,b}c_1 o_{p,b} + r_{2,b}c_2 o_{g,b}\right) - y\right)^{12}}{12}$$
$$\Longrightarrow {c_1}^* = \frac{(wx + b - y) + \left(\alpha v_w x + r_{2,w}c_2 o_{g,w}x + \alpha v_b + r_{2,b}c_2 o_{g,b}\right)}{-\left(r_{1,w}o_{p,w}x + r_{1,b}o_{p,b}\right)} = \frac{\delta_3 + \left[\alpha(v_w x + v_b) + c_2\left(r_{2,w}o_{g,w}x + r_{2,b}o_{g,b}\right)\right]}{-\left(r_{1,w}o_{p,w}x + r_{1,b}o_{p,b}\right)} \quad (22)$$

$${c_2}^* = \underset{0\le c_2\le1}{arg\ min}\, G_3(\alpha, c_1, c_2)$$
$$= \underset{0\le c_2\le1}{\arg\min} \frac{\left(\left(w + \alpha v_w + r_{1,w}c_1 o_{p,w} + r_{2,w}c_2 o_{g,w}\right)x + \left(b + \alpha v_b + r_{1,b}c_1 o_{p,b} + r_{2,b}c_2 o_{g,b}\right) - y\right)^{12}}{12}$$
$$\Longrightarrow {c_2}^* = \frac{(wx + b - y) + \left(\alpha v_w x + r_{1,w}c_1 o_{p,w}x + \alpha v_b + r_{1,b}c_1 o_{p,b}\right)}{-\left(r_{2,w}o_{g,w}x + r_{2,b}o_{g,b}\right)} = \frac{\delta_3 + \left[\alpha(v_w x + v_b) + c_2\left(r_{1,w}o_{p,w}x + r_{1,b}o_{p,b}\right)\right]}{-\left(r_{2,w}o_{g,w}x + r_{2,b}o_{g,b}\right)} \quad (23)$$

公式(15)、公式(16)、公式(21)、公式(22)、公式(23)可以观察到，当误差(即δ_1、δ_2、δ_3)越大时，超参数惯性速度权重α、个体最优解权重c_1、群体最优解权重c_2都适合采用更大的值，以快速降低误差。因此，在训练初期，建议采用较大的超参数值来降低误差。

2. 公式(7)、公式(14)、公式(21)中的分母负号主要用于修正方向，可以通过绝对值观察修正幅度。当个体最优解权重c_1、群体最优解权重c_2的系数值绝对值总和大于惯性速度权重α的系数值绝对值总和时(如：$\left|c_1\left(r_{1,w}o_{p,w}x + r_{1,b}o_{p,b}\right) + c_2\left(r_{2,w}o_{p,w}x + r_{2,b}o_{p,b}\right)\right| > \left|-(v_w x + v_b)\right|$)，则需要增大惯性速度权重$\alpha$以维持惯性速度的影响力。其中，当输入变量数(即$x$的数量)增加时，由于分子和分母中都存在输入变量，所以主要影响仍是惯性速度权重α、个体最优解权重c_1、群体最优解权重c_2的系数值，并以系数值决定最佳惯性速度权重α^*。

3. 公式(8)、公式(15)、公式(22)中的分母负号主要表示修正方向，可以通过绝对值观察修正幅度。当惯性速度权重α、群体最优解权重c_2的系数值绝对值总和大于个体最优解权重c_1的系数值绝对值总和时(如：$\left|\alpha(v_w x + v_b) + c_2\left(r_{2,w}o_{p,w}x + r_{2,b}o_{p,b}\right)\right| > \left|-\left(r_{1,w}o_{p,w}x + r_{1,b}o_{p,b}\right)\right|$)，则需要加大个体最优解权重$c_1$以维持个体最优解的影响力。其中，当输入变量数(即$x$的数量)增加时，由于分子和分母都有存

在输入变量，所以主要影响仍是惯性速度权重α、个体最优解权重c_1、群体最优解权重c_2的系数值，以系数值决定最佳个体最优解权重$c_1{}^*$。

4. 公式(9)、公式(16)、公式(23)中的分母负号主要用于修正方向，可以通过绝对值观察修正幅度。当惯性速度权重α、个体最优解权重c_1的系数值绝对值总和大于群体最优解权重c_2的系数值绝对值总和时(如：$\left|\alpha(v_w x + v_b) + c_2\left(r_{1,w} o_{p,w} x + r_{1,b} o_{p,b}\right)\right| > \left|-\left(r_{2,w} o_{g,w} x + r_{2,b} o_{g,b}\right)\right|$),则需要加大群体最优解权重$c_2$以维持群体最优解的影响力。其中，当输入变量数(即$x$的数量)增加时，由于分子和分母中都存在输入变量，因此主要影响仍是惯性速度权重α、个体最优解权重c_1、群体最优解权重c_2的系数值，并以系数值决定群体最优解权重的最佳值$c_2{}^*$。

4. 实验分析与讨论

为验证本研究提出的方法效率和优化结果，将本研究方法与采用默认超参数值所得的结果分别在不同的适应函数下比较收敛结果。其中，本节中的超参数初始值和变量设置如下：粒子数量为 2、惯性速度权重$\alpha = 0.5$、个体最优解权重$c_1 = 1$、群体最优解权重$c_2 = 1$，以及函数 3中的输入变量$x = 0.5$且输出变量$y = 0.1 \times x + 0.2$。

根据上述超参数初始值和变量设置，在不同适应函数下的收敛结果如图 2、图 3、图 4所示。从实验结果可以观察到，本研究提出的方法所得超参数最优解，即使在只有2 个粒子的情况下，对于每个适应函数都能在 20个 训练轮次 内收敛到接近零误差。如果超参数采用固定权重值，在只有 2个粒子的情况下，可能无法收敛到较低的误差。因此，与采用默认超参数或随机生成的超参数相比，本研究方法的超参数可以更快地收敛到最小误差(即最小损失)。除此之外，本研究方法未来也可以求解不同的适应函数，并且也适用于求解最大值。

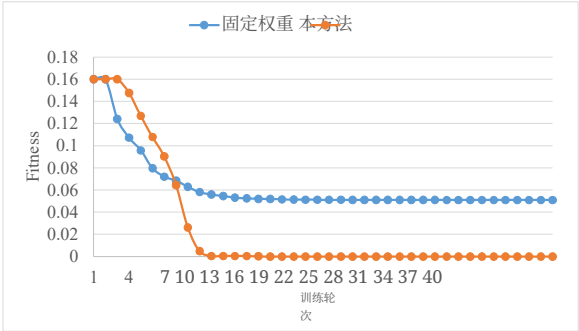

图 2 函数 1 收敛结果

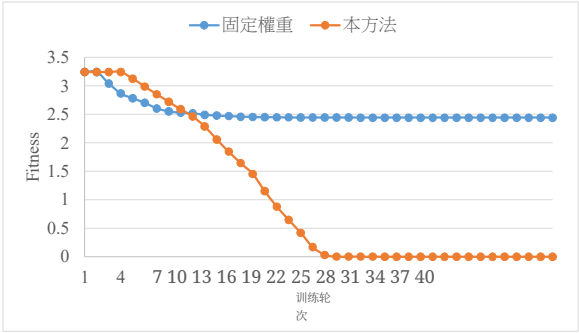

图 3 函数 2 收敛结果

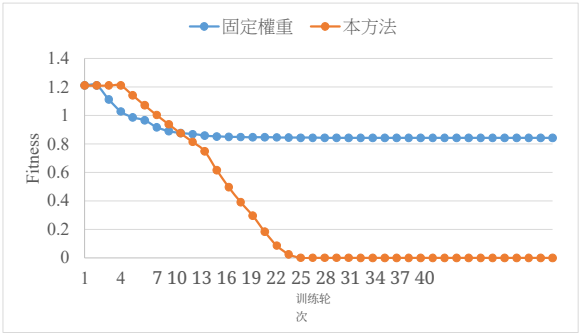

图 4 函数 3 收敛结果

5. 结论与未来研究

本研究旨在提出一个分析框架，分别分析三个适应函数的平均适应函数值，并通过最小化平均适应函数值来获得每个超参数的最优解。此外，通过实验证明，本研究提出的方法能够取得更快、更好的收敛结果。此外，在 3.5节中归纳并总结了粒子群优化算法超参数在不同适应函数下的变化和设置原则，可作为未来设置其他模型超参数时的参考。未来可以修改适应函数，并将其应用于深度学习模型优化。

参考文献

1. C. L. Camacho-Villalón、M. Dorigo、T. Stützle,

“PSO-X：用于自动设计粒子群优化算法的基于组件的框架”，*IEEE*进化计算汇刊，第26卷，第3期，第402–416页，2022年。

2. T. Wu、M. Yao、J. Yang，“海豚群算法”，《信息技术与电子工程前沿》，第17卷，717–729页，2016年。

3. A. Routray, R. K. Singh, R. Mahanty，《使用改进灰狼优化进行混合级联多电平逆变器的谐波消除》，*IEEE*工业应用汇刊，第56卷，第2期，第1827–1838页，2020年。4. M. A. Kamel, X. Yu, Y. Zhang，《基于混合遗传算法–粒子群优化算法的多台轮式移动机器人的实时容错编队控制》，*IEEE*自动化科学与工程汇刊，第18卷，第3期，第1263–1276页，2021年。5. L. Zhang等，《针对区间多目标优化问题的多种群协同人工蜂群算法》，*IEEE*模糊系统汇刊，第27卷，第5期，第1052-1065页，2019年。6. G. Dhiman, M. Garg, A. Nagar, V. Kumar, M. Dehghani，《一种用于全局优化的新算法：鼠群优化器》，《环境智能与人性化计算杂志》，第12卷，第8457–8482页，2021年。7. A. C. H. Chen，《探索各种梯度下降算法中每个超参数的优化值》，*arXiv*，2212.12279，2022年。DOI: 10.48550/arXiv.2212.12279 8. J. Kennedy, R. Eberhart，《粒子群优化》，*ICNN'95*论文集——国际神经网络会议，第4卷，第1942–1948页，澳大利亚西澳大利亚州珀斯，1995年。9. Y. Shi, R. Eberhart，《改进粒子群优化器》，*1998 IEEE*进化计算国际会议论文集。*IEEE*计算智能世界大会（目录号*98TH8360*），第69-73页，美国阿拉斯加州安克雷奇，1998年。

如何证明粒子群优化超参数的优化值?

Abel C. H. Chen 中华电信股份有限公司
chchen.scholar@gmail.com ORCID 0000-0003-3628-3033

摘要

近年来，人们提出了多种群体智能优化算法，用于解决各类优化问题。然而，仍需确定若干超参数的取值。例如，尽管粒子群优化（PSO）已应用于多种场景并取得较高的优化性能，但仍需确定惯性速度、粒子已知的最优位置以及群体已知的最优位置所对应的权重。因此，本研究提出一种分析框架，基于各类适应度函数的数学模型分析优化后的平均适应度函数值（AFFV）。此外，对于最小化问题，可以通过较低的 AFFV 确定优化后的超参数值。实验结果表明，采用所提方法得到的超参数值能够实现更高效的收敛，并获得更低的 AFFV。

关键词：粒子群优化、超参数、优化值