

基于MYA-LSTM的课堂表现预测

王 藏¹, 秦 学¹, 袁有树², 齐 睿¹

¹贵州大学大数据与信息工程学院, 贵州 贵阳

²北京中医药大学中医学院, 北京

收稿日期: 2023年2月2日; 录用日期: 2023年4月1日; 发布日期: 2023年4月7日

摘 要

本文基于长短期记忆(LSTM)模型分析学习过程中生成的行为数据并预测下一阶段的课堂行为表现,使教师能够基于此对学生采取精细的学习干预。考虑到传统的LSTM-Attention模型注意层参数计算方式优化空间不足,导致模型性能低下,本文提出基于改进的飞蛾扑火优化算法寻找注意层参数的长短期记忆(MYA-LSTM)分类预测模型,首先将注意力机制引入到LSTM网络前,其次,针对MFO算法容易陷入局部最优、收敛精度低的缺点,本文提出MYMFO算法,在种群初始化阶段加入混沌策略以及寻优后期引入柯西变异,最后利用MYMFO对注意层参数进行寻优。通过分析12个基准测试函数的仿真结果,MYMFO算法对比MFO算法的寻优精度得到了有效的提升,同时在课堂行为表现的预测实验中,MYA-LSTM对比使用未改进的MFO算法来寻找注意层参数的MA-LSTM模型在“上课积极性”、“课堂参与度”和“知识掌握度”三种课堂行为表现上的F1值分别提升了3.66、3.77和3.14,而MYA-LSTM对比LSTM-Attention在三种课堂行为上的F1值分别提升了4.53、4.46、4.56,充分证明了MYA-LSTM模型的有效性。

关键词

MFO, 注意力机制, LSTM, 课堂表现预测

Prediction of In-Class Performance Based on MYA-LSTM

Cang Wang¹, Xue Qin¹, Youshu Yuan², Rui Qi¹

¹College of Big Data and Information Engineering, Guizhou University, Guiyang Guizhou

²School of Chinese Medicine, Beijing University of Chinese Medicine, Beijing

Received: Feb. 2nd, 2023; accepted: Apr. 1st, 2023; published: Apr. 7th, 2023

Abstract

This study uses the behavior data generated during the Long Short Term Memory (LSTM) model analysis learning process to predict the classroom performance in the next stage, which will en-

able teachers to take more refined learning interventions for students. Considering that the traditional LSTM-Attention model pays attention to the insufficient optimization space of the calculation method of layer parameters, resulting in the low performance of the model. In this paper, a Long Short Term Memory (MYA-LSTM) classification model based on the improved MFO algorithm to find the parameters of the attention layer is proposed. Firstly, the attention mechanism is introduced before the LSTM network. Secondly, we propose a MYMFO algorithm that adds chaos strategy in the population initialization stage and introduces Cauchy mutation in the later stage of optimization. Finally, MYMFO is used to optimize the parameters of the attention layer. By analyzing the simulation results of 12 benchmark test functions, the optimization accuracy of MYMFO algorithm compared with MFO algorithm has been effectively improved. At the same time, in the prediction experiment of classroom behavior performance, MYA-LSTM compared with the MA-LSTM model, which uses the improved MFO algorithm to find the attention layer parameters, has increased the F1 values of “class enthusiasm”, “class participation” and “knowledge mastery” in three classroom behavior performances by 3.66, 3.77 and 3.14, respectively, compared with LSTM-Attention, MYA-LSTM increased the F1 value of three classroom behaviors by 4.53, 4.46 and 4.56 respectively, which fully proves the effectiveness of MYA-LSTM model.

Keywords

MFO, Attention Mechanism, LSTM, Prediction of In-Class Performance

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着人工智能和大数据技术的蓬勃发展，以深度学习和数据挖掘为代表的大数据技术可以发现一些数据模式，提取有价值的信息和知识，为解决各个领域的问题提供服务，这已成为当今工业界和学术界的共识。目前，大数据等技术已广泛应用于金融[1]、医疗[2] [3]、电子商务[4]、能源与制造[5]、交通[6] [7]等多个领域。在此背景下，教育数据的价值和巨大潜力引起了各国的关注，美国、印度、中国等主要国家相继出台了相关政策。目的是利用大数据技术推动教育领域深化创新、改革，探索新的教育方式，提高教学质量。

目前，许多学者基于教育数据做了大量研究，包括数据挖掘和预测。例如，Xu 等人[8]使用卡方检验对已完成复习和未完成复习的两类学生的评论进行统计分析，结果表明，完成者倾向于通过发布评论和事后出示证书来表达对课程相关内容的赞赏，而非完成者倾向于通过回复对平台建设的技术问题持否定态度。Bhalchandra 等人[9]使用主成分分析方法探索运动员的人格属性，认知能力，注意力水平，社会经济背景，地区和其他因素对运动表现的影响。另一方面，对预测的研究包括决策树[10]，SVM [11] [12] [13]，贝叶斯[11]，逻辑回归[14]和其他传统机器学习。例如，Zhang 等人[14]使用逻辑回归对 MOOC 上学生的学习行为数据进行分类，以达到学生辍学预测的目的。Chen 等[13]将学习者在 MOOC 上的学习行为数据输入到 SVM 中，以预测他们的课程成绩，作者认为 SVM 比其他传统分类算法具有更好的性能。此外，还有利用反向传播神经网络(BP)，递归神经网络(RNN)和长短期记忆(LSTM)等网络来实现预测。由于 LSTM 在处理时间序列数据方面具有较好的效果，因此被该领域的许多学者使用。例如，Tang 等人[15]从 MOOC 上学生的学习活动日志中选择 43 维行为特征，通过卷积神经网络(CNN)进行选择，最后输入

LSTM 网络, 实现辍学预测。实验结果表明, 利用 CNN 的 LSTM 网络性能有了很大的提高。Qu 等人[16] 将时间序列学习行为数据输入 LSTM 以预测课程表现。我们发现大多数研究人员都使用的 MOOC 的数据, 且由于 MOOC 面临高辍学率的问题[17], 因此他们都专注于辍学和课程表现的预测。

注意力机制起源于对人类视觉的研究[18], 然而, 它在自然语言处理(NLP)领域取得了巨大的成功[19]。如今, 它已被广泛应用于语音、图像、文本等领域, 注意力层中的参数通常通过反向传播来计算。另一方面, 智能优化算法的出现为求解复杂问题的最优值或近最优解提供了新的思路和方法。飞蛾扑火优化算法(MFO)具有收敛速度快、局部搜索能力强等优点。因此, 一些学者将其应用于分类模型[20]和神经网络[21]中的参数优化。

然而, 上述研究中仍然存在一些未解决的问题:

1) 目前关于线下课堂的研究较少, 但近年来对混合教学的研究表明, 在线学习只能达到与课堂教学相当或更差的教学效果, 而混合教学比在线教学或课堂教学具有更好的教学效果[22], 因此对课堂教学的研究也是必不可少的。

2) 教师无法根据辍学预警或课程表现预测对学习者的有效干预。

3) 注意力层参数计算方法单一且优化困难。

基于此, 本文提出 MYA-LSTM, 将注意力层引入 LSTM 网络前并利用改进后的 MFO 算法对该参数进行优化。最后, 该模型被用于预测学生的课堂行为表现。本文的贡献可以总结如下:

1) 收集线下课堂的学习行为数据, 并将其制作成数据集。首次提出使用学生的学习行为数据来预测他们下一阶段的课堂行为表现。

2) 考虑到注意力机制的重要性和缺点, 其在模型中的位置被重新考虑添加在 LSTM 网络前。

3) MYA-LSTM 旨在预测学生的课堂行为表现, 其预测结果可以为教育管理工作提供更有针对性的参考。

4) 基于 MFO 算法提出 MYMFO 算法, 增强其寻优能力, 以提高预测模型的精度。

最终实验结果表明, 使用未改进的 MFO 算法寻找注意力层参数的预测模型 MA-LSTM 在“上课积极性”、“课堂参与度”和“知识掌握度”三种课堂行为表现上的 F1 值对比传统 LSTM-Attention 分别提升了 0.87、0.69 和 1.42, 而针对 MFO 算法存在的问题进行优化, 并利用其寻找注意力层参数的预测模型 MYA-LSTM 对比 LSTM-Attention 在三种课堂行为上的 F1 值分别提升了 4.53、4.46、4.56。这证明了使用 MFO 算法寻找注意力层参数方法的有效性以及通过优化 MFO 算法提升预测模型性能的可行性。

本文的其余部分组织如下: 第 2 节给出了本文所做的相关工作, 第 3 节提供了 MYA-LSTM 的模型结构和 MFO 算法改进策略, 第 4 节讨论了 MFO 和预测模型的实验分析和预测模型, 在第 5 节给出了结论和未来工作方向。

2. 相关工作

2.1. LSTM

BP 神经网络由 Rumelhart 和 McClelland 于 1986 年提出, 具有较强的非线性映射能力和灵活的网络结构[23]。如今, BP 神经网络的研究和应用发展迅速, 并渗透到各个学科[24], 但在处理时间序列数据方面并没有太多优势。

随着 LSTM 的提出, 时间序列数据的预测得到了更好地解决。在训练过程中, LSTM 不仅对每个样本进行垂直学习和分类, 而且还学习样本本身特征之间的关系, 即特征的横向学习。目前, LSTM 和时间序列的研究和应用领域已经相当广泛, 如预测库存趋势[25], 设备故障[26], 能耗和使用[27] [28] [29]。此外, 与传统分类算法相比, LSTM 完全摒弃了复杂的特征工程, 降低了模型的复杂性。但存在一个问

题,就是对特征的关注能力不够,难以充分挖掘特征的重要性,导致模型性能不足。因此,鉴于注意力机制在 NLP、图像处理等领域的突出贡献,研究人员开始将注意力引入 LSTM 网络。例如,Shi [30]和 Chen [31]等人通过向具有注意力机制的 LSTM 提供股票数据来预测股价走势。肖[32]将 LSTM 与注意力机制相结合,提出了 LSTM-注意力温度预测模型。以上实验结果证明了注意力机制在基于 LSTM 的预测方法中的有效性。

2.2. MFO 算法

MFO 算法[33]是由 Mir Jalili *et al.*于 2015 年提出的一种智能优化算法。通过对 7 个经典工程问题的试验,结果表明,MFO 算法对求解未知搜索空间的参数问题非常有效[33]。近年来,该算法逐渐被许多学者证明在解决电力系统[34] [35],经济效益[36]和网络参数优化[20] [21]等问题方面具有良好的效果。在基本的 MFO 算法中,每只飞蛾围绕一个火焰进行搜索,并在找到更好的解时对其进行更新[33],这使得它具有局部搜索能力强的优势,但也存在后期容易进行局部优化的问题,导致其收敛精度低。

3. MYA-LSTM 模型框架

注意力机制对于 LSTM 网络的性能非常重要[30] [31],但其参数计算方法单一,改进角度不大。本文构建 MYA-LSTM 预测模型,重新考虑注意力层在模型中的位置,并采用 MYMFO 算法对其参数进行优化。通过这种方式,模型可以通过改进 MFO 算法来提高其性能。模型的整体框架如图 1 所示,包括五个功能模块:预处理层、注意力层、隐藏层、输出层和参数优化层。

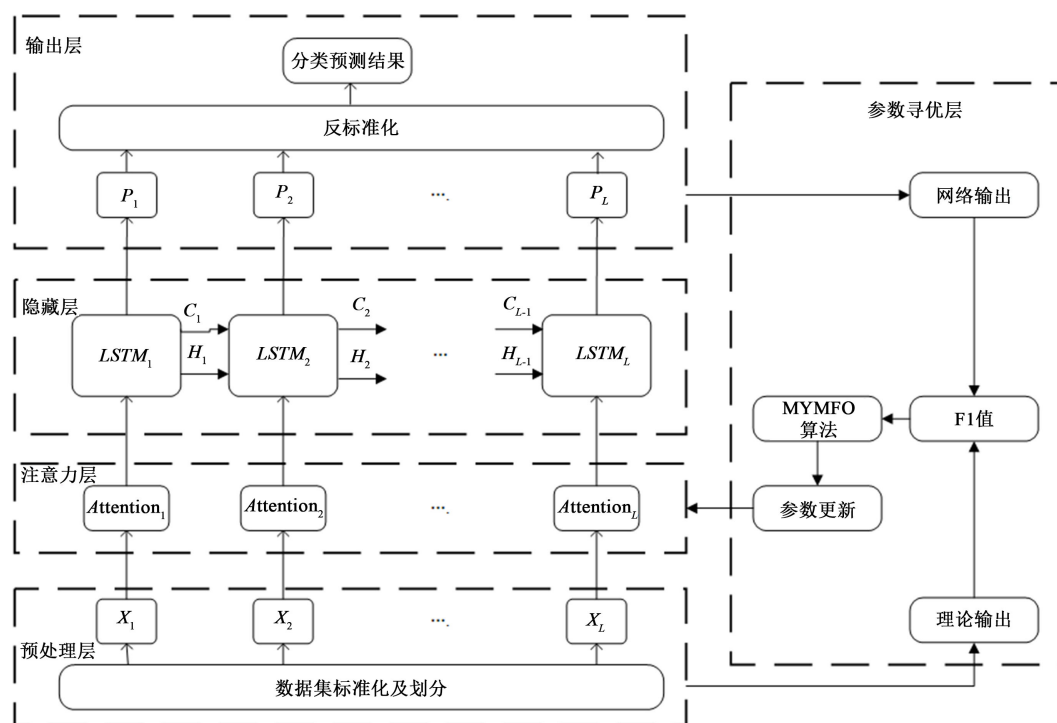


Figure 1. Model structure of MYA-LSTM

图 1. MYA-LSTM 模型框架结构图

3.1. 模块简介

预处理层首先切割和集成数据,以满足模型对输入的形状要求。然后划分训练集和测试集,最后对

数据进行标准化。数据标准化可以在一定程度上提高后续隐含层的梯度下降速度。

注意力层是该模型的核心模块。它的主要功能是给 LSTM 网络的输入时间序列特征分配不同的权重，增强模型对输入特征的注意力能力，从而提高模型的性能。

隐藏层是 LSTM 网络，它是由 LSTM 单元构建的单层循环神经网络。输出层是网络的分类预测结果。

参数优化层是为注意力层设计的，用于优化注意力层中的参数。该模块由网络输出、理论输出和 MYMFO 算法组成。为了更好地平衡模型的召回率和准确性，将网络输出和理论输出计算出的 F1 值作为 MYMFO 算法的目标函数，然后根据目标函数值迭代更新注意力层参数。

3.2. MFO 算法优化策略

3.2.1. 混沌策略

在传统 MFO 算法中，初始种群是随机生成的。然而，不理想的初始种群分布会影响算法的早期优化精度和收敛速度。与盲目的随机不同，混沌是随机的和符合工程学的，这种遍历性可以大大增强种群的多样性[37] [38]。本文采用正弦映射策略生成混沌变量，并通过混沌变量初始化飞蛾。混沌映射和初始化飞蛾种群的数学表达式分别为公式(2)和(3)：

$$z_1 = \text{random}(-1,1) \quad (1)$$

$$z_{i+1} = \sin(\pi z_i) \quad (2)$$

$$M_i = z_i * \text{upper} \quad (3)$$

其中 z_i 表示第 i 个混沌映射值，“upper”表示搜索区间的上限， M_i 表示初始化种群中的第 i 只飞蛾。在初始化种群时，通过公式(1)在 $[-1, 1]$ 区间内随机生成与候选解相同维数的第一代混沌变量，通过公式(2)生成下一代混沌变量，最后利用公式(3)的载波原理将每一代混沌变量映射到搜索区间得到初始化种群。

3.2.2. 柯西变异

扰动策略是解决局部最优的有效手段。柯西变异可以产生较大的突变步长，可用于增强全局探索能力，克服算法容易陷入局部最优和过早收敛的缺陷[39] [40]。本文引入它，使算法在后期跳出局部优化。添加柯西变异后，飞蛾通过公式(4)对其进行更新：

$$M_i = (D_i * e^{bt} * \cos 2\pi t + F_j) * (1 + \text{cauchy}(0,1)) \quad (4)$$

其中 D_i 表示第 j 火焰的第 i 个飞蛾的距离， b 是定义对数螺旋形状的常数， t 是 $[-1, 1]$ 中的随机数， $\text{cauchy}(0,1)$ 表示柯西分布。

4. 实验结果与分析

本实验分为两部分。首先，分析了改进的 MFO 算法和传统 MFO 算法在基准测试函数上的实验结果；其次，对比分析了传统分类算法 SVM、BP、LSTM、LSTM-Attention、MA-LSTM 以及 MYA-LSTM 模型在课堂表现数据集上的分类预测性能。

4.1. MFO 算法实验

4.1.1. 算法实验设计

为了探究混沌策略和柯西突变对算法性能的影响，本节设计了四个测试实验：MFO (原始 MFO)；

CHMFO (具有混沌策略的 MFO 算法); KCMFO (具有柯西突变的 MFO 算法); MYMFO (具有柯西突变的混沌策略的 MFO 算法)。在实验中, 选取了 12 个常用于算法性能比较的基准测试函数。所采用的测试函数的数学公式如表 1 所示。

考虑到预测模型实验的输入向量维数, 将多维测试函数维度设置为 60。四种算法的参数设定相同: 种群大小设置为 30, 最大迭代次数为 500, 每个算法独立运行 20 次。该算法的性能由两个指标描述: 最优解的平均值(Ave_pre)和第 100 次迭代时解的平均值(Ave_pre100)。公式如公式(5)和公式(6)所示:

$$\text{Ave_pre} = \frac{\sum_{i=1}^t |f(X_i^*) - f(X_{opt})|}{t} \quad (5)$$

$$\text{Ave_pre100} = \frac{\sum_{i=1}^t |f(X_i^{100}) - f(X_{opt})|}{t} \quad (6)$$

其中 X_i^* 表示得到的最优解, X_{opt} 是理论最优值, X_i^{100} 表示第 100 次迭代时得到的解, t 是算法的独立运行次数。解的平均值是独立运算算法得到的最优值与理论最优值之差的平均值, 平均值越小, 算法寻优性能越好。

Table 1. Information of benchmark function

表 1. 基准测试函数信息

函数表达式	定义域	维度	最优值
$f_1(x) = 0.5 + \frac{\sin^2(x_1^2 - x_2^2) - 0.5}{[1 + 0.001(x_1^2 + x_2^2)]^2}$	$[-100, 100]$	2	0
$f_2(x) = -\frac{1 + \cos(12\sqrt{x_1^2 + x_2^2})}{0.5(x_1^2 + x_2^2) + 2}$	$[-5.12, 5.12]$	2	-1
$f_3(x) = -20 \exp\left(-0.2\sqrt{\frac{1}{d}\sum_{i=1}^d x_i^2}\right) - \exp\left(\frac{1}{d}\sum_{i=1}^d \cos(2\pi x_i)\right) + 20 + \exp(1)$	$[-32, 32]$	60	0
$f_4(x) = \sum_{i=1}^d x_i^{i+1}$	$[-1, 1]$	60	0
$f_5(x) = \sum_{i=1}^d [x_i^2 - 10\cos(2\pi x_i) + 10]$	$[-5.12, 5.12]$	60	0
$f_6(x) = \sum_{i=1}^d x_i^{10}$	$[-10, 10]$	60	0
$f_7(x) = \sum_{i=1}^d i x_i^2$	$[-10, 10]$	60	0
$f_8(x) = \sum_{i=1}^d x_i $	$[-100, 100]$	60	0
$f_9(x) = 0.5 + \frac{\sin^2(\cos x_1^2 - x_2^2) - 0.5}{[1 + 0.001(x_1^2 + x_2^2)]^2}$	$[-100, 100]$	2	0
$f_{10}(x) = \sum_{i=1}^d (x_i - 1)^2$	$[-5.12, 5.12]$	60	0
$f_{11}(x) = x_1^2 + 2x_2^2 - 0.3\cos(3\pi x_1) - 0.4\cos(4\pi x_2) + 0.7$	$[-100, 100]$	2	0
$f_{12}(x) = (x_1 + 2x_2 - 7)^3 + (2x_1 + x_2 - 5)^2$	$[-10, 10]$	2	0

Table 2. Experimental results of benchmark function
表 2. 基准测试函数实验结果

算法及指标		函数			
		f_1	f_2	f_3	f_4
MFO	Ave_pre100	1.247e-03	3.869e-01	2.003e+01	5.226e-01
	Ave_pre	0.000e+00	3.869e-01	1.993e+01	4.513e-06
CHMFO	Aver_pre100	7.816e-04	3.507e-01	1.634e+01	2.438e-03
	Ave_pre	0.000e+00	0.000e+00	8.313e+00	2.024e-06
KCMFO	Ave_pre100	1.096e-03	3.761e-01	1.686e+01	6.069e-03
	Ave_pre	0.000e+00	0.000e+00	7.824e+00	3.109e-10
MYMFO	Ave_pre100	0.000e+00	6.661e-15	1.422e+01	3.484e-03
	Ave_pre	0.000e+00	0.000e+00	1.768e+00	4.038e-12

Table 3. Experimental results of benchmark function
表 3. 基准测试函数实验结果

算法及指标		函数			
		f_5	f_6	f_7	f_8
MFO	Ave_pre100	7.480e+02	1.655e+10	2.709e+04	1.856e+03
	Ave_pre	5.412e+02	2.501e+09	8.211e+03	5.324e+02
CHMFO	Aver_pre100	4.345e+02	2.770e+08	5.700e+03	8.141e+02
	Ave_pre	3.891e+02	2.571e+06	5.456e+02	2.155e+01
KCMFO	Ave_pre100	4.204e+02	3.418e+08	6.908e+03	8.269e+02
	Ave_pre	1.593e+02	2.777e+04	7.172e+01	1.294e+02
MYMFO	Ave_pre100	4.009e+02	2.354e+08	4.101e+03	4.569e+02
	Ave_pre	8.686e+01	1.374e+03	1.022e+01	7.849e+00

Table 4. Experimental results of benchmark function
表 4. 基准测试函数实验结果

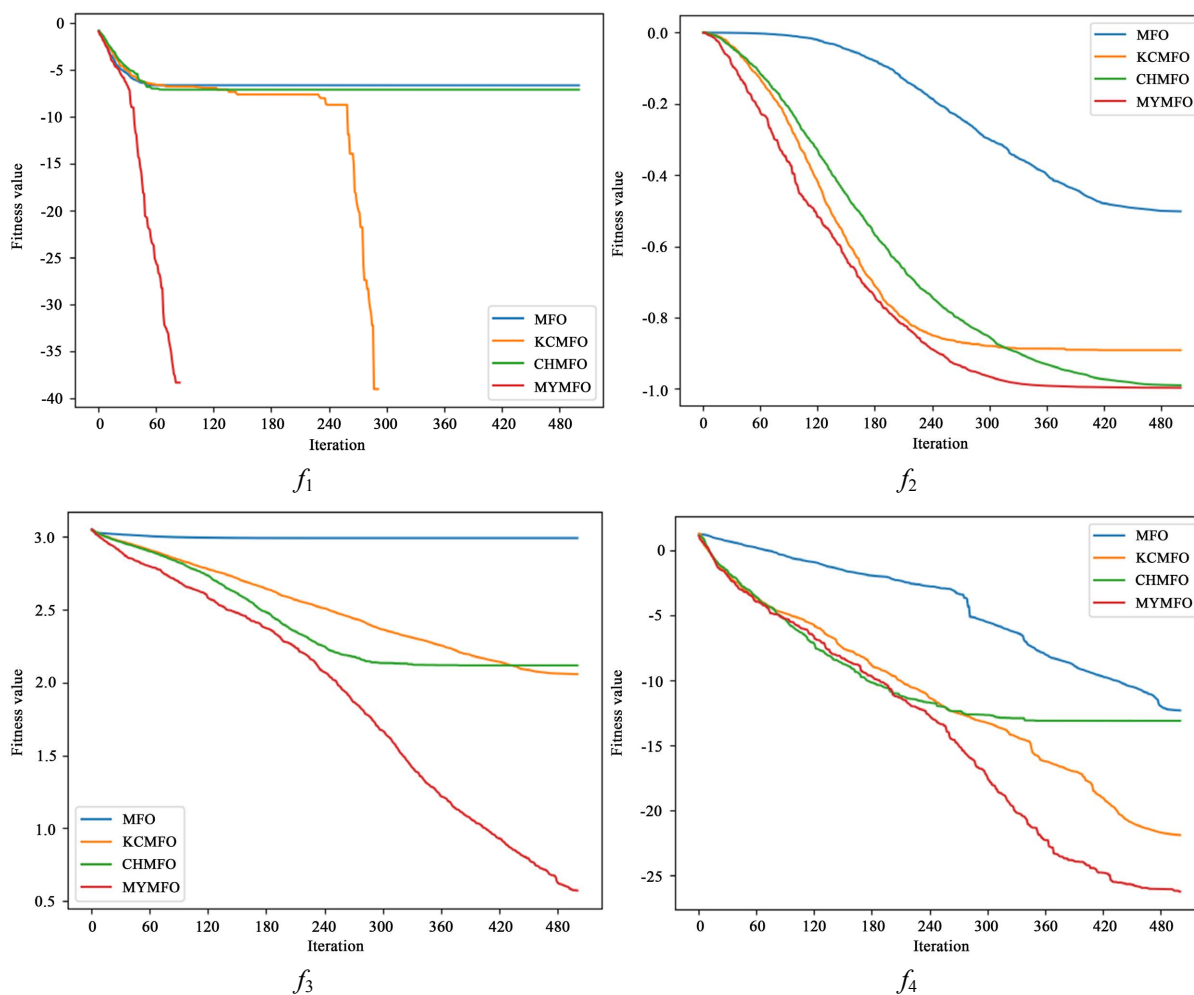
算法及指标		函数			
		F_9	f_{10}	f_{11}	f_{12}
MFO	Ave_pre100	2.360e-03	2.881e+02	5.551e-12	7.478e-19
	Ave_pre	1.622e-03	7.067e+01	0.000e+00	3.993e-18
CHMFO	Ave_pre100	1.844e-03	7.067e+01	0.000e+00	3.993e-18
	Ave_pre	1.722e-03	3.334e+01	0.000e+00	0.000e+00
KCMFO	Ave_pre100	2.267e-03	7.180e+01	0.000e+00	6.785e-19
	Ave_pre	1.567e-04	2.842e+01	0.000e+00	0.000e+00
MYMFO	Ave_pre100	1.606e-04	4.056e+01	0.000e+00	1.118e-21
	Ave_pre	1.266e-06	1.822e+01	0.000e+00	0.000e+00

4.1.2. 算法性能实验分析

如表 2~4 所示, 可以看出, MFO 和 CHMFO 直接找到了 f_1 , f_{11} 和 f_{12} 的全局最优值, 而 KCMFO 和 MYMFO 除了 f_1 , f_{11} 和 f_{12} 以外还找到了 f_2 的全局最优值。

从第 100 次迭代时解的平均值来看, 与 MFO 相比, CHMFO 在除测试函数 f_{12} 以外的其他测试函数上都具有更好的表现, 特别是在测试函数 $f_1, f_4, f_6, f_7, f_8, f_{10}$ 和 f_{11} 。这有力地表明, 使用混沌策略初始化种群是提高算法前期优化能力的可行方案。值得一提的是, 在比较 KCMFO 与 MFO 时, 我们发现 KCMFO 的精度也高于 MFO, 但略差于 CHMFO。这说明柯西变异在前期也能提高算法的优化能力, 但其效果不如混沌策略。对于 MYMFO, 无论是与 MFO, CHMFO 还是 KCMFO 相比, 其性能都得到了较大的提升, 这在测试函数中 $f_1, f_2, f_7, f_8, f_9, f_{10}, f_{12}$ 中很明显, 我们认为这是混沌策略与柯西突变相互作用的结果。

从最优解的平均值来看, CHAO、KCMFO 和 MYMFO 的性能明显优于 MFO。结合以上对第 100 次迭代时解的平均值的分析, 我们可以认为, 无论是混沌策略还是柯西突变, 它们对算法优化能力的提升贯穿于整个搜索过程, 不局限于早期或后期。然而, 我们可以看到 KCMFO 的性能明显高于 CHMFO, 因此, 我们可以得出结论, 尽管混沌策略和柯西突变在整个搜索过程中增强了算法的优化能力, 但它们的主要作用阶段是不同的。混沌策略往往更早, 而柯西突变往往更晚。为了更直观地展示结果, 本文给出了一些函数收敛曲线(图 2), 其中包括两个二维函数和四个高维函数。为了便于读者区分收敛曲线, 本文对一些结果进行了对数转换($f_1, f_3, f_4, f_5, f_6, f_7, f_8$)。可见, 混沌策略可以增强前期优化能力, 但也面临后期局部优化的困境, 柯西突变对后期优化能力的提升是显而易见的, 同时, 本文针对 MFO 算法提出的两个改进策略也是十分有效的。



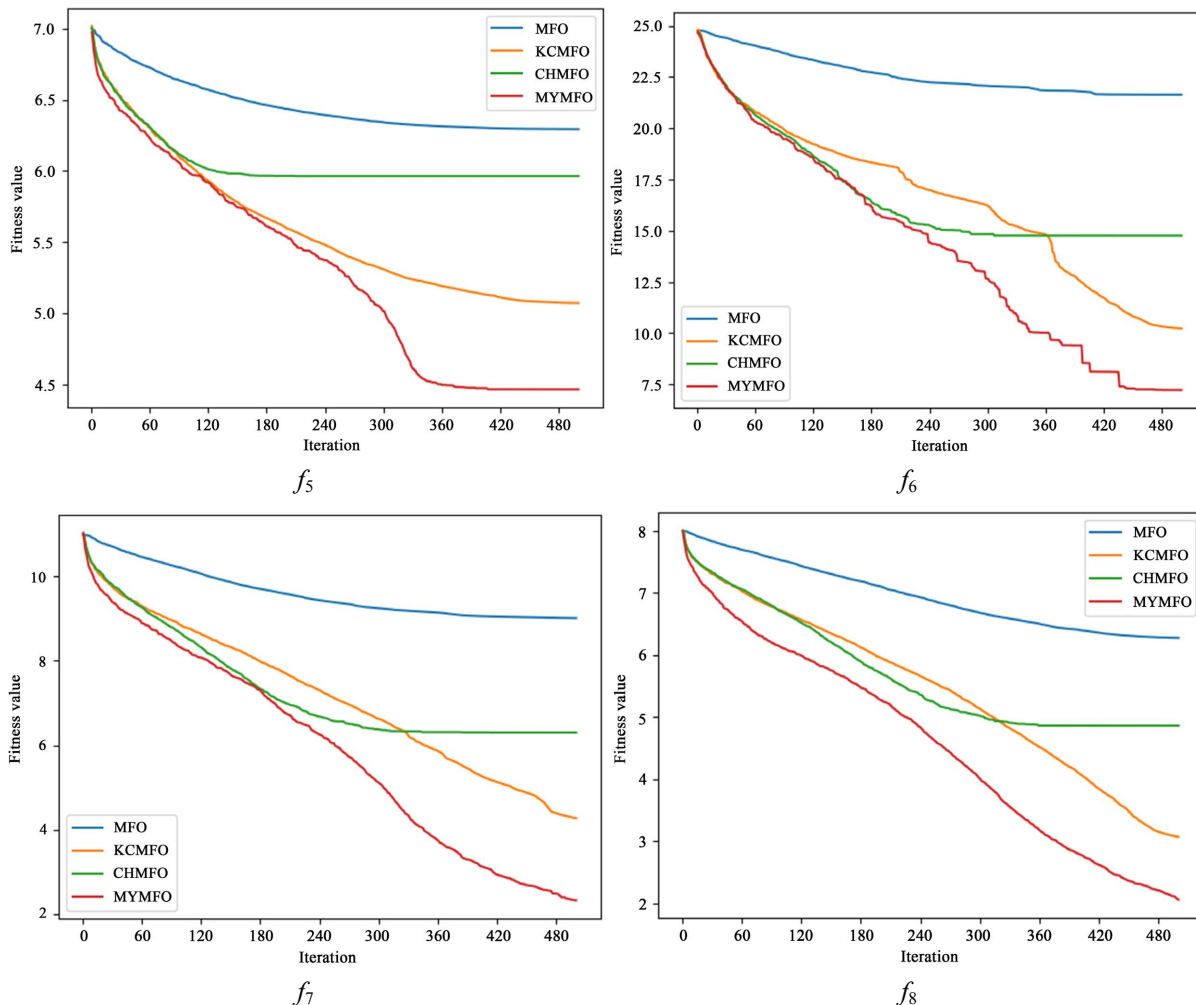


Figure 2. Convergence curve of benchmark function

图 2. 测试函数收敛曲线图

4.2. 行为表现预测实验

4.2.1. 数据描述与处理

这部分实验收集了某大学某专业 2018~2019 年 7 门本科课程的学习行为数据。这些数据通过智能教学管理平台采集，覆盖 460 名学生，包括在线预习和传统课堂行为。通过对源数据的预处理，在预习阶段得到 7 个行为特征，在传统课堂阶段得到 6 个行为特征。具体说明见表 5 和表 6。

Table 5. Behavior characteristics and instructions for preview

表 5. 预习阶段的行为特征和说明

行为特征名	描述
章节	当前预习章节序号
滞留时间	预习任务发布到打开的时间间隔(以分钟为单位)
预习时长	预习持续时间(以分钟为单位)
课件完成率	课件 ppt 查阅的页数与总页数的比率，值范围[0, 1]

Continued

答题率	课件中习题的作答率, 值范围[0, 1]
答题正确率	作答正确率, 取值范围[0, 1]
得分率	所得分数与课件总分的比率, 值范围[0, 1]

Table 6. Behavior characteristics and instructions in the classroom

表 6. 课堂阶段的行为特征和说明

行为特征名	描述
章节	当前课堂的章节序号
签到间隔	发出签到二维码与学生扫描二维码之间的时间间隔 (分钟为单位)
课堂互动得分	在课堂上积极回答问题或参与讨论的次数
作答率	课堂随机测验中的答题率, 取值范围[0, 1]
答题正确率	作答的正确率, 取值范围为[0, 1]
得分率	课堂答题得分率和题目总分的比值, 取值范围[0, 1]

这部分实验使用前四章的预览行为数据和前三章的课堂行为数据来预测第四章的课堂行为表现。为了使教学管理者更容易理解模型输出, 从而做出相应的学习干预。本文重新组合了被预测章节的课堂行为, 然后根据实际情况和聚类算法对其进行聚类。其中, 将“课堂互动得分”与“作答率”相结合得到“课堂参与度”, “答题正确率”和“得分率”结合得到“知识掌握度”, “签到间隔”更名为“上课积极度”。有关组合行为名称及其类别, 请参阅表 7。通过上述数据处理和整合方法, 本文得到了课堂表现数据集, 包括 2244 个有效数据, 并且将预测问题转化为分类问题。

Table 7. Predicted chapter classroom behavior performance and its description

表 7. 被预测章节课堂行为表现及其描述

原始课堂表现变量	组合后行为名称	描述	类别	样本数
签到间隔	上课积极度	学生对这门课的热情程度	积极	279
			一般	1624
			不积极	341
课堂互动得分	课堂参与度	学生在课堂上的主动性	积极	1322
			一般	427
			不积极	495
答题正确率 得分率	知识掌握度	学生对本章知识点的掌握情况	好	1099
			不好	1148

4.2.2. 实验参数设定

课堂表现预测实验涉及三种实验: 传统分类算法 SVM、BP 神经网络和基于 LSTM 的实验。SVM 使用多项式核函数和一对一来构造多类分类器。BP 神经网络的参数: 迭代次数为 600, 隐藏层中包含的神经元个数设置为 64, 学习率设置为 0.001。LSTM 网络有五个基本参数, 包括时间步长、隐藏层中的单元个数、dropout、批大小和迭代次数。其中, 时间步长为 4, 经过多次实验, 当其他参数分别设置为 64、

0.001、16 和 300 时，实验性能最佳。MFO 算法部分使用 10 只飞蛾和 20 次迭代。最后，为了增强模型的鲁棒性，本节实验全部采用 5 倍交叉验证法。

4.2.3. 实验结果及分析

为了验证所提出的 MYA-LSTM 课堂表现预测模型的有效性，本文将前四章的预习行为变量和前三章的课堂行为变量作为模型输入。模型输出表示第四章中某种课堂行为的表现。选取 80% 的数据作为训练集，20% 的数据作为测试数据集。表 8 显示了三种课堂行为在不同模型上的 F1 值，其中 LSTM-Attention 代表利用梯度下降求解注意层参数的方法，MA-LSTM 代表参数寻优层使用 MFO 算法查找参数，MYA-LSTM 表示使用的是 MYMFO 算法。

Table 8. Performance of different models
表 8. 不同模型的性能表现

组别	模型	课堂行为	F1 值
1	SVM	上课积极性	43.77
		课堂参与度	58.59
		知识掌握度	60.49
2	BP	上课积极性	54.60
		课堂参与度	65.56
		知识掌握度	70.10
3	LSTM	上课积极性	63.92
		课堂参与度	70.57
		知识掌握度	74.72
4	LSTM-Attention	上课积极性	65.55
		课堂参与度	71.98
		知识掌握度	75.60
5	MA-LSTM	上课积极性	66.42
		课堂参与度	72.67
		知识掌握度	77.02
6	MYA-LSTM	上课积极性	70.08
		课堂参与度	76.44
		知识掌握度	80.16

从表 8 可以看出，MYA-LSTM 模型性能最好。比较第 1、2 和 3 组的实验结果，BP 在“上课积极性”、“课堂参与度”和“知识掌握度”上的 F1 值分别比 SVM 高 10.83、6.79 和 9.61，而 LSTM 对比 BP 在三种课堂行为表现上分别高出了 9.32、5.01 和 4.62。LSTM 的性能明显优于其他两组，另一方面，BP 的性能优于 SVM。这证明，在处理复杂数据集时，神经网络比传统分类算法具有更强的函数拟合能力，性能更好。同时，LSTM 在训练时间序列数据方面比普通神经网络具有优势。

从第 3、4、5 和 6 组的实验结果来看，LSTM-Attention 在“上课积极性”、“课堂参与度”和“知识掌握度”上分别比 LSTM 高 1.63、1.41 和 0.88，MA-LSTM 对比 LSTM-Attention 分别提升了 0.87、0.69 和 1.42，而 MYA-LSTM 对比 LSTM-Attention 在三种课堂行为上的 F1 值分别提升了 4.53、4.46、4.56。

这说明注意力机制能够提升 LSTM 模型精度，此外，使用 MFO 算法寻找注意力层的参数比传统方式更有优势，因为通过优化 MFO 算法提升预测模型性能具有一定的可行性。

比较第 5 组和第 6 组的实验结果，我们发现“上课积极度”、“课堂参与度”和“知识掌握度”的 F1 值分别提高了 3.66、3.77 和 3.14。图 3 显示了 MA-LSTM 和 MYA-LSTM 的 F1 值提升过程，其中“--”和直线分别表示使用原始 MFO 和 MYMFO 寻找注意力层参数。可以看出，MYA-LSTM 的 F1 值在前期和后期均高于 MA-LSTM，充分说明改进算法能够提高注意力层参数的精度，同时也提升了模型的训练效率。

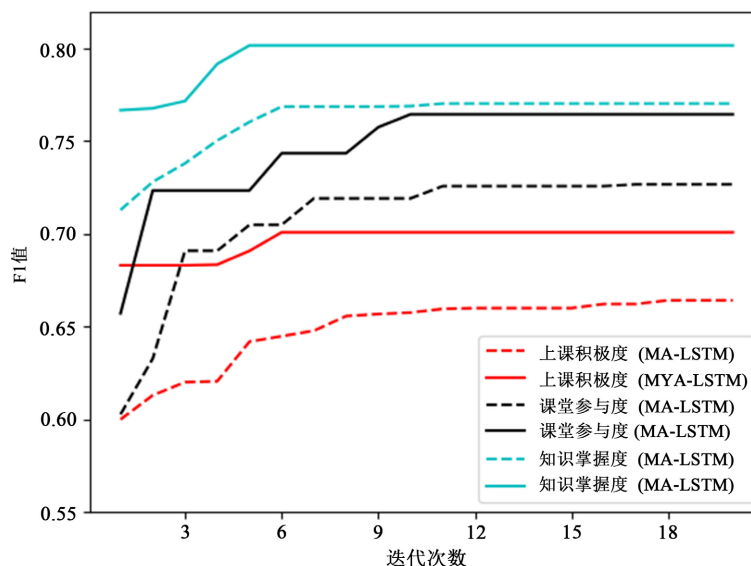


Figure 3. F1 value lift curve
图 3. F1 值提升曲线图

从所有的模型结果来看，有一个问题值得注意。“课堂参与度”的 F1 值始终高于“上课积极度”，而“知识掌握度”的 F1 值最高。通过对数据集的分析，我们认为这是类别分布不均匀造成的，此外，“知识掌握度”是二元分类问题也是一个重要原因。

5. 结论与展望

本文利用学生的学习行为数据预测学生的课堂表现，在模型实现过程中，针对模型表现较弱的问题提出解决方案。首先，在 LSTM 网络前面添加注意力层，然后利用 MFO 算法对注意力层的参数进行优化，以提高模型性能。在此基础上，在 MFO 算法中引入混沌和柯西突变两种优化策略，以提高其优化能力。最后，通过算法在 12 个基准测试函数上的仿真实验和模型在课堂表现数据集上的实验结果分析，可以看出两种优化策略都有显著的效果，此外，与传统的 LSTM 模型相比，带有注意力层的 LSTM 网络在课堂表现预测中有效提高了其 F1 分数，与原始 MFO 相比，改进后的 MFO 也提高了其寻找注意力层参数的效率和精度。

然而，我们发现除了“知识掌握度”的 F1 值达到 80.16%，“上课积极度”和“课堂参与度”的 F1 值仍然远非理想，这使得我们无法完全依靠模型输出的结果来为学生采取精细的学习干预。我们分析了原因。首先，多分类问题比二元分类问题更复杂，我们还没有针对多分类问题提出改进建议。其次，类别不平衡也制约了模型的性能。这将是我们的下一个研究内容。

参考文献

- [1] Yang, J., Zhao, C.C. and Xing, C.X. (2019) Big Data Market Optimization Pricing Model Based on Data Quality. *Complexity*, **2019**, Article ID: 5964068. <https://doi.org/10.1155/2019/5964068>
- [2] Stufi, M., Bacic, B. and Stoimenov, L. (2020) Big Data Analytics and Processing Platform in Czech Republic Healthcare. *Applied Sciences-Basel*, **10**, 1705. <https://doi.org/10.3390/app10051705>
- [3] Niu, B., et al. (2017) Analysis and Modeling for Big Data in Cancer Research. *Biomed Research International*, **2017**, Article ID: 4649191. <https://doi.org/10.1155/2017/1972097>
- [4] Li, B. and Lei, Q. (2022) Hybrid IoT and Data Fusion Model for e-Commerce Big Data Analysis. *Wireless Communications & Mobile Computing*, **2022**, Article ID: 2292321. <https://doi.org/10.1155/2022/2292321>
- [5] Zhang, F.C. and Lin, H. (2022) Energy Economy and Energy Structure Optimization Countermeasures Based on Big Data Mining. *Wireless Communications & Mobile Computing*, **2022**, Article ID: 4711551. <https://doi.org/10.1155/2022/4711551>
- [6] Zhao, L. and Jia, Y.H. (2021) Cluster Coordination between High-Speed Rail Transportation Hub Construction and Regional Economy Based on Big Data. *Complexity*, **2021**, Article ID: 6610882. <https://doi.org/10.1155/2021/6610882>
- [7] Liu, S.Y. (2022) Application of Big Data Technology in Urban Greenway Design. *Security and Communication Networks*, **2022**, Article ID: 4826523. <https://doi.org/10.1155/2022/4826523>
- [8] Peng, X. and Xu, Q.M. (2020) Investigating Learners' Behaviors and Discourse Content in MOOC Course Reviews. *Computers & Education*, **143**, Article ID: 103673. <https://doi.org/10.1016/j.compedu.2019.103673>
- [9] Bhalchandra, P., et al. (2017) Discovering Significant Performing Variables of Athlete Students through Data Mining Techniques. *International Conference on Computing, Analytics and Networking (ICCAN)*, Bhubaneswar, 3-5 July 2017, 613-623. https://doi.org/10.1007/978-981-10-7871-2_59
- [10] Hooshyar, D. and Yang, Y. (2021) Predicting Course Grade through Comprehensive Modelling of Students' Learning Behavioral Pattern. *Complexity*, **2021**, Article ID: 7463631. <https://doi.org/10.1155/2021/7463631>
- [11] Mengash, H.A. (2020) Using Data Mining Techniques to Predict Student Performance to Support Decision Making in University Admission Systems. *IEEE Access*, **8**, 55462-55470. <https://doi.org/10.1109/ACCESS.2020.2981905>
- [12] Yang, Y.W., et al. (2020) Predicting Course Achievement of University Students Based on Their Procrastination behaviour on Moodle. *Soft Computing*, **24**, 18777-18793. <https://doi.org/10.1007/s00500-020-05110-4>
- [13] Chen, J.M., et al. (2018) Course Performance Prediction for Basic Courses of Universities Based on Support Vector Machine. *Journal of Physics: Conference Series*, **1168**, Article ID: 032066.
- [14] Zhang, Y., Zhang, Q. and Liu, X. (2020) Dropout Predictions of Ideological and Political MOOC Learners Based on Big Data. *19th Annual Wuhan International Conference on E-Business (WHICEB)*, Wuhan, 29-31 May 2020, 27-32.
- [15] Tang, X.Q., et al. (2022) Dropout Rate Prediction of Massive Open Online Courses Based on Convolutional Neural Networks and Long Short-Term Memory Network. *Mobile Information Systems*, **2022**, Article ID: 8255965. <https://doi.org/10.1155/2022/8255965>
- [16] Qu, S.J., et al. (2019) Predicting Student Achievement Based on Temporal Learning Behavior in MOOCs. *Applied Sciences-Basel*, **9**, 5539. <https://doi.org/10.3390/app9245539>
- [17] Khalil, H. and Ebner, M. (2014) MOOCs Completion Rates and Possible Methods to Improve Retention—A Literature Review. *World Conference on Educational Multimedia, Hypermedia and Telecommunications (ED-Media)*, Tampere, 23-27 June 2014, 1305-1313.
- [18] Lin, L., et al. (2019) Recurrent Models of Visual Co-Attention for Person Re-Identification. *IEEE Access*, **7**, 8865-8875. <https://doi.org/10.1109/ACCESS.2018.2890394>
- [19] Bahdanau, D., Cho, K. and Bengio, Y.J.C.S. (2014) Neural Machine Translation by Jointly Learning to Align and Translate.
- [20] Kalita, D.J., Singh, V.P. and Kumar, V. (2021) A Dynamic Framework for Tuning SVM Hyper Parameters Based on Moth-Flame Optimization and Knowledge-Based-Search. *Expert Systems with Applications*, **168**, Article ID: 114139. <https://doi.org/10.1016/j.eswa.2020.114139>
- [21] Ouyang, W.M. (2021) Classroom Education Effect Evaluation Model Based on MFO Intelligent Optimization Algorithm. *Journal of Intelligent & Fuzzy Systems*, **40**, 6791-6802. <https://doi.org/10.3233/JIFS-189512>
- [22] Wu, G.F., Zheng, J. and Zhai, J. (2021) Individualized Learning Evaluation Model Based on Hybrid Teaching. *International Journal of Electrical Engineering Education*. <https://doi.org/10.1177/0020720920983999>
- [23] Hodo, E., et al. (2016) Threat Analysis of IoT Networks Using Artificial Neural Network Intrusion Detection System. *2016 International Symposium on Networks, Computers and Communications (ISNCC)*, Yasmine Hammamet, 11-13 May 2016, 1-6. <https://doi.org/10.1109/ISNCC.2016.7746067>

- [24] Yang, A.M., *et al.* (2019) Design of Intrusion Detection System for Internet of Things Based on Improved BP Neural Network. *IEEE Access*, **7**, 106043-106052. <https://doi.org/10.1109/ACCESS.2019.2929919>
- [25] Song, D., *et al.* (2020) Forecasting Stock Market Index Based on Pattern-Driven Long Short-Term Memory. *Economic Computation and Economic Cybernetics Studies and Research*, **54**, 25-41. <https://doi.org/10.24818/18423264/54.3.20.02>
- [26] Ma, M., *et al.* (2022) Predicting Machine's Performance Record Using the Stacked Long Short-Term Memory (LSTM) Neural Networks. *Journal of Applied Clinical Medical Physics*, **23**, e13558. <https://doi.org/10.1002/acm2.13558>
- [27] Chen, Y. (2020) Voltages Prediction Algorithm Based on LSTM Recurrent Neural Network. *Optik*, **220**, Article ID: 164869. <https://doi.org/10.1016/j.ijleo.2020.164869>
- [28] Kim, T.Y. and Cho, S.B. (2019) Predicting Residential Energy Consumption Using CNN-LSTM Neural Networks. *Energy*, **182**, 72-81. <https://doi.org/10.1016/j.energy.2019.05.230>
- [29] Mu, L., *et al.* (2020) Hourly and Daily Urban Water Demand Predictions Using a Long Short-Term Memory Based Model. *Journal of Water Resources Planning and Management*, **146**, Article ID: 05020017. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001276](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001276)
- [30] Shi, Y., *et al.* (2021) Improved ACD-Based Financial Trade Durations Prediction Leveraging LSTM Networks and Attention Mechanism. *Mathematical Problems in Engineering*, **2021**, Article ID: 7854512. <https://doi.org/10.1155/2021/7854512>
- [31] Chen, S. and Ge, L. (2019) Exploring the Attention Mechanism in LSTM-Based Hong Kong Stock Price Movement Prediction. *Quantitative Finance*, **19**, 1507-1515. <https://doi.org/10.1080/14697688.2019.1622287>
- [32] Xiao, K., *et al.* (2021) Attention-Based Long Short-Term Memory Network Temperature Prediction Model. 2021 *7th International Conference on Condition Monitoring of Machinery in Non-Stationary Operations (CMMNO)*, Guangzhou, 20-22 May 2021, 278-281.
- [33] Mirjalili, S. (2015) Moth-Flame Optimization Algorithm: A Novel Nature-Inspired Heuristic Paradigm. *Knowledge-Based Systems*, **89**, 228-249. <https://doi.org/10.1016/j.knosys.2015.07.006>
- [34] Abdel-Mawgoud, H., *et al.* (2021) Optimal Allocation of DG and Capacitor in Distribution Networks Using a Novel Hybrid MFO-SCA Method. *Electric Power Components and Systems*, **49**, 259-275. <https://doi.org/10.1080/15325008.2021.1943066>
- [35] Trivedi, I.N., *et al.* (2018) Optimal Power Flow with Voltage Stability Improvement and Loss Reduction in Power System Using Moth-Flame Optimizer. *Neural Computing & Applications*, **30**, 1889-1904. <https://doi.org/10.1007/s00521-016-2794-6>
- [36] Jangir, P., *et al.* (2016) Economic Load Dispatch Problem with Ramp Rate Limits and Prohibited Operating Zones Solve Using Levy Flight Moth-Flame Optimizer. 2016 *International Conference on Energy Efficient Technologies for Sustainability (ICEETS)*, Nagercoil, 7-8 April 2016, 442-447.
- [37] Yang, L., *et al.* (2018) An Improved Chaotic ACO Clustering Algorithm. 20th *IEEE International Conference on High Performance Computing and Communications (HPCC)*/16th *IEEE International Conference on Smart City (Smart-City)*/4th *IEEE International Conference on Data Science and Systems (DSS)*, Exeter, 28-30 June 2018, 1642-1649. <https://doi.org/10.1109/HPCC/SmartCity/DSS.2018.00267>
- [38] Feng, J.H., *et al.* (2017) A Novel Chaos Optimization Algorithm. *Multimedia Tools and Applications*, **76**, 17405-17436. <https://doi.org/10.1007/s11042-016-3907-z>
- [39] Chen, H.W., *et al.* (2021) Application of Distributed Seagull Optimization Improved Algorithm in Sentiment Tendency Prediction. 15th *International Conference on Complex, Intelligent and Software Intensive Systems (CISIS)*, Asan, 1-3 July 2021, 90-99. https://doi.org/10.1007/978-3-030-79725-6_9
- [40] Jiang, K.Q., Jiang, M.Y. and IEEE (2021) Lion Swarm Optimization Based on Balanced Local and Global Search with Different Distributions. *IEEE International Conference on Progress in Informatics and Computing (IEEE PIC)*, Shanghai, 17-19 December 2021, 276-280. <https://doi.org/10.1109/PIC53636.2021.9687052>