

基于深度学习的智能学习资源推荐算法

宋菲菲¹, 隋 栋², 周湘贞³

(1. 内蒙古电子信息职业技术学院 计算机与网络安全学院, 内蒙古 呼和浩特 010011;
2. 北京建筑大学 电气与信息工程学院, 北京 102406; 3. 北京航空航天大学 计算机学院, 北京 100191)

摘 要: 为了提高资源推荐性能, 采用广义回归神经网络完成资源推荐。首先, 提取推荐系统的用户和资源特征, 选择两者的特征差异值之和作为推荐系统目标函数, 然后构建广义回归神经网络 (Generalized regression neural network, GRNN) 资源推荐模型。考虑到 GRNN 训练效果对平滑因子和核函数中心的依赖性强的特点, 引入差分进化 (Differential evolution, DE) 算法对 GRNN 的平滑因子和核函数中心偏移因子进行优化求解; 选择最小特征差异值求解函数作为 DE 算法适应度函数, 通过 DE 算法的多次交叉、变异和选择操作, 获得最优平滑因子和偏移因子。最后采用优化后的平滑因子和偏移因子进行 GRNN 资源推荐, 生成特征差异较小的候选资源序列作为资源推荐序列。试验证明, 选择合理的 DE 算法交叉速率和差分缩放因子, 能够获得较好的平滑因子和偏移因子, GRNN 也能够获得更好的推荐效果。和常用资源推荐算法比较, 对于 3 种不同的训练样本, 该文算法能够获得更优的资源推荐准确率, 且 RMSE 值较低。

关键词: 深度学习; 智能学习; 在线环境; 资源推荐; 广义回归神经网络; 差分进化

中图分类号: TP391 **文章编号:** 1005-9830(2022)02-0185-07

DOI: 10.14177/j.cnki.32-1397n.2022.46.02.009

Intelligence learning resource recommendation algorithm based on deep learning

Song Feifei¹, Sui Dong², Zhou Xiangzhen³

(1. Computer and Network Security College, Inner Mongolia Electronic Information Vocational Technical College, Hohhot 010011, China; 2. School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 102406, China;
3. School of Computer Science and Engineering, Beihang University, Beijing 100191, China)

Abstract: In order to improve the performance of resource recommendation, generalized regression neural network is used to complete resource recommendation. Firstly, the features of users and resources in the recommendation system are extracted, and the sum of their feature differences is

收稿日期: 2021-09-17 修回日期: 2021-11-09

基金项目: 内蒙古自治区教育科学研究“十三五”规划课题 (NZJJGH2019119); 国家自然科学基金青年基金 (61702026)

作者简介: 宋菲菲 (1983-), 女, 讲师, 主要研究方向: 计算机应用技术, E-mail: 9187054@qq.com。

引文格式: 宋菲菲, 隋栋, 周湘贞. 基于深度学习的智能学习资源推荐算法 [J]. 南京理工大学学报, 2022, 46(2): 185-191.

投稿网址: <http://zrxuebao.njust.edu.cn>

selected as the objective function of the recommendation system. Then, a generalized regression neural network (GRNN) resource recommendation model is constructed. Considering the strong dependence of GRNN training effect on smoothing factor and kernel function center, differential evolution (DE) algorithm is introduced to optimize the smoothing factor and kernel function center offset factor of GRNN. The minimum feature difference value solving function is selected as the fitness function of DE algorithm, and the optimal smoothing factor and offset factor are obtained through multiple crossover, mutation and selection operations of DE algorithm. Finally, the optimized smoothing factor and offset factor are used for GRNN resource recommendation, and the candidate resource sequence with small feature difference is generated as the resource recommendation sequence. Experiments show that a reasonable cross rate and differential scaling factor of DE algorithm can obtain better smoothing factor and offset factor, and GRNN can obtain better recommendation effect. Compared with common resource recommendation algorithms, it can obtain better resource recommendation accuracy for three different training samples, and the RMSE value is lower.

Key words: deep learning; intelligence learning; online environment; resource recommendation; generalized regression neural network; differential evolution

互联网数据日益增长,学习用户从网络中获取有效学习资源的难度提升。从用户端来讲,用户需要花费更多的时间来完成有价值数据的全网检索^[1]。因此,主流方法是通过服务端的记录、统计和计算来实现用户的个性化资源推荐,以解决用户从海量数据快速获取价值数据的问题。当前,基于云端的在线学习得到了学习用户的青睐,但是在线环境资源量大、资源形式多样化以及资源可用平台限制等原因造成资源推送难度提升。为了实现资源精准推荐,需要细粒度地分析学习用户和所供资源的特征属性,尽量去寻找两者属性的最小差异,为用户匹配特征差异最小的资源。

当前,关于智能推荐的研究较多,温占考等^[2]采用 Hadoop 平台对资源进行管理和推荐,借助 Hadoop 来完成海量资源管理与推送。冯兴杰等^[3]将 Spark 平台用于大量资源推荐,以此来提高资源推荐的效率。前两者均是基于云计算环境海量资源的推荐研究,更侧重于云计算的数据推送平台建设,对微观的资源细节及方法未深入展开。莫恭钊等^[4]和崔建双等^[5]分别采用协同过滤算法和多分类支持向量机算法来构建智能推荐系统,并进行了具体实证分析,推荐准确度得到提升,但是仍无法有效体现微观的资源细节。

近期,深度学习技术在人工智能领域表现十分突出,樊海玮等^[6]提出基于深度神经网络的学习资源推荐算法,对资源细节的多特征属性具有较好的识别和分类效果。因此,本文借助于深度学习算法,运用深度学习中的广义回归神经网络 (Generalized regression neural network, GRNN) 算

法对在线教学环境的资源进行推荐。为了提高 GRNN 算法的资源推荐适应度,对 GRNN 算法进一步优化,从而更进一步提高资源推荐的准确度。

1 资源推荐

学习资源推荐问题模型包含学习者和学习资源两个实体模型。学习资源推荐实际上是根据学习者和学习资源的多个维度之间的特征,发现两者之间特征差异最小值,进行精准匹配,推荐个性化学习资源序列的过程。考虑到在线环境下学习资源自身属性和平台特点,需要选择在线环境下用户和资源的最关键特征。关于用户和资源特征主要如表 1 所示^[7]。

表 1 用户及资源特征描述

特征		描述
用户特征	学习能力	按年级和学历分级
	学习内容	学科所属类别
	学习媒介	视音频、文字、图片等
	学习的热点	知识点及关键词
	学习时间	学习的时间段
	学习设备终端	手机、电脑及对应的操作系统
资源特征	资源难度等级	资源难度定级
	资源内容类别	资源所属学科类别
	资源可提供媒介	资源的提供形式
	资源热度分类	平台按学习次数的资源热度分级
	资源开放时间	资源可用时间
资源使用平台支持		支持的操作系统平台

学习的热点可以根据学习-资源次数矩阵获得,根据表 1 用户及资源特征,对特征进行数字编码,然后进行特征对比求差异值,可以获得 6 个对应的特征差异函数 $S_i(i=1,2,\dots,6)$,那么推荐模型目标函数为

$$\min S = \sum_{i=1}^6 \omega_i S_i \quad (1)$$

式中: ω_i 表示不同差异函数对整体推荐模型的权重。

根据推荐模型目标函数,针对每一个用户,将所有可用资源特征与用户特征差异进行升序排序,选择排名靠前的资源生成候选资源推荐序列。

2 深度学习算法极其优化

2.1 广义回归神经网络

GRNN 是一个包含 4 层结构的神经网络结构,相比于普通神经网络,GRNN 增加了求和层^[8],其主要结构如图 1 所示。

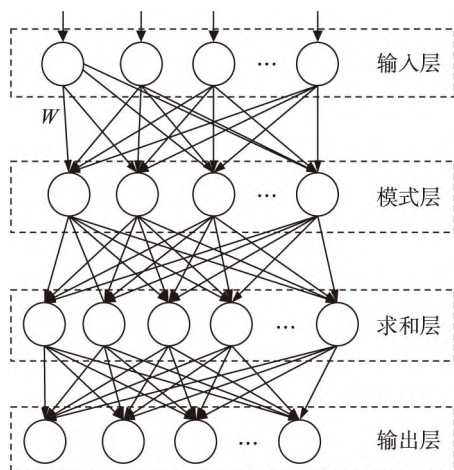


图 1 GRNN 结构

下面对 GRNN 网络结构进行数学描述,其网络结构主要参数如表 1 所示。首先,设 $X =$

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}, Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & & \vdots \\ y_{k1} & y_{k2} & \cdots & y_{kn} \end{bmatrix} \text{ 分}$$

别表示 GRNN 的输入和输出矩阵。

其中 x_{ij} 表示第 j 个样本的第 i 个输入属性, y_{ij} 表示第 j 个样本的第 i 个输出, m 表示输入属性总数, n 表示样本总数, k 表示输出值维度。

输入信号经过输入层,通过激活函数,获得模

式层结果^[9]

$$\alpha_j = \exp \left[-\frac{(X' - X_j)^T (X' - X_j)}{2\sigma^2} \right] \quad j=1,2,\dots,n \quad (2)$$

式中: σ 表示平滑因子 ($\sigma > 0$), α_j 表示第 j 个模式层神经元输出, X' 为任意输入的一个样本, X_j 为第 j 个训练样本。

在求和层,将所有模式层结果进行加权求和^[10]

$$S_{N_i} = \sum_{j=1}^n w_{ij} \alpha_j \quad i=1,2,\dots,k \quad (3)$$

令 $S_D = \sum_{j=1}^n \alpha_j$, 则有^[11]

$$\hat{y}_i = \frac{S_{N_i}}{S_D} \quad (4)$$

式中: \hat{y}_i 表示第 i 个输出层神经元对输入数据的估计量结果。

σ 和核函数中心对 GRNN 算法的训练影响最大,引入中心偏移因子 λ ,以调节核函数中心位置^[12],则式(2)变为

$$\alpha_j = \exp \left[-\frac{(X' - \lambda X_j)^T (X' - \lambda X_j)}{2\sigma^2} \right] \quad (5)$$

2.2 差分进化算法的 GRNN 优化

上一节提到影响 GRNN 算法性能的主要是平滑因子 σ 和核函数中心偏移因子 λ ,在 GRNN 进行资源推荐时,考虑到 σ 和 λ 的设置难度,采用差分进化 (Differential evolution, DE) 算法对 σ 和 λ 进行优化求解,DE 算法的适应度函数为 GRNN 资源推荐结果目标函数的倒数。下面将对 DE 算法进行数学描述。

设种群规模为 N ,属性维度为 D ,交叉速率 CR ,每个个体的取值为 $[U_{\min}, U_{\max}]$,则第 i 个个体的 j 维属性可表示为

$$x_{ij} = U_{\min} + \text{rand} \times (U_{\max} - U_{\min}) \quad (6)$$

式中: $i=1,2,\dots,N; j=1,2,\dots,D; \text{rand}$ 为 $(0,1)$ 随机数。

设第 G 代个体 $x_i^G (i=1,2,\dots,N)$ 的变异操作得到的第 $G+1$ 代个体为^[13]

$$v_i^{G+1} = x_{r_1}^G + F \times (x_{r_2}^G - x_{r_3}^G) \quad (7)$$

式中: $i \neq r_1 \neq r_2 \neq r_3, r_1, r_2$ 和 r_3 为第 G 代中除了编号为 i 的个体之外随机的 3 个个体, F 为差分缩放因子。

个体交叉方法为^[14]

$$u_{ij}^{G+1} = \begin{cases} v_{ij}^{G+1} & rand(0,1) \leq CR \\ x_{ij}^G & otherwise \end{cases} \quad (8)$$

对比 x_i^G 与 u_i^{G+1} , 分别求取两个个体的适应度值, 选择两者中值较高的个体进行进化, 具体方法为^[15]

$$x_i^{G+1} = \begin{cases} u_i^{G+1} & f(u_i^{G+1}) > f(x_i^G) \\ x_i^G & f(u_i^{G+1}) \leq f(x_i^G) \end{cases} \quad (9)$$

式中: f 表示适应度函数。当达到最大代数 G_{\max} 时, DE 算法停止。

F 常见取值为 $[0, 2]$, DE 的优化过程与 F 值密切相关, F 值不合适将会导致差分进化算法的优化性能不高, 因此在计算时引入自适应 F 值策略^[16]

$$F = F_{\min} + (F_{\max} - F_{\min}) \times e^{\frac{1 - G_{\max}}{G_{\max} - G + 1}} \quad (10)$$

式中: F_{\min} 和 F_{\max} 范围为 $[0, 2]$ 。

2.3 DE-GRNN 推荐流程

首先, 输入学习记录样本, 然后进行用户和资源的特征数字编码, 接着将特征变量输入 GRNN 网络, 并通过 DE 算法进行 GRNN 核心参数优化, 获得稳定的 GRNN 学习资源推荐模型, 最后通过多次训练, 获得稳定的 DE-GRNN 资源推荐模型。迭代停止的具体条件是特征差异是否满足要求, 也就是说特征差异是否达到最小值, 主要流程如图 2 所示。

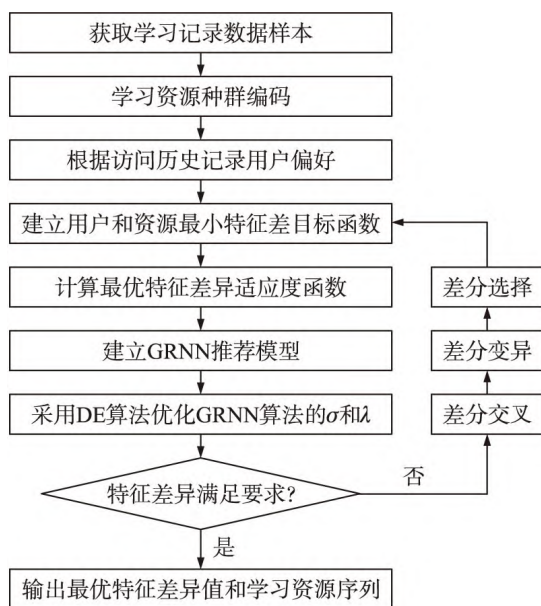


图 2 基于 DE-GRNN 的资源推荐

3 实例仿真

为了验证 DE-GRNN 算法在资源推荐中的性能, 进行实例仿真。数据来源为某大型在线学习平台, 具体样本数据见表 2。为了充分验证 DE-GRNN 的推荐性能, 首先验证 DE 对 GRNN 的优化性能, 分别采用 GRNN 和 DE-GRNN 算法进行推荐性能仿真, 其次分别采用常用资源推荐算法和本文算法进行推荐性能仿真。

表 2 仿真样本集

样本集编号	学习资源数量	学习记录条数
1	30	7528
2	50	12016
3	100	25012

3.1 DE 的优化性能

3.1.1 特征差异最小值性能仿真

为了验证 DE 算法对 GRNN 资源推荐的优化性能, 分别采用 GRNN 算法和 DE-GRNN 算法对表 2 中的 3 个样本集进行归一化特征差异平均最小值性能仿真, 以验证 2 种算法生成候选推荐序列的准确度。

图 3~图 5 均反映出 DE-GRNN 求解的特征差异值比 GRNN 算法更小, 对于 3 个样本集, DE-GRNN 算法的特征差异平均最小值均不大于 0.1, 而 GRNN 算法特征差异平均最小值均大于 0.15。从收敛性方面看, 在样本集 1 的特征差异最小值求解过程中, GRNN 更早完成收敛, 而在样本集 2 和样本集 3 的求解过程中, GRNN 和 DE-GRNN 算法表现非常接近。横向对比发现, 随着学习记录条数增多, 特征差异值更小。

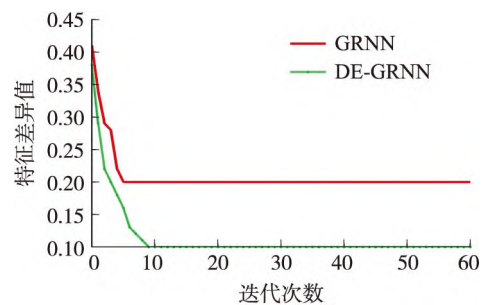


图 3 GRNN 与 DE-GRNN 的特征差异最小值 (样本集 1)

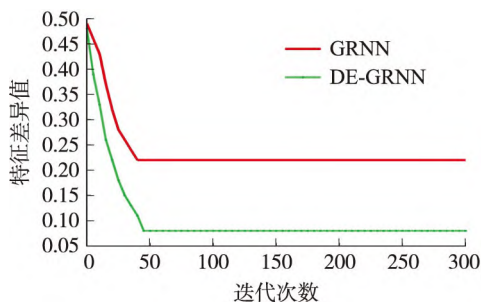


图 4 GRNN 与 DE-GRNN 的特征差异最小值 (样本集 2)

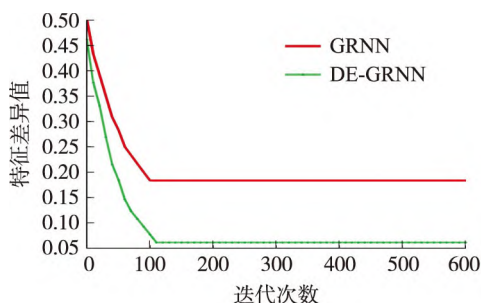


图 5 GRNN 与 DE-GRNN 的特征差异最小值 (样本集 3)

3.1.2 稳定性及效率仿真

为了进一步验证 DE 算法对 GRNN 的优化性能,对 GRNN 和 DE-GRNN 的特征差均值标准差和候选推荐序列求解时间进行性能仿真,结果如表 3 所示。

表 3 GRNN 和 DE-GRNN 算法的标准差及推荐时间

样本编号	算法	特征差异均值	标准差	耗时/s
1	GRNN	0.2002	0.0047	12.927
	DE-GRNN	0.1013	0.0032	14.033
2	GRNN	0.2297	0.0081	28.921
	DE-GRNN	0.0839	0.0066	30.785
3	GRNN	0.1834	0.0129	63.462
	DE-GRNN	0.0662	0.0092	65.562

从表 3 可知,DE-GRNN 的特征差异均值标准差性能明显比 GRNN 算法更小,其中 DE-GRNN 算法在样本集 1 的标准差性能最优,仅为 0.003 2,而 GRNN 的最小标准差为 0.004 7。从 2 种算法的候选推荐序列求解时间来看,学习记录条数越多,推荐耗时越长。对比发现,相同样本集,DE-GRNN 算法需要更多的推荐时间,但两者运算时间差距并不大,这是因为经过 DE 优化后,

GRNN 算法能够更快获得更小的特征差异值,所以经过 DE 优化后,并未耗费大量计算时间。

3.2 不同算法的推荐性能

为了验证不同算法对 3 个样本集的资源推荐性能,分别采用协同过滤^[4]、支持向量机 (Support vector machine, SVM)^[5]、深度神经网络^[6] 和 DE-GRNN 算法对 3 个样本集进行训练。对比用户实际选择的学习资源和经过 4 种算法推荐的候选资源序列,计算 4 种资源推荐系统的准确率、召回率和 F1 值。

从表 4 可知,对于 3 个样本集,DE-GRNN 的资源推荐准确率最高,均高于 0.91,深度神经网络次之,准确率都收敛于 0.8 以上,协同过滤和 SVM 表现较差,均未超过 0.7;而在召回率和 F1 值方面,DE-GRNN 算法的性能表现也高于其他 3 种算法。

表 4 4 种算法的推荐性能

算法	样本集	准确率	召回率	F1 值
协同过滤	1	0.6922	0.7047	0.6892
	2	0.6610	0.6652	0.6285
	3	0.6101	0.6122	0.6041
SVM	1	0.6814	0.7132	0.6653
	2	0.6523	0.6722	0.6126
	3	0.6071	0.6139	0.5825
深度神经网络	1	0.8891	0.8737	0.8762
	2	0.8549	0.8511	0.8516
	3	0.8028	0.8290	0.8188
DE-GRNN	1	0.9527	0.9132	0.9167
	2	0.9316	0.9027	0.9047
	3	0.9162	0.8631	0.8824

为了更充分地验证 4 种算法资源推荐准确率的稳定性,分别对 3 个样本集求解 4 种算法资源推荐准确率的均方根误差 (Root mean square error, RMSE) 值,其结果如图 6~图 8 所示。

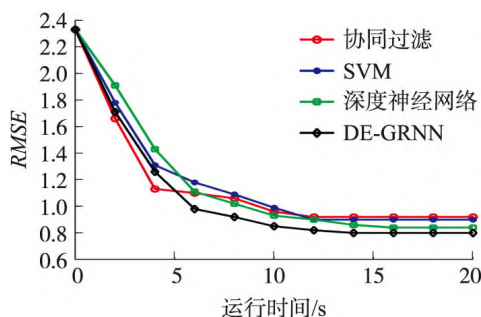


图 6 4 种算法的推荐准确率 RMSE 性能 (样本集 1)

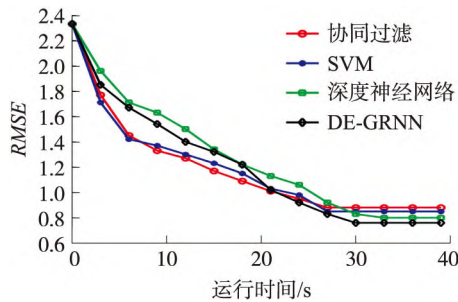


图 7 4 种算法的推荐准确率 RMSE 性能 (样本集 2)

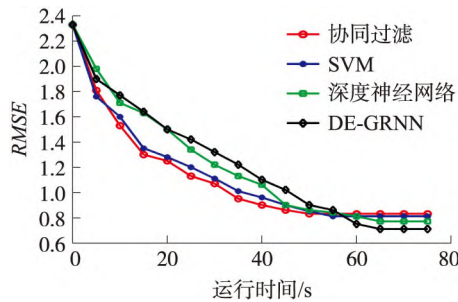


图 8 4 种算法的推荐准确率 RMSE 性能 (样本集 3)

从图 6、图 7 和图 8 可知,对于 3 个样本集,从推荐准确率的 $RMSE$ 值方面看,DE-GRNN 算法的性能最优,深度神经网络次之,协同过滤表现最差,SVM 和协同过滤算法的 $RMSE$ 接近。从推荐效率方面来看,协同过滤和 SVM 表现了更优越的性能,DE-GRNN 次之,深度神经网络表现最差。

在样本 2、3 中,也就是学习记录条数增多时,DE-GRNN 的准确率 $RMSE$ 值在没稳定前都是基本高于其他对比方法的,且其稳定速度也慢于其他方法。这是因为 DE-GRNN 不同于其他方法,其性能受到特征差异最小值求解过程的直接影响,且当特征差异接近最小值时才能趋于稳定。

4 结束语

将广义回归神经网络应用于推荐系统,在训练样本量充足时,合理设置 DE 算法的交叉速率和差分缩放因子,通过 DE 优化,能够对 GRNN 算法的平滑因子和核函数中心进行优化求解,获得合适的平滑因子和核函数中心,提高 GRNN 算法的资源推荐性能。试验证明,相比于常见资源推荐算法,本文算法能够获得更小的用户-资源特征差异度值,学习资源推荐更精准。后续研究将进一步对 GRNN 核心参数更新进行优化计算,以提高 GRNN 推荐效率,提高这一算法在学习资源推荐中的适用度。

参考文献:

- [1] Chaabi Y, Ndiaye N, Lekdioui K. Personalized recommendation of educational resources in a MOOC using a combination of collaborative filtering and semantic content analysis[J]. International Journal of Scientific & Technology Research, 2020, 9(2): 3243-3248.
- [2] 温占考, 易秀双, 刘勇, 等. 基于属性向量协同过滤推荐算法并行化[J]. 计算机工程与设计, 2018, 39(2): 425-429.
Wen Zhankao, Yi Xiushuang, Liu Yong, et al. Distributed collaborative filtering recommendation algorithm based on property vector [J]. Computer Engineering and Design, 2018, 39(2): 425-429.
- [3] 冯兴杰, 贺阳. 改进的 Hadoop 作业调度算法[J]. 计算机工程与应用, 2017, 53(12): 85-91.
Feng Xingjie, He Yang. Improvement of job scheduling algorithm on Hadoop [J]. Computer Engineering and Applications, 2017, 53(12): 85-91.
- [4] 莫恭钿, 陈正振. 基于协同过滤算法的课程资源平台的设计与实现[J]. 信息通信, 2019, 32(7): 78-79.
Mo Gongdian, Chen Zhengzhen. Design and implementation of curriculum resource platform based on collaborative filtering algorithm [J]. Information & Communications, 2019, 32(7): 78-79.
- [5] 崔建双, 车梦然. 基于多分类支持向量机的优化算法智能推荐系统与实证分析[J]. 计算机工程与科学, 2019, 41(1): 153-160.
Cui Jianshuang, Che Mengran. An intelligent recommendation system for optimization algorithms based on multi-classification support vector machine and its empirical analysis [J]. Computer Engineering & Science, 2019, 41(1): 153-160.
- [6] 樊海玮, 史双, 张博敏, 等. 基于 MLP 改进型深度神经网络学习资源推荐算法[J]. 计算机应用研究, 2020, 37(9): 2629-2633.
Fan Haiwei, Shi Shuang, Zhang Bomin, et al. Improved learning resource recommendation algorithm of deep neural network based on MLP [J]. Application Research of Computers, 2020, 37(9): 2629-2633.
- [7] Song Chenyang, Wang Liguang, Hou Jian, et al. The optimized GRNN based on the FDS-FOA under the hesitant fuzzy environment and its application in air quality index prediction [J]. Applied Intelligence, 2021, 51(11): 8365-8376.

- [8] 薛培. GRNN 神经网络在粮食产量预测中的应用[J]. 数据挖掘, 2020, 10(4): 247-253.
Xue Pei. Application of GRNN neural network in grain yield prediction [J]. Hans Journal of Data Mining, 2020, 10(4): 247-253.
- [9] Fannas L Y, Ben Sasi A Y. Off line signature recognition based on angle features and GRNN-neural-networks [J]. International Journal of Parallel Programming, 2019, 7: 229-233.
- [10] Zhang Yongli, Niu Jianguang, Na S. A novel nonlinear function fitting model based on FOA and GRNN [J]. Mathematical Problems in Engineering, 2019, 2019: 1-10.
- [11] Bani-Hani D, Khasawneh M. A recursive general regression neural network (R-GRNN) oracle for classification problems [J]. Expert Systems With Applications, 2019, 135: 273-286.
- [12] Ochoa P, Castillo O, Soria J. High-speed interval type-2 fuzzy system for dynamic crossover parameter adaptation in differential evolution and its application to controller optimization [J]. International Journal of Fuzzy Systems, 2020, 22(2): 414-427.
- [13] Wang S L, Morsidi F, Ng T F, et al. Insights into the effects of control parameters and mutation strategy on self-adaptive ensemble-based differential evolution [J]. Information Sciences, 2020, 514: 203-233.
- [14] Liu Qingxue, Du Shengzhi, van Wyk B J, et al. Double-layer-clustering differential evolution multimodal optimization by speciation and self-adaptive strategies [J]. Information Sciences, 2021, 545: 465-486.
- [15] 张维, 黄勃, 张娟, 等. 面向少数类用户兴趣演化的推荐算法 [J]. 南京理工大学学报, 2021, 45(2): 214-222.
Zhang Wei, Huang Bo, Zhang Juan, et al. Recommendation algorithm oriented to the interested evolution of minority users [J]. Journal of Nanjing University of Science and Technology, 2021, 45(2): 214-222.
- [16] Gu Tianlong, Liang Haohong, Bin Chenzhong, et al. Combining user-end and item-end knowledge graph learning for personalized recommendation [J]. Journal of Intelligent & Fuzzy Systems, 2021, 40(5): 9213-9225.