

基于胶囊网络与深度置信网络融合模型的手写汉字识别

管小卫, 丁琳

(江苏财经职业技术学院, 江苏 淮安 223003)

✉56491644@qq.com; 372369299@qq.com



摘要: 针对离线手写汉字的特征提取困难、不能准确识别等问题, 提出了一种胶囊网络与深度置信网络的融合模型。首先从CASIA-HWDB1数据集中随机选择了一些文本分别训练胶囊网络和深度置信网络, 然后采用胶囊网络和深度置信网络的融合策略进行了手写汉字识别实验。实验结果表明, 在不确定方向上使用汉字融合模型的错误率降低了5.2%, 与单独使用胶囊网络和深度置信网络相比, 具有更好的识别效果。

关键词: 手写汉字; 深度学习; 胶囊网络; 深度置信网络

中图分类号: TP399 **文献标识码:** A

Handwritten Chinese Character Recognition based on the Fusion Model of Capsule Network and Deep Belief Network

GUAN Xiaowei, DING Lin

(Jiangsu Vocational and Technical College of Finance & Economics, Huai'an 223003, China)

✉56491644@qq.com; 372369299@qq.com

Abstract: Aiming at the difficulties in feature extraction of offline handwritten Chinese characters and inaccurate recognition, this paper proposes a fusion model of Capsule Network (CapsNet) and Deep Belief Network (DBN). First, some texts from CASIA-HWDB1 data set are randomly selected to train CapsNet and DBN respectively. Then, handwritten Chinese character recognition experiments are conducted using the fusion strategy of CapsNet and DBN. The experimental results show that the error rate of fusion model is reduced by 5.2% for Chinese characters in uncertain direction, and it has better recognition effect than using CapsNet and DBN alone.

Keywords: handwritten Chinese characters; deep learning; capsule network; deep belief network

1 引言(Introduction)

汉字识别的研究涉及人工智能、模式识别、图像处理、统计决策理论等学科, 在办公、银行、邮政等自动分类领域具有重要的理论意义和实践价值^[1]。汉字识别可分为印刷汉字识别和手写汉字识别两大类。手写汉字识别可分为在线手写汉字识别和离线手写汉字识别。在线汉字识别是指在通过触摸屏等输入设备手写汉字过程中, 计算机根据书写汉字的笔画走向、笔画顺序、书写速度等多种信息进行识别, 由于信息量多且具有连续性, 因此识别难度较小, 识别准确率也较高^[2]。离线汉字识别提供的信息量少, 仅仅通过识别一个汉字的二维图像来提取汉字特征, 所以识别难度较大, 识别准确率也较低^[3]。各大高校和研究所致力于汉字识别的研究, 由于汉字具有类别多、字形复杂、相似等特点, 因此每个人手写汉字千差万别, 导致手写汉字识别困难, 故手写汉字识别一

直是研究的热点和难点^[4-5]。

手写汉字识别是一个极具挑战性的模式识别与机器学习问题, 特征如下: 一是汉字类别中的汉字数量很多。二是字体结构复杂。三是字形变化很大。离线手写汉字主要用在日常生活中。手写字是任意的, 缺乏规范性, 水平、垂直、点等笔触容易变形, 例如笔直变弯, 笔触成圆弧, 短横、短竖成点等。四是有很多类似的词。汉字集中有很多相似词, 如: “已一己一巳, 盲一育, 兔一免”等, 可能由于预处理不当从而导致字符错误。

总之, 诸多研究人员已经做了大量的离线手写汉字识别工作。针对离线手写汉字识别的难点, 文献[6]提出了一种CNN-DBN手写汉字融合模型, 比单独分别使用CNN和DBN的识别效果好。文献[7]从GoogLeNet网络构建了卷积神经网络, 并使用随机弹性变换算法扩展了训练数据。文献[8]结合

ResNet网络，通过使用作业图像中汉字笔迹测试结果中的每个有效检测区域作为输入，实现基于深度学习模型的汉字笔迹识别。文献[9]提出了一种基于PCCG-GAN的手写汉字归一化方法，实现了从手写到打印的汉字生成任务。利用对称网络提取手写汉字多尺度信息并进行特征融合，减少了输入输出之间共享的底层信息量，减少了欠采样过程中信息的丢失。文献[10]由于卷积神经网络对手写汉字的识别速度较慢，二维主成分分析(2DPCA)和卷积神经网络相结合来识别手写汉字与基于Alexnet的CNN模型相比，时间减少了78%，与基于ACNN的模型相比，时间减少了80%。文献[11]针对传统手写汉字识别特征提取过程复杂、识别率低、分类模型能力弱的问题，设计了一种多通道交叉融合的深度残差网络模型，并设计了中心损失函数，比之前的算法提高了2.3%识别率。

近年来，随着深度学习的兴起和不断发展，各种智能算法在计算机视觉领域取得了突破性的成果，使得手写汉字的识别率也越来越高。为了进一步探索线下手写汉字的识别方法，本文拟提出一种胶囊网络与深度置信网络融合模型，以提高手写汉字的识别能力。

2 相关工作研究(Related Work Research)

2.1 胶囊网络

Hinton等人在2017年提出了胶囊网络(Capsule Network, CapsNet)。CapsNet是建立在卷积神经网络(Convolutional Neural Networks, CNN)基础上的图像分类识别技术，CNN的缺陷是对于物体间的空间识别能力及物体旋转后的识别能力不强，而CapsNet能很好地解决这两个问题，如图1所示。CapsNet结构层次浅，由卷积层、主胶囊层、数字胶囊层构成，其结构如图2所示。

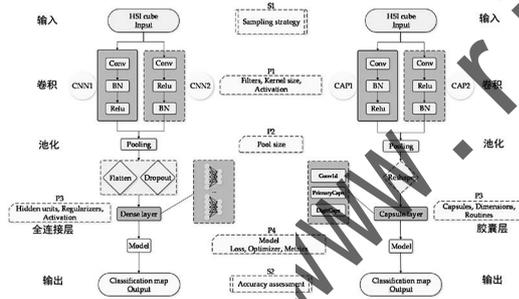


图1 CNN与CapsNet结构模型对比图

Fig.1 Comparison of CNN and CapsNet structure models

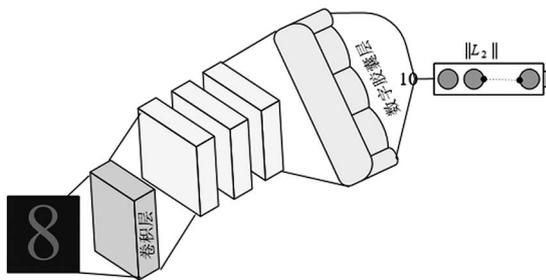


图2 胶囊网络结构模型

Fig.2 CapsNet structure model

传统的神经网络是由神经元组成的，神经元表示对象中的各种属性。胶囊(Capsule)称为向量神经元，向量长度表示对象存在的概率，向量方向表示对象的属性，它包含多个神经元。CapsNet用胶囊代替传统神经元，它的输入输出都是一个向量。低层封装需要将输出传递给该输出的高层封装。具

体地说，路由数据的传输是通过低层胶囊的输入与高层胶囊的输出的相似性来确定的。如果低层胶囊的预测矢量与高层胶囊的有效矢量具有高度相似性，则表明这两个胶囊高度相关。

CapsNet中使用了迭代动态路由算法，以与输出向量相似的方向在胶囊中获取向量，并且与向量的数量和模块的长度呈正相关。为了避免将内部乘积用作无上限情况的度量，将矢量压缩到输出之前。

2.2 深度置信网络

深度置信网络(DBN)由多层有限Boltzmann机器(RBM)和一层分类器组成，经典的DBN网络结构是由多层RBM和一层BP组成的深度神经网络。该深度模型广泛应用于图像分类识别、语音识别等领域。

DBN是基于生物神经网络的研究和浅层神经网络的发展而来的，并且从联合概率分布中推断出概率样本模型的数据样本分布。DBN生成模型通过训练网络结构中神经元之间的权重，整个神经网络根据最大概率生成训练数据，形成高级抽象特征，并提高了模型的性能。

DBN使用自下而上的传输，底层神经元接收原始特征向量，不断抽象到更高的水平，顶层神经网络形成易于组合的特征向量。通过添加层，可以抽象出更大的特征向量，并且网络的每一层都会削弱前一层的错误信息和辅助信息，以确保深度网络的准确性。DBN结构如图3所示。受限的Boltzmann机器由两个神经元节点组成，即隐藏层和可见层。

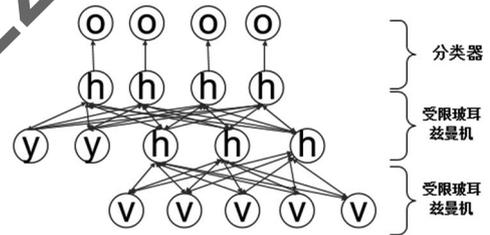


图3 DBN模型结构

Fig.3 DBN model structure diagram

2.2.1 受限玻耳兹曼机(RBM)

首先，DBN模型是基于人工神经网络的，由接收输入数据的显性神经元和特征提取的隐性神经元组成。DBN的关键组件是RBM，它通过将多层RBM与最终分类器结合在一起检测、识别和分类输入数据。RBM组成结构包含两层神经元，每一层都可以由一个向量表示，向量的维数由每一层中神经元的数量确定，如图4所示。

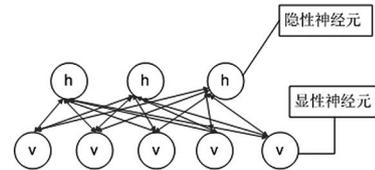


图4 RBM结构图

Fig.4 RBM structure diagram

由图4可知，为了保证层中神经元的独立性，RBM各层中的神经元之间无连接，层间神经元双向连接，即与给定元素相对应的隐藏元素的值无关，并且当该元素被赋予隐藏元素值时，将保留相同的特性。

2.2.2 DBN模型的构建

DBN就是一些堆叠在一起的RBM，前一个RBM的输出就是后一个RBM的输入。本文以两层RBM和Softmax分类层为

例构建DBN模型。如图3所示，v代表显示层神经元，h代表隐藏层神经元，y代表标签已知的样本，o代表分类结果输出。DBN模型的构造如下：固定第一个训练后的RBM的权重和偏差，将其隐藏元素的状态作为第二个RBM的输入；训练后将第二个RBM堆叠在第一个RBM上；重复上述过程进行多次逐层学习。如果训练数据集包含标记的样本，则需要在第二次RBM训练期间将其添加，最后使用Softmax对数据进行分类。

深度置信网络在检测识别领域的应用效果分析，如表1所示。样本数据主要是手写数字、频谱图像和语音。

表1 DBN检测识别领域优劣分析

Tab.1 Advantage and disadvantage analysis of DBN detection and recognition field

作者	方法	应用领域	优势	劣势
张绪冰等 ^[12]	DBN	频谱仪高光谱影像	提高分类精度	缺乏实验对比结果
郑益勤等 ^[13]	改进DBN	静止卫星图像海上强对流云团图像	提高识别精度	时间成本高
何群等 ^[14]	变分模态分解+DBN	运动想象分类识别	优化模型参数	分类性能低
徐宇恒等 ^[15]	DBN+不同优化算法	复杂体制雷达信号识别	自适应提取差异	训练时间长
陈亮等 ^[16]	IC-GAN+DBN	面部表情识别	降低误识率	时间成本高
张华丽等 ^[17]	CNN+DBN	肺结节影像分类识别	优化模型和参数，识别时间短	训练时间长
宋文豪 ^[18]	ResNet-BLSTM+DBN	语音识别、说话人信息识别	提高分类进度	分类前的样本处理与去噪
陈海霞 ^[19]	DBN+MTL	多维说话人信息识别	大幅度提高识别正确率	复杂度偏高

2.3 CapsNet与DBN融合模型

CapsNet模型和DBN模型都可以应用于线下手写汉字识别任务，CapsNet解决了CNN的缺点，能捕捉特征位置、相对大小、特征方向等属性间的关系。DBN通过采用逐层训练的方式为整个网络赋予了较好的初始权值，以重构数据为目标，使网络只要微调就可以达到最优解，这在无监督的环境中是非常有用的。

由于CapsNet和DBN网络结构不同，因此提取特征的优势不同，如字形特征不明显，或与其他字形相似，那么不同的人识别的结果可能也不一样。本文提出的CapsNet和DBN融合模型结合了两者的优势，比单独使用CapsNet和DBN识别汉字具有更高的识别能力，这种模型称为CapsNet-DBN融合模型，包括训练和识别过程，如图5所示。

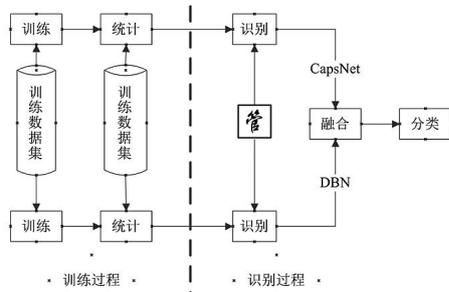


图5 CapsNet-DBN融合模型

Fig.5 CapsNet-DBN fusion model

3 仿真与测试(Simulation and Test)

3.1 实验准备

使用数据集CASIA-HWDB 1.1进行训练和测试，该数据集于2010年5月公布，为手写单字，分别由300人书写，包含171个英文数字符号，3,755个GB2312一级汉字，共300套，总计1,172,907个有效样本。表2列出了CapsNet-DBN模型中的各层参数。

表2 实验中使用的网络结构

Tab.2 Network structure used in the experiment

输入	隐藏层参数	输出
64*64	(32*3*3)-(2*2)-(64*2*2)-(2*2)-(96*2*2)-(2*2)-(128*2*2)-(2*2)-(512)	3,755

3.2 实验结果与分析

本文实验环境：Windows 10 64位操作系统，处理器为Intel(R) Core(TM)i7-6500U CPU，主频为2.5 GHz，内存为16 GB，实验平台为Python 3.8版本。实验数据来自http://www.nlpr.ia.ac.cn/databases/handwriting/Download.html。

表3为CapsNet网络、DBN网络及其融合网络的准确率。CapsNet、DBN、CapsNet-DBN对不同写法的汉字识别能力的比较如图6所示。图7表明三种算法对不同汉字的识别能力，CapsNet-DBN在大部分情况下都比CapsNet和DBN对不同汉字的识别能力要高。图8和图9分别显示了CapsNet-DBN网络训练过程准确率和训练损失。

表3 模型的准确率

Tab.3 Accuracy of model

模型	准确率(%)
CapsNet	97.12
DBN	95.42
CapsNet-DBN	97.57

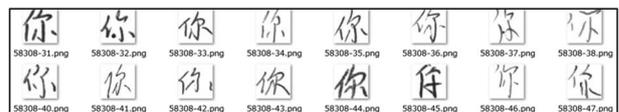


图6 待检测手写汉字

Fig.6 Handwritten Chinese characters to be detected

注：第一排下标1-8，第二排下标9-16。

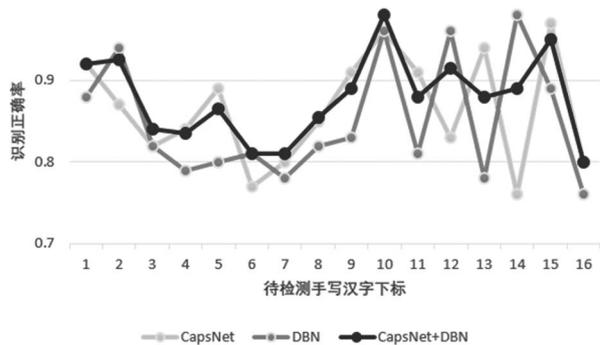


图7 CapsNet、DBN和CapsNet-DBN的识别性能对比

Fig.7 Recognition performance comparison of CapsNet, DBN and CapsNet-DBN

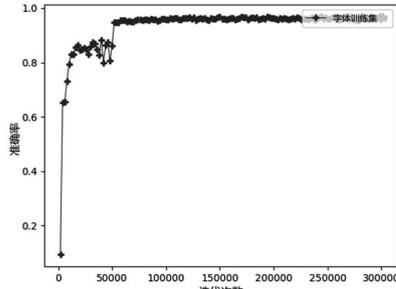


图8 CapsNet-DBN网络训练过程准确率

Fig.8 Accuracy of CapsNet-DBN network training process

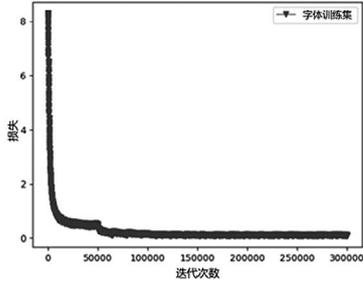


图9 CapsNet-DBN网络训练过程损失

Fig.9 Loss of CapsNet-DBN network training process

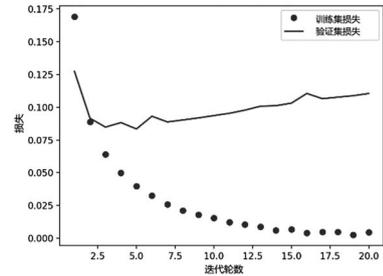
使用相同数据集验证并比较表1中的各种模型算法和本文所述的融合模型的识别准确率，从表4中可知，ResNet-BLSTM+DBN方法对手写汉字的识别效果良好，本文提出的CapsNet-DBN融合模型也可以获得较高的精度。

表4 不同模型准确率对比

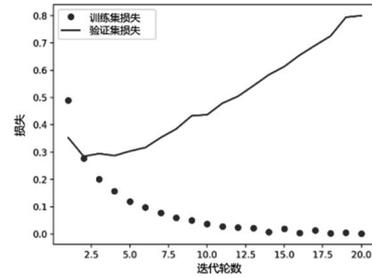
Tab.4 Accuracy comparison of different models

作者	模型	识别准确率(%)
张绪冰等 ^[12]	DBN	92.78
郑益勤等 ^[13]	改进DBN	93.12
何群等 ^[14]	变分模态分解+DBN	91.2
徐宇恒等 ^[15]	DBN+不同优化算法	91.25
陈亮等 ^[16]	IC-GAN+DBN	93.32
张华丽等 ^[17]	CNN+DBN	93.54
宋文豪 ^[18]	ResNet-BLSTM+DBN	94.12
陈海霞 ^[19]	DBN+MTL	90.21
本文	CapsNet-DBN	94.86

本文在CapsNet-DBN模型的基础上尝试增加隐藏层和隐藏单元数目及在不同的迭代次数中进行测试，分别增加了隐藏层个数及隐藏单元数。图10(a)隐藏层参数为(32*3*3)-(16*2*2)-(2*2)-(2*2)-(64*2*2)-(2*2)-(512)，图10(b)隐藏层参数为(32*3*3)-(2*2)-(16*2*2)-(2*2)-(32*2*2)-(2*2)-(64*2*2)-(2*2)-(512)，图10(b)比图10(a)增加了2个隐藏层。图11(a)隐藏层参数为(32*3*3)-(2*2)-(32*2*2)-(2*2)-(96*2*2)-(2*2)-(128*2*2)-(2*2)-(512)，图11(b)隐藏层参数为(32*3*3)-(2*2)-(64*2*2)-(2*2)-(96*2*2)-(2*2)-(128*2*2)-(2*2)-(512)，图11(b)其中一个隐藏层单元数为图11(a)的2倍。实验结果说明了隐藏层个数和隐藏单元数目的增加可以将大量信息扩展到维度较大的中间空间，这样会将模型的验证精度提高7.2%。当然，隐藏层个数和隐藏单元数目也不是越多越好，需要根据具体问题进行调整。



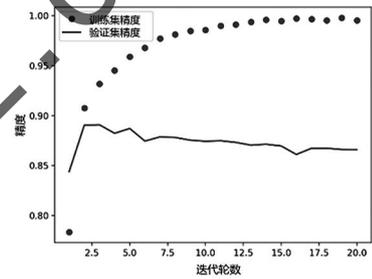
(a)7层隐藏层



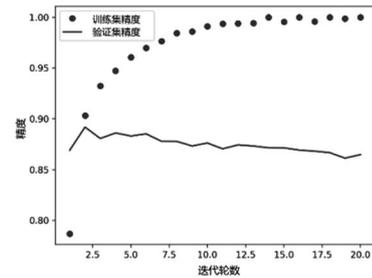
(b)9层隐藏层

图10 训练损失和验证损失对比

Fig.10 Comparison of training loss and verification loss



(a)其中一个隐藏层数为32层



(b)对应隐藏层数为64层

图11 训练精度和验证精度对比

Fig.11 Comparison of training accuracy and verification accuracy

4 结论(Conclusion)

本文针对离线手写汉字的特征提取困难、不能准确识别等问题，提出了一种胶囊网络与深度置信网络的融合模型，从CASIA-HWDB 1.1数据集中随机选择了一些文本进行手写汉字识别实验。实验结果表明，与单独的CapsNet和DBN及传统的机器学习方法相比，CapsNet-DBN融合模型在数据集上获得了更好的识别结果。

参考文献(References)

[1] SHEN X, MESSINA R. A method of synthesizing handwritten Chinese images for data augmentation[C]// International Conference on Frontiers in Handwriting Recognition(ICFHR). 2016 15th International Conference on Frontiers in Handwriting

- Recognition. Shenzhen, China: IEEE, 2016:114-119.
- [2] WU P L, WANG F Y, LIU J Y. An integrated multi-classifier method for handwritten Chinese medicine prescription recognition[C]// International Conference on Software Engineering and Service Science (ICSESS). 2018 IEEE 9th International Conference on Software Engineering and Service Science. Beijing, China: IEEE, 2018:1-4.
- [3] WANG Z R, DU J. Writer code based adaptation of deep neural network for offline handwritten Chinese text recognition[C]// International Conference on Frontiers in Handwriting Recognition(ICFHR). 2016 15th International Conference on Frontiers in Handwriting Recognition. Shenzhen, China: IEEE, 2016:548-553.
- [4] 闫喜亮,王黎明.卷积深度神经网络的手写汉字识别系统[J].计算机工程与应用,2017(10):246-250.
- [5] 金连文,钟卓耀,杨钊,等.深度学习在手写汉字识别中的应用综述[J].自动化学报,2016,42(8):1125-1141.
- [6] 李兰英,周志刚,陈德运.DBN和CNN融合的脱机手写汉字识别[J].哈尔滨理工大学学报,2020,25(3):137-143.
- [7] 林恒青,郑晓斌,王麟珠,等.基于深度卷积神经网络和随机弹性变换的脱机手写形近汉字识别[J].兰州工业学院学报,2020,27(3):62-67.
- [8] 徐穗越.面向作业智能评阅的汉字手写体检测与识别[D].石家庄:河北师范大学,2020.
- [9] 张红蕊.基于生成式对抗网络的汉字生成方法研究[D].天津:天津师范大学,2020.
- [10] 郑延斌,韩梦云,樊文鑫.基于二维主成分分析与卷积神经网络的手写体汉字识别[J].计算机应用,2020,40(8):2465-2471.
- [11] 张秀玲,周凯旋,魏其珺,等.多通道交叉融合的深度残差网络脱机手写汉字识别[J].小型微型计算机系统,2019,40(10):2232-2235.
- [12] 张绪冰,王贤敏,王凯,等.基于深度置信网络的CRISM影像火星表面矿物识别方法[J].地质科技通报,2020,39(4):189-200.
- [13] 郑益勤,杨晓峰,李紫薇.深度学习的静止卫星图像海上强对流云团识别[J].遥感学报,2020,24(1):97-106.
- [14] 何群,杜硕,王煜文,等.基于变分模态分解与深度置信网络的运动想象分类识别研究[J].计量学报,2020,41(1):90-99.
- [15] 徐宇恒,程嗣怡,董晓璇,等.基于DBN特征提取的雷达辐射源个体识别[J].空军工程大学学报(自然科学版),2019,20(6):91-96,108.
- [16] 陈亮,吴攀,刘韵婷.面向类内差距表情的深度学习识别[J].中国图象图形学报,2020,25(4):679-687.
- [17] 张华丽,康晓东,冉华,等.用于肺结节影像分类识别的DBN与CNN的比较研究[J].计算机科学,2020,47(z1):254-259.
- [18] 宋文豪.基于深度学习的说话人识别技术应用[D].成都:电子科技大学,2020.
- [19] 陈海霞.基于神经网络的多维说话人信息识别研究[D].南京:南京邮电大学,2019.

作者简介:

管小卫(1981-),男,硕士,讲师.研究领域:图像处理,软件工程.
丁琳(1980-),女,硕士,讲师.研究领域:物联网技术.

(上接第30页)

的并行随机微粒群算法的收敛率随进化代数的增加而增加,单纯形法的优化快速性在改进算法中得以体现.实验结果表明:对同一个测试函数,在其维数、进化代数均相同的情况下,由于全部初始点均随机产生,因此算法多样性好.在进化的过程中,因为很好地结合了单纯形法和随机微粒群算法的特点,新算法的收敛速度和收敛率均明显优于基本随机微粒群算法.据此说明改进算法是保证全局收敛的有效算法.

(2)通过构造开放实验室能耗多目标优化函数,并使用改进随机微粒群算法求解该优化函数,仿真实验表明了算法的有效性,为规划开放实验室、优化实验室资源、提高设备利用率提供了一种参考.

参考文献(References)

- [1] 刘宏炳.中医药学实验教学平台资源共享优化模型建设与探索[J].新疆医学,2020,8(5):879-882.
- [2] KANEMASA M, AIYOSHI E. Algorithm tuners for PSO methods and genetic programming techniques for learning tuning rules[J]. IEEE Trans on Electrical and Electronic

Engineering, 2014, 9(4):407-411.

- [3] SUGIHARA K, TANAKA H. Interval evaluation in the analytic hierarchy process by possibility analysis[J]. Computational Intelligence, 2001, 17(3):567-579.
- [4] 杨亮.群体动画行为自动控制的微粒群优化算法[J].现代电子技术,2020,43(19):106-110.
- [5] 林棋.微粒群进阶算法在天然气管道优化运行中的应用[J].石油工业技术监督,2020,36(8):33-39.
- [6] 黄扬飞.基于微粒群算法对城乡应急避难场所规划的研究[J].地震工程学报,2020,42(1):236-241.
- [7] 曾建潮,介婧,崔志华.微粒群算法[M].北京:科学出版社,2004:10-25.
- [8] 夏桂梅,曾建潮.一种基于单纯形法的随机微粒群算法[J].计算机工程与科学,2007,29(1):90-92.

作者简介:

王建丽(1983-),女,硕士,实验员.研究领域:最优化理论及其应用.