

改进粒子群 BP 算法的四六级翻译评分模型^①

唐泽¹, 吴敏^{1,2}, 吴桂兴², 郭燕²

¹(中国科学技术大学 现代教育技术中心, 合肥 230026)

²(中国科学技术大学 苏州研究院, 苏州 235123)

摘要: 针对四六级考试翻译题型, 给出了一种基于改进 PSO-BP 神经网络的评分方法. 通过 BLEU 和 SVD 等算法获取到文本特征值以及老师评分作为输入集, 然后用该集合对改进 PSO-BP 神经网络进行训练, 训练好的 BP 神经网络可以用来预测翻译分数. 从惯性权值计算和适应度函数两方面优化了 PSO-BP 算法, 在全局范围内寻找最优解, 使得实验效果更加稳定. 用 Matlab 进行了仿真实验, 结果表明, 在翻译评分中, 使用改进 PSO-BP 神经网络比采用多元线性回归能获得更好的相关性, 与人工评分的皮尔逊相关系数平均提高了 12%.

关键词: 翻译评分; BP 神经网络; 粒子群算法; Matlab 仿真

Translation Scoring Model in CET Based on Improved PSO-BP Neural Network

TANG Ze¹, WU Min^{1,2}, WU Gui-Xing², GUO Yan²

¹(Center of Modern Educational Technology, University of Science and Technology of China, Hefei 230026, China)

²(Suzhou Institute of University, Science and Technology of China, Suzhou 235123, China)

Abstract: This paper presents a new scoring method for the translation in CET-4 and CET-6 based on improved PSO-BP neural network. We get the text feature values by using the algorithm of BLEU and SVD, together with the scores the teacher have scored, are gathered as input set. We use it to train the improved PSO-BP neural network, which can be reversely used to predict the translation score. This paper improves the PSO-BP neural network by the calculation of the inertia weight and the adaptive value function. We use Matlab to make simulation, the result shows that, in the translation scoring, the use of improved PSO-BP neural network is better than using multiple linear regression to obtain better correlation, and the Pearson correlation coefficient of artificial scoring average increased by 12%.

Key words: translation scoring; BP neural network; PSO algorithm; Matlab simulation

1 引言

英语学习已成为大家逐渐关注的内容, 大学四六级考试是一项全国的大规模考试. 而每年国家都需要投入大量的人力物力集中批改试卷, 这在每年报考人数近千万的规模下, 确实是一项不少的开支. 随着计算机辅助技术和 e-learning 在教育领域的大量普及, 有关计算机自动阅卷的研究也正日渐成熟.

计算机主观题主要分为对翻译和写作, 在写作评分领域, 梁茂成博士综合国内外作文评分系统, 构建了一个写作评分模型^[1]. 而在翻译评分领域, 依然没有一个权威的参考模型. 最初对翻译评分的研究来自

机器翻译评分领域, 使用 BLEU 算法^[2]获得了不错的成效. 俞士汶教授在其开发的 MTE 系统^[3]中, 采用了基于测试点的方法, 体现了人工翻译评分中“信, 达, 雅”的标准. 江进林在“近五十年来自动评分研究综述”^[4]中提出了一些主要变量, 采用了多元线性回归的思想去拟合分数.

本文吸取了前人的经验, 分别提取了和翻译评分相关的六个特征, 提出了一种基于改进 PSO-BP 神经网络的新的翻译评分模型, 可以体现其中的部分非线性关系, 来尽可能提高和人工评分的相关度. 作文评分研究中梁茂成提出线性回归方法, 以一条直线来拟

^① 收稿时间:2014-12-29;收到修改稿时间:2015-02-05

合各特征与得分之间的关系. 本文研究翻译评分过程中, 发现很多特征并不可能很好的以直线来拟合, 比如拼写错误数, 很可能是决定两篇优秀的译文那关键的一分, 但是如果是两篇结构语法都乱七八糟的文章, 错三个字和错四个字其实对分数的影响不大. 显然, 此时拼写错误数和分数之间采用线性回归会削弱相关性. 另一方面, 采用单纯 BP 神经网络算法实验结果会产生一定范围内的波动, 主要是因为实验结果有多个局部极值点造成干扰, 而使用改进的 PSO-BP 神经网络算法, 在全局范围内寻优, 实验结果相对更加稳定.

2 文本特征提取及相关算法

2.1 文本特征提取

四六级评分标准主要强调语义表达, 用词, 连贯性以及语法和拼写错误数目. 根据评分标准, 统计出下述六个文本特征变量.

2.1.1 语义

2.1.1.1 SVD 值

LSA(潜语义分析)^[5]用来分析两篇译文的潜在语义相关性. 文章由词语组成, 在一篇文章中很可能出现两个同义词, 比如计算机和电脑, 在一般的词向量空间发现不了他们之间的相似性, LSA 就是为了解决这种问题而出现的. LSA 一般采用 SVD 奇异值分解的方法来挖掘文章单词矩阵的内在关联, 我们可以求出的内容相似性, 以此来预测译文的水平等级以及最终的分值.

2.1.1.2 BLEU 值

SVD 值代表了译文潜在的语义相关性, 而 BLEU 则更加的表面直观一些, 它代表译文和标准译文之间的语句相似度.

2.1.2 语言基础

评分标准中对译文的语言基础要求不是放在首位, 更看重译文的整体性, 但是如果一篇译文想要得到高分, 太多的语言错误也是不能容忍的. 一个英语能力比较好的学生, 语言功底往往是比较扎实的, 很少会犯基本的语言错误. 本文将语言基础以拼写错误和语法错误来体现, 使用 CMET 实验室工具进行检错纠错. 实验发现其中单词拼写错误呈现比较大的相关性, 而语法错误的相关性几乎可以忽略不计, 可能是由于语法错误检测成功的成功率比较低导致的.

2.1.3 连贯性

翻译更注重整体映象, 一篇文章, 即使其中有些许语言错误, 如果整体上读起来比较通顺, 整个篇章非常连贯, 一气呵成, 也可以看成是一篇比较好的译文.

在检测连贯性方面, 采用的是和连接词表进行匹配的方法, 研究发现, 一篇文章是否连贯, 和其使用的连接词比例多少成正比. 本文从词典和网络中收集了大量的连接词构成连接词表, 以匹配比例作为文本特征, 此参数呈微弱相关性.

2.1.4 测试点

像大多数题型一样, 每次翻译考试中也会有很多考点. 学生对考点的掌握程度会大大影响其对应的分数. 翻译对考点的考察可以分为两个方面. 其一是对考点的命中, 即学生有没有翻译该内容, 有没有漏掉某一块内容. 其二是对考点的把握程度, 同样一个词, 会有很多种翻译方式, 哪一种翻译方式最为准确, 这更能反映学生的英文水平.

在系统评分前, 由相关老师制作标准译文, 将制作好的译文进行批注, 标识出里面的测试点, 按照已经定义好的测试点规则制成 xml 形式的规则文件, 在系统初始化的时候载入系统.

2.2 重要算法

2.2.1 BLEU 算法

先介绍下 N-gram 的概念, 一个 N-gram 指的是一个字符序列中连续的 N 个连续字符组成的元组. 给定一个字符序列, 比如 I like to play football in the playground. 其中 like to 就是一个 2-gram, like to play 就是一个 3-gram.

BLEU 算法最初由 IBM 提出, 其认为机器译文和标准译文越相似, 则该机器翻译得分越高. 而这种相似度可以通过比较被测译文和标准译文的 N-gram 匹配数目. BLEU 算法采取的做法是计算各阶 N-gram 的片段数目和总片段数的比值, 并进一步加权平均求出其 BLEU 值. 为了避免短句子得分偏高的情况, 引入了长度惩罚因子.

其计算公式^[6]为

$$BLEU = BP * e^{\left(\sum_{n=1}^N w_n \log p_n\right)}$$

其中 w_n 为权重系数, p_n 为 n 阶元组的相似度比值, BP 为长度惩罚因子.

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{-\frac{r}{c}} & \text{if } c \leq r \end{cases}$$

c 为待测译文的长度, r 为待测译文中最短句的长度. 当 c 比较小时, 需要通过长度惩罚因子削弱其得分.

2.2.2 奇异值分解算法

根据语料构建单词文档矩阵 X, 矩阵的每一行表示某个单词在文章中出现的次数, 矩阵的每一列表示一篇文章中各个单词的出现次数, 这样矩阵中的每一个元素就表示一个单词在一篇文章中的出现次数.

通过公式 $X = U * \epsilon * V$ 对其进行奇异值分解, 得到三个矩阵. 其中 U 是单词特征矩阵, 矩阵每一行为一个单词的特征向量, V 是译文特征矩阵, 每一列为一篇译文的特征向量, ϵ 为奇异矩阵. 这里关注其中的 V 矩阵, 要想求解两篇译文的潜语义相似度, 只需要在分离出对应矩阵之后, 比较 V 矩阵的两个列向量的相似度即可, 而这可以通过求解两个向量的余弦. 假设两篇文章对应的列向量分别为 x, y, 则有

$$\cos(x, y) = \frac{x * y}{|x| * |y|}$$

3 改进的PSO-BP神经网络

3.1 BP 神经网络

在建立变量之间关系的模型上, 神经网络已越来越受到关注. 其中 BP 神经网络^[7]最为常用, 最初引入多级网络是为了解决线性不可分问题, 但隐藏层的神经元误差却无法得知, BP 神经网络是一种采用了误差反向传播算法的多层前馈网络.

一般的 BP 神经网络采用三层结构, 分为输入层、隐含层和输出层^[8].

输入层:

在本文中, 输入层即为提取出来的文本特征项, 具体来说分别为 SVD 值, BLEU 值, 拼写错误数, 语法错误数, 连贯度, 测试点匹配率一共为六个输入神经元. 这六个神经元分别涵盖了翻译评分标准中语义, 语言, 流畅度和得分点的内容, 非常具有说服力.

隐含层:

BP 神经网络的隐含层通常使用单层结构, 多层隐含层增大了计算量却不一定能提高网络的精度和效果. 隐含层的节点数对预测精度有比较大的影响, 一方面, 如果节点数太少, 网络学习的速度会下降, 并且还可能会影响训练的准确率; 另一方面, 如果节点数太多,

则会加大网络的负担, 导致网络过拟合. 本文采用 MATLAB 神经网络 30 个案例分析^[9]一书中的公式 $L < n - 1$ (n 为输入层神经元个数) 来确定隐层神经元的个数, 故此处隐含层个数取 4.

输出层:

本文旨在进行翻译评分, 故输出层神经元只有一个即译文分数, 根据翻译评分标准, 译文分数按标准分从 1 到 15 分分布.

其网络结构如图所示.

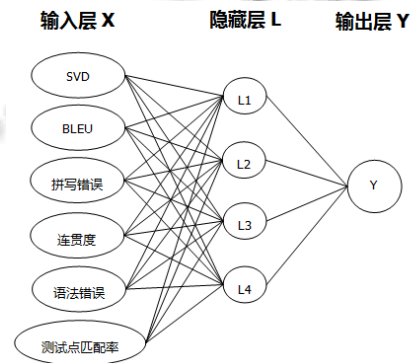


图 1 BP 网络结构图

3.2 粒子群算法

粒子群算法(PSO)起源于复杂适应系统(CAS), 最早由 Kennedy 和 Eberhart 受到自然界群体效应的启发而提出^[10]. 粒子群算法研究一个群体系统的效应, 群体中的每一个粒子都有位置和速度, 同时根据粒子的位置有一个适应度函数来判定该位置的优劣, 粒子群算法的目的就是为了找到最优位置.

在算法的初期会初始化一群粒子, 给予它们随机的位置和速度, 粒子通过跟踪自己的个体极值和种群的全局极值, 不断迭代更新自己的位置和速度, 直至找到最优位置, 也即适应度函数最大的位置.

假设每个粒子的维数空间为 N, 则第 i 个粒子的位置可以表示为 $x_i = (x_{i1}, x_{i2}, \dots, x_{iN})$, 其速度可以表示为 $v_i = (v_{i1}, v_{i2}, \dots, v_{iN})$, 同理用 p_i 表示该粒子的历史最优位置, 用 g 来表示种群历史最优位置, 上标 k 表示第 k 轮迭代, 迭代公式为:

$$v_i^{k+1} = w * v_i^k + c1 * a * (p_i^k - x_i^k) + c2 * b * (g - x_i^k)$$

$$x_i^{k+1} = x_i^k + r * v_i^{k+1}$$

其中 a, b 为从 0 到 1 之间的随机数, c1, c2 分别为保持个体历史最优位置和群体最优位置而设置的系数. w 为惯性权值, 改变惯性权值的大小可以平衡粒子的全

局寻优和局部寻优的能力。

3.3 改进的粒子群优化 BP 算法及其过程

3.3.1 PSO-BP 算法改进

(1) 惯性权值的自适应调整^[11]。

惯性权值越大, 则粒子的全局寻优能力越强, 惯性权值越小, 则粒子的局部寻优能力越强。在算法初期, 为了避免陷入局部最优解, 将惯性权值设置的比较大, 这样更容易在全局范围内寻找最优解。而在算法后期为了提高最优解的精度和准确度, 将惯性权值设置的比较小。陈贵敏^[12] 在研究粒子群算法权值递减策略时发现, 对于多数连续优化问题, 基于凹函数的权值递减策略要优于基于线性函数的递减策略, 因而本文采用的公式为:

$$w = (w_{start} - w_{end})(t_{max} - t)^2 / t_{max}^2 + w_{end}$$

其中 w_{start} 和 w_{end} 分别为最初权值和最终权值, 本文分别设为 1.5 和 0.3, t 为当前迭代次数, t_{max} 为总迭代次数。权值的迭代过程如下图所示:

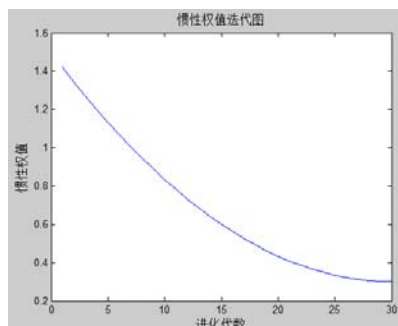


图 2 惯性权值非线性递减图

(2) 适应度函数更具针对性。

在作文评分中, 一般使用皮尔逊相关系数来评估评分性能的好坏, 翻译评分系统也可以引入这个系数。从这个角度设置适应度函数, 使得寻找最优解的过程变得更有针对性。

取训练集中的译文, 在适应度函数中运用种群中的参数训练 BP 网络, 用该 BP 网络求出预测值, 并计算预测值和期望值的相关度, 相关系数的计算公式为:

$$\rho_{X,Y} = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}}$$

3.3.2 基于改进 PSO-BP 算法的翻译评分过程

(1) 准备输入输出矩阵并确定神经网络的拓补结构。在这里输入矩阵为从译文中提取出来的文本特征

值, 输出矩阵为老师批改的分数。

(2) 为了取消各维数据间数量级差异, 神经网络预测前往往要对数据归一化处理, 一般使用最大最小法。

$$x_k = \frac{(x_k - x_{min})}{(x_{max} - x_{min})}$$

其中, x_{min} 和 x_{max} 分别为数据集中的最小值和最大值。

(3) 建立一个 30 个粒子的种群, 设置种群的参数, 即计算公式中的 w 、 $c1$ 、 $c2$ 、 r 等, 并随机初始化每个粒子的位置和速度。

(4) 在适应度函数中, 根据粒子的位置初始化 BP 网络的权重矩阵和阈值, 并根据相关度计算公式计算出评分相关度作为粒子的适应度。

(5) 计算粒子的个体历史最优值以及群体的历史最优解并保存最优解的位置。

(6) 检查是否超过迭代次数, 若超过则算法结束, 否则检查误差是否满足要求, 满足则算法结束, 否则按照计算公式更新每个粒子的位置和速度并转到第 4 步。

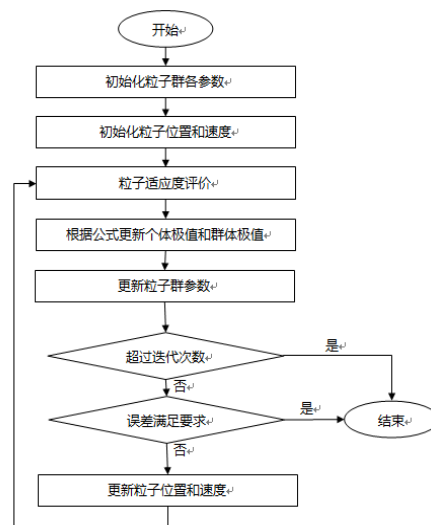


图 3 改进 PSO-BP 算法流程图

3.4 翻译评分系统的构建

(1) 利用 CMET 实验室的作文引擎处理语料并将单词错误和语法错误都集成到一个 xml 文件作为输出。

(2) 分析所得到的 xml 文件, 利用 python 脚本进行处理, 计算出其中的各个文本特征值。

(3) 组织老师对译文进行评分, 并将所有文本特征值和分数格式化输出到一个记事本文件中, 用来作为神经网络的输入数据。其中一行数据依次为 BLEU

值, 测试点匹配率, 连贯度, 拼写错误数, 语法错误数, svd 值以及分数.

```

0.731623326028 0.5 0.0780141843972 0 4 0.808399759669 12
0.769901927757 0.5 0.05 1 6 0.793218262434 9
0.829021956444 0.625 0.0423728813559 1 7 0.761508702443 14
0.834068545179 0.5 0.0792079207921 0 10 0.952674273757 10
0.772860867478 0.3125 0.0841121495327 1 6 0.463772299126 10
0.810981826604 0.5 0.0695652173913 1 4 0.684327595725 11
0.825230406982 0.5 0.0803571428571 1 4 0.834141088105 11
0.781547320078 0.375 0.0733944954128 2 3 0.735574769302 10
0.788224157393 0.3125 0.0576923076923 4 5 0.997800554744 10
0.810892292461 0.4375 0.0720720720721 2 7 0.90087816492 13
0.777023082802 0.375 0.0636363636364 0 6 0.769700303001 12
0.739769541859 0.375 0.0654205607477 0 7 0.956591561045 9
0.804027322394 0.5625 0.0648148148148 1 8 0.623918251542 11
0.843986105075 0.5625 0.0727272727273 0 7 0.511661636669 13
0.823306983697 0.5 0.06 0 8 0.713778619831 11
0.899518177207 0.5625 0.0816326530612 0 4 0.995894296796 10

```

图 4 输入数据图

(4) 将 223 组数据分为两组, 其中 180 篇为训练集, 后 43 篇为测试集. 以训练集中的数据对改进 PSO-BP 神经网络进行训练, 再用训练好的网络对测试集中的分数进行预测, 最后求出其皮尔逊相关系数来评估评分性能的优劣.

4 实验结果验证与对比

4.1 实验结果验证

本文在将训练集 180 篇译文数据对改进 PSO-BP 神经网络进行训练后, 接着对预测集中 43 篇译文进行了分数预测.

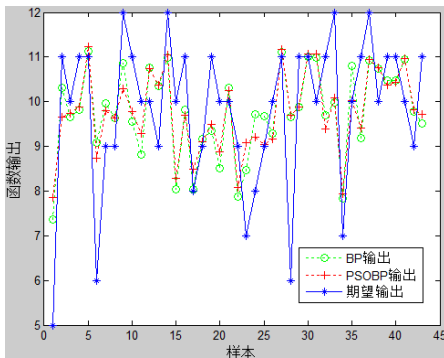


图 5 预测结果图

可以看出改进 PSO-BP 神经网络预测的结果和期望输出大体一致, 呈现很强的相关性.

4.2 实验结果对比

4.2.1 改进 PSO-BP 神经网络的优越性

PSO-BP 神经网络具有全局寻优的特点, 经过种群多次迭代能逐步发现最优适应度, 下图为本次实验的迭代图:

另一方面, 分别用 BP 神经网络的随机权重矩阵和阈值与改进 PSO-BP 神经网络算法经过 30 轮种群迭

代之后的权重矩阵和阈值去初始化神经网络, 发现改进 PSO-BP 神经网络更不易陷入局部极小值, 能够收敛于更优解.

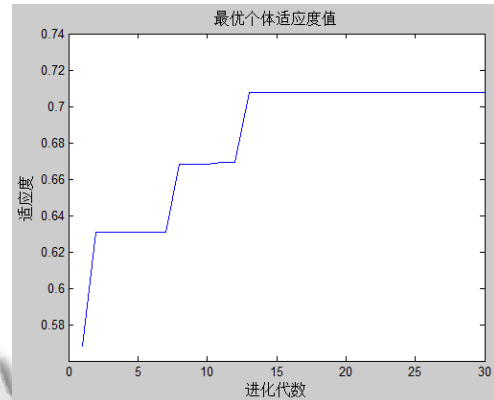


图 6 改进 PSO-BP 神经网络适应度迭代图

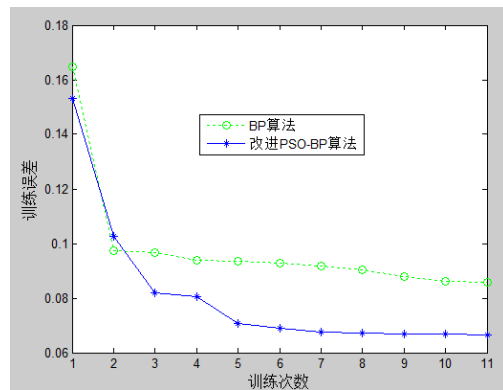


图 7 两种网络学习误差曲线

4.2.2 评分效应对比

作为对比, 本文按照多元线性回归的方法也对分数做了一个预测, 其最终实验结果如下图所示.

表 1 三种算法效果对比

	平均值 [□]	方差 [□]	与人工评分相关度 [□]
人工评分 [□]	9.8621 [□]	2.3374 [□]	
多元线性回归 [□]	9.7236 [□]	0.7880 [□]	0.5210 [□]
BP 神经网络 [□]	9.8531 [□]	0.7505 [□]	0.53-0.65 [□]
改进 PSO-BP 网络 [□]	9.8644 [□]	0.9996 [□]	0.64 以上 [□]

多元线性回归方法的皮尔逊系数为 0.5210, 用 BP 神经网络的方法, 其系数在 0.53 到 0.65 之间波动, 这是因为 BP 神经网络是用随机值初始化的, 但是其平均效果要优于多元线性回归方法. 采用改进 PSO-BP 算法, 因为其全局寻优的特性, 更容易找到最优解, 评分相关度基本保持在 0.64 以上. 另外从方差来看, BP 神经网络<多元线性回归<改进 PSO-BP 神经网络<人工评分, 故从数据稳定性来讲, BP 神经网络取得的

数据集是最稳定的,而改进 PSO-BP 神经网络为了与人工评分的值更加接近,牺牲了一点点方差。

5 结语

可以看出基于改进 PSO-BP 神经网络的算法在翻译分数预测上已经比多元线性回归和单纯的 BP 网络更加准确,皮尔逊相关系数稳定在 0.64 以上,但是这种相关度还是不够好,仍然达不到大规模使用的要求。究其原因,主要有两个。一、文本特征的计算还不够准确。比如依赖的语法纠错模块,并不能很好的发现语法错误,这给整个模型植入了一些不确定性。另外关于 SVD 值的直接计算并不科学,实验表明它对于评分的相关系数不是很高,还需要进一步研究发现其中的规律。二、特征的种类有限,应该发掘出更多新的特征,来提高评分相关度。下一步的工作将是优化各特征提取算法以及挖掘出新的特征,使得基于 BP 神经网络的大规模四六级评分能够用于实际。

参考文献

- 1 梁茂成.中国学生英语作文自动评分模型的构建[学位论文].南京:南京大学,2005.
- 2 Papineni K, Roukos S, Ward T, et al. BLEU: a method for automatic evaluation of machine translation. Proc. of the 40th Annual Meeting on Association for Computational Linguistics. Association for Computational Linguistics. 2002. 311-318.
- 3 俞士汶,姜新,朱学锋,等.机译译文质量自动评价原理.多语种机器翻译国际研讨会,1991.
- 4 江进林.近五十年来自动评分研究综述—兼论中国学生英译汉机器评分系统的新探索.现代教育技术,2013,23(6): 62-66.
- 5 Dumais S, Furnas G, Landauer T, et al. Using latent semantic analysis to improve access to textual information. Proc. of Computer Human Interaction. New York. ACM, 1988. 28-285.
- 6 孙连恒,杨莹,姚天顺.OpenE:一种基于 n2gram 共现的自动机器翻译评测方法.中文信息学报,2004,18(2).
- 7 蒋宗礼.人工神经网络导论.北京:高等教育出版社,2011.
- 8 冯继妙,胡立芳.基于 BP 神经网络的车祸库预警技术.计算机系统应用,2011,20(9):127-130,106.
- 9 MATLAB 中文论坛. MATLAB 神经网络 30 个案例分析.北京:北京航空航天大学出版社,2010.
- 10 Eberhart RC, Shi Y. Particle swarm optimization: development, application and resources. Proc. of the IEEE Congress on Evolutionary Computation. Piscataway. IEEE Service Center, 2001. 81-86.
- 11 张玲,王玲,吴桐.基于改进的粒子群算法优化反向传播神经网络的热舒适度预测模型.计算机应用,2014,34(3).
- 12 陈贵敏,贾建援,韩琪.粒子群优化算法的惯性权值递减策略研究.西安交通大学学报,2006,40(1):53-56.