

基于遗传算法的 BP 神经网络 在企业资信评估中的应用

孟凡超, 张洪伟, 徐 剑
(四川大学 计算机学院, 成都 610065)

摘要: 提出了一种新的企业资信评估方法。通过把神经网络和遗传算法有机地结合起来, 既克服了传统 BP 网络训练时间长、易陷入局部极值的缺点, 又利用遗传算法提高了网络全局收敛的效率。该模型采用 C#. NET + SQL server 2000 实现。实验结果表明, 基于遗传算法的 BP 神经网络系统对企业资信评估有着良好的性能。

关键词: 遗传算法; 反向传播神经网络; 资信等级评估

中图分类号: TP311.5 文献标志码: A 文章编号: 1001-3695(2007)08-0301-03

Application in credit rating evaluation based on BP neural network

MENG Fan-chao, ZHANG Hong-wei, XU Jian
(College of Computer Science, Sichuan University, Chengdu 610065, China)

Abstract: This paper presented a new credit rating evaluation algorithm by combining genetic algorithm and back propagation (BP) neural network. It not only conquered the demerits such as long time net-training and easily running into local optimum in traditional BP network, but also improved the convergent efficiency of network by using GA. The model was implemented with C#. NET and SQL server 2000. The rationality of experiment results indicate that BP-based neural network can evaluate corporations accurately and efficiently by credit rating.

Key words: genetic algorithm(GA); BP neural network; credit rating evaluation

随着我国市场经济和金融改革的不断深化, 电子商务高速发展, 企业客户资信评估体系的建设和完善已经迫在眉睫。

根据著名的“二八原则”^[1], 即企业 20% 的客户创造了 80% 的价值。要对这 20% 的客户实施最优质的服务, 发掘出他们对公司利润的最大化, 其前提便是从企业众多的客户中发现这 20% 的重点客户。所谓重点客户是指对企业具有重要意义、主要包括以下几类: 潜在客户, 即有价值的新客户; 交叉销售, 即同一客户有更多的消费需求; 客户保持, 即保持客户的忠诚度。重点客户的发现通常由一系列的数据挖掘来实现。例如通过分析客户对产品的应用频率、持续性等指标来判别客户的忠诚度; 通过对交易数据的详细分析来鉴别哪些是企业希望保持的客户^[2]。

客户资信度无疑是衡量客户是否是重点客户中极其重要的一环。本文提出了基于遗传算法的神经网络对客户资信进行评估分类的一种方法, 并在一大型粮食企业中实施应用, 取得了预期的效果。

1 基于遗传算法的神经网络

1.1 遗传算法

遗传算法(GA)由美国密执安大学的 Holland 教授提出。

它是模拟生物在自然环境中的遗传和进化过程而形成的一种自适应全局优化概率搜索算法。主要包括选择、交叉和变异等操作, 具有简单通用、鲁棒性强、适于并行处理的优点^[3]。

通过把神经网络和遗传算法有机地结合起来, 既利用神经网络较强的学习能力, 又结合遗传算法全局搜索的特点。本文在 BP 神经网络模型的基础上, 引入遗传算法, 提出融入遗传算法的神经网络预测模型, 是对 BP 神经网络的进一步优化应用。其基本思路是: 在 BP 算法训练网络出现收敛速度缓慢时启用遗传算法优化网络的运行参数, 把优化结果作为 BP 算法的初始值再用 BP 算法训练网络, 这样交替运行 BP 算法和遗传算法, 直到达到问题要求的精度。

1.2 遗传算法神经网络系统结构

BP(back propagation) 网络是 1986 年由 Rumelhart 和 McClelland 为首的科学家小组提出的, 是一种按误差逆传播算法训练的多层前馈网络, 是目前应用最广泛的神经网络模型之一。BP 网络的学习规则是使用最速下降法, 通过反向传播来不断调整网络权值, 使网络的误差平方和最小。如图 1 所示, 首先对客户资信指标进行分类提取, 然后抽取一定特征的指标, 作为输入信息送入一个具有由输入层、中间层(隐层)和输出层构成的三层网络进行训练。利用遗传算法根据训练目标函数对网络权值进行迭代学习, 找到最佳分类权值, 网络成为

收稿日期: 2006-05-23; 修返日期: 2006-07-19

作者简介: 孟凡超(1981-), 男, 山东人, 研究生, 主要研究方向为智能信息系统、数据库与计算机网络(ganzalaz@gmail.com); 张洪伟(1955-), 男, 四川人, 教授, 西德博士, 西德博士后, 主要研究方向为智能信息系统、数据库与计算机网络; 徐剑(1981-), 男, 贵州人, 研究生, 主要研究方向为智能信息系统、数据库与计算机网络。

稳定的模式评估器,即可输出评估结果。

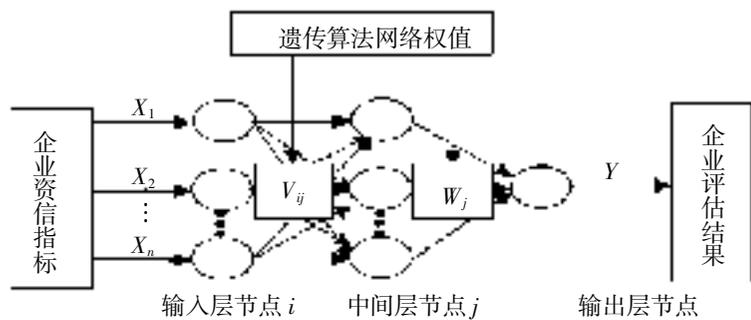


图1 网络训练模型图

模型的输入节点数目为 n (企业资信评估指标), 中间层节点数目为 n , 输出层节点数目为 1, 即该企业的资信评估值。输入层节点将信息直接传递给中间层节点, 故输入层节点输出与输入相等; 中间层节点的输出信息作为输出层节点的输入信息, 输出层只有一个节点, 接收中间层节点的输入并输出企业资信的评估结果。

a) 输入层节点 $X_i, i = \{1, 2, \dots, n\}$, 即企业的 n 个资信评估指标。

b) 中间层节点。

$$I_j = \sum_{i=1}^n v_{ij} X_i \quad (1)$$

输出为

$$O_j = 1 / \{1 + [(\sum_{i=0}^{n-1} v_{ij} X_i)^{-1} - 1]^2\} = 1 / \{1 + (I_j^{-1} - 1)^2\} \quad (2)$$

其中: v_{ij} 为输入层节点 i 到中间层节点 j 的权重; X_i 表示样本的第 i 个因子, 即企业第 i 个资信指标。

c) 输出层节点。输出层只有一个节点, 输入为中间层节点的输出:

$$y = 1 / \{1 + [(\sum_{j=0}^{n-1} w_j O_j)^{-1} - 1]^2\} \quad (3)$$

其中: y 为样本的实际输出值, 也就是企业计算所得出的资信值; w_j 为中间层节点 j 到输出层节点的连接权重。

1.3 遗传算法

1) 编码方式

常用的遗传编码方式有二进制编码和浮点数编码。二进制编码符合高等生物染色体为双倍体的重要生物特性, 有长期记忆作用, 便于实行各种遗传操作^[4,5]; 浮点数编码精度高, 便于大空间搜索。为了方便遗传操作, 对权系数采用二进制编码。

2) 训练目标函数

定义 P 个样本训练的网络实际输出与样本评估值的误差平方和的均值为训练目标函数:

$$E = (1/p) \sum_{p=1}^p [\bar{y} - y]^2 = (1/p) \sum_{p=1}^p E_j \quad (4)$$

其中: p 为训练样本模式总对数; y 为样本的网络实际输出值; \bar{y} 为样本的网络样本期望值。把被辨识对象(企业资信评估模型)的实际输出 y 作为反馈信号, 将其与神经网络的样本期望值 \bar{y} 进行比较, 使 $E < \epsilon$ 。其中 ϵ 为一个很小的数。如果不满足要求, 则不断调整权系数以达到相应要求。

3) BP神经网络预测系统权重调整模型

网络学习的目的是通过调整网络中的连接权重使 E 最小。应用梯度下降法, 连接权重的调整量为

$$\begin{cases} v_{ij} = - (E / V_{ij}) \\ w_j = - (E / W_j) \end{cases} \quad (5)$$

其中: η 为学习效率, $0 < \eta < 1$ 。则输入层节点与中间层节点的连接权重调整量为

$$v_{ij} = x_i w_j O_j^2 [1 - \sum_{i=1}^n v_{ij} x_i] \quad (6)$$

其中: $\delta_j = 2y^2 [1 - \sum_{j=1}^j w_j O_j] [\bar{y} - y]^2$ 。中间层节点与输出层节点的连接权重调整量为

$$w_j = y^2 O_j [1 - \sum_{j=1}^j w_j O_j] [\bar{y} - y]^2 \quad (7)$$

权重调整公式为

$$v_{ij}(t+1) = v_{ij}(t) + \eta \delta_j O_j x_i + \alpha v_{ij}(t) \quad (8)$$

$$w_j(t+1) = w_j(t) + \eta \delta_j O_j y + \alpha w_j(t) \quad (9)$$

其中: α 为动量系数, 且 $0 < \alpha < 1$ 。

应用上述模型, 根据通常神经网络的迭代算法可确定网络的连接权重值, 使实际输出与样本值误差减小。

4) 优选神经网络的连接权重优化问题描述

$$\min(E) = f(w_1, w_2, \dots, w_n)$$

其中: E 为网络训练的总误差; (w_1, w_2, \dots, w_n) 为统一编号后的连接权重(包括输入层节点与中间层节点的连接权重 v_{ij} 、中间层节点与输出层节点的连接权重 w_j); N 为连接权重总个数。约束条件为 $a_i \leq w_i \leq b_i (i = 1, 2, \dots, n)$ 。其中: b_i 和 a_i 分别为变量 w_i 变化的上界和下界。

5) 初始种群的产生

关于初始种群大小的选取, 根据研究表明认为^[1,5], 所求问题的非线性越大, P 越大。本文 $P = 30$, 初始种群由随机产生的 P 个串构成。

6) 适合度的计算

群体适合度函数的计算一直是遗传算法应用的一个瓶颈问题^[1,5]。由于个体的选择概率与适合度成正比, 适合度函数的定义方法对遗传算法具有极大的影响。适合度函数与目标函数之间存在严密的映射关系, 因此采用如下的适合度计算方法为

$$f_i = \begin{cases} C - E & E < C \\ 0 & E \geq C \end{cases} \quad (10)$$

其中: E 为训练目标函数; C 为当前代中所有 E 中的最大值。

7) 遗传操作产生新群体

遗传操作主要包括复制、交叉和变异。复制操作的目的是为了提高算法的全局收敛性和效率。其主要思想是串的复制概率 P_r 正比于其适合度, 即适合度大的个体将以较大的概率复制到下一代中。交叉操作的作用是组合出新的个体, 在串空间进行有效的搜索, 避免有效基因缺失。只有在发生了有效基因缺失的情况下突变操作的作用才较为明显。为了兼顾全局收敛的高效性又避免造成有效基因缺失所导致的早熟收敛的发生, 在遗传操作参数的选择上采用以下策略:

$$P_r = \begin{cases} 2(f_i / C)^2 & f_i / C \in [0, 0.5] \\ [1 - 2(1 - f_i / C)^2] & f_i / C \in [0.5, 1] \end{cases} \quad (11)$$

其中: C 为定义适合度时选取的参数; f_i 为适合度。式(11)使复制概率随适合度变化呈明显趋势变化, 即适合度大的个体将以更大概率遗传到下一代, 适合度较小的个体将以更小的概

率遗传到下一代。研究表明^[4], 交叉概率 P_c 通常取 [0.65, 0.9] 较合适, 而变异概率 P_m 通常为 [0.001, 0.1] 较合适。

1.4 GA 训练的多层前向神经网络算法步骤

a) 选定训练样本数据, 随机产生网络输入层和中间层、中间层和输出层的连接权重 v_{ij} 、 w_{jo} 。

b) 按照式 (1) ~ (3) 计算样本的实际输出 y 。

c) 按照式 (4) 计算样本实际输出值与期望值误差平方和的价值函数 E 。

d) 若 E 满足误差要求, 训练结束, 转至步骤 k); 否则, 进入步骤 e)。

e) 按照式 (6) (7) 计算输入层和中间层、中间层和输出层的权重调整值。

f) 按照式 (8) (9) 重新计算新的连接权重 v_{ij} 、 w_{jo} 。

g) 根据新的权重和样本数据, 按照式 (1) ~ (4) 重新计算样本的实际输出 y 和网络的平均误差 E 。

h) 若 E 满足误差要求, 训练结束, 转到步骤 k); 否则进入步骤 i)。

i) 设定种群大小 n 、权重变化范围、交叉概率 P_c 以及变异概率 P_m , 根据式 (10) (11) 确定复制概率 P_r 。

j) 复制、交叉、变异, 进行遗传算法计算。根据遗传算法产生的网络权重作为神经网络模型下一次训练的初始权重值, 转步骤 e)。

k) 训练结束, 输出满足训练精度的权重 v_{ij} 、 w_{jo} 。

l) 根据确定的权重值 v_{ij} 、 w_{jo} , 输入待评估的实验数据, 输出评估结果。

固定已完成训练任务的神经网络权值, 该网络模型便构成了一个模式评估器。由于网络固有的容错性, 即使训练样本组以外的企业客户资信指标输入该 BP 网络时, 网络都会给出一个实际评估结果。

2 企业资信评估

2.1 建立企业指标体系

企业资信评估包括资产和信用两个方面的评估。它是对独立经营企业或经济主体进行企业资产实力、偿债能力和信用风险程度的分析, 确定受信企业的信用条件, 为社会营造一个良好的经营环境。常用的工商业企业信用指标如表 1 所示。

表 1 商业信用指标

| 一级指标 | 二级指标 | 一级指标 | 二级指标 |
|---------|-------|----------|-----------|
| 企业经营者素质 | 经历 | 企业社会信用状况 | 借款偿付率 |
| | 业绩 | | 应付款偿付率 |
| | 品质 | | 纳税情况 |
| | 能力 | | 合同履行情况 |
| | 资产负债率 | | 企业发展前景 |
| | 流动比率 | | 新产品经营开发能力 |
| | 速动比率 | | 市场预期影响 |
| | 销售利润率 | | |
| | 资产报酬率 | | |
| | 存货周转率 | | |

2.2 根据理论初始化各企业指标体系和分类范围

根据中国工商管理学会和中国企业信用评估中心^[6,7] 以

及相关的理论研究^[8], 建立如表 2 所示的指标体系和等级分类范围。

表 2 各指标体系分值分布和等级范围

| 分类等级 | 资信评估指标 | | | | | | | | 评估结果分类标准(企业综合得分) | |
|------|-----------|----------|----------|-----------|-----------|-----------|-----------|-----------|------------------|------|
| | 资产 负债率 | 流动 比率 | 速动 比率 | 销售 利润率 | 资产 报酬率 | 存货 周转率 | 借款 偿付率 | 贷款 回收率 | | |
| 优 | 0.15 | 1.60 | 0.90 | 0.25 | 0.50 | 2.95 | 0.99 | 0.95 | ... | 89 |
| 良 | 0.35 | 1.55 | 0.85 | 0.18 | 0.25 | 2.85 | 0.96 | 0.85 | ... | 79 |
| 中 | 0.60 | 1.40 | 0.80 | 0.15 | 0.08 | 2.50 | 0.90 | 0.80 | ... | 60 |
| 差 | > 0.60 | < 1.40 | < 0.80 | < 0.15 | < 0.08 | < 2.50 | < 0.90 | < 0.80 | ... | < 60 |

2.3 实验结果与分析

为了检验上述 GA 神经网络训练模型的性能, 首先从标准样本库中选择一个样本输入网络训练模型初始化模型参数, 然后从某粮食业集团 DRP 系统中随机抽取的八个企业客户, 作为待评估分类样本, 如表 3 所示。设定网络训练的误差精度为 $=1 \times 10^{-3}$ 。设置神经网络的运行参数为: 学习效率 0.9、动量系数 0.9; 设置遗传算法的运行参数为: 权重变化范围 0.5^[4]、交叉概率 0.75^[4]、变异概率 0.005。

表 3 企业客户资信评估调查表

| 企业编号 | 资信评估指标 | | | | | | | | ... |
|-------|-----------|----------|----------|-----------|-----------|-----------|-----------|-----------|-----|
| | 资产 负债率 | 流动 比率 | 速动 比率 | 销售 利润率 | 资产 报酬率 | 存货 周转率 | 借款 偿付率 | 贷款 回收率 | |
| X^1 | 0.12 | 1.70 | 0.95 | 0.20 | 0.45 | 2.85 | 0.99 | 0.98 | ... |
| X^2 | 0.29 | 1.40 | 0.80 | 0.18 | 0.14 | 2.10 | 0.92 | 0.95 | ... |
| X^3 | 0.32 | 0.94 | 0.75 | 0.05 | 0.12 | 1.20 | 0.85 | 0.86 | ... |
| X^4 | 0.25 | 1.52 | 0.73 | 0.16 | 0.29 | 2.13 | 0.92 | 0.88 | ... |
| X^5 | 0.23 | 1.13 | 0.88 | 0.26 | 0.19 | 2.24 | 0.96 | 0.89 | ... |
| X^6 | 0.19 | 1.79 | 0.85 | 0.25 | 0.50 | 2.70 | 0.93 | 0.84 | ... |
| X^7 | 0.33 | 1.39 | 0.86 | 0.17 | 0.20 | 2.22 | 0.79 | 0.91 | ... |
| X^8 | 0.30 | 1.29 | 0.76 | 0.15 | 0.22 | 2.35 | 0.91 | 0.90 | ... |

当将上述八个企业客户的指标分别作为实验数据输入后, 产生的评估结果输出如表 4 所示。

表 4 实验输出结果

| 企业客户 | 输出结果 y | 企业客户 | 输出结果 y |
|-------|----------|-------|----------|
| X^1 | 90.016 7 | X^5 | 80.758 9 |
| X^2 | 7.882 2 | X^6 | 79.249 5 |
| X^3 | 57.949 5 | X^7 | 70.019 5 |
| X^4 | 67.860 5 | X^8 | 72.564 5 |

根据已定义的企业资信等级分类范围, 输出企业资信的评估结果如表 5 所示。

表 5 企业客户资信分类结果

| 优 | 良 | 中 | 差 |
|-------|------------|----------------------|-------|
| X^1 | X^5, X^6 | X^2, X^4, X^7, X^8 | X^3 |

3 结束语

本文对基于遗传算法的 BP 网络功能进行了分析和研究, 网络连接系数采用遗传算法进行训练寻优。其基本思路是在 BP 算法训练网络出现收敛速度缓慢时启用遗传算法优化网络的运行参数, 把优化的结果作为 BP 算法的初始值再用 BP 算法训练网络。这样交替运行 BP 算法和遗传算法, 直到达到问题要求的精度。实验采用一大型粮食业集团的八个企业客户的多个资信指标数据进行测试, 建立了一个基于遗传算法的网络模型。通过对神经网络连接权值的寻优, 亦即初始种群的代代遗传进化, 实现了对企业客户的资信等级评估分类。实验结果表明, 基于遗传算法的 BP 神经网络评估器对企业客户的资信等级评估分类有良好的性能。 (下转第 305 页)

始 CT 图像(图 3) 经过医学图像观察软件处理后,得到骨骼最清晰的图像,如图 4 所示;然后将图像进行分割并转换成二值图像,再进行图像去噪、边缘提取^[2~5]等处理后得到图 5 所示的骨骼轮廓图像,以测量颅骨的几何特征。

由于拍摄 CT 时头部位置摆放不标准等原因,使得很多拍摄出的 CT 图像有一定的倾斜,如图 3 所示,也就是镰膜(中线)与水平轴不垂直。从图中可以看出,颅骨水平断面 CT 图像中前额骨部分即夹在两条水平直线 L_1 与 L_2 中间的部分可以近似为抛物线,如图 5 所示。直线 L_1 与抛物线的交点为 E , F , 直线 L_2 与抛物线的交点为 G , H , 则线段 EF 和 GH 是抛物线的两条平行弦。记 EF 的中点为 A , GH 的中点为 B , 则直线 AB 一定平行于抛物线的对称轴, 因此根据上述抛物线的几何原理, 求出直线 AB 与中线的夹角 θ , 将图像旋转 θ 角, 就可以对图像进行纠偏校正。



图 3 颅骨 CT 图像

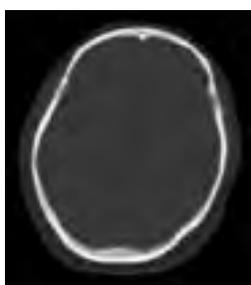


图 4 经过预处理的颅骨图像

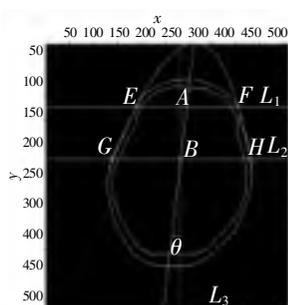


图 5 颅骨轮廓图像

1.3 纠偏算法的实现

利用抛物线原理对图 5 所示的颅骨轮廓图像进行纠偏时,不必求出抛物线的方程,而只需要求出水平直线 L_1 、 L_2 和颅骨轮廓边界的交点坐标,即用直线与颅骨轮廓边界的交点来代替直线与抛物线的交点。但是由于颅骨轮廓边界的方程未知,无法求出交点坐标的解析解。考虑到轮廓边界是一些离散点,直线 L_1 必定从颅骨轮廓边界上的某相邻两点间穿过,或经过边界上的某个点,这样就可能不存在精确的交点,此时只能求一个近似的交点。将图 5 中 E 点附近的离散图像进行局部放大,如图 6 所示。可以很清楚地看到,水平直线 L_1 从图像中颅骨轮廓边界上的 M 、 N 两点中间穿过。也就是说,直线 L_1 与图像的边界没有精确的交点,但是 M 、 N 的连线与直线 L_1 有一个交点 E , 考虑到 M 、 N 是 CT 轮廓图像边界上相邻的两点,可把 E 点作为直线 L_1 与 CT 图像边界的近似交点。

对于一般情况,设点 M 、 N 的坐标分别为 $M(x_1, y_1)$ 和 $N(x_2, y_2)$, 水平直线 L_1 的方程为

$$y = h \quad (2)$$

则直线 MN 的方程为

$$y - y_1 = (y_2 - y_1) / (x_2 - x_1) (x - x_1) \quad (3)$$

由式(2)(3)解得 E 点坐标为 $E[x_1 + (h - y_1)(x_2 - x_1) / (y_2 - y_1), h]$ 。

类似地,还可求出图 5 中点 F 、 G 、 H 的坐标,从而求出点 A 和 B 的坐标,计算出直线 AB 与中线的夹角 θ , 将图像旋转 θ 角, 就可以对偏斜的图像进行纠偏校正,如图 7 所示。

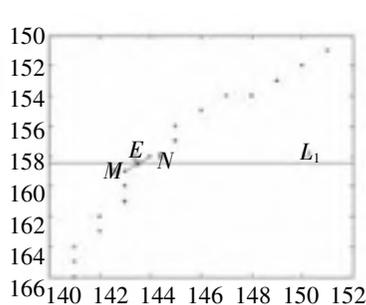


图 6 交点局部放大图

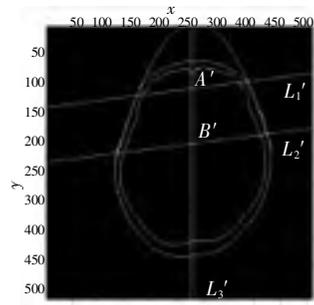


图 7 纠偏后的颅骨轮廓图像

2 结果

依据上述原理和方法,借助自编的 MATLAB 程序,有效地实现了对偏斜颅骨 CT 图像的自动纠偏校正,结果如图 7 所示。大量实验证明,偏斜角度在 30° 之内的图像均能很好地校正,纠偏后能够满足镰膜(中线)与水平轴垂直的要求。

3 结束语

该研究依据抛物线的一族平行弦的中点的轨迹是一条直线,并且该直线平行于抛物线的对称轴,或者与抛物线的对称轴重合这一几何原理。经过纠偏后的图像为实现基于 CT 图像的颅骨几何特征的自动测量提供了统一的数据。该方法不仅适用于对颅骨 CT 图像的纠偏校正,也适用于其他具有抛物线特征的图像纠偏校正。

参考文献:

- [1] 吕林根,许子道. 解析几何[M]. 北京:高等教育出版社,2001:184-230.
- [2] 李银波,洪波,高上凯,等. 人体 CT 切片图像中骨骼的分割[J]. 生物医学工程学杂志,2004,21(2):169-173.
- [3] TANG H, WUA E X. MRI brain image segmentation by multi-resolution edge detection and region selection[J]. Computerized Medical Imaging and Graphics,2000,24:349-357.
- [4] 吕庆文,李树祥. CT 图像中大脑边界的提取与跟踪及其算法实现[J]. 北京生物医学工程,1995,14(2):111-113.
- [5] 张金宏,陆光华. 基于颅脑 CT 图像的边缘提取[J]. 北京生物医学工程,2002,21(1):24-26.
- [6] 企业信用等级评定标准[EB/OL]. [2006-03-12]. <http://www.cccrec.com/info/gshowinfo.asp?id=276>.
- [7] 中国外贸企业诚信评级标准[EB/OL]. [2006-03-17]. <http://www.ibdibd.com/manage/shownews.asp?newsid=1784>.
- [8] 邹新月,王建成. 企业信用等级模糊综合评判[J]. 系统工程,2001,19(4):72-75.
- [9] 朱丽,张洪伟,谭辉. 基于软计算的资信评估研究[J]. 计算机应用研究,2005,22(5):74-77.
- [10] 恽为民,席裕庚. 遗传算法的运行机理[J]. 控制理论与应用,1996,13(3):297-303.

(上接第 303 页)

参考文献:

- [1] 巴诺斯 J G. 客户关系管理成功奥秘——感知客户[M]. 北京:机械工业出版社,2002.
- [2] LEUNG Y. Intelligent spatial decision support system[M]. Berlin: Springer-Verlag,1997:229-268.
- [3] WHITLY D. Genetic algorithms and neural networks[M]. Mississauga: Wiley,1995:203-216.
- [4] 张文修,梁怡. 遗传算法的数学基础[M]. 西安:西安交通大学出版社,2000.
- [5] 黄席樾,马笑潇. 基于方向基元集描述的指纹方向图求取算法