

上市公司财务危机预警系统的研究

王广正

(安徽工业大学 计算机学院, 安徽 马鞍山 243002)

摘 要:企业财务危机预测是非线性预测,各个影响因素之间又存在着复杂的组合决策关系,并且现实中的数据多为连续的,很难直接用于机器分类学习。因此文中从分析财务预警问题的特点出发,融合了智能软计算的多种方法建立完整的预测模型。首先以粗糙集决策表一致性水平、区间平均信息熵、离散化程度等因素为离散化结果的评价标准;然后利用遗传算法全局、并行搜索的优点,以上面提到的3个因素作为启发信息对所有条件属性的割点集合进行最优搜索。得到离散化的数据后,用BP神经网络对数据进行分类学习。最终网络学习训练后对企业财务状况进行了预测,实验结果表明:系统的预测正确率达93%。

关键词:财务预警;粗糙集;遗传算法;神经网络

中图分类号:TP183

文献标识码:A

文章编号:1673-629X(2008)11-0100-03

Research of Financial Crisis Early - Warning System of Listed Company

WANG Guang-zheng

(School of Computer Science, Anhui University of Technology, Maanshan 243002, China)

Abstract: Enterprise's financial crisis predicts is the non-linear prediction, there is a complicated association decision relation between each influence factor, and the data in reality are continuous, it is very difficult to be used in the categorized machine to study directly. After analyzing the characteristic of the early warning problem, merged many kinds of soft computing methods to construct the prediction model. Firstly, take consistency level of decision, average information entropy and degree of discretization as evaluation criteria of the result of discretization. Then utilize the overall search of genetic algorithm to find the optimized cut points. After discretization by the optimized cut points, train the BP neural network with the samples in training set. When finishing the training, use the BP neural network to predict financial crisis of listed company, and the experimental result indicates, the prediction rate is up to 93%.

Key words: financial early - warning; rough set; genetic algorithm; neural network

0 引 言

财务危机预警是管理科学与工程领域中研究的一个前沿和热点问题,是一项多学科交叉的边缘性研究课题,难度比较大。发展企业财务风险预警理论,积极应对21世纪新经济环境所带来的冲击,乃是大势所趋。所以建立企业财务危机预警系统,对财务营运做出预测,对各方面来说,都是十分必要的。

通过对上市公司的详细分析,决定采取下列方法来建立更好的预测模型:首先,可以从上市公司提供的财务年报提取重要的财务比率指标来构建粗糙集决策

表中的条件属性集合,并选用一个公司是否有财务危机作为粗糙集决策表的决策属性。其次,以多个财务比率指标进行决策分析,显然是一个非线性判别问题,决定采用BP神经网络来对问题进行建模,因为BP神经网络的输入神经元和输出神经元均可以在文中研究的问题中找到很好的对应。再次,财务属性的取值很显然是连续的,考虑到属性之间的关系和决策表的相容性,采用了从粗糙集决策表的一致性水平、属性区间信息熵、属性离散化区间数量等角度来考虑一种新的离散化算法。最后,遗传算法具有全局搜索、并行处理等特点,可以大大降低搜索算法的时间复杂度,所以采用遗传算法并将上面提到的三个离散化评价因素作为启发信息进行全局优化搜索。

1 基于粗糙集理论的计算模型的改进

文中财务危机预测模型的知识表示,采用了决策

收稿日期:2008-02-18

基金项目:国家自然科学基金(60473142);安徽工业大学科研项目(200704)

作者简介:王广正(1980-),男,山东郓城人,硕士,讲师,研究方向为自然语言处理、信息检索等。

表的形式化模型 $K = (U, A, V, Q)^{[1,2]}$ 。将各项财务比率指标作为决策表的条件属性集;将公司是否存在财务危机作为决策表的决策属性。将决策表一致性水平作为文中提出的(LEGAD)离散化算法的遗传启发信息之一,而且对一致性水平的计算模型做了改进。

决策表的一致性水平 L_c 定义^[2]为:

$$L_c = \frac{POS_C(D)}{|U|}, \text{ 其中 } POS_C(D) = \bigcup_{x \in U/D} C_-(x),$$

$POS_C(D)$ 为 D 的 C 正域,表示论域 U 中所有根据分类 U/C 的信息可以正确地划分到关系 D 的等价类中去的对象集合, $C_-(X)$ 为 X 的下近似集。

通过分析可以发现要计算决策表的一致性水平必须先求出分类 U/C 和 U/D ,而在对决策表搜索最优离散化割点集合的过程中,每次改变条件属性的割点集合都将重新计算 U/C ,算法计算量过于庞大^[3]。

因此文中改进了决策表的一致性计算模型,基本思想如下:重新构建决策表,将决策系统 $S = \langle U, C \cup d, V, f \rangle$ 转化为一个新的信息系统 $S^* = \langle U^*, KB_C, V^*, f^* \rangle$ 。其中, U^* 是由原系统 S 中决策属性取值不同的所有元组对构成的新论域。文中用割点集合能够区分元组对的个数来衡量粗糙集的一致性水平。如果该割点集合能够区分的元组对个数多,说明该离散化方式对于属于不同类别的对象区分能力强,决策表一致性水平高。反之,如果该割点集合能够区分的元组对个数少,则说明该离散化方式,对属于不同类别的对象的区分能力降低了,即破坏了决策表的一致性。在第 3 节,将给出算法的详细步骤。

2 标准遗传算法的不足及其改进

标准遗传算法的选择算子,采用按比例选择方法。其选择操作的原则是:每一个体被选中的概率与其适应度大小成正比^[4]。但由于是随机操作的原因,这种选择方法的选择误差比较大,有时候甚至连适应度较高的个体都可能在选择操作中被淘汰^[5,6]。

文中从所研究问题的特点出发,对标准遗传算法进行了一些优化处理。

1) 根据个体适应度的数学期望来选择下一代种群中个体,这样可以避免由按比例选择方法中的不确定因素所带来的影响,适应度较高的个体将肯定能被选中,并且有些适应度非常高的个体在下一代种群中的期望生存数量会大于 1。

2) 引入了最优保存策略模型,即当前群体中适应度最高的个体不参与交叉、变异运算,而是用它来替换本次进化产生的种群中适应度最低的个体。在具体算法实现中,系统每次最优保存的个体数量可以自由调

整,加强了算法的灵活性。

3) 在算法终止条件的设定上,定了迭代进化次数的上界 1000,如果在迭代过程中连续 10 代的种群平均适应度无明显变化,则认为种群进化成熟,算法可以提前终止。

3 基于粗糙集一致性和信息熵的遗传方法的离散化算法

(1) 算法说明与执行流程。

① 算法说明:

输入参数 测试集和训练集中包含原始数据的决策表;

输出参数 各个条件属性经离散化处理以后的决策表;

基本思想 采用遗传算法对所有条件属性的候选割点集进行全局搜索,启发信息采用粗糙集决策表一致性水平、离散化区间平均信息熵、离散化区间数量^[7-9]等 3 个离散化效果评价标准。当遗传算法最终搜索到所有条件属性的最优割点集合后,利用该割点集合对原始数据进行离散化处理。

② LEGAD 算法执行流程。算法执行流程见图 1。

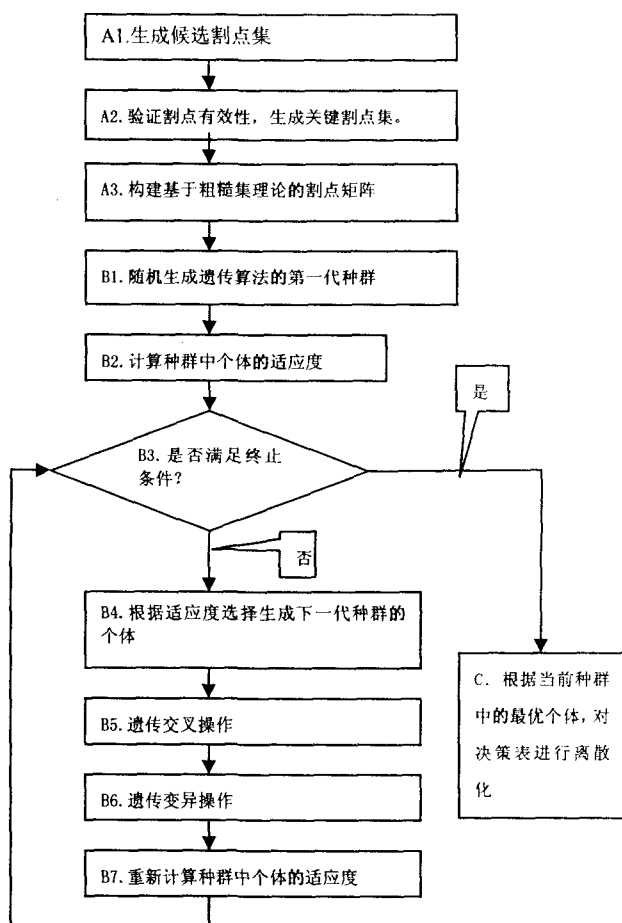


图 1 LEGAD 算法执行流程

(2) 算法改进后的分析。

①全面的启发式信息:采用了决策表一致性、信息熵和割点数目等 3 个启发式信息,使得最终得到的离散化结果更加合理。

②空间换时间的策略:根据每一个条件属性 c 生成按相应的 c 升序排序后的视图。这样可以大大减少生成“粗糙集割点矩阵”的时间复杂度。并且由于生成的仅仅是视图,而不是实际存在的 Table,所以耗费的空间也相当小。

③简化了决策表一致性的计算模型:构建了新的信息系统 S^* ,并用 $\beta^* \frac{D_v}{\text{card}(U^*)}$ 来衡量决策表的一致性水平,和传统的计算一致性水平的算法相比,大大降低了复杂度。

④改进的遗传选择操作:摒弃了传统的按比例选择,而是根据数学期望来决定哪些个体能够被选择到下一代,这样可保证适应度大的一些个体一定能够被保留到下一代群体中。同时,又采用了最优保留策略,可由用户自由指定最优保留的个体数目,增强算法的灵活性,方便试验测试。

4 实验分析

4.1 连续属性离散化

由于大多数分类器都要求连续属性数据必须预先离散化之后才能进行,这样有利于获得更好的预测结果,同时也将大大提高模型的收敛速度。因此,先用文中提出的 LEGAD 算法对决策表进行离散化处理。

该方法分为两步:

(1)生成基于粗糙集理论的割点矩阵。

(2)利用遗传算法寻找最优割点集并对样本离散化。

4.2 数据归一化和 BP 神经网络分类

由于某些属性之间取值过于悬殊,如果直接进行神经网络的训练,会造成某些属性权重过大。因此,先对离散化后的决策表进行一次数据归一化处理。将所有属性的取值限制在 $[0,1]$ 区间中。

最终采用的 BP 网络的结构是 6-9-1 的三层网络结构,其中输入层 6 个神经元,对应于所选的 6 个财务比率指标。

5 结束语

以上市公司年度财务报表为依据,采用粗糙集、遗传算法、BP 神经网络等智能软计算方法,对上市公司财务危机建立预测模型。为了提高离散化算法执行效率,先构建粗糙集割点矩阵,改进了决策表一致性水平

的计算模型。然后以粗糙集决策表一致性水平、区间平均信息熵、离散化程度等 3 个因素作为遗传算法的启发信息,对决策表条件属性的所有候选割点进行全局最优搜索。进而利用遗传算法搜索得到的最优割点对训练集、测试集的数据样本进行离散化操作。最终将离散化的决策表归一化处理后用 BP 神经网络对训练集中的数据进行示例学习,网络学习训练完毕后就可以对企业财务状况进行预测。实验结果表明文中构建的上市公司财务危机预警模型预测成功率达到 93%,说明该模型是有效、可行的。

文中以上市公司年报数据作为基础对其未来财务状况进行预测。但没有考虑上市公司财务年报的数据是真实可靠和公司年报数据以外的因素,如当前行业基本面、相关产业发展情况等都会对公司财务状况产生一定的影响。因此以后的一个工作方向可以是对上市公司历史财务报表进行置信度分析,剔除可能的虚假信息。而且在建立预测模型时还可以结合一些其他的机器学习算法来减少现有模型中的一些缺点和不足,如支持向量机就可以很好地解决局部收敛的问题,目前 SVM 方法也已经取得了很多成功应用的案例。

参考文献:

- [1] 王国胤. Rough 集理论与知识获取[M]. 西安:西安交通大学出版社,2001.
- [2] 张文修,吴伟志,梁吉业,等. 粗糙集理论与方法[M]. 北京:科学出版社,2001.
- [3] 黄金杰,李士勇,蔡云泽. 一种建立粗糙数据模型的监督模糊聚类方法[J]. 软件学报,2005,16(5):744-753.
- [4] Hsu Chen-Chien, Wang Wei-Yen, Yu Chih-Yung. Genetic Algorithms - Derived Digital Integrators and Their Applications in Discretization of Continuous Systems[C]//In Proceedings of the 2002 Congress on Evolutionary (CEC2002). [s.l.]:[s.n.],2002:443-448.
- [5] Jiang Hao, Yan Pu-Liu. A New Deduction Algorithm - Difference-Similitude Matrix[M]. [s.l.]:IEEE,2003.
- [6] 谢宏,程浩忠,牛东晓. 基于信息熵的粗糙集连续属性离散化算法[J]. 计算机学报,2005,28(9):1570-1574.
- [7] Li Ren-Pu, Wang Zheng-Ou. An Entropy-Based Discretization Method For Classification Rules With Inconsistency Checking[M]. Beijing:[s.n.],2002:243-246.
- [8] Dai Jian-Hua, Li Yuan-Xiang. Study on Discretization Based on Rough Set Theory[C]//In Proceedings of the first International Conference on Machine Learning and Cybernetics. Beijing:[s.n.],2002:1371-1373.
- [9] 李军,刘艳,顾雪平. 基于信息熵的属性离散化算法在暂态稳定评估中的应用[J]. 电力系统自动化,2005,29(8):26-31.