Journal of Computer Applications
计算机应用,2013,33(11):3094-3096,3148
ISSN 1001-9081
CODEN JYIIDU
2013-11-01
http://www.joca.cn

文章编号:1001-9081(2013)11-3094-03
doi:10.11772/j.issn.1001-9081.2013.11.3094

基于稀疏贝叶斯学习的个人信用评估

李太勇[1,2*],王会军[1],吴　江[1,2],张智林[3],唐常杰[4]

(1.西南财经大学 经济信息工程学院,成都 610074;　2.西南财经大学 中国支付体系研究中心,成都 610074;
3. Emerging Technology Lab, Samsung Research and Development Institute America-Dallas, Richardson TX 75082, USA;
4.四川大学 计算机学院,成都 610064)

(*通信作者电子邮箱 litaiyong@gmail.com)

摘　要:针对传统信用评估方法分类精度低、特征可解释性差等问题,提出了一种使用稀疏贝叶斯学习方法来进行个人信用评估的模型(SBLCredit)。SBLCredit充分利用稀疏贝叶斯学习的优势,在添加的特征权重的先验知识的情况下进行求解,使得特征权重尽量稀疏,以此实现个人信用评估和特征选择。在德国和澳大利亚真实信用数据集上,SBLCredit方法的分类精度比传统的*K*近邻、朴素贝叶斯、决策树和支持向量机平均提高了4.52%,6.40%,6.26%和2.27%。实验结果表明,SBLCredit分类精度高,选择的特征少,是一种有效的个人信用评估方法。

关键词:稀疏贝叶斯学习;分类;信用评估;金融风险;特征选择

中图分类号:F830.5;TN911.7　　**文献标志码**:A

Sparse Bayesian learning for credit risk evaluation

LI Taiyong[1,2*], WANG Huijun[1], WU Jiang[1,2], ZHANG Zhilin[3], TANG Changjie[4]

(1. *School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu Sichuan* 610074, *China*;
2. *Institute of Chinese Payment System, Southwestern University of Finance and Economics, Chengdu Sichuan* 610074, *China*;
3. *Emerging Technology Lab, Samsung Research and Development Institute America-Dallas, Richardson TX* 75082, *USA*;
4. *School of Computer Science, Sichuan University, Chengdu Sichuan* 610064, *China*)

Abstract: To solve the low classification accuracy and poor interpretability of selected features in traditional credit risk evaluation, a new model using Sparse Bayesian Learning (SBL) to evaluate personal credit risk (SBLCredit) was proposed in this paper. The SBLCredit utilized the advantages of SBL to get as sparse as possible solutions under the priori knowledge on the weight of features, which led to both good classification performance and effective feature selection. SBLCredit improved the classification accuracy of 4.52%, 6.40%, 6.26% and 2.27% averagely when compared with the state-of-the-art *K*-Nearest Neighbour (*K*NN), Naïve Bayes, decision tree and support vector machine respectively on real-world German and Australian credit datasets. The experimental results demonstrate that the proposed SBLCredit is a promising method for credit risk evaluation with higher accuracy and fewer features.

Key words: Sparse Bayesian Learning (SBL); classification; credit risk evaluation; financial risk; feature selection

0　引言

信用评估是商业银行控制风险的关键技术,发生在美国的"次贷危机"就是信用风险的大爆发,因此信用评估方法的研究具有非常重要的现实意义。信用评估实质上是数据挖掘中的分类问题——将贷款者根据其属性分成能够按期还本付息的可信的"好"客户(正类)和违约的"坏"客户(负类)两类,进而预测未来贷款人的违约风险,为消费信贷决策提供科学依据。

由于信用评估的重要性,它已成为近年来的研究热点,信用评估的方法主要有决策树[1]、朴素贝叶斯(Naïve Bayes)[2]、*K*近邻(*K*-Nearest Neighbour, *K*NN)[3]、支持向量机(Support Vector Machine, SVM)[4-6]、自然计算[7-8]及这些方法的集成[9-11]等。但已有方法大多存在分类精度低,不能有效进行特征选择以致模型可解释性差[8]等问题。本文将最近几年在信号处理、模式识别中的研究热点——"稀疏学习"引入到信用评估中,提出了一种基于稀疏贝叶斯学习(Sparse Bayesian Learning, SBL)的个人信用评估模型(Sparse Bayesian Learning-based Credit, SBLCredit)。SBLCredit模型首先添加各属性的权重先验知识,然后在该先验知识约束下求解属性权重,以此建立信用评估模型;对于一个待分类样本,先计算各权重与属性值乘积的累加和,然后将得到的值映射到正、负类。据作者所知,这是首次将稀疏学习方法应用到信用评估上。在取自德国和澳大利亚的真实信用数据集上做了实验,结果表明,相对于传统的*K*NN、Naïve Bayes、决策树和SVM分类方法,SBLCredit算法具有更高的分类精度且选出的特征更稀疏。

1　稀疏贝叶斯学习框架

稀疏学习是近年的研究热点,在信号处理、模式识别和机器学习等领域得到了广泛研究,而稀疏贝叶斯学习框架是一种典型的稀疏学习方法,相对于传统的基于L1惩罚项稀疏学

收稿日期:2013-05-20;**修回日期**:2013-07-16。　**基金项目**:教育部人文社会科学研究青年基金资助项目(11YJCZH084);中央高校基本科研业务专项资金资助项目(JBK130142, JBK130503);西南财经大学科研基金资助项目(2011XG130)。

作者简介:李太勇(1979-),男,四川安岳人,副教授,博士,CCF高级会员,主要研究方向:数据挖掘;　王会军(1988-),男,山东潍坊人,硕士研究生,主要研究方向:金融风险;　吴江(1980-),男,浙江衢州人,副教授,博士,主要研究方向:数据库与知识工程;　张智林(1980-),男,湖南武陵人,博士,主要研究方向:稀疏贝叶斯学习;　唐常杰(1946-),男,重庆人,教授,博士生导师,主要研究方向:数据库与知识工程。

习方法(比如 Lasso、Basis Pursuit),SBL 具有明显的优势[12-13]:1)在无噪声情况下,除非满足特定的条件,L1 算法的最优解并不是真正的最稀疏解。因此,当真实解是最稀疏解的应用场合,SBL 是最佳选择。2)当感知矩阵的列与列之间相关性很强时,L1 算法的性能非常差。但在这种情况下,SBL 仍然能获得良好的解。3)已经有研究表明,SBL 等价于一种迭代加权 L1 最小化算法,L1 算法仅仅是其第一步,因此,SBL 完全可以获得比 L1 更优的稀疏解。

稀疏学习/压缩感知的一般模型可描述为:

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{x} + \boldsymbol{v} \tag{1}$$

其中:$\boldsymbol{D}$ 为 $N \times M$ 的感知矩阵,$\boldsymbol{y}$ 为 $N \times 1$ 维压缩信号,$\boldsymbol{x}$ 为 M 维待求解向量,$\boldsymbol{v}$ 是噪声。为了得到稀疏的 $\boldsymbol{x}$,SBL 假设 $\boldsymbol{x}$ 中的每个元素都服从一个参数化的均值为 0,方差为 γ_i 的高斯分布:

$$p(x_i;\gamma_i) = N(0,\gamma_i);\quad i = 1,2,\cdots,M \tag{2}$$

其中:x_i 是 $\boldsymbol{x}$ 中的第 i 个元素,γ_i 是未知参数,其值将由算法自动估计出来。在算法的运行过程中,部分 γ_i 变成 0 或趋于 0,SBL 通常将小于某个阈值的 γ_i 置为 0,此时对应的 x_i 也为 0,以此达到稀疏的效果。在有噪声的情况下,通常假设 $\boldsymbol{v}$ 为方差为 λ 的高斯白噪声向量,即:

$$p(\boldsymbol{v};\lambda) = N(0,\lambda I) \tag{3}$$

根据贝叶斯规则很容易获得噪声的后验分布,也为一高斯分布。当所有的未知参数 γ_i 和 λ 都被估计出来后,$\boldsymbol{x}$ 的最大后验估计由这个高斯分布的均值给出,而这些未知参数由第二类最大似然估计[12] 获得。

2　基于稀疏贝叶斯学习的个人信用评估

SBL 本身的模型就是一个稀疏线性回归模型,在求得式(1)中的解向量(回归系数) $\boldsymbol{x}$ 后,对于一个测试样本 T,可以计算其观测值(自变量或属性值) $\boldsymbol{A} = [a_1,a_2,\cdots,a_M]$ 与压缩信号(因变量) r 之间关系的表达式,如式(4):

$$r = \boldsymbol{A}\boldsymbol{x} = \sum_{i=1}^{M} a_i x_i \tag{4}$$

其中:a_i 是测试样本的第 i 个属性值;r 是一个实数,为了将 SBL 用于分类问题,必须将其映射为类标签,在本文,分别采用 1,-1 表示正、负类标签,将正数和 0 的 r 映射为 1,负数映射为 -1,即:

$$Label(T) = sign(r) = sign\left(\sum_{i=1}^{M} a_i x_i\right) \tag{5}$$

其中 $sign(r)$ 表示取数值 r 的符号,即:

$$sign(r) = \begin{cases} 1, & r \geqslant 0 \\ -1, & r < 0 \end{cases} \tag{6}$$

算法 1　基于稀疏贝叶斯学习的个人信用评估(SBLCredit)。

输入　N 个训练样本构成的训练集 $\{\boldsymbol{D},\boldsymbol{y}\}$,其中:$\boldsymbol{D} = [\boldsymbol{D}_1,\boldsymbol{D}_2,\cdots,\boldsymbol{D}_N]^{\mathrm{T}} \in \mathbf{R}^{N\times M}$ 为训练样本属性构成的矩阵,$\boldsymbol{y} = [y_1,y_2,\cdots,y_N]^{\mathrm{T}} \in \{-1,1\}^{N\times 1}$ 为训练样本的类标签构成的向量;测试样本 T 的属性构成的向量 $\boldsymbol{A} = [a_1,a_2,\cdots,a_M]$。

输出　测试样本的类标签 $Label(T)$。

步骤如下:

1)对训练样本属性矩阵 $\boldsymbol{D}$ 按列进行归一化;

2)根据式(1) ~ (3)求解,得到最优解向量 $\boldsymbol{x}$;

3)根据式(4) ~ (6)计算测试样本类标签 $Label(T)$。

由于在信用评估数据中,某些属性的值与其他值相比,数值相差太大,会造成分类性能下降,因此,对训练样本 $\boldsymbol{D}$ 中的数据按列进行归一化。

需要特别指出的是,在稀疏信号恢复和压缩感知领域中,绝大多数稀疏算法都要求模型(1)中的矩阵 $\boldsymbol{D}$ 满足 $N < M$。但是 SBL 并无这种要求,它在 $N < M$,或者 $N \geqslant M$ 的情况下均可使用。在个人信用评估问题中,$N > M$(即训练样本个数一般大于样本的特征数),这种情况下 SBL 仍然能得到满意的效果。另外,本文提出的 SBLCredit 方法也不同于相关向量机(Relevance Vector Machine, RVM)[14],RVM 是一种典型的核方法,而 SBLCredit 并未采用核策略。

3　实验结果及分析

3.1　实验环境

本文采用加州大学欧文分校(UCI)提供的机器学习公开数据集中[15]的德国信用数据集和澳大利亚信用数据集对本文方法进行验证。

德国信用数据集中,采用"german. data-numeric"文件内的数据,该数据集被广泛用于个人信用评估方法的验证。包含 1 000 个样本,其中包括 700 条无违约记录的正样本和 300 条有违约记录的负样本,每个样本包括 24 个属性和一个表示正或负的类标签。前 20 个属性涉及用户的基本信息、工作信息、财产信息、信用记录等[7],后 4 个属性未给出确切含义,如表 1 所示。

澳大利亚信用数据集包括 690 个样本,其中正、负样本分别为 307 个和 383 个,每个样本有 14 个属性,其中 8 个属性具有离散值,6 个属性具有连续值,由于涉及到保密,该数据集并未给出每个属性的具体含义,仅用符号表示。其中 37 个样本部分属性具有缺失值,对于离散型的缺失值,用同类别样本中该值最频繁的值替代;对于连续型的缺失值,用同类样本在该值上的平均值替代。

为了减少较大属性值对分类效果的影响,本文对所有数据进行了预处理,将所有属性值线性映射到[0,1]区间。

实验平台为:Pentium 双核 2.60 GHz CPU,4 GB 内存,Windows 7,Matlab 2012。为了评估 SBLCredit 的性能,与 *K*NN、Naïve Bayes、决策树和 SVM 进行比较。其中:*K*NN 算法中的 K 取 10,决策树采用 C4.5 算法,SVM 采用径向基函数(Radius Basis Function, RBF)核函数。

3.2　实验结果

k-折交叉验证是基于给定数据随机选样划分的,是常用的评估分类方法准确率的技术。因此,为了更准确地评估 SBLCredit 算法的准确率,把每个数据集分成 5 个互不相交的子集,把这 5 个子集按 4:1 的比例组成训练集和测试集,进行 5-折交叉验证,再将分类精度与传统的 *K*NN、Naïve Bayes、决策树和 SVM 进行比较。

在德国信用数据集上的结果如图 1 所示。

实验结果表明,在德国信用数据集上,相对于传统分类方法,SBLCredit 算法表现出了更高的分类精度,比 *K*NN、Naïve Bayes、C4.5 和 SVM 的分类精度分别提高了 5.52%,4.14%,6.68% 和 3.26%。同时,该数据集中的 24 个特征在 SBLCredit 分类中的重要程度如图 2 所示,可见,24 个属性中的第 8,13,21,23,24 个属性对分类几乎没有贡献,而第 1,2,3,15,18,19 个属性对分类非常重要,由此可以看出,SBL 能进行有效的特征选择。

表 1 德国信用数据库中的客户属性

属性编号	属性名	属性值(离散化)
1	经常账户状况	1:账户余额 < 0 马克;2:0 马克≤账户余额 < 200 马克;3:≥200 马克;4:无经常账户记录
2	账户持续时间(月)	—
3	贷款历史状况	1:无贷款记录或所有贷款均按时返还;2:在本银行的所有贷款均按时返还;3:迄今为止现存贷款按时返还;4:过去曾延迟还款;5:存在危帐或仍存在贷款(非本银行)
4	贷款用途	1:新车;2:二手车;3:家具设备;4:收音机或电视机;5:家庭用品;6:维修;7:教育;8:度假;9:接受再培训;10:经商;11:其他用途
5	贷款数额	—
6	储蓄存款账户状况	1:账户余额 < 100 马克;2:100 马克≤账户余额 < 500 马克;3:500 马克≤账户余额 < 1 000 马克;4:账户余额≥1 000 马克;5:未知或无储蓄存款
7	现工作就业时间	1:失业;2: < 1 年;3:[1 年,4 年);4:[4 年,7 年);5:≥7 年
8	分期付款占月收入百分比	—
9	个人状况及性别	1:男性离异或分居;2:女性离异、分居或结婚;3:男性单身;4:男性结婚或鳏居;5:女性单身
10	其他债务或保证金	1:无;2:联合申请人;3:保证人
11	现居住状况	—
12	财产状况	1:有房产不动产;2:无房产不动产,有社保储蓄协议或养老保险;3:无房产不动产,无社保或养老保险,有汽车或其他(不在属性 6 范围内);4:未知或无财产
13	年龄	—
14	其他分期付款计划	1:银行;2:商店;3:无
15	房屋状况	1:租住;2:自有;3:免费使用
16	在本银行已有存款数目	—
17	工作状况	1:失业、无技能或非本地居民;2:无技能的本地居民;3:技术工人或公务员;4:经理、自由职业者、高级雇员或官员
18	应抚养人数	—
19	电话	1:无;2:有或已注册
20	是否外籍劳工	1:是;2:否
21 ~ 24	未知	0 或 1

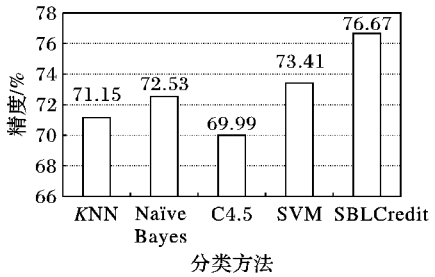

图 1 德国信用数据集分类精度对比

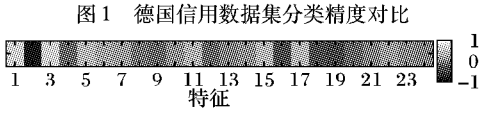

图 2 SBLCredit 在德国信用数据集的特征选择效果

在澳大利亚信用数据集上的结果如图 3 所示。

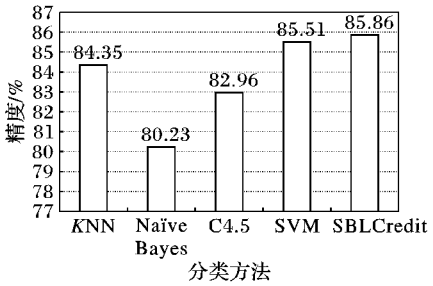

图 3 澳大利亚信用数据集分类精度对比

实验结果表明,在澳大利亚信用数据集上,相对于传统分类方法,SBLCredit 算法仍然表现出了更好的分类效果。较 *K*NN、Naïve Bayes、C4.5 和 SVM 的分类精度分别提高了 1.51%,5.63%,2.90% 和 0.35%。其特征选择效果如图 4 所示,可见,第 1,4,11 个属性对分类的贡献非常小。因为该数据集总的属性只有 14 个,因此,特征选择效果没有在德国信用数据集上那么明显。

图 4 SBLCredit 在澳大利亚信用数据集的特征选择效果

以上实验表明,SBLCredit 是一种有效的信用评估算法,除了能得到较好的分类精度外,还能进行有效的特征选择。因为 SBLCredit 选出的特征较稀疏,所以增强了特征的可解释性。

4 结语

信用评估对于银行业务具有十分重要的意义,传统的信用评估方法存在一些缺陷。本文将稀疏贝叶斯学习引入到信用评估中,取得了较传统信用评估方法更好的分类效果,同时,该方法还能有效地进行特征选择。下一步将研究 SBLCredit 与其他方法的结合使用,比如研究 SBLCredit 的组合分类器模型等。

参考文献:

[1] YAP B W, ONG S H, HUSAIN N H M. Using data mining to improve assessment of credit worthiness via credit scoring models [J]. Expert Systems with Applications, 2011, 38(10): 13274 – 13283.

[2] BAESENS B, van GESTEL T, VIAENE S, *et al*. Benchmarking state-of-the-art classification algorithms for credit scoring [J]. Journal of the Operational Research Society, 2003, 54(6): 627 – 635.

[3] 姜明辉, 王雅林, 赵欣, 等. *k*-近邻判别分析法在个人信用评估中的应用[J]. 数量经济技术经济研究, 2004, 21(2): 143 – 147.

[4] HUANG C L, CHEN M C, WANG C J. Credit scoring with a data mining approach based on support vector machines [J]. Expert Systems with Applications, 2007, 33(4): 847 – 856.

[5] LESSMANN S, STAHLBOCK R, CRONE S F. Genetic algorithms for support vector machine model selection [C]// Proceedings of 2006 International Joint Conference on the Neural Networks. Washington, DC: IEEE Computer Society, 2006: 3063 – 3069.

[6] WANG Y, WANG S, LAI K. A new fuzzy support vector machine to evaluate credit risk [J]. IEEE Transactions on Fuzzy Systems, 2005, 13(6): 820 – 831.

[7] 吴江, 唐常杰, 段磊, 等. 基于基因表达式编程的信用评估模型挖掘方法[J]. 计算机应用, 2007, 27(4): 877 – 880.

(下转第 3148 页)

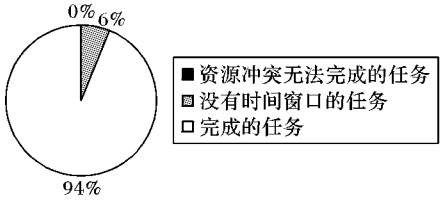

图 5 侧摆角度为 10°时任务完成情况

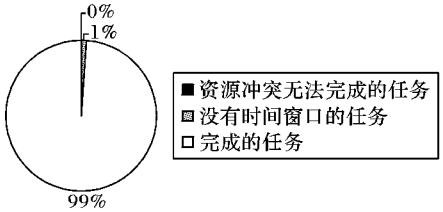

图 6 侧摆角度为 25°时任务完成情况

从表 1 和图 4 ~ 6 可以看到,载荷通过侧摆,能明显地增加目标的完成数量,从而提高对应调度方案的总收益值,理论上而言,载荷侧摆能力越强,所完成的目标点的数目越多,对应的收益值越高,但是由于载荷侧摆需要能量,而且载荷侧摆的能力对设备要求比较高,因而载荷侧摆角度比较大时,所需要的成本花费比较高。同时,结果显示,当载荷侧摆达到一定的角度后,提高载荷侧摆的能力对最后的总收益值影响不会很大。

5 结语

为充分利用各种资源,最大限度地发挥资源的价值,获取最大的成像效益,本文以最大化成像的综合效益为第一原则、最小化侧摆次数为第二原则、最小化总的侧摆角度为第三原则,建立了带侧摆的多星点目标调度模型;并基于演化算法,提出了一种载荷侧摆情况下多星点目标调度算法。本文重点设计了演化算法的种群初始化、交叉算子、变异算子、个体评价、冲突减少算子、选择算子以及冲突消除算子。最后,本文给出了一个具体的测试实例,给出了 5 星 100 个点目标在侧摆情况下的调度及仿真结果,并对侧摆角度分别为 0°、10°、25°时的调度性能进行了分析。

参考文献:

[1] 戴光明,王茂才. 多目标优化算法及在卫星星座设计中的应用[M]. 武汉:中国地质大学出版社,2009:1 - 116.

[2] LEMAITRE M, VERFAILLIE G. Daily management of an earth observing satellite: comparison of ILOG solver with dedicated algorithms for valued constraint satisfaction problems [C]// Proceedings of the 3rd ILOG International Users Meeting. Paris: [s. n.], 1997: 1 - 9.

[3] PEMBERTON J. Towards scheduling over-constrained remote sensing satellites [C]// Proceedings of the 2nd NASA International Workshop on Planning and Scheduling for Space. San Francisco: Space Telescope Science Institute, 2000: 1 - 13.

[4] BENSANA E, VERFAILLIE G, LEMAITRE M. Earth observing satellite management [J]. Constraints, 1999, 4(3): 293 - 299

[5] FRANK J, JONSSON A, MORRIS R, *et al.* Planning and scheduling for fleets of earth observing satellites [C]// Proceedings of the 6th International Symposium on Artificial Intelligence, Robotics and Automation in Space. Montreal: The Press of Carnegie Mellon University, 2002: 1 - 8.

[6] DUNGAN J, FRANK J, JONSSON A, *et al.* Advances in planning and scheduling of remote sensing instruments for fleets of earth orbiting satellites [EB/OL]. [2013-02-23]. http://www.isprs.org/proceedings/XXXIV/part1/paper/00001.pdf.

[7] CHIEN S, CICHY B, DAVIES A, *et al.* An autonomous earth-observing sensorweb [J]. IEEE Intelligent Systems, 2005, 20(3): 16 - 24.

[8] ABRAMSON M, CARTER D, KOLITZ S, *et al.* Real-time optimized earth observation autonomous planning [EB/OL]. [2013-02-22]. http://www.esto.nasa.gov/conferences/estc-2002/Papers/A5P1(Abramson).pdf.

[9] BIANCHESSI N, RIGHINI G. Planning and scheduling algorithms for the COSMO-SkyMed constellation [J]. Aerospace Science and Technology, 2008, 12(7): 535 - 544.

[10] FLORIO S D. Performances optimization of remote sensing satellite constellations: a heuristic method [EB/OL]. [2013-03-02]. http://www.stsci.edu/largefiles/iwpss/20069151043Paper.pdf.

[11] 贺仁杰,高鹏,白保存,等. 成像卫星任务规划模型、算法及其应用[J]. 系统工程理论与实践,2011,31(3):411 - 422.

[12] 李菊芳,白保存,陈英武,等. 多星成像调度问题基于分解的优化算法[J]. 系统工程理论与实践,2009,29(8):134 - 143.

[13] 王钧,李军,陈健,等. 多目标 EOSs 联合成像调度方法[J]. 宇航学报,2007,28(2):354 - 359.

[14] 白保存,徐一帆,贺仁杰,等. 卫星合成观测调度的最大覆盖模型及算法研究[J]. 系统工程学报,2010,25(5):651 - 658.

[15] 樊鹏山,熊伟,李智. 载荷侧摆情况下卫星覆盖区域计算方法研究[C]// 系统仿真技术及其应用学术会议论文集. 合肥:中国科学技术大学出版社,2009:536 - 540.

[16] 宋志明. 面向区域目标的卫星调度问题的研究与仿真[D]. 武汉:中国地质大学,2011.

(上接第 3096 页)

[8] BERLANGA F J, RIVERA A J, del JESUS M J, *et al.* GP-COACH: Genetic Programming-based learning of COmpact and ACcurate fuzzy rule-based classification systems for high-dimensional problems [J]. Information Sciences, 2010, 180(8): 1183 - 1200.

[9] YU L, YUE W, WANG S, *et al.* Support vector machine based multiagent ensemble learning for credit risk evaluation [J]. Expert Systems with Applications, 2010, 37(2): 1351 - 1360.

[10] HSIEH N C, HUNG L P. A data driven ensemble classifier for credit scoring analysis [J]. Expert Systems with Applications, 2010, 37(1): 534 - 545.

[11] WANG G, MA J, HUANG L H, *et al.* Two credit scoring models based on dual strategy ensemble trees [J]. Knowledge-Based Systems, 2012, 26: 61 - 68.

[12] WAN J, ZHANG Z, YAN J, *et al.* Sparse Bayesian multi-task learning for predicting cognitive outcomes from neuroimaging measures in Alzheimer's disease [C]// Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Science, 2012: 940 - 947.

[13] WIPF D P, RAO B D. Sparse Bayesian learning for basis selection [J]. IEEE Transactions on Signal Processing, 2004, 52(8): 2153 - 2164.

[14] TIPPING M E. Sparse Bayesian learning and the relevance vector machine [J]. The Journal of Machine Learning Research, 2001, 1: 211 - 244.

[15] UCI Machine Learning Repository [EB/OL]. [2013-05-07]. http://archive.ics.uci.edu/ml/datasets.html.