

Personal Credit Risk Assessment Based on Rough Set Attribute Reduction and Multiple Classification Models

Ning Cao, Shujin Li

College of Economics, Hangzhou Dianzi University, Hangzhou Zhejiang

Email: relmaning@163.com, jslsj@163.com

Received: May 12th, 2018; accepted: May 29th, 2018; published: Jun. 5th, 2018

Abstract

Personal credit is one of the most important businesses of commercial banks, and there is information asymmetry in this business, and banks that are in information disadvantaged face enormous credit risks. "Knowledge is the ability to classify objects", based on this view, this article proposes that "personal credit assessment is a method of classifying the borrower's credit". SMOTE has improved the unbalanced category of German credit data sets. Boolean reasoning approaches objectively discretize some continuous indicators, overcoming the subjectivity of artificial discretization. Based on the rough set genetic attribute reduction algorithm, the evaluation index of German credit data set has a significant reduction effect. The index is reduced from 20 to 10 and the reduction rate is as high as 50%, which realizes that the performance of the classifier is approximately constant, improving the interpretability of the classifier, alleviating overfitting, enhancing the generalization capability of the classifier, and greatly reducing the time-consuming training of the classifier. After attribute reduction, the performance of the C4.5 decision tree has improved. Logistic regression is the optimal model for attribute reduction of German credit data sets.

Keywords

Personal Credit Assessment, Data Mining, Rough Set, Attribute Reduction

基于粗糙集属性约简与多种分类模型的个人信用评估

曹 宁, 李淑锦

杭州电子科技大学, 经济学院, 浙江 杭州

Email: relmaning@163.com, jslsj@163.com

摘要

个人信贷是商业银行最重要的业务之一, 而该业务存在着信息不对称, 处于信息劣势方的银行面临着巨大的信用风险。“知识就是一种对对象进行分类的能力”, 依据这一观点本文提出“个人信用评估就是一种对借款人信用的分类方法”。SMOTE改善了德国信用数据集类别不平衡的情况, 布尔逻辑离散化技术客观地对一些连续的指标离散化, 克服了人工离散化的主观性。基于粗糙集遗传属性约简算法对德国信用数据集的评估指标约简效果显著, 指标由20个约简为10个, 约简率高达50%, 实现了在分类器性能近似不变的前提下, 提高分类器的可解释性、缓解了过拟合、增强了分类器的泛化能力, 并且大大缩短了训练分类器的耗时。经过属性约简, C4.5决策树的性能有所提高。对于经过属性约简的德国信用数据集, Logistic回归是最优的模型。

关键词

个人信用评估, 数据挖掘, 粗糙集, 属性约简

Copyright © 2018 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 研究背景

个人信贷是商业银行最重要的业务之一, 而该业务存在着信息不对称, 处于信息劣势方的银行面临着巨大的信用风险。关于什么是“信用风险”, 巴塞尔委员会认为信用风险是债务人未能履约, 从而对经济主体未来活动产生负面影响的风险。对于商业银行而言, 信用风险是债务人未能履约, 使得商业银行遭受损失的可能性。

产生个人信用风险的主要原因是借款人与贷款人(商业银行)之间的信息不对称。在资金的用途、还款能力等方面, 借款人有着绝对的信息优势, 贷款人若想了解与监督借款人的全部信息需要付出极大的成本, 这甚至是不可能的。显然, 商业银行对借款人的信息的掌握是有限的, 如果商业银行能够根据有限的信息, 通过某种方法预测每一位借款人的信用, 为商业银行的授信决策提供相对准确的、客观的依据, 那么将大大地提升商业银行的竞争力。人工智能领域普遍认为“知识就是一种对对象进行分类的能力”, 依据这一观点本文提出“个人信用评估就是一种对借款人信用的分类方法”。数据挖掘领域有许多分类模型, 其可以根据借款人的信息数据集通过训练得到个人信用分类器, 以区分某客户的信用是“好的”(即能够履行协议)或“坏的”(即违约)。如今大数据与人工智能时代的来临为商业银行对个人信用的评估开启了新思路。商业银行可以利用自身巨大体量的数据和先进的数据挖掘技术有效地降低在个人信贷领域因信息不对称带来的风险、提升自身盈利能力。

数据挖掘领域的“分类模型”是一种能够找出区分数据类别的模型, 人们能够基于分类模型将类别未知的对象划分到已知类别的集合中, 从而确认该对象的类别[1]。分类模型使得构建一套高效的、准确的、稳定的与客观的个人信用评估器成为可能。上世纪八九十年代, 波兰科学家 Z. Pawlak 发明了粗糙集理论。利用粗糙集属性约简算法对数据集中的个人信用评估指标进行处理, 可以使得个人信用分类器在

分类预测能力近似不变的前提下, 减少数据集中不必要的评估指标, 这能够提高分类器的可解释性、泛化能力, 并且大大缩短分类器的训练时间。

Brenda Mak 和 Toshinori Munakata (2002)利用神经网络和决策树结合粗糙集对信用进行了建模与分析[2]。J. Chen 等(2007)指出对个人信用影响最大的评估指标为收入, 其次为受教育水平和年龄, 而国籍、居住地区和婚姻情况的影响非常小[3]。Yao Ping 和 Lu Yongheng (2011)先利用粗糙集将属性约简, 然后利用支持向量机进行建模并与线性判别分析、Logistic 回归和神经网络等模型进行比较, 得出粗糙集与支持向量机结合的混合分类器拥有最好的信用评估能力[4]。杜婷(2012)首先对评估指标进行了约简, 然后利用支持向量机将个人信用分类, 大大地缩短了分类器的训练耗时, 且平均分类正确率在选取合适的约简率时会有明显的提升[5]。胡来丰(2015)针对多种粗糙集属性约简算法进行了详细的介绍并提出改进。经过评估指标约简的 BP 神经网络的分类正确率相对于约简之前略有下降, 但评估指标由 20 个约简为 8 个, 使得神经网络训练耗时显著缩短[6]。杨蕴涵(2015)运用 Logistic 回归、C5.0 决策树和随机森林模型, 对 UCI 数据库中“德国信用数据”进行了学习、预测, 通过对比得出“三种模型中没有任何一种模型具有绝对优势”[7]。陈慧(2016)利用 C4.5 决策树、朴素贝叶斯、支持向量机和 BP 神经网络对多个个人信用数据集进行了评估, 其结果显示, 对于单一分类模型, 非统计模型的分类精度优于统计模型, 但统计模型的稳定性与可解释性均更优秀; 对于集成分类模型, 其性能优于以上任一单一模型[8]。

本文首先采用合成少数类过采样技术与布尔逻辑离散化技术对德国信用数据集进行类别平衡与离散化处理, 其次利用遗传属性约简算法对德国信用数据集中的评估指标进行约简, 然后基于 Logistic 回归、C4.5 决策树、BP 神经网络和支持向量机模型训练个人信用分类器, 最后对各种分类器的性能进行比较、评价与分析。本文的创新点是通过将粗糙集属性约简与数据挖掘分类模型相结合, 使得在分类器的性能近似不变的前提下, 减少数据集中不必要的评估指标, 旨在提供一种个人信用评估的新思路。

2. 信用风险评估方法

2.1. Logistic 回归

形如: $f(x) = \omega^T * x + b$, 其中 $\omega = (\omega_1, \omega_2, \dots, \omega_m)^T$, $x = (x_1, x_2, \dots, x_m)^T$ 称为线性回归模型。该模型可以依据每个样本中的评估指标的线性组合, 将该样本的信用情况映射为一个实数, 然而信用情况的取值在实数范围, 这并没有为信用的好或坏提供一个明确的“界线”。因此需要对线性回归模型因变量的值域进行“压缩”, 即再映射到区间[1,2], 本文中采用的德国信用数据集将信用好定义为 1, 信用坏定义为 2。因此, 当某个样本的因变量的取值小于 1.5 时, 该样本的信用情况为好, 反之为坏。形如: $f(x) = g(\omega^T * x + b)$, 其中 $g(\cdot)$ 称作联系函数称为广义线性回归模型。Logistic 回归模型是广义线性回归模型的一个特化, 常用 Sigmoid 函数 $f(x) = \frac{1}{1 + e^{-(\omega^T * x + b)}}$ 作为联系函数[9]。

2.2. C4.5 决策树

“树”是一种抽象的数学概念, 通过递归可以构建一棵树, 树的每个结点都可以存储数据。对于德国信用数据集, 最初将数据集存储在根节点上, 若通过某种“分裂准则”指定某个评估指标为“分裂特征”, 使得基于这个特征分裂构建出的子树结点所存储的数据子集更加的“纯”, 这样递归地生成的叶子结点就代表了信用类别。对于新样本, 可以从根节点起沿着各个分支结点所代表的评估指标找到一个叶子结点, 这样就实现了对新样本信用的分类。建立决策树的关键是如何确立分裂准则和分裂程度。C4.5 分裂准则采用了一种启发式算法来选择候选分裂特征中信息增益高于平均信息增益的特征中拥有最大信息增益率的评估指标作为分裂特征[1]。

2.3. BP 神经网络

人工神经网络是一种仿照生物神经网络构造的具有并行性和适应性的网络, 它由众多人工神经元互连组成。在 M-P 神经元模型中, 评估指标信息是通过带权重的连接进行传递的, 某个神经元接收来自其他神经元传递过来的带权重的评估指标信息作为输入, 该神经元将输入信息进行某种处理, 然后将结果与阈值进行比较, 最后通过激活函数的处理以产生某种评估指标信息输出。对于人工神经网络, 隐层是输入层与输出层之间层级的统称, 输入层神经元没有激活函数, 而隐层神经元与输出层神经元均具有激活函数。多层前馈神经网络(Multilayer Feedforward Neural Network)每层的各个神经元与其下一层的各个神经元互连, 同层神经元之间的连接和跨层连接是不被允许的。如果每个隐层和输出层神经元代表一个评估指标, 那么对个人信用评估 BP 神经网络的训练就是根据训练集数据和某种准则不断地调整神经元之间的权值以及阈值的过程[9]。

2.4. 支持向量机

对于德国信用数据集中的任一样本, 其几何本质为 m 维空间中的一点。支持向量机试图寻找一个超平面 $w^T x + b = 0$, $w = (w_1, w_2, \dots, w_m)^T$, 其能够将 m 维空间划分成两部分, 使得信用好的样本和信用坏的样本按类别分散在超平面的两侧。优良的超平面指的是既能够将 m 维空间中的样本信用情况尽可能正确地划分, 又使得各个样本点到超平面的距离尽可能的远[9]。

3. 实证研究

3.1. 数据介绍

德国信用数据(German Credit Data), 来源于美国的加州大学欧文分校(University of California, Irvine)所建立 UCI Machine Learning Repository。数据集包含样本(实例)数目 1000 个, 其中信用为“好”的样本有 700 个, 信用为“坏”的样本有 300 个。数据集完整, 但类别不平衡, 包含 20 个评估指标和 1 个决策属性, 其中 3 个为连续型(F2、F5、和 F13), 17 个为离散型, 详情见表 1。

3.2. 合成少数类过采样技术

鉴于德国信用数据集的类别不平衡现状, 本文采用合成少数类过采样技术(Synthetic Minority Over-sampling Technique, SMOTE)对少数类进行过采样处理, 以使得数据集类别平衡。SMOTE 是一种基于“插值”的为少数类合成新样本的技术。利用合成少数类过采样技术对德国信用数据集进行类别平衡(算法参数 $N = 133.5 \times 300$ 和 $k = 5$), 使得处理后的数据集具有信用好的样本 700 个, 信用坏的样本 700 个。

3.3. 布尔逻辑离散化技术

本文的研究主题是基于粗糙集属性约简, 是否能够在分类器的性能近似不变的前提下, 减少数据集的评估指标。粗糙集属性约简算法是对决策表进行操作的, 它要求决策表中的值必须为离散的。然而, 德国信用数据集中有 3 个评估指标的值是连续的, 这就要求在属性约简前必须先将这 3 个评估指标的值离散化。对于德国信用数据集中值为连续的评估指标, 如 F2 (支票账户持续时间)、F5 (贷款金额)和 F13 (年龄), 将持续时间或贷款金额人工的划分为几个整数区间未免过于主观, 因此本文将采用布尔逻辑离散化技术, 并对比基于不同的离散化处理对分类器性能的影响。Nguyen H.S.和 Skowron 发明的布尔逻辑离散化是一种“客观的”离散化技术, 这是一种启发式算法, 它遵循贪心原则根据断点值的重要性依次把断点加入到断点集中[10]。德国信用数据集的离散化结果见表 2。

Table 1. German credit data**表 1.** 德国信用数据

评估指标	符号	类型	值域
现有支票账户情况	F1	离散	F11 F12 F13 F14
支票账户持续时间	F2	连续	[4, 72]
信用历史记录	F3	离散	F30 F31 F32 F33 F34
贷款目的	F4	离散	F41 F42 F43 F44 F45 F46 F47 F48 F49 F410
贷款金额	F5	连续	[250, 18424]
储蓄账户/债券情况	F6	离散	F61 F62 F63 F64 F65
在职时间	F7	离散	F71 F72 F73 F74 F75
分期付款占可支配收入的比率	F8	离散	F81 F82 F83 F84
性别与婚姻状况	F9	离散	F91 F92 F93 F94 F95
债务人/担保人	F10	离散	F101 F102 F103
在现居住地居住时间	F11	离散	F111 F112 F113 F114
资产情况	F12	离散	F121 F122 F123 F124
年龄	F13	连续	[19, 75]
拥有其他信贷情况	F14	离散	F141 F142 F143
住房情况	F15	离散	F151 F152 F153
在本银行的贷款次数	F16	离散	F161 F162
工作情况	F17	离散	F171 F172 F173 F174
法定抚养人数	F18	离散	F181 F182
电话	F19	离散	F191 F192
是否为外国人	F20	离散	F201 F202
信用情况	D	离散	D1 (好) D2 (坏)

Table 2. Discretization of German credit data sets**表 2.** 德国信用数据集的离散化

评估指标	原始连续取值区间	人工离散化取值区间	布尔逻辑离散化取值区间
F2 (支票账户持续时间)	[4, 72]	[0, 6], (6, 12], (12, 18], (18, 24], (24, *)	[0, 21), [21, *)
F5 (贷款金额)	[250, 18424]	[0, 1000], (1000, 5000], (5000, 10000), (10000, *)	[0, 1286), [1286, 2067), [2067, 4152), [4152, *)
F13 (年龄)	[19, 75]	[18, 35), [35, 60), [60, *)	[18, 27), [27, 33), [33, 42), [42, *)

3.4. 德国信用数据集的最小属性约简

交叉概率 $P_c = 0.8$, 变异概率 $P_m = 0.05$, 最大繁殖代数 $g_{\max} = 100$ 万代的遗传属性约简算法得到的德国信用数据集的多个最小属性约简(最小评估指标集), 评估指标由原来的 20 个约简为 10 个, 约简率达 50% [11]。下面给出的最小评估指标集经过测试, 是使得分类器性能下降最少的。德国信用数据基于遗传约简算法的最小评估指标集见表 3。

Table 3. German credit data is based on the minimum evaluation index set of the genetic reduction algorithm**表 3.** 德国信用数据基于遗传约简算法的最小评估指标集

符号	评估指标
F1	现有支票账户情况
F3	信用历史记录
F4	贷款目的
F5	贷款金额
F7	在职时间
F8	分期付款占可支配收入的比率
F9	性别与婚姻状况
F11	在现居住地居住时间
F12	资产情况
F13	年龄

3.5. 个人信用风险评估结果及其对比分析

本文采用的建模软件是由新西兰的怀卡托大学(The University of Waikato)开发的怀卡托智能分析环境(Waikato Environment for Knowledge Analysis, WEKA)。训练上述分类器的训练集与验证集的划分采用 10 次 10 折交叉验证。在本文使用的各种技术中,有些技术只能处理离散的数据,因此本文使用的第一个数据集(以下简称“1”)是经过布尔逻辑离散化处理的数据集;第二个数据集(以下简称“2”)是先经过主观的人工离散化,再进行 SMOTE 类别平衡处理的数据集;第三个数据集(以下简称“3”)是先经过布尔逻辑离散化,再经过 SMOTE 类别平衡处理的数据集;第四个数据集(以下简称“4”)是在 3 的基础上对评估指标进行了约简(表 4~7)。

通过比较上述四张表中的 1、3 可以得出,基于 SMOTE 对德国信用数据的类别平衡处理能够提升分类器的分类精度、PRC Area 和 AUC,且大大地降低了第二类错误率,这是符合预期的,因为原数据集信用好的样本远远多于信用坏的样本,原数据集训练出的分类器会以非常大的概率把真实坏的样本也分类预测为好的样本,即第二类错误率非常高。这样的分类器即使其总分类精度再高,也是没有任何意义的。SMOTE 提升了所有的评价指标,这印证了“Garbage in, garbage out。”。

基于 3 训练的分类器的性能显著高于基于 2 训练的,这说明布尔逻辑离散化技术较人工离散化对分类器的性能是有所提升的,因此后续的评估指标约简应在 3 的基础上进行。

基于遗传约简算法得到了德国信用数据集的多个最小评估指标集,这里挑选了能够使得分类器性能下降最少的那个最小评估指标集作为代表,即数据集 4。不幸的是,基于 4 训练的各种分类器的性能大多所下降,但下降幅度很小,如 C4.5 决策树分类器的总分类精度仅仅下降了 0.1%,见表 5,性能下降幅度最大的 BP 神经网络分类器的总分类精度下降了 4.5%,见表 6。

令人欣慰的是,德国信用数据集由最初的 20 个评估指标约简为 10 个,约简率高达 50%,这使得分类器的训练耗时大幅度的缩短,在这里本文并没有给出实验中建模的时长,这是没有必要的,因为本文并没有针对各种模型理论进行改进,也就是说模型的时间复杂度和空间复杂度并没有任何改变,而且同样的实验在不同的环境下的运行的速度是不同的,单纯地比较模型的训练时长是没有意义的。

最后,本文发现经过评估指标约简的 C4.5 决策树分类器的性能在某程度上有所提升,见表 5, ROC Area 提升了 0.016, PRC Area 提升了 0.023,这从侧面反映了评估指标的减少缓解了 C4.5 决策树分类器的过拟合。

Table 4. Performance comparison of personal credit evaluation classifier based on logistic regression
表 4. 基于 Logistic 回归的个人信用评估分类器的性能比较

	总分类精度(%)	第二类错误率(%)	F-Measure	ROC Area	总 PRC Area
1	75.0	49.0	0.744	0.776	0.793
2	76.2	22.4	0.762	0.841	0.832
3	76.6	22.4	0.766	0.841	0.829
4	75.1	23.7	0.751	0.824	0.813

Table 5. Performance comparison of personal credit evaluation classifier based on C4.5 decision tree
表 5. 基于 C4.5 决策树的个人信用评估分类器的性能比较

	总分类精度(%)	第二类错误率(%)	F-Measure	ROC Area	总 PRC Area
1	73.1	55.3	0.721	0.685	0.692
2	74.1	27.9	0.741	0.756	0.706
3	75.2	26.1	0.752	0.757	0.708
4	75.1	24.4	0.751	0.773	0.731

Table 6. Performance comparison of personal credit evaluation classifier based on BP neural network
表 6. 基于 BP 神经网络的个人信用评估分类器的性能比较

	总分类精度(%)	第二类错误率(%)	F-Measure	ROC Area	总 PRC Area
1	72.5	49.0	0.722	0.736	0.763
2	79.4	18.4	0.793	0.870	0.861
3	79.2	18.9	0.792	0.865	0.849
4	74.7	22.1	0.747	0.822	0.807

Table 7. Performance comparison of personal credit evaluation classifier based on support vector machine
表 7. 基于支持向量机的个人信用评估分类器的性能比较

	总分类精度(%)	第二类错误率(%)	F-Measure	ROC Area	总 PRC Area
1	74.5	48.7	0.740	0.679	0.684
2	75.5	23.7	0.755	0.755	0.693
3	77.2	22.4	0.772	0.772	0.710
4	75.0	23.1	0.750	0.750	0.688

4. 结论

个人信用风险评估问题是信用风险领域的热点问题。本文基于布尔逻辑离散化技术客观地对一些连续的指标离散化, 克服了人工离散化的主观性; 利用属性约简方法降阶, 去除冗余的信用风险评估指标从而建立有效的评估指标体系, 这是两点重要的创新, 对个人信用风险评估有很强的指导意义。利用常用的评估方法对德国的信用数据进行评估, 结果表明, 评估效果显著, 详情见以下几点。

1) 对德国信用数据集的 SMOTE 类别平衡处理是必要的。

基于类别不平衡的数据集训练得到的分类器的分类必然是不公平的, 典型的例子就是用只包含信用为好的数据集训练分类器, 这样的分类器会将之后见到的一切样本均划分到信用好的类别中去, 显然这是错误的。

2) 布尔逻辑离散化技术在某种程度上优于人工离散化。

布尔逻辑离散化技术客观地对一些连续的指标离散化, 克服了人工离散化的主观性, 经它处理的数据集能够训练出性能更高的分类器。

3) 德国信用数据集经过属性约简, 评估指标由 20 个约简为 10 个, 约简率高达 50%。

经过评估指标约简的数据集, 较未经过约简的数据集训练的分类器的性能下降幅度非常小(下降至多不超过 4.5%, 最少仅为 0.1%), 甚至还提升了 C4.5 决策树分类器的 ROC Area 与总 PRC Area。分类器性能的下降是可接受的, 因为评估指标的减少提升了模型的可解释性、缓解了 C4.5 决策树分类器的过拟合、提升了分类器的泛化能力、大幅度的缩短了分类器的训练耗时。

4) 对于数据集 4, Logistic 回归分类器是最好的。

若以 Logistic 回归分类器的性能作为基准, Logistic 回归与 C4.5 决策树分类器拥有最高的总分类精度; BP 神经网络分类器拥有最低的第二类错误率; Logistic 回归分类器的 ROC Area、总 PRC Area 最高。针对“寻找分类性能最好的分类器”这一目的, 可以说 Logistic 回归分类器是最好的。因为 Logistic 回归模型的可解释性强, 每个评估指标对分类结果的影响可以从评估指标的权重上直观地看出。Logistic 回归和 C4.5 决策树相比于支持向量机和 BP 神经网络易于理解, 且基于前两者的分类器训练耗时显著短于后两者。根据奥卡姆剃刀原则, 对德国信用数据集的属性约简是应该被接受的, Logistic 回归分类器是最好的。

最后, 本文必须强调, 针对特定的问题、特定的先验信息、特定的数据的分布、特定的训练样本的数目、特定的代价或奖励函数等而言, 不存在绝对优秀的算法。算法是否能够取得优良的效果多数情况下取决于该算法的归纳偏好是否与待解决的问题相匹配[9]。本文旨在提供粗糙集属性约简与多种分类模型相结合的思路, 期望这种方法能够在其他信用数据集和其他领域的分类任务中有所建树。

基金项目

国家社会科学基金项目(17BJY233)。

参考文献

- [1] Jiawei Han, Micheline Kamber, Jian Pei. 数据挖掘: 概念与技术[M]. 第 3 版. 北京: 机械工业出版社, 2012.
- [2] Mak, B. and Munakata, T. (2002) Rule Extraction from Expert Heuristics: A Comparative Study of Rough Sets with Neural Networks and ID3. *European Journal of Operational Research*, **136**, 212-229. [https://doi.org/10.1016/S0377-2217\(01\)00062-5](https://doi.org/10.1016/S0377-2217(01)00062-5)
- [3] Huang, C.L., Chen, M.C. and Wang, C.J. (2007) Credit Scoring with a Data Mining Approach Based on Support Vector Machines. *Expert Systems with Applications*, **33**, 847-856. <https://doi.org/10.1016/j.eswa.2006.07.007>
- [4] Ping, Y. and Lu, Y. (2011) Neighborhood Rough Set and SVM Based Hybrid Credit Scoring Classifier. *Expert Systems with Applications*, **38**, 11300-11304. <https://doi.org/10.1016/j.eswa.2011.02.179>
- [5] 杜婷. 基于粗糙集支持向量机的个人信用评估模型[J]. 统计与决策, 2012(1): 94-96.
- [6] 胡来丰. 基于粗糙集 BP 神经网络个人信用评估模型[D]: [硕士学位论文]. 成都: 电子科技大学, 2015.
- [7] 杨蕴涵. 多种分类模型在个人信用评估中的应用[D]: [硕士学位论文]. 重庆: 重庆大学, 2015.
- [8] 陈慧. 基于数据挖掘的个人信用风险评估单一模型与集成模型的研究[D]: [硕士学位论文]. 南宁: 广西大学, 2016.
- [9] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
- [10] 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.
- [11] 颜艳, 杨慧中. 基于遗传算法的粗糙集属性约简算法[J]. 计算机工程与应用, 2007, 43(31): 156-158.

知网检索的两种方式：

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择：[ISSN]，输入期刊 ISSN：2161-0967，即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：<http://www.hanspub.org/Submission.aspx>

期刊邮箱：fin@hanspub.org