

# 基于核的人人贷的信贷风险建模

高 见

( 华中科技大学, 湖北武汉 430074 )

摘 要：人人贷作为一种新型借贷，近年来得到广泛的关注。但现有的理论研究都是对人人贷进行定性研究，风险评估作为人人贷的重要部分仍沿用传统借贷的风险评估方法，但由于人人贷的无抵押性及投资者的非专业性，传统借贷中的方法并不适用。本文提出针对人人贷的基于核的信贷风险评估模型，有效地预测贷款的风险和收益率。

关键词：人人贷；信贷风险；核回归

中图分类号：TP309 文献标识码：A 文章编号：1003-9767 ( 2013 ) 01-0007-02

## 1. 引言

人人贷，即个人对个人的借贷。贷款者提交申请，各投资者对其投标，标明投资额度。当贷款额度等于所有标的额度之和，借贷过程开始，贷款者分期还款。投资者通常将资金分散到不同的贷款以减小风险，每笔贷款由多个投资者投资。作为借贷的新模式人人贷吸引了越来越多的注意。Hulme[1]、Wang[3]从多个方面对民间借贷进行了深度剖析；Jin[2]检验了在线社交网络能否缓和市场的信息不对称问题；Puro[4]研发贷款者决策支持系统，帮助贷款者进行策略选择。上述研究都是定性研究，缺乏对人人贷上贷款风险的定量研究，人人贷仍沿用传统借贷的风险评估方法。

风险的决定因素很多，传统的回归方法能以较高的准确度解决二元问题，即定性问题；对于定量问题准确度远远不够。本文提出基于核的信贷风险评估模型，利用核函数将原特征空间变换到某一新的高维特征空间，然后在这个空间上进行线性回归。结果表明，与传统借贷中常用的逻辑回归（线性回归）相比，该模型能更准确地预测贷款风险。

## 2. 基于核的信贷风险评估模型

本文提出基于核的风险评估模型，首先利用核函数代替最优分类面中的点积，把原特征空间变换到一个新的特征空间，然后在这个新的特征空间上对训练样本进行回归，得到贷款及借款人的特征对贷款风险和收益率的影响大小，即回归函数的系数。利用训练后的基于核函数的回归模型即可得到该贷款的风险和收益率。

### 核函数

为了更好地认识这种能够将原特征空间映射到高维特征空间的“特别”函数，下面给出它的数学描述：设  $\chi$  是  $R^n$  中的一个子集。称定义在  $\chi \times \chi$  上的函数  $K(\chi, \chi')$  是核函数，如果存在着从  $\chi$  到某一个 Hilbert 空间  $H$  的映射：

$$\Phi: \begin{matrix} \chi \rightarrow H \\ x \rightarrow \phi(x) \end{matrix}$$

使得

$$K(x, x') = (\phi(x) \cdot \phi(x'))$$

其中， $(\cdot)$  表示  $H$  中的内积。

根据这个思想，将核函数  $K(\chi, \chi')$  代替最优分类面中的点积，这样就相当于把原特征空间变换为高维特征空间。通过上述两式可以发现，尽管通过非线性函数将样本数据映射到具有高维甚至于无穷维的特征空间，并在特征空间构造最优分类超平面，但在求解最优化问

题和计算决策函数时并不需要显示计算该非线性函数，而只需计算该和函数，从而避免特征空间维数灾难问题，也不会增加计算复杂性。这里我们采用较易计算的径向基函数内核：

$$K(|x - x_i|) = \exp\left(-\frac{\|x - x_i\|^2}{\sigma^2}\right)$$

### 高维空间回归

假设给定训练样本  $(x_i, \mu_i, \sigma_i), x \in R^d, \mu \in R, i = 1, \dots, n$  其中  $\chi$  是贷款和借款人的个人特征变量， $\mu$  是贷款收益率， $\sigma$  则是贷款的风险，即收益率的标准差，回归的目标就是求下列回归函数：其中  $\langle \cdot, \cdot \rangle$  为  $\chi$  的内积

$$f(x) = \langle \omega \cdot x \rangle + b.$$

利用样本训练可得：

$$\omega_\mu = \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i, \quad \omega_\sigma = \sum_{i=1}^n (\beta_i - \beta_i^*) x_i.$$

其中， $\alpha_i, \alpha_i^*, \beta_i, \beta_i^*$  是使得回归均方误差最小的系数。将训练结果代入回归式，并引进核函数，则得贷款风险的回归方程：

$$\mu = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x, x') + b_\mu, \quad \sigma = \sum_{i=1}^n (\beta_i - \beta_i^*) K(x, x') + b_\sigma.$$

## 3. 实验验证

人人贷市场提供了大量真实的交易数据，本实验是基于活跃的人人贷市场 Lending Club 的公开数据集进行的。为了验证基于核的回归模型（KRM）的有效性，我们将他与传统的线性回归模型（LRM）进行比较。将每个数据集分为 12 个子集，使用 12 折交叉验证来证明模型的有效性。每次使用一个作测试，其余作训练。按如下步骤比较：

(1) 训练每个模型，用训练得到的模型预测测试集中每笔贷款的收益率  $\mu_i$  和风险  $\sigma_i$ ；

(2) 比较预测收益率  $\mu_i$  和真实收益率  $R_i$ 、预测风险  $\sigma_i$  和真实风险，检验模型的预测准确度。

(3) 将每个模型的预测收益率  $\mu_i$  和风险  $\sigma_i$  输入到投资组合决策模型 [5]，设置 9 组总投资额  $M$  和期望收益率  $R^*$ ，验证模型在投资组合决策中是否能取得较好的效果。9 组参数如表 1 所示。

表 2 给出了每个模型的预测收益率  $\mu_i$  和实际收益率  $R_i$ 、预测风险  $\sigma_i$  和实际风险  $\sigma_i$  的差别，具体来说，我们计算了欧氏距离作为准确性的度量，如表 2 所示，KRM 的欧氏距离要小，表明其准确度更高。

图 1 不同模型的输出结果输入投资组合决策模型后的表现，可以看出无论总投资额  $M$  和期望收益率  $R^*$  是多少，基于 KRM 的投资组

# 计算机语言识别技术的应用分析

张统豪

(福建师范大学协和学院, 福建福州 350108)

**摘要:**在现代科技高速腾飞的今天,计算机信息技术已经成为人们生活办公中必不可少的一部分,计算机信息技术的不断丰富毫无疑问地改变和发展了人们生活,推动了现代文明的进步。现如今,各国在发展计算机信息技术方面的投入逐渐加大,计算机信息技术已经成为综合国力竞争的重要组成部分。在此趋势下,加强对计算机信息技术科学的发展趋势探析有利于我们更好地了解其发展历史和发展现状,推动计算机语言识别技术的发展与进步,发展现代化的语音通信,造福人类生活,促进社会发展进步。

**关键词:**计算机科学;信息网络;关键技术;趋势

**中图分类号:**TP18 **文献标识码:**A **文章编号:**1003-9767(2013)01-0008-02

计算机信息技术是一种现代化科学技术产物,是将现代的通信技术与计算机技术相结合的技术性产物。计算机信息技术是按照网络协议,集合了地球上分散的、独立的计算机信息技术,使其相互连接,发展成各种技术性的集合。连接的介质主要有电缆、双绞线、光纤、微波、载波或通信卫星。此外,计算机信息技术不但具有共享硬件、软件功能,还有具有处理庞大的数据信息资源的功能,集中处理共享数据信息资源,并且对其进行管理和维护。目前,随着计算机语言的各种关键技术的不断发展,计算机语言识别技术已经深入到社会、生活的各个领域,体现在各个国家政府机关、企业部

门、家庭通信等,计算机网络技术不仅从公众视野逐步迈向私人领域,各种关键技术的研发和问世体现着现代科技的进步性,具有着超强生命力和发展前景,近年来,随着科技的不断创新,计算机语言识别等技术在运算速度、软件开发、关键技术性能等都取得了突破,创造了关键网络技术质与量的极大飞跃,并且逐步推广开来。

## 1. 计算机语言识别技术的发展现状

### 1.1 普及性与深入发展性

随着科学技术的不断发展与进步,目前,计算机语言识别技术也在不断地更新,作为当今社会技术发展进步的潜在力量,计算机语言

表 1 投资组合的基准参数

组别	期望收益率 ( $R$ )	投资总额 ( $M$ )
1	5.5%	\$15,000
2	5.5%	\$20,000
3	5.5%	\$25,000
4	6.5%	\$15,000
5	6.5%	\$20,000
6	6.5%	\$25,000
7	7.5%	\$15,000
8	7.5%	\$20,000
9	7.5%	\$25,000

表 2 模型准确度比较

欧氏距离	LRM	KRM
收益率	14.55	14.23
风险	12.49	11.89

合模型输出的投资组合收益率要更高,这间接反映了 KRM 模型具有更高预测准确度。

## 4. 总结

我们提出基于核的信贷风险评估模型来预测人人贷的贷款风险和收益率,该模型将原有的特征空间映射到高维的特征空间,使得在低维空间不可分的特征在高维特征空间变得线性可分;在此基础上运用线性回归对收益率和风险进行回归计算,与直接进行回归运算相比更加准确。基于真实世界数据的实验从单笔贷款和贷款组合两个层次表明,该模型与传统借贷中常用的线性回归模型相比预测准确度更高。

**参考文献:**

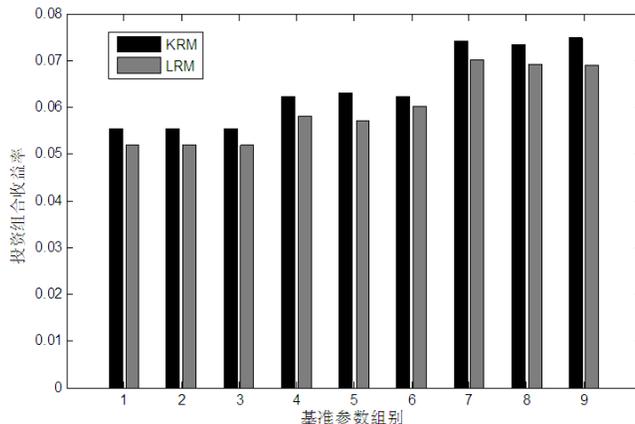


图 1 投资组合收益率

[1] M. Hulme. Internet based social lending. Social Futures Observatory, 1-115, 2006.

[2] G. Jin. Do social networks solve information problems for P2P lending? Working Papers, 2008.

[3] H. Wang, J. Aronson. People-to-People Lending: The Emerging E-Commerce Transformation of a Financial Market. Value Creation in E-Business Management, 182-195, 2009.

[4] L. Puro. Borrower Decision Aid for people-to-people lending. Decision Support Systems, 52-60, 2010.

[5] 雒春雨. P2P 网络借贷中的投资决策模型研究. 博士研究生学位论文. 大连理工大学, 2012, 06.