

学校编码: 10384

学号: 200431060

分类号\_\_\_\_\_密级\_\_\_\_\_

UDC\_\_\_\_\_

厦 门 大 学

硕 士 学 位 论 文

基于 SVM 的满意特征选择及其在企业  
信用评估中的应用

Satisfactory Feature Selection based on SVM and Its  
Application in Enterprise Credit Assessment

凌 健

指导教师姓名: 林 成 德 教授

专 业 名 称: 控制理论与控制工程

论文提交日期: 2007 年 5 月

论文答辩时间: 2007 年

学位授予日期: 2007 年

答辩委员会主席: \_\_\_\_\_

评 阅 人: \_\_\_\_\_

2007 年 5 月

## 厦门大学学位论文原创性声明

兹呈交的学位论文，是本人在导师指导下独立完成的研究成果。本人在论文写作中参考的其他个人或集体的研究成果，均在文中以明确方式标明。本人依法享有和承担由此论文产生的权利和责任。

声明人（签名）：

年 月 日

## 厦门大学学位论文著作权使用声明

本人完全了解厦门大学有关保留、使用学位论文的规定。厦门大学有权保留并向国家主管部门或其指定机构送交论文的纸质版和电子版，有权将学位论文用于非赢利目的的少量复制并允许论文进入学校图书馆被查阅，有权将学位论文的内容编入有关数据库进行检索，有权将学位论文的标题和摘要汇编出版。保密的学位论文在解密后适用本规定。

本学位论文属于

1. 保密（ ），在年解密后适用本授权书。
2. 不保密（ ）

（请在以上相应括号内打“√”）

作者签名：                      日期：    年        月        日

导师签名：                      日期：    年        月        日

## 摘要

信用评估是银行贷款业务的核心工作，对贷款客户的信用评价是否科学可靠，关系着银行经营的成败。

国内银行的信用评估方法目前还停留在传统的比例分析阶段，亟需建立一套更为科学的评估模型。近年来，神经网络（NN）、支持向量机（SVM）等智能模型已被引入信用评估领域，并取得了令人鼓舞的研究性成果。但前人的研究着重于智能评估模型的建立，对评估指标体系的选定却未作充分的探讨。本文采用数据挖掘领域中的特征选择方法，结合 SVM 分类技术，对评估指标的选取进行了较为深入的研究，并以此为基点，对信用评估的建模方法进行了新的拓展。

本文将满意优化思想引入特征选择方法中，并针对国内银行信用数据的特点，综合考虑选出特征子集的维数及其在各个信用等级上的预测能力，设计了独特的满意特征选择算法。为使选出的特征子集具有算法针对性及适用性，本文采用 SVM 交叉验证方法评估特征子集的预测性能。通过与三种不同的特征选择方法的比较，发现基于 SVM 的满意特征选择能选出维数较低、总体预测性能较高、等级间预测性能较为均衡的特征子集，证实了该算法的适用性及优越性，并最终选定了一套较为简洁、实用的信用评估指标体系。

在此基础上，本文还将特征选择方法应用到建模领域，探索构建多分类 SVM 组合评估模型的有效方法。对比实验表明，基于满意特征选择方式的多分类 SVM 组合模型，在预测性能上优于基于 Bagging 方法的组合、基于特征分组方式的组合、以及另外三种基于 Wrapper 特征选择方法的组合，并大大好于单个多分类 SVM 分类器，是一种有效且高性能的信用评估模型。

**关键词：**信用评估；特征选择；支持向量机 (SVM)

## Abstract

Credit assessment is the key step in loan business of commercial banks. The rationality and reliability of credit assessment will greatly affect the achievements of a bank.

The commercial banks in china need some much better assessment methods to improve their competition ability, in which, nowadays, the traditional ratio-analysis method is popularly adopted. Recently, the intelligent models, such as Neural Networks (NN) and Support Vector Machines (SVM), have been introduced into the credit assessment domain and have already achieved some promising results. The existing researches are mainly focused on the establishment of the intelligent assessment models, but the optimal selection of the evaluating index system has still been rarely studied. In this paper, a new feature selection scheme, based on SVM, is adopted to perform the task of selecting the evaluating indexes. Furthermore, new modeling method is developed based on this scheme.

We have designed the Satisfactory Feature Selection (SFS) scheme by applying satisfactory optimization technique into feature selection method. In SFS, the classification performance of the selected feature-subset and its size are considered compromisingly based on the characteristic of the credit data in domestic banks. Three other different feature selection schemes are presented to contrast with SFS. The experiment results show that SFS is superior to three other ones in the qualities such as the size of the selected subset, the predict performance on the total sample, and the predict performances equilibrium within the individual classes. A terse and reasonable evaluating index system is presented after the contrastive experiments.

The ensemble methods of multi-classification SVMs are also studied in this paper. We manipulate the input feature-set to construct the classifier-ensemble, and create base-classifiers with large diversity by SFS. The contrastive experiments show that the ensemble based on SFS is superior to the ensemble based on bagging method, the ensemble based on feature grouping method and the ensembles based on three other feature selection methods. The multi-classification SVMs ensemble based on SFS also performances much better than a single SVM classifier.

**Keywords:** Credit Assessment; Feature Selection; Support Vector Machine (SVM)

# 目 录

前 言 .....	1
<b>第一章 企业信用评估及评估方法.....</b>	<b>3</b>
1.1 企业信用评估概述 .....	3
1.2 信用评估的指标体系 .....	4
1.3 企业信用评估方法与模型 .....	6
1.3.1 5C 要素分析法（专家判断法） .....	6
1.3.2 百分制计分方法 .....	7
1.3.3 信用模型分析法 .....	7
<b>第二章 特征选择及支持向量机（SVM）技术简介 .....</b>	<b>12</b>
2.1 特征选择 .....	12
2.1.1 特征选择概述 .....	12
2.1.2 特征选择过程 .....	13
2.1.3 特征选择方法分类 .....	14
2.2 支持向量机（SVM）简介 .....	17
2.2.1 最优分类超平面 .....	17
2.2.2 软间隔分类超平面 .....	19
2.2.3 支持向量机 .....	19
2.2.4 内积核函数的选择 .....	20
2.2.5 SVM 的优点 .....	20
2.2.6 SVM 的相关研究概述 .....	21
<b>第三章 基于 SVM 的满意特征选择及其在信用指标选取中的应用 ...</b>	<b>23</b>
3.1 待选指标体系的确定 .....	23
3.2 样本预处理 .....	25
3.2.1 信用等级的划分 .....	25
3.2.2 总体样本的筛选与抽样 .....	26
3.2.3 样本输入归一化 .....	27
3.3 基于 SVM 的满意特征选择方法 .....	28

3.3.1	特征集满意度函数的设计 .....	30
3.3.2	满意特征选择算法及其参数选择 .....	32
3.4	对比实验及评估指标体系的选定 .....	34
<b>第四章 基于多分类 SVM 组合的企业信用评估模型 .....</b>		<b>39</b>
4.1	分类器组合 .....	39
4.1.1	分类器组合概述 .....	39
4.1.2	构造组合分类器的方法 .....	40
4.2	基于满意特征选择的多分类 SVM 组合信用评估模型 .....	43
4.2.1	多分类 SVM 组合方法的比较研究 .....	43
4.2.2	基于满意特征选择的多分类 SVM 组合模型 .....	49
<b>第五章 总结及展望 .....</b>		<b>52</b>
<b>参考文献 .....</b>		<b>54</b>
<b>攻读硕士学位期间发表的论文 .....</b>		<b>58</b>
<b>致 谢 .....</b>		<b>59</b>

# Contents

<b>Preface .....</b>	<b>1</b>
<b>Chapter 1 Credit Assessment of Enterprise and Its Approaches.....</b>	<b>3</b>
1.1 Introduction of Credit Assessment.....	3
1.2 The Index System for Credit Assessment.....	4
1.3 Methods and Models for Credit Assessment.....	6
1.3.1 5C Factors Analysis .....	6
1.3.2 Scoring Method Based on Percentage .....	7
1.3.3 Analysis on Credit Model Methods .....	7
<b>Chapter 2 Introduction of Feature Selection and Support Vector Machine (SVM) .....</b>	<b>12</b>
2.1 Feature Selection Technique.....	12
2.1.1 Introduction of Feature Selection Technique .....	12
2.1.2 The Process of Feature Selection.....	13
2.1.3 The Framework of Feature Selection Methods.....	14
2.2 Support Vector Machine (SVM).....	17
2.2.1 The Optimal Separating Hyperplane .....	17
2.2.2 The Soft Margin Separating Hyperplane .....	19
2.2.3 Support Vector Machine.....	19
2.2.4 The selection of Inner Product Kernel Function.....	20
2.2.5 Advantages of SVM.....	20
2.2.6 Related Researches on SVM.....	21
<b>Chapter 3 Satisfactory Feature Selection (SFS) Method Based on SVM and Its Application in Selecting Credit Indexes.....</b>	<b>23</b>
3.1 The Index System for selecting.....	23
3.2 Data Pretreatment.....	25
3.2.1 The Classification of Credit Rating .....	25
3.2.2 Data Sampling.....	26
3.2.3 Scaling of Input-Data.....	27



<b>3.3</b>	<b>Satisfactory Feature Selection Method (SFS) Based on SVM.....</b>	<b>28</b>
3.3.1	The Design of Satisfactory-rate Functions .....	30
3.3.2	Satisfactory Feature Selection Algorithm and Its Parameter Selection ..	32
<b>3.4</b>	<b>Contrastive Experiments and the evaluating index system.....</b>	<b>34</b>
<b>Chapter 4</b>	<b>A Credit Assessment Model Constructed by</b>	
	<b>Multi-classification SVMs Combination Based on SFS.....</b>	<b>39</b>
<b>4.1</b>	<b>Classifiers Combination.....</b>	<b>39</b>
4.1.1	An Introduction of Classifiers Combination.....	39
4.1.2	Methods of Classifiers Combination .....	40
<b>4.2</b>	<b>A Credit Assessment Model Constructer by Multi-classification SVMs</b>	
	<b>Combination Based on SFS.....</b>	<b>43</b>
4.2.1	Comparison of Multi-classification SVMs Combination Methods .....	43
4.2.2	Multi-classification SVMs Combination Model Based on SFS .....	49
<b>Chapter 5</b>	<b>Conclusion and Perspective .....</b>	<b>52</b>
	<b>Reference .....</b>	<b>54</b>
	<b>Published Papers .....</b>	<b>58</b>
	<b>Acknowledgement.....</b>	<b>59</b>

厦门大学博硕士学位论文摘要库

## 前言

商业银行在金融体系中占有举足轻重的地位,在创造货币存款、提高金融政策效率、实现社会投资等方面都发挥着核心作用。然而,商业银行在营运过程中无时无刻不面临着各种金融风险,包括信用风险、市场风险、利率风险、流动性风险、操作风险、法律风险和声誉风险等。其中,信用风险占有特殊的地位,世界银行对全球银行业危机的研究表明,导致银行破产的最常见原因就是信用风险。信用风险是指信贷资金安全系数的不确定性,表现为企业或贷款人由于各种原因,不愿意或无力偿还银行贷款本息,使银行贷款无法回收,形成呆账的可能性。信用评估是商业银行信用风险管理的首要工作和关键环节,事关银行的生存和社会的稳定。

在经济发达国家,信用评估已有近百年的历史;而在我国,企业信用评估自起步起,也只有十几年的时间,还属新生事物。大多数的国内商业银行目前还停留在传统的比例分析阶段,普遍采用的是“打分法”这一评估方式。对信用评估工作的基础——评估指标体系的确定,也一直没有统一和规范起来,各家评估机构各有一套体系,独树一帜,自主评级。对于处于新兴市场和转型经济时期的我国商业银行而言,亟需建立一套更为科学的评估体系及评估模型,本文的选题就是在这样的背景之下产生的。

本课题是与福建省某咨询公司的合作项目,在此之前,课题组已采用多种机器学习的方法对信用评估模型的建立进行了较为深入的研究:叶武<sup>[1]</sup>、陈雄华<sup>[2]</sup>、高洪涛<sup>[3]</sup>使用神经网络技术建立较为可行及有效的信用等级评估模型;张柳园<sup>[4]</sup>、金剑<sup>[5]</sup>将基于符号处理的专家系统技术和神经网络技术相结合,融入了更为智能化的推理解释机制,使得信用评估系统更加的实用化;刘闯<sup>[6]</sup>、阳维<sup>[7]</sup>分别采用支持向量机(SVM)技术及基于人工神经网络分类器组合技术,进一步提高了评估模型的性能。

在上述的研究当中,前人着重于评估模型的建立,而对信用评估的基础——评估指标体系的选择,却显得过于简单。他们大多数依从财务专家建议,或只是进行简单的筛选,并没有在此领域进行深入的探讨。本文针对于这一缺陷,将基于SVM的满意特征选择技术应用于评估指标体系的选取中,提出了新方法,并在

此基础上构造 SVM 多分类组合模型。研究表明,该方法确定的指标体系简洁、有效,新的组合模型在预测精度及等级预测性能方面较单一 SVM 评估模型均有一定程度的提高。

本文内容具体组织如下:

第一章 介绍了企业信用评估的原则和方法,阐述已有的建模技术及其优缺点与适用性。

第二章 着重描述本文所依据的技术基础。本章包括两个部分:第一部分是介绍特征选择技术的概念、意义及一些主流的方法;第二部分对 SVM 技术的理论基础和算法做了一个简要介绍。

第三章 详细阐述了基于 SVM 的特征选择技术在信用指标选取中的应用。在确定待选指标体系及样本预处理之后,详细描述了基于 SVM 的满意特征选择算法的设计及运行流程。并研究了算法中的参数选择问题,对优化算法起了关键的作用。最后将其与三种不同的特征选择方法进行对比,列出了相关的实验结果,并最终选定了一组较好的评估指标体系。

第四章 对构造多分类 SVM 的组合评估模型进行了深入的讨论。利用第三章中的特征选择方法,采用不同的组合方式,分别进行了对比实验。实验结果表明,基于满意特征选择的多分类 SVM 组合分类器较其它方法产生的模型优越,是一种较为有效的评估模型。

第五章 总结了本课题的研究工作,分析了研究过程中存在的不足、以及将来可以继续深入的方向,并对企业信用评估研究的前景作了展望。

## 第一章 企业信用评估及评估方法

### 1.1 企业信用评估概述

信用评估，就是由资信评估机构，根据科学的指标体系，采用严谨的分析方法，运用明确的文字符号，对被评估单位履行经济责任的能力及其可信任程度进行客观公正的评价，并确定其资信等级的一种经济活动，是对债务偿还风险的综合评价。按照评估对象来分，信用评估可分为企业资信评估、证券资信评估等，详见图1.1。此外，按照评估方式来分，可分为公开评估和内部评估；按照收费与否可分为有偿评估和无偿评估；按照评估内容可分为综合评估和单项评估。我们在研究制定具体的评估标准和方法时，要充分注意到不同类型之间的差别，从而更加客观的反映被评估对象的资信状况。本文研究的重点是企业资信评估中的公司企业信用评估。

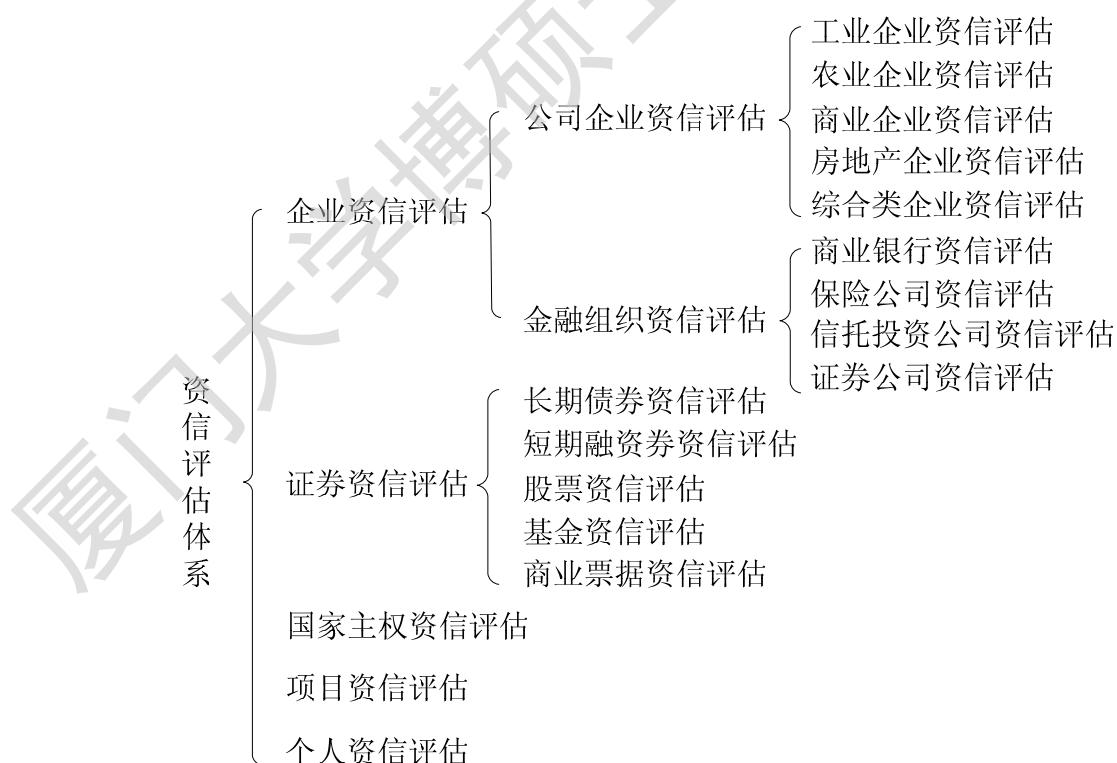


图1.1 按照评估对象来划分的资信评估体系

信用评估能提供市场信息，帮助投资者选择投资项目，减少投资风险，而且

也有利于拓宽筹资者的融资渠道，降低筹资费用。信用评估也是现代金融监管机制中非常有用的工具。金融监管部门可以通过信用评估方法，提高金融机构的信息披露标准，增强透明度；根据信用级别限制受监管机构的投资范围，控制风险；利用信用评估监测金融机构的资产质量和资本充足率，提高监管效率；利用信用评估监测金融机构偿债能力的变化，提高信贷资产质量<sup>[8]</sup>。

国际信用评估制度已有一百多年的历史。信用评估制度经过长时间的发展，逐步形成了一套较为完善的理论和方法。目前，国际上已有许多著名的评估公司，如穆迪、标准普尔、惠誉等，这些公司的评估业务已覆盖全球，在推动世界金融及经济的发展中，扮演着极其重要的角色。

中国的信用评估业产生于1987年，迄今已有二十年的历史。1988年，我国成立了第一家独立的资信评估机构，此后资信评估开始受到重视，有了一定的发展，但总体说来水平不高，缺乏一套科学的评估管理体系，难以树立国际声誉。目前，在国际资本市场上，根本听不到我国资信评估机构的声音，国外评估机构实际上垄断了中国信用评估市场，掌握了中国进入国际资本市场的“通行证”。尽管这些国际评估机构的大多数结论较真实的反映了一些国家和经济组织的状况，有利于避免信贷危机，保证借贷人利益，有助于有关国家和经济组织正视问题，加速金融整改，但由于这些评估机构均是西方的跨国公司，其意识形态、政治立场、经济理论、价值观念、文化观念、评定标准等难免带有西方国家的主观色彩，这就使得我国在对外的经济活动中处于一种十分被动的局面，增加了我们的交易成本，随着我国对外开放力度的加大，这一问题就越发的突显，因此，在我国建立一套符合我国国情的、科学的评估理论和方法体系，拥有我们自己的国际性信用评估机构，是极其重要的。

## 1.2 信用评估的指标体系

信用评估指标体系是信用评估机构和评估人员从事信用评估工作的基础和依据，没有一套科学的信用评估指标体系，很难设想，它能保证评估结果的客观公正与相应的可信程度。但我国对发债和贷款企业的信用评估指标体系，一直没有统一和规范起来，各家评估机构各有一套体系，独树一帜，自主评级。近年来将智能分类模型应用于信用评估领域也是研究的热点，但对于智能模型的输入，

即信用评估指标体系选取的研究，却还处于起步阶段。大多数的智能评估模型只是将某个评估机构的指标体系直接拿过来用，或只做简单的处理。这样得到的指标体系，不是数量太多，影响评估模型的效率，就是不够全面，遗漏了某些关键指标，使得模型的精度受到影响。可见，评估指标体系的选择已经成为信用评级工作的首要问题，它关系着评估工作的成败。

寻找一套较为科学的指标选取的方法，是本文研究和探索的重点之一。为使选出的指标体系保留原始的财务意义，本文采用特征选择的方法，在一套数量较多、覆盖较为全面的待选信用评估指标体系中，选出数量恰当、构成合理的核心评估指标。在后续的章节中，将对此问题进行深入的探讨。在此之前，为使指标的选取更为客观、可信，待选指标体系的确定必须在正确的指导原则下进行，该原则可归纳为：

(1) 全面性：待选指标体系的内容应该全面地反映所有影响评估对象资信状况的各项要素，不但要考核过去的业绩，还要预测未来的发展趋势；不但要考虑评估对象本身的情况，而且还要研究周边的环境及其产生的影响。

(2) 科学性：建立待选信用评估指标体系，各项指标必须有机配合，形成体系，相互之间既不重复，又无矛盾；同时，指标的计算和评价方法必须科学，要有一定的依据。整个指标体系的建立，要在不断实践的基础上逐步充实提高，要经得起实践的考验，逐步增强指标体系的科学性。

(3) 针对性：待选评估指标体系必须具有针对性，不同的评估对象和评估目的，指标体系应该有所区别。信用评估在国内目前分为贷款企业评估、证券评估和特定资信关系评估三类，就应根据评估要求的不同，分别制定各自的信用评估指标体系。又由于不同企业具有各自的经营特点，因而有一部分指标又要结合企业的经营特点来确定，不能千篇一律。

(4) 公正性：待选评估指标体系的建立，要符合客观事实，能正确反映评估对象信用等级真实面貌。指标体系和计算方法不能偏向评估对象或投资主体的任何一方，评估机构和评估人员必须态度公正，评价客观，以事实为依据，不能根据个人爱好，任意改变指标项目、计算方法和评价标准。

(5) 合法性：信用评估必须遵守国家有关政策、法律和法规，待选评估指标体系要体现国家宏观经济政策的导向，有些经济效益指标和风险监管指标，国家政府机关规定有标准值的，必须体现规定要求。

(6) 可操作性: 待选评估指标体系的建立, 要具有实用性, 便于操作和设计计算机程序。既要符合我国国情, 具有本国特色; 又要参照国际惯例, 考虑今后同国际惯例接轨, 逐步走向世界。

### 1.3 企业信用评估方法与模型

早期的资信调查, 比较简单, 主要是定性判断, 随着资信评估业的发展, 资信评估工作逐步健全起来, 定性分析和定量分析工具得到了综合运用, 形成了多种不同特点、不同风格的评估方法, 为银行和企业做出信贷和财务决策提供了科学的依据。现将几种使用较为普遍的信用评估方法作一简要介绍。

#### 1.3.1 5C 要素分析法 (专家判断法)

专家判断法<sup>[9]</sup>在银行业中的应用由来已久, 目前在许多银行仍居于主导地位。它是指依靠信贷专家的专业技能、主观判断和某些关键因素的权衡来评判借款人的违约可能性, 进而确定是否发放贷款和发放贷款的规模。这些信贷专家一般是银行或其分支机构的信贷审批官 (Loan Approving Officer), 不同的信贷专家从不同的视角来看待企业的经营和财务情况, 其考虑的潜在因素和方法很多。但是, 以下五个方面是多数专家评估企业必须要考虑的因素。

- 品格 (Character): 对企业声誉的一种度量, 考察其偿债意愿和偿债历史。品格包含多个方面, 既包括其在不同银行、工商局和税务海关等部门的资信记录, 也包括其在对于上下游产业的履约记录。

- 偿债能力 (Capacity): 还款能力, 反映借款者受益的易变性。

- 资本 (Capital): 所有者的股权投入及其对债务的比率, 这一类指标是预示企业破产可能性的良好指标, 同时也可以反映出自身经营头寸的充足与否和对外源融资如银行借款的依赖性。

- 抵押品 (Collateral): 典型的贷款风险缓解手段, 如果发生违约, 银行对借款人的抵押物品就具有要求权。

- 经营环境 (Condition): 包括经济周期因素的影响, 商业周期的状态是决定信用风险损失的一项重要因素。

尽管许多银行仍然在其信贷决策中使用专家方法, 但是这些方法具有两个重



Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to [etd@xmu.edu.cn](mailto:etd@xmu.edu.cn) for delivery details.

厦门大学博硕士论文摘要库