

学校编码: 10384

分类号____密级____

学号: 13920080150050

UDC____

厦 门 大 学

博 士 学 位 论 文

数据挖掘技术在政府绩效评估指标设计中的
应用研究

A Study on the Application of Data Mining Methods in Indicator
Development of Government Performance Evaluation

刘典文

指导教师姓名: 卓越教授

专业名称: 行政管理

论文提交日期: 2011年4月

论文答辩时间: 2011年 月

学位授予日期: 2011年 月

答辩委员会主席: _____

评 阅 人: _____

2011年4月

厦门大学学位论文原创性声明

本人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其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（）课题（组）的研究成果，获得（）课题（组）经费或实验室的资助，在（）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年 月 日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

厦门大学博硕士学位论文摘要库

摘要

科学的绩效指标设计是政府绩效评估有效实施的核心和关键环节，也是政府绩效评估理论和实践中的难点问题所在。而绩效指标设计方法和思路的选择作为政府绩效指标设计的灵魂，在很大程度上决定着政府绩效指标的科学性、合理性和有效性。因为，只有采取科学的政府绩效评估指标设计方法，才能确保绩效评估的目的和内容得以实现，也才能保证政府绩效评估结果的客观性、全面性和公正性。关于政府绩效评估指标设计的思路和方法，国内外学者一直在努力探寻和摸索。尽管如此，我国政府绩效评估指标设计仍面临着诸如理论认识储备不足和设计方法运用不充分、绩效评估指标构建多处于自发被动状态、定性分析与定量分析方法未能很好地结合等问题，很大程度上制约着政府绩效评估指标设计的科学性、客观性和有效性。数据挖掘技术的出现和应用，为政府绩效评估指标的设计提供了一种全新的、有效的方法和技术，利用该技术对政府部门相关数据进行挖掘分析，可以在大量关系复杂的数据中挖掘出隐含的深层次信息，并有助于形成以知识管理和知识发现来直接或间接辅助政府绩效评估指标的设计。本文旨在收集有关政府数据资源作为研究对象的基础上，选取众多数据挖掘方法和技术中的主成分因子回归分析、决策树分类模型和聚类分析等方法，进行实证分析并探索性地构建有关的绩效评价指标或者为绩效指标设计提供导向性思路。

通过本研究可以发现，数据挖掘中的主成分因子回归分析方法能够用于设计具有预测性评价功能的政府绩效评估模型指标。决策树分类模型运用于政府部门相关业务数据分析得出的有用知识，则可为政府绩效评估指标设计直接提供有效信息和依据。而聚类分析则为政府绩效评估指标设计提供了一个崭新的思路和视角。同时，数据挖掘技术在政府绩效评估指标设计中应遵循一定的步骤，不同的数据挖掘技术在政府绩效评估指标设计中也有各自的适用范畴。总之，数据挖掘技术应用于政府绩效评估指标设计方面可以有效凸显其使用价值，并且在公共管理领域也具有广阔的应用前景。

本文的创新或研究贡献之处则主要体现在：

1. 把数据挖掘技术引入政府绩效评估指标设计领域，利用挖掘所获得的信

息用于政府绩效评估指标的设计，使绩效指标的设计更具针对性、科学性。通过本文的研究，可以充分地认为，数据挖掘技术作为一种与社会调查研究方法、平衡计分卡法、绩效凌柱模型、政府职能目标分解法等并列且有效的绩效指标设计方法，完全是可行的和合理的，它丰富了政府绩效评估指标设计方法类型。

2. 应用数据挖掘的主成分因子回归方法构建了土地利用结构合理性评价的模型指标。作为通过构建数据挖掘统计模型而设计出的模型指标，从其准确性、应用的广度和可操作性等方面均要优于传统方法设计的评价指标，并且模型指标还具备重要的预测性评价功能。

3. 应用数据挖掘的决策树分类模型设创设了建设项目闲置用地整治评价的指标体系。土地管理部门关于建设项目闲置用地整治的业务数据中，隐藏着大量有关土地类型特征、项目准入产业、投资强度等与土地闲置之间密切关系的信息，利用决策树分类模型对政府部门业务数据进行挖掘分析，以问题为导向直接设计出闲置土地整治评价的若干创设性指标。这在政府绩效指标的来源及其设计方法看，一定程度上也是一种创新。

4. 应用数据挖掘的聚类分析方法设计政府绩效评估间接性建议指标。通过对政府绩效指标的文本数据集进行聚类挖掘，可以实现从指标到指标的筛选性指标设计的功能。而本研究主要是针对大量的政府扶贫开发监测数值型数据，采用聚类分析方法对众多的扶贫开发对象进行聚类分析，并对不同的扶贫开发项目类型提出了建议性的评价指标。这可以使绩效指标的设计更具针对性，并体现各自扶贫对象的特色，也能使扶贫开发绩效的横向对比更具操作性和合理性。

5. 归纳和提炼了数据挖掘技术应用于政府绩效评估指标设计的一般步骤，并通过比较分析研究，总结和发现数据挖掘方法中主成分因子回归方法、决策树分类模型和聚类分析方法在政府绩效评估指标设计应用中的适用性，为数据挖掘技术在政府绩效评估指标设计中的推广应用提供可能。

关键词： 数据挖掘；主成分因子分析；决策树；聚类分析；政府绩效指标

Abstract

Scientific development of performance indicator is the core and critical process of the implementation of government performance, also the difficulty that has to be conquering in the theory and practice of government performance evaluation. As the spirit of government performance evaluation development, to a great extent, the method and clue for performance indicator development decides the scientific, rationality and effectiveness of government performance indicator. Therefore, only scientific development method be selected that the aim and content of performance evaluation could be achieved, and the objectiveness, comprehensiveness and fairness also could be realized. Scholar whatever domestic and abroad are exploring industriously for the designing technique and approach. However, we still face problems such as narrow-minded theories, development method not be used completely, quality and quantity methods can not be combined effectively, which constraint the development of government performance evaluation indicator. Making use of Data Mining technology introduces a completely new and effective method and technology for government performance evaluation indicator development. Potential information can be excavated from compliantly related data when using this technique to dig government data, and also good to form a method to aid the development of the government performance evaluation indicator directly and indirectly by knowledge management and knowledge discovery. This dissertation aims to develop related performance evaluation indicator or supply the guide idea for performance evaluation indicator development based on collecting government data resource and doing empirical analysis by methods of Principal Component Regression analysis, Decision Tree method and Cluster Analysis.

This study demonstrated that Principal Component Regression analysis method of Data Mining could be used to design indicators of government performance evaluation which has forecast function. Making use of Decision Tree model in analyzing data of government operating department could affords effective information and basis for government performance evaluation indicator development.

Cluster Analysis supplies a completely new perspective for that. At the same time, there would be some fixed procedures that must be obeyed when using Data Mining technique; different kinds of Data Mining techniques have its categories in government performance evaluation indicator development. In conclusion, Data Mining technique has its values when used in public management, especially government performance evaluation, and has vast application foreground.

The innovation of this paper lies in:

First of all, introducing Data Mining technique into government performance evaluation indicator development, using information got by digging in indicator development to make assure of the scientific and pertinence of the indicator. Through this dissertation, it's approved that as a method used in performance evaluation indicator development; Data Mining is feasible and reasonable and can be compared to methods such as social investigation, BSC, Performance Prism. It enriches the methods of government performance evaluation indicator development.

Secondly, using Principal Component Regression Analysis method of Data Mining construct an indicator model of land use evaluation. As a model indicator designed through contriving Data Mining statistic model, this model indicator is much better in accuracy, application extent and maneuverability, especially, it has function of forecast evaluation.

Thirdly, making use of Decision Tree methods to design creative evaluation indicator system when remediate idle land in construction projects. There is much information closely related to the reasons of idle land such as land type, project approval industries, investment intencity. Applying Decision Tree method to analyzing business data of government, designing creative indicators for idle land's evaluation directly and problem-oriented. This is a kind of innovation of development method and origin of government performance evaluation indicator.

Fourthly, using Cluster analysis method to design indirect constructive indicator in Poverty alleviation and development performance evaluation. By clustering excavate text data of government performance indicator, can realize the function that

select indicator from indicator to indicator. This study advanced propositional evaluation indicator for different poverty alleviation projects by using cluster analysis method to analyze many monitored numeric data of government alleviate poverty project. This could make performance evaluation indicator development much more pertinent and reflected the characteristics of objects and improve the maneuverability and rationality of the performance of poverty alleviation projects when compared laterally.

Fifthly, this study concluded and abstracted the general procedure of government performance evaluation indicator development when using Data Mining technique. Summarizing the applicability of Principal Component Regression method, Decision Tree method and Cluster Analysis method of Data Mining technique in government performance evaluation indicator development, approving the possibility of spread and application of Data Mining technique in government performance evaluation indicator development.

Key Words: Data Mining; Principal Component Analysis; Decision Tree; Cluster Analysis; Government Performance Evaluation Indicator

厦门大学博硕士学位论文摘要库

目 录

第一章 绪论	1
1.1 导言	1
1.1.1 选题背景.....	1
1.1.2 研究意义.....	6
1.2 相关概念界定	9
1.2.1 政府绩效评估.....	9
1.2.2 政府绩效评估指标.....	15
1.2.3 数据挖掘.....	17
1.3 相关文献综述	22
1.3.1 政府绩效评估指标设计研究综述.....	22
1.3.2 数据挖掘技术应用有关研究.....	37
1.3.3 简评.....	40
第二章 研究设计	42
2.1 研究目标与研究思路	42
2.1.1 研究目标.....	42
2.1.2 研究思路.....	43
2.2 研究对象的选择	46
2.2.1 备选研究对象.....	47
2.2.2 研究对象筛选依据.....	52
2.2.3 研究对象的确定.....	52
2.3 研究技术和研究工具的选择	54
2.3.1 数据挖掘技术方法的选择.....	54
2.3.2 数据预处理技术的选择.....	61
2.3.3 数据挖掘工具的选择.....	64
2.4 研究方法与研究内容	71
2.4.1 研究方法.....	71

2.4.2 研究内容与结构安排.....	72
第三章 主成分因子回归分析在土地利用结构合理性评价模型指标构建中的应用	75
3.1 理论准备与业务理解	75
3.1.1 土地利用结构及其合理性的理解.....	75
3.1.2 土地利用结构合理性评价研究的意义.....	77
3.1.3 土地利用结构及其合理性评价研究现状.....	79
3.1.4 土地利用结构变化的影响因素.....	81
3.2 实证挖掘方法的选择	84
3.2.1 主要的统计方法.....	85
3.2.2 主成分分析法.....	86
3.2.3 普通最小二乘法.....	87
3.3 实证挖掘过程步骤	88
3.3.1 数据收集.....	89
3.3.2 数据预处理——变量标准化处理.....	89
3.3.3 对标准化后的数据进行主成分分析与因子分析.....	93
3.3.4 利用土地利用强度进行回归与预测.....	98
3.4 土地利用结构合理性评价模型指标总结分析	103
3.4.1 土地利用规模系数模型指标的构建及其特征.....	103
3.4.2 土地利用规模系数模型指标的预测性评价功能.....	105
第四章 决策树分类模型在建设项目闲置用地整治评价指标设计中的应用.....	107
4.1 理论准备与业务理解	107
4.1.1 建设项目闲置用地及其产生原因的理解.....	107
4.1.2 建设项目闲置用地清理整治评价研究的紧迫性.....	109
4.1.3 闲置土地有关问题的研究现状.....	111
4.2 实证挖掘方法的选择	113
4.2.1 决策树分类方法的原理、结构与步骤.....	113

4.2.2 几种常见的决策树算法比较.....	117
4.2.3 决策树模型应用于建设项目闲置用地整治评价指标设计的优势.....	121
4.3 实证挖掘过程步骤	123
4.3.1 数据准备.....	123
4.3.2 数据预处理.....	123
4.3.3 数据理解.....	126
4.3.4 决策树分析.....	132
4.4 指标创设——决策树分类知识的应用	140
4.4.1 决策树分类知识与指标创设.....	141
4.4.2 决策树分类模型创设的建设项目闲置用地整治评价指标.....	142
第五章 聚类分析在政府扶贫开发绩效评价指标设计中的应用.....	144
5.1 理论准备和业务理解	144
5.1.1 我国政府扶贫开发绩效评估的类型及特点.....	144
5.1.2 政府扶贫开发绩效评估的重要性及紧迫性.....	146
5.1.3 政府扶贫开发绩效评估的研究现状.....	148
5.2 实证挖掘方法的选择	151
5.2.1 聚类分析的原理.....	151
5.2.2 聚类分析的意义.....	152
5.2.3 主要聚类算法比较.....	153
5.2.4 聚类分析在政府扶贫开发绩效评估指标设计应用的优势.....	159
5.3 实证挖掘过程步骤	160
5.3.1 数据准备与预处理.....	161
5.3.2 挖掘模型选取及运行.....	163
5.3.3 聚类结果分析.....	169
5.4 聚类结果对政府扶贫开发绩效评价指标设计的启发和建议	170
5.4.1 聚类分析结果知识汇总.....	170
5.4.2 聚类结果对政府扶贫开发绩效评价指标设计的启发.....	171
5.4.3 扶贫开发绩效评估指标设计的建议.....	173

第六章 结 语	177
6.1 研究结论与启示	177
6.1.1 研究结论.....	177
6.1.2 管理启示.....	180
6.2 研究贡献与研究局限	182
6.2.1 研究贡献.....	182
6.2.2 研究局限.....	183
6.3 研究展望	185
参考文献.....	187
致 谢.....	199

Contents

Chapter I Introduction	1
1.1 Preface.....	1
1.1.1 Study Background.....	1
1.1.2 Significance of the Study	6
1.2 Definition of the Key Concepts	9
1.2.1 Government Performance Evaluation.....	9
1.2.2 Government Performance Evaluation Indicator	15
1.2.3 Data Mining	17
1.3 Study Reviews.....	22
1.3.1 Study Reviews of Government Performance Evaluation Indicator	22
1.3.2 Study Reviews of Data Mining Application	37
1.3.3 Comments	40
Chapter II Study designation	42
2.1 Study Objectives and Study Clue.....	42
2.1.1 Study Objectives	42
2.1.2 Study Clue.....	43
2.2 Selection of Study Objection	46
2.2.1 Study Objection Alternatives	47
2.2.2 Reasons of Study Objection Selection.....	52
2.2.3 Study Objection Confirmation.....	52
2.3 Selection of the Study Techniques and Study Tools	54
2.3.1 Selection of Data Mining Methods	54
2.3.2 Selection of Data Pretreatment Methods	61
2.3.3 Selection of Data Mining Tools.....	64
2.4 Study Methods and Study Contents	71

2.4.1 Study Methods	71
2.4.2 Study Contents and Structural Arrangements	72
Chapter III Principal Component Regression Using in Designing the Model Indicator to Rationality of Land Use Structure Evaluating...	75
3.1 Theoretical Preparation and Business Understanding.....	75
3.1.1 Understanding of Land Use Structure Rationality	75
3.1.2 Significance of Land Use Structure Rationality Evaluating	77
3.1.3 Study Status of Land Use Structure Rationality Evaluating	79
3.1.4 Factors of Land Use Structure Changing.....	81
3.2 The Selection of Empirical Methods	84
3.2.1 The Main Statistical Methods	85
3.2.2 Principal Component Analysis.....	86
3.2.3 Ordinary Least Square	87
3.3 Empirical Process and the Steps	88
3.3.1 Data Collection	89
3.3.2 Data Preprocessing---Variables Standardized	89
3.3.3 Principal Component Analysis on Standardized Data	93
3.3.4 Regression and Prediction Using Land Use Intensity.....	98
3.4 Analysis of the Land Use Structure Rationality Evaluating Model Indicator	103
3.4.1 Characteristic of Land Use Proportion Coefficient Model Indicators	103
3.4.2 Predictive Capabilities of Land Use Proportion Coefficient Model Indicators.....	105
Chapter IV Decision Tree Model Using in Designing the Indicator of Remediation Evaluation to Idle Land in Construction Projects.....	107
4.1 Theoretical Preparation and Business Understanding.....	107
4.1.1 Understanding of Idle Land in Construction Projects.....	107
4.1.2 Urgency of Evaluation Study to Idle Land in Construction Projects.....	109

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 for delivery details.

厦门大学博硕士论文摘要库