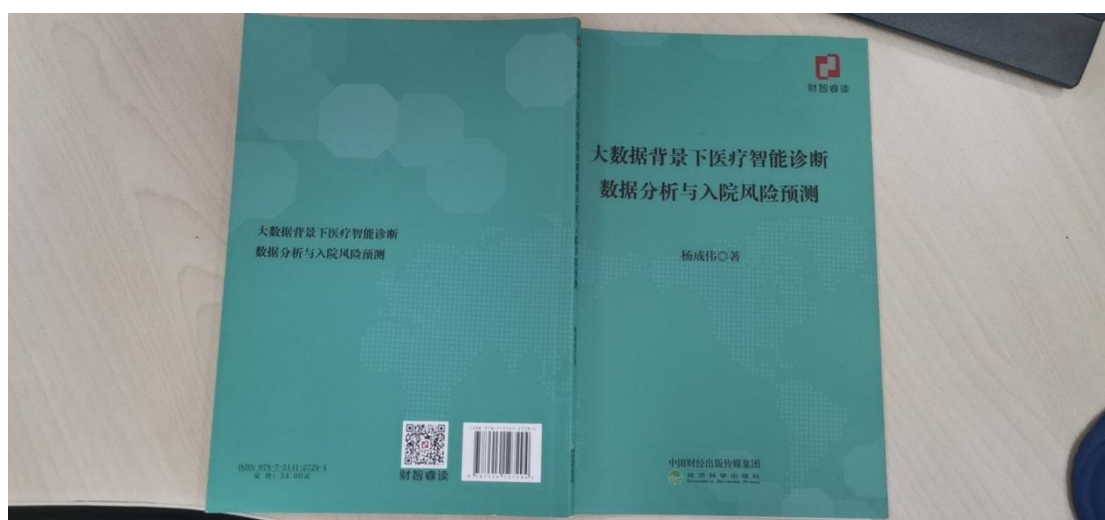
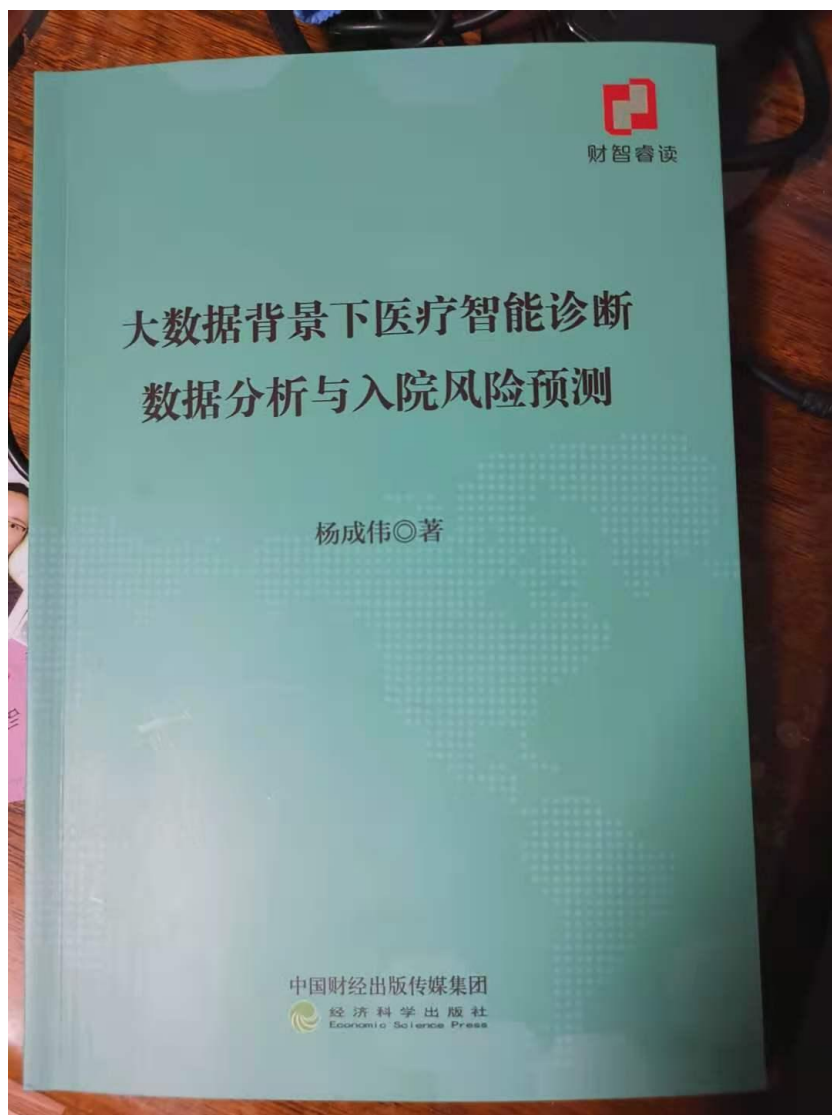






#### 作者简介

杨成伟，山东财经大学管理科学与工程学院讲师、硕士生导师。主要研究领域为数据挖掘、大数据商务分析、推荐算法等。先后主持参与国家自然科学基金课题，中央财政课题，山东省自然科学基金课题，山东省经信委课题，山东省教育厅课题，中国博士后课题，山东省计算机学会开放课题等10余项。参与面向对象分析与设计、J2EE框架开发技术、大数据分析与预测的MOOC视频建设，主持山东财经大学优质课程，山东财经大学信息系统实战课程，担任Hadoop大数据分析技术、Hadoop大数据实战等课程讲授。主要的社会兼职有中国计算机学会青年学会（CCF YOCSEF）济南AC委员兼学术秘书（2014-2016年度）、副主席（2017-2018年度）、中国计算机学会会员分部委员、ACM协会济南秘书长、中国计算机学会（CCF）会员代表、中国计算机学会计算机应用专委会委员、山东省人工智能学会委员、山东省计算机学会委员。近年来，发表学术论文20余篇，其中SCI、EI收录15篇，其他统计源文章5篇。软件著作权3项，专利2项。

## 大数据背景下医疗智能诊断 数据分析与入院风险预测

杨成伟 著

中国财经出版传媒集团  
经济科学出版社  
ECONOMIC SCIENCE PRESS

## 目录 CONTENTS

|                             |    |
|-----------------------------|----|
| 第1章 医疗数据挖掘与智能辅助诊断系统概述 ..... | 1  |
| 1.1 研究背景及意义 .....           | 1  |
| 1.2 研究目的及方法 .....           | 3  |
| 1.3 国内外研究现状 .....           | 4  |
| 1.4 本书的主要创新之处 .....         | 9  |
| 1.5 本书组织结构 .....            | 9  |
| 1.6 本章小结 .....              | 10 |
| 第2章 医疗数据挖掘方法与关键技术 .....     | 1  |
| 2.1 医疗数据的基本特征 .....         |    |
| 2.2 数据预处理方法 .....           |    |
| 2.3 医疗数据挖掘的基本流程 .....       |    |
| 2.4 医疗数据挖掘的关键技术 .....       |    |
| 2.5 本章小结 .....              |    |
| 第3章 基于统计方法的医疗数据分析与挖掘 .....  |    |
| 3.1 基于统计分析的模型及方法 .....      |    |
| 3.2 医疗数据挖掘的主要软件和工具 .....    |    |
| 3.3 基于心脏病数据的医疗统计分析与挖掘 ..... |    |
| 3.4 本章小结 .....              |    |
| 第4章 基于BP神经网络模型的医疗分类预测 ..... |    |
| 4.1 神经网络 .....              |    |

|      |                 |     |
|------|-----------------|-----|
| 4.2  | 心脏病分类预测实验过程     | 30  |
| 4.3  | 本章小结            | 63  |
| 第5章  | 基于随机森林模型的人院风险预测 | 64  |
| 5.1  | 随机森林            | 64  |
| 5.2  | 入院风险预测实验过程      | 69  |
| 5.3  | 模型学习及参数调优       | 80  |
| 5.4  | 模型评估            | 98  |
| 5.5  | 实验总结            | 109 |
| 5.6  | 本章小结            | 110 |
| 第6章  | 心脏病辅助诊断原型系统实现   | 111 |
| 6.1  | 背景及功能           | 111 |
| 6.2  | 开发环境            | 112 |
| 6.3  | 系统分析与设计         | 113 |
| 6.4  | 系统实现            | 114 |
| 6.5  | 系统问卷评估          | 115 |
| 6.6  | 本章小结            | 116 |
| 第7章  | 总结与展望           | 117 |
| 7.1  | 本书总结            | 117 |
| 7.2  | 不足与展望           | 118 |
| 附录   |                 |     |
| 附录1  | 调查问卷            |     |
| 附录2  | 对用户输入分词预处理代码    |     |
| 参考文献 |                 |     |

## 第 1 章

# 医疗数据挖掘与智能 辅助诊断系统概述

### 1.1 研究背景及意义

21 世纪,随着人口数量的快速增加,人口老龄趋势的日益严重,社会医疗资源已经求过于供,尤其是公共医疗资源,在医学领域主要表现为:每天医院里有大量病人不断涌入,需要医生给予医治,然而,技艺高超的医生数量极为短缺,造成许多疾病难以得到及时有效的医治而被延迟,给病人造成无法弥补的巨大的损失,也使医患关系更加紧张,最终的危害无疑是巨大的。所以,如何提升医生的医疗效率以及病情判断准确性成为亟待解决的问题。在此同时,随着新一代信息技术尤其是 Internet 大数据、人工智能、移动互联技术的快速发展,各个行业都积累了大量数据,如何从这些海量的、复杂的数据中提取出有价值的、能帮人们辅助分析决策的知识,已经成为一个亟须解决的任务。为了达到这的,人们逐步研究和开发了一种新型的数据处理的工具——数据挖掘(Data Mining),迅猛发展,已得到了广泛的接受与应用。数据挖掘是分析数据并从中获取有效信息的一种技术,它使用了统计科学中的假设检验等知识,以及机器学习中的部分算法,超出若干学科研究。将统计科学、数据仓库技术、机器学习、知识工程、人工智能、等等一些最新技术科学成果融合在一起,应用于社会的多个领域,治、商务分析、电子商务等,已经取得了很好的效果。

机器学习(machine learning)是近年来新兴的、热度非常