

学校编码: 10384

分类号_____密级_____

学号: 23220070153995

UDC_____

厦 门 大 学

博 士 学 位 论 文

基于互联网的决策理论与方法
的若干问题研究

Research on web-based decision-making theory and method

洪 文 兴

指导教师姓名: 李 茂 青 教授

专 业 名 称: 系 统 工 程

论文提交日期: 2010 年 月

论文答辩时间: 2010 年 月

学位授予日期: 2010 年 月

答辩委员会主席: _____

评 阅 人: _____

2010 年 月

厦门大学博硕士学位论文摘要库

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

厦门大学博硕士学位论文摘要库

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

我们正在离开信息时代，进入推荐时代。

互联网的高速发展，使得某些领域的信息，已经超出“充裕”、进入“泛滥”的处境。在推荐时代，如何从信息洪流中，发现和最大化数据的价值，实现信息各取所需、智能聚合，成为众多网络用户关心的问题。在此背景下，数据价值链分析和推荐系统应运而生。

数据价值链分析，就是探讨如何从数据中获得信息，再从信息中提取知识，进而将知识用于决策和管理，并最终形成完整的价值链。一个完整的数据价值链与数据应用场景密切相关，但一般包括四个具体实施环节，并构成一个循环：获取正确的数据；数据质量管理；从数据抽取信息，自信息中提取知识；知识管理和应用。

推荐系统是数据价值链分析的典型应用，发源于信息提取和信息过滤，是网络时代的主要决策方法和工具。推荐模型可以模式化的表示为一个优化模型，寻找用户信息和项目（商品）信息的最佳组合。基于内容的推荐、基于协同过滤的推荐和混合推荐是推荐系统的三大类型。其中，又以混合推荐系统最为流行。推荐算法是描述统计、聚类分析、关联分析、预测和优化等数据挖掘算法的有效延伸。对推荐系统的评价分为两个方面：推荐算法的评价和推荐系统的评价。虽然，推荐算法是推荐系统的核心，但是算法评价和系统评价并不完全一致。

借助系统工程和软件工程的方法，可以给出推荐系统的实施模型。这种实施模型，可以将推荐系统的应用范围从电子商务领域推广到其他信息处理领域，为数据价值链分析提供有效的工具，对实际工业应用有借鉴意义。

本文的创新研究成果主要体现在如下几个方面：

- (1) 从数据生命周期的角度，对数据价值链分析进行了扩展。知识管理并不是数据价值链的末端。在知识管理的基础上，还可以做知识分析等其他价值拓展服务。价值链的发展是由用户的具体需求确定的。

专家系统、决策支持系统、推荐系统都分别只是数据价值链的一种典型应用，相同之处在于三者都是对知识管理的拓展，因此都适用关联分析、聚类、预测和优化等基础数据分析方法；不同之处在于三者的知识口径不一，专家系统的知识来源于专家，群体决策支持系统的知识来源于群体，

推荐系统的知识来源于系统的所有用户。

数据价值链分析的基础是正确的数据的采集。研究以无线传感器网络中的数据处理为例，提出一种新的“预测—修正”模型，用以获取数据、提升数据准确度。

- (2) 针对推荐时代，总结了推荐系统在数据价值链分析的作用，首次针对网上商城提出了一个完整决策支持系统框架。这个框架包含了作为外部决策的搜索引擎优化模块和作为内部决策的推荐系统模块。通过这个框架，将决策支持系统和推荐系统有机的衔接起来。论文并进一步针对垂直型电子商务网站设计了一种新的混合型推荐系统框架，称为 *HRS-VCO*。该框架采用改进的 *X-means* 方法对用户和产品进行聚类，然后用 *Logistic* 回归对交易数据进行关联分析，最后使用线性信息融合模型对所有规则进行综合判断，给出合理的推荐结果。
- (3) 给出了推荐系统较为完整的评估体系。相比较其他的研究文献，研究归纳总结了现在流行的各类评价指标，包括算法准确度、多样性、覆盖率和新鲜度等。本文指出，许多研究文献由于对推荐算法和推荐系统的概念蕴含混淆，将推荐系统的评价等同于推荐算法的评价。研究修正了这种错误，并说明时下流行的离线评价和在线评价分类方法，体现了一种对应关系，即离线评价对应于算法评价，而在线评价则对应于系统实施评价。此外，本文为推荐系统实施修订了一个较为完整的评价体系，并提出了一种基于竞争框架的在线评价模型。
- (4) 对推荐系统案例进行综合分析，给出系统实施的螺旋模型，延伸了推荐系统的应用范围。本文收集和整理了不同工业背景的典型推荐系统应用，整理出系统实施的参考模型。针对现有大部分推荐系统局限于电子商务的情形，用实际案例延伸了推荐算法和推荐系统的应用领域，例如人才信息推荐等。

关键词：基于互联网的决策；数据价值链；推荐系统

Abstract

Leaving the information age, we are entering into the era of recommendation.

The rapid development of internet has made some areas of information go beyond what is “adequate” and turned into “flood”. In the recommendation era, identifying and maximizing the value of data in the torrent of information to realize what it is needed for and gather intelligence is a concern of many users. In this context, data value chain analysis and recommendation system came into being. Data value chain analysis is to explore how to get information from data, then extract knowledge from information for decision-making and management, and eventually form a complete value chain. A complete data value chain is closely related with the data scenarios and consists of four specific implementation links that form a cycle: obtaining the correct data, data quality management, extracting information from the data, extracting knowledge from information, knowledge management and application.

Originated from information extraction and filtration, recommendation systems are a typical application of data value chain analysis. They are the main method and tool for decision making in the network era. Recommendation model can be stereotyped as an optimization model which seeks the best combination of user information and item (product) information. Content based recommendation, collaborative filtering based recommendation and hybrid recommendation are the three major types of recommendation systems with the hybrid recommendation being the most popular type. Recommendation algorithms are an effective extension of data mining algorithms such as descriptive statistics, cluster analysis, correlation analysis, forecasting and optimization. The evaluation of recommendation systems is divided into two areas that are: the evaluation of the recommendation algorithms and the evaluation of the recommendation systems. Although the recommendation algorithm is the core of a recommendation system, the algorithm evaluation and system evaluation are not entirely consistent.

With the help of system engineering and software engineering, the implementation model of recommendation systems can be given. This implementation model can expand the application range of recommendation systems from e-commerce to other areas of information processing, serve as an effective tool for data value chain analysis and provide practical reference for industrial application. This innovative research results mainly in the following aspects:

- (1) It expanded the data value chain analysis from the data life-cycle perspective.

Knowledge management is not the end of the data value chain. Additional value excavating services like knowledge analysis can be provided based on knowledge management. The development of value chain is determined by the user's specific needs. Expert system, decision support system, recommendation systems are each a typical application of data value chain. In common is that all the three are on the expansion of knowledge management and, as a result, all applicable to basic data analysis methods such as correlation analysis, clustering, forecasting and optimization. Their difference lies in the different sources of knowledge where knowledge of expert system is from experts, knowledge of group decision support system from groups, knowledge of recommendation systems from all users. The basis of data value chain analysis is correct data collection. Research in data processing in wireless sensor networks, for example, proposes a new "forecasting - revision" model that is used to collect data and improve data accuracy.

- (2) For the recommendation era, it summed up recommendation systems' role in data value chain analysis and proposed a complete decision support system framework first ever for online mall.

This framework includes search engine optimization module for external decision and recommendation system module for internal decision. Through this framework, the decision support systems are organically joined with recommendation systems. This paper further designed a new mix-type recommendation system framework for vertical e-commerce web sites. This module clusters users and products by using an improved $X - means$ approach, then performs correlation analysis of transaction data by using *Logistic*, and finally makes an integrated determination of all the rules by using linear information fusion models to give a reasonable result.

- (3) It gave a more complete evaluation system for recommendation systems.

Compared in other research literature, research summarized various types of evaluation indicators that are now popularly used including accuracy, degree of diversity, coverage and freshness of the algorithm, and innovatively pointed out that many papers take the evaluation of recommendation systems as identical to the evaluation of recommendation algorithms as a result of conceptual confusion between recommendation systems and recommendation algorithms. This paper fixed the error and explained that the currently popular classification methods for off-line evaluation and on-line evaluation embody a corresponding relationship

in which off-line evaluation corresponds to algorithm evaluation whereas on-line evaluation corresponds to system implementation evaluation. In addition, this paper, for the first time, used analytical hierarchy process to revise and recommend a more complete evaluation system for recommendation systems.

- (4) Comprehensive analysis of recommendation system cases provided an implementation model for reference and expanded the application range of recommendation systems.

This paper collected and collated typical recommendation system applications in different industrial backgrounds and provided a referential system implementation model. As most existing recommendation systems are limited to e-commerce, the practical cases expanded the application range of recommendation algorithms and recommendation systems to, for instance, talent recommendation.

Key Words: Web-based decision-making, data value chain, recommender system

厦门大学博硕士学位论文摘要库

目 录

摘要	I
Abstract	III
第一章 绪论	1
1.1 研究的背景和意义.....	1
1.1.1 研究的背景.....	1
1.1.2 研究的意义.....	2
1.2 互联网应用发展综述	2
1.3 决策理论的研究现状	4
1.3.1 决策科学的发展.....	4
1.3.2 决策支持系统的发展.....	6
1.4 基于互联网的决策理论综述.....	7
1.5 本文的主要工作和结构安排.....	8
第二章 决策理论的数据价值链分析	11
2.1 数据价值链的形成.....	12
2.2 获取准确的数据	13
2.2.1 数据分类和来源.....	13
2.2.2 数据管理技术	14

2.3 数据质量管理	17
2.3.1 数据的质量特征.....	17
2.3.2 案例：用“预测-修正”数据管理方法提升传感网数据质量.....	18
2.4 从数据抽取信息、提炼知识	22
2.5 知识管理和应用	24
2.5.1 专家系统.....	25
2.5.2 决策支持系统.....	26
2.5.3 推荐系统.....	27
第三章 经典决策数据分析方法	29
3.1 描述统计	29
3.2 关联规则挖掘	30
3.2.1 定义.....	30
3.2.2 算法描述.....	32
3.3 决策树分类	33
3.3.1 定义.....	34
3.3.2 算法描述.....	35
3.4 聚类分析	39
3.4.1 定义.....	39
3.4.2 K 均值聚类模型	39
3.5 预测与序列模式	40
3.5.1 序列模式的数据挖掘.....	40

3.5.2 Logistic 回归	41
3.6 优化与遗传算法	42
3.7 本章小结	43
第四章 基于X-means聚类的混合型推荐系统.....	45
4.1 推荐系统的研究动机和数学描述.....	45
4.1.1 推荐系统的研究动机.....	45
4.1.2 推荐系统的数学描述.....	46
4.2 推荐方法	47
4.2.1 基于内容的推荐方法.....	47
4.2.2 基于协同过滤的推荐方法.....	51
4.2.3 混合推荐方法	54
4.3 基于X均值聚类和信息融合的混合型推荐系统.....	55
4.3.1 改进的X均值聚类方法	56
4.3.2 信息融合方法	59
4.4 系统评估实验	60
4.4.1 实验数据集描述.....	60
4.4.2 实验结果和讨论.....	62
4.5 小结	63

第五章 推荐系统的评价体系	65
5.1 推荐算法的评价体系	65
5.1.1 单项预测准确度	67
5.1.2 列表预测准确度	69
5.1.3 排序准确度	71
5.1.4 推荐满意度的评价	73
5.2 一种基于竞争框架的在线评价模型	75
5.2.1 区分离线和在线评价	75
5.2.2 改进的基于竞争框架的在线评价模型	76
5.3 从用户的角度制定综合评价体系	78
5.4 本章小结	80
第六章 推荐系统的实施	81
6.1 推荐系统案例分析	81
6.2 推荐系统实施模型	84
6.3 用于人才信息的推荐系统	86
6.3.1 系统定位与总体目标	86
6.3.2 系统概念模型	86
6.3.3 数据价值链分析流程及核心问题分析	87
第七章 总结与展望	91
7.1 总结	91

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.

厦门大学博硕士学位论文摘要库