

# 文本复制检测报告单(全文标明引文)

№:ADBD2018R\_2018053015312720180530154846440174249827

检测时间:2018-05-30 15:48:46

检测文献: 53141211\_陈午阳\_计算机科学与技术(网络与信息安全)\_专车平台客户需求挖掘与实现

作者: 陈午阳

检测范围: 中国学术期刊网络出版总库

中国博士学位论文全文数据库/中国优秀硕士学位论文全文数据库

中国重要会议论文全文数据库

中国重要报纸全文数据库

中国专利全文数据库

图书资源

优先出版文献库

大学生论文联合比对库

互联网资源(包含贴吧等论坛资源)

英文数据库(涵盖期刊、博硕、会议的英文数据以及德国Springer、英国Taylor&Francis 期刊数据库等)

港澳台学术文献库

互联网文档资源

CNKI大成编客-原创作品库

个人比对库

时间范围: 1900-01-01至2018-05-30

## 检测结果

总文字复制比: 13.1%

跨语言检测结果: 0%

去除引用文献复制比: 7.7%

去除本人已发表文献复制比: 13.1%

单篇最大文字复制比: 6.8% (《推荐系统》基于标签的用户推荐系统 - Thinkgamer博客 - CSDN博客)

重复字数: [2995]

总段落数: [7]

总字数: [22869]

疑似段落数: [5]

单篇最大重复字数: [1546]

前部重合字数: [191]

疑似段落最大重合字数: [2005]

后部重合字数: [2804]

疑似段落最小重合字数: [72]



指标: ☐ 疑似剽窃观点 ☒ 疑似剽窃文字表述 ☐ 疑似自我剽窃 ☐ 疑似整体剽窃 ☐ 过度引用

表格: 0 公式: 0 疑似文字的图片: 0 脚注与尾注: 0

0% (0) 中英文摘要等 (总2328字)

7.2% (191) 第1章绪论 (总2655字)

11.4% (417) 第2章数据库营销与推荐技术 (总3660字)

0% (0) 第3章数据库与关联规则挖掘 (总6203字)

47.3% (2005) 第4章一种基于标签的推荐算法的实现 (总4243字)

12.1% (310) 第5章信息安全问题 (总2559字)

5.9% (72) 第6章总结与展望 (总1221字)

(注释: 无问题部分 文字复制比部分 引用部分)

## 1. 中英文摘要等

总字数: 2328

相似文献列表 文字复制比: 0%(0) 疑似剽窃观点: (0)

原文内容 红色文字表示存在文字复制现象的内容; 绿色文字表示其中标明了引用的内容

### 摘要

专车平台客户需求挖掘与实现

充分预测客户需求以提升服务质量一直以来是企业间争取客户资源的重要课题, 互联网+时代, 数据库营销思想深入各商

业领域，共享经济模式方兴未艾。网约车专车平台作为互联网+时代共享经济背景下的新兴产业，对数据挖掘与分析有着巨大的潜在需求。

推荐系统作为数据挖掘与分析在商业领域的重要应用，目前主要依靠三种推荐算法，即基于内容的推荐算法、协同过滤算法和基于知识的推荐算法。本文在分析现有推荐系统和推荐技术的基础上，探讨了对于专车平台适用性较高的几种推荐算法，并通过python实现了一种基于标签的推荐算法，同时通过频繁项集和相似函数挖掘出部分关联规则，为之后基于知识的推荐进行准备。最后分析了专车平台推荐系统可能涉及到的信息安全问题并提出了一些解决方案。

接下来的工作需要在现有数据库基础上对数据进行扩充并检验获得的关联规则是否发生变化，进而利用这些规则进行基于知识的推荐。更长远的工作应包含通过Hadoop云计算平台等技术实现大数据下基于协同过滤和知识的推荐。

关键字：专车平台，推荐技术，关联规则，信息安全

Abstract

Customer demand mining and implementation for online car-hailing platforms

Properly predicting customer needs to improve service quality has always been an important issue for enterprises to win customer resources. In the era of Internet+, database marketing thoughts have penetrated into various business fields. Sharing economy is developing rapidly. As an emerging industry under the background of the Internet + era sharing economy, the Internet dating car car platform has enormous potential for data mining and analysis.

The recommendation system, as an important application of data mining and analysis in the business field, currently relies on three recommendation algorithms, namely content-based recommendation algorithm, collaborative filtering algorithm, and knowledge-based recommendation algorithm. Based on the analysis of the existing recommendation system and recommendation technology, this paper discusses several recommended algorithms for higher applicability of the car platform, and implements a tag-based recommendation algorithm through python, while passing frequent item sets and similar functions. Dig out some of the association rules to prepare for knowledge-based recommendations. Finally, it analyzes the information security problems that the car platform recommendation system may involve and proposes some solutions.

The following work needs to expand the data on the basis of the existing database and test whether the obtained association rules have changed, and then use these rules to conduct knowledge-based recommendation. Longer-term work should include recommendations for collaborative filtering and knowledge based on big data through technologies such as the Hadoop cloud computing platform.

Keywords：Online car-hailing platforms，Recommended algorithm，Association rules，Information security

目录

第1章绪论	4
1.1 网约车专车平台的诞生与发展	4
1.1.1 互联网+与共享经济	4
1.1.2 传统巡游出租车行业颓势难阻	5
1.1.3 专车平台应运而生	6
1.2 专车平台客户需求挖掘的必要性	7
第2章数据库营销与推荐技术	9
2.1 数据库营销	9
2.1.1 数据库营销是时代的选择	9
2.1.2 数据库营销现有技术	9
2.2 推荐技术	11
2.2.1 推荐技术的发展	11
2.2.2 推荐技术的广泛应用	11
2.3 三种流行的推荐算法	12
2.3.1 基于内容的推荐算法	12
2.3.2 协同过滤算法	13
2.3.3 基于知识的推荐算法	13
第3章数据库与关联规则挖掘	15
3.1 数据需求	15
3.1.1 基于标签的推荐算法的数据需求	15
3.1.2 关联规则挖掘的数据需求	15
3.2 数据来源分析	16
3.2.1 客户端录入	16
3.2.2 过往订单记录	16

3.2.3 合作平台数据集 .....	16
3.2.4 社交渠道 .....	17
3.3 数据库的建立 .....	17
3.3.1 表项确定 .....	17
3.3.2 数据库界面 .....	18
3.4 专车平台关联规则挖掘 .....	19
3.4.1 关联规则的典型应用 .....	19
3.4.2 思想与算法 .....	20
3.4.3 频繁项集 .....	23
3.4.4 关联规则 .....	24
第4章一种基于标签的推荐算法的实现 .....	25
4.1 标签系统的典型代表 .....	25
4.2 用户如何打标签 .....	26
4.2.1 标签流行度 .....	26
4.2.2 标签选取 .....	27
4.3推荐算法 .....	28
4.3.1.试验设置 .....	29
4.3.2.一个简单的算法 .....	31
4.3.3 算法的改进 .....	33
4.3.4.标签清理 .....	34
4.4 标签推荐 .....	37
4.4.1 为什么给用户推荐标签 .....	37
4.4.2 标签推荐的四种简单算法 .....	37
第5章信息安全问题 .....	40
5.1 传统数据安全问题 .....	40
5.1.1 数据安全事故 .....	40
5.1.2 数据泄露 .....	40
5.2 推荐系统带来的新威胁 .....	42
5.3 专车平台数据安全方案 .....	43
5.3.1 安全的设计架构 .....	43
5.3.2 数据库安全技术 .....	43
第6章总结与展望 .....	45
6.1 成果归纳 .....	45
6.2 问题与解决方案 .....	45
6.2.1 数据量与可靠度不足 .....	45
6.2.2 关联规则及推荐算法针对性低 .....	46
参考文献 .....	48
致谢 .....	49

2. 第1章绪论		总字数：2655
相似文献列表 文字复制比：7.2%(191) 疑似剽窃观点：(0)		
1	滴滴专车平台的竞争策略分析 周换换(导师：沈固朝) - 《南京大学博士论文》 - 2016-05-01	3.6% ( 96 ) 是否引证：是
2	Uber在中国 毛军; - 《商业文化》 - 2016-07-25	2.1% ( 56 ) 是否引证：否
3	1322666-季凌钰-专车市场现状分析及发展前景-汽车服务工程 季凌钰 - 《大学生论文联合比对库》 - 2017-04-05	2.1% ( 55 ) 是否引证：否
4	11469845_钟健_滴滴打车运营管理现状分析 钟健 - 《大学生论文联合比对库》 - 2017-05-10	2.1% ( 55 ) 是否引证：否
5	基于合约理论的网约车商业模式分析——以滴滴出行为例	1.7% ( 45 )

刘飞 - 《大学生论文联合比对库》 - 2017-04-28	是否引证：否
6 基于合约理论的网约车经营合约分析——以滴滴出行为例	1.7% ( 45 )
刘飞 - 《大学生论文联合比对库》 - 2017-05-25	是否引证：否
7 重新定义集团财务管理	1.2% ( 32 )
刘鸥; - 《首席财务官》 - 2016-11-10	是否引证：否

原文内容 红色文字表示存在文字复制现象的内容; 绿色文字表示其中标明了引用的内容

## 第1章绪论

### 1.1 网约车专车平台的诞生与发展

网约车专车平台作为互联网+时代背景下诞生的新兴产业，其发展过程充分体现了市场对资源配置的决定性作用。本节重点阐述国内网约车专车平台产生背景和发展历程。

#### 1.1.1 互联网+与共享经济

伴随着1995年Amazon, Yahoo两大互联网巨头的成立，近25年来，互联网渗透进人们包括社交、生产、消费等在内的几乎所有领域，极大的影响了人们的日常生活和消费习惯。互联网从简单的资讯门户网站开始，逐渐实现了移动终端多样化与智能终端普及，目前已能够满足人们方方面面的消费需求，以社交网络，电子商务，在线出游为代表的行业从中受益匪浅。伴随着消费互联格局的确立，互联网+与web2.0时代已经到来，网络与信息技术的飞速发展使虚拟化进程从个人向企业蔓延，企业逐渐成为互联网的重要参与者。行业经验、生产与消费渠道、网络、产品价值等壁垒使得传统消费互联网巨头不再拥有垄断级别的优势，并购与合作并存、自主与发展齐飞，成为传统企业搭乘互联网+这趟快车的主要方式。

与此同时，随着互联网技术的广泛应用和社交网络生态逐渐成熟，共享经济（sharing economy）这一全新商业模式已初露锋芒，众多共享网站如雨后春笋般层出不穷。政府的高度重视则为这种新兴模式提供了制度保障，2016年政府工作报告指出要大力推动包括共享经济等在内的“新经济”领域快速发展。目前，共享经济模式已经渗透进了社会生活和行业生产的方方面面，有力地提高了人民的生活质量和推进了产业的创新。在互联网+的时代背景下，在全面深化改革的关键时期，物联网、云计算的发展，传统与移动终端的普及与升级，都为共享模式的应用与创新提供了更多可能，充分利用共享模式进行商业创新，是众多新兴产业的战略性问题。

#### 图1-1 共享经济市场主要集中领域

#### 1.1.2 传统巡游出租车行业颓势难阻

传统巡游出租车作为城市公共交通的重要参与者，其在国内的发展已经历了百余年历程。尽管长时间的积累已使传统出租车行业拥有了较为完善的行业规范与监管体系，但其弊端也在社会主义新时代逐渐显现，比较突出的是以下三个方面：

- 1.出租车数量管制虽然保证了行业稳定和道路交通正常运转，但随之而来的供给不足问题亟待解决，高峰时段市民一车难求，司机为追逐高利润拒载挑客的情况饱受诟病；
- 2.行业利益分配不均，出租车公司的收入来源于与运营实际收入无关的基础份子钱，运营风险则由司机独自承担，间接导致了出租车司机成为劳动强度极大收入却低于社会平均水平的职业；
- 3.行业缺乏规范的退出机制，地方交通部门对出租车公司的处罚形式局限于约谈、整改、减少或禁止其获得新运力指标等，毒瘤公司无法根除，严重危害行业整体健康。

#### 1.1.3 专车平台应运而生

网约车专车作为互联网+与传统出租车行业碰撞的产物，其不足十年的发展历程可谓一波三折，不过伴随一系列相关文件的出台，专车合法化已经基本实现。

#### 1.网约车专车平台发展大事记

2010年5月，易到用车面市。

2012年8-9月，快的、滴滴打车面市。

2013年5月，国内各种打车平台已有近50种，随之而来的监管问题日渐突出。

2014年2月，国际打车软件优步（Uber）进入中国。

2014年6-9月，多个平台陆续推出“专车”服务，并从一线城市向全国各地普及。

2015年1月8日，交通部首次在正式场合使用“专车”一词，专车对道路交通的积极意义得到认可。

2015年10月8日，上海市交通委向滴滴快的专车平台颁发网络约租车平台经营资格许可，这是国内第一张专车平台牌照。

。

2016年7月28日，《网络预约出租汽车经营服务管理暂行办法》出台，“专车合法化”终被认可

#### 2. 专车平台竞争优势

网约车专车作为新兴产物，要进入出租车行业争夺传统巡游出租车公司的既得市场，竞争优势必不可少，其现有的核心优势大致包含以下几点：

- （1）可选车型多，车况普遍较高，乘客可根据喜好预约车辆，体验惬意而个性化的旅程；
- （2）专车司机的收入普遍与乘客评价正相关，服务质量得到保障；
- （3）乘客预约专车后，通过专车平台实时获取预约车辆位置信息，使等车时间“可视化”；
- （4）采用线上支付，方便快捷；

(5) 服务完成后可通过平台查看行车路线及费用,降低了恶意绕路等欺骗行为出现的概率

## 1.2 专车平台客户需求挖掘的必要性

在中国特色社会主义进入新时代,社会主要矛盾发生变化,社会生产力达到一定水平,总体小康已经实现的大背景下,无论线上还是线下,无论出行还是餐饮,满足基本需求已不再是服务行业的市场主流,以人为本、以用户体验为核心的理念成为越来越多企业的基本价值观之一。

在逐步进入规范化的网约车行业,市场乱象将不复存在,专车平台市场前景广阔,越来越多的公司与平台加入到“分蛋糕”的行列中,市场竞争日趋激烈,南京大学周焕焕以国内专车行业的龙头企业滴滴专车作研究对象,通过调查问卷的方式,分析专车行业竞争现状,从多个角度分析平台所处竞争环境,进而为滴滴专车平台提出竞争策略规划与建议,搭乘移动互联网“快车”,积极遵守与践行相关政策与意见,最大化规避竞争风险,最终达成让出行更美好的战略目标[1]。不难察觉的是,服务是专车的最大优势与竞争核心。客户日益增长的完善服务需要与专车平台不具体、不准确的客户数据是该服务的根本矛盾与主要矛盾,能否妥善处理好这一矛盾是检验一个公司是否能从新时代社会主义市场经济下企业间竞争中脱颖而出的试金石。显然,不具体不准确的数据是这一矛盾的主要方面,尤其在面对高端商务类客户时,由于该类客户往往具有单次服务时间较长,服务需求频繁等特点,如果能在服务开始前通过数据挖掘与分析对具体客户潜在的服务偏好、个人兴趣等作出分析,以相关信息为导向,要求服务者——即平台签约司机,在服务前进行有针对性的学习,并分配最为匹配的资源(主要是车辆),将使服务效果实现飞跃性提升。在数据积累和技术完善达到一定水平后,可将这种推荐技术普及至更广泛的客户群体中,这在竞争日趋激烈的专车市场将成为企业的核心优势。

然而相比于衣食住行中其他三个方面:淘宝、京东等C2C、B2C平台的商品推荐系统,携程、去哪儿等旅行网站的酒店推荐系统,美团、口碑等消费平台的商户推荐系统,客户需求挖掘的本领还未在各专车平台得以广泛彰显。本文的二、三章将在分析这些现有推荐技术的基础上重点论述如何建立专车平台自己的推荐系统。

### 指标

#### 疑似剽窃文字表述

1. 2015年10月8日,上海市交通委向滴滴快的专车平台颁发网络约租车平台经营资格许可,这是国内第一张专车平台牌照。

## 3. 第2章数据库营销与推荐技术

总字数: 3660

相似文献列表 文字复制比: 11.4%(417) 疑似剽窃观点: (0)

1	改进的关联规则算法在慢性病数据挖掘中的研究 杨余垒(导师: 张云华) - 《浙江理工大学博士论文》 - 2017-03-07	4.2% ( 155 ) 是否引证: 是
2	网络数据库营销的优势_东狮营销_整合网络营销专家 - 《网络 ( <a href="http://blog.tianya.c">http://blog.tianya.c</a> )》 - 2016	3.1% ( 115 ) 是否引证: 否
3	东狮营销_整合网络营销专家 - 《网络 ( <a href="http://blog.tianya.c">http://blog.tianya.c</a> )》 - 2016	3.1% ( 115 ) 是否引证: 否
4	企业经典管理知识汇总&#40;中国质量管理在线总群版&#41;- 豆丁网 - 《互联网文档资源 ( <a href="http://www.docin.com">http://www.docin.com</a> )》 - 2016	3.1% ( 115 ) 是否引证: 否
5	网络营销的基本知识_图文 - 《互联网文档资源 ( <a href="http://wenku.baidu.c">http://wenku.baidu.c</a> )》 - 2016	3.1% ( 115 ) 是否引证: 否
6	专业企业管理知识全面汇编【白金珍藏版】 - 豆丁网 - 《互联网文档资源 ( <a href="http://www.docin.com">http://www.docin.com</a> )》 - 2017	3.1% ( 115 ) 是否引证: 否
7	企业经典管理知识汇总&#40;中国质量管理在线总群版&#41;- 豆丁网 - 《互联网文档资源 ( <a href="http://www.docin.com">http://www.docin.com</a> )》 - 2017	3.1% ( 115 ) 是否引证: 否
8	电商企业基于数据进行精准营销的探讨 罗红梅;- 《武汉商业服务学院学报》 - 2013-06-28	3.1% ( 115 ) 是否引证: 否
9	[NO&#46;EFU-8-网络营销]网络视频营销漫路求索 - docin.com豆丁网 - 《互联网文档资源 ( <a href="http://www.docin.com">http://www.docin.com</a> )》 - 2012	2.7% ( 98 ) 是否引证: 否
10	基于移动设备端音乐平台的实时个性化推荐实现方法比较 张燕 - 《大学生论文联合比对库》 - 2017-05-18	2.7% ( 98 ) 是否引证: 否
11	网络数据库营销研究 乐斌辉(导师: 任天飞) - 《湘潭大学硕士论文》 - 2006-11-24	2.6% ( 95 ) 是否引证: 否
12	我国商业银行信用卡数据库营销研究 韩霄(导师: 陈晶萍) - 《哈尔滨工程大学硕士论文》 - 2007-05-01	2.3% ( 83 ) 是否引证: 否
13	我国商业银行信用卡数据库营销研究-百度文库	2.3% ( 83 )



- 《互联网文档资源 ( <a href="http://wenku.baidu.c">http://wenku.baidu.c</a> ) 》 - 2012	是否引证：否
14 营销渠道管理 - 豆丁网	2.3% ( 83 )
- 《互联网文档资源 ( <a href="http://www.docin.com">http://www.docin.com</a> ) 》 - 2015	是否引证：否
15 浅谈数据库营销在邮政报刊发行中的应用	2.3% ( 83 )
周晓燕;张瑞凤;陈军须; - 《当代经济》 - 2011-04-08	是否引证：否
16 浅谈企业客户管理对市场营销的影响	2.3% ( 83 )
付波; - 《今日科苑》 - 2011-05-23	是否引证：否
17 试论市场营销中的客户档案管理	2.3% ( 83 )
丁伟; - 《中国集体经济》 - 2012-01-25	是否引证：否
18 电子商务环境下企业市场营销方式的转变	2.3% ( 83 )
孙永建; - 《才智》 - 2013-09-15	是否引证：否
19 数据库的“百亿蓝海”	2.1% ( 77 )
熊莉; - 《成功营销》 - 2011-04-15	是否引证：否
20 数据挖掘技术在营销策略中的应用	2.1% ( 77 )
刘国香; - 《中国商贸》 - 2011-10-21	是否引证：否
21 浅谈企业客户关系管理	2.1% ( 76 )
陈利; - 《发展》 - 2011-01-05	是否引证：否
22 数据化营销拉动业绩成长	1.9% ( 71 )
- 《医学美学美容(财智)》 - 2011-08-15	是否引证：否
23 数据库营销在企业中的应用研究	1.9% ( 71 )
董国民; - 《企业导报》 - 2012-03-15	是否引证：否
24 我国芒果产业营销对策浅析	1.3% ( 46 )
许能锐;张德生;罗海燕; - 《热带农业工程》 - 2010-12-26	是否引证：否
25 新媒体背景下的出版营销之道探析	1.1% ( 42 )
谢文辉; - 《出版发行研究》 - 2014-05-15	是否引证：否
26 基于Web使用挖掘的协同过滤推荐算法研究	1.0% ( 37 )
蔡浩(导师：贾宇波) - 《浙江理工大学博士论文》 - 2010-03-11	是否引证：否

原文内容 红色文字表示存在文字复制现象的内容; 绿色文字表示其中标明了引用的内容

## 第2章数据库营销与推荐技术

数据库营销作为数据分析与挖掘技术在商业领域的重要应用，已经历了20余年的发展，相关成果极大的推动了产业创新与商业模式升级。推荐技术则是将数据分析的结果通过多种手段加以聚类实现个性化信息定制的智能系统。本章主要回顾并分析了现有数据库营销技术与推荐技术的思想与应用。

### 2.1 数据库营销

#### 2.1.1 数据库营销是时代的选择

为准确而高效的预测客户的服务偏好，时下已深入各销售行业的数据库营销是一个很好的模式。数据库营销 ( Database Marketing ) 是为了实现接洽、交易和建立客户关系等目标而建立、维护和利用顾客数据库与其他顾客资料的过程。它是在 Internet与Database技术发展上逐渐兴起和成熟起来的一种市场营销推广手段，主要目的是通过收集和积累消费者大量的信息，经处理后预测消费者有多大可能去购买某种产品，以及利用这些信息给产品精确定位，有针对地制作营销信息以说服消费者进行消费[2]。相较于大数据、云计算等相对新兴的技术，数据库营销已至少有20年的历史，沈池俊于1999年发表于商业经济与管理期刊的《对数据库营销功能的战略思考》及Peter C.Verhoef和JannyC.Hoekstra同年发表的《Status ofDatabase Marketingin the Dutch Fast Moving Consumer Goods Industry》是对数据库营销模式的早期探索。经过多年的发展，数据库营销不论在数据库建立还是数据分析方面都有已积累了大量实战经验。

#### 2.1.2 数据库营销现有技术

##### 1.数据库营销架构图

##### 图2-1 数据库营销架构

##### 2.数据挖掘算法与技术的广泛使用

与此同时，基于客户数据库的各种分析算法的相关研究同样成果丰富，大量成熟的算法可供专车平台客户需求挖掘所使用。在CSDN上《十大数据挖掘算法及详解》一文中对现有的算法进行了概括，而各类文献则提出了相关拓展。关联规则挖掘是数据挖掘的主要分支之一,其核心思想一言以蔽之，就是从大量数据中发现数据项之间隐藏的相关关系由于挖掘产生的规则形式规范、观察方便,关联规则挖掘成为了应用最为广泛的挖掘技术之一[3]。浙江理工大学杨余垒为了有效利用医疗数据为预防和管控慢性病提供科学依据,选取了慢性病之一的高血压进行数据挖掘方面的研究，其主要探究高血压患者体征与心血管风险水平之间的相关性,以及高血压与其他慢性病之间的关联性,系统能够探究慢性病医疗数据背后隐藏的知识,辅助医生决策,具有很高的实用价值。而武汉大学蔡皎洁对基于语义集成的客户知识挖掘进行了研究，对客户数据的语义异构问题进行了处理，为挖掘工具的有效运行的打好了基础,提升了客户知识模式获取的有效性[4]。此外，北京邮电大学马万智提出了基于本体的

数据模型概念。在中国电信客户关系管理系统中引入了基于本体的数据模型,建立了一整套数据信息本体体系对客户信息进行处理,该体系能被相关从业人员随时随地共享与使用[5]。

由此可见,基于客户数据库的数据库营销及其关联算法在国内外研究和应用领域都已积累大量理论和实践经验,为挖掘与实现专车平台客户需求,需要在此基础上确定数据库各表及其内属性,做好数据准备工作,选取时间和空间上与之匹配的算法,最终得出高准确性与普适性的结论。

## 2.2 推荐技术

### 2.2.1 推荐技术的发展

推荐系统是一种智能个性化信息服务系统,其主要功能是通过用户建模技术对用户的信息需求进行描述,并根据用户模型通过各种推荐算法与推荐技术实现高度个性化的信息定制[5]。推荐系统的早期探索可以追溯到1995年3月斯坦福大学的MarkoBalabanovic等人在美国人工智能协会上推出的LIRA——一个相对完善的个性化推荐系统。在之后的二十余年中推荐系统快速发展,其用户需求驱动、主动服务和信息个性化程度高等优点使之在电子商务、线上学习等领域得以广泛普及,目前已被视为最有前途的信息个性化技术发展方向[6]。本章将简要分析推荐技术在数据库营销领域的现有应用,第3、4章将在此基础上从算法思想、典型应用、算法拓展、知识挖掘等方面论述并实现适合专车平台的推荐算法。

### 2.2.2 推荐技术的广泛应用

今日头条、知乎、抖音等新兴的新闻、论坛、视频类软件席卷全国,推荐技术在其中功不可没。其实,这些软件内的内容和传统同类软件差异不大,不同之处在于,百度贴吧、天涯这些传统的论坛是通过看得见的墙将用户加以区分和聚集,达到聚集相同兴趣用户的目的,而大多新兴的软件则通过推荐算法这双看不见的手完成了这一工作。

推荐系统的目的是通过不同的媒介将用户的兴趣和物品相关联。在云计算和大数据等技术飞速发展的背景下,目前较为主流的推荐算法可归纳为以下三种:

- 1.基于内容的推荐算法
- 2.协同过滤推荐算法
- 3.基于知识的推荐算法

#### 图2-2 推荐系统联系用户和物品的几种途径

当然,推荐算法并不局限于上述三大类,大量机器学习、数据挖掘中的方法往往也被应用于推荐技术中。在实际应用时往往会同时使用多种算法使推荐系统具备更好的健壮性。

本章余下几节将简单介绍上述三种算法的思想和应用。

## 2.3 三种流行的推荐算法

### 2.3.1 基于内容的推荐算法

一言以蔽之,基于内容的推荐算法就是先找出客户喜欢过的物品,然后为其推荐与这些物品相似的物品。

基于内容的推荐算法又被称为基于item的系统过滤推荐算法,该类算法的基本思想是推荐给用户在内容上与其喜欢和曾关注过的商品类似的商品。举例来说,某用户曾观看《蜘蛛侠》和《美国队长》,此时基于内容的推荐算法发现《奇异博士》和《复仇者联盟3》等产品,其与该客户曾观看的两部电影在内容等方面有很大关联性,于是将《奇异博士》和《复仇者联盟3》进行推荐,这种方法看似最为简单,但也具备其独特优势,即避免了Item的冷启动问题。

[注]冷启动:如果一个物品从未被关注,则推荐算法很少能够推荐该物品。

然而其弊端同样显而易见,推荐的Item重复几率高,这也是新闻推荐系统较少采用该类算法的原因:设想你看了一则关于汶川地震十周年的报道,很可能推荐的新闻与你看过的有大量重复内容。

限制该算法的另一个问题是,在面对音乐、电影、图片等多媒体推荐时,由于产品内容难以提取,推荐难度增大,由此衍生出的一种基于标签的推荐算法逐渐成为该类算法的主要应用形式。第三章将实现一种通过该算法进行专车平台车辆推荐的方案。

### 2.3.2 协同过滤算法

一言以蔽之,协同过滤算法就是先找出具有相似兴趣的用户,然后为用户推荐和其兴趣相似的用户喜欢的物品。

该类算法的思想是用户可能喜欢那些具有相似兴趣的用户喜好的物品。举例来说,某用户的“朋友”喜欢电影《复仇者联盟3》,则推荐系统将该电影推荐给该用户,这里的“朋友”不是现实意义上的朋友,而是一个形象的说法,在协同过滤算法中,这些“朋友”有一个专有名词——邻居用户,如何确定邻居用户是协同过滤算法及很多具有相同思想算法的核心,最近邻选择算法是否合理往往决定了一个协同过滤算法的优劣性。该类算法的弊端是训练过程较为复杂、训练周期长,但训练完成后其出色的速度和健壮性是其其他算法所不具备的。

协同过滤算法在电子商务领域最为经典的应用之一是亚马逊网络书店的书推荐系统,在顾客选择一本喜好书籍后,推荐系统会快速生成一行“Customer Who Bought This Item Also Bought”,如今几乎所有的电子商务平台都提供这一服务。

### 2.3.3 基于知识的推荐算法

一言以蔽之,基于知识的推荐算法就是在挖掘用户和用户兴趣间的内部联系(知识)的基础上,引导用户选择其可能的喜欢的物品。

基于知识的推荐系统又可以分为两个大类,区别二者的依据是如何使用所提供的知识:

- 1.基于约束的推荐算法依赖明确定义的推荐规则集合,其核心在于如何确定这些规则;

2.基于实例的推荐系统主要根据不同的相似度衡量方法检索出相似的物品，其核心是相似算法[7]。

下面以基于约束的推荐中基于关联规则的推荐举例，某用户喜欢《复仇者联盟3》，根据对大量用户喜好的数据挖掘发现，喜欢《复仇者联盟3》的用户中有超过某阈值的用户同时喜欢《甄嬛传》，这个关联听起来很荒谬，但可能真实存在，推荐系统据此将《甄嬛传》推荐给该用户。

有时基于知识的推荐算法也被归纳为基于内容的推荐算法，这是因为该类算法的知识追根究底来源于物品内容，但目前大部分学着将其视作为单独的一类算法，这是因为，基于内容的推荐算法是对内容本身进行处理的，它属于一种过滤性系统，而基于知识的推荐算法则不同，它是在挖掘用户和物品内容所包含的知识后进行处理的一种带有很大引导性的系统。与之前讨论的两种算法相比，基于知识的推荐算法需要较大的代价来获取和维护所需的知识，但其不受评分数据和时间跨度的影响，灵活性与健壮性极高。

## 4. 第3章数据库与关联规则挖掘

总字数：6203

相似文献列表 文字复制比：0%(0) 疑似剽窃观点：(0)

原文内容 红色文字表示存在文字复制现象的内容; 绿色文字表示其中标明了引用的内容

### 第3章数据库与关联规则挖掘

在第二章我们已经介绍了基于知识的推荐算法，该类算法的核心工作之一是进行知识挖掘，本章在此基础上介绍了知识挖掘中最为常见的关联规则挖掘，分析了其思想和应用。同时通过MYSQL建立了专车平台数据库，并通过模拟数据手段进行了专车平台客户数据关联规则挖掘，得出了相关频繁项集和一些关联规则。

#### 3.1 数据需求

在使用不同的数据分析与挖掘技术时，对数据类型和数据量有不同的需求，本节将结合本文第三章将要使用的两种数据挖掘算法，对专车平台客户数据挖掘所需求的数据作出分析。

##### 3.1.1 基于标签的推荐算法的数据需求

在第三章我们将会介绍，由于基于标签的推荐算法属于基于内容的推荐算法，其推荐的依据是物品本身的内容，因此这一算法并不过分重视客户自身的数据，重点应放在如何获得有价值的车辆标签数据。由于客户在描述某车辆的标签时并不一定按照我们的期望提供有用的信息，所以在用户打标签时，应提供参考项对客户进行引导。

##### 3.1.2 关联规则挖掘的数据需求

关联规则挖掘在本实验中作为基于知识的推荐算法中的第一步——知识挖掘，成员属性的选取和数据量的大小对最终分析能否得到潜在且可信的关联项起重要作用，属性应尽量全面但又不宜过分冗余而使得后期分析时占用大量空间与时间。由于现有专车平台客户数据集基本只包含最基本的客户数据，而根据这些数据难以挖掘出有用的知识，加之本文着重论述专车平台客户数据挖掘的实用性和可行性，故本次实验选择自建数据库并通过调查问卷和模拟的方式保证关联挖掘算法的正常运行。

将关联规则挖掘的思想用于专车平台，就是要挖掘出在面对某特定客户群体时车辆和驾驶员情况与服务效果的联系，预测客户的服务偏好，对资源作出有针对性的分配，最大化现有资源的价值，提升企业市场竞争力。本次实验的主要内容是在搜集整理客户、司机、车辆三者数据的基础上，对相关数据集进行分析与挖掘，获得与服务效果有关的关联项，这些关联项包括但可能不仅限于客户属性与车辆属性间的关联，客户属性与司机属性间的关联，客户、车辆、司机三者属性间的关联。

#### 3.2 数据来源分析

正如绪论中提到的，目前，数据挖掘与分析由于多种多样的原因，在网约车专车行业中并未像其他电子商务领域内一样大显身手，各专车平台的客户数据集并不完善，客户数据来源问题也是一个重要课题，本节探讨几种较为有效的数据获取方式。

##### 3.2.1 客户端录入

通过移动或传统客户终端，在用户访问平台注册时通过交互页面获取用户一些简单却最为实用的基本信息，包括年龄、性别、职业等。

##### 3.2.2 过往订单记录

专车服务是一种日常的频繁服务，且客户在使用后的一段时间往往会继续使用相同平台进行预订，因此整理订单记录以获取信息对尤其是基于知识的推荐技术的推荐系统具有重要价值。

##### 3.2.3 合作平台数据集

目前大部分专车平台都提供合作平台登录的功能，对于通过微信、微博等合作平台登录的客户，可以申请访问其合作平台的数据集。同时，在新用户注册时也可以申请访问合作平台的数据集寻找是否有这一客户的数据，在客户同意的情况下对相关数据进行调用。

##### 3.2.4 社交渠道

互联网+时代下，每一个客户的线上活动都留下大量潜在信息。在足够的网页爬虫、语义分析等技术的支撑下，推荐系统的数据来源可以是客户在各大社交平台公开的任何看似没有价值的数据。当然，这是一个复杂的工程，在具体实现时，可先在高端商务型客户进行实践，并在技术成熟后逐步普及。



### 3.3 数据库的建立

#### 3.3.1 表项确定

本次实验数据库各实体的属性确定如下：

客户：客户编号，客户姓名，客户性别，客户年龄，职业，

司机：司机编号，司机姓名，司机年龄，驾驶习惯

车辆：车辆编号，车型，品牌，种类，漆色，内饰，香水，

订单：订单编号，服务效果，司机编号，车辆编号

\* 为主键， 为外键

在操作时，上述属性应根据数据获取阶段实际情况作出有针对性的调整，尤其是对于客户数据的统计，应注意一些容易获取易被忽略的属性，如客户的常住城市、身高甚至体型等等，或许这些意想不到的细节对客户的服务偏好有举足轻重的影响。但这些数据可能涉及客户的信息安全问题，这些将在本文第四章进行分析。

同时，由于实际数据需要完整的平台终端或长期的企业数据集支持，本次实验通过问卷调查和模拟数据补充的方式获取保障推荐算法有效运行的最低数据量进行实际操作以论述算法的可行性。

#### 3.3.2 数据库界面

图 3-1 E-R图

图3-2 数据库界面

### 3.4 专车平台关联规则挖掘

#### 3.4.1 关联规则的典型应用

关联规则挖掘包括在数据集的特征值组合之间找到值得关注的“if-then”规则。一个关联规则 $X \rightarrow Y$ ,  $X$ 和 $Y$ 是组合的特征值(模型),意味着如果 $X$ 出现在一个对象 $Y$ 也出现在同一个对象。通常，如果关联规则的频率和可信度不小于用户指定的频率和置信阈值，则认为它是值得关注的。规则 $X \rightarrow Y$ 的频率的频率模型数据集 $\Omega_{XY}$ ;它的可信度是数据集中的对象的分数，如果 $X$ 出现， $Y$ 也会出现。关联规则挖掘包括两个基本步骤：

1.搜索频繁模型(频率不小于频率阈值的模型)

2.从频繁模型构建关联规则

第一步(也称为频繁项集挖掘)本身是非常重要的,因为规律(模型)的数据发现,这些模型,来源于应用程序代表用户配置文件,做法,常见的综合症,风险因素,等等,领域包含市场营销、生物信息学、医药、网络安全等。就像绪论中所介绍的一样，关联规则挖掘作为数据挖掘领域最早也最经典的研究领域，如今已衍生出各种挖掘算法并广泛应用于医疗、科研、公共管理和市场营销等在内的几乎所有需要数据分析的领域中。关联规则挖掘是数据挖掘与分析中知识发现这一步的重要任务。已被用于市场营销、犯罪分析、入侵检测、欺诈检测、疾病诊断或分析等几乎所有社会科学与商业领域。

#### 3.4.2 思想与算法

频繁项集意味着什么?这意味着相同的特征值组合在数据集中出现一定次数。例如，给定表3-1中所示的数值和非数字特征(混合数据)所描述的数据集，假设0.6作为频率阈值和0.9作为置信阈值，则特征值的唯一频繁项集是(婚姻状况为未婚)，在数据集的6个对象中出现4次，没有值得关注的关联规则。

表3-1 混合数据集的示例

$\Omega$  年龄汽车婚姻状况

o1 23 紧凑型未婚

o2 25 大型未婚

o3 25 中型未婚

o4 29 中型未婚

o5 34 大型已婚

o6 38 豪华已婚

相似的概念或相反的概念，不同的概念(并不一定是距离)是一种自然的工具，通常用于软科学中作出决定。如果相似度函数不同的频繁项集,也称为频繁类似的模型,是一个组合的特性值的研究对象,这样,类似模型的相似性积累不小于一个阈值指定的频率。

可以注意到，相似函数的使用与等式(特征值和对象描述之间)的不同产生了频繁的模式和值得关注的关联规则，这些规则隐藏在使用等式作为相似函数的算法中。这篇论文的初步结果发表在罗德里格斯-冈萨雷斯等人(2008)。在目前的工作中，我们关注的是关联规则挖掘，在混合数据上使用相似函数。这一过程分为两步:(I)频繁项集的挖掘;(II)产生值得关注的关联规则。对于第一步，我们提出了两种算法:一种用于保持 $f$ 向下闭包属性的相似函数，另一种用于不具有该属性的相似函数。第二步，我们提出了GenRules算法(Agrawal & Srikant, 1994)的改编。重要的是要强调，在Rodriguez-Gonzalez、Martinez-Trinidad、Carrasco-Ochoa和Ruiz-Shulcloper(2011)中，一种常见的类似模型挖掘算法(称为RP-Miner)用于布尔相似函数，而这些函数不符合向下闭包性。尽管实验结果报道Rodriguez-Gonzalez et al .(2011)表明,RP-Miner比本文中提出的算法更高效的相似性函数不满足向下关闭属性(STreeNDC-Miner)和更有效的算法提出了相似性函数,完成下行关闭属性(STreeDC-Miner)。在Rodriguez-Gonzalez等人(2011)中也发现，在那些我们知道相似函数满足向下闭合属性的问题中，STreeDC-Miner的速度更

快。在这些问題中，我们知道相似函数不能满足向下的闭包属性，STreeNDC-Miner发现了所有的模型，而RP-Miner只找到一个子集。因此，本文提出的算法是一种替代方案，在这些情况下，RP-Miner不能提供良好的结果。因此，本文的研究结果完成了基于布尔相似函数不同于等式的混合数据挖掘项集算法的研究。

ObjectMinerDanger、Ruiz-Shulcloper和Berlanga(2004)是第一个使用相似函数挖掘频繁模型的算法。为了让修剪的搜索空间频繁项集相似，相似度函数，该算法设计：如果两个对象不是类似的对功能集合S然后他们没有类似的关于任何的超集S ObjectMiner是先验的启发算法(Agrawal & Srikant,1994)。它遵循广度优先搜索策略：首先，对于每个特性，所有频繁相似的值(频繁相似的子描述只有一个特性)，都是确定的。结合常见的长度1的相似模型，可以得到长度为2的相似模型的候选。之后，对每一对 $(\pi, P_j)$ 频繁类似subdescriptions  $k-1$ 特性，发现在迭代 $k-1$ ，如果 $\pi$ 和 $P_j$ 有共同的subdescription  $k-2$ 的特性，他们结合，以创建一个新的候选人subdescription P和 $k$ 特性(见图1)。在这个步骤中，对每一对 $(\pi, P_j)$ 频繁类似subdescriptions  $k-1$ 特性和常见subdescription  $k-2$ 功能，接下来的流程是：

- 结合P\*获得从 $\pi$ 和 $P_j$ 。
- 一组候选人获得类似于P\*相交的集合subdescriptions类似于 $\pi$ 和subdescriptions类似 $P_j$ 的集合。
- 设置的候选人，一组P\*subdescriptions相似。
- P\*的频率计算的数据集。
- 如果P\*的频率大于或等于minFreq则P\*subdescription经常相似。

当迭代不产生任何与 $k$ 特性相似的子描述时，这个过程就结束了。

ObjectMiner的主要缺点是，尽管对对象的描述或子描述通常在数据集中重复，但它并没有使用这个事实来减少后续步骤的操作数。由于这个原因，相同的子描述重复的相似度被计算出来，从而导致额外的不必要的计算工作量。同时，在处理多个对象的数据集时，为每个频繁的子描述存储一组相似的子描述(包括它的重复)会影响ObjectMiner的性能。

在Danger等人的研究中，我们发现，在计算特征值组合的频率时，使用相似函数的方法不同于计算特征值组合的频率，从而可以发现使用相等函数时隐藏的频繁项集。同样，，值得关注的关联规则从频繁项集中得到。然而，目前并没有相关值得关注的生成关联规则的算法。

为了从Agrawal和Srikant(1994)的频繁项集中生成值得关注的关联规则，提出了GenRules算法。GenRules算法通过将两个分离子集的频繁模型中的特征分离出来，并验证规则的可信度是否大于或等于一个指定的最小置信阈值，从而为每个频繁模型生成所有可能的规则。

在Agrawal和Srikant(1994)中描述了一种快速生成值得关注关联规则的算法。该算法基于以下属性。对于每个频繁模型，如果规则的置信度 $X \rightarrow Y$ 分离获得的功能在两个不相交的子集小于指定的最低置信度阈值，然后所有的置信度规则 $X-Z \rightarrow YZ$ ，Z是X的子模型，也小于指定的最低门槛。由于这个属性，如果规则 $X \rightarrow Y$ 不是值得关注的所有规则 $X-Z \rightarrow YZ$ ，Z是X的子模型，也不感兴趣，没有必要验证他们的置信度。在快速算法中，对于每个频繁模型，首先在结果中只有一个特征的所有规则，以及前一个特征中的其余特征。后来，对于每一个规则，先行词的特征被递归地转移到结果中，从而产生新的规则，直到规则的置信度小于最小的置信阈值。

在文献中已经报道了从频繁模型生成值得关注关联规则的其他算法，但是它们是为特定种类的关联规则或域而设计的。

让 $\Omega = \{O_1, O_2, \dots\}$ 是一个数据集。每个对象O是描述一组特性 $R = \{r_1, r_2, \dots, r_m\}$ ，表示为一个元组 $(v_1, v_2, \dots, v_m)$   $v_i \in D_i$  ( $D_i$ 的域特性 $r_i$ ) (1im)的一个对象O subdescription特性的一个子集 $S \subseteq R$ 表示(O)，是O的特性的描述； $O[r]$ 表示的价值特性 $r \in O$ ；和 $f_S(O, O')$ 表示O，O'之间的相似性用subdescriptions(O)表示，(O')分别(Martinez-Trinidad、Ruiz-Shulcloper & Lazo-Cortes,2000)。给定两个subdescriptions(O)、(O')和 $O, O' \in \Omega$ ， $f_S(O, O') = 1$ 意味着O类似于O'， $f_S(O, O') = 0$ 意味着并不类似于O'。两个基于集合S相似理论的例子是：

(1)

$Cr: x \rightarrow \{0, 1\}$ 是一个比较函数r值之间的特性，和 $\alpha \in [0, 1]$ 。比较函数的两个例子是：

(2)

(3)

令(O)是一个子数据库， $O \in \Omega$ ， $S \subseteq R$ ， $S \neq \emptyset$ 和f是一个相似性函数，取决于特性集S，则 $\Omega(O)$ 的频率f被定义为：

给定一个数据集对象 $\Omega$ ，所描述的一组特性R，相似性函数f，取决于功能集，和一个频率阈值minFreq，频繁相似模型挖掘问题在于找到 $\Omega$ 中所有类似的频繁模型。如果信任阈值minConf也已知，使用相似函数的关联规则挖掘问题就在于从 $\Omega$ 频率相似模型中找到所有值得关注的关联规则。

### 3.4.3 频繁项集

考虑到最后一个相似的模型定义，假设：(I)两个年龄相仿，如果其差异的绝对值最多为5年；紧凑型轿车与中型轿车相似，中型车与大型轿车相似；大型汽车与豪华轿车相似；然后，从表3-1中挖掘出的频繁项集和值得关注的关联规则以及它们的频率和置信值将会显示在表3-2和表3-3中。

表3-2 频繁项集

频繁项集频率

年龄= 25 0.66

年龄 29岁 0.66

车=中型 0.83  
 车=大型 0.83  
 结婚= No 0.66  
 年龄= 25 , 汽车=中型 0.66  
 年龄= 29 , 汽车=中型 0.66  
 年龄= 25 , 结婚=No 0.66  
 车=中型 , 结婚= No 0.66  
 频繁项集频率  
 年龄= 25 , 车=中 , 已婚= N 0.66

#### 3.4.4 关联规则

表3-3 关联规则

值得关注的关联规则置信度

(年龄= 25)→(车=中型) 1  
 (年龄= 29)→(车=中型) 1  
 (年龄= 25)→(结婚= No) 1  
 (结婚= No)→(年龄=25岁) 1  
 (车=中型)→(结婚=No) 0.8  
 (年龄= 25)→(车=媒介,结婚= No) 1  
 (年龄= 25 车=中型)→(结婚= No) 1  
 (年龄= 25结婚= No)→(车=中型) 1  
 (车=媒介,结婚= No)→(年龄=25岁) 1  
 (结婚=No)→(年龄=25岁车=中型) 1

## 5. 第4章一种基于标签的推荐算法的实现

总字数：4243

相似文献列表 文字复制比：47.3%(2005) 疑似剽窃观点：(0)

1	《推荐系统》基于标签的用户推荐系统 - Thinkgamer博客 - CSDN博客 - 《网络 ( <a href="http://blog.csdn.net">http://blog.csdn.net</a> ) 》 - 2017	36.4% ( 1546 ) 是否引证：否
2	《推荐系统》基于标签的用户推荐系统 - sysmedia的博客 - CSDN博客 - 《网络 ( <a href="http://blog.csdn.net">http://blog.csdn.net</a> ) 》 - 2017	35.4% ( 1502 ) 是否引证：否
3	基于标签的个性化推荐系统的研究 杨丹(导师：赵朋朋) - 《苏州大学博士论文》 - 2014-10-01	10.8% ( 457 ) 是否引证：是
4	20131060276_邹鑫胤_推荐系统 邹鑫胤 - 《大学生论文联合比对库》 - 2017-04-26	10.4% ( 440 ) 是否引证：否
5	20131060276_邹鑫胤_推荐系统 邹鑫胤 - 《大学生论文联合比对库》 - 2017-04-27	9.3% ( 395 ) 是否引证：否
6	基于图书馆业务数据分析服务的个性化推荐系统设计与实现 张通(导师：王丹志) - 《北京邮电大学博士论文》 - 2012-12-30	8.6% ( 363 ) 是否引证：是
7	点餐软件点餐推荐算法设计与实现 魏皓 - 《大学生论文联合比对库》 - 2017-05-29	8.1% ( 343 ) 是否引证：否
8	基于Android的智能收藏APP 王文贤 - 《大学生论文联合比对库》 - 2017-06-02	5.5% ( 235 ) 是否引证：否
9	11648080_王文贤_基于Android的智能收藏APP 王文贤 - 《大学生论文联合比对库》 - 2017-06-05	5.5% ( 235 ) 是否引证：否
10	机器学习->推荐系统->利用用户标签数据 - Mr - 《网络 ( <a href="http://blog.csdn.net">http://blog.csdn.net</a> ) 》 - 2017	4.9% ( 210 ) 是否引证：否
11	基于个性化推荐的移动应用管理系统的设计与实现 潘宇彬(导师：李兵兵;郭春燕) - 《西安电子科技大学博士论文》 - 2012-12-01	4.5% ( 190 ) 是否引证：是
12	点餐软件点餐推荐算法设计与实现 魏皓 - 《大学生论文联合比对库》 - 2017-06-01	4.4% ( 188 ) 是否引证：否
13	基于Android的智能收藏APP 王文贤 - 《大学生论文联合比对库》 - 2017-06-05	3.8% ( 162 ) 是否引证：否
14	基于图模型的个性化标签推荐技术研究	3.7% ( 159 )

	吴幸良(导师：涂风华) - 《重庆大学博士论文》 - 2014-04-01	是否引证：是
15	算法应用 - HadoopSpark - CSDN博客 - 《网络 ( <a href="http://blog.csdn.net">http://blog.csdn.net</a> ) 》 - 2017	1.8% ( 75 ) 是否引证：否
16	基于标签的推荐系统的研究与实现 刘海浪(导师：陈先桥) - 《武汉理工大学博士论文》 - 2014-04-01	1.5% ( 62 ) 是否引证：否
17	基于用户行为的推荐系统关键技术研究 刘天毅 - 《大学生论文联合比对库》 - 2016-05-22	1.4% ( 59 ) 是否引证：否
18	201092048_吕望_基于协同过滤的电子商务推荐系统研究与应用 吕望 - 《大学生论文联合比对库》 - 2014-06-12	0.9% ( 39 ) 是否引证：否
19	201092048_吕望_基于协同过滤的电子商务推荐系统研究与应用 吕望 - 《大学生论文联合比对库》 - 2014-06-12	0.9% ( 39 ) 是否引证：否
20	201092048_吕望_基于协同过滤的电子商务推荐系统研究与应用 吕望 - 《大学生论文联合比对库》 - 2014-06-12	0.9% ( 39 ) 是否引证：否
21	201092048_吕望_基于协同过滤的电子商务推荐系统研究与应用3 吕望 - 《大学生论文联合比对库》 - 2014-06-18	0.9% ( 39 ) 是否引证：否
22	14_付浩静 付浩静 - 《大学生论文联合比对库》 - 2016-03-13	0.8% ( 36 ) 是否引证：否

原文内容 **红色文字**表示存在文字复制现象的内容; **绿色文字**表示其中标明了引用的内容

#### 第4章一种基于标签的推荐算法的实现

**第二章我们介绍了基于标签的推荐算法，它是一种基于内容的推荐算法。本章将在此基础上分析一些平台现有的标签系统应用，然后选取了一组专车平台可用的标签并实现了一种适合专车平台车辆推荐的推荐算法，最后分析了进行标签推荐的必要性和几种可行的标签推荐算法。**

##### 4.1 标签系统的典型代表

目前几乎所有的图书、音像、电子商务等平台都拥有自己的标签系统，下图为某图书平台和音乐平台的标签界面：

图4-1 [豆瓣图书标签界面](#)

图4-2 [网易云音乐标签界面](#)

**大量实践证明标签系统在帮助用户发现他们感兴趣的项目时非常有效。**

##### 4.2 用户如何打标签

###### 4.2.1 标签流行度

**在用户表面的随机行为之后隐藏大量规律，因此我们需要对标签进行统计，引入标签流行度这个变量：**当用户为某物品打上标签时，该标签的流行度 tagfreq [tag] 加1。

图4-2是一个典型的符合长尾分布的标签流行度分布图（横坐标是标签流行度K，纵坐标是流行度K对应的标签数目），其双对数曲线趋于一条直线。

图4-3 标签流行度的长尾分布

###### 4.2.2 标签选取

**在客户选择一个车辆时，为了使他打的标签是能够准确的描述车辆属性的关键词，**有必要人工编纂一些特定的标签供客户选择。我们先来看一些现有推荐系统的标签包含哪些信息：

- 物品是什么的信息：比如物品是一只猫，就会有“猫”这个标签；是亚马逊的首页，就有“亚马逊”这一标签。**
- 种类信息：比如在qq音乐的标签中，有流行、古典、现代等表示音乐种类的标签。
- 归属信息：比如很多论坛的标签中包含作者标签。
- 用户观点信息：包含各种形容词，如有趣的、平庸的、完善的等等。
- 用户任务信息：比如即将阅读、搜集素材、找工作等。

结合上述标签并回顾第三章所讨论的数据相关问题，本次实验选择提供如下标签：

表4-1 一种专车平台标签方案

车型  
小型紧凑中型大型豪华 SUV MPV 跑车  
车系  
国产德系美系日系法系英系韩系其他  
漆色  
黑色银灰白色红色蓝色粉色其他  
内饰  
皮革运动桃木机械其他

##### 4.3推荐算法

标签作为接洽用户和物品的纽带，将用户对**物品的描述规范化与系统化，是了解用户兴趣的主要数据源，本节探讨如何**



利用用户标签数据进行具有较高个性化程度的推荐。

我们以豆瓣的标签数据为分析对象进行探讨，它将标签系统融入产品线的主要为一下三步：

1.在每部电影的页面上，提供“豆瓣成员常用标签”，该应用程序的主要功能是给出用户最频繁给该电影打上的标签。

2.每当用户评价一部电影时，豆瓣通过提供标签对其进行提示。

3.在最终的个性化推荐结果里，豆瓣将用户的推荐结果做了聚类，这个过程正是通过标签实现的。它显示了对不同标签下用户的推荐结果，极大提升了系统效果。

将上述步骤推广化，用三元组的集合表示用户标签行为的数据集，数据集中的记录 $(u, i, b)$ 表示客户 $u$ 给物品 $i$ 打上了标签 $b$ 。这样的三元组表示法虽然无法详尽描述客户复杂的真实标签行为，比如时间、客户属性、物品属性等。但是为了集中讨论标签数据，只考虑上面定义的三元组形式的数据，即客户的每一次打标签行为都用一个三元组（客户、物品、标签）表示[10]。

#### 4.3.1. 试验设置

本小节将数据集分为10份，这个分割应保证随机性，要注意分割的键值不含有标签，仅仅包含客户和物品这样做的目的在于保证客户对单一物品多个标签不会出现一部分在训练集其余一部分在测试集的情况。数据集准备工作完成后，任选一个测试集，其余自动成为测试集，算法的目的是学习训练集中的用户标签数据所包含的知识，一次对测试集上用户会为哪些物品打标签作出预测。令 $R(u)$ 为客户 $u$ 的推荐列表，该列表的长度为 $N$ ，表项是物品，这些物品都是系统预测用户将要打标签的，然后用 $T(u)$ 存放测试集中用户 $u$ 实际打标签的物品。完成上述工作后，利用准确率（precision）和召回率（recall）检验个性化推荐算法的可信度。

(1)

(2)

将数据集中的每一份都作为测试集运行上述程序，取这十次实验的准确率和召回率计算平均值，并将其作为最终的评测结果。下面从三个方面评测系统的性能。

1.覆盖率，其计算公式如下：

(3)

3. 多样性

首先，使用物品标签向量的余弦相似度衡量物品的相似度。计算过程如下，其中 $item\_tags[i]$ 存放每个物品 $i$ 的标签向量。

然后，用如下公式计算一个推荐列表的多样性：

(4)

Python实现为：

推荐系统的多样性为所有用户推荐列表多样性的平均值。

3.新颖度

推荐系统的新颖性是一个不易量化的指标，推荐结果的平均热门程度度量可以作为一个参考。而对推荐系统，定义它的平均热门度如下：

(5)

$item\_pop(i)$ 表示给物品 $i$ 打过标签的用户数。

#### 4.3.2. 一个简单的算法

本小节根据用户标签行为数据介绍一个个性化推荐算法，该算法包含以下几个步骤：

1.对于每一用户，统计其最常用标签。

2.对于每个标签，统计被打过这个标签次数最多的物品。

3.对于一个用户，调取其常用的标签，然后找到最符合这些标签的物品当中最为热门的。

用户 $u$ 对物品 $i$ 的兴趣公式如下：

(6)

其中， $B(u)$ 为 $u$ 打过的标签集合， $B(i)$ 为 $i$ 被打过的标签集合，是 $u$ 使用标签 $b$ 的次数，是物品 $i$ 被打过标签 $b$ 的次数。在下述程序中：

$records[i] = [user, item, tag]$ ;  $user\_tags[u][b]$  表示存储的数组， $tag\_items[b][i]$  表示存储的数组。

用该程序从 $records$ 中统计出 $user\_tags$ 和 $tag\_items$ ：

统计出 $user\_tags$ 和 $tag\_items$ 之后，通过如下程序对用户进行个性化推荐：

#### 4.3.3 算法的改进

回顾上一节提出的算法

(7)

该算法存在处理热门商品和解决数据稀疏性问题时表现不佳，而这两点对推荐系统的性能影响较大，本小节对其进行一些改进。

1.TF-IDF

4.3.2中的算法给包含热门标签的人们商品分配了极高权重，因此这些物品很容易被推荐，这个问题类似于第二章提到的

冷启动问题。这会使系统的新颖性不足并且上述算法难以体现用户个性化兴趣。公式8利用TF-IDF思想对公式7进行改进：

(8)

同理，公式9运用相同的思想对热门物品进行惩罚：

(9)

记录物品*i*被多少个不同的用户打过标签。

## 2. 数据稀疏性

在上一节的算法在面对新用户时表现不佳，为了解决这个问题我们可以对标签进行扩展。例如，某用户使用过“关联规则”这一标签，则同时向该用户的标签集合中加入“频繁项集”、“相似函数”这些标签。

下面介绍一种基于邻域的标签扩展方法。

标签扩展的主要工作是计算标签之间的相似度，从而依据此找到相似标签。如果认为同一个物品上的不同标签具有某种相似度，那么当两个标签同时出现在很多物品的标签集合中时，就可以认为这两个标签具有较大的相似度[11]。对于标签*b*，令*N(b)*为有标签*b*的物品的集合，通过公式(10)计算标签*b*和标签*b'*的相似度：

(10)

### 4.3.4. 标签清理

用户使用的标签并不一定都表示其偏好。下面是一个有些好笑的例子：在一个电影网站中，某用户对一部电影打了“平淡无奇”这一标签，标签系统因此认为该用户对“平淡无奇”感兴趣，进而为该用户推荐其他平淡无奇的电影。相反，如果用户对某电影打过“李小龙”这一标签，标签系统应该能够据此得知用户对李小龙的电影感兴趣并为其推荐李小龙的其余作品。同时，标签系统里还有形如“famous”和“well-known”这样词形不同、词义相同的标签。

此外，标签清理还能使推荐解释更加合理。目前主流的标签清理方法可分为以下三种：

1. 去除词频很高的停止词；

2. 去除词形不同的同义词，比如famous和well-known；

3. 去除因分隔符造成的同义词，比如 well\_known和well-known。

为了提升标签的质量，也可以通过交互系统让用户能够对标签进行反馈，即让用户告知某个标签是否对评价物品有所帮助。

运行结果：

('912', 610)

('3763', 394)

('52503', 238)

('39051', 154)

('45647', 147)

('21832', 144)

('1963', 143)

('1237', 140)

('33815', 140)

('5136', 138)

## 4.4 标签推荐

### 4.4.1 为什么给用户推荐标签

正如之前讨论的一样，进行标签推荐主要有以下两个好处：

1. 方便用户输入标签：用户从键盘输入标签增加了用户打标签的时间，这可能导致部分用户放弃打标签，因此我们需要进行标签推荐来减小用户打标签的难度，从而提高用户打标签的参与度。

2. 提高标签质量：同一个语义，尤其在汉语中，不同的用户可能用不同的词语来表示。这样的同义标签既增大了数据量与计算量，还会相似度计算的精确度。在引入标签推荐后，大部分标签都是我们所推荐的，只有少部分有特别想法的用户会加入新的标签，这样一来在对词表进行选择后，既能保证其中不包含过多同义词，又能保证出现的很大几率是一些比较流行的、有代表性的词[12]。

### 4.4.2 标签推荐的四种简单算法

1. PopularTags：为客户*u*推荐系统的最热门标签，令tags[b]为标签*b*的热门度，算法实现如下。

2. ItemPopularTags：为客户*u*推荐物品*i*上最热门的标签，令item\_tags[i][b]为物品*i*被打上标签*b*的次数。

3. UserPopularTags：为客户*u*推荐该客户自己最常用的标签，令user\_tags[u][b]为用户*u*使用标签*b*的次数，

4. HybridPopularTags：综合使用两种算法，将算法2和算法3的结果线性加权，将加权后的结果作为最终结果。

在上面的实现中，两个列表线性相加时都将两个列表按最大值做了归一化，归一化的目的在于控制两个列表对结果的影响，避免一方所占权重过高，具体来说就是避免物品热门程度过高而掩盖了用户对推荐结果的作用，或者由于用户过于活跃而掩盖了物品本身对推荐结果的作用[13]。

指 标	
疑似剽窃文字表述	
1. 基于标签的推荐算法的实现	第二章我们介绍了基于标签的推荐算法，它是一种基于内容的推荐算法。本章将在此
2. 豆瓣图书标签界面	图4-2 网易云音乐标签界面
3. 覆盖率，其计算公式如下：	大量实践证明标签系统在帮助用户发现他们感兴趣的项目时非常有效。
(3)	
3. 多样性	首先，使用物品标签向量的余弦相似度衡量物品的相似度。
4. 如下公式计算一个推荐列表的多样性：	
(4)	
Python实现为：	
推荐系统的多样性为所有用户推荐列表多样性的平均值。	
3.新颖度	推荐系统的新颖性是一个不易量化的指标，推荐结果的平均热门程度
5. 相反，如果用户对某电影打过“李小龙”这一标签，标签系统应该能够据此得知用户对李小龙的电影感兴趣并为其推荐李小龙的其余作品。同时，标签系统里	
6. 清理方法可分为以下三种：	
1.去除词频很高的停止词；	
2.去除词形不同的同义词，	

6. 第5章信息安全问题		总字数：2559
相似文献列表 文字复制比：12.1%(310) 疑似剽窃观点：(0)		
1	舒泰神信息安全体系保障核心数据资产 本报记者 张兰兰 - 《中国信息化周报》- 2017-01-16	9.1% ( 233 ) 是否引证：否
2	基于ThinkPHP的在线音乐播放网站 商家赫 - 《大学生论文联合比对库》- 2017-06-07	2.5% ( 64 ) 是否引证：否
3	基于ThinkPHP的在线音乐播放网站 商家赫 - 《大学生论文联合比对库》- 2017-06-15	2.5% ( 64 ) 是否引证：否
4	数据库安全研究现状与展望[1]-百度文库 - 《互联网文档资源 ( <a href="http://wenku.baidu.c">http://wenku.baidu.c</a> ) 》- 2012	2.5% ( 64 ) 是否引证：否
5	便携式数据库审计系统设备 吴纪芸;陈志德;李金寨; - 《网络安全技术与应用》- 2015-01-15	1.4% ( 35 ) 是否引证：否

原文内容 红色文字表示存在文字复制现象的内容; 绿色文字表示其中标明了引用的内容

第5章信息安全问题	
<p>信息安全问题自数据分析与挖掘技术被应用于市场营销范畴起就一直困扰着相关企业。大数据时代,大数据安全和隐私保护在安全架构、数据隐私、数据管理和完整性、主动性的安全防护等方面需要更多技术支持。专车平台数据挖掘与推荐系统的实现应在足够的信息安全保障下进行，这里所讨论的信息安全问题主要是信息泄露问题。本章论述推荐系统中一些已知的和潜在的威胁，并提出几种可行的解决方案。</p> <p>5.1 传统数据安全问题</p> <p>5.1.1 数据安全事故</p> <p>数据的保护只要有一点疏漏，就有可能造成极大的影响，特别是对企业而言，平台数据一旦泄露往往规模极大，需要进行重点防范[15]。然而近年来全球数据泄露事故发生率仍不断上升，以下是一些著名的数据安全事故：</p> <ol style="list-style-type: none"> <li>1.雅虎30亿帐号泄露，政监机构介入调查</li> <li>2.土耳其重大数据泄露事件约三分之二的公民信息遭到泄露</li> <li>3.uber客户信息泄露</li> </ol> <p>5.1.2 数据泄露</p> <p>信息泄密事件的发生大多数和人密切相关，泄密的途径和方式也多种多样，专车平台的数据泄露同时涉及到商业秘密和个人隐私的泄密，这两类泄露的泄密渠道可分为以下三类：</p> <ol style="list-style-type: none"> <li>1.主动泄密隐患</li> </ol>	



主动泄密已经成为当前企业信息安全的首要问题，在中国国家信息安全测评认证中心最新一期调查结果中，信息安全的现实威胁主要是主要泄密隐患，即内部信息泄露和内部人员犯罪造成的泄密，被动泄密和第三方泄密造成的影响相对较小。主动泄密隐患包括：

- (1) 将企业内部文档私自拷贝外带及复用泄密（USB/网络/即时通讯/刻录）；
- (2) 越权访问非授权数据泄密；
- (3) 盗用他人账号及设备非法访问数据泄密；
- (4) 伙同他人实现敏感数据跨安全域转移泄密；
- (5) 通过打印机、传真机等将敏感数据进行介质转换泄密；
- (6) 私自携带笔记本设备接入内部网络非法下载数据泄密；
- (7) 对敏感数据的恶意传播及扩散泄密；
- (8) 对核心应用系统的非安全接入及访问泄密；
- (9) 不遵守管理制度的其他导致数据泄密的行为等。

## 2.被动泄密隐患

被动泄密是指导致信息泄密的人员在无意识或不知情的情况下所发生的泄密隐患，而个人电脑中的重要数据泄密是被动泄密的主体。目前存在的被动泄密隐患包括：

- (1) 移动笔记本、USB存储设备遗失或失窃导致数据泄密；
- (2) 邮件或网络误操作、误发送引起的泄密；
- (3) 对敏感数据保管不当，如随意共享等；
- (4) 感染病毒、木马后引发的敏感数据泄密；
- (5) 移动笔记本、USB存储设备和硬盘等维修、废弃时引发的泄密。

## 3.第三方泄密隐患

第三方泄密是指重要数据交付给合作伙伴、客户或其他关系密切人员后保管不当或故意分发所引起的第三方扩散泄密，是目前企业商业化合作需求背景下逐渐体现的信息安全问题。目前存在的第三方泄密隐患包括：

- (1) 合作伙伴、外协人员接入内部网络非法获取敏感数据泄密；
- (2) 发送给客户、合作伙伴及其他关系密切人员的敏感数据保管不当或恶意扩散引起的泄密；
- (3) 合同、图纸以及科研、学术报告等敏感资料外部打印、复印时副本拷贝泄密；
- (4) 应用系统、邮件服务器以及数据中心外部托管时非法窃取泄密等。

## 5.2 推荐系统带来的新威胁

尽管推荐系统确实为企业和客户带来了双赢，但其同时也带来了一些新的信息安全隐患，这些隐患也可以认为是为第一节列举的信息泄露方式提供了新的渠道。本节以一则专车平台真实案例探讨推荐系统带来的新威胁。

2018年五月，21岁空姐李某搭乘滴滴平台顺风车遇害事件一度将推荐算法与推荐系统在类似案件中是否处于中立的问题腿上舆论的风口浪尖。虽然本次案件涉及的顺风车并非本文探讨的专车，但其所反应的数据挖掘与推荐技术带来的潜在威胁是所有专车平台在使用这些技术时不可忽视的重要问题。事件发生后，滴滴平台的推荐技术和安全保障机制成为了社会关注焦点，除了对司机的审核和监管问题，平台和算法在这样一起案件中是否处于中立引发了大量争论，在滴滴平台现有的功能支持下，顺风车车主可获取乘客的诸多有用信息，包括乘车时间、地点等，这些信息大部分情况下都是为了帮助车主与乘客双方节省时间。同时双方还能以“印象标签”的形式相互评价。据媒体披露，这些标签中不乏“颜值爆表”、“安静的美少女”等与服务无关的词汇，这些标签会显示在乘客的个人信息中，司机尚未接单之前就能看到这些过往评价。也就是说，司机与乘客双方可以根据这些评价，来选择其想要搭载的司机或乘客。

这些标签功能原本的目的是给司机和乘客提供参考，比如这个乘客是否按时付费、司机是否存在违约行为等，这些信息应当包含且仅包含承运人与乘客合同关系认可的正常评价信息。但如果平台共享标签信息超出了这个范畴，涉及到乘客的泛社交属性，比如是否漂亮、穿着如何等，那该类标签共享功能就涉及对乘客和承运人的隐私权侵权问题，更有甚者会对双方的人身安全构成威胁。

## 5.3 专车平台数据安全方案

信息泄露随时可能发生，信息泄露新渠道也层出不穷。对于专车平台而言，应该做到把安全策略放在整个业务流程的始末。

### 5.3.1 安全的设计架构

具体而言，在总体架构设计阶段，有明确的授权管理要求、用户认证要求、日志审计要求等。

对于专车平台，应当明确网络安全、应用安全、数据安全相关的安全防护手段。常见的安全设计包含以下三个方面：

- 1.在技术架构上，应采用分层的架构，实现对底层业务逻辑进行有效隔离，避免将底层实现细节暴露给最终用户；
- 2.在部署架构上，应采用站库分离的思想。将数据库应用服务器分离部署，这样在应用服务器被攻击时，可以避免或减少核心数据损失。

3.在外部接口设计方面，应采用最小接口暴露的原则，避免因开发不必要的服务而带来的安全隐患。

### 5.3.2 数据库安全技术



针对数据库系统的攻击将直接导致敏感与隐私信息泄露，开展数据库安全理论与技术研究,是实现数据库系统安全的迫切需求[16]。

- 图5-1 一个典型外包数据库
- 目前常用的数据库安全技术主要包含以下几个方面
- 1.使用代理保护数据库免受攻击；
  - 2.建立审计和健壮的日志记录；
  - 3.实行严格的用户帐户管理；
  - 4.保证数据库软件和操作系统是最新的；
  - 5.对敏感数据进行加密

指 标		
疑似剽窃文字表述		
1. 泄密的人员在无意识或不知情的情况下所发生的泄密隐患，而个人电脑中的重要数据泄密是被动泄密的主体。		
7. 第6章总结与展望		总字数：1221
相似文献列表 文字复制比：5.9%(72) 疑似剽窃观点：(0)		
1	基于Hadoop云的数据库营销海量数据处理与挖掘的研究 王海飞(导师：何利力) - 《浙江理工大学博士论文》 - 2013-03-12	3.0% ( 37 ) 是否引证：否
2	基于协同过滤的推荐方法的研究 邓星(导师：邓珍荣) - 《桂林电子科技大学博士论文》 - 2016-04-01	2.6% ( 32 ) 是否引证：否
原文内容 红色文字表示存在文字复制现象的内容; 绿色文字表示其中标明了引用的内容		

第6章总结与展望

6.1 成果归纳

本文在分析现有推荐技术及相关数据挖掘技术的基础上，探讨了如何利用并改进这些技术进行专车平台信息推荐。目前主要成果包括：

1.归纳总结了现有推荐技术，分类讨论了基于内容的推荐，基于协同过滤的推荐和基于知识的推荐

2.建立了专车平台企业数据库，数据库包含车辆、客户、驾驶员、订单等信息，录入了可保证相关推荐和挖掘算法正常运行的原始数据

3.通过模拟数据进行了基于专车平台客户数据的关联挖掘，实现了基于知识的推荐的第一步

4.实现了一种适用于专车平台的基于标签的推荐算法，并对算法进行了改进

5.探讨了专车平台推荐系统可能会涉及的信息安全问题，尤其分析了推荐技术带来的新威胁，并提出了相关解决方案

上述成果充分论证了专车平台客户数据挖掘与信息推荐的可行性与必要性，为相关实践提供了一定的理论基础和技术支持。

6.2 问题与解决方案

6.2.1 数据量与可靠度不足

由于本文的侧重点在于讨论专车平台客户推荐系统的必要性和可行性，且目前各专车平台企业数据集所包含的数据与本文所用推荐技术所需数据交叉性较小，故本文所用数据主要通过问卷与模拟手段获得，数据量较小，所获得知识的可靠性仍待进一步检验。

为进一步探讨专车平台客户数据挖掘及推荐技术实现，必然需要更大的数据量与更高的数据可信度，为解决这一问题，接下来将有两项工作：

进一步研究现有专车平台企业数据集，学习相关技术手段筛选可用数据，并结合所获数据改进算法

学习网络爬虫技术或邀请相关技术人员爬取各大汽车论坛、app上的相关内容，根据获得内容决定是否采用语义分析、机器学习等手段。为满足爬取大量数据的需求，需要根据爬虫调度端信息调取爬虫程序获取目标数据。

图5-2 网络爬虫的工作程序

在实际操作时这是一项复杂的工程，其实现有待在之后的具体工作中探讨。

6.2.2 关联规则及推荐算法针对性低

本文在进行关联规则挖掘和实现基于标签的推荐算法时选取的数据和算法普适性较高但针对性弱，可以实现推荐但推荐效果预期不高。此外，在本文所讨论的三种推荐算法中，第四章实现的基于内容的推荐算法虽然最易理解且应用广泛，但基于协同过滤的推荐和基于知识的推荐在进年来大数据和云计算技术的支持下发展势头正劲，相比而言基于内容的推荐受益较少。

为提高算法针对性以提升实用价值，接下来的工作重点之一是学习与研究Hadoop等云计算平台并探讨如何利用这些平台实现基于协同过滤的推荐和基于知识的推荐。例如针对经典Apriori算法及其改进算法无法处理大规模数据集,探讨基于Hadoop-MapReduce编程模型的改进算法。大量相关经验可供学习和参考，如浙江理工大学王海飞提出的基于Hadoop云的数据库营销

系统的架构，实现了海量数据的处理与存储，在系统初步完成并投入红塔集团运行后，集团产品销量同比大幅提升，数据库营销在其中起了至关重要的刺激作用。

#### 参考文献

- [1] 周换换. 滴滴专车平台的竞争策略分析[D]. 南京大学, 2016.
- [2] 沈池俊. 对数据库营销功能的战略思考[J]. 商业经济与管理, 1998(03):36-39.
- [3] 杨余垒. 改进的关联规则算法在慢性病数据挖掘中的研究[D]. 浙江理工大学, 2017.
- [4] 蔡皎洁. 基于语义集成的客户知识挖掘研究[D]. 武汉大学, 2012.
- [5] 马万智. 基于本体方法的客户关系管理实现[D]. 北京邮电大学, 2009.
- [6] 孙倩. 基于LM-BP神经网络的推荐算法的研究与应用[D]. 北京交通大学, 2016.
- [7] 皮佳明. 基于用户兴趣变化的协同过滤推荐算法研究[D]. 云南财经大学, 2014.
- [8] Ansel Y. Rodríguez-González, José Fco. Martínez-Trinidad, Jesús A. Carrasco-Ochoa, José Ruiz-Shulcloper. Mining frequent patterns and association rules using similarities[J]. Expert Systems With Applications, 2013, 40(17).
- [9] 易明, 邓卫华. 基于标签的个性化信息推荐研究综述[J]. 情报理论与实践, 2011, 34(03):126-128.
- [10] 杨丹. 基于标签的个性化推荐系统的研究[D]. 苏州大学, 2014.
- [11] 潘宇彬. 基于个性化推荐的移动应用管理系统的设计与实现[D]. 西安电子科技大学, 2013.
- [12] 吴幸良. 基于图模型的个性化标签推荐技术研究[D]. 重庆大学, 2014.
- [13] 张通. 基于图书馆业务数据分析服务的个性化推荐系统设计及实现[D]. 北京邮电大学, 2013.
- [14] 靳延安. 社会标签推荐技术与方法研究[D]. 华中科技大学, 2011.
- [15] 梁晓燕, 王如龙, 王军丽, 杨健. 信息安全保密中信息泄密途径及其防护[J]. 微计算机应用, 2004(04):406-410.
- [16] 张敏. 数据库安全研究现状与展望[J]. 中国科学院院刊, 2011, 26(03):303-309.

说明：1.总文字复制比：被检测论文总重合字数在总字数中所占的比例

2.去除引用文献复制比：去除系统识别为引用的文献后，计算出来的重合字数在总字数中所占的比例

3.去除本人已发表文献复制比：去除作者本人已发表文献后，计算出来的重合字数在总字数中所占的比例

4.单篇最大文字复制比：被检测文献与所有相似文献比对后，重合字数占总字数的比例最大的那一篇文献的文字复制比

5.指标是由系统根据《学术论文不端行为的界定标准》自动生成的

6.红色文字表示文字复制部分;绿色文字表示引用部分

7.本报告单仅对您所选择比对资源范围内检测结果负责



 [amlc@cnki.net](mailto:amlc@cnki.net)

 <http://check.cnki.net/>

 <http://e.weibo.com/u/3194559873/>