



信息系统协会中国分会第七届学术年会
(CNAIS 2017)

论文集

主办单位： 信息系统协会中国分会 (CNAIS)

承办单位： 复旦大学 管理学院

支持单位： 国家自然科学基金委员会
清华大学出版社

中国·上海 2017年10月20-22日

信息系统协会中国分会第七届学术年会 (CNAIS2017)

中国·上海 2017年10月20-22日

大会荣誉主席:

陈国青(清华大学)

大会主席:

毛基业(中国人民大学)

程序委员会领域主席(以姓名拼音为序):

黄丽华(复旦大学)

李一军(国家自然科学基金委员会)

马费成(武汉大学)

王刊良(中国人民大学)

徐心(清华大学)

杨善林(合肥工业大学)

程序委员会(以姓名拼音为序):

蔡剑(□京大学)

陈福集(福州大学)

陈晓红(中南大学)

党延忠(大连理工大学)

葛世伦(江苏科技大学)

郭迅华(清华大学)

黄丽华(复旦大学)

黄伟(西安交通大学)

霍佳震(同济大学)

寇纲(西南财经大学)

李纲(武汉大学)

李敏强(天津大学)

李一军(国家自然科学基金委员会)

梁昌勇(合肥工业大学)

刘渊(浙江大学)

吕廷杰(□京邮电大学)

鲁耀斌(华中科技大学)

马费成(武汉大学)

毛基业(中国人民大学)

戚桂杰(山东大学)

邵培基(电子科技大学)

孙建军(南京大学)

王刊良(中国人民大学)

谢康(中山大学)

徐升华(江西财经大学)

徐心(清华大学)

严建援(南开大学)

杨善林(合肥工业大学)

叶强(哈尔滨工业大学)

张朋柱(上海交通大学)

张新(山东财经大学)

赵晶(中国地质大学(武汉))

赵捧未(西安电子科技大学)

仲伟俊(东南大学)

程序委员会执行主席: 王刊良(中国人民大学)

组委会主席: 黄丽华(复旦大学)

组委会副主席: 徐云杰(复旦大学)

院长/系主任论坛主席: 左美云(中国人民大学)

主办单位: 信息系统协会中国分会(CNAIS)

承办单位: 复旦大学 管理学院

支持单位: 国家自然科学基金委员会 清华大学出版社

目录

ID 与注册号对应。可以通过本文件书签快速查找文章内容。

ID	论文标题	作者	页码
002	IT 外包与云计算情境下的 IT 业务匹配过程研究	邓春平 宋琦 王滢 孙玥璠	1
003	云制造平台工业企业准入指标体系研究	聂会星 邓旭	8
005	国内社科领域大数据研究知识扩散特征分析	曹玲 吴其静	15
007	移动虚拟社区治理对组织公民行为影响研究——从关系质量视角出发	迟铭 毕新华	22
008	农产品物流金融融通仓盈利模式演化博弈分析	杨波 吴燕	34
009	基于 S-O-R 模型的青年消费者“双十一”购物意愿形成机制研究	陈传红 李小倩	41
010	基于科学知识图谱的国内政策评估可视化研究	宁静 王亚民 马续 补	48
011	比特币矿工的最优选择策略设计	郑阳 杜荣	53
014	基于贝叶斯网络推理的干散货疏运偷盗风险分析	崔维平 黄磊 宋容 嘉	61
015	UGC 社媒前向商业变现动因研究	郭明君 吴俊 欧阳 书凡	67
016	当出版遇上互联网+——中南传媒集团“互联网+”的实践	许媛 张美娜 余艳	74
017	基于依存句法关系的在线评论情感属性的降维	王洪伟 蔡文嘉	84
018	话题情感对众筹投资者行为的影响研究	疏卉 张昊 尤薇佳	93
019	电子商务与企业绩效：供应链敏捷与集成的中介作用	张馨 李琦 张宁	101
020	食品在线个性化定制情景下营养信息呈现方式对用户行为的影响研究	富露 周良 王刊良	105
021	风险驱动的智慧建筑系统开发案例——基于万达	张克迪 傅湘玲 齐 佳音	117
022	iPad 和 Kindle 在电子阅读器市场的竞争与合作	姚忠 徐彪	123
023	基于论据的电子商务人机谈判模型	曹慕昆 庞俊杰	131
024	社会化推荐研究综述	王刚 蒋军 王含茹	137
025	基于知识元的科技文本资源组织方式研究	刘杰 秦春秀 赵捧 未 刘怀亮	147
026	大数据背景下威胁评估对网络隐私顾虑的影响：组织隐私政策的调节	曲静 谢卫红 屈喜 凤 张延林	154
027	在线评论对 O2O 模式创新扩散的影响研究	杜宾	164
028	技术压力对用户满意度的影响机制研究——角色压力的中介作用	王玮 喻亚琴 宋宝 香	171
031	移动 O2O 模式下消费者购买意愿影响因素分析——基于公平理论论和 BRA 模型的实证研究	刘百灵 孙文静 夏 惠敏 徐伟	178
032	不确定需求下双寡头软件厂商两时期发布策略	王宇 李敏强 冯楠 陈富赞 田津	184

033	企业社交工作平台对工作—非工作边界影响研究	孙元 吴丽霞 潘绵臻	190
034	活跃用户影响力对投资组合绩效的影响——基于雪球网投资社群的探索	张偲妮 吴俊 殷雯	195
035	共享单车扩散过程中规范的采纳行为研究	马源鸿 曹云忠 刘佩雯 李敏	202
037	大数据与企业绩效关系研究——基于资源与分析洞察能力视角	谢卫红 钟苏梅 苏芳 王永健 王田绘	208
038	基于 CNKI 网络口碑领域文献的计量分析	张欢 张宁 尹乐民 李娜	219
039	一种考虑评论特征权重的在线评论子集提取方法	倪乃晨 张瑾 任明	226
040	基于公众兴趣的政务微博传播效果因素研究	冯小东 汤志伟 张会平	232
041	基于二部图的电子商务退货风险预测研究	张亮 刘冠男 马宝君	239
042	社交媒体用户倦怠与消极使用：基于扎根理论的探索性研究	李旭 张冰倩 刘鲁川	246
043	特定兴趣领域的社交媒体用户影响力研究	滕德宁 芦鹏宇	254
044	Optimal Dynamic Pricing of Online Platforms with Network Externalities	Guofang Nan, Lin Chen, Tianyu Wang, Mingqiang Li	260
048	双边网络外部性下平台企业最优产品线设计	金治州 冯海洋 李敏强	266
049	三维全景漫游系统用户使用行为影响因素研究	胡丽雪 方佳明 邵培基 曹云忠	272
050	降价式拍卖消费者学习效应研究——以贡天下为例	徐亚男 杨波 杜书彧	277
051	企业大数据治理模式的多案例研究	田金英 杨波 胡梦可	289
052	网络降价式限量拍卖中的一元效应研究——以贡天下为例	陈宇菲 杨波 张晶	297
053	降价式拍卖中消费者重复购买的影响因素研究	王星 杨波 马茜	306
054	网约车转型中的司机激励因素探讨	侯婷 程絮森	315
055	众筹项目融资绩效影响因素的计量经济分析——以淘宝众筹为例	张文涛 闫相斌 陈越	319
056	不同阶段微博口碑情感对票房影响的研究	俞一凡 黄京华 宋婷	326
057	金钱激励能导致更多的在线用户参与行为吗	匡丽妮 杨涵 颜志军	332
058	虚拟品牌社群人际互动对产品购买决策的影响	赵晓燕 沈波	337
059	基于隐私设计的信息系统安全研究	步飞 王能民	343
060	品牌转换意愿的形成机理研究：基于品牌和替代者的双重视角	丁晓燕 张新 张戈	349
061	产品众筹资助人发起人情情绪与融资动态关系研究	齐子豪 郑海超 李立婷 李万庚	355
062	基于 TOPSIS 的知识密集型众包任务人才选择模型实证研究	赵杨 袁析妮	362

063	支持数据驱动决策的情境建模：基于本体的方法研究	宋容嘉 王英 崔维平 黄磊	369
064	信息系统实施后工作特征及系统特征变化对员工工作绩效的影响机制研究	王玮 肖春华 宋宝香	377
066	基于组织承诺与组织控制的信息安全遵从研究	刘晨晖 王能民	384
067	基于 SERVPERF 模型的电商快递服务质量评价指标研究	金丹 任烜毅	394
068	在线问答社区信息价值影响因素的扎根理论分析	孙晓宁 齐云飞 赵宇翔 朱庆华	399
069	基于知识管理的电子政务信息资源建设	常金玲 刘青青	406
074	生产成本依赖规模经济效应的供应链需求扰动管理	李陶然 刘东苏	413
075	中国大陆内外信息资源管理研究现状分析	张甦 王宇	419
077	Understanding Interpersonal Work Connections and Personal Relationships between Co-workers through Electronic Communication Networks: A Case Study in China	WANG Youying, HUANG Qian, Robert M.DAVISON	425
078	虚拟社会资本对网红名人商业价值的影响分析研究	黄月涵 谷钰 洪帆	435
079	基于博弈论的企业信息安全技术评价	李晓彤 李华 杜黎	447
080	企业微信影响力及服务质量研究：以物流企业为例	杜松华 柯晓波 朱琳	452
081	基于 Word2vec 的微博多类别情感分析及电影票房预测	关琳 俞一凡 黄京华	458
082	协同消费参与意向的社会多维价值影响因素研究	蔡舜 彭志伟 张意成 庞晓 丁国维	464
083	基于深度学习 CNN 与协同过滤的在线商品推荐方法	管悦 陈国青 卫强	474
084	患者如何选择线上医生——医生贡献和患者回馈视角	洪紫映 张韦 邓朝华 刘汕	481
085	社会化商务意愿的实证研究：S-O-R 视角	甘春梅 林恬恬 许嘉仪	487
086	云计算对上市公司财务绩效影响的实证研究	李正华 王念新 葛世伦	493
087	基于语义查询扩展的关联主题推荐研究	霍辰辉 刘东苏	500
088	微信朋友圈用户潜水意向影响因素研究	刘晓丹 闵庆飞 刘子龙	506
089	智慧居家养老感知数据预处理研究	左美云 侯静波 蒋立新	512
090	基于机器学习的公益众筹项目融资能力评价模型研究	赵杨 武立强 谭道勋	521
091	电子政务云吸收与价值影响机制研究	梁乙凯 戚桂杰 周蕊	528
092	对未来关系的重视与 IT 企业创新合作	曾维君 艾宏峰	537
094	农产品电子商务顾客满意度和忠诚度的影响因素——基于产品类型的对比研究	李婷 罗志梅 林家宝 江飞	542
095	制度环境下企业农产品电子商务吸收的影响因素分析	李蕾 罗志梅 胡倩 林家宝	548

096	负面评论的商家回复与潜在消费者购买意向的关系研究	徐琬月 李瀛 陈昊 李文立	554
097	社会化媒体在多语言翻译资源构建中的应用研究	刘伟成 MISCHO William H.	561
098	我国“涉农”电子商务的研究回顾——基于 CSSCI 期刊的统计分析	李良强 邵培基 杨 锐 曹云忠	566
100	移动信息技术与组织控制：文献综述与理论框架	周黎 杨琪 王祎	572
102	在线教育平台中竞争对手类型对用户学习绩效的影响研究	邓泓舒语 郭迅华 陈国青	583
103	社会化问答社区用户知识贡献的影响因素探究	刘子齐 吴鼎 郭迅 华 李纪琛	590
104	基于深度学习的学业状态预测模型研究	王兴芬 孙彦超	597
105	外部压力对政府内部组织要素与开放政府数据质量的调节效应研究	赵玉攀 樊博	604
106	我国互联网境外上市公司的聚类、回归与演化分析	楼润平 孙鹏 毛彧	618
107	基于可变密钥强度的射频支付卡	姜琛凯 潘松洁	628
108	知识付费产品销量影响因素探究	石海荣 蔡舜 傅馨	641
109	XBRL 技术扩散与“一带一路”下财务报告的国际沟通研究	乔鹏程	651
110	基于使用行为分析的共享单车管理优化研究	傅哲 余力	662
112	社交媒体特性对于用户焦虑情绪影响的实验研究	张冰倩 李旭 刘鲁 川	671
113	移动广告价值对消费者态度及购买意向的影响	黄丽娟 李林子 李 成蹊	679
114	移动营销中消费者感知价值构成维度研究——概念界定与量表开发	黄丽娟 贾琳 李成 蹊 李林子	690
115	信息交互能力：概念界定与研究框架	孙璐 李力 乐承毅	701
116	基于心理防御视角的个体网络知识分享研究	李玉豪 王刊良	715
118	传统企业的“互联网+”转型：公司治理与并购经营绩效	张悦悦	723
120	大体积耐用品适用于 B2C 电子商务吗？以家具为例	李英 陈振环 龚敏 赵越	731
121	网络疑病影响因素及对策研究	饶倩雯 李艳红	742
122	基于社会资本的知识基于社会资本的知识产品收入影响因素研究	邱婷 蔡舜 傅馨	747
123	基于微博内容分析的共享单车评价研究	张凌 罗曼曼 罗鹏 程	753
124	基于 Citespace 的企业知识库研究知识图谱	李慧 田亚丹	758
125	智慧养老服务体系及平台构建研究	李彩宁 毕新华	766
126	产品众筹发起人性格对项目融资与执行结果的影响机理研究	李立婷 郑海超 刘 琛	772
127	CIO 拼创行为对组织 IS 创新战略的影响研究	张延林 李礼 白海 青 吴学雁	780
128	微信使用影响了生活满意度吗？基于社会资本视角的解释	郑大庆 黄林 王雨	786
129	国内基于本体的知识服务研究进展：核心内容	孙雨生 白璧娇 廖 盼	791
130	共享房屋平台使用意愿的影响因素：基于 TAM 的实证研究	满溪柳 王洪伟	796

131	社交媒体中企业突发事件的舆情分析研究	周鹤 李良强 袁华 钱宇 侯伦	803
133	地理距离、投资经验与在线众筹投资决策:行业匹配度的调节作用	陈文波 宾颖 焦会 玲	810
135	网络中立, 补贴和内容创新	李枝勇 孙为政	817
136	权限请求界面设计:通过情境信息线索降低授权不确定性	柳君 刘子龙	826
137	社交媒体问题性使用行为的实证研究	王天华 刘子龙	833
139	基于在线农产品评论的消费者情感标签抽取方法研究	白梨霏 李开明 李 良强 邹芳	840
140	公司披露文本分析研究进展	王洪伟 朱林源	847
141	在线短租房源图片对消费者行为意愿的影响	吴江 靳萌萌	853
142	基于网络数据的投资者情绪和股市相关性分析	刘赫 尚维 孙毅 汪 寿阳	861
143	电子商务环境下多平台消费者价格偏好研究	张友莎 钱宇 袁华	867
144	全球数字贸易格局及影响因素研究:基于社会网络分析	陆菁 傅诺	872
146	职业认同与职业倦怠对教师网络实践社用户采纳意愿的影响-基于情感的中介作用	徐光 刘鲁川	882
147	在线客户评论的属性提取与细粒度情感分析:基于深度学习和层次聚类方法的视角	马宝君 陈璐 万岩	891
149	项目描述的欺诈性与众筹投资意愿: 基于文本分析的方法	沈倪 王洪伟	899
152	基于旅游数字足迹的游客时空特征研究——以南京为例	廉同辉 余菜花	904
153	移动信息技术与组织结构、员工授权的跨层次关系研究	杨琪 周黎 王祎	911
154	基于在线招聘大数据的劳动力市场分析	刘耘	924
155	大学生对智能手机平面广告认知偏好研究	柴亚青 李慧	930
156	企业业务软件的云化迁移决策研究	何梦娇 任南 苗虹	937
157	网络效应在信息产品“免费 + 增值”模式中的作用机制	李伟	945
159	基于贝叶斯网络的食品安全风险预警研究	莫名垚 姜同强	950
160	网站质量测量评估及其影响的元分析研究	叶许红 韩芳芳 陈 慧栋	957
161	基于层次结构知识元的文本资源语义空间	秦春秀 李祯静 刘 杰 谢庆球	963
162	制造企业 IT 能力对创新绩效的影响: 数字化转型的中介作用	池毛毛 王伟军 陈 秋阳	972
163	信息技术对科研论文引用的影响:基于经济学发表的实证	马鹏飞 张丹煜 张 诚 徐云杰	982
164	颜值对网上交友影响的研究	张琦 张诚	988
165	新兴技术背景下的高校教育变革	赵钊 贺荣戈	995
166	基于云技术新型架构的育种数据服务平台	岳媛 赵刚	1000
167	基于 SNA 和 DMR 的慢病社群探测与主题演化趋势研究---以高血压为例	周利琴 潘建鹏 张 斌	1006
168	众筹中顾客参与和产品市场表现的关系研究	苏颖 林丽慧	1016
169	产品共享对企业定价策略的影响研究	丁灵 廖貅武 杨橹	1023
170	社交媒体中食品安全风险沟通效果研究	史丰源 何德华 方 雯琳	1030

171	SHARING DATA ON FITNESS APPLICATIONS: THE IMPACT OF SOCIAL FACTORS	ZHOU Ya, KANKANHALLI Atreyi, PHANG Chee Wei	1039
172	G-S 匹配机制在招聘网站的可用性研究	付自强 孙永洪 姜 红丙	1045
173	远程问诊中不良医患关系的影响因素研究	晏梦灵 谭鸿瀛 梁 嘉熠 李晨昱	1053
174	人民币国际化之区域经济联动与金融科技发展	王超 王坚	1061
175	全球政府开放数据在数字经济中的贡献评估报告	范佳佳	1069
176	基于 TAM 的众包物联网模型在灾难响应中的应用	黄虎 韩水华	1081
177	基于深度学习 LSTM 的电商销量预测研究	武玉英 严勇 何喜 军 蒋国瑞	1088
178	物流信息系统“不当可视化”对顾客信息焦虑影响研究	李曼宁 袁月 董宜 斐 于美婷	1097

基于二部图的电子商务退货风险预测研究*

张亮¹, 刘冠男¹, 马宝君²

(1. 北京航空航天大学经济管理学院, 北京 100191; 2. 北京邮电大学经济管理学院, 北京 100876)

文 摘: 退货是电子商务企业越来越需要重视的问题, 据统计, 电商企业产品的平均退货率高达三分之一。因此, 本文聚焦于电子商务中的退货问题, 深入地分析用户的退货行为, 进而研究电商交易中的退货风险, 这对于提高电子商务企业的管理决策水平具有重要意义。针对退货风险的预测, 本文定义了一笔交易的退货风险, 并通过引入二部图结构来组织历史退货记录, 将问题形式化为二部图上的排序问题。同时, 基于实际退货数据的观测, 发掘出描述产品和用户的特征属性, 将其转化为先验信息引入模型, 使得模型具有较好的可扩展性。通过在实际数据集上的实验表明, 本研究所提出的模型相较其他方法有更良好的性能。

关键词: 电子商务; 退货; 随机游走; 推荐

中图分类号: 分类号 1; 分类号 2

在互联网不断渗透进各行各业的今天, 网上购物因其方便快捷的属性已越来越成为用户生活方式的一部分, 这极大地推动了电子商务的发展。然而, 电子商务因其固有的弊端使得顾客无法获得商品的实际信息, 以确定是否满足自己的需要, 只能依赖于商品描述、照片等媒介。所以顾客更加容易对商品的价值估计产生偏差, 这将会增加实际产品与顾客需求不一致的现象。在这种情况下, 顾客便可能产生退货行为。一笔意料之外的退货可能会给制造商和零售商带来巨大的损失, 例如零售商必须根据退回产品的情况进行库存调整; 制造商可能要调整其生产计划等。有数据表明, 在美国, 每年因为产品退货产生的成本损失达到了上千亿美元^[1]。我国的情况也同样不容乐观, 据零售咨询公司嘉思明统计, 电商企业产品的平均退货率高达三分之一^[2]。所以退货是一个需要高度重视的问题。

对于该问题, 以往的研究主要是从营销和运作管理的角度出发的。如 Pasternack 等人研究了定价策略和退货政策, 提出了一种对于短期寿命商品的层次定价模型^[3]。Mukhopadhyay 等人发现提供友好的退货政策能够增加收入, 但同时也会由于高昂的退货和设计费用增加成本, 并基于此提出了一种优化退货政策的最大化模型^[4]。Y.Li 等人设计了不同的模型检验在线购物中退货政策、商品价格、商品质量对于消费者购买意愿和退货意愿的影响, 发现这些要素的影响是相互作用和耦合的^[5]。特别地, De 等人实证地研究了电商平台中信息技术的使用

对于退货的影响, 包括图片, 网站排版、文字描述等^[6]。本文的研究内容则与之前的文献稍有不同, 从更为微观和个性化的角度出发, 通过挖掘用户在退货过程中的行为模式, 预测了一笔交易最终被退货的风险, 这些结果将有益于指导电子商务企业的营销实践。

1 退货风险的定义

退货是用户的一项综合的决策过程, 反映了消费者的偏好特征和个性化的行为模式。不同的消费者对于商品评价的侧重点不同, 这与购物过程存在内在的一致, 二者都是对消费者未知的偏好的深度挖掘。因而可以借鉴推荐问题的相似定义来定义退货的风险。

故本研究问题可以被形式化地描述如下: 令某电子商务平台中的用户集合为 U , 平台上的产品集合为 I , 用户的特征属性集合为 F_U , 产品的特征属性集合为 F_I 。每个用户 $u \in U$ 产生的多次退货记录为集合 T , 形如 (u, i) , 其中 $u \in U, i \in I$ 。对每一个用户 U , 通过推荐算法可以得到一个长度为 L 的产品列表 R_u , 其中 $R_u \in I$ 。当用户购买的产品 $i \in R_u$ 时, 即可以被认为是一笔高风险交易。

为了更清晰地描述用户与产品之间的关系, 本研究引入了二部图的概念。令二部图为 $G=(U \cup I, E)$, U, I 分别为用户集合和产品集合, 是两类不同的节点。 E 是该二部图的边集, 由历史退货记录集合 T 生成, 形如 (u, i, w_{jk}) 。 w_{jk} 表示用户对产品 i 的偏好程度, 其中 $j \leq |U|, k \leq |I|$, 分别表示用户 u 与产品 i

*作者简介: 张亮 (1995-), 男 (汉族), 硕士研究生。

通讯联系人: 刘冠男, 讲师, E-mail: liugan@buaa.edu.cn

的索引。所有的 w_{jk} 共同组成了边集的权重矩阵 $W = \{w_{jk}\}$ 。对每一个用户节点和每一个产品节点而言，图上的重要属性是度，所以引入两个由权重矩阵 W 生成的对角矩阵 D_U 和 D_I 。例如索引为 j 的用户，度的大小为 $(D_U)_{jj}$ ；同理索引为 k 的产品，度的大小为 $(D_I)_{kk}$ 。

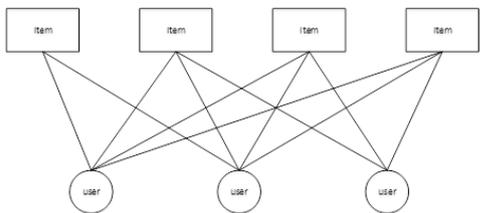


图1 二部图结构示意图

由于特征属性的存在，该二部图中的节点具有异质性。至此，问题被形式化为异质二部图上的推荐问题。

上述定义中涉及到的数学符号及其意义见下表。

表1 本研究数学符号定义

符号	描述
$U = \{u\}$	用户集合
$I = \{i\}$	产品集合
F_U	用户的特征属性集合
F_I	产品的特征属性集合
$T = \{(u, i)\}$	历史退货记录集合
R_u	长度为L的推荐列表
$G = (U \cup I, E)$	用户产品二部图
E	二部图的边集
$W = \{w_{jk}\}$	二部图的权重矩阵
D_U, D_I	用户节点和产品节点的度矩阵

2 退货行为的观测

为了分析用户的退货行为，得到合理的 F_U 和 F_I ，我对一家天猫网店的实际退货数据进行了数据观测。

2.1 用户特征的观测

观测 1：用户的信用评分对于退货行为可以产生很强的影响，不同信用评分区段的用户具有不同的退货特征，高信用评分的用户显然对退货行为更为谨慎。

图2展示了拥有不同信用评分的用户的退货次数分布，其中高退货次数的用户主要集中在低信用评分区段；当信用评分超过 2000 时，退货次数基本稳定在 1 次。总体呈现出负相似的趋势。

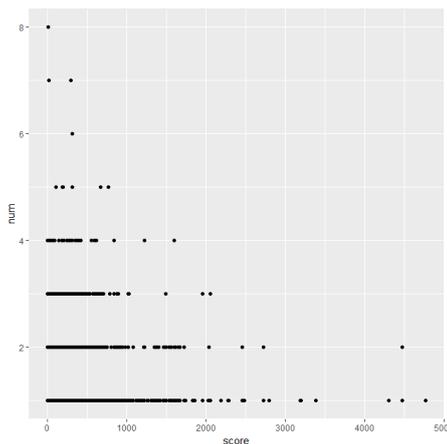


图2 不同信用评分用户的退货次数分布

2.2 产品特征的观测

该数据集用来刻画产品的属性主要包括价格、是否有证书、是否有发票和运费支付方这几项。

观测 2：不同价格的产品具有不同的退货特征，且被退货的可能性随产品价格的升高而降低。

图3展示了产品价格与退货次数之间的关系，随着产品价格的升高，退货次数逐渐降低，二者呈现负相似的关系。因而产品价格可以作为预测退货风险的一大特征。

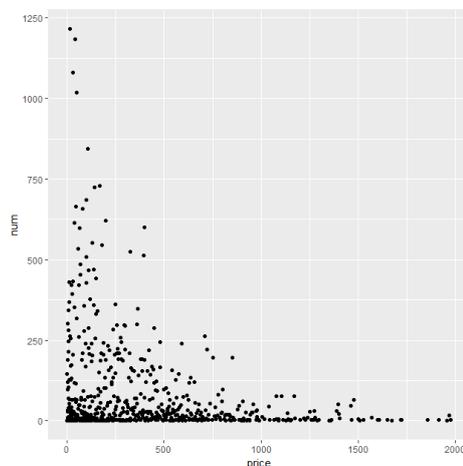


图3 不同价格产品的退货次数分布

观测 3：产品的运费支付方是影响退货风险的重要因素。

图4展示了产品的运费支付方与相对退货次数之间的关系。提出相对退货次数是为了消除产品属性本身的数量差异，如在退货数据中卖家支付运费的记录数为 200，而在产品集合中，拥有该属性的产品数为 400，则相对退货次数为 0.5。

从图4可以看出，当运费支付方为买家时，用户对该笔交易的评价会更为苛刻，退货的可能性急剧增加。

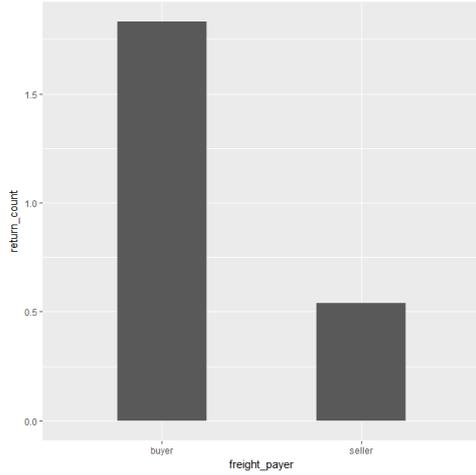


图4 产品的运费支付方的相对退货次数分布

观测 4: 产品是否拥有证书对退货行为具有很强的影响。

图5展示了产品是否拥有证书与相对退货次数之间的关系。可以看出, 没有证书的产品被退货的风险远高于拥有证书的产品。这一方面可能因为没有证书的产品总体上质量较差, 另一方面也可能因为用户对没有证书的产品心存芥蒂, 拥有相对不公正的评价标准。

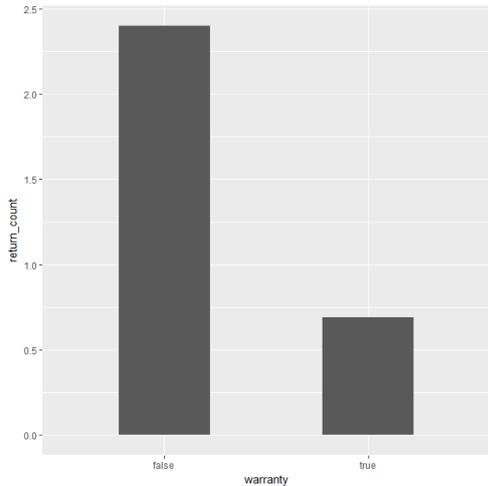


图5 产品是否拥有证书的相对退货次数分布

3 基于二部图的退货风险预测模型

在推荐系统中, 协同过滤是应用最成功且应用最为广泛的个性化推荐方法。协同过滤基于拥有相似兴趣的用户会喜好相似的产品这个假设来挖掘历史交易数据, 其核心是如何度量不同用户之间的相似度和不同商品之间的相似度。例如基于邻近的协同过滤直接使用统计学上的相似性作为度量指标, 而基于潜在因子的协同过滤则通过把用户/产品映射到一个隐含空间来计算相似性等。

在退货数据中, 用户之间的相似性来自于两方面: 退了同一种产品的两个用户是相似的; 信用评分相近的两个用户是相似的。相应的, 产品之间的相似性也来自于两方面: 被同一用户退的两个产品之间是相似的; 特征属性取值接近的两个产品是相似的。其中, 第一种相似性是根据历史退货记录的不同而动态确定的, 而第二种相似性为用户、产品自身固有的, 是一种静态的先验信息。

3.1 静态相似性的度量

1. 用户静态相似性的度量

在退货数据集中, 用户的特征属性主要是信用评分, 根据观测 1 可知, 不同信用评分的用户具有不同的退货行为特征, 可以很好地用来量化用户的静态相似性。为了将量纲归一化为[0,1], 设计如下的相似性函数。

$$S_U(x, y) = \frac{|S_x - S_y|}{\max(S_x - \min_{1 \leq i \leq |U|} S_i, \max_{1 \leq i \leq |U|} S_i - S_x)} \quad (1)$$

其中, x 表示被推荐的用户, y 表示尚未评价的用户, S_x, S_y 分别表示用户 x 和用户 y 的信用评分。当用户 y 是目标用户时, $S_U(x, y)$ 值为 0; 当用户 y 与目标用户信用评分差距最大时, $S_U(x, y)$ 的取值为 1。且 $S_U(x, y)$ 在 0 与 1 之间具有良好的线性变化性质。

2. 产品静态相似性的度量

根据之前的观测, 产品的特征属性主要包括价格、是否有证书、是否由消费者支付运费三项。其中价格是连续性变量, 其他两项是二值变量。为了消除量纲的影响, 先对价格属性进行归一化处理, 归一化函数与上小节相同。

$$f(x, y) = \frac{|P_x - P_y|}{\max(P_x - \min_{1 \leq j \leq |I|} P_j, \max_{1 \leq j \leq |I|} P_j - P_x)} \quad (2)$$

x 表示目标产品, y 表示尚未评价的产品, P_x, P_y 分别表示产品 x 和产品 y 的价格。

令产品 x, y 经过归一化后的特征向量为 $(F_I)_x$ 和 $(F_I)_y$, 采用调整的相似系数作为产品之间相似性的度量函数。

$$S_I(x, y) = 0.5 \cdot \frac{\text{Cov}((F_I)_x, (F_I)_y)}{\sqrt{D((F_I)_x)} \sqrt{D((F_I)_y)}} + 0.5 \quad (3)$$

3.2 动态相似性的度量

动态相似性的度量需要利用历史退货记录和二部图的结构。考虑一个真实的推荐情境, 对于目标用户 u , 可以构造产品集合 $I(u) = \{i\}$, 其中

$(u, i) \in T$ 。 $I(u)$ 产品与目标用户具有较强的相似性。于是，退了 $I(u)$ 中产品的用户 u_k 则与目标用户具有较强的相似性。与此同时，用户 u_k 所退的产品集合 $I(u_k)$ 也与目标用户产生了相似性，循环迭代，则可以生成与目标用户最相似的用户节点集以及产品节点集。上述过程可形式化地描述为以下两个迭代规则：

$$u_j = \sum_{k=1}^{|I|} w_{jk} i_k \quad (4)$$

$$i_k = \sum_{j=1}^{|U|} w_{jk} u_j \quad (5)$$

然而根据 Zhou 等人的研究，上述形式的迭代规则不容易平稳地收敛，很容易受到异常点和参数设置的影响，所以需要进行形式上的正则化处理 [7][8]。本研究采用类似的图的对称正则方法进行平滑处理，正则化后的迭代规则如下：

$$u_j = \sum_{k=1}^{|I|} \frac{w_{jk}}{\sqrt{d_j} \sqrt{d_k}} i_k \quad (6)$$

$$i_k = \sum_{j=1}^{|U|} \frac{w_{jk}}{\sqrt{d_j} \sqrt{d_k}} u_j \quad (7)$$

迭代收敛后的 u 与 i 就是与目标用户最相似的用户节点集以及产品节点集。

3.3 静态相似性和动态相似性的结合

动态相似性来自于对二部图的图结构的挖掘，而静态相似性来自于对二部图的节点的特征信息的挖掘。为了提高推荐的准确性，应该将静态相似性作为算法的先验信息输入，同时使用动态相似性和静态相似性以提高信息的利用率。

1. 用户先验信息的构建

在实际的推荐问题中，通过 $S_U(x, y)$ 函数计算所有用户与目标用户 u 的静态相似性，可以生成用户的先验信息 u^0 。

$$u^0 = \{S_U(u_i, u)\}, 1 \leq i \leq |U| \quad (8)$$

2. 产品先验信息的构建

由于问题的目标是寻找目标用户最可能退货的产品列表，所以产品的先验信息应该表达出产品与目标用户之间的相似性，这里采用退货次数占比作为这种相似性的度量，记为 $r(u, i)$ 。

$$r(u, i) = \frac{|T_{u,i}|}{|T_u|} \quad (9)$$

$T_{u,i}$ 表示退货记录中用户为 u 产品为 i 的记录集合， T_u 表示退货记录中用户为 u 的记录集合。但用户的退货记录总体上是较为稀疏的，目标用户对很多产品的退货次数可能是 0，难以进行有效的区

分。所以本研究基于产品的静态相似性引入了平均退货次数占比的概念，记为 $C(u, i)$ 。

$$C(u, i) = \begin{cases} \frac{\sum_{k=1}^{|I(u)|} S_I(i, i(u)_k) \cdot r(u, i)}{|I(u)|} & i \notin I(u) \\ r(u, i) & i \in I(u) \end{cases} \quad (10)$$

计算目标用户 u 所有退货次数为 0 的产品的平均退货次数，可以生成产品的先验信息 i^0 。

$$i^0 = \{C(u, i_k)\}, 1 \leq k \leq |I| \quad (11)$$

3. 加入先验信息的迭代规则

引入超参数 α, β 对动态相似性和静态相似性信息进行线性组合，可以得到最终的迭代规则。

$$u_j = \alpha \sum_{k=1}^{|I|} \frac{w_{jk}}{\sqrt{d_j} \sqrt{d_k}} i_k + (1 - \alpha) u_j^0, 0 \leq \alpha \leq 1 \quad (12)$$

$$i_k = \beta \sum_{j=1}^{|U|} \frac{w_{jk}}{\sqrt{d_j} \sqrt{d_k}} u_j + (1 - \beta) i_k^0, 0 \leq \beta \leq 1 \quad (13)$$

α, β 分别表示了用户先验信息与产品先验信息的重要性。上述规则可以使用向量形式更为简洁地表达如下。

$$u = \alpha (D_U^{-\frac{1}{2}} W D_I^{-\frac{1}{2}}) i + (1 - \alpha) u^0, 0 \leq \alpha \leq 1 \quad (14)$$

$$i = \beta (D_U^{-\frac{1}{2}} W D_I^{-\frac{1}{2}})^T u + (1 - \beta) i^0, 0 \leq \beta \leq 1 \quad (15)$$

上述迭代规则是基于二部图的退货风险预测模型的核心，算法输入用户先验信息、产品先验信息、权重矩阵以及超参数 α, β ，经过多次的迭代，收敛后输出按相似性排序的用户集合 u 和产品集合 i ，排名前 L 的产品集合作为推荐列表 $R(t)$ 。

算法 1. 基于二部图的退货风险预测模型

Input: User prior information u^0 , Product prior information i^0 , Weight matrix W , hyperparameters α, β ;

Output: User vector u , Product vector i ;

1: Symmetrically normalize W : $S = D_U^{-\frac{1}{2}} W D_I^{-\frac{1}{2}}$;

2: Randomly initialize i and u ;

3: while Stopping criteria is not met do:

4: $u = \alpha (D_U^{-\frac{1}{2}} W D_I^{-\frac{1}{2}}) i + (1 - \alpha) u^0$;

5: $i = \beta (D_U^{-\frac{1}{2}} W D_I^{-\frac{1}{2}})^T u + (1 - \beta) i^0$;

6: end

7: return i and u

4 实验结果

4.1 实验设计

本研究从淘宝网某在线商家中获取了交易数据,淘宝网是阿里巴巴旗下的电子商务 B2C 购物网站,是我国目前最大的电子商务平台之一。该商家主要经营护肤产品,包括面霜、面膜、香水等产品。该数据集包含了用户记录、产品记录、以及 2013 年全年的退货记录。为了更好地发现退货行为中隐藏的模式,本文对发生频繁退货的用户和产品进行了采样,保留了退货次数超过 3 的用户以及退货次数超过 3 次的产品。并抽取了用户的信用评分作为用户特征,产品价格、运费支付方和证书状态作为产品特征。抽样后形成的新数据集的相似统计数据如表 2 所示。

表 2 数据集描述

描述	数值
用户数	1258
产品数	1955
退货记录数	6456
用户平均退货次数	5.13
产品平均退货次数	3.31

将每个用户的退货记录按照 3:1 的比例进行划分,取其中的三份划入训练集,其余的划入测试集。这种划分方法避免了将某个用户的所有记录都划入了训练集或测试集的极端问题,在此基础上,可以进行实验如下。

4.2 实验比较方法与评价指标

1. 实验比较方法

为了验证本研究提出的算法 Rrank 的实际推荐效果,选取了如下一些常用的推荐方法作为基准比较方法。

- 基于物品的协同过滤 (ItemCF)

基于物品的协同过滤的基本思想是向用户推荐和他们之前偏好的物品相似的物品。该算法认为,物品 A 与物品 B 具有很强的相似性是因为偏好物品 A 的用户也更倾向于偏好物品 B。记物品 A 的退货向量为 I_A , B 的退货向量为 I_B , 用余弦夹角计算二者之间的相似度可以得到物品的相似度矩阵。对于目标用户,利用物品相似度对用户偏好程度进行加权平均,经排序后可输出推荐列表 $R(u_i)$ 。

$$S_I(I_A, I_B) = \frac{\sum_{j=1}^{|I_A|} (I_{A_j} \cdot I_{B_j})}{\sqrt{\sum_{j=1}^{|I_A|} (I_{A_j})^2} \sqrt{\sum_{j=1}^{|I_B|} (I_{B_j})^2}} \quad (16)$$

- 基于用户的协同过滤 (UserCF)

基于用户的协同过滤的基本思想是向用户

推荐与其相似的用户所偏好的物品。该算法认为,用户 A 与用户 B 很相似是因为二者偏好同样的产品。记用户 A 的退货向量为 U_A , 用户 B 的退货向量为 U_B , 采用余弦夹角计算二者之间的相似度可以得到用户之间的相似度。对于目标用户,利用用户相似度对产品偏好程度进行加权平均,经排序后可输出推荐列表 $R(u_i)$ 。

$$S_U(U_A, U_B) = \frac{\sum_{j=1}^{|U_A|} (U_{A_j} \cdot U_{B_j})}{\sqrt{\sum_{j=1}^{|U_A|} (U_{A_j})^2} \sqrt{\sum_{j=1}^{|U_B|} (U_{B_j})^2}} \quad (17)$$

- 奇异值分解 (SVD)

奇异值分解是一种矩阵分解的方法,它可以将推荐问题映射到一个隐含空间进行求解。其基本思想是根据已有的退货记录,分析出用户对各个隐含因子的偏好程度以及产品包含各个隐含因子的程度,再反过来根据隐含空间中的分析结果预测用户偏好。典型的奇异值分解如下式:

$$A = U \Sigma V^T \quad (18)$$

- 不含信息先验信息的 Rrank (NF-Rrank)

在本研究提出的退货风险预测模型中,用户和产品的属性特征作为模型的先验信息均被纳入了模型中。为了验证先验信息的重要性和有效性,我们在这里舍弃了这部分信息,将原模型简化为 NF-Rrank,并在下一节将得到的结果与原模型的结果进行对比。

2. 评价指标

- 平均排序值 (RankScore)

该指标用来测量推荐算法能否将用户偏好的项目置于推荐列表的前端位置^[9]。对于用户 u_i , 推荐算法得到的推荐列表为 $R(u_i)$, 长度为 L 。当用户选择的项目 i_k 位于推荐列表的第 n 位时, $r_{ik} = n/L$ 。计算用户 u_i 选择的所有产品的算术平均值,可以得到关于用户的 r_i 。当有多个目标用户时,计算不同用户 r_i 的均值,可以得到算法的平均排序值。

$$r_{ik} = n/L, 0 \leq r_{ik} \leq 1 \quad (19)$$

- 准确率 (Precision)

准确率是反映推荐算法性能的单值指标。推荐列表中命中的项目排名越靠前,准确率越高^[9]。对于用户 u_i , 推荐算法得到的推荐列表为 $R(u_i)$, $R(u_i)$ 中命中的项目为 $hits(u_i)$ 。

$$Precision_{u_i} = \frac{|hits(u_i)|}{|R(u_i)|} \quad (20)$$

- 召回率 (Recall)

召回率是衡量推荐列表中成功命中的项目的

个数占测试集中用户选择的项目个数的比例^[9]。对于用户 u_i ，其选择的项目列表记为 $I(u_i)$ ，推荐命中的项目列表记为 $hits(u_i)$ 。

$$Recall_{u_i} = \frac{|hits(u_i)|}{|I(u_i)|} \quad (21)$$

4.3 退货推荐实验结果

1. 参数敏感性分析

模型包含 α 和 β 两个超参数，分别用来衡量用户先验信息和产品先验信息的重要性，可以根据实际的使用情景自由设置。不同的参数设置可以导致不同的推荐结果，本小节着重分析模型对超参数的敏感性。令推荐列表长度为 10，计算超参数取不同值时算法的性能。

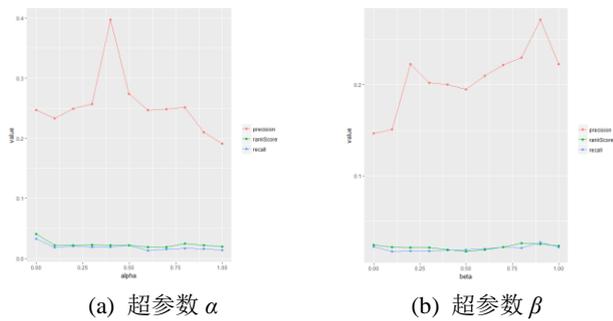


图 6 超参数敏感性分析

取超参数 $\alpha = 0.8$ ，分析参数 β 对模型性能的影响。可以发现随着 β 值的不断增加，模型的准确率和召回率总体上呈上升的趋势，而模型的平均排序值则呈先降后升的趋势。当 $\beta = 0.8$ 时，召回率取到最大值；而当 $\beta = 0.9$ 时，准确率和平均排序值同时取到最大值。考虑到准确率曲线的变化幅度远高于召回率曲线，对 β 的取值更为敏感，所以设置参数 β 的值为 0.9。

取超参数 $\beta = 0.9$ 并保持不变，分析参数 α 对模型性能的影响。可以发现随着 α 值的增大，模型的召回率和平均排序值呈现不断下降的趋势。而模型的准确率则先升后降，当 $\alpha = 0.4$ 时，取得最大值。当 $\alpha = 1$ ，即无任何用户先验信息时，模型的三项参数都下降到了最低点，可见用户信用评分信息对于模型性能的巨大影响。参数 α 的选择依然要在三项指标中进行取舍，不难发现当 α 的值等于 0.4 时，模型的准确率远高于其他情况，极大地提高了模型的性能，在这里取参数 α 的值为 0.4。

2. 算法性能分析

将参数 α 设置为 0.4，参数 β 设置为 0.9，对所有用户进行推荐，比较各个方法在评价指标上的取值如图所示。整体上来看，本研究提出的算法在所有指标上均表现得最好。具体来看，在此数据集上，SVD 算法表现出了极差的性能，这可能时因为评分矩阵十分稀疏，矩阵的分解在重构的过程种积累了

很大的误差；而相比基于用户的协同过滤，基于产品的协同过滤表现出了更好的性能，这可能是因为在该数据集中，用户的退货次数比较接近，而产品的退货次数差距较大，所以基于用户的相似性计算区分度不高；另一方面，缺乏先验信息引导的 NF-Rrank 算法表现并不好，这足以说明用户和产品的特征属性在一定程度上提高了模型与数据的拟合，也提高了推荐的质量，对于迭代的引导作用是不可忽视的。

值得注意的是，在准确率这项指标上，Rrank 算法的取值随着推荐列表长度的增加而先升后降，这可以说明目标用户最偏好的产品几乎都位于推荐列表的较前端，表现出了很高的推荐质量。

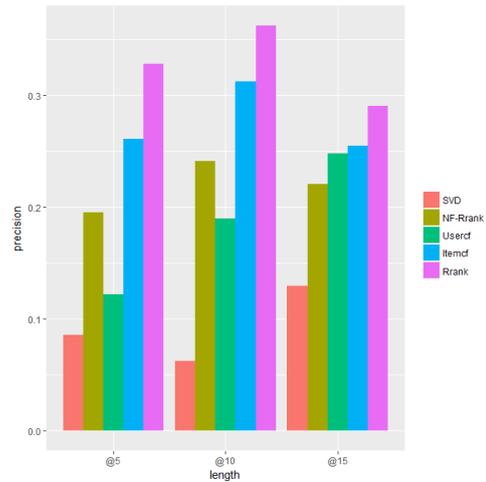


图 7 不同推荐算法的准确率

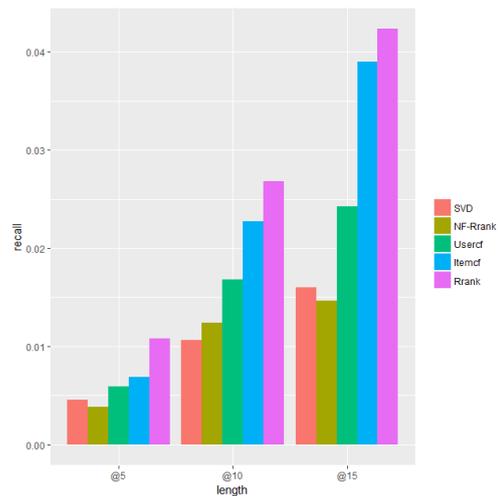


图 8 不同推荐算法的召回率

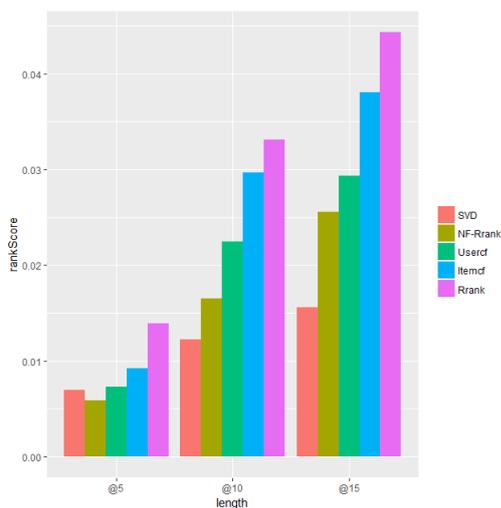


图9 不同推荐算法的平均排序值

5 结论

退货过程和购物过程存在内在的一致，都反映了消费者的偏好特征和个性化的行为模式。而消费者行为模式的挖掘与发现实际上是对消费者隐含偏好的挖掘。因此，本研究针对电子商务交易中的退货数据进行研究，定义了消费者对产品的偏好程度来定义一笔交易被退货的风险。为了表示消费者的退货结构，引入了二部图来组织历史退货记录，从而将问题形式化为二部图上的推荐问题。基于实际退货数据的观测，对与退货相关的产品和用户的特征属性进行探索，将其转化为先验信息引入模型，使得模型拥有更高的信息利用率。通过在真实数据集上的实验表明，本研究提出的模型具有相较其他方法更高的性能。

参考文献

- [1] Anderson E T, Hansen K, Simester D. The Option Value of Returns: Theory and Empirical Evidence[J]. *Marketing Science*, 2009, 28(3):405-423.
- [2] 聂林海. 我国电子商务创新与规范发展[J]. *中国流通经济*, 2016, 30(6):52-57.
- [3] Pasternack B A. Commentary---Optimal Pricing and Return Policies for Perishable Commodities[J]. *Marketing Science*, 1985, 4(2):166-176.
- [4] Mukhopadhyay S K, Setoputro R. Optimal return policy and modular design for build-to-order products[J]. *Journal of Operations Management*, 2005, 23(5):496-506.
- [5] Li Y, Xu L, Li D. Examining relationships between the return policy, product quality, and pricing strategy in online direct selling[J]. *International Journal of Production Economics*, 2013, 144(2):451-460.
- [6] De P, Hu Y J, Rahman M S. Product-Oriented Web Technologies and Product Returns: An Exploratory Study[J]. *Information Systems Research*, 2013, 24(4):998-1010.
- [7] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Scholkopf, "Learning with local and global consistency," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 321-328.
- [8] D. Zhou, J. Weston, A. Gretton, O. Bousquet, and B. Scholkopf, "Ranking on data manifolds," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 169-176.
- [9] Adomavicius G, Tuzhilin a. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(6):734-749.

Predicting the Product Risk in Electronic Commerce based on Bipartite Graph

ZHANG Liang¹, LIU Guannan¹, MA Baojun²

(1. School of Economics and Management, Beihang University, Beijing 100191, China;

2. School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: Recently product returns has become a focal issues in e-commerce platforms, with up to one-third return rate according to statistics from the industry. This paper investigates the product return phenomenon with regards to the return risk of an individual transaction. To that end, in this paper, a novel model based is proposed to better support managerial decisions in practice. With respect to predicting the return risk, this paper studies the driving force of defining the return risk of a single transaction and organize the historical return records as bipartite graph. Thus, the prediction problem is formulated as a ranking problem on the bipartite graph. Based on the observations of the actual return data, the features of products and the users are extracted and transformed into prior information of the model, which makes the model more generalized. The proposed model has been experimented on real world datasets, and it shown to have great performance and strong feasibility.

Key words: electronic commerce; product return; random walk; recommendation