

物流服务评论与顾客购买后行为的关联性研究

王世雄,蓝一珍,李 雪

(浙江理工大学经济管理学院,杭州 310018)

摘 要: 在网络购物平台中,已购顾客的物流服务评论通常影响着新顾客购买决策行为。利用网络爬虫获取京东商城的女装在线评论数据,借助中文分词、情感分析等技术提取评论中物流服务相关要素和顾客购买后行为,得到顾客最关注的物流服务要素和最普遍的顾客购买后行为,并运用FP-growth算法挖掘物流服务要素与顾客购买后行为之间的强关联规则。研究结果揭示了不同顾客购买后行为对物流服务感知的侧重点差异。该结果有助于B2C企业寻找物流服务优化的关键点,为改进物流服务质量提供借鉴。

关键词: 物流服务;内容分析;购买后行为;关联挖掘

中图分类号: F252

文献标志码: A

文章编号: 1673-3851(2017)04-0285-08

天猫、京东等B2C电子商务提供评论功能,顾客发表的评论内容丰富,包含顾客对物流服务的反馈等海量的信息,其背后蕴含着顾客消费规律、消费习惯和购买行为。在网络购物的过程中,顾客通常会关注已购物顾客对商品和服务的评论,并将它们作为购买决策的依据之一。从商务运营角度而言,B2C平台上海量的在线评论是商务大数据的重要组成部分,具有Volume(数量大)、Variety(多样化)、Velocity(快速化)、Value(价值大)等典型特征,已经成为企业的核心生产要素和商务资源。

随着交易量的持续增加,已购买商品的顾客对电子商务物流服务的评论越来越多,他们常常利用电子商务平台提供的评论功能,对物流服务的评论做出正面或负面的评价,数据量大,内容的情绪化、信息的破碎化比较突出,常常影响顾客购买决策行为。如何从海量的在线评论中分离出产品质量、价格变化、物流服务相关内容,分析它们对顾客消费行为的影响,是学术界和产业界均十分关注的热点问题。为此,本文利用网页爬虫工具获取京东商城女装的在线评论内容,采用文本分析的方法提取物流服务评价和顾客购买后行为信息;结合关联挖掘技

术,研究物流服务评论与顾客购买后行为的关联性,为企业准确快速地响应顾客对物流服务的评论、消除负面影响提供支持。

一、文献综述

Byrant等^[1]在2008年提出的大数据在商务、科学和社会领域的应用,是数据科学应用的革命性突破,引起了广大学者和业界高管的重视。最初,数据科学主要应用于科学研究、商品零售业、金融投资业、电信业以及保险业。在B2C电子商务领域,数据科学主要通过捕捉、整合消费者的线上行为、互动方式,使得企业了解顾客的消费行为,进而为企业向顾客推荐产品、优化产品策略、预测供应链模型等提供参考依据。

电子商务蕴含价值链,又依托供应链,物流服务是电子商务服务的核心内容之一。研究物流服务对顾客消费行为、消费心理等方面的影响是当前的热点课题。Gefen等^[2]提出物流服务是影响电子商务顾客忠诚度的重要因素,并根据顾客物流感知建立了服务质量测评量表。Stank等^[3]研究发现,物流服务感知会影响顾客满意度,通过提升物流服务质量

量可以提高顾客满意度,进而提高市场份额。Hofacker等^[4]研究发现,顾客对物流服务的感知可以影响B2C企业的口碑和顾客重购的意愿。张岩岩^[5]结合我国B2C电子商务的特点和物流服务质量理论研究,将B2C物流服务质量定义为:网络购物用户在B2C商城购物后产生的实际物流服务感知与期望物流感知的差距感知。宗蕊等^[6]从顾客角度出发,采用在线客户评论的数据挖掘方法,发现不同类型的顾客对物流服务感知的侧重点有所不同。王家琦等^[7]通过文献分析,得出量化B2C物流服务质量的7个维度,分别为订购质量、定制化服务质量、时间性质量、交付质量、信息质量、补救服务质量和经济性质量。

购买后行为^[8](posting-purchase behavior)是顾客行为的一部分,是指顾客购物活动后,基于对购物活动的感知或评价而产生的心理活动及行为,主要包括购物后给出的评价、重购、推荐等行为。在电子商务的评论系统中,顾客的评论包含大量的对商品的评价、重购决策以及推荐行为等。Seiders等^[9]首次提出网络购物用户的重复购买意愿,将其定义为顾客重购行为,即顾客在网络购物之后,愿意再次购买的一种意愿。李泓荣^[10]则将网购重购行为定义为与商家首次交易承购后,顾客再次浏览该商铺,并有重购行为或重购意向。对于影响顾客重购意愿的因素,Lilijander等^[11]通过研究得出影响顾客重购的三个驱动因素为:顾客的满意度、对品牌的情感度和之前购物或产品使用经历。推荐行为包括线下的“口头”推荐和线上的“在线评论”推荐。现有文献对推荐行为的研究较少,因此,本文通过在线评论中的重购频率和推荐意愿来研究顾客的推荐行为。

前述研究多数以问卷调查的形式获取研究数据,本文则借助信息技术获取在线评论数据,主要包括网络爬虫技术、文本分析技术、中文分词技术和关联挖掘技术等。现阶段,获取在线评论数据主要依靠网络爬虫技术,常用的挖掘软件有Metaseeker、八爪鱼采集系统、乐思信息采集系统等。获得数据之后,需要对评论进行文本分析,判断评论的极性,获取顾客的态度、情绪等数据。

Turney^[12]首先提出运用PMI-IR算法计算词语间的互信息,通过计算与种子词语之间的互信息,判断该词语的情感极性。Dave等^[13]首次提出“观点挖掘”的概念,并研发了系统“ReviewSeer”用来识别评论文本的情感倾向。Hu等^[14]利用关联规则挖掘商品评论中隐藏的商品特征属性,并研发了系统

“Opinion Observer”对网络评论信息进行情感极性识别。

近年来,中文文本分析技术不断取得进展。Tsou^[15]通过研究中文期刊上的600多篇关于政治人物的报道,先标记语料库得到文本中的极性元素,然后统计极性元素的分布、极性元素的密度和极性元素的语义强度,最后计算文本褒贬分类和强度大小。王根等^[16]采用随机对句子进行情感挖掘,并提出基于多重冗余标记的CRF模型。张成功等^[17]通过分析副词、否定词和情感词的不同组合模式,提出在不同组合模式下的具体极性短语的情感计算公式。这些研究为本文的研究奠定良好基础。关联挖掘技术目前相对成熟,Agrawal等^[18]提出的关联规则挖掘被广泛运用于商业领域,其经典算法有:Apriori算法、FP-growth算法等。

总而言之,相对于传统的问卷调查方法,以大数据相关技术获取电子商务平台中的顾客消费、评论数据,以此分析顾客对商家提供的信息、物流、金融等服务的评价,获取顾客消费行为、消费心理以及购买决策等重要商业信息,这种方法具有数据量更大、内容更真实、时空跨度更大等特点,是当前的研究热点。本文以现有研究成果和开源工具技术挖掘京东商城中有关评论,分析物流服务与顾客购买后行为之间的关联性。

二、模型与算法

(一)关联模型

假设 $I = \{I_1, I_2, \dots, I_m\}$ 是项的集合^[19]。设任务相关的数据 D 是数据库事务的集合,其中每个事务 T 是一个非空项集,使得 $T \subseteq I$ 。每一个事务都有一个标识符,称为TID。设 A 是一个项集,事务 T 包含 A ,当且仅当 $A \subseteq T$ 。设顾客购买后行为项集为 W ,物流服务要素项集为 U ,则可定义 A_w 为顾客购买后行为的子项目集, A_u 为物流服务要素的子项目集,其中 $A \subset W, A \subset U$ 。同样的可以定义 B 的项目集,分别为 B_w 和 B_u 。

关联规则是形如 $A \Rightarrow B$ 的蕴含式,其中 $A \subset W, B \subset W, A \neq \phi, B \neq \phi$,并且 $A \cap B = \phi$ 。关联规则 $A \Rightarrow B$ 在事务集 D 中成立,具有支持度 s 、置信度 c 。其中, s 是 D 中包含 $A \cup B$ (即集合 A 和集合 B 的并或 A 和 B 二者)的百分比,如包含物流服务要素 A 和顾客购买后行为 B 的评论占有所有评论的百分比,其值为 $P(A \cup B)$,即:

$$\text{support}(A \Rightarrow B) = P(A \cup B).$$

c 是 D 中包含 A 的事务同时也包含 B 的事务的百分比,即物流服务评论 A 包含顾客购买后行为 B 的评论占所有评论的百分比,其值为 $P(B|A)$,即:

$$confidence(A \Rightarrow B) = P(B|A).$$

根据物流服务要素以及顾客购买后行为的定义,给出关联规则集合如下:

$A_U \Rightarrow B_W$:该规则显示的是物流服务要素与顾客购买后的行为之间的相关性。主要挖掘不同物流服务要素对顾客购买后的行为的关联性强弱,是本文的研究重点。

$A_U \Rightarrow B_U$:该规则显示的是不同的物流服务要素之间的相关性。

$A_W \Rightarrow B_W$:该规则显示的是不同的购物后的行为之间的相关性。

后面两个规则是第一个规则的补充。例如,根据规则 $A_U \Rightarrow B_U$ 挖掘出 $\{u_5\} \rightarrow \{u_4\}$,说明 u_5 和 u_4 两个物流服务要素之间有强关联关系,那么如果 $A_U \Rightarrow B_W$ 中有含有规则 $\{u_5\} \rightarrow \{w_1\}$,则说明具有 w_1 购物行为的顾客在接受物流服务的时候,不仅关注 u_5 的物流服务要素,而且关注 u_4 的物流服务要素。

根据关联规则定义,同时满足最小支持度阈值(min_sup)和最小置信度阈值(min_conf)的规则称为强规则,即:

$$support(A \Rightarrow B) \geq min_sup;$$

$$confidence(A \Rightarrow B) \geq min_conf.$$

(二) 关联性指标

现有研究表明,由于支持度和置信度不足以完全过滤无意义的关联规则,以此为基础的关联性分析存在一定的局限性。为了解决上述问题,可以从增加相关性度量的角度补充关联性指标,如提升度、 χ^2 分析等。本文采用提升度作为关联性分析的补充指标,即关联规则表示为:

$$A \Rightarrow B[support, confidence, correlation].$$

其中,提升度是指在含有 A 的条件下同时含有 B 的可能性与没有这个条件下项集中含有 B 的可能性之比。具体而言,若 $P(A \cup B) = P(A)P(B)$,则项集 A 的出现独立于项集 B 的出现;否则,项集 A 和项集 B 就是相关的。因此,提升度亦可以表示为:

$$lift(A, B) = \frac{P(A \cup B)}{P(A)P(B)}.$$

由于提升度表示的还是当项集 A 出现的提升时,项集 B 出现的程度,所以它还可以定义为:

$$lift(A, B) = \frac{P(B|A)}{P(B)} = \frac{confidence(A \Rightarrow B)}{support(B)}.$$

当 $lift(A \Rightarrow B) < 1$ 时,说明 A 与 B 是负相关的;当 $lift(A \Rightarrow B) > 1$ 时,说明 A 与 B 是正相关的;当 $lift(A \Rightarrow B) = 1$ 时,说明 A 与 B 是独立的。以物流服务要素 A 和顾客后购买行为 B 在评论中出现的情况而言,当 $lift(A \Rightarrow B) < 1$ 时,评论中仅出现线物流服务要素而没有出现相应的顾客购买后行为;当 $lift(A \Rightarrow B) > 1$ 时,评论中出现线物流服务要素的情况下也出现了相应的顾客购买后行为;当 $lift(A \Rightarrow B) = 1$ 时,物流服务要素和顾客购买后行为独立出现在评论中。由此可见,提升度可以增加衡量规则的可靠性,是置信度的一种互补指标。

本文采用提升度来度量物流服务要素与顾客购买行为之间的相关性、不同物流服务要素之间的相关性和不同的购物后的行为之间的相关性,删去提升度小于 1 的关联规则,保留提升度大于 1 的关联规则。

(三) 算法

关联挖掘算法的关键步骤包括:

a) 找出所有的频繁项集:根据关联规则的定义,这些项集的每一个频繁出现的次数都要大于或等于最小支持计数;

b) 有频繁项集产生强关联规则:根据关联规则的定义,这些规则必须满足最小支持度和最小置信度。

其主要的算法有 Apriori 算法、FP-growth 算法等。其中,Apriori 算法使用逐层搜索的迭代方式扫描。首先,通过扫描数据库,累计每个项的技术,并收集满足最小支持度的项,找出频繁为 1 的项集的集合,记为 L_1 ,然后使用 L_1 找出频繁为 2 的项集的集合,记为 L_2 ,使用 L_2 找出 L_3 。以此类推,直到不能再找到频繁 k 的项集。该算法的优点是简单、易理解、对数据要求低,缺点是要多次扫描原始数据,当原始数据较大时,磁盘 I/O 次数太多,效率较低。为了解决 Apriori 算法效率低下的缺点,Han 等^[19]提出了基于 FP 树生成频繁项集的 FP-growth 算法,该算法采用分而治之的策略。考虑到本文获取的数据量较大,为保证效率,采用 FP-growth 算法进行关联性分析。

首先,通过扫描数据库,得到数据库中的频繁项集 F 和频繁项的支持度 s ,把频繁项集 F 按支持度递降排序记为 L 。

然后,创建频繁模式树(FP-tree)的根节点,记为 T ,标记为“null”。根据 L 中的顺序,对 D 中每个事务 T 进行排序,将 D 中排好序的事务项列表记为

$[p|P]$, 其中 p 是第一个元素, P 是列表的剩余部分, 调用递归函数。

三、方法与实现

(一)数据收集

在当前的 B2C 平台中, 京东商城提供的物流服务一直领先于其他竞争对手, 这源于其强大的自营物流体系。该平台中顾客对其物流服务的评论无论是数量还是质量, 都相对客观, 因而具有较好的研究价值。因此, 本文以京东商城的评论内容作为研究对象。另外一方面, 由于京东商城出售的产品种类较多, 不同产品的需求量和物流服务水平差异较大, 考虑到女装的销售量和评论量均较大, 评论内容更具可比性, 因而仅获取与女装销售相关的物流服务和顾客购买后行为的评论内容, 以提高研究的准确性和可靠性。

本文在 2016 年 7 月 1 日至 2016 年 8 月 1 日期间, 利用深圳乐思软件技术有限公司提供的网页信息采集工具——乐思信息采集系统, 获得 2013 年 10 月 16 日至 2016 年 6 月 30 日期间的女装销售相关的数据, 共计 7349 条, 包括商品名、详细页面网址等信息。以此为基础, 利用详细页面网址, 采集到商品的其他详细信息, 包括商品名称、评论总数、格式化评论信息(包括好评数、中评数、差评数、好评率)、评价客户、顾客等级、评价内容、评价日期等信息, 共 705719 条数据。这些数据具有较好的结构, 所承载的信息量很大, 具有较好的研究价值。

(二)数据处理

获取的数据包含的内容丰富, 包括产品名称、总评论数量、好评数量、中评数量、差评数量、好评率、晒图数、评价顾客、顾客等级、评价内容和评价日期等。为了更准确地研究物流服务评论与顾客购买后行为的关联性, 分几个步骤对数据进行处理。

首先, 对抓取的数据进行清洗, 删除缺失值。然后, 利用关键词“物流”、“配送”、“送货”、“快递”、“包装”、“服务”、“态度”、“送货员”、“给力”、“神速”、“完整”、“盒子”、“完整”、“售后服务”、“完好无损”等关键词对清洗后的数据进行筛选, 获得与物流服务和顾客购买后行为相关的评论, 共 74924 条。

接着, 使用 ICTCLAS 分词系统对筛选后的评论进行内容分析, 对筛选出的数据进行中文分词, 得到分词词组, 并进行词频统计, 获取分词结果; 根据分词结果提取关键词, 获得顾客重购行为和推荐行为相关的评论内容, 得到 10923 条。

最后, 根据分词结果判断评论情绪极性, 筛选出情绪极性为正向的数据, 得到 10805 条。根据词频统计结果, 筛选出与物流服务质量相关的词频结果, 并按顺序从高到低进行排序, 得到表 3。

表 3 关于物流服务要素词频统计

词语	词性	词频	词语	词性	词频
物流	n	6375	快速	d	16
服务	v	2735	送货员	n	14
态度	n	2299	到货	v	13
给力	a	2276	送货上门	v	10
速度	n	1957	中通	n	10
发货	v	1823	圆通	n	8
快递	v	1795	效率	n	8
包装	n	1599	外包装	n	7
送货	v	174	咨询	v	5
快递员	n	142	运输	v	5
神速	a	58	退换	v	5
负责	v	57	速	a	5
售后服务	n	31	配	v	5
完好无损	v	28	派	v	5
完整	a	20	订单	n	5
配送	v	20	打电话	v	5
盒子	n	19	帅哥	n	5

依据表 3, 可以看到, 与物流服务质量相关的高频词(出现次数大于 100)有: “物流”、“服务”、“态度”、“给力”、“速度”、“发货”、“快递”、“包装”、“送货”、“快递员”等, 即对应在线评论中顾客最为关注的物流服务要素, 与物流服务质量相关的中频词(出现次数 10~100 之间)有“神速”、“负责”、“售后服务”、“完好无损”、“完整”、“配送”、“盒子”等, 即代表着顾客比较关注的物流服务要素, 与物流服务质量相关的低频词(出现次数小于 10)有“圆通”、“效率”、“外包装”、“咨询”、“运输”、“退换”、“速”、“配”、“派”、“帅哥”等, 即顾客关注的物流服务要素, 但不具有普遍性。整理得到顾客关注的物流服务要素, 即包装完整、发货速度快、运送速度快、送货速度快、快递员服务态度好、售后服务态度和退换货速度。

根据词频统计结果, 筛选出与顾客购买后行为相关的词频结果, 并按顺序从高到低进行排序, 得到表 4。

依据表 4, 可以发现高频词(出现次数大于 300)有“值得”、“购买”、“光顾”、“推荐”、“购物”、“朋友”、“继续”、“下手”、“光临”等; 中频词(出现次数 100~300 之间)有“同事”、“介绍”、“机会”、“赶快”、“合作”、“放心”、“入手”等; 低频词(出现次数小于 100)有“MM”等。组合这些词汇, 可以提炼出以下几种

信息:“帮家人/同学/同事也买了一条”、“有机会再次光临/购买”等关于顾客再次购买的意向或行为;“值得推荐”、“值得购买”、“值得拥有”、“喜欢的MM可以放心购买”、“赶紧入手”、“介绍给朋友/同事/同学”、“值得推荐的商家”等关于顾客推荐产品的行为信息。

表 4 关于顾客购买后行为的词频统计

词语	词性	词频	词语	词性	词频
会	v	4087	赶紧	d	300
来	v	3936	购	v	273
再	d	2316	同事	n	224
值得	v	2176	介绍	v	217
购买	v	1707	机会	n	216
光顾	v	1662	赶快	d	202
推荐	v	1355	合作	v	184
购物	v	876	放心	v	170
朋友	n	817	入手	v	157
继续	v	679	商家	n	154
下手	v	420	店铺	n	151
光临	v	419	准备	v	132
需要	n	361	收藏	v	125
关注	v	344	MM	n	34

词频分析结果显示,与物流服务评论相关的顾客购买后行为主要包括顾客购物后的重购行为和推荐行为。其中,顾客重购行为是指顾客重新购买该商品或顾客有再次购买的意向;顾客推荐行为是指顾客在购买后向朋友或家人推荐购买的商品或所购买的商家以及在网络上向其他顾客进行的推荐性行为。因此,后续研究的重点是物流服务评论与这两种行为的之间关联性。

(三)评论矩阵构建

为了方便关联规则的挖掘,首先从顾客评论中提取出相应的物流服务要素和顾客购买后行为。例如,截取下面一条评论:“全五分好评,店家人好耐心服务,物流快速也不含糊,神速到位,裙子颜色亮丽,款式新颖物美价廉,下次还会再来,并推荐朋友们都来买。”从该评论中可以提取出以下 4 个信息:a)发货速度快;b)快递速度快;c)会重购;d)会推荐朋友购买。然后,将提取出有效的文本信息数据标准化,并将其制成二进制的矩阵形式。设定集合 $R=\{r_1, r_2, r_3, \cdots, r_m\}$ 表示 m 条在线评论的集合,集合 $W=\{\omega_1, \omega_2, \omega_3, \cdots, \omega_n\}$ 表示顾客在购物后的 n 种行为的集合,集合 $U=\{u_1, u_2, u_3, \cdots, u_k\}$ 表示在线评论中的 k 个物流服务要素。

根据顾客购买后行为和物流服务要素的划分,集合 W 包含 2 个元素,集合 U 包含 7 个元素: ω_1 表

示顾客重购行为; ω_2 表示顾客推荐行为; u_1 表示产品包装完整; u_2 表示发货速度快; u_3 表示运送速度快; u_4 表示送货的速度快; u_5 表示快递员服务态度良好; u_6 表示售后服务良好; u_7 表示退换货的速度快和便捷性。则上述评论例子可记录为 $\omega_1 \rightarrow \omega_2 \rightarrow u_2 \rightarrow u_3$ 。

在评论内容中提取物流服务要素取时,可以发现有一些意思表达模糊的词语,例如“快递给力”、“物流给力”、“快递速度给力”等。由于“物流”、“快递”等词语未指明特定的物流服务要素,则本文将类似上述的表述定义为对物流整个运作过程的评价,即“快递给力”表示的是顾客对发货速度、运送速度、送货速度、快递员的服务态度等的满意度,“快递速度给力”表示的是顾客对发货速度、运送速度和送货速度的满意度。同时,类似“送货上门”等词语,虽然没有提及有关的物流服务要素,但是这些词语表达了对快递员的肯定。因此,在整理的时候,需要对上述评论信息进行补充、修改,最后得到评论矩阵。

表 5 列举 5 条不同内容的评论矩阵,矩阵的行表示每条评论,列设为顾客行为和物流服务要素,用 0 或 1 来显示每条评论是否提到某一顾客后行为和物流服务要素,其中 1 表示提到,0 表示未提到。例如,上述评论例子中提到的顾客购买后行为有: ω_1 、 ω_2 、物流服务要素有: u_2 、 u_3 ,则在评论矩阵中记为 $\{1,1,0,1,1,0,0,0,0\}$,即表 5 中的 r_1 。

表 5 在线评论矩阵举例

评论行为 和要素	ω_1	ω_2	u_1	u_2	u_3	u_4	u_5	u_6	u_7
r_1	1	1	0	1	1	0	0	0	0
r_2	1	0	1	0	1	0	0	0	0
r_3	1	1	1	0	0	0	1	1	0
r_4	0	1	1	1	0	0	1	0	0
r_5	1	1	0	0	0	0	0	0	1

(四)分析结果

运用 FP-growth 算法挖掘在线评论矩阵中的关联规则。首先,设置 $min_sup=0.01$, $min_conf=0.3$,保证关联规则之间的强关联性;然后,过滤掉 $lift \leq 1$ 的规则,保证出现物流服务要素的情况下也出现相应的顾客购买后行为;得到 179 条关联规则。滤掉如 $\{\omega_1\} \Rightarrow \{u_6\}$ 类型的规则,筛选出符合研究的关联规则,并将形如 $\{u_2\} \Rightarrow \{u_3\}$ 和 $\{u_3\} \Rightarrow \{u_2\}$ 两种互反规则合并,最终汇总成 68 条数据;其中与顾客重购行为关联的规则有 14 条,如表 6 所示;与顾客推荐行为关联的规则有 9 条,如表 7 所示。

表6 关于顾客重购行为的关联规则

Rule_ w_1 #	<i>lhs</i>	<i>rhs</i>	<i>support</i>	<i>confidence</i>	<i>lift</i>
1	$\{u_5\}$	$\{w_1\}$	0.2091624	0.7084639	1.025583
2	$\{u_3\}$	$\{w_1\}$	0.5231837	0.7001486	1.013546
3	$\{u_4\}$	$\{w_1\}$	0.5332716	0.7016561	1.015728
4	$\{u_3, u_5\}$	$\{w_1\}$	0.1776030	0.7083795	1.025461
5	$\{u_4, u_5\}$	$\{w_1\}$	0.1798242	0.7099014	1.027664
6	$\{u_2, u_5\}$	$\{w_1\}$	0.1806571	0.7026638	1.017187
7	$\{u_3, u_4\}$	$\{w_1\}$	0.5198519	0.6998505	1.013114
8	$\{u_2, u_3\}$	$\{w_1\}$	0.5193892	0.7005368	1.014108
9	$\{u_2, u_4\}$	$\{w_1\}$	0.5194817	0.7010992	1.014922
10	$\{u_3, u_4, u_5\}$	$\{w_1\}$	0.1766775	0.7096654	1.027322
11	$\{u_2, u_3, u_5\}$	$\{w_1\}$	0.1765849	0.7079777	1.024879
12	$\{u_2, u_4, u_5\}$	$\{w_1\}$	0.1764924	0.7099777	1.027774
13	$\{u_2, u_3, u_4\}$	$\{w_1\}$	0.5186488	0.7010258	1.014816
14	$\{u_2, u_3, u_4, u_5\}$	$\{w_1\}$	0.1763073	0.7097615	1.027462

表7 关于顾客推荐行为的关联规则

Rule_ w_2 #	<i>lhs</i>	<i>rhs</i>	<i>support</i>	<i>confidence</i>	<i>lift</i>
1	$\{u_1\}$	$\{w_2\}$	0.059972	0.4167203	1.141649
2	$\{u_2\}$	$\{w_2\}$	0.314021	0.3674066	1.006549
3	$\{u_1, u_3\}$	$\{w_2\}$	0.014993	0.4475138	1.226011
4	$\{u_1, u_4\}$	$\{w_2\}$	0.014160	0.4421965	1.211444
5	$\{u_1, u_2\}$	$\{w_2\}$	0.024711	0.4405941	1.207053
6	$\{u_1, u_3, u_4\}$	$\{w_2\}$	0.014068	0.4457478	1.221173
7	$\{u_1, u_2, u_3\}$	$\{w_2\}$	0.014438	0.4457143	1.221081
8	$\{u_1, u_2, u_4\}$	$\{w_2\}$	0.013882	0.4437870	1.215801
9	$\{u_1, u_2, u_3, u_4\}$	$\{w_2\}$	0.013790	0.4421365	1.211279

根据表6,可以得到以下结论:

a)与顾客重购行为 w_1 具有强关联关系的物流服务要素有:发货速度 u_2 、运送速度 u_3 、送货速度 u_4 和快递员服务态度 u_5 。

b)支持度 *support* 最高的关联规则是 Rule_ w_1 _3,说明在线评论中同时出现顾客重购行为和发货速度的概率最大,即当 B2C 企业满足送货速度快并且准时,顾客重购的可能性最大;

c)置信度 *confidence* 最高的关联规则是 Rule_ w_1 _12,说明在线评论中包含顾客重购行为的情况下含有发货速度、送货速度和快递员的服务态度的概率最大,即当 B2C 企业满足发货速度迅速、送货速度快并且准时和快递员的服务态度良好时,顾客重购的可能性最大;

d)提升度 *lift* 最高的关联规则是 Rule_ w_1 _12,说明在线评论中出现顾客重购行为的情况下也出现发货速度、送货速度和快递员的服务态度的概率最大,即当 B2C 企业满足发货速度迅速、送货速度快并且准时和快递员的服务态度良好时,顾客重购的

可能性最大。

根据表7,可以得到以下结论:

a)与顾客推荐行为 w_2 具有强关联关系的物流服务要素有:包装完整 u_1 、发货速度 u_2 、运送速度 u_3 和送货速度 u_4 。

b)支持度 *support* 最高的关联规则是 Rule_ w_2 _2,说明在线评论中同时出现顾客推荐行为和发货速度的概率最大,即当 B2C 企业满足送货速度快并且准时时,顾客推荐的可能性最大;

c)置信度 *confidence* 最高的关联规则是 Rule_ w_2 _3,说明在线评论中包含顾客推荐行为的情况下含有包装和运送速度的概率最大,即当 B2C 企业满足包装完整和运货速度快时,顾客推荐的可能性最大;

d)提升度 *lift* 最高的关联规则是 Rule_ w_2 _3,说明在线评论中出现顾客推荐行为的情况下也出现包装和运送速度的概率最大,即当 B2C 企业满足发货速度迅速、送货速度快并且准时和快递员的服务态度良好时,顾客推荐的可能性最大。

四、结 论

随着电子商务的交易量持续增加,已购买商品的顾客对电子商务物流服务的评论增多,物流服务评论成为顾客购买决策行为的重要影响因素之一,因此,本文对大量的在线评论进行内容分析提取出顾客最关注的物流服务要素,并运用 FP-growth 算法从中挖掘出物流服务评论与顾客购买后行为间的强关联规则,得到以下结果:

a)顾客最关注的物流服务要素有:包装完整、快递员服务态度好、发货速度快、售后服务态度、运送速度快、退换货速度和送货速度快。顾客关注的物流服务要素贯穿于服务的全过程,除了对发货速度有要求之外,还包括前期的包装、中途的运送、终端的配送、收货后的退换货等,因而物流服务企业提高服务质量需要全过程改善,关注整体优化。

b)与顾客重购行为具有强关联规则的物流服务要素有:发货速度、运送速度、送货速度和快递员服务态度。这说明,提高发货速度、运送速度、送货速度能够促使顾客重购,而快递员服务态度亦显著影响顾客重购。因此,对于物流企业而言,除了物流服务的中间环节,最后一公里、最后 100 米的物流服务至关重要。

c)与顾客推荐行为具有强关联关系的物流服务要素有:包装完整、发货速度、运送速度和送货速度。这说明,提高发货速度、运送速度、送货速度能够促使顾客重购,而包装的完整性亦显著影响顾客推荐。因此,对于物流企业而言,除了物流服务的中间环节,快递的包装的优劣至关重要。

综上所述,B2C 企业可以从顾客最关注的物流服务要素来提升物流服务质量;从改善与顾客重购行为具有强关联规则的物流服务要素来增强顾客重购意愿,增加顾客重购次数,降低企业成本,增加企业利润;从改进与顾客推荐行为具有强关联关系的物流服务要素来增强顾客推荐意愿,提升企业或产品的市场知名度。

参考文献:

- [1] BRYANT R, KATZ R H, LAZOWSKA E D. Big-data computing: Creating revolutionary breakthroughs in commerce[J/OL]. Computing Research Association, 2008: 1-15 [2016-12-20]. <https://doi.org/10.1145/2487575.2488188>.
- [2] GEFEN D. Customer loyalty in e-Commerce[J].

- Journal of the Association for Information Systems, 2002, 3:27-51.
- [3] STANK T P, GOLDSBY T J, VICKERY SK, et al. Logistics service performance: Estimating its influence on market share[J]. Journal of Business Logistics, 2003, 24(1):27-55.
- [4] HOFACKER C F, GOLDSMITH R E, BRIDGES E, et al. E-Services: A synthesis and research agenda[J]. International Journal of Value Chain Management, 2007, 1(1):13-44.
- [5] 张岩岩. B2C 电子商务物流服务质量测度模型及其应用[D]. 长春:吉林大学, 2011.
- [6] 宗蕊, 葛泽慧. 消费者对 B2C 网购物流服务因素的感知分析: 基于京东商城在线客户评论的实证研究[J]. 消费经济, 2014(2):53-58.
- [7] 王家琦, 张耀荔, 陈静. B2C 网络购物模式下物流服务质量与顾客满意的关系研究[J]. 中国商贸, 2013(1):112-114.
- [8] 杨永清, 于本海. 基于物流服务的消费者在线购买后行为实证分析[J]. 管理学报, 2014, 11(3):434-438.
- [9] SEIDERS K, VOSS G B, GREWAL D, et al. Do satisfied customers buy more? Examining moderating influences in a retailing context [J]. Journal of Marketing, 2005, 69(4):26-43.
- [10] 李泓蓉. 网络团购模式下顾客重复购买意向影响因素研究[D]. 北京:北京工业大学, 2013.
- [11] LILJANDER V, ROOS I. Customer-relationship levels-from spurious to true relationships[J]. Journal of Services Marketing, 2002, 16(7):593-614.
- [12] TURNEY P D. Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL[C]// Machine Learning: ECML 2001. Berlin Heidelberg: Springer, 2001:491-502.
- [13] DAVE K, LAWRENCE S, PENNOCK D M. Mining the peanut gallery: opinion extraction and semantic classification of product reviews [C]//International World Wide Web Conference. ACM, 2003:519-528.
- [14] HU M, LIU B. Mining opinion features in customer reviews [C]//Proceedings of the 19th national conference on Artificial intelligence. AAAI Press, 2004:755-760.
- [15] TSOU B K Y. Polarity Classification of celebrity coverage in the Chinese Press[C/OL]//Proceedings of International Conference on Intelligence Analysis, Virginia, USA, 2005. [2016-12-20]. <https://www.mendeley.com/research-papers/polarity-classification-celebrity-coverage-chinese-press/>.
- [16] 王根, 赵军. 基于多重冗余标记 CRFs 的句子情感分析

- 研究[J]. 中文信息学报, 2007, 21(5): 51-55.
- [17] 张成功, 刘培玉, 朱振方, 等. 一种基于极性词典的情感分析方法[J]. 山东大学学报(理学版), 2012, 47(3): 50-53.
- [18] AGRAWAL R, IMIELINSKI T, SWAMI A. Mining association rules between sets of items in large databases[J]. *Acm Sigmod Record*, 1993, 22(2): 207-216. Mining association rules between sets of items in large databases [C]// *ACM SIGMOD International Conference on Management of Data*. ACM, 1993: 207-216.
- [19] 韩家炜, 范明, 孟小峰. 数据挖掘概念与技术[M]. 北京: 机械工业出版社, 2007: 158-159.

Association Analysis Between Logistics Service Review and Customers' Post-purchase Behavior

WANG Shixiong, LAN Yizhen, LI Xue

(School of Economics and Management, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: In online shopping platform, customers' logistics service review usually influences purchase decision behavior of new customers. Online reviews data about women's dress were obtained from Jingdong eMall with web crawler. The factors about logistics service and customers' post-purchase behaviors were extracted from the review through Chinese words segmentation and emotion analysis. Then, logistics service factors that customers pay most attention to and the most widespread customers' post-purchase behaviors were gained. Meanwhile, FP-growth algorithm was used to mine the strong association rules between the logistics service factors and the post-purchase behavior. The result shows that different post-purchase behaviors bring different emphasis on logistics service perception. This research will help B2C merchants to find the key point to optimize the logistics service and provide reference for improving the quality of logistics services.

Key words: logistics service; content analysis; post-purchase behavior; association mining

(责任编辑: 钱一鹤)