

## 线上OR线下，谁与争锋？

Yang Yue (杨玥), Fei Qi (费淇), Ma Siyu (马思宇), Wang Kun (王堃)

沈阳工学院 信息与控制学院

**摘要:** 近些年随着全球经济一体化与信息技术的发展，电子商务正飞速兴起，开启了购物新模式。一种全新的、高效的购物方式正在全球范围内迅速普及。如今，在互联网下的各种直播带货正流行起来，许多人足不出户就能购买各种东西。那么与传统的购物方式相比，人们更倾向传统购物还是网上购物呢？针对以上背景，本文提出了用数据挖掘技术分析人们更倾向线上购物还是线下购物。

**关键词:** 大数据；数据挖掘

**Abstract:** in recent years, with the development of global economic integration and information technology, e-commerce is rising rapidly, opening a new mode of shopping. A new and efficient way of shopping is rapidly spreading all over the world. Nowadays, all kinds of live broadcasting with goods on the Internet are becoming popular, and many people can buy all kinds of things at home. So compared with the traditional way of shopping, do people prefer traditional shopping or online shopping? In view of the above background, this paper proposes to use data mining technology to analyze whether people prefer online shopping or offline shopping.

**Key words:** Big data; Data mining.

### 1 研究背景与意义

近年来，实体店受到网络零售巨大冲击，经营陷入困难，电商企业流量红利消退，发展遇到瓶颈，加之消费诉求发生深刻变化，实体零售与网络电商正逐步从独立、对抗走向融合、协作，深度融合是优势互补、实现共赢的发展方向。据支付宝数据显示，2016年全国有近1000家大型购物中心、5万家超市便利店、55万家餐厅参加了双十二活动。在未来零售市场中，以互联网、物联网、人工智能及大数据等领先技术为驱动，数字化技术将虚拟与现实深度融合，传统零售在物理空间和时间维度上将获得极大延展，消费者不再受区域、时段和店面等因素限制，零售行业终将发展成面向线上线下全客群，提供全渠道、全品类、全时段、全体验的新型零售模式。

伴随着数字技术成长起来的15-34岁的年轻人是线上消费的主力军。这些消费者热衷并精通新技术，据调查，青年群体对线上消费的需求不仅推动了线上销售的增长，还促使这一产业日趋成熟。这些消费者对数字技术的使用有着空前的热情，线上购物对他们而言不仅是一个习惯，而且逐渐成为了一种新的生活方式。

虽然网上购物已经成为了年轻人的一种习惯，但是人们面对网购的各种问题仍非常担心。那么不同年龄段的人们面对网上购物都持有怎么样的态度？是否信任网络店家？线上与线下购物的购物差距是多少，哪一种购物方式更受人们喜欢都值得我们去发掘。

### 2 相关技术与方法

#### 1、数据获取方式-网络爬虫

利用网络爬虫，从网站某一个页面(通常是首页)开始，读取网页的内容，找到在网页中的其它链接地址，然后通过这些链接地址寻找下一个网页，这样一直循环下去，直到把这个网站所有的网页都抓取完为止。如果把整个互联网当成一个网站，那么网络蜘蛛就可以用这个原理把互联网上所有的网页都抓取下来。

**PR 值计算方法:** 一般说来，PR 值越高，反映在在搜索结果中的排名越靠前 (重要性越高)，当前有很多人通过很多方式来提高自己网站的 PR 值，除网站内部的优化外，大多数人采用的是外部链接。如下：

$$PR(A) = (1-d) + d(PR(t1)/C(t1) + \dots + PR(tn)/C(tn))$$

其中 PR(A)表示的是从一个外部链接站点 t1 上，依据 Pagerank 系统给你的网站所增加的PR 分值;PR(t1)表示该外部链接网站本身的PR 分值;C(t1)则表示该外部链接站点

所拥有的外部链接数量。

## 2、科学计算和数据分析

数据学和数据科学是关于数据的科学，定义为研究探索 Cyberspace 中数据界奥秘的理论、方法和技术。其中包括的内涵：一个是研究数据本身；另一个是为自然科学和社会科学研究提供一种新方法，称为科学研究的数据方法。大量应用到多维矩阵和矢量计算。关系图如图 2.1所示。



图 2.1 关系图

## 3、数据挖掘技术

数据挖掘技术是一种数据处理的技术，是从大量的、不完全的、有噪声的、模糊的、随机的数据集中识别有效的、新颖的、潜在有用的，以及最终可理解的模式的非平凡过程。它是一门涉及面很广的交叉学科，包括机器学习、数理统计、神经网络、数据库、模式识别、粗糙集、模糊数学等相关技术。

### 3 数据预处理与分析

数据清理例程通过填写缺失的值、光滑噪声数据、识别或删除离群点并解决不一致性来“清理”数据。主要是达到如下目标：格式标准化，异常数据清除，错误纠正，重复数据的清除。

数据挖掘时往往数据量非常大，在少量数据上进行挖掘分析需要很长的时间，数据归约技术可以用来得到数据集的归约表示，它小得多，但仍然接近于保持原数据的完整性，并结果与归约前结果相同或几乎相同。通过了解数据集中的每日数据字段解释，将重复的属性进行归约删除，如订单单笔价，商品交易数量等。

### 4 预测模型的构建

分别采用决策树和贝叶斯分类方法进行货物的预测。预测模型流程图如图4.1所示。

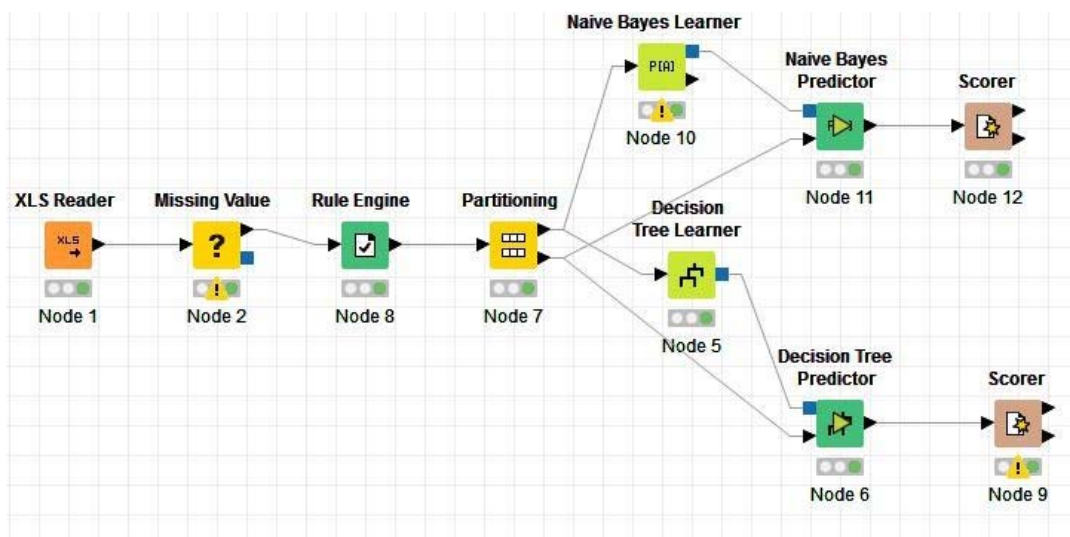


图4.1 预测模型流程图

低价格是实现高GMV的主要原因。带货的商品售价一般低于线上及线下的商品售价，商品数量有限，并且仅在某一时间段内享受折扣，头部主播对上游供应链的议价权是核心竞争力，一般来说，粉丝量越高的主播能拿到的商品价格越低。

通过数据挖掘，我们将年初以来头部主播直播带货价格区间挖掘得到了以每日 TOP20 主播直播数据来看，今年3月全月带货商品价格区间超过5000元以上的占比约8.9%；1000-5000元价格区间的商品占比约6.4%；5001000元价格区间的商品占比约8.9%；100-500元价格区间的商品占比约44.6%；100元以下价格区间的商品占比约31.1%。

通过大数据比对和调查结果反馈，我们发现网上购物似乎已经成为了人们更多的购物方式。人们对于不同的类型的商品，在两种购买途径上也有所不同。对于相对来说比较轻巧的服装、不急用的零食来说，人们更倾向于在网络上购买；而对于比较沉重电器、着急用的药品来说，人们更倾向于在线下购买。总体来说，线上的交易数量要高于线下的交易数量，电商的发展前景较好，平台可以为其提供更多的资源。人们对于各方面的满意程度线上高于线下，线上购物人们更省心。

### 5 结论与展望

如今，各大互联网和电商纷纷准备发力在直播电商的领域，已经逐渐成为人们购物的一种趋势。网上购物正在一步步的扩大，受到更多人的欢迎。线上购物商品类别更加的繁多，而且随着世界的发展，人们更没有足够的大片时间亲自去商场购买商品，可以利用碎片化的时间通过手机来选择自己心仪的商品。因此线上购物的推广十分收到大家的欢迎，前景一片美好。

虽然线上购物发展趋势较好，但线上的商品信息繁杂多样，真假难辨，需要花费的精力会更多。而且对于年龄大的老人来说过程有些繁琐，较难在全年龄段推广，所以也很难取代传统的线下购物模式。

### 参考文献：

1. 曾宪宇. 电商用户消费行为预测与心理建模方法研究[D].中国科学技术大学, 2017. 4
2. 谭晶.O2O 电商用户行为预测应用研究[J].信息通信, 2019(01): 276-277.