

Credit Evaluation of Supplier Based on Big Data for E-Commerce Platform

Ming Yan

Logistics Institute, Beijing Wuzi University, Beijing
Email: 1099885031@qq.com

Received: Aug. 1st, 2018; accepted: Aug. 13th, 2018; published: Aug. 20th, 2018

Abstract

From the perspective of customers, based on big data analysis, this paper obtains the evaluation value index of supplier credit under the e-commerce platform. Among them, python is mainly used to implement the crawler of the commodity comment of the corresponding supplier on the e-commerce platform, and the after-sales evaluation text is obtained. Finally, data processing and analysis are carried out on the text to construct the supplier credit evaluation index under the e-commerce platform. The evaluation index constructed by this method is more suitable for the credit evaluation of suppliers under the e-commerce platform, so as to effectively manage the credit of suppliers under the e-commerce platform.

Keywords

Big Data, The Crawler, Python, E-Commerce Platform, Supplier Credit

基于大数据的电商平台供应商信用评价

闫 明

北京物资学院物流学院, 北京
Email: 1099885031@qq.com

收稿日期: 2018年8月1日; 录用日期: 2018年8月13日; 发布日期: 2018年8月20日

摘 要

本文从客户的角度, 基于大数据分析, 得到电商平台下的供应商信用的评价价值指标。其中主要运用python实现对电商平台对应供应商商品评论的爬虫, 得到售后评价文本。最后在对文本进行数据的处理与分析, 以此作为依据来构建电商平台下供应商信用评价指标。以此构建的评价指标更适用于电商平台下供应商

的信用评价，以便于对电商平台下的供应商信用进行有效的管理。

关键词

大数据，爬虫，Python，电商平台，供应商信用

Copyright © 2018 by author and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

2013 年,中商协会与北师大电商研究中心就中国电子商务的诚信情况作了首次调查报告,报告显示:71.1%的调查者对一些网站的合法性与真实性等问题持有怀疑态度,56.4%的调查者曾遇到过网上商品信息失真的情况,还有40.9%的调查者遇到过承诺的在线服务没有兑现的情况。2016 年据中国电子商务投诉与维权公共服务平台监测数据显示,虚假促销、商品质量、疑似售假、货不对板等供应商信用问题占了零售电商投诉问题的24.47%。由于电子商务具有交易的电子化、信息不对称、时间差、匿名等特点,使得信用问题频发。

目前,一些平台信用度的变化仅仅是依据交易的次数,而与每次交易的金额无关,因此使得一些卖家可以通过出售低价值商品在短期内提高自己的信用值[1]。以京东为代表的电商平台,建立了由质控、运营、售后、技术等多部门联合监管机制并自主研发了“反作弊识别系统”,系统利用京东大数据平台,从订单、商品、用户、物流、支付、评论、浏览等多个维度进行分析,识别各个交易环节的异常数据,结合多种智能算法进行数据建模,从而对虚假交易进行精准定位。

2. 相关理论介绍

2.1. Python 语言介绍

Python 具有源代码的高度可阅读性,通过简单的标点符号与英文字母的计算机设计语言实现面向对象的解释。python 采用的是动态类型系统,所以在编译的时候,python 需要在运行过程中才能对对象进行检查[2]。同时,python 内置数据类型丰富,能实现数据代码长度的有效减少。

2.2. Python 的特点

1) 时效性高

由于 python 的底层是基于 C 语言所构建的,且大多数标准库和第三方库也都是基于 C 语言写的,所以,python 在运行过程中效率非常的高。

2) 代码开源

Python 作为开放源码软件之一。使用者可以自由地发布这个软件的拷贝、阅读它的源代码、对它做改动、把它的一部分用于新的自由软件中。

3) 可移植性

由于它的开源本质,python 已经被移植在许多平台上(经过改动使它能够工作在不同平台上)。

4) 面向对象

Python 支持面向对象与过程的编程。在“面向过程”的语言中，程序是由过程或仅仅是可重用代码的函数构建起来的。在“面向对象”的语言中，程序是由数据和功能组合而成的对象构建起来的。

3. 爬虫设计与大数据文本分析

由于本文采用的是从客户的角度对供应商的信用进行分析评价，所以，需要去对应网页进行客户评论的提取以此作为文本数据分析依据。基于 python 的诸多优点，利用 python 设计爬虫，根据爬虫的功能，需要确定网页初始 URL，由于客户评论不止一页，根据 URL 特点得到 URL 的列表，通过分析每个 URL 地址中的链接获取对应内容，最后以 csv 的文件形式保存。

3.1. 爬虫设计

由于 Python 具有强大的数据库支持，根据网页爬虫的抓取方式，需要的数据包有 requests、re、lxml、json、csv [3]。带入代码如下：

```
import requests
import re
from lxml import etree
import json
import csv
import time
```

1.1 输入网页进行对应评论文本进行输出：product_url = ‘爬虫网页’

1.2 获取 commentVersion，用于构造评论页的 url

1.3 解析网页内容，获取下一页的链接

1.4 根据参数生成商品评论 url

1.5 爬取单页的评论信息

1.6 输出 csv

3.2. 大数据文本分析

通过网页爬虫会得到大量的文本数据，需要对这些海量数据进行有效的分析。同样是利用 python 的强大数据解释功能，做到的是对大量文本的关键评论的词的提取[4]，以此作为供应商信用评价指标的构建的依据。

1) 数据预处理

对于爬虫出来的售后评论，首先需要进行文本数据的预处理。预处理主要指的是对文本进行切分、分词以及去停用词的处理。由于本文是对每条售后文本数据进行分析处理，所以文本分割可以省略。主要是对文本进行停用词(stop words)的提出，对于停用词是有专门的词库组成的，可以利用“jieba”通过 python 来实现处理。

2) 关键词的提取

关键词指的是对所需要的能表现出文本核心信息的词汇，以一定的特征项来代表文档并能表达相应的意义。在文本的挖掘时只需要对这些关键词进行有效的分析，从而实现对非结构化文本的处理。

3) 匹配分类

将文本转化成向量的形式，并计算出对应的权重，便可以进行分类挖掘，既匹配分类。

4. 指标构建分析

4.1. 正则分析

由于本文的目的性较强，所以在文本分析时根据实际需要，对大量文本进行有效的分词，并通过数据文本的分析过程，输出结果如表 1。

4.2. 电商平台供应商信用评价指标

1) 及时满意度

在线会话过程中或结束时，客户对于服务评价：(非常满意 + 满意)/(非常满意 + 满意 + 一般 + 不满意 + 非常不满意)目标值为 90%。

2) 咚咚应答率

咚咚应答率指的是在电商平台中供应商对消费者的问题应答情况。对于应答率规定，咨询量 ≥ 1000，咚咚应答率 ≥ 95%，店铺周平均在线客服应答率应不低于 90%。

3) 产品丰富度

同一类商品下有多个单品(sku)，并且及时更新，以保证不会出现断货、缺货等情况，影响消费者体验。

4) 虚假宣传

指在平台页面展示含有虚假信息内容，欺骗和误导消费者的行为，具体情形以广告法规定为准；包括但不限于通过文字或图片等方式明示或暗示与商品实际内容不相符的虚假信息，使消费者对商品的质量、制作成分、性能、用途、生产者、有效期限、产地等产生误解的行为。

5) 售后服务

售后服务指的是在终端客户收到产品之日起，供应商开始售后服务工作。供应商保证售后服务符合承诺标准以及国家相关规定，例如因商品质量问题需要退、换、维修时供应商按生产规定办理延迟退、换、维修时间。

Table 1. Output result
表 1. 输出结果

问题	关键词	权重	对应评论数
及时满意度、咚咚应答率	售后 差评 回电 客服 来电 不回话 耽误 沟通 从来 不回 回复 咨询 挂机 不理人 回复慢 咚咚 效率低	0.288	26,983
产品丰富度	缺货 种类 采购中 配套 抢购 很少 不全	0.0043	407
虚假宣传	假货 正品 信任 怀疑 宣传 图片 描述 相差 实物 谨慎 地摊货 水货 不符	0.1396	13,141
售后服务	售后服务 赔偿 投诉 纠纷 交易 折价 运费 二手 维修 态度 服务 承诺 电话 打不通 马虎	0.0761	7168
价格变动	价格 便宜 骗人 降价 不值 原价 买完 降价 性价比 比底 虚高 不合理 满减	0.0542	5101
刷单炒新	刷单 疑似 慎重 刷单 虚假 信用值 同类 空包 炒作 不诚信	0.0031	289
48 小时揽件及时率	没到 发货 时效 未发货 速度慢 拖延 收不到	0.0168	1579
商品退换货	换货 残不退 退换货 返修 退货 退款 次品 补发	0.0216	2056

6) 价格变动

价格变动主要指的是，电商平台上的商品会不定期的推出优惠活动，或者是一些品牌日的时候，会通过满减、领取购物券等方式对商品进行促销。但此时会涉及到商家是否将商品大幅度提价然后在进行促销活动，所以需要对商品的价格变动幅度进行管理。

7) 刷单炒信

是指供应商通过不正当方式获取虚假的商品销量、店铺评分、商品评论等，妨害、干扰消费者购物权益的行为。这种不正当方式主要包括但不限于以下几种形式：自买自评、第三方炒作、空盒等。

8) 商品退换率

实际发生退换货返修的服务单量/已完成商品件数(服务单生成时间是 30 天内的数据)。

9) 48 小时揽件及时率

48 小时揽件及时率指的是，交易达成后 48 小时内买家可以查询到订单的快递公司揽件跟踪信息情况。

5. 总结

由于电商平台的特殊性，使得对其供应商的管理也不同于传统的供应商。本文从客户的角度通过大量售后文本数据的分析，来构建电商平台供应商信用评价指标，使指标的输出更具有代表性。同时，利用大数据文本语义分析也可以增强指标的辨识性，提高精准性。能做到对供应商的信用进行重点的考察，从而弥补因电商特性所带来的供应商信用难以管理的问题。

但是数据的处理过程会占用计算资源工作，作为一种额外的资源消耗业务，需要尽量加强不同系统间的数据规范化和标准化，保证系统数据的规范以及编码的一致性。而本文在数据的处理中由于实际的需要与现实情况，需要首先保证对文本数据的预处理实现规范化，最后在系统的调试运行中保证编码的一致性即可[5]。总之对于电商平台下供应商信用的管理问题还是要以大数据作为支撑的。

参考文献

- [1] 荣飞琼, 郭梦飞. 基于大数据的跨境电商平台供应商信用评估研究[J]. 统计与信息论坛, 2018, 33(3): 100-107.
- [2] 白雪丽. 浅析基于 Python 爬虫技术的特性及应用[J]. 山西科技, 2018, 33(2): 53-55.
- [3] Kevin. 网络爬虫技术原理[J]. 计算机与网络, 2018, 44(10): 38-40.
- [4] 黄文杰, 姚庚梅. 大数据时代下爬虫技术应用与研究——以标讯快车项目为例[J]. 科技创新与应用, 2018(6): 37-38+41.
- [5] 郭丽蓉. 大数据环境下的网络爬虫设计[J]. 山西电子技术, 2018(2): 50-52+94.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2169-2556, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: ass@hanspub.org