

生鲜产品的多变量 SVR 需求预测 ——基于在线评论的顾客感知因素提取

张炎亮 代沛沛

(郑州大学 管理工程学院, 郑州 450001)

摘要 为探索在线评论数据中蕴含的顾客感知产品信息对生鲜产品需求量预测准确度的影响,针对生鲜产品电商平台中大量评论数据,利用网络爬虫技术和 Word2vec 模型建立产品特征词库,提取主要需求预测影响因素,并基于产品特征词库对评论文本分类将影响因素量化,构建多变量 SVR 需求预测模型,同时运用粒子群算法对 SVR 模型中的主要参数进行优化,在此基础上进行实证分析。结果表明:1) Word2vec 模型能挖掘在线评论数据中顾客关注的产品特征,有效提取顾客感知的需求预测影响因素;2) 与单变量 SVR 模型相比,加入评论中顾客感知因素的多变量 SVR 在预测产品需求量时误差更小。利用在线评论中顾客感知因素建立多变量 SVR 需求预测模型能有效提高生鲜产品需求量预测准确度。

关键词 生鲜产品; 在线评论; 顾客感知; 多变量 SVR; 需求预测

中图分类号 F252

文章编号 1007-4333(2022)07-0275-08

文献标志码 A

Multi-variable SVR demand forecast model for fresh products: Factor extraction of customer perception based on online reviews

ZHANG Yanliang, DAI Peipei

(School of Management Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract In order to explore the influence of customer perceived product information contained in online review data on the accuracy of fresh product demand prediction, based on the large amount of review data of a fresh product e-commerce platform, web crawler technology and Word2vec model are used to build a product feature database, extract the main factors influencing demand forecast, and quantify the influencing factors based on the product feature database to classify the review text. A multi-variable SVR demand forecast mode is then constructed and the particle swarm algorithm is adopted to optimize the main parameters of the SVR model. On this basis, an empirical analysis and verification is carried out. The results show that: 1) The Word2vec model can mine the product features that consumers pay attention to in online review data, and the influencing factors on consumer perception for demand forecasting can be extracted effectively. 2) Compared with the univariate SVR model, the multivariate SVR with customer perception factors added has smaller error in predicting product demand. In conclusion, using customer perception factors in online reviews to establish the multivariate SVR demand forecasting model can improve the accuracy of fresh product demand forecasting effectively.

Keywords fresh products; online reviews; customer perception; multivariate SVR; demand forecasting

电子商务背景下,网络零售迅速发展,越来越多的电商巨头进入生鲜领域,拓展了生鲜产品线上销售渠道,推动了生鲜行业的加速发展。2020 年我国

生鲜产品电商行业规模达 4 584.9 亿元,预计到 2023 年,生鲜产品电商行业规模将超万亿^[1]。在广阔的市场需求下,生鲜产品电商行业却仍面临着货

收稿日期: 2021-10-06

基金项目: 国家自然科学基金青年项目(71801195)

第一作者: 张炎亮,教授,主要从事质量管理、物流与供应链管理研究,E-mail: yanliang@zzu.edu.cn

损率高、盈利少等诸多问题^[2],究其原因,部分生鲜产品电商企业通过资金投入、扩大宣传盲目拓展市场,但若缺乏准确的市场需求信息,生鲜产品极易造成积压,导致货损率增大,企业成本增加。基于此,生鲜产品电商企业须对产品需求做出迅速响应,此外相比线下销售,线上销售能为顾客提供表达购买体验、情感态度的在线评论平台,顾客在平台上可主动分享自身对产品的全方位评价,从而让潜在顾客更多了解产品信息,帮助其执行购买决策,因此在线评论在一定程度上影响了产品需求量。

在线评论是开放式表达渠道,评论内容是顾客对产品 & 服务的感知性评价,此评价无意中会形成一种口碑效应,影响后续产品销售^[3-4]。目前国内对在线评论数据与产品销量之间的关系进行了较多探讨:纪雪等^[5]量化从评论文本数据中提取出的产品属性,计算产品属性的用户满意度来确定下一代产品的开发需求。张梦莹等^[6]基于有用性排序的方法,探究出评论总数、评论时效性、情感倾向对产品销量都存在着不同程度的显著性影响。沈超等^[7]提取出产品的关键属性和非关键属性,利用决策树模型分析出了客户偏好趋势。王英等^[8]考虑品牌效应因素,基于采纳信息模型,验证了品牌强度、评论评价及评论时效度对产品销量会存在显著相关关系。胡雅淇等^[9]以农产品为研究对象,运用逐步回归的方法分析出在线评论数量、可视化评论、差评数量及评论长度均会正向或负向的影响产品销量。

国外对在线评论数据与产品销量关系的研究较多:Lau等^[10]基于情感分析方法,挖掘出评论数据中消费者情绪,并以此进行销售预测,提高了销售预测的准确度。Chen等^[11]基于实验研究认为正面或负面的补充评论的顺序会影响消费者的购买意愿,而产品的涉入程度会调整它们之间的关系。Hu^[12]以京东商城在线评论数据为例,以问卷调查和访谈的方式,探讨出经济相关评论和服务相关评论对网购行为有显著影响。Gopinath等^[13]运用动态分层线性模型,发现口碑数量能够影响口碑极性进而影响品牌绩效。Ruiz-Mafe等^[14]发现评论顺序对销量有着不同影响,当在线评论以正面评论开始时对企业销量的提升更有益处。

目前针对在线评论数据与销量关系的研究仍在继续发展,现有文献主要针对评论数据中的信息与产品销量之间的关系进行探讨,但从时间维度上看,评论信息反映了随着时间的变化顾客对产品需求的

变化,如何利用挖掘出的评论信息把握产品未来需求变化、预测产品未来需求量,鲜少有研究对其进行更加深入的讨论,所以利用在线评论中顾客感知信息预测产品需求量的研究仍然较为缺乏。在此过程中,评论数据中顾客感知因素的提取是对产品需求量进行准确预测的前提。由于在线评论中数据量巨大且评论内容多是由不同词语组成,传统方法很难准确提取出顾客感知因素,因此,出现了许多用于处理大量非结构化文本数据的方法,如隐狄利克雷分布(LDA)、Word2vec模型、K-means聚类等^[15-18]。在众多评论文本提取方法中,相对于其他数据处理方法,Word2vec模型能够基于深度学习中的循环神经网络,结合词语上下文关系,理解文本中的语义和语法信息,在大量语料库中进行无监督学习,通过计算词向量确定两个词语之间关系远近,因此更加贴合评论文本特征,能够更全面获取产品的特征词语。此外在需求预测方面,由于提取出的产品特征数据具有多元化、非线性的特点,而SVR作为SVM的一种,在处理非线性数据上具有很强的处理能力^[19-20],其能根据预测对象与其他影响因素的关系解决冗余属性问题,并且在小样本需求预测上具有较高的准确度^[21],在处理复杂的产品特征数据上具有较大优势,因此本研究选取多变量SVR预测生鲜产品需求量。

本研究旨在以生鲜产品为研究对象,基于Word2vec模型挖掘评论文本中顾客感知的产品特征因素,并对特征因素量化处理,建立包含多个特征因素的多变量SVR需求预测模型,以期对产品需求量进行准确预测,从而使企业准确了解顾客需求,及时调整产品决策。

1 研究方法与模型

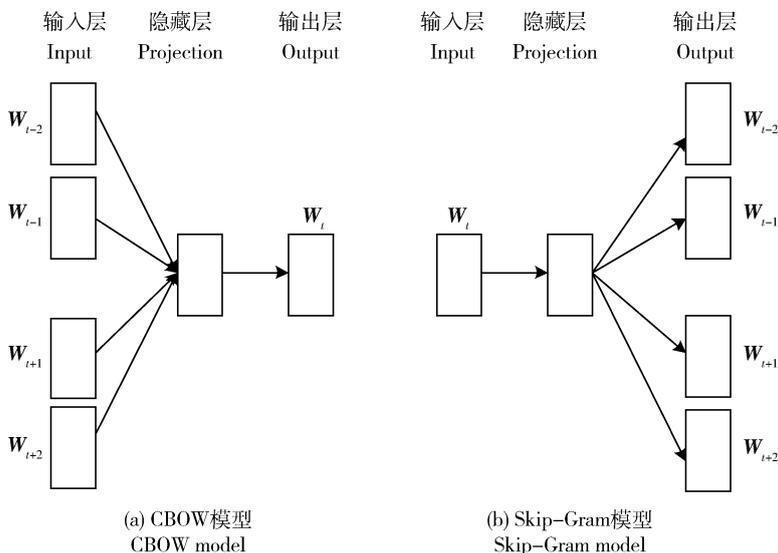
为利用在线评论中顾客感知因素对生鲜产品需求量进行准确预测,本研究构建了2个模型,分别是Word2vec模型和多变量SVR需求预测模型。

1.1 Word2vec核心思想

Word2vec是用于高效训练词向量的工具,其基本思想是一篇文章中上下文相似的2个词语,它们的词向量也应该相似^[22-23]。该模型主要包含2种训练方式:CBOW(Continuous Bag-of-Word)模型以及Skip-Gram模型(图1),两者均是通过单层神经网络为每个词语分配一个恒定的向量,该向量并不代表词语特征,而只是显示一个数字,构建句子或文

本中每个词对应的数值向量后,将单个词的向量进行组合,可显示整个句子的数值向量。其中 CBOW 模型是根据句子中目标词语的相邻词($W_{t-1}, W_{t-2}, W_{t+1}, W_{t+2}$)信息对目标词语 W_t 进行预测,训练词向量速度较快;Skip-Gram 模型是根据目标词语 W_t

所在的语境推测出上下文中与它相似的词语,训练速度相对较慢,但其训练效果比 CBOW 模型训练效果好^[24]。鉴于本研究需要提取产品主题,对结果准确要求较高,因此采用 Skip-Gram 模型,将所爬取的网站评论数据作为该模型的训练集。



W_t 为目标词语; $W_{t-1}, W_{t-2}, W_{t+1}, W_{t+2}$ 分别为目标词语的相邻词。

W_t is the target word; $W_{t-1}, W_{t-2}, W_{t+1}$ and W_{t+2} are the adjacent words of the target word, respectively.

图 1 CBOW 模型和 Skip-Gram 模型

Fig. 1 CBOW model and Skip-Gram model

1.2 影响因素特征集构建

从在线评论数据中提取出顾客感知的产品特征因素,可使企业了解消费者所关注的产品特性,是对需求进行预测的前提。在线评论数据中顾客对产品特征的描述多是以词语形式呈现,因此本研究通过 Word2vec 模型词向量聚类的方法从大量评论数据中提取用户所关注的多个产品特征集,作为需求预测模型的输入变量。提取步骤如下:

1) 中心特征词提取。提取过程中,一方面通过查阅相关文献,分析影响顾客感知、顾客满意度的相关要素,进而初步预判影响需求的主要特征;另一方面对在线评论的文本数据进行词频以及词性统计,结合所分析的顾客感知要素选取高频词语作为影响因素特征集的中心特征词。

2) Word2vec 模型训练。运用 python 中的“jieba”分词工具对评论文本进行分词和停用词处理,并设置模型中的上下文窗口参数和词向量空间维度对模型进行训练,在设置不同参数对模型进行

多次训练后,本研究将上下文窗口参数和词向量空间维度分别设为 5 和 100。

3) 特征集构建。依据中心特征词,通过 Word2vec 模型对词语进行多次聚类,选取相似度较高的词语作为特征词库,形成需求预测影响因素的特征集,将其作为预测变量输入到 SVR 模型中。

1.3 SVR 预测模型构建

1.3.1 数据预处理

Word2vec 模型训练出来的结果是由一个个词语所组成的特征词库,为更好的将其输入到预测模型中,需将文本因素转化为数值因素。因每条评论中都包含着顾客所感知的产品特征,因此本研究运用经 Word2vec 模型聚类出来的各个特征词来表达顾客观点,实现文本因素的数值化转换。假设在线评论数据经过聚类之后形成的影响因素共 $X_i (i = 1, 2, \dots, n)$ 类,当某条评论文本中包含与 X_i 因素相似的词语时可将该条评论文本归为 X_i 类,从而得到各个因素所包含的评论文本数量,实现文本因素的

量化,在 $t(t=1,2,\dots,T)$ 时刻第 i 类因素可表示为 $X_i(t)$,为消除量纲对预测结果的影响,需要对数据进行归一化处理,利用 Z-score 标准化的方法将数据限制在一定范围之内,计算公式如下:

$$X_i(t)^* = \frac{X_i(t) - \overline{X_i(t)}}{\sigma} \quad (1)$$

式中: $\overline{X_i(t)}$ 为原始数据的均值; σ 为原始数据的方差。

1.3.2 支持向量回归模型原理

SVR 作为 SVM 的一种,引入了不敏感损失函数 ϵ ,在解决非线性回归问题方面具有很强的处理能力,它的优化目标函数可表示为:

$$\begin{cases} \min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^T l_\epsilon(f(x_i), y_i) \\ l_\epsilon(z) = \begin{cases} 0 & |z| \leq \epsilon \\ |z| - \epsilon & \text{其他} \end{cases} \\ \text{s. t. } |y_i - (\omega x_i + b)| \leq \epsilon + \zeta \quad \epsilon > 0 \end{cases} \quad (2)$$

式中: C 表示惩罚因子, C 值越大对误差分类的惩罚越大,越小对误差分类的惩罚越小; x_i 为模型需要输入的变量值; y_i 为相应的预测值; ω 为函数 $f(x_i)$ 中自变量 x_i 的系数; l_ϵ 表示损失函数,即允许超平面外存在样本点,但需要使损失函数尽可能小; $(f(x_i), y_i)$ 表示超平面中的样本点; ζ 为松弛变量,当样本点位于超平面以内或其边缘上时 $\zeta=0$,当样本点位于超平面上方时 $\zeta>0$,当样本点位于超平面下方时 $\zeta<0$ 。为求解式(2),一般采用拉格朗日对偶变换,公式为:

$$\begin{cases} \min_{\alpha_i^*, \alpha_i} \frac{1}{2} \sum_{t_1, t_2=1}^T (\alpha_{t_1}^* - \alpha_{t_1})(\alpha_{t_2}^* - \alpha_{t_2})(x_{t_1} x_{t_2}) + \\ \epsilon \sum_{i=1}^T (\alpha_i^* - \alpha_i) - \sum_{i=1}^T y_i (\alpha_i^* - \alpha_i) \\ \text{s. t. } \sum_{i=1}^T (\alpha_i^* - \alpha_i) = 0, 0 \leq \alpha_i^* \leq C, \\ t = 1, 2, \dots, T \end{cases} \quad (3)$$

式中: α_i 和 α_i^* 为拉格朗日系数。对于单变量 SVR 预测模型,通常仅将时间序列的 p 个数据作为预测模型的输入变量,对未来某个时刻的值进行预测,此情况下, y_t 与其前几个时刻的值之间存在着如下函数关系:

$$y(t+T+1) = F(y(t), y(t+1), \dots, y(t+m)) \quad (4)$$

在本研究中除了将时间序列历史数据作为预测

模型的输入变量外,还提取了其他相关影响因素,因此需将单变量 SVR 转换为多变量 SVR,即上述函数关系转换为需求量 y_t 与多个输入变量之间的关系,函数关系表示为:

$$y_t = F(x_1(t), x_2(t), \dots, x_n(t)) \quad (5)$$

引入核函数将非线性回归问题转换为一个近似线性回归问题,仍然采用拉格朗日对偶变换法进行求解,具体公式表示为:

$$\begin{cases} \min_{\alpha_i^*, \alpha_i} \frac{1}{2} \sum_{t_1, t_2=1}^T (\alpha_{t_1}^* - \alpha_{t_1})(\alpha_{t_2}^* - \alpha_{t_2})(K(x_{t_1})K(x_{t_2})) + \\ \epsilon \sum_{i=1}^T (\alpha_i^* + \alpha_i) - \sum_{i=1}^T y_i (\alpha_i^* - \alpha_i) \\ \text{s. t. } \sum_{i=1}^T (\alpha_i^* - \alpha_i) = 0, 0 \leq \alpha_i^* \leq C, \\ t = 1, 2, \dots, T \end{cases} \quad (6)$$

在核函数选择过程中,可供选择的核函数主要有径向基函数(Radial basis kernel function, RBF)、高斯核函数、多项式核函数、Sigmoid 核函数等,其中 RBF 函数能够将特征空间映射到无穷维,与其他核函数相比参数较少,方便计算,因此本研究引入 RBF 函数作为 SVR 模型的核函数。最后将所提取的影响因素特征集输入到此回归函数中可对产品的需求量进行预测。

1.3.3 PSO 算法优化 SVR 模型参数

在 SVR 模型中惩罚因子 C 以及不敏感损失函数 ϵ 是影响模型预测准确度的重要参数,其中 C 值过大曲线容易过拟合,反之容易欠拟合; ϵ 反映了数据映射到新的特征空间后的分布, ϵ 值越大支持向量个数越少,反之支持向量个数越多。为较快速确定参数 C 和 ϵ 的值,采用粒子群算法对参数进行优化,操作步骤如下:

1) 种群初始化。设置迭代次数、种群大小、学习因子、惯性权重等基本参数。

2) 计算该算法的适应度函数值,并将 R^2 作为算法的适应度函数。

3) 根据适应度函数计算粒子个体及全局最优解,并与历史值进行比较更新群体中粒子的速度和位置。

4) 判断是否满足终止条件,如满足则停止迭代,若不满足则不断更新粒子速度和位置直至满足终止条件。

2 研究数据与结果分析

2.1 数据获取及预处理

为验证评论数据中的产品特征因素对生鲜产品需求量预测准确度的影响,本研究采用京东生鲜网站中产品评论数据进行实证分析。利用谷歌插件 web scraper 分别获取了苹果、火龙果和小龙虾 3 种生鲜产品 2021-06-25—2021-08-17 共 54 天的评论数据,得到的 3 种生鲜产品评论数据量分别为 10 312 条、9 685 条和 9 064 条。根据网站的评论规则,每个顾客在确认收货后均会形成一条文本评论数据,基于此本研究将每天的评论数量作为生鲜产品的日需求量,从而建立以天为周期的历史需求量数据集。从网站爬取出来的评论数据是顾客对产品使用体验的随意性表达,其中会存在着图片、符号等非文本信息,因此需要删除评论文本中的表情符号、数字等无用评论信息,确保评论文本的纯文字性;其次对于一些不包含任何主题信息的超短评论,也需将其删除;最后将处理后的评论数据输入到 python

软件中,运用“jieba”分词工具对评论数据进行分词、去停用词处理,形成最终的文本语料库。

2.2 在线评论产品特征因素提取

挖掘评论数据中顾客感知的产品及服务信息是提取生鲜产品需求影响因素的关键步骤。对经过预处理后的评论文本进行词频统计分析,选取词频排名前 100 的词语作为产品特征词库,在该词库中由于会存在着同义词,需要对其进行人工筛选,将同义词合并为一类,从而形成生鲜产品的中心特征词,作为 Word2vec 模型的输入语料。

经过词频统计分析及人工筛选合并同义词,最终确定生鲜产品的 5 类中心特征词,分别为“口感”、“包装”、“物流”、“性价比”和“服务”。将此中心特征词分别输入到 Word2vec 模型中,借助模型中的相似度函数,计算得到与各个中心特征词的相似词语,以火龙果评论数据中的特征词“口感”为例,得到的相似度排名前 30 的词语集合见表 1。按此步骤最终可形成生鲜产品需求预测影响因素特征词集。

表 1 “口感”语义相关特征词集及其与中心特征词的相似度

Table 1 Semantic related feature word set of “Taste” and its similarity to the central feature word

词语 Word	相似度 Similarity	词语 Word	相似度 Similarity	词语 Word	相似度 Similarity	词语 Word	相似度 Similarity
味道 Smell	0.955 7	好吃 Good to eat	0.812 3	挺大 Quite big	0.797 2	个儿 Size	0.778 3
很甜 Very sweet	0.873 8	个个 All	0.812 0	糖分 Sugar	0.795 2	超值 Value	0.777 0
甜美 Sweet	0.856 7	口味 Taste	0.811 9	坏果 Bad fruit	0.793 9	爽口 Tasty	0.777 0
新鲜 Fresh	0.850 8	均匀 Evenly	0.811 8	适中 Moderate	0.791 5	磕碰 Bump	0.776 2
不小 Not small	0.849 4	很大 Very big	0.808 3	偏小 Too small	0.786 5	一顿 A meal	0.775 4
皮薄 Thin skin	0.848 0	个头 Size	0.807 7	蛮大 Quite big	0.785 0	很正 Very positive	0.774 6
品相 Appearance	0.828 0	个子 Size	0.801 8	蛮甜 Quite sweet	0.784 9	中等 Medium	0.771 6
挺甜 Quite sweet	0.822 1	浓郁 Rich	0.800 8	成功 Success	0.784 6	打算 Intend	0.771 5
大果 Big fruit	0.813 6	正好 Just right	0.798 0	甘甜 Sweet	0.781 4	保存 Save	0.769 7
不算 Not count	0.812 7	太甜 Too sweet	0.797 5	超甜 Super sweet	0.779 2	清爽 Refreshing	0.765 6

2.3 多变量 SVR 预测结果分析

经 Word2vec 模型词语聚类后得到 5 类影响生鲜产品需求的影响因素,分别为口感(X_1)、包装(X_2)、物流(X_3)、性价比(X_4)和服务(X_5),在此基础上利用各个特征词的相似词语对评论文本数据进行

归类,得到各影响因素的数据集合,并运用式(1)对各个数据集进行 Z-score 标准化处理。以火龙果产品为例,将 2021-6-25—2021-8-11 共 48 天的产品特征因素及对应的历史需求量数据组成训练样本,用于训练模型,求解模型中的最优参数,建立本研究的

需求预测模型;将 2021-8-12—2021-8-17 共 6 d 的数据组成测试样本,用于检验模型,预测产品需求量。在训练样本数据中得到的单变量 SVR 预测值与多变量 SVR 预测值对比见图 2:单变量 SVR 模

型得出的需求量预测数据波动较大,在某些时间段甚至偏离了真实值,而加入在线评论影响因素的多变量 SVR 模型具有更好的预测效果,预测值与真实值相比波动小,更接近真实市场需求量数据。

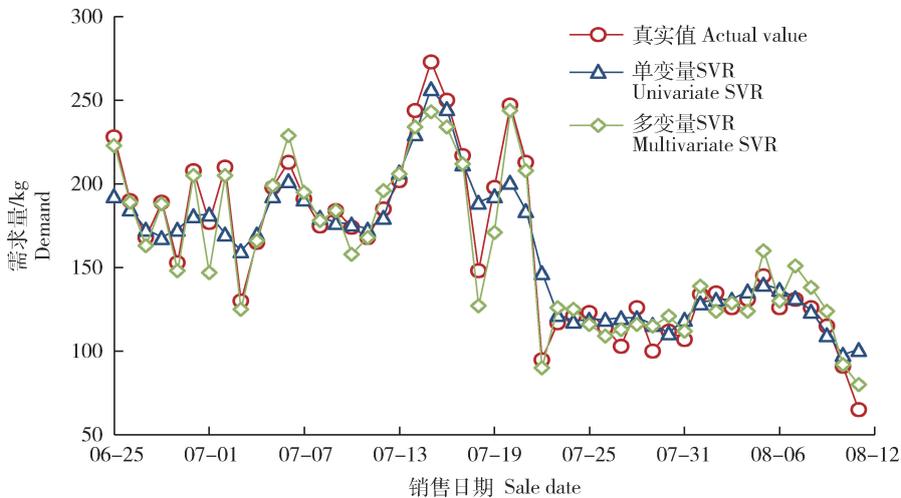


图 2 训练样本数据中火龙果需求量真实值与单变量 SVR 和多变量 SVR 的预测值

Fig. 2 The true value and the predicted value of the univariate SVR and multivariate SVR of pitaya demand in the training sample data

利用建立好的多变量 SVR 模型对测试样本的火龙果需求量数据进行预测,结果见图 3。

为更准确验证 2 种模型需求量预测的准确度,针对 3 种产品分别计算 2 种模型需求量预测值的均方误差 MSE、平均绝对误差 MAE 和判定系数 R^2 ,其中 MSE 和 MAE 数值越小、 R^2 数值越大说明预测结果越好,所得到的计算结果见表 2。可见多变量 SVR 模型的需求量预测误差均优于单变量 SVR 模型,预测精度更高。因此,在实际生产运作中,加入在线评论顾客感知产品特征因素的多变量 SVR 需求预测模型能更准确预测产品需求量。

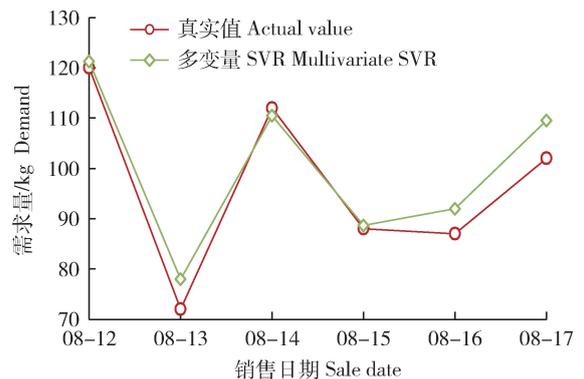


图 3 测试样本数据中火龙果需求量真实值与多变量 SVR 预测值

Fig. 3 The true value and the forecast value of multivariable SVR of pitaya demand in the test sample data

表 2 单变量 SVR 与多变量 SVR 火龙果需求量预测误差比较

Table 2 Comparison of pitaya demand forecast errors between univariate SVR and multivariate SVR

产品 Product	模型 Model	均方误差 Mean square error	平均绝对误差 Mean absolute error	判定系数 R^2 Coefficient of determination
火龙果 Pitaya	单变量 SVR	0.132 9	0.245 1	0.877 1
	多变量 SVR	0.052 5	0.176 1	0.943 4
苹果 Apple	单变量 SVR	0.122 9	0.196 9	0.877 1
	多变量 SVR	0.037 0	0.101 8	0.963 9

表 2(续)

产品 Product	模型 Model	均方误差 Mean square error	平均绝对误差 Mean absolute error	判定系数 R^2 Coefficient of determination
小龙虾 Crayfish	单变量 SVR 多变量 SVR	0.151 3 0.013 2	0.219 6 0.084 2	0.848 7 0.986 8

3 结束语

本研究针对社交媒体中在线评论数据信息,以生鲜农产品为研究对象,将文本挖掘技术和需求预测模型相结合,预测了生鲜产品下一阶段的市场需求量。利用谷歌插件 web scraper 爬取大量在线评论数据,构建 Word2vec 模型提取评论数据中顾客感知的产品特征因素,最大程度挖掘在线评论中产品特征信息,从而考虑多个因素对需求量预测准确度的影响。在此基础上,将提取出的产品特征因素量化,引入多变量支持向量回归的需求预测方法,实现生鲜产品多变量需求预测。在 python 软件上进行算例仿真分析,结果表明:在线评论中顾客感知的产品特征因素能有效用于生鲜产品需求量预测中,并且与只根据历史数据进行需求预测的单变量 SVR 相比,本研究提出的预测模型能提高产品需求量预测的准确度。

参考文献 References

[1] 张春玲. 2021 年中国生鲜电商行业研究报告[R]. 北京: 艾瑞咨询, 2021
Zhang C L. 2021 China Fresh Food E-commerce Industry Research Report[R]. Beijing: iResearch, 2021 (in Chinese)

[2] Shen L X, Li F C, Li C C, Wang Y M, Qian X Q, Feng T, Wang C. Inventory optimization of fresh agricultural products supply chain based on agricultural superdocking[J]. *Journal of Advanced Transportation*, 2020: 1-13

[3] Hu N, Liu L, Zhang J J. Do online reviews affect product sales? The role of reviewer characteristics and temporal effects [J]. *Information Technology and Management*, 2008, 9(3): 201-214

[4] 李玉鹏, 曾丽娟, 曹进. 基于顾客满意度感知要素的需求预测模型[J]. 计算机集成制造系统, 2017, 23(2): 404-413
Li Y P, Zeng L J, Cao J. Demand forecasting model based on perception factors of customer satisfaction [J]. *Computer Integrated Manufacturing Systems*, 2017, 23(2): 404-413 (in Chinese)

[5] 纪雪, 高琦, 李先飞, 高菲. 考虑产品属性层次性的评论挖掘及需求获取方法[J]. 计算机集成制造系统, 2020, 26(3):

747-759
Ji X, Gao Q, Li X F, Gao F. Reviews mining and requirements elicitation methodology considering product attributes' hierarchy[J]. *Computer Integrated Manufacturing Systems*, 2020, 26(3): 747-759 (in Chinese)

[6] 张梦莹, 邓三鸿, 王昊, 王丽娟. 基于有用性排序的在线评论与销量的关系研究[J]. 现代情报, 2019, 39(2): 152-160
Zhang M Y, Deng S H, Wang H, Wang L J. Research on the relationship between online reviews and sales based on usefulness of online reviews [J]. *Journal of Modern Information*, 2019, 39(2): 152-160 (in Chinese)

[7] 沈超, 王安宁, 方钊, 张强. 基于在线评论数据的产品需求趋势挖掘[J]. 中国管理科学, 2021, 29(5): 211-220
Shen C, Wang A N, Fang Z, Zhang Q. Product demand trend mining based on online review data[J]. *Chinese Journal of Management Science*, 2021, 29(5): 211-220 (in Chinese)

[8] 王英, 杨丹, 王乐, 张卓, 谢鹤鹏, 郝会成. 基于品牌效应的在线评论对稻米销量影响的实证研究[J]. 中国农业大学学报, 2018, 23(1): 178-187
Wang Y, Yang D, Wang L, Zhang Z, Xie H P, Hao H C. Impact of online reviews on the sales of rice under the influence of brand effect: An empirical study[J]. *Journal of China Agricultural University*, 2018, 23(1): 178-187 (in Chinese)

[9] 胡雅淇, 林海. 在线评论特征对生鲜电商农产品销量的影响: 来自淘宝羊肉大数据的证据[J]. 中国农业大学学报, 2021, 26(6): 206-218
Hu Y Q, Lin H. Effects of online-review characteristics on agricultural products sales of fresh e-commerce: Evidence from the big data about mutton products on Taobao platform[J]. *Journal of China Agricultural University*, 2021, 26(6): 206-218 (in Chinese)

[10] Lau R Y K, Zhang W P, Xu W. Parallel aspect-oriented sentiment analysis for sales forecasting with Big Data[J]. *Production and Operations Management*, 2018, 27(10): 1775-1794

[11] Chen H J, Yan Q, Xie M J, Zhang D, Chen Y P. The sequence effect of supplementary online comments in book sales[J]. *IEEE Access*, 2019, 7: 155650-155658

[12] Hu F G. The relationship analysis between online reviews and online shopping based on B2C platform technology[J]. *Cluster Computing*, 2019, 22(2): 3365-3373

[13] Gopinath S, Thomas J S, Krishnamurthi L. Investigating the relationship between the content of online word of mouth, advertising, and brand performance[J]. *Marketing Science*,

- 2014, 33
- [14] Ruiz-Mafe C, Bigné-Alcañiz E, Currás-Pérez R. The effect of emotions, eWOM quality and online review sequence on consumer intention to follow advice obtained from digital services[J]. *Journal of Service Management*, 2020, 31(3): 465-487
- [15] Hu M Q, Liu B. Mining and summarizing customer reviews [C]// Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining-KDD '04. August 22-25, 2004. New York: ACM Press, 2004
- [16] Iwata T, Hirao T, Ueda N. Topic models for unsupervised cluster matching[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2018, 30(4): 786-795
- [17] 李良强, 袁华, 叶开, 钱宇, 唐小我. 基于在线评论词向量表征的产品属性提取[J]. *系统工程学报*, 2018, 33(5): 687-697
Li L Q, Yuan H, Ye K, Qian Y, Tang X W. Extraction product features from online reviews based on word-vector-representation[J]. *Journal of Systems Engineering*, 2018, 33(5): 687-697 (in Chinese)
- [18] 王仁武, 宋家怡, 陈川宝. 基于 word2vec 的情感分析在品牌认知中的应用研究[J]. *图书情报工作*, 2017(22): 6-12
Wang R W, Song J Y, Chen C B. Application of sentiment analysis based on Word2vec in brand awareness[J]. *Library and Information Service*, 2017, 61(22): 6-12 (in Chinese)
- [19] 梁昌勇, 马银超, 陈荣, 梁焱. 基于 SVR-ARMA 组合模型的日旅游需求预测[J]. *管理工程学报*, 2015, 29(1): 122-127
Liang C Y, Ma Y C, Chen R, Liang Y. The daily forecasting tourism demand based on SVR-ARMA combination model[J]. *Journal of Industrial Engineering and Engineering Management*, 2015, 29(1): 122-127 (in Chinese)
- [20] 蔡方中, 俞婷婷, 朱国荣. 一种基于 RFE 特征选择的 PSO-SVR 用电需求预测模型[J]. *计算机应用研究*, 2020, 37(S1): 105-107
Qi F Z, Yu T T, Zhu G R. PSO-SVR electricity demand forecasting model based on RFE feature selection [J]. *Application Research of Computers*, 2020, 37(S1): 105-107 (in Chinese)
- [21] 荆浩, 刘垭, 唐金环. 基于多变量支持向量机的供应链需求预测分析[J]. *系统工程*, 2018, 36(11): 121-126
Jing H, Liu Y, Tang J H. Supply chain demand forecasting based on multivariate support vector machine [J]. *Systems Engineering*, 2018, 36(11): 121-126 (in Chinese)
- [22] Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space[EB/OL]. (2021-10-06). <https://arxiv.org/pdf/1301.3781.pdf>
- [23] Mikolov T, Sutskever I, Chen K, Corrado G Dean J. Distributed representations of words and phrases and their compositionality[J]. *Advances in Neural Information Processing Systems*, 2013, 26
- [24] 王仁武, 陈川宝, 孟现茹. 基于词向量扩展的学术资源语义检索技术[J]. *图书情报工作*, 2018, 62(19): 111-119
Wang R W, Chen C B, Meng X R. Semantic retrieval technology of academic resources based on word embedding extension[J]. *Library and Information Service*, 2018, 62(19): 111-119 (in Chinese)

责任编辑: 刘迎春