

DOI:10.3969/j.issn.1003-5060.2017.09.024

考虑品牌情感的汽车销量预测模型

刘业政^{1,2}, 章 旭^{1,2}, 王锦坤^{1,2}

(1. 合肥工业大学 管理学院, 安徽 合肥 230009; 2. 合肥工业大学 过程优化与智能决策教育部重点实验室, 安徽 合肥 230009)

摘 要:现有汽车销量预测研究大多面向整体汽车市场而非单一汽车品牌,而且缺乏对用户口碑数据的深入挖掘,忽略了口碑因素对购买意愿的影响,从而导致预测模型预测精度较低。文章提出一种改进的考虑品牌情感的自回归模型(brand opinion aware autoregressive model, BOAR),对于单一品牌,在考虑该品牌不同时间窗口历史销量基础上,结合用户在线评论中挖掘的用户情感值进行汽车销量预测。对多个汽车品牌的实验表明,该模型可以准确预测单一汽车品牌的销量,具有更好的稳定性。研究结果可以为汽车制造厂商生产规划和控制提供更为有效的决策支持。

关键词:汽车销量预测;时间序列分析;在线评论情感分析;自回归模型

中图分类号:C931.6 **文献标识码:**A **文章编号:**1003-5060(2017)09-1276-07

An automobile sales prediction model incorporating brand opinion

LIU Yezheng^{1,2}, ZHANG Xu^{1,2}, WANG Jinkun^{1,2}

(1. School of Management, Hefei University of Technology, Hefei 230009, China; 2. Key Laboratory of Process Optimization and Intelligent Decision Making of Ministry of Education, Hefei University of Technology, Hefei 230009, China)

Abstract: Existing researches on automobile sales prediction mainly focus on the whole automobile market instead of one particular brand. Moreover, current researches fail to effectively use the online reviews for the automobile, resulting in the low accuracy of sales prediction due to the discard of the influence of word-of-mouth on purchase intention. To address these issues, an improved brand opinion aware autoregressive model(BOAR) is introduced to predict each specific automobile. The proposed model considers historical sales within various time windows, and incorporates users' opinion for the certain automobile mining from online reviews to predict the sales of that brand. Results of empirical studies show that this model can accurately predict the sales for an individual automobile brand with better stability. This can provide effective decision support for the production planning and control of manufacturers.

Key words: automobile sales prediction; time series analysis; online reviews sentiment analysis; autoregressive model

中国汽车行业过去几十年取得了突飞猛进的发展,已经成为支撑和拉动中国经济持续快速增长的主导产业之一。2015 年中国汽车产销双双达到 2 450 万辆,在全球汽车制造业的市场份额已从 2000 年的 3.5% 提高到 28.0%^[1]。面对当前复杂的国内外经济环境和不断加大的经济下行

压力,准确地预测汽车销量无论是对于从宏观上整体掌控汽车市场的发育与成长态势的政策制定者,还是对于从微观上研究市场行情以制定营销策略的汽车厂商,都具有极其重要的作用。

现有汽车销量预测研究主要关注某一地区的市场总体汽车销量的宏观预测^[2-4],文献[2]基于

收稿日期:2016-12-08;修回日期:2017-02-20

基金项目:国家自然科学基金资助项目(71490725;71371062;91546114)

作者简介:刘业政(1965—),男,安徽合肥人,博士,合肥工业大学教授,博士生导师;

章 旭(1992—),男,安徽六安人,合肥工业大学硕士生,通讯作者,E-mail:zhangxu@mail.hfut.edu.cn.

时间序列分析中的差分自回归移动平均 (autoregressive integrated moving average model, ARIMA) 模型利用每月历史销售数据预测中国市场总体汽车需求量;文献[3]利用 GDP、人口、折扣率等 12 个经济指标作为自变量建立线性回归模型预测美国市场总体汽车销量;文献[4]采用时间序列分析和数据挖掘算法预测德国和美国市场总体汽车销量。对于细化至汽车品牌粒度的销量预测研究较少,因此本文提出一种汽车销量预测模型以解决单一汽车品牌的销量预测问题。

现有的汽车销量预测模型主要利用历史销售数据,缺乏对专业汽车论坛用户评论数据的深入挖掘。由于大众对产品的看法会影响产品的销售情况,因此对在线评论进行深入分析对于预测产品未来销量具有重要作用。在电影票房预测、图书销量预测等领域的研究^[5-6]表明,在线评论数据有助于提高产品销量预测模型的准确率。文献[5]从在线评论表达的情感和文本质量出发,基于时间序列分析模型有效地预测了电影票房走势,说明在线评论的情感因素会影响潜在用户的消费决策;文献[6]的研究也论证了评论情感因素对于提高图书销量预测精度有一定的作用。因此,本文所提预测模型对汽车领域不同品牌的在线口碑评论情感倾向进行分类和量化,并将品牌情感作为影响汽车销量的重要因素用于预测汽车销量。

此外,我国汽车市场受气候、节假日、产品更迭等诸多因素的影响,汽车月度销售数据呈现稳定的季节性特征,因此本文的模型进一步考虑了汽车的历史同期销量数据。

综合上述问题,本文面向汽车领域,基于汽车专业论坛在线口碑评论大数据,综合考虑历史同期销量、前期销量以及口碑评论的情感倾向对汽车销量的影响,提出一种考虑品牌情感的自回归模型 (brand opinion aware autoregressive model, BOAR) 用于预测各汽车品牌月度销量。本文利用汽车专业论坛爬取的真实数据并结合多个汽车品牌的实际销量数据进行实验,结果表明对于所有汽车品牌,所提 BOAR 模型平均预测误差小于 10%,比传统时间序列预测中的 ARIMA 模型具有更高的预测精度,表明在线评论情感倾向有助于提升汽车销量预测模型的精度。

1 模型构建

1.1 前期销量的影响

现有销量预测研究主要利用产品前期销售数

据预测未来销量。前期销售数据包含了众多复杂且未知的变量对销量的影响,将前期销量作为自变量预测未来销量避免了复杂因素对销量的影响。文献[2]基于历史前期的销售数据,利用自回归移动平均模型有效预测了中国市场总体汽车需求量,在利用前期销量预测未来销量的方法中,最常用的就是自回归模型。

自回归模型是用自身做回归变量的过程,即利用前期若干时刻的随机变量的线性组合来描述以后某时刻随机变量的线性回归模型,它是时间序列中的一种常见形式^[7]。自回归模型被广泛应用于包括销量预测在内的时间序列分析中。显然汽车品牌的销量会受到该品牌前期销量的影响,因此本文采用自回归模型作为基础模型预测汽车品牌销量。

具体地,用 y_t 表示第 t 个月的汽车品牌销量, $t=1, 2, \dots, N$, 并用 $\{y_t\}$ 表示整个时间序列 y_1, y_2, \dots, y_N 。对汽车销量时间序列建立 p 阶自回归销量预测模型 (记为 $AR(p)$), 即

$$y_t = \sum_{i=1}^p \varphi_i y_{t-i} + \varphi_0 + \varepsilon_t \quad (1)$$

其中, p 为要考察的第 t 个月之前 p 个月的销售情况对第 t 个月的影响; y_t 为第 t 期销量; φ_0 为常数项; φ_i 为最小二乘回归得到的模型参数; ε_t 为第 t 期的误差项。

自回归模型可以有效捕获前期销量对于销量预测的影响,本文在自回归模型的基础上引入更多变量,提高销量预测精度。

1.2 历史同期销量的影响

产品销售的季节性是指产品的销量随着季节的变化而呈现出高峰低谷的规律性周期变动趋势。这一特性已在其他领域得到证明。文献[8]在预测鞋子的销量时考虑了产品销售的季节特性,在传统预测模型基础上融入了时间维度,提高了销量预测的精度。

由于受气候、节假日、产品更迭等诸多因素的影响,我国汽车月度销售数据呈现出明显的季节性特征,主要为年末至年初为旺季,年中为淡季。文献[9]的研究论证了我国汽车市场月度销量数据具有周期为 12 的季节性波动。中国汽车市场 2013-09 至 2016-08 月度销售数据如图 1 所示。

由图 1 可以看出,每年的 12 月和 1 月为汽车销售旺季,2 月和 7 月为汽车销售淡季,有着明显的季节性波动。因此本文在自回归模型基础上引入历史同期销量因素 y_{t-12} , 得到考虑历史同期销

量的自回归预测模型:

$$y_t = \sum_{k=1}^r \lambda_k y_{t-12k} + \sum_{i=1}^p \varphi_i y_{t-i} + \varphi_0 + \varepsilon_t \quad (2)$$

其中, r 为要考察的第 t 个月之前 r 个历史同期销量对第 t 个月的影响; p 为要考察的第 t 个月之前 p 个月的销售情况对第 t 个月的影响; y_t 为第 t 期销量; φ_0 为常数项; φ_i 为最小二乘回归得到的模型参数; ε_t 为第 t 期的误差项。

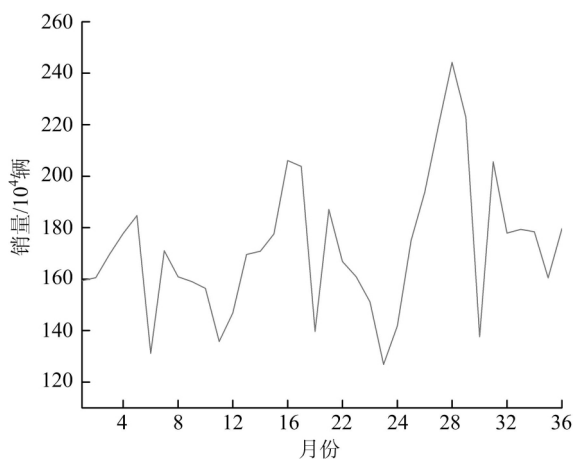


图1 中国汽车市场 2013-09 至 2016-08 月度销售数据

1.3 在线口碑评论中情感因素的影响

在线商品评论数据的快速增长为了解普通消费者的观点提供了一个很好的契机,有很多专家学者开始研究在线评论数据中隐藏的价值。情感分析在微博分类^[10]、客户意见聆听^[11]等诸多领域得到了广泛应用。文献[12-14]的研究表明在评论文本中产品被提及的次数以及这些数据中所蕴含的情感倾向对于预测产品的销量有明显的效果,尤其是类似汽车、住房等需要顾客投入时间和精力的高度介入产品,在线评论数据中的情感倾向对于销量预测的效果更加显著。文献[15]利用社交媒体文本情感和汽车关键词被提及的频率提高汽车销量预测精度;文献[16]在考虑预发布营销、影院可用性和专业评论的基准模型中增加在线产品评论指标,大大提高了电影票房收入预测准确性。鉴于以上原因,本文对汽车领域不同品牌的在线口碑评论情感倾向进行分类和量化,并将品牌情感作为影响汽车销量的重要因素用于预测汽车销量。用 ω_t 表示第 t 个月汽车品牌口碑情感值, $t=1, 2, \dots, N$ 。 ω_t 为不可直接观测的变量,需要从评论文本中去挖掘。

因此在(2)式基础上进一步引入汽车品牌口碑情感 ω_t , 得到考虑品牌情感的自回归模型(记

为 BOAR):

$$y_t = \sum_{k=1}^r \lambda_k y_{t-12k} + \sum_{i=1}^p \varphi_i y_{t-i} + \sum_{j=1}^q \rho_j \omega_{t-j} + \varphi_0 + \varepsilon_t \quad (3)$$

其中, r 为要考察的第 t 个月之前 r 个历史同期销量对第 t 个月的影响; p 为要考察的第 t 个月之前 p 个月的销售情况对第 t 个月的影响; q 为要考察的第 t 个月之前 q 个月的口碑情感对第 t 个月的影响; y_t 为第 t 期销量; ω_t 为第 t 期品牌口碑情感; φ_0 为常数项; λ_k 、 φ_i 和 ρ_j 为用最小二乘回归得到的模型参数; ε_t 为第 t 期的误差项。

BOAR 模型包含如下 3 个部分:

- (1) $\sum_{k=1}^r \lambda_k y_{t-12k}$ 反映前 r 年历史同期销量数据对第 t 期销量的影响。
- (2) $\sum_{i=1}^p \varphi_i y_{t-i}$ 反映前 p 期的历史销量数据对第 t 期销量的影响。
- (3) $\sum_{j=1}^q \rho_j \omega_{t-j}$ 反映前 q 期的品牌口碑情感数据对第 t 期销量的影响。

2 数据与变量

2.1 销量数据

本文使用的月度销量数据是利用网络爬虫从搜狐汽车采集汽车工业协会发布的中国市场 10 个常见汽车品牌(丰田、奥迪、大众、别克、本田、比亚迪、福特、奇瑞、雪佛兰和起亚)从 2011-05 至 2015-10 月的月度汽车销量 y_t ^[17], 如图 2 所示。

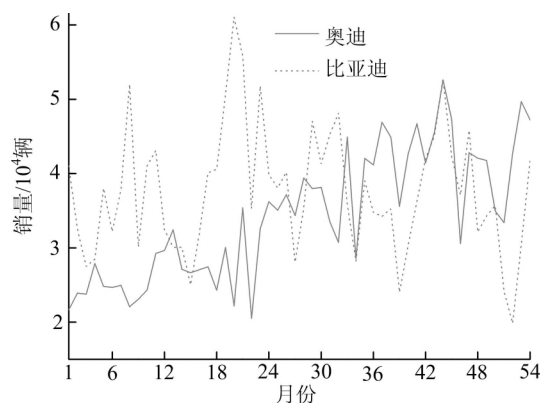


图2 随时间波动的汽车月度销售数据

2.2 在线口碑评论情感数据

本文使用的在线口碑评论数据源自于“汽车之家”口碑数据^[18]。使用网络爬虫从汽车之家网站采集 472 346 条口碑评论,将评论文本过少的

舍弃,共计有效口碑数据 330 877 条。

2.2.1 情感词典构造方法

情感词典是本文口碑评论文本情感值计算的基础。目前已经有很多学者提出通用的情感词典应用于文本情感的分析,但汽车领域的情感词典目前的研究还很少。汽车领域情感词典构造存在着诸多的困难:一是汽车领域专业化程度高,涉及到大量的专业术语,如油耗高、加速快,这些文本的情感通用的情感词典难以准确描述;二是目前通用的情感词典采用的语料内容主要是传统的词典、字典以及正规的长文本内容。汽车论坛的口碑数据全部由购车的网民发表,文本组织大多比

较自由,没有特定的行文规范。

为解决上述情感词典构造的难题,本文基于情感词典构造的一般思路^[19],针对汽车之家口碑数据的特点提出基于点互信息(PMI)的情感词典构造方法,并对在线口碑评论情感分类和量化。本文使用的汽车之家网站 1 条完整的口碑数据如图 3 所示。图 3 共包含如下 3 类数据:① 评分数据(1~5 分),如图 3 中区域 A 所示;② 评论文本数据,如图 3 中区域 B 所示;③ 口碑发表时间数据,如图 3 中区域 C 所示。对于评论文本数据,本文利用开源中文分词工具 HanLP^[20]进行词汇切分并去除常见停用词。

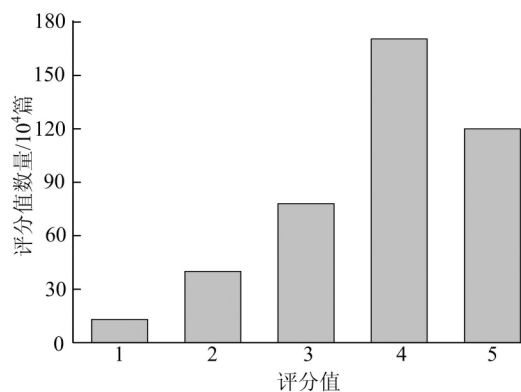
图 3 展示了汽车之家网站上的一条完整口碑数据。该数据包含以下部分：

- 区域 A (评分数据):** 显示了购车信息（车型、地点、经销商、时间、价格、油耗、里程）以及各项性能指标的评分（空间、动力、操控、油耗、舒适性、外观、内饰、性价比）。
- 区域 B (评论文本):** 包含购车目的、购车时间、购车地点、购车经销商、购车时间、购车价格、购车油耗、购车里程、购车空间、购车动力、购车操控、购车油耗、购车舒适性、购车外观、购车内饰、购车性价比等详细评价。
- 区域 C (口碑发表时间):** 显示了购车时间、购车地点、购车经销商、购车时间、购车价格、购车油耗、购车里程、购车空间、购车动力、购车操控、购车油耗、购车舒适性、购车外观、购车内饰、购车性价比等详细评价。

图 3 汽车之家网站 1 条完整的口碑数据

具体地,将 8 条评分项分别记为 G_i , 其中 $i \in$ (空间,动力,操控,油耗,舒适性,外观,内饰,性价比), 12 条评论文本项分别记为 R_j , 其中 $j \in$ (最满意的一点,最不满意的一点,空间,动力,操控,油

耗,舒适性,外观,内饰,性价比,为什么最终选择这款车,其他描述);口碑发表时间记为 t 。其中评分项中的每一个 G_i ,在评论文本项中都会有对应的 R_i 。所有口碑数据的评分项分布情况如图 4 所示。

图 4 所有评分项 G_i 得分分布情况

由图 4 可以明显得知网络消费者更加倾向于给出较高的评分。这也是合理的,因为汽车作为一种高介入商品,消费者在购买前通常会深入了解相关信息并做出自己最满意的决策,所以购车后的评价更多倾向于正面。因此本文认为,当消费者对某项评价为 1 分或 2 分时,表示消费者对该项非常不满意;而给出 5 分时,则认为消费者对该项满意。对于 1 条口碑数据,若 $G_i \leq 2$,则认为 R_i 为负向的,属于负向文档集 D_n ;若 $G_i = 5$,则认为 R_i 为正向的,属于正向文档集 D_p ;同时本文认为 $R_{\text{最不满意的一点}}$ 为负向的,属于 D_n ; $R_{\text{最满意的一点}}$ 为正向的,属于 D_p 。文档中每个词 w 的情感值 S_w 计算公式如下:

$$S_w(w) = P(w, \text{pos}) - P(w, \text{neg}) \quad (4)$$

其中, P 为点互信息,即

$$P(w, \text{pos}) = \lg \frac{f(w, \text{pos})N}{f(w)f(\text{pos})} \quad (5)$$

其中, $f(w, \text{pos})$ 为词 w 在正向文档集 D_p 中出现的频次; $f(w)$ 为词 w 在整体文档集 D ($D = D_p + D_n$) 中总共出现的频次; $f(\text{pos})$ 为正向文档的数量; N 为整体文档集中所有文档的数量。同理可以计算 $P(w, \text{neg})$ 。由此, (4) 式可以简化为:

$$S_w(w) = \lg \frac{f(w, \text{pos})f(\text{neg})}{f(w, \text{pos})f(\text{pos})} \quad (6)$$

由于点互信息对于低频次词语估计效果较差,将正向文档集和负向文档集中出现频次低于 5 的词全部舍弃。

2.2.2 汽车品牌情感计算方法

根据 2.2.1 构造的情感词典对汽车品牌的口碑评论情感进行量化。对每一篇口碑评论文本仅抽取情感词典中包含的词汇,则第 k 篇口碑评论文本情感值 $S_{\text{rev}}(r_k)$ 的计算公式如下:

$$S_{\text{rev}}(r_k) = \sum_{w=1}^{M_k} S_w(w) \quad (7)$$

其中, M_k 为第 k 篇口碑评论中共含有 M_k 个情感词典中的词,即口碑评论文本的情感值由该评论文本中所有情感词的情感值累加得出。

由于口碑数据是针对具体车型,计算品牌情感值时,该品牌下所有车型对应的口碑数据均视为针对该品牌的口碑评论。则第 t 期品牌情感值 ω_t 为:

$$\omega_t = \sum_{k=1}^{K_t} S_{\text{rev}}(r_k) \quad (8)$$

其中, K_t 为该汽车品牌下第 t 期用户发表的所有口碑数据。即汽车品牌的情感值为该品牌下所有车型的口碑评论文本情感值累加得出。

3 实验验证

3.1 实验流程

如前文所述,汽车销量序列存在一定的季节性波动,因此销量序列可能是非平稳的。在目前的时间序列分析中,对于非平稳的时序分析问题通常采用 ARIMA 模型模拟生成时间序列的随机过程,因此本文将 BOAR 模型和 ARIMA 模型的预测效果进行对比。本文主要实验流程如下:

- (1) 构造情感词典,计算每条口碑的情感值。
- (2) 计算上述 10 个品牌各月的口碑情感值。
- (3) 利用 ARIMA 模型预测每个月的销量。
- (4) 利用 BOAR 模型预测每个月的销量。

由于在汽车品牌认可度比较稳定的情况下,前一年的历史同期销量基本可以反映行业变化,假定 $r=1$ 。另外,本文认为 1 a 内的前期销量基本可以反映前期销量对于销量预测的影响,且用户计划购车时查看他人评论的时间与实际购车时间通常相差在 1 a 以内,所以共考察 $p=1, 2, \dots, 11, q=1, 2, \dots, 12$ 时模型的预测效果。第 $t-12$ 的前期销量即为 $r=1$ 时的历史同期销量,故不考察 $p=12$ 。

(5) 分别计算 2 种模型的预测误差。本文采用平均绝对百分误差(mean absolute percentage error, MAPE)来衡量模型预测准确率:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|\text{Pred}_i - \text{True}_i|}{\text{True}_i} \times 100\% \quad (9)$$

其中, n 为总的预测期数; Pred_i 为第 i 期模型的预测值; True_i 为真实值。从 (9) 式可看出,预测值与实际值越接近,MAPE 值越低,即预测得越准确。

由于涉及的口碑评论数据量较大,且多为非结构化的文本内容,本文采用 Hadoop 平台的

MapReduce 编程实现数据的并行化计算,快速构建汽车领域情感词典并计算汽车品牌口碑情感值,有效地节省了时间资源,提高了计算效率。

3.2 实验结果与分析

3.2.1 情感词典构造结果

利用 2.2.1 节所述方法构建情感词典,词典中共包含 2 045 个词。情感词典部分结果见表 1、表 2 所列。

表 1 情感词典中情感值最小的 10 个词

w	$S_w(w)$
有异	-6.495
太薄	-6.457
漏水	-6.271
太差	-5.889
共振	-5.578
响声	-5.574
烦人	-5.526
太弱	-5.446
异响	-5.292
太慢	-5.104

表 2 情感词典中情感值最大的 10 个词

w	$S_w(w)$
回头率	5.366
吸引	5.097
很棒	5.091
没的说	5.025
很漂亮	4.956
大气	4.955
精准	4.740
OK	4.697
性感	4.446
漂亮	4.301

表 1 中最小的 10 个词全部为负向词,表 2 中最大的 10 个词全部为正向词。这也验证了这种情感词典构造方法的有效性。

3.2.2 预测效果分析

分别利用 ARIMA 模型和 BOAR 模型预测上述 10 个常见汽车品牌每个月的销量。对于 BOAR 模型,考虑 $r=1, p=1, 2, \dots, 11, q=1, 2, \dots, 12$ 时模型预测效果,使用网格搜索法(grid search)选择每个汽车品牌最优的参数组合,得到 MAPE 最小的参数组合,实验结果见表 3 所列。

从表 3 可以看出,考虑情感因素的 BOAR 模型预测精度更高,有效地降低了预测误差。证明口碑情感因素在预测销量问题中确实发挥了作用。其中,奇瑞汽车精度提高最大,达 13.51%;大众汽车预测精度提高最小,仅为 0.63%。ARIMA 模型平均预测误差为 12.17%,而 BOAR 模型平均预测误差仅为 5.93%,预测误差减少了

6.24个百分点。

表 3 ARIMA 模型及 BOAR 模型预测效果

汽车品牌	ARIMA		BOAR		误差减少量/%
	(p, d, v)	MAPE/%	(r, p, q)	MAPE/%	
丰田	(9, 1, 7)	16.06	(1, 1, 10)	7.30	8.76
奥迪	(5, 1, 5)	11.77	(1, 6, 10)	2.79	8.98
大众	(7, 0, 5)	9.07	(1, 2, 4)	8.44	0.63
别克	(9, 1, 7)	9.10	(1, 2, 10)	6.32	2.78
本田	(10, 0, 7)	16.37	(1, 7, 10)	3.68	12.69
比亚迪	(9, 0, 5)	11.25	(1, 8, 9)	3.68	7.57
福特	(7, 1, 6)	9.89	(1, 5, 7)	5.44	4.45
奇瑞	(6, 0, 7)	16.50	(1, 7, 11)	2.99	13.51
雪佛兰	(6, 0, 4)	10.41	(1, 1, 9)	8.99	1.42
起亚	(4, 1, 3)	11.25	(1, 2, 10)	9.69	1.56

通过对 BOAR 模型最优的参数组合观察发现,不同品牌的 BOAR 模型最优参数组合的 q 取值大多集中在 10 附近。这说明口碑情感对销量的影响滞后期约为 10 个月,即用户计划购车前查看他人评论的时间与实际购车时间通常相差在 10 个月以内,这与常识相符合。

3.2.3 参数稳定性分析

在回归分析的研究中,模型参数是否具有稳定性对所建立的模型来说具有非常重要的意义。本文基于回归分析对销量预测建立 BOAR 模型,设 BOAR 模型的最优参数组合为 (r', p', q') ,下面分别考察参数选取 $(r', p'-1, q')$ 和 $(r', p'+1, q')$ 以及 $(r', p', q'-1)$ 和 $(r', p', q'+1)$ 时模型的预测情况,结果如图 5、图 6 所示。

结合图 5、图 6 不难发现,在参数 p 和 q 各自波动 ± 1 时,BOAR 模型预测结果比较一致、稳定,预测误差 MAPE 值在正常范围内波动,没有出现异常变化,这也进一步验证了 BOAR 模型参数具有稳定性。

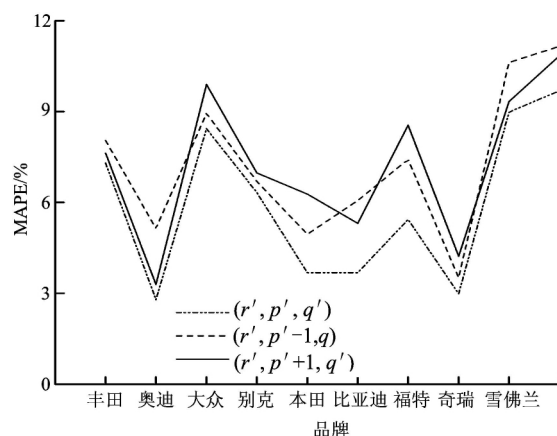


图 5 $(r', p'-1, q')$ 和 $(r', p'+1, q')$ 时模型预测结果

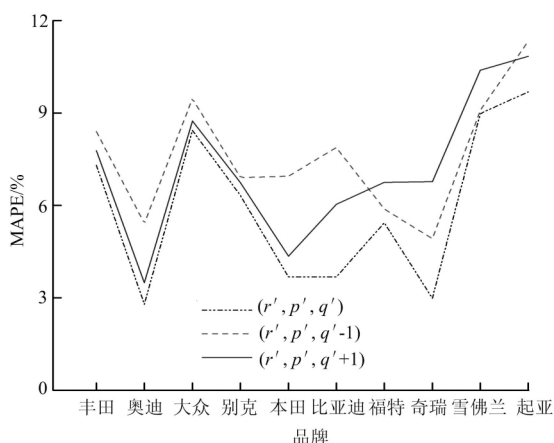


图 6 $(r', p', q'-1)$ 和 $(r', p', q'+1)$ 时模型预测结果

4 结 论

在现有研究的基础上,本文利用在线评论数据和销量数据探讨了历史销量和品牌情感在汽车预测方面发挥的作用。综合历史同期销量、前期销量和品牌情感的汽车销量预测模型平均预测误差为 5.93%,比 ARIMA 模型降低了 6.24 个百分点,这也验证了历史同期销量和在线评论情感倾向在汽车销量预测研究中具有重要作用。该模型可以为行业相关决策者提供可靠的销量预测数据,从而辅助制定生产、销售和营销计划。

同时,本文提出的预测模型对于销量波动大的时间点预测结果不甚理想。在时间序列相关预测问题研究中,神经网络、SVR 等数据挖掘算法已得到了广泛应用^[21-22],在后续工作中拟采用 BP 神经网络等数据挖掘算法,考虑更多的影响因素,如消费者搜索行为数据、汽车厂商广告投放数据等,以进一步降低预测的误差。

【参 考 文 献】

- [1] 中国汽车工业协会. 中国汽车工业发展年度报告(2016) [M]. 北京: 社会科学文献出版社, 2016.
- [2] CHEN D. Chinese automobile demand prediction based on ARIMA model[C]//International Conference on Biomedical Engineering and Informatics, BMEI 2011, Shanghai, China. [S. l. : s. n.], 2011: 2197-2201.
- [3] SHAHABUDDIN S. Forecasting automobile sales[J]. Management Research News, 1978, 32(32): 670-682.
- [4] HÜLSMANN M, FRIEDRICH C M, REITH D, et al. General sales forecast models for automobile markets and their analysis[J]. Ibai Publishing, 2011, 5(5): 65-86.
- [5] YU X, LIU Y, HUANG X, et al. Mining online reviews for predicting sales performance: a case study in the movie domain[J]. IEEE Transactions on Knowledge & Data Engineering, 2010, 24(4): 720-734.

- [6] 李雪妮, 张绍武, 杨亮, 等. ARES: 用于预测的情感感知自回归模型[J]. 计算机研究与发展, 2013, 50(8): 1722-1727.
- [7] SHUMWAY R H, STOFFER D S. Time series analysis and Its applications [J]. Springer Texts in Statistics, 2009, 102(479): 1079-1079.
- [8] XIONG L, CHEN X, HUANG T K, et al. Temporal collaborative filtering with bayesian probabilistic tensor factorization[C]//Siam International Conference on Data Mining, SDM 2010, April 29-May 1, 2010, Columbus, Ohio, USA. [S. l. : s. n.], 2010: 211-222.
- [9] 王旭天, 李政远, 舒慧生. 基于 SARIMA 的我国汽车销量预测分析[J]. 中国市场, 2016(1): 71-74.
- [10] 张金伟, 刘晓平. 基于心理预警模型的微博情感识别研究[J]. 合肥工业大学学报(自然科学版), 2013, 36(11): 1318-1322.
- [11] 何建民, 刘明星. 面向特定主题的客户抱怨文本分类识别方法[J]. 合肥工业大学学报(自然科学版), 2010, 33(12): 1884-1888.
- [12] CHOI H, VARIAN H. Predicting the present with google trends[J]. Economic Record, 2009, 88(S1): 2-9.
- [13] WU L, LIN C Y, ARAL S, et al. Value of social network-a large-scale analysis on network structure impact to financial revenue of information technology consultants [C]//The Winter Conference on Business Intelligence. [S. l. : s. n.], 2009: 1-25.
- [14] GEVA T, OESTREICHER-SINGER G, EFRON N, et al. Using forum and search data for sales prediction of high-involvement products[J]. MIS Quarterly, 2017, 41(1): A1-A4.
- [15] AHN H I, SPANGLER W S. Sales prediction with social media analysis[C]//Srii Global Conference, Washington, D. C. : IEEE Computer Society, 2014: 213-222.
- [16] DELLAROCAS C, ZHANG X, AWAD N F. Exploring the value of online product reviews in forecasting sales: The case of motion pictures[J]. Journal of Interactive Marketing, 2007, 21(4): 23-45.
- [17] 汽车工业协会. 搜狐汽车销量[EB/OL]. [2016-09-10]. <http://db.auto.sohu.com/carsales/>.
- [18] 汽车之家. 汽车之家口碑频道[EB/OL]. [2016-09-10]. <http://k.autohome.com.cn/#pvareaid=103420>.
- [19] KIRITCHENKO S, ZHU X, MOHAMMAD S M. Sentiment analysis of short informal text[J]. Journal of Artificial Intelligence Research, 2014, 50: 723-762.
- [20] 上海林原信息科技有限公司. HanLP[EB/OL]. [2016-09-10]. <http://hanlp.linrunsoft.com/>.
- [21] 姜平, 石琴, 陈无畏, 等. 基于 Elman 型回归神经网络的公交客流预测[J]. 合肥工业大学学报(自然科学版), 2008, 31(3): 340-342.
- [22] 陈荣, 梁昌勇, 谢福伟. 基于 SVR 的非线性时间序列预测方法应用综述[J]. 合肥工业大学学报(自然科学版), 2013, 36(3): 369-374.

(责任编辑 张 镔)