沈 桐,向 菲. 基于用户评论情感分析的医院评分模型构建[J]. 中华医学图书情报杂志,2020,29(9):46-51.

DOI:10.3969/j.issn.1671-3982.2020.09.007

·情报研究与方法·

基于用户评论情感分析的医院评分模型构建

沈 桐,向 菲

[摘要] 目的:构建基于用户评论情感分析的医院评分模型。方法:运用 ROST EA 工具对文本评论进行分析和利用相关分析对星级评分与文本评论的拟合程度进行分析,构建基于线性回归的文本评论自动打分模型,对星级评分与文本评论的相关性进行检验。结果:当情感强度的比值是"一般:中度:高度=1:1:1"时,两者具有最高的相关性,相关性系数为 0.680。结论:最终回归模型为 A=0.961exp(E0)+7.348,星级评分可以在一定程度上代表文本评论,说明该模型具有一定应用价值。

[关键词]情感分析;在线评论;医院评价;意见挖掘

[中图分类号]R197.3 [文献标志码]A [文章编号]1671-3982(2020)09-0046-06

Establishment of hospital scoring model based on emotion analysis of users´ reviews

SHEN Tong, XIANG Fei

(School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 400030, Hubei Province, China)

Corresponding author: XIANG Fei

[**Abstract**] **Objective** To establish the hospital scoring model based on emotion analysis of users´ reviews. **Methods** The linear regression-based automatic scoring model of text reviews was established by analyzing the text reviews with ROST EA and the fitting degree of star scoring and text reviews by correlation analysis. The correlation between star scoring and text reviews was verified. **Results** The highest correlation coefficient of star scoring and text reviews was 0.680 when the emotion intensity ratio was "generality : moderation : height = 1 : 1 : 1. **Conclusion** The final regression model is A = 0.961exp (EO) +7.348, indicating that star scoring can represent the text reviews in a certain extent and the model is of a certain application value.

[**Key words**] Emotion analysis; Online review; Hospital evaluation; Opinion mining

2018 年 4 月 28 日,国务院办公厅印发《关于促进"互联网+医疗健康"发展的意见》,就促进医疗健康与互联网深度融合发展进行部署。"互联网+医疗健康"带来了新的医疗理念。根据第 45 次《中国互联网络发展状况统计报告》的统计结果,截至 2020 年 3 月,我国互联网普及率达到 64.5%,较 2018 年底增长了 4.9%,网民规模达 9.04 亿,较 2018 年底新增了 7 508 万[1]。随着互联网发展而出现的网上预约挂号平台延续了电子商务的部分特点,产生了大量的用户评价,成为了用户在进行决策时所考虑的重要因素。本文对用户线下就诊体验的文本评论进行分析,进而对星级评分与文本评论的拟合程度进行研究并找出两者相关系数最高时的条件,构建基于线性回归的文本评论自动打分模型。本文的研究结果一方面提供了一个对医疗服务较为客观准确的打分,有利于为用户选择医院就诊提供

[作者单位]华中科技大学同济医学院医药卫生管理学院,湖北 武汉 430030

[作者简介]沈 桐(1997-),男,湖北武汉人,在读硕士研究生。

[通讯作者]向 菲(1981-),男,湖北武汉人,博士,副教授,系副主任,研究方向为健康信息资源管理、社区健康信息服务、健康信息需求与行为、卫生知识管理。E-mail:xiangfei@hust.edu.cn

决策参考;另一方面,使医院对用户的评价有一个整体的了解,促使医院提高服务效率,改进服务质量,创造更为良好的医疗环境,从而最大程度地满足用户需求。

1　文献回顾

目前,关于用户评价的研究主要集中在文本挖掘、产品口碑、评论有用性 3 个方面[2]。

文本挖掘方面的研究主要包括方法学及应用的研究。如有学者开发了一个基于语义词典的用户在线评论的文本挖掘方法[3],有学者利用基于概率图模型理论的潜在狄利克雷分布提出了文本的潜层结构特征的概念[4],也有学者基于文本挖掘的方法对用户反馈的服务质量及其影响进行了研究[5],还有学者利用文本挖掘对影响消费者差评行为的重要因素进行了研究[6]。

产品口碑的研究主要是对网络口碑产生的原因及作用的研究。有研究结果显示,激励政策会对消费者评论产生影响[7];有研究则从评论对价格的影响着手研究了网络口碑所产生的作用[8]。但研究者不仅关注网络口碑产生的前因与后果,同时也更关注网络口碑如何才能得到更好的利用,以及企业等第三方如何更好地借助网络口碑这一载体创造价值[9]。如有学者运用基于细粒度的情感分析技术并结合自回归滑动平均模型(autoregresssive moving average model, ARMA)预测了商品的销量,并进行了实证分析[10]。

评论有用性的相关研究主要集中在影响评论有用性的因素及其影响机制的研究。如有学者以信息采纳理论为研究框架,从社会网络视角探讨了用户评论有用性的两类影响因素,即评论本身特征和评论者要素[11];有学者从在线评论的形式特征和内容特征两个方面提取出影响在线评论有用性的 7 个指标属性进行量化计算,构建了在线评论有用性指标体系和排序模型[12]。

目前情感分析的方法主要包括基于语义词典的方法和基于机器学习的方法。基于机器学习的方法是利用机器学习的各种分类方法识别情感;基于语义词典的方法一般在构造情感词词典或列表的基础上,将分词后的评论中所包含的词语与情感词词典中的词进行匹配,从而判断评论的情感倾向[13]。

综上所述,无论产品口碑、评论有用性还是评论文本挖掘的研究,都是从用户评论中挖掘有价值的信息,最终服务于商家或用户。本文的研究最终服务的是医院和患者,对用户评论的研究既包括其内容挖掘又包括其本身特性或其产生的后果、涉及的影响因素等。文本挖掘为社会科学领域的研究提供了新形式的数据来源,也符合大数据驱动的研究模式。从以往的研究来看,现阶段的用户评价研究主要集中在酒店、电影、图书三大领域[14]。本文将在线评论的情感分析用于医疗领域,在“互联网+医疗”的背景下,运用计算机技术与统计学方法,充分挖掘用户对医院的评价,以期为用户和医院提供有价值的参考。

2　数据与方法

2.1　研究内容

首先对用户的文本评论与星级评论的一致性进行检验,其次分析情感强度的改变对两者相关性的影响,最后建立一个基于线性回归的文本评论自动打分模型并对模型进行验证。其技术路线如图 1 所示。

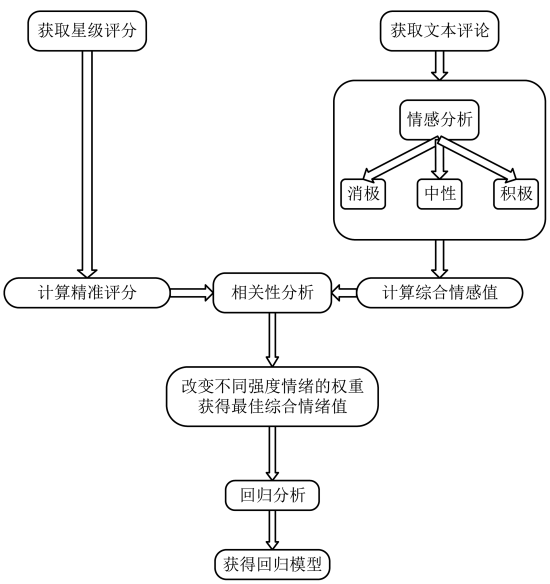


图 1　基于线性回归的文本评论自动打分模型技术路线

2.2　数据来源

“微医挂号网”(以下简称“微医网”)是原国家卫生和计划生育委员会批准的全国就医指导及健康咨询平台,是国际领先的移动医疗服务平台。其主要功能之一就是挂号,用户可以通过“微医网”挂号

并在线下就诊完成后对就医体验进行反馈,因而产生了大量的线下就诊用户反馈数据。“微医网”的用户评价主要分为文本评论和满意度评分(本文称为“星级评分”)。文本评论是患者根据其就医的经历体验用文字叙述的方式对医生、医院或就医结果等进行评价,满意度评分则是患者对就医的满意度进行主观的打分,用星级体现,最低为 1 星,最高为 5 星。本文运用“爬虫”工具随机获取了该平台上 300 多家医院的线下就诊用户反馈数据。由于所获取的数据中部分医院患者评价数过少,可能会影响后续步骤的准确性,因此剔除评价数少于 50 条的医院后,将余下的 193 家医院作为本文的研究样本。

2.3 研究方法

本文的主要分析方法是基于情感词词典的情感分析方法,所选用的文本分析工具是武汉大学的基于情感词词典的 ROST EA 情感分析工具。首先,运用 ROST EA 情感分析工具对文本评论的情感极性进行分析的基础上,计算代表文本评论的综合情感值和代表星级评分的精准评分;其次,运用相关分析对星级评分与文本评论的拟合程度进行研究;最后,在两者相关程度最高的条件下构建线性回归模型,并结合多种检验方法对模型进行修正并再次检验。

2.4 变量计算

根据 ROST EA 情感分析工具的分析结果,用 P、M、N 分别代表每家医院评论中积极、中性、消极情感评论占比,那么原始综合情感值 E0 的计算公式为 $E0=P\times1+M\times0+N\times(-1)$,以此代表文本评论的评价结果。

根据对每家医院的评价中星级评分占比的统计结果,用 S1、S2、S3、S4、S5 分别代表 1 星到 5 星评分占比,那么精准评分 A 的计算公式为 $A0=(S1\times1+S2\times2+S3\times3+S4\times4+S5\times5)/100\%$。由于原评分为 5 分制,转化成 10 分制则为 $A=(S1\times1+S2\times2+S3\times3+S4\times4+S5\times5)/100\%\times2$。

3 结果

3.1 星级评分与文本评论的一致性检验

运用 SPSS 22.0 对原始综合情感和精准评分进行相关性分析。软件分析结果显示两者相关系数为 0.680,显著性检验值为 0.000,说明原始综合情感和精准评分显著相关,星级评分可以在一定程度上代表文本评论。

ROST EA 工具除了可将评论的情感极性分为“消极”“中性”“积极”外,还可根据分析得出的评论情感值进一步将“积极”分为“一般积极”“中度积极”“高度积极”,将“消极”分为“一般消极”“中度消极”“高度消极”。在前面的 E0 计算中,一般、中度和高度 3 种强度的权重比默认为“1 : 1 : 1”。为了探讨情感强度的改变对星级评分和文本评论相关性的影响,现通过调整 3 种强度的权重比,比较不同情感强度(即对不同程度的情感倾向赋予不同的权值)下得到的综合情感与精准评分的相关性。

综合情感的计算公式为 $E=Np\times X1+Mp\times X2+Hp\times X3+M\times0+Nn\times(-X1)+Mn\times(-X2)+Hn\times(-X3)$。公式中 Np、Mp、Hp 分别代表“一般积极占比”“中度积极占比”“高度积极占比”,M 为中性情感占比,Nn、Mn、Hn 分别代表“一般消极占比”“中度消极占比”“高度消极占比”。X1:X2:X3 代表情感强度为“一般”“中度”“高度”的权重比。本文设置了 7 种不同的权重比,分别为“1 : 1 : 1”“1 : 2 : 3”“1 : 2 : 4”“1 : 3 : 5”“3 : 2 : 1”“4 : 2 : 1”“5 : 3 : 1”。

E0、E1、E2、E3、E4、E5、E6 分别对应以上不同权值下计算所得的综合情感值。将不同权值下计算所得的综合情感值分别与精准评分进行相关性分析,结果如表 1 所示。

表 1 不同情况下综合情感值与精准评分的相关性

综合情感值	相关系数	显著性检验值
E0	0.680	0.000
E1	0.661	0.000
E2	0.607	0.000
E3	0.616	0.000
E4	0.450	0.000
E5	0.410	0.000
E6	0.373	0.000

结果显示,在各种权值下,对应的综合情感值与精准评分都显著相关,因此文本评论与星级评分具有较高的一致性。在“一般 : 中度 : 高度 = 1 : 1 : 1”的情况下,综合情感值和精准评分具有最高的相

关性。因此在接下来的步骤中,所有的分析将以此为前提进行。

3.2　模型检验

基于线性回归的文本评论自动打分模型调整过程如图 2 所示。经过多次调整后,最终以 *exp*(E0)为自变量,以精准评分 A 为因变量建立模型(A = *f*(*exp*(E0)))进行回归分析,结果如表 2 至表 5 所示。

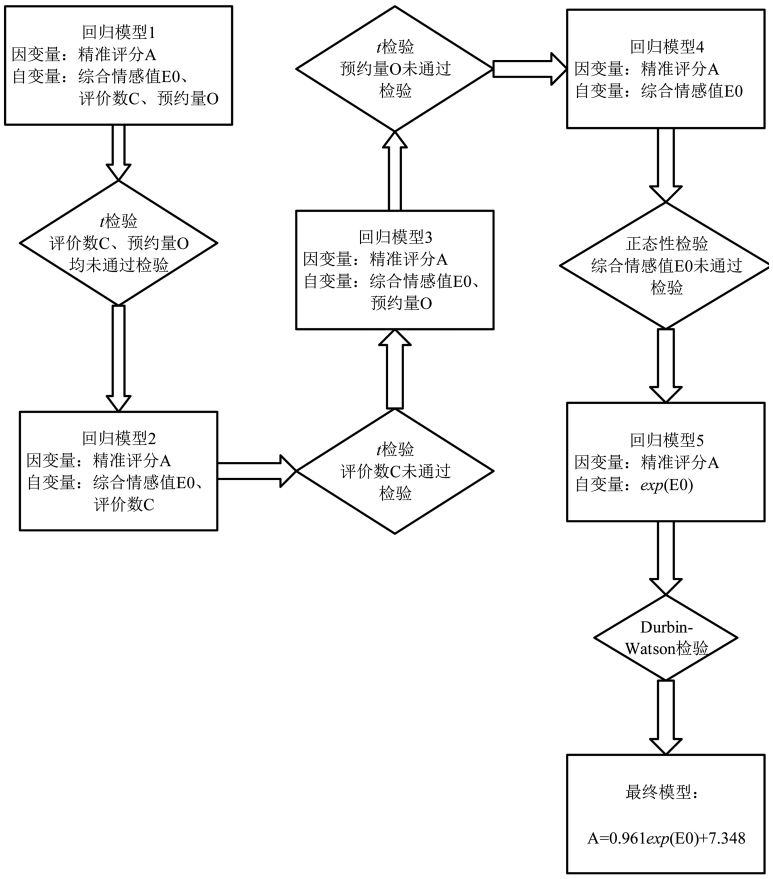


图 2　基于线性回归的文本评论自动打分模型调整过程

表 2　基于线性回归的文本评论自动打分模型参数

参数	数值
R 值	0.717
R^2 值	0.514
调整 R^2 值	0.511
标准估计的误差	0.2052379
Durbin-Watson 值	1.878

表 3　基于线性回归的文本评论自动打分模型整体回归方程的假设检验

变异来源	平方和	自由度	均方	*F* 值	*P* 值
回归	7.130	1	7.130	169.263	0.000
残差	6.740	160	0.042		
总计	13.869	161			

表 4　基于线性回归的文本评论自动打分模型系数

变量	相关系数	标准误差	*t* 值	*P* 值
常量	7.348	0.151	48.508	0.000
exp(E0)	0.961	0.074	13.010	0.000

表 5　基于线性回归的文本评论自动打分模型正态性检验结果

变量	Kolmogorov-Smirnov 检验			Shapiro-Wilk 检验		
	统计量	自由度	*P* 值	统计量	自由度	*P* 值
精准评分	0.047	162	0.200	0.994	162	0.750
exp(E0)	0.052	162	0.200	0.989	162	0.272

此时,模型的显著性水平和 *exp*(E0)回归系数的显著性水平都小于 0.001,故最终纳入模型的变量

为 *exp*(E0)。由此得到回归模型为:A = 0.961*exp*(E0)+7.348。式中,A 为精准评分,E0 为最佳综合情感值。由表 2 可知,本模型的最终 R^2 值为0.514,调整 R^2 的值为 0.511,即 51.1%,Durbin-Watson 检验值为 1.878,按照 α=0.05、样本量 N=162、自变量 k=1,查 Durbin-Watson(DW)检验界值表后得到对应的 DW 检验临界值 dL 和 dU 分别位于[1.720, 1.75]和[1.747, 1.779]范围内。因为 dU≤DW≤4-dU,故认为模型通过 DW 检验,残差不存在自相关。由表 5 可知,自变量 *exp*(E0)的 P 值为 0.272,大于 0.05,说明该变量服从正态分布。

上一步研究获得了基于用户评论文本的自动评分模型,现利用余下的数据对模型进行验证。将训练数据中对应的 E0 值代入上一步所获得的模型中,获得 162 个对应的 A 值,经计算的理论值与实际值的差值位于[-0.8161, 0.5752]区间内,将此区间作为回归方程计算结果的可接受偏差范围。随后将测试数据中对应的 E0 值代入上一步所获得的模型中,获得个 31 个对应的 A 值,测试数据计算的理论值与实际值的差值位于[-0.4843, 0.5665]区间内,包含在预期的偏差范围内,说明回归方程通过了验证测试。

4 讨论

4.1 星级评分仅能在一定程度上代表其文本评论

相关分析结果显示,精准评分和综合情感值的相关性并不算特别高,也就是说星级评分并不能完全代表其文本评论,主要有 3 个方面原因。一是由于每个人的教育背景、知识储备或爱好等情况不同,加之人的评价通常具有主观性[15],因此两个人不同的感受也可能打相同的分数。二是用户可能未准确表达自己的真实想法。一方面,如有学者在研究中提到的,社交网站中人们发布信息的活动具有积极性的倾向,进而导致实际表达出来的评论与真实想法不一致[16];另一方面,由于医学专业性比较强,非医学专业的用户可能会很难通过文字完整准确地表达其真实想法。三是评论质量不够高。由于缺乏有效监控机制,网络上的用户评价良莠混杂,大量低质量甚至不实的信息夹杂其中[17],部分用户评论随意,在一定程度上导致了文本评论与星级评论的相关度不高。

4.2 情感强度的增强或减弱均会降低一致性

本文的研究结果显示,各种权值下星级评分与文本评论的相关性并不完全相同,这一结果与学者马松岳、许鑫[2]的研究结果基本一致。无论是情感强度增强时(即各强度情感“一般:中度:高度”的权重比由“1:1:1”升高至“1:3:5”的过程),还是情感强度减弱时,文本评论与星级评分的相关性都减弱,仅当情感强度的比值为“一般:中度:高度=1:1:1”时,文本评论与星级评分具有最高的相关性。

其原因可能是当某一医院的总体评价情感为中性时,其星级评价也更接近于评分体系的平均值,而计算综合情感值的时候,本文将中性情感的权重记为 0。如果将不同情感强度的权重比向偏离“1:1:1”的方向调整,会使不同的情感倾向之间有较大区分度,因此降低了星级评分与文本评论的相关性[2]。

根据分析结果,当情感强度“一般”的权重小于“高度”时的相关性大于“一般”的权重大于“高度”时的相关性,其原因可能是极端情感评论所占比例较小。而越靠近中性情感,对应评论所占比例越大,因此“一般”的权重对结果的偏离影响较大,“一般”的权重大于“高度”时相关性更小。

4.3 自变量对变异的解释力较低

在打分模型的构建过程中,经过多次修改模型,发现仅纳入综合情感值一个自变量,回归方程的调整 R^2 值仅为 0.511。而在马松岳、许鑫[2]等关于电影评分的研究中,其对电影的打分模型纳入了综合情绪值和评论数两个自变量,且调整 R^2 值较为理想。经过调查发现,部分综合实力较强的医院目前并未在该平台开通预约挂号,或曾开通但现在已经关闭,这可能是导致预约量和评论量无法纳入模型的原因之一。但与马松岳、许鑫[2]的研究相比,本文的测试结果相对来说较为理想,说明该模型仍旧具有一定的应用价值。今后的研究中将纳入更多变量,以进一步提高打分的准确性。

5 结论与展望

通过分析基于用户评论情感分析的回归模型,发现星级评分可以代表文本评论。而在用情感值计算综合情感值时,当不同情感强度的权重比值为

“1∶1∶1”时,星级评分与文本评论具有最高相关性。最终,研究得出的回归模型通过检验,可以利用用户反馈的文本评论对医院进行评分,说明该模型具有一定的应用价值。

本文的研究虽然取得了一些成果,但仍有待进一步改进。一是在相关性分析中,虽然结果显示显著性水平较高,但相关性并不算特别理想。本文对此进行了猜想和解释,希望以后可以通过增加样本量等方法,使获得的结果更接近真实情况,同时可以对猜想进行验证。二是在建立回归模型时,回归方程仅纳入一个自变量,回归方程的调整 R^2 值仅为 0.511。在今后的研究中将纳入更多的变量,以提高模型预测的准确性。三是由于用户评论的质量难以控制,容易出现研究结果与事实偏离的情况。因此在进行用户评价研究时,需要考虑多种因素,如评价数据的来源、商家声誉等因素,力图获取质量较高、能在最大程度上代表用户真实观点的评价,从而使结果更贴近事实。四是本文选取的情感分析工具 ROST EA是基于情感词词典进行情感分析的,具有较大的局限性。以后可以考虑采用效率和准确度更高的机器学习方法进行情感分析,以获得更准确的结果和结论。

【参考文献】

[1] 第 45 次《中国互联网络发展状况统计报告》[EB/OL]. (2020-04-28)[2020-07-20]. http://www.cac.gov.cn/2020-04/27/c_1589535470378587.htm.

[2] 马松岳,许　鑫. 基于评论情感分析的用户在线评价研究:以豆瓣网电影为例[J]. 图书情报工作,2016,60(10):95-102.

[3] 崔永生. 在线评论文本挖掘对电商的影响研究[J]. 中国商论,2018(33):17-23.

[4] 贺佳玮. 基于图模型的文本内容挖掘技术[D]. 北京:北京邮电大学,2017.

[5] 刘福珍. 基于文本挖掘的用户反馈服务质量及影响研究[D]. 武汉:武汉大学,2019.

[6] 毕达天,楚启环,曹　冉. 基于文本挖掘的消费者差评意愿的影响因素研究[J]. 情报理论与实践,2020,43(10):137-143.

[7] Guo G Q, Wang X F, Zeng Y. A study on the impact of external incentives on p-e WOM [C]// Marketing Science Innovations and Economic Development. Australia: Orient Acad Forum, 2009:40.

[8] Pavlou P A, Dimoka A, The nature and role of feedback text comments in online marketplaces: Implications for trust building, price premiums, and seller differentiation[J]. Information Systems Research, 2006, 17(4):392-414.

[9] 龚艳萍,张晓丹,张　琴. 行为视角下的网络口碑国外研究综述与展望[J]. 情报杂志,2016,35(5):161-166,160.

[10] 孟　园,王洪伟,王　伟. 网络口碑对产品销量的影响:基于细粒度的情感分析方法[J]. 管理评论,2017,29(1):144-154.

[11] 殷国鹏,刘雯雯,祝　珊. 网络社区在线评论有用性影响模型研究:基于信息采纳与社会网络视角[J]. 图书情报工作,2012,56(16):140-147.

[12] 张艳丰,李贺,彭丽徽,等. 基于情感隶属度模糊推理的网络口碑舆情监测评估实证研究:以手机品牌在线评论数据为例[J]. 图书情报工作,2016,60(23):119-127.

[13] 张　林,钱冠群,樊卫国,等. 轻型评论的情感分析研究[J]. 软件学报,2014,25(12):2790-2807.

[14] 肖　洁. 基于文本挖掘的在线评论情感强度分析[D]. 广州:暨南大学,2018.

[15] 李　慧. 一种适用于信誉系统的主观性消除方法[C]// 工程和商业管理国际学术会议文集. 上海:美国科研出版社,2012:3235-3238.

[16] 王水珍,马红宇. 社交网站情感表露的积极性偏差研究[J]. 华中科技大学学报:社会科学版,2019,33(3):123-134.

[17] 聂　卉. 基于内容分析的用户评论质量的评价与预测[J]. 图书情报工作,2014,58(13):83-89.

[收稿日期:2020-07-23]

[本文编辑:黄思敏]