

基于 LSTM 和 Word2Vec 的情感分析在商品评论中的应用

周 萌

(上海对外经贸大学, 上海 201620)

摘 要: 我国电子商务高速发展,由此产生了大量浏览、购买、评论记录。商品评论的情感分析对于深刻理解商家和用户、挖掘用户的情感等方面有着重要的价值,在提升用户的满意度、改进产品设计等方面有着重要的应用。本文用python爬虫抓取京东商城冰箱商品评论,对于冰箱进行典型意见分析,利用深度学习方法word2vec和LSTM建立二分类器,对抓取的冰箱评论进行分类,模型的准确率能达到96.7%。分类器可以用于评估商品的满意度水平,对于改进电商服务和产品设计具有重要意义。

关键词: 情感分析; word2vec; LSTM; 用户满意度

中图分类号: TP391.1

文献标识码: A

文章编号: 1674-537X (2019) 12.0081-04

DOI: 10.16722/j.issn.1674-537x.2019.12.017

一、引言

电子商务高速发展,2018年上半年,国内网络零售交易规模占社会零售总额的22%。由此产生了大量浏览、购买、评论记录。情感分析^[1]是自然语言处理的一种应用,能提取文本情感内容,进行文本分类,可以帮助企业分析消费者对产品的反馈,检测在线评论中的差评,极大提升人们对于大量文本的理解效率,帮助改进服务和产品,提升用户满意度。

近年来中文情感分析主要应用有:谢丽星等^[2]基于支持向量机的层次结构多策略对微博进行情感分析;李实等^[3]基于特征挖掘算法和支持向量机对手机评论挖掘特征和情感倾向;肖璐等^[4]对各品牌手机评论进行情感分析,用于企业竞争对手研究;魏慧玲^[5]基于情感词典对手机评论进行情感分类;黄仁等^[6]用word2vec构建情感词典对手机评论进行情感分析;赵明等^[7]用word2vec和LSTM对饮食文本分类;张英^[8]用LSTM对微博做情感分析。大多数用于商品评论情感分析的模型具有良好的表现,而在后续分析文本的具体特征以及提出改进意见方面有所欠缺,因此本文的主要工作是利用性能优秀的模型对商品评论进行情感分析,分析各产品评论特征,提出建议,以提升用户满意度。

二、情感分析概述

(一) 情感分析的基本方法

在情感分析领域,文本可以划分为积极和消极两类,或者积极、消极和中性(或不相干)的多类。分析方法主要分为:(1)基于词典的分析;(2)基于机器学习的分析;(3)基于深度学习的分析。

(二) 基于词典的分析

基于词典的分析运用了由预标记词汇组成的字典^[9]。用词法分析器将文本转化为单词序列,把每个单词和字典里的单词匹配。如果有一个积极匹配,将分数加到输入文本分数的总池中,如果有一个消极匹配,总分减少。Minqing Hu等^[10]将形容

词编制成情感词典,根据形容词的极性判断评论的极性,利用此方法能达到84.2%的准确率。Ding Xiaowen等^[11]考虑情感词与主题词的距离,在亚马逊数据上达到92%的准确度。比较成熟的开源情感词典有:(1)GI(the general inquirer)^[12],对每个情感词都给出词性,情感极性,以及客观指数等;(2)LIWC(linguistic inquiry and word count)^[13]给出了情感词类别和对应正则表达式;(3)MPQA(multi-perspective question answering)^[14]有2718个正面情感词以及4912个负面情感词,包括情感极性、情感强度、词个数、词性以及时态;(4)Opinion Lexicon^[10]有2006个正面情感词和4783个负面情感词,包含其标准形式、俚语、拼写错误及社交媒体标记等形式;(5)SentiWordNet^[15]对WordNet中的词汇进行情感分类并给出量化分数。随着字典大小的增加,该方法的性能(时间复杂度和精度)会迅速下降。

(三) 基于机器学习的分析

机器学习技术在情感分析中主要使用监督学习的方法,分为三个阶段:数据收集和预处理、特征构造、训练分类,训练数据为有标记语料库,使用一系列特征向量对目标数据进行分类。Pang等^[16]首先提出利用标准机器学习方法来处理情感分类问题,对电影评论进行情感分类,支持向量机与Unigrams特征结合可达82.9%准确率。Dave等^[17]把Unigrams特征和Bigrams特征进行比较,发现使用Unigrams特征在情感分类中准确率更高。Mullen等^[18]发现混合特征的分类方法Hybrid SVM(Turney/Osgood and Lemmas)的准确率比其他特征组合的方法更高。Zhang等人^[19]使用基于二值的Bigram特征和朴素贝叶斯能实现95.67%的准确率,使用基于频率的Bigram特征和支持向量机可实现94.83%的准确率。相比于基于词典的分析方法,机器学习技术在词典大小呈指数增长的时候仍然工作得很好。

(四) 基于深度学习的分析

2006年以来,Geoffrey Hinton^[20]在神经网络领域取得了重

大突破。应用于情感分类的深度学习方法首先从语料中学习出语义词向量,然后通过语义合成词向量对句子文档进行特征表达。^[21] Bessalov 等^[22]用 n-gram 特征进行线性组合得到文档的特征向量。Glorot 等^[23]利用除噪堆叠自编码器处理情感分类中的领域适应问题,实现跨领域情感分类。Socher 等^[24]基于递归神经网络(RecNN)分类模型来处理情感分类问题。Kim^[25]使用卷积神经网络(CNN)进行情感分类,其分类性能比递归神经网络更优。Kalchbrenner 等^[26]提出一种新的卷积神经网络模型,结合动态 k-max 池化。循环神经网络(RNN)在自然语言处理有成熟应用,如 Zhu 等^[27]用长短期记忆网络(LSTM)来处理情感分类问题,将语句转化为词序列进行情感分类。

三、研究方法

本文主要工作包括:京东在线评论的抓取和数据预处理、Word2vec 词向量模型训练、基于 LSTM 的商品评论情感分析。本文以冰箱为例。

(一) word2vec模型

word2vec 是一款将词表征为实数向量的工具,采用 CBOW 或 Skip-Gram 模型。CBOW 模型由背景词预测中心词, Skip-Gram 模型由中心词预测背景词。

word2vec 可以把对文本内容的处理简化为 K 维空间的向量运算,向量空间上的相似度可以表示文本语义的相似度。CBOW 模型训练速度快,本文选用其进行训练。

CBOW 的背景词很多,背景词向量取平均, $v_i \in \mathbb{R}^d$ 和 $u_i \in \mathbb{R}^d$ 为索引为 i 的词的背景词和中心词向量,设中心词 w_i 在词典中索引为 t,背景词 w_{o1}, \dots, w_{o2m} , 给定背景词生成中心词的条件概率为:

$$P(w_i | w_{o1}, \dots, w_{o2m}) = \frac{\exp\left(\frac{1}{2m} u_i^T (v_{o1} + \dots + v_{o2m})\right)}{\sum_{i \in V} \exp\left(\frac{1}{2m} u_i^T (v_{o1} + \dots + v_{o2m})\right)}$$

似然函数为 $\prod_{i=1}^T P(w_i | w_{o1}, \dots, w_{o2m})$, 训练中最大化似然函数。

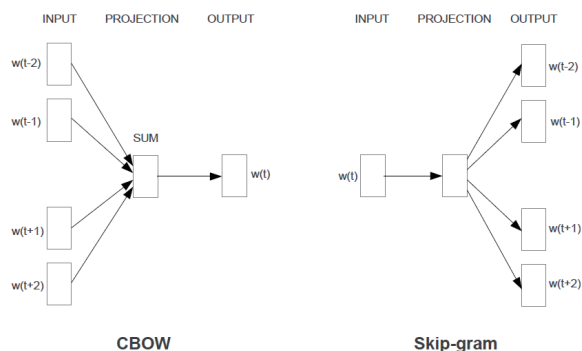


图1: CBOW与Skip-gram模型

(二) LSTM网络

循环神经网络(RNN)能够通过原有神经网络基础上

增加记忆单元,对历史有记忆功能,理论上可以处理任意长度的序列,比一般神经网络更加能处理序列相关的问题。但实际中无法记忆时间跨度大的信息。RNN 变体 LSTM 模型可以学习到长期依赖。LSTM 通过门的结构选择性让信息通过。 x_t 表示不同时间点的输入序列, h_t 为每个时间点的输出, c_t 为记忆单元。

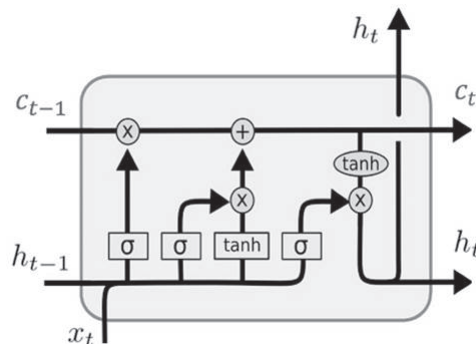


图2: LSTM 单元

$$i_t = \sigma(w^{(i)} x_t + U^{(i)} h_{t-1}) \quad (\text{Input gate})$$

$$f_t = \sigma(w^{(f)} x_t + U^{(f)} h_{t-1}) \quad (\text{Forget gate})$$

$$o_t = \sigma(w^{(o)} x_t + U^{(o)} h_{t-1}) \quad (\text{Output / Exposure gate})$$

$$\tilde{c}_t = \tanh(w^{(c)} x_t + U^{(c)} h_{t-1}) \quad (\text{New memory cell})$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \quad (\text{Final memory cell})$$

$$h_t = o_t \cdot \tanh(c_t)$$

LSTM 的第一步要决定从单元中抛弃哪些信息。这个决定由叫做遗忘门(Forget gate)的 sigmoid 层决定,以 h_{t-1} 和 x_t 为输入,输出的 f_t 介于 0 和 1 之间的数,1 代表完全保留,0 代表完全遗忘。下一步将旧的 c_{t-1} 更新到 c_t , f_t 与 c_{t-1} 相乘更新记忆单元,输入门(Input gate)的输出 i_t 乘 \tilde{c}_t , 表示对最新 c_t 的更新程度,随后更新 c_t 。最终基于当前信息决定输出 h_t 。

四、实验验证与结果分析

(一) 数据采集与处理

利用 python 的 urllib 库和 json 库抓取和解析京东商城各主流冰箱的用户评论,汇总数据如下:

表1: 评论及数量

评论	数量	评论预览
积极	10180条	冰箱很大,可以放很多东西。西门子品牌值得信赖。送货师傅很靠谱,还帮我拆门,帮我放好。
消极	9600条	冰箱噪音太大,从来没有见过有这么大噪音的冰箱,无法忍受,退货很不顺畅。

在对文本进行分析前,需要对语料进行分词、去除停用词的操作。本文用 Jieba 加载中文停用词表,进行分词。

(二) word2vec词向量模型

本文基于 CBOW 训练 word2vec 词向量模型,词向量维度为 300,训练窗口为 7。词向量可以以余弦距离或欧式距离表示语义相似度,很好与满意的余弦距离为 0.593,愉快与满意

的余弦距离为 0.603，语义相似距离比较接近。以下是词向量的可视化及降维后的词向量。



图3：词向量的可视化

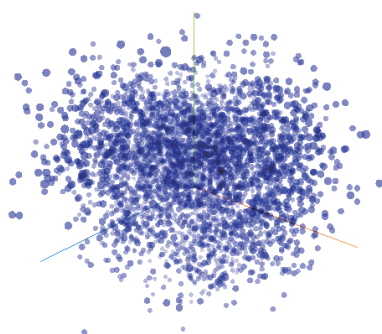


图4：经过PCA降维后的词向量

(三) 基于LSTM的情感分析

数据集 80% 划分为训练集，20% 为测试集。分别用不同的文本表示方法和分类器进行实验，二分类器的评价指标有精确率 (Precision)、召回率 (Recall)、F1 (精确率、召回率的调合平均数)。SVM 用 RBF 核构造判别函数，使特征空间的间隔最大。构造的 LSTM 有一层 50 个 LSTM 神经元，一层有 2 个神经元的全连接层进行分类。可以看出用 word2vec 表示与 bag-of-words 表示、tf-idf 表示相比，分类器的表现更好。LSTM 与传统 SVM 相比，性能有提升。

表2：不同文本表示方法的模型表现

方法	Precision	Recall	F1
bag-of-words+SVM	0.9300	0.9000	0.9148
tf-idf+SVM	0.9446	0.9198	0.9320
Word2vec+SVM	0.9501	0.9597	0.9549

表3：不同分类方法的模型表现

方法	Precision	Recall	F1
Word2vec+LSTM	0.9629	0.9686	0.9667
Word2vec+SVM	0.9501	0.9597	0.9549

参考文献：

[1]宗成庆.统计自然语言处理[M].北京:清华大学出版社,2007.

因此对于自然语言处理任务，LSTM 表现性能更佳。接下来用 LSTM 对抓取的卡萨帝、西门子、松下、三星冰箱（价位在 6 千~1.2 万）进行情感分析，评估用户的满意度水平。关键词由 tf-idf 算法得到。

表4：各品牌分类后满意度

关键词	西门子	三星	卡萨帝	松下
声音	0.9228	0.8906	0.9794	0.9601
空间	0.9948	0.9852	1.0000	0.9974
外观	0.9827	0.9634	0.9922	0.9801
质量	0.9343	0.8558	0.9689	0.9900
容量	0.9915	1.0000	1.0000	0.9899
保鲜	0.9847	1.0000	0.9437	0.9780
功能	0.9819	0.9412	1.0000	0.9945
价格	0.8939	0.8605	0.9837	0.9108
售后	0.8403	0.5211	0.9085	0.5714

通过情感分析，发现中高价位的冰箱品牌中，用户普遍反应比较满意，对于卡萨帝的满意度相对较高。用户对于容量、外观、空间满意度更高，对于声音、质量、价格、售后满意度较低。再通过分析关键词所在的具体评论，发现不同的品牌评价的异同：1、选择西门子、松下、三星冰箱的用户，是其品牌的忠实消费者，购有其电器产品，对于品牌的形容词，西门子是信赖高端，松下专业，三星信赖，卡萨帝认为其是高端品牌，能信赖。2、在购买前会去实体店，认为网购具有性价比，会确认预留尺寸，安装过程中，可能没有电梯，厨房门小，要拆门才能进入，冰箱尺寸可能与预期有差异。3、冰箱外观容量普遍好评，其中松下压缩机置顶，小体积大容量适合小户型。4、功能方面，用户很注重保鲜，看重杀菌、净味、速冻功能，松下、卡萨帝分区合理，对松下的制冰功能很满意，卡萨帝智能。5、用户会注重声音、质量，对于售后、客服比较不满。因此厂家可以从产品的声音、质量入手，打造用户体验更好的产品；电商可以从售后、价格方面提升服务水平，提高消费者的满意度水平。

五、总结

本文利用爬虫抓取京东平台下的冰箱评论数据，利用深度学习中 word2vec 和 LSTM 的方法建立分类模型，对冰箱的评论进行分类，评估数据清洗后的好评率，二分类器的准确率可达 96.7%。本文的意义在于将深度学习的方法应用于评论的情感分类，相比于传统的机器学习模型，深度学习模型表现出更佳的性能，该分类器可以用于评估商品的满意度水平；评论的情感分析发现冰箱差评集中在产品质量、声音以及价格售后方面，对于改进电商服务和产品设计具有重要意义。

- [2]谢丽星,周明,孙茂松.基于层次结构的多策略中文微博情感分析和特征抽取[J].中文信息学报,2012,26(01):73-83.
- [3]李实,叶强,李一军,罗嗣卿.挖掘中文网络客户评论的产品特征及情感倾向[J].计算机应用研究,2010,27(08):3016-3019.
- [4]肖璐,陈果,刘继云.基于情感分析的企业产品级竞争对手识别研究——以用户评论为数据源[J].图书情报工作,2016,60(01):83-90+97.
- [5]魏慧玲.文本情感分析在产品评论中的应用研究[D].北京交通大学,2014.
- [6]黄仁,张卫.基于word2vec的互联网商品评论情感倾向研究[J].计算机科学,2016,43(S1):387-389.
- [7]赵明,社会芳,董翠翠,陈长松.基于word2vec和LSTM的饮食健康文本分类研究[J].农业机械学报,2017,48(10):202-208.
- [8]张英.基于深度神经网络的微博短文本情感分析研究[D].中原工学院,2017.
- [9]Nasukawa T, Yi J. Sentiment analysis: capturing favorability using natural language processing[C]. international conference on knowledge capture, 2003: 70-77.
- [10]Hu M, Liu B. Mining and summarizing customer reviews[C]. knowledge discovery and data mining, 2004: 168-177.
- [11]Ding Xiaowen, Liu Bing, Yu P S. A holistic lexicon-based approach to opinion mining[C] //Proc of Int Conf on Web Search and Web Data Mining. New York: ACM, 2008: 231-240.
- [12]Stone P J, Dunphy D C, Smith M S. The general inquirer: A computer approach to content analysis[J]. American Journal of Sociology, 1968, 73(5): 375-376.
- [13]Pennebaker J W, Francis M E, Booth R J. Linguistic inquiry and word count 2001[J].
- [14]Lawrence Erlbaum Associates Mahwah Nj, 2001, 10(2): 22-32.
- [15]Wilson T, Wiebe J, Hoffmann P, et al. Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis[C]. empirical methods in natural language processing, 2005: 347-354.
- [16]Pang B, Lee L, Vaithyanathan S, et al. Thumbs up? Sentiment Classification using Machine Learning Techniques[J]. empirical methods in natural language processing, 2002: 79-86.
- [17]Dave K, Lawrence S, Pennock D M, et al. Mining the peanut gallery: opinion extraction and semantic classification of product reviews[J]. the web conference, 2003: 519-528.
- [18]Mullen T, Collier N. Sentiment Analysis using Support Vector Machines with Diverse Information Sources.[C]. empirical methods in natural language processing, 2004: 412-418.
- [19]Ye Q, Zhang Z, Law R, et al. Sentiment classification of online reviews to travel destinations by supervised machine learning approaches[J]. Expert Systems With Applications, 2009, 36(3): 6527-6535.
- [20]Hinton G E, Osindero S, Teh Y W, et al. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554.
- [21]Mitchell, Jeff, and Mirella Lapata. "Composition in distributional models of semantics." Cognitive science 34.8 (2010): 1388-1429.
- [22]Bespalov D, Bai B, Qi Y, et al. Sentiment classification based on supervised latent n-gram analysis[C]. conference on information and knowledge management, 2011: 375-382.
- [23]Glorot X, Bordes A, Bengio Y, et al. Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach[C]. international conference on machine learning, 2011: 513-520.
- [24]Socher R, Huval B, Manning C D, et al. Semantic Compositionality through Recursive Matrix-Vector Spaces[C]. empirical methods in natural language processing, 2012: 1201-1211.
- [25]Kim Y. Convolutional Neural Networks for Sentence Classification[J]. empirical methods in natural language processing, 2014: 1746-1751.
- [26]Kalchbrenner N, Grefenstette E, Blunsom P, et al. A Convolutional Neural Network for Modelling Sentences[J]. meeting of the association for computational linguistics, 2014: 655-665.
- [27]Zhu X, Sobihani P, Guo H, et al. Long Short-Term Memory Over Recursive Structures[C]. international conference on machine learning, 2015: 1604-1612.