

请参阅此出版物的讨论、统计数据和作者简介：<https://www.researchgate.net/publication/344487215>

情绪分析方法、实践和应用综述

研究 · 2020 年 10 月

奖状

1

2 位作者：



沙尼尔·潘迪亚

瑞典林奈大学

93 篇出版物 3,218次引用

查看简介

阅读

18,004



普贾·梅塔

甘地讷格尔理工学院

7 篇出版物 270次引用

查看简介

# 情感分析方法论回顾， 实践与应用

Pooja Mehta, Sharnil Pandya 博士

摘要:情感分析是一种分析文本形式的信息并从文本中确定情感内容的技术。它也被称为情感或情感挖掘。如今, Twitter、Facebook、YouTube 等在线交流渠道对人类生活产生了很大的影响。人们在上面分享他们的想法或感受。在这篇评论文章中,我们倾向于关注情感挖掘或情感分析,这是网络数据挖掘和机器学习的一个领域。本文通过使用不同的 ML 和词典调查方法来展示调查的后果。分析结果以进行评估研究并检查当前作文的估计。通过这种方式,它将帮助未来的研究人员了解可能性分析配置中的当前开端。

索引术语:情感分析、意见、情感、机器学习、准确性、NLP、支持向量机

## 1. 简介

人类的行为很大程度上受到抽象情感和信念的影响,例如举止、感觉、评价或倾向。我们做出的选择可能会在很大程度上受到他人对世界的印象的影响,因为传达他人的评价通常与每个人息息相关,并将我们描绘成“社会生物”。

如今, Twitter、Facebook 和 YouTube 等社交沟通渠道已经非常受欢迎。

观点挖掘是情感分析的另一个名称,属于机器学习和数据挖掘的范畴。

从不同社交媒体的使用来看,观点挖掘或情感分析技术必须从人们的数据开始,用于分析政治、经济或生物等不同领域[2]。每天都会以数字形式记录与不同个体实体相关的大量信息。因此,该领域的快速发展与其他社交媒体相关内容(例如论坛讨论)共存

、博客、客户评论、Twitter 和社交网站。

情绪分析包括将数据分类为各种类别,例如乐观(即良好判断力)或消极(即不良判断力)或中立(即无效)。情感分析是感知给定意见总体上是正面还是负面的任务[3](例如,电影评论、个人、政党或政策或产品功能评论。由于消息和产品的格式自由)由于微博平台易于访问,社交媒体上的大多数数据都是非结构化的[3],当需要做出结论或最终输出时,从不同人的经验和评论中获取包含重要内容的意见非常重要。最近的工作是在促进和具有挑战性的领域进行意见挖掘,以满足个人的愿望[5],需要适当的训练集来进行情感分析以获得更好的性能,并需要准确的数据集来进行不正确的文本分析。为了更好的手段和准确性,语言分析被认为是机器

机器学习。本文的结构如下:下一部分对情感挖掘和观点挖掘的介绍进行解释。从那时起,将显示不同阶段的意见分析。第四节中对之前工作的描述是关于情感分析技术的。

第五节介绍了情感分析来源/资源。第六部分解释了情感分析的挑战。最后,第七节阐述了讨论和结论

## A. 情绪:

情绪可以用不同的方式表达。它可以用各种情绪来表达,通过判断、观点、洞察力或人们的观点来表达。情绪可以表达为一个人根据情况有意识或无意识的突然反应。如果我们以文本格式研究情绪,那么可以从两个不同的角度来看待它。首先,它可以研究对作者的影响他选择词语来表达某些情绪的方式。其次是对读者的影响他如何根据自己的心态和分析事物的能力来解读所写的内容。

## B 情绪分析:

情感分析或意见挖掘是识别和识别或分类用户对任何服务(如电影、产品问题、事件或任何属性是积极、消极或中性)的情绪或意见的过程。该分析的来源是社交传播渠道,即网站,其中包括评论、论坛讨论、博客、微博、Twitter 等。该研究领域如今非常流行,因为它具有观点数据,用户可以在其中找到对任何服务的评论对他们的日常生活有用。大量固执己见的数以数字形式存储。对于特定主题或任何观点,与数据挖掘相关的情感分析起作用并给出输出。对于情感分析,研究工作基于情感的总结、感觉或心灵提取。情感分析也称为意见挖掘,它使用 NLP(自然语言处理)来跟踪公众舆论对任何产品或服务的特定主题的情感和感受。由于情感分析非常出名,它在调查和广告活动中也有很多用途,可以通过人们的意见或建议来获取任何产品或服务成功率。它还提供有关人们喜欢和不喜欢的信息,并且公司对其产品功能有更清晰的了解。情感分析有

· Pooja Mehta 目前正在印度古吉拉特邦苏仁德拉纳格尔 Wadhvan 的 C. U Shah 大学技术与工程学院攻读计算机工程博士学位。

电

子邮件: poojamehta810e@gmail.com

· Sharnil Pandya 博士目前是印度马哈拉施特拉邦浦那市 Symbiosis 大学的副教授,电子邮箱: sharnil.pandya@sitpune.edu.in

有效学习自然语言处理(NLP)和计算语言学是否有主要语言

在政治 [9]、商业 [10] 和营销/销售和广告（用于估计特定产品的销售额）等各个领域，这种观点的接受度大大提高。因此，识别句子的类型是意见挖掘中最重要的部分。我们必须将句子分为主观的或客观的。

最近或现有的研究正在使用监督和无监督学习技术来提供不同的

情绪分析的几个目的。在初始阶段  
研究采用以下全部或组合的监督技术。

- 1.支持向量机
  - 2. 最大熵
  - 3.朴素贝叶斯
- 初步研究使用的无监督技术包括 :1. 利用情感词典
- 2. 语法分析
  - 3. 句法模式

一般来说，“情绪分析包括高级过程。  
该分析有一系列完全不同的任务：情感分类（监督或无监督）、主观或客观分析以及意见提取。对于任何文本文档或句子，可以通过主题级别分析来评估主观或客观的标签。情感分类包括探测过滤后句子的情感极性的任务”。所有的句子都分为中性、消极或积极类型，这取决于我们从评论中得到的情绪。

1.主观/客观分类  
情感分析（SA）的主要任务之一是主观和客观分类。主要目的是将整个文档或句子分为两类之一；客观或主观[11]。在意见挖掘中，文本中的句子被标记为主观或客观。通过使用情感分析，我们可以提取主观句子。客观句子产生的事实信息需要去除。当它包含情感或感觉时，主观句子对于情感分析过程很重要。主观句子由用户对句子级别的以下看法、观点、想法、评论和意见组成。

示例- 1)主  
观：《盗梦空间》是一部很棒的电影。  
（该句子带有情感（棒了），因此是主观的）

2.)客观:莱昂纳多是《盗梦空间》的男主角。（这句话中没有感情,它是一个实际的事实,因此它是客观的)[13]。

对于主观句子,各种研究正在进行中。对于任何组织来说,客观线条都会被消除,而在研究中会考虑主观句子,因为它包含观点和情感。英语主要用于主观分类研究,但其他各种研究也在多种语言中进行。阿拉伯语和乌尔都语是其中之一,使用不同的监督学习分类方法,支持向量机（SVM）是其中之一-主观性和情感分析。

C分类级别  
分类级别有多种类型:1)文档级别 2)句子级别 3)属性/方面级别

1)文档级:这是仅基于文档的意见挖掘或情感分析的第一级。在这个特定的层面上,我们考虑整个文档并找出极性。通过这个级别或在这个级别的帮助下,我们可以对可用的意见或情绪是否为我们提供积极情绪或消极情绪进行分类[4]。考虑到这一点,文档应该是单一主题。本文件要考虑的主要来源是情绪或情感。例如,在一篇文本中,文件仅包含单个产品的评论,现在系统开始计算整个评论是否表达了对该产品的总体正面或负面意见。因此,对于许多产品来说,这种级别的审查是无效的。主要也是唯一的优点是我们可以得到特定特征的大部分极性,而这个级别的缺点是人们的喜欢和不喜欢并不能通过这个来获得。

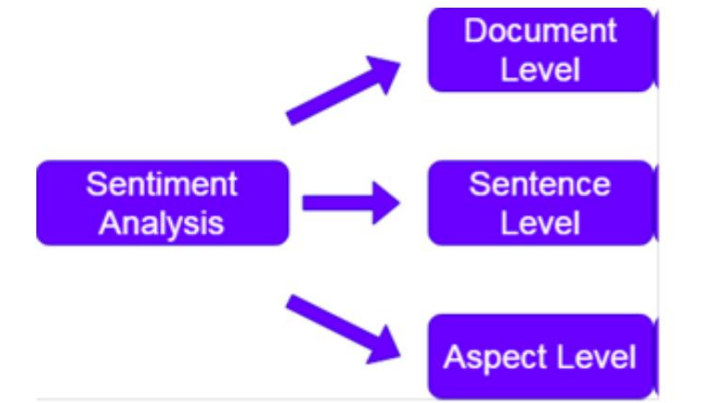


图1情感分析的种类

2)句子级别:情绪分析的另一个类别是句子级别,其中处理和分析确定每个句子的极性,从而对该句子给出积极、消极或中性的意见。

主观句子由用户对句子的看法、观点和意见组成。当句子不暗示任何意见时,句子是中性的。当句子是中性的时,它更有可能被称为客观句子,它在主观分类下提供事实信息,称为主观句子,它显示主观观点和意见。主观性和客观性分类是句子分析的主要优势。一般来说,主观句子是在机器学习中检测到的。但是,在句子层面,情绪分析存在局限性

3)方面/特征级别:方面级别的另一个名称也是特征级别或实体级别。我们可以在文档和句子级别的分析中找到人们的好恶。最后,将输出表达为意见的方面级别的输出是最详细的情绪分析类型。两个结果被认为是积极或消极的,以及目标值。要找出对实体及其特征的情绪,目标意见有助于认识到这一级别的重要性。评论、反馈、评论和投诉等都在此级别进行。

2. 应用

- A. 决策支持：  
建立一个可以进行决策的网站是非常关键的部分。分析有其自身的优点；它可以引发不同的想法，帮助我们在日常生活中做出决定，例如选择一家好的餐厅去吃晚饭，或者购买一辆新车或选择一部好电影来观看等。
- B. 业务相关申请：  
因为市场每天都在变化，合作世界中的竞争也随之加剧。每个人都想创造一种创新且最新的产品，以充分满足客户的需求。为了提高产品的价值，组织可以收集用户的所有需求，并通过从客户那里收集的反馈来提高产品的效率。
- C. 预测和趋势分析：  
通过情绪审查跟踪公众的观点，使任何人都可以预测市场情况，这有助于任何人进行交易和民意调查。通过利用所有这些观点，用户可以预测市场趋势。

3. 材料和方法  
方法

- 有多种方法可用于意见挖掘，但主要使用两类。 SA 的问题将由第一组通过实施机器学习方法来解决。第二组使用基于词典的方法，这是一种语言学倾向的方法。在这两个群体中，都存在许多技术。通过以下方式，我们可以提取文本或句子的特征。
- 1) N-Gram:一次只能取一个单词（一元语法）或两个单词（二元语法），最多得出 n 个单词。
- 有些观点无法捕获一元语法特征。例如，这本书很有趣。如果仅在一元模型中将其组合在一起是令人着迷的并且是负面的，那么这是一个乐观的评论。
- 2)词性标注： 是词语在内容（语料库）中标记它的方式，它根据词语的定义和内涵与其词性相联系。
- 名词、代词、形容词、副词等是不同词性的例子”。
- 3) 词干提取 其中，消除前缀和后缀是主要过程。例如，“running”、“sleeping”、“ran”分别来自“run”和“sleep”。它基本上有助于编目，但有时也会导致编目准确性下降。
- 4) 停用词 停用词是代词（他/她、它）、冠词（a、an、the）、介词（above、in、near、under、besides）。
- 这些话只是提供很少或根本没有提供任何关于

- 的情绪。在互联网上，您可以访问停用词列表。在预处理步骤中，可以使用它来去除它们。
- 5) 连词处理 一般来说，每个句子一次只有一个意思。但有一些可用的连词，如 But.And.while、although.然而，改变了句子的整个含义。例如，尽管骑行很好，但没有达到我的期望。通过使用这些规则，吞吐量可以提高 5% [6]。
- 6) 否定处理 像“not”这样的否定词颠倒了整个句子的主旨。例如，这部电影并不“好”，它是乐观的，但“不”将分裂转为消极。
- 识别情绪或观点词是观点挖掘中许多应用中的重要任务。根据给定的特征，对极性进行分类是一项基本的重要任务。积极、消极和中性是极性分类的三个类别。从极性识别中，可以使用词典技术计算情绪强度、情绪分数等。“观点挖掘有多种方法和技术，主要使用两类。1) 使用词典方法和 2) 解决 SA 问题的机器学习方法。

- 1)基于词典的方法:在当前的方法中,当对给定的文本使用可用的词典技术时,将分离单词。一般来说,它通过分数的聚合来执行:例如主观单词分数为积极、消极和中性等分别总结为相同的。它为每个单词分配一个分数。最后生成四个分数。获得最高分数的给出了文本的整体分割[10]。主要分为两部分:a)基于词典;b)基于语料库。
- A)基于词典的方法 - 在这个系统中,用户收集一组情感词并准备种子列表。之后,用户开始搜索短语手册和词典来查找特定文本的同义词和反义词。完成后,新创建的替代词将添加到种子列表中。直到用户找不到新词,该过程才会继续。
- 缺点:很难找到上下文或面向领域的情感词。
- B) 基于语料库的方法 - 语料库基本上是一个术语,它是一组文字,就像一些文字的组,通常涉及非常精确的问题。在这种情况下,用户使用语料库文本的帮助来绘制种子列表这是有组织的情况[9]。

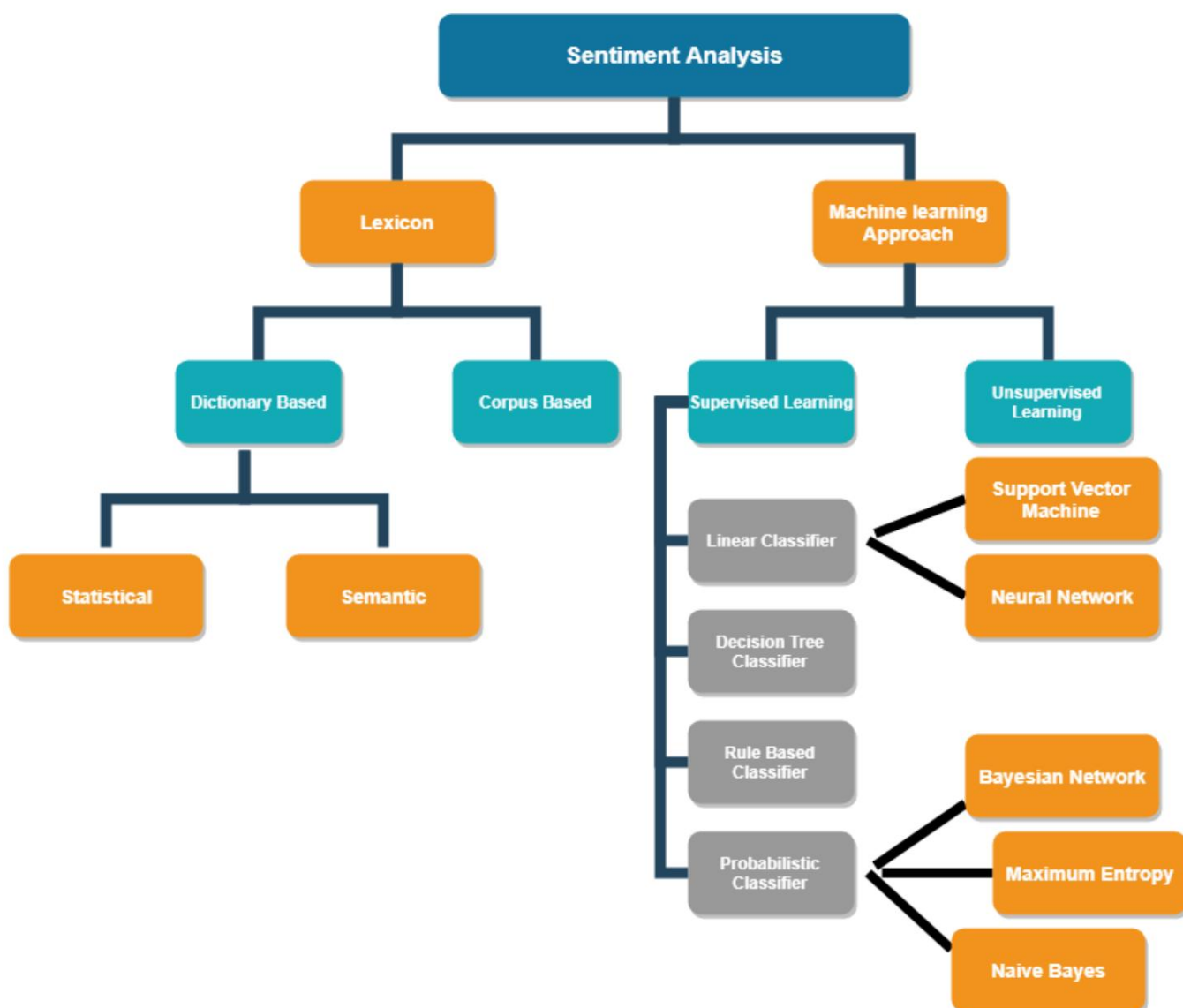


图 2 情绪分析的各种方法

$$P(\text{标签/特征}) = P(\text{标签}) * P(\text{特征/标签}) / P(\text{特征})$$

## 2) 机器学习方法:

在这种方法中,最初通过对文档进行两种不同的组合来执行分类。训练数据和测试数据是其中的一部分。这被称为非自愿分类。从特征中提取进一步的文本,并将其分类为 I) 监督和 II) 无监督。

a) 监督系统 - 在各种数据集中,标记训练数据集是用于监督系统的数据集之一。每种类型的类别都有自己的属性和优势,并有与之相关的标签,可用于该系统。每个单词到达后都会根据其类型和与之相关的特征归类到标签下。

### 1. 概率分类器:预测或预期概率

与不同模块之间输入记录相关的功能。

a) 朴素贝叶斯:在此,仅使用文本文档作为输入,借助贝叶斯定理来生成一组可能性,以提供一组属性属于一个特定标签的预测。BOW - Bag of Words 是一种使用机器学习方法提取文本的方法,简单且易于实现。该现有模型使得这些所有功能都被赋予自主性。[20]

b) 贝叶斯网络:它用于表示不同特征之间的关系。它可以与非循环图进行比较,其中节点表示随机变量,边表示依赖关系。该模型非常昂贵,因此很少使用。

c) 最大熵:通过编码,使用分类器将标记的特征集转换为向量。该向量被转换并用于决定这些特征的权重,这些权重可用于假设和预测每个特征集的标签。

2. 线性分类器:线性分类的特点可以通过使用该分类器来实现,该分类器用于显示预测器作为结果,并可分为两类:

A) 支持向量机 (SVM):这种学习模型在监督下用于分类。该特定模型最重要的目的是确保这是分类的最佳线性分离器。这将使用 SVM 训练创建一个模型,该模型将新信息分为一类或两类。

B) 神经网络 (NN):它是具有神经元电子网络的大脑神经结构。在这个网络中,神经元

国际科技研究杂志第9卷第02期,2020年2月

ISSN 2277-8616

是基本组成部分。神经元分为三部分 :输入、隐藏和输出。

3.决策树分类器 :为了划分数据,有一个条件可用。一个类由那些缓和条件的数据组成,另一个类由其余数据组成。这种技术称为递归技术,它有两个部分 :单属性分割和多属性分割。

4. 基于规则的分类器 它是基于条件的分类器,它使用条件或规则,如 IF、THEN。它可以写成

IF 条件 THEN 决策》

我们可以根据训练阶段的要求制定规则[2]。

4. 情感分析资源

收集数据是情绪分析的主要目的,其中使用社交沟通渠道,如 Twitter、Facebook 或任何预先存在的资源。

A) 博客和论坛 :它是观点和情感的来源,我们可以从中获取用于研究的信息,研究人员可以通过网络论坛和博客使用所有这些信息。“通常,只设计单一主题的论坛 ;因此,通过使用论坛,我们可以确保在单一领域进行情感挖掘。此外,博主主们在他们所在地区、国家和世界各地开展活动后,每天都会更新他们的博客和评论,这是一种趋势。

B)评论 :有许多可用的研究仅专注于评论,因为它们对意见和情绪的可用性。在任何研究中,研究人员主要研究电影和产品评论,其主要目的是从情绪和意见中获得反馈。

C)新闻文章 :新闻文章,如金融文章和政治评论,是情绪分析的流行来源[51] 》。新闻文章文本的主要格式是结构化的和正式的。

D) 社交网络 :我们可以从许多社交网站上获取意见和评论进行情感分析,例如 Twitter、Facebook 等。

· 推特:

推文是不同用户发布的消息,限制为 140 个字符。用户可以阅读彼此的消息 (称为推文) 。提供此功能的微博客服务称为 Twitter。通过使用这些推文,我们可以将其作为未来模式的意见和评论,从而生成民意调查结果。

· 脸书:

提供发布个人资料、照片、视频

以及其他相关信息由最著名的社交网络工具 Facebook 提供,Facebook 在 2004 年推出后就广受欢迎。

因此,这些以用户消息形式提供的大量信息,依赖于此消息背后情绪的计算机技术被称为情绪分析。

5.相关工作

学者们已经做了很多研究来分析情绪或观点。有许多方法用于提取数据。

各种方法和手段的比较

在[22]和[24]的研究工作中使用前馈神经网络来:

1) 识别经常表达感受、表达观点和发推文的在线用户。 2) 为了根据正面和负面关键词将这些推文划分为不同的类别,为此他们还使用了 Twitter API。

卷积神经网络是用于情感分析的方法之一。通过结合情感分析和形态句子模式模型,我们可以获得许多好的结果。其他技术,如分词、词干提取和预处理、自组织映射 (SOM) 和递归神经网络也可用于情感分析。在[21]中,所提出的系统表明,在数据预处理后,它会分类为句子级别,然后提取数据的特征。然后应用连贯性解析和 SentiWordNet 的使用。采用SVM机器学习方法来统计产品特征的准确性。最后,我们找到有关产品功能的总体情绪或准确性。在[23]中,研究工作表明使用不同的机器学习方法来提取情感。它利用不同的twitter数据进行预处理、主观分类和特征分类。最后,它计算了所有机器学习方法 (如 SVM、朴素贝叶斯和 ME)的准确性。在[25]中,提出了适用于人工智能的不同方法。提出模型适用于VADER方法,这与SVM或ME等传统方法不同。通过使用 VADER 方法 (即 效价感知词典和情绪推理器) ,将意见分为正面、负面和中性。用户情绪按 a 分为乐观、负面和无偏见。结果显示三个人工智能的排名最高。 SA 的最佳资源是审核数据。在[26]中,研究人员使用自然语言处理和计算语言学技术对酒店数据评论的情绪进行分类。这一结果体现了游客满意度、安全性、舒适性、豪华性和住宿服务的结果。它将帮助酒店管理者了解顾客的需求,发现进一步发展的领域并提高服务质量。 NLP 技术用于评论数据,与文本数据处理一起进行情感分析。在这里,研究人员在他们的工作中使用了基于情感极性的模型 (SPBM) 。它使用朴素贝叶斯方法的多项式算法,与其他分类算法相比,该算法给出了良好的预测结果。在意见挖掘中,可以使用电子商务和新闻类型数据集。在[27]和[28]研究中,他们从亚马逊和BBC (他们提出作品的在线新闻频道)获取了数据集。在处理所有数据集时,它分别从特征评论和文章中划分出目标的正面和负面文本,并执行不同的分析方法,例如清理数据以进行分析的预处理。首先,目标是计算文本数据的极性,无论是正数还是负数。朴素贝叶斯和支持向量机方法用于寻找数据的准确性和精度,这是监督学习方法。

已知和已定义的单词的集合称为情感词典。情感分析有两种类型 :1)基于词典和 2)机器学习。极性转变是任何方面或特征层面的主要关注点。诸如 [29][31] 之类的研究就是为了发现这种极性转变。在他们的研究中,词袋将文本数据处理为不同单词的向量。并使用各种 ML 技术对这些词进行分类。



该模型还主要缺乏极性转移问题。PSDEE-极性转移检测、消除和集成技术用于解决文档级情感分析中的极性转移问题。然后,在对数据进行预处理后,使用朴素贝叶斯和 SVM 等 ML 技术来确定情感极性(积极、消极或神经)。Twitter 站点是微博客,因为它的数据库不是结构化格式的。这些数据由不同的用户以他们的感受或日常生活的形式共享。Twitter 数据适合数据流挖掘,因为数据或消息较小且连续。[30][32][33][36][41],对 Twitter 数据进行研究,从短文本中分析情绪。文本数据可以分为正面或负面意见。一般或特定的项目特征称为情感分类。不同人对特定产品的看法可以分为积极情绪、消极情绪或中性情绪。有监督的机器学习方法是情感分析最可靠的方法。可用于情感分析的机器学习算法有朴素贝叶斯、最大熵和支持向量机(SVM),用于查找产品特征的准确性。情感分析是机器学习一项非常具有挑战性和重要的任务。在[35]中,提出了阿拉伯语的情感分类工作。他们表示,阿拉伯语推文为意见挖掘研究打开了大门,但由于缺乏情感分析资产或阿拉伯语文本分析困难,这些推文被推迟。阿拉伯语有两个级别可供分类。第一个主观分析器基于监督方法,过滤相关和不相关的评论,第二级情感分析器也基于监督方法和集成技术,将相关评论分类为正面、负面和中立。为了利用不同的权重计划,进行了词干提取和 n-gram 程序测试,结果表明,与朴素贝叶斯分类器相比,通过二元组使用 TF-IDF 的 SVM 分类器更好。朴素贝叶斯

(NB)、支持向量机(SVM)和 Rocchio 分类器都属于此分类。在 [34] 中,他们提出了情绪分析来处理来自相关文档的任何主题,从而给出正面或负面的输出。对于意见检索,主题相关的结构是通过查询依赖的帮助来构建的。为了实现数据检索的排名算法,研究人员使用 SVMRank。为了计算排名的性能,可以使用许多方法。

平均精度 (MAP)作为 TREC 社区使用的评估指标。在[37]中,通过使用朴素贝叶斯分类器,它可以检测英语推文的极性,无论推文是积极的、消极的还是中性的。构建了朴素贝叶斯分类器的两个独特变体:1) 特定基线 和 2)二元 (使用词典方法和组作为正负)

通过这种方法可以识别来自各种来源的多词和价移位者。[38] 讨论了社交媒体网站,如 twitter、Facebook,它们在社交媒体网络中非常有名。他们提出了一个新的框架来从网络数据集中找出意见或情绪的极性。它将这个系统与来自 Twitter 的手动数据结合起来。Twitter API 用于收集数据。它将数据分析为积极、消极和中性。为此使用了 Naive bayes 方法的子方法 Unigram Naive Bayes。在 [39] 中,他们提出了一个监督情绪分类框架,该框架基于来自 Twitter 的数据来查找数据的准确性。对于 Twitter 客户端的特征,它在推文中包含单词中的哈希标签、n-gram,然后将它们合并为一个单独的元素向量以进行情绪排序。K 最近邻算法用于通过为准备和测试集中的每个模型构建一个元素向量来分配情绪名称。

表1.0情感分析技术对比研究

序号	年份	论文标题 使用的方法 准确性	查看数据集	
1	(2017)	使用 SVM 和共指解析的基于特征的情感分析方法[21]	SVM 和共同参考分辨率	训练数据集 产品审核73.6%
2	(2015)	情感神经网络 推特分析[22]	带 Feed 的神经网络 前向法	Twitter 数据集74.15%
3	(2017)	Twitter 情绪研究 使用机器进行分析 学习 算法 在 蟒蛇[23]	朴素贝叶斯, 支持向量机 最大熵	Twitter 数据集 86.4% 73.5% 88.97%
4	(2018 年)	情绪分析使用 神经网络:一种新的 方法[24]	卷积神经网络	产品数据审核 推特数据74.15% 64.69%
5	(2018 年)	Twitter 的情感分析 与人工相关的语料库 情报助理[25]	价意识词典和 情绪推理机 (VADER)	电子产品评论87.4%
6	(2018)	带有酒店评论意见挖掘的情感分析框架[26]	朴素贝叶斯	酒店点评 观点排名83.5%
7	(2018 年)	方面级别 电子商务分析 数据[27]	朴素贝叶斯 支持向量机	亚马逊 顾客 审查数据90.423% 83.43%
8	(2017)	文档层面的情绪 分析自 消息	机器学习方法	BBC新闻数据集57.7%

		文章[28]			
9	(2017 年)	极性 转移 检测 情绪方法 分析:一项调查[29]	基于词典和监督 基于机器学习	产品审核	84.6%
10	(2017 年)	一种微博短文本文情感分析方法[30]	语言技术平台 (LTP)用于依赖句法分析	COAE2014(BBC 数据集	86.5%
11	(2016 年)	SEMEval-2016 任务 4 情绪 分析 进 推特[32]	支持向量机	Twitter 数据集	84.5%
12	(2016 年)	基于主题的方法 Twitter 上的情绪分析 数据[33]	支持向量机	Twitter 数据集	74.09%
13	2013年	分类系综 主观性算法和 阿拉伯语情感分析 客户评价[35]	朴素贝叶斯 支持向量机	来自jeeran.com 的阿拉伯语评论 (服 务和产品评论)	97.06% 89.1%
14	(2014)	城市:朴素贝叶斯情感分析策略 英文推文[37]	朴素贝叶斯	训练数据集 by 推文 2014年SEMEVAL	76.54%
15	(2013 年)	社交意见挖掘 媒体数据[38]	朴素贝叶斯	Twitter 数据集	76.8%
16	(2010)	情绪 知识 Twitter 流媒体中的发现 数据[41]	多项式朴素贝叶斯	推特API	82.45%
17 号	(2010)	Twitter 作为语料库 情绪分析和 意见挖掘[42]	SRF	推特数据集	56.4%

在 [40] 中,他们提出了一种三路线显示方法,将情绪分为正面、负面和中性类别。在这种分析中,使用了基于 Unigram 的模型和基于树核的模型。对于基于树核的模型,它使用口头推文作为推文树。基于亮点的模型使用 100 个亮点,而 Unigram 使用超过 10,000 个亮点。基于树核的模型击败了其他两个模型。在参考文献 [41] 中,Firehouse API 提供 Twitter 流数据来获取每个用户的所有消息,这些数据是公开的。使用了随机角度下降、Hoeffding 树和多项式朴素贝叶斯等不同方法来测试这些数据。他们得出结论,当使用基于 SGD 的模型并采用适当的学习率时,它比其他方法更好。微博网站是意见挖掘和情绪分析的丰富数据来源。在 [42] 中,提出了一个模型将用户的消息 (又名推文)分为正面和负面。我们推荐了一些收集海量数据的方法,这些数据可用于情绪分析和意见挖掘。他们使用 Twitter API 收集了大量带有情绪的推文,随后对其进行了评论。他们依靠多项式朴素贝叶斯等技术 (使用 N-gram 和 POS 标签)并使用大量推文创建了一个情绪分类器。由于使用的数据集只有带有表情符号的推文,因此还达不到有效水平。成功的情绪分类基于丰富的数据集,可以为商业、政治或产品特征等提供独特的数据。在 [43] 中,为了区分推文,他们使用了两个模型,一个朴素贝叶斯贝叶斯分类器的性能远胜于最大熵所展示的性能。在 [44] 中,一种远程监督技术 (拥有由带有表情符号的推文组成的数据集)被用于 Twitter 的情绪分析。他们使用的模型利用了 MaxEnt 和支持向量机 (svm) 朴素贝叶斯,整个元素空间由 unigram,bigram 和 POS 组成。在所有这些中,SVM 都表现优异。

在 [45] 中,提出了一种使用两阶段自动情绪分析对推文进行排序的方法。在第一阶段,他们收集推文并将其排列为客观的,然后在第二阶段,抽象的推文被命名为正面的或负面的。使用的元素空间包括与单词的极端和 POS 等亮点相关的转发、标签、链接、重音和感叹号。

6 结论和未来范围

在本文中,我们主要关注情绪/观点挖掘的基础知识及其层次。有多种方法和手段可以从内容中识别情绪。在本文中,我们的研究代表了机器学习程序。

情感分析通过各种分类方法将结果分为积极、消极和中性分数。

研究表明,SVM、朴素贝叶斯和神经网络等机器学习方法具有最高的准确性,可以被视为基线学习方法,并且在某些情况下基于词典的方法非常有效。在未来的工作中,可以发现文本数据的各种其他组合的结果以及其他对预测准确性的影响。未来需要做更多的工作来改进绩效衡量标准。

参考

[1] Surnar, Avinash 和 Sunil Sonawane。 “使用各种方法进行 Twitter 情绪分析的回顾。”IJARCET-VOL 6-ISSUE 5,2017 年。

[2] 埃利亚西克·阿尔帕斯兰·布拉克和埃尔多安·埃尔多安。 “Twitter 上金融界的用户加权情绪分析。”信息技术创新 (IIT),2015 年第 11 届国际会议。 IEEE,2015。

[3] Preslav Nakov,Alan Ritter,Sara Rosenthal,Fabrizio Sebastiani|、Veselin Stoyanov。 “SemEval-2016 任务 4:Twitter 中的情感分析” ,SemEval-2016 论文集,计算语言学协会。

[4] Ahmed, Khaled, Neamat El Tazi 和 Ahmad Hany Hossny。 “社交网络上的情绪分析:



概述。”系统、人与控制论 (SMC),IEEE 国际会议。IEEE,2015 年。

[5] Rasika Wagh,Payal Punde。“使用 Twitter 数据集进行情绪分析调查”第二届电子、通信和航空航天技术国际会议论文集 (ICECA 2018)IEEE Xplore ISBN:978-1-5386-0965-1

[6] Anchal Kathuria,Saurav Upadhyay 博士。” 一篇小说评论各种情感分析技术”国际计算机科学与移动计算杂志,第 6 卷,第 4 期,2017 年 4 月,第 17 页。17-22。

[7] DME-DM Hussein,“情绪分析挑战调查” ,J. King Saud Univ.-Eng. Sci.,第 34 卷,第 4 期,2016 年。

[8] Liu, B. 情感分析:挖掘观点、情绪和情感。剑桥大学出版社,2015 年。

[9] Bilal Saberi,Saidah Saad。“情绪分析或观点挖掘:评论”。国际先进科学工程信息技术杂志,第 7 卷 (2017) ,ISSN:2088-5334。

[10] J. Bollen,H. Mao 和 X. Zeng “Twitter 情绪预测股市”。计算科学杂志,2(1):1-8 2011。

[11] 徐涛,彭强,程勇。 “使用 S-HAL 识别术语的语义方向进行情感分析”。基于知识的系统,35 279-289,2012

[12] TT Dang,NTX Huong,AC Le 和 VN Huynh。“越南语主观性分类的自动学习模式”。知识与系统工程。Springer,第 629-640 页,2015 年。

[13]阿罗拉,皮尤什。 “印地语情感分析。”Diss.海得拉巴国际信息技术研究所,2013 年。

[14] T. Wilson,P. Hoffmann,S. Somasundaran,J. Kessler,J. Wiebe,Y. Choi,C. Cardie,E. Riloff 和 S. Patwardhan。“意见发现者:一种主观性分析系统”。ht/emnlp 交互式演示会议论文集,第 34-35 页。

[15] E. Riloff,J. Wiebe 和 W. Phillips。“利用主观性分类来改进信息提取”。《全国人工智能会议论文集》,第 1106 页。

[16] P.D. Turney,“赞成还是反对?:语义取向在评论无监督分类中的应用”。载于第40届计算语言学协会年会论文集,第417-424页。

[17] Math Alrefai,Hossam Faris,Ibrahim Aljarah。“阿拉伯语情感分析:方法和技术简要调查”。2018

[18] Emma Haddia,Xiaohui Liua,Yong Shib,《文本预处理在情感分析中的作用》,ELSEVIER,Procedia Computer Science 17 (2013)26-32。

[19] Jagdale,Rajkumar S.,Vishal S. Shirsat 和 Sachin N. Deshmukh。“使用开源工具对 Twitter 事件进行情感分析。” (2016 年)。

[20] Kang Hanhoon,Yoo Seong Joon,Han Dongil,“用于餐厅评论情感分析的 Senti-lexicon 和改进的朴素贝叶斯算法” ,Expert Syst Appl,39 :6000–10,2012

[21] Hari Krishna M,Rahamathulla K,Ali Akbar,“使用 SVM 和共指解析进行情感分析的特征方法” ,国际情感分析会议

创新通信技术,JICCT2017。和 计算型

[22] Brett Duncan 和 Yanqing 张,“Twitter 上情感分析的神经网络” ,IEEE 第 14 届认知信息学与认知计算国际会议 (ICICCC 2015)

[23] Monika Negi,Kanika Vishwakarma,Goldi Rawat,Priyanka Badhani,Bhumika Gupta,“使用 Python 机器学习算法进行 Twitter 情绪分析的研究” ,国际计算机应用杂志 (0975 – 8887) 第 165 卷 – 第 9 期,5 月2017年

[24] Shiv Dhar,S.Pednekar,K.Borad,Prof.Ashwini Save,“使用神经网络进行情感分析:一种新方法” ,国际创造性通信和计算技术会议 (ICICCT 2018)

[25] Chae Won Park,Dae Ryong Seo,“与人工智能助理相关的 Twitter 语料库的情感分析” ,第五届工业工程与应用国际会议,2018 年。

[26] Kudakwashe Zvarevashe,Oludayo O. Olugbara,“酒店评论观点挖掘的情感分析框架”,会议信息通信技术与社会 (ICTAS) 2018。

在

[27] Satuluri Vanaja.Meena Belwal,“电子商务数据的方面级情感分析” ,计算应用发明研究国际会议 (ICIRCA 2018) 。

[28] Vishal S. Shirsat,Rajkumar S. Jagdale,SN Deshmukh,“新闻文章的文档级情感分析” ,2017 年国际计算、通信、控制和自动化会议 (ICCUBE)。

[29] Sayali Zirpe,Bela Joglekar,“情绪分析中的极性转变检测方法:一项调查” ,国际发明系统与控制会议,2017 年。

[30] 李杰;邱立荣,“一种微博短文本情感分析方法” ,计算科学与工程国际会议 (CSE)和EEE嵌入式与普适计算国际会议 (EUC)2017。

[31] Erik Cambria,南洋理工大学,“情感计算与情感分析” ,IEEE 智能系统,2016。

[32] Preslav Nakov,Alan Ritter,Sara Rosenthal,Fabrizio Sebastiani,Veselin Stoyanov,“SemEval-2016 任务 4:Twitter 中的情绪分析” ,SemEval-2016 论文集。

[33] Pierre FICAMOS,刘岩,“基于主题的 Twitter 数据情绪分析方法” ,国际先进计算机科学与应用杂志 2016。

[34]罗准臣,Miles Osborne,“一种有效的推特观点检索方法” ,Springer Science+Business Media New York 2013。

[35] Nazlia Omar,Mohammed Albared,Adel Qasem Al-Shabi,Tareq Al-MosImi,“阿拉伯客户评论主观性和情感分析的分类算法集合” ,国际计算技术进步杂志 (IJACT),2013 年。

[36] Neha Upadhyay,Angad Singh 教授,“使用机器学习技术对 Twitter 进行情绪分析”,

国际应用科学与工程技术研究期刊 (IJRASET) ,2016。

[37] Pablo Gamallo,Marcos Garcia,“Citius:英语推文情感分析的朴素贝叶斯策略”,

国际语义评估研讨会 (SemEval 2014) 。

[38] Po-Wei Liang,Bi-Ru Dai,“社交媒体数据意见挖掘”,IEEE 第 14 届移动数据管理国际会议,2013 年。

[39] Dmitry Davidov,Oren Tsur,Ari Rappoport,“使用 Twitter 标签和表情增强情感学习”,Coling 2010,北京,2010 年 8 月。

[40] Apoorv Agarwal Boyi Xie Ilia Vovsha Owen Rambow Rebecca Passonneau,“Twitter 数据的情绪分析”。

[41] Albert Bifet 和 Eibe Frank,“Twitter 流数据中的情感知识发现”,2010 年国际发现科学会议。

[42] Alexander Pak,Patrick Paroubek,“Twitter 作为情绪分析和观点挖掘的语料库”,国际语言资源与评估会议,LREC 2010。

[43] Ravi Parikh 和 Matin Movassate,“使用各种分类技术对用户生成的 Twitter 更新进行情绪分析”,2009 年。

[44] Alec Go,Richa Bhayani,Lei Huang,“使用远程监督的 Twitter 情绪分类”2009。

[45] Luciano Barbosa,冯俊兰,《基于有偏差和噪声数据的 Twitter 鲁棒情感检测》,Coling 2010,北京,2010 年 8 月。

[46]叶倩文,张哲瀚,R. Law. “通过监督机器学习方法对旅游目的地在线评论进行情感分类”. 专家系统及其应用,36(3): 6527-6535,2009.

[47] AB Goldberg 和 X. Zhu,“在星星不多的时候看到星星:基于图的半监督学习进行情绪分类”。《第一届基于图的自然语言处理方法研讨会论文集》,第 45-52 页。

[48] R. Prabowo 和 M. Thelwall. “情感分析:一种组合方法”。信息计量学杂志,3(2):143-157,2009。

[49] J. Liu 和 S. Seneff. “通过解析和释义范式审查情绪评分”。自然语言处理经验方法会议记录:第 1 卷-第 1 卷,第 161-169 页,2009 年。

[50] Y. Chen和J. Xie “在线消费者评论:口碑作为营销传播组合的新元素”。管理科学,54(3): 477-491,2008。

[51] M. Abdul-Mageed,MT Diab 和 M. Korayem “现代标准阿拉伯语的主观性和情感分析”,载于《计算语言学协会第 49 届年会论文集:人类语言技术》短文第 2 卷,2011 年。

[52] S. Yu 和 S. Kak. “利用社交媒体进行预测的调查”. arXiv preprint arXiv:1203.1647, 2012 年。

[53] L. Pan. “中文情感分析”. 布兰迪斯大学, 2012 年。