**基于深度卷积神经网络的情绪观点挖掘**

作者：Soujanya Poria , Erik Cambria , Alexander Gelbukh

绪论：

本文首先介绍了在观点挖掘方面的一个深度学习方法。相位提取是情绪分析的一个子任务，它包括在有主见的文本中识别观点，即:在检测某一产品或服务的特定方面时，持意见者要么是赞扬，要么是抱怨。我们使用了一个7层的卷积神经网络来将每个单词标记为含有观点的句子，作为一个方面或非方面的单词。我们也为同样的目的开发了一套语言模式，并将它们与神经网络相结合。由此产生的集成分类器，再加上一个用于情感分析的模型，使得我们的方法获得了比最先进的方法更精确。

1 介绍：

公众舆论对社会事件,政治运动,公司战略,市场活动和产品的喜好，引起了科学界的兴趣(因为激动人心的开放的挑战)和商业界的兴趣(因为可以有益于营销和金融市场预测)。今天，情绪分析研究在几个不同的场景中都有它的应用。有很多公司，不论是大公司还是小公司，都把重点放在对意见和情感的分析上，作为他们任务的的一部分[1]。

意见挖掘技术可以用来创建和自动维持具有评论和意见的网站，在这些网站上，人们不断地收集意见，这些意见不仅仅局限于产品评论，更是更广泛的话题，比如政治问题和品牌评论。情绪分析作为其他系统的子组件技术也具有的巨大潜力。它可以增强顾客与管理人员的联系和推荐系统。举个例子，允许用户发现哪些特性是客户特别喜欢的，或者排除那些推荐列表中收到明显负面反馈的项目。同样，它也可以用于社交网络中的钓鱼邮件过滤和增强反垃圾邮件系统。商业智能对情绪分析领域的兴趣也是主要因素之一[2]。

在观点挖掘中，提出了不同层次的分析粒度，每个层次都有各自的优缺点[3]。基于方面的意见挖掘[4,5]侧重于关系的层面和文档的极性。一个方面，也称为意见目标，是在给定的文档中表达意见的一个概念。例如，在句子中，“我手机的屏幕非常好，它的分辨率非常高”，因为手机评论包含正极性，即。作者喜欢手机。然而，更具体地说，积极的意见是关于它的屏幕和分辨率;因此，这些概念被称为意见目标或观点。在给定的武断文本中识别各方面的任务称为方面抽取。在基于情感的观点中定义了两种类型的方面:显式方面和隐式方面。显式的方面是在有观点的文档中明确表示意见目标的词。例如，在上面的例子中，在文本中显式地提到了对获取屏幕和解析的意见。相比之下，隐含的方面是一个概念，它代表一个观点，但在文本中没有明确指定。人们可以推断出，“这款相机是圆滑的，而且价格实惠”，含蓄地包含了实体相机的外观和价格的正面的观点。这些相同的方面将在一个等价的句子中明确:“这个相机的外表是光滑的，它的价格是非常便宜的。”

之前的大多数方面的研究都使用了条件随机字段(CRFs)[6,7]或语言模式[4,8]。这两种方法都有自己的局限性:CRF是一个线性模型，因此它需要大量的特性才能正常工作;语言模式需要手工制作，关键在于句子的语法准确性。在本文中，我们通过使用卷积神经网络(CNN)来克服这两种限制，这是一种非线性的监督分类器，它可以更容易地适应数据。以前，[9]使用这样的网络来解决一系列的任务(而不是面向方面的)，在这些任务中，它的表现优于其他最先进的NLP方法。此外，我们使用语言模式来进一步提高方法的性能，尽管在这种情况下，语言模式中会有上述问题来影响框架。

本文首先介绍了深度学习法在相位提取任务中的应用。我们的实验结果表明，深度的CNN比现有的方法更能有效地解决问题。我们还介绍了具体的语言模式，并结合了一种语言模式方法和一种深度学习方法来进行方面的提取任务。

2 相关工作：

从观点提取方面首先由胡和刘[4]研究。他们介绍了显性和隐形的区别。然而，作者只讨论了显性的方面，并使用了一套基于统计观察的规则。Hu和Liu的方法后来得到Popescu和Etzioni[10]和blair - goldensohn等[11]的改进。Popescu和Etzioni[10]假设产品是预先知道的。他们的算法检测一个名词或名词短语是否是一个产品特征，通过在名词短语和产品之间进行点间互信息。

Scaffidi等[12]提出了一种使用语言模型识别产品特征的方法。他们认为产品特性在产品评论中比一般的自然语言文本更频繁。然而，他们的方法似乎具有低的精度，因为检索的方面受到噪声的影响，将这方面的术语提取视为序列标记，并使用CRF。即使在跨域实验中，这些方法在数据集上也表现得非常好[6,7]。

主题建模已被广泛应用于各个方面的执行和分组[13,14]。有两种模型为:pLSA[15]和LDA[16]。两种模型都引入了可观察变量“文档”和“单词”之间的潜在变量“主题”来分析文档的语义主题分布。在主题模型中，每一个文档被表示为一个随机的混合，在潜在的主题上，每个主题的特征是一个不同的词。这种方法在社交媒体分析中越来越流行，比如Twitter上出现的政治话题检测[17]。LDA模型为文档-主题分布定义了一个Dirichlet概率性的生产过程;在每个文档中,一个潜在的方面是选择根据多项分布,控制狄利克雷先前的α。个根据一个方面,由一个多项分布提取来控制另一个狄利克雷先驱β。在现有的作品中这些模型的使用策略是全球方面(如产品的品牌)和具体的方面(如产品[18])的性质,提取关键短语[19],评等多个角度[20]的总结[21]。[22]采用最大熵方法对基于词性标记的开关变量进行训练，并将其用于单独的层面和情感词。

Mcauliffe和Blei[23]将用户反馈添加到LDA，作为与每个文档相关的响应变量。Lu和Zhai[24]pro -提出了半监督模型。DF-LDA 也代表了半监督模式，允许用户设置必须链接和不能链接的约束。一个必须链接的约束意味着两个术语必须在同一个主题中，而一个不能链接的约束意味着两个术语不能在同一个主题中。Poria .[26]在LDA算法中计算单词分布的综合常识计算[27]，从而使基于方面的情绪分析中的语法和语义的转换成为可能。

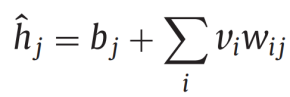
Wang等[28]提出了基于播种方面的两种半监督的产品相位提取模型。在监督方法的范畴中，[29]使用种子词汇指导主题模型，学习用户具体感兴趣的主题，[20]和[30]使用播种词从产品评审中提取相关产品方面。

另一方面，近年来使用深部CNNs的方法[9,31]在自然语言处理(NLP)任务的范围内表现出了显著的性能改进。[9]

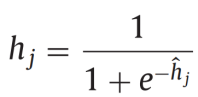
3 关于CNN的背景

一个深层的神经网络(DNN)可以被看作是一个简单的、无监督的模型的合成，比如限制玻尔兹曼机械RBMs)，在这里，每一个RBM的隐藏层都可以作为下一个RBM的可见层。RBM是由两层神经元组成的两部分图:可见的和隐藏的层;同一层神经元之间的连接是不允许的。

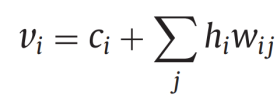
为了训练这种多层系统，需要计算所有层的总能量函数E的梯度。为了学习这些权值，并最大化全局的能量函数，可以使用近似的最大似然反求方法。该方法使用每个训练样本来初始化可见层。其次，利用吉布斯-布林算法对隐层进行更新，然后连续重构可见层，直到收敛发生[32]。作为一个例子，考虑一个逻辑回归模型来学习二次隐藏神经元。每个可见神经元被认为是一个正态分布的三倍[33]。连续状态隐藏的神经元,偏见,加权求和持续可见神经元V。

 （1）

其中是指从可见神经元到隐藏神经元j的连接的重量。隐藏神经元的二进制状态可由激活函数定义:

 (2)

类似地，在下一个迭代中，每个可移植的神经元v的连续状态被重构。在这里，我们用偏压来确定可见神经元的状态，作为一个随机样本，从分布中，均值是所有二进制隐藏神经元的加权总和:

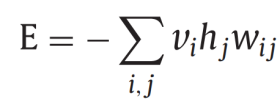
 (3)

其中是连接从可见神经元i到隐藏的j的权重。这连续的状态是一个正态分布的随机样本N(,σ)是所有可见神经元的方差。与隐藏神经元不同，在高斯RBM中可见的神经元可以取连续的值。然后，将权重更新为原始数据v数据与重构可见层之间的差异:

 （4）

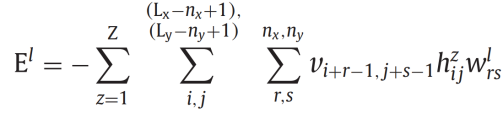
α是的学习速率和的期望频率可见神经元i和隐藏的神经元j活跃在一起,当可见向量样本训练集和隐藏神经元计算根据(1)-(3),再进行k迭代之后。

最后，DNN的能量可以从最后一层(输出层前的一层)确定为:

 （5）

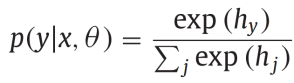
为了将深度神经网络扩展到深度CNN，只需将隐藏层划分为Z组。每个Z组与一个×过滤器,其中是内核的高度和是内核的宽度。假设输入尺寸×,它在我们的案例中是由单词在句子中的特性,嵌入等词,每个词。然后Z的卷积将导致一个隐藏层组,每个维度(−+ 1)×(−+ 1)。

这些内核的学习权值是在一个特定群体中所有隐藏的神经元之间共享的。l层的能量函数现在是对单个块的能量的求和:

 (6)

4 训练CNN的顺序数据

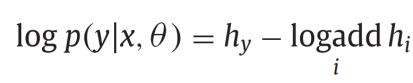
我们使用了一种特殊的训练算法，适用于由胶伯特等[9]提出的序列数据。我们在这里总结一下，主要是下面的[34]。该算法通过反向传播训练神经网络，使训练句子的可能性最大化。考虑到网络参数θ。我们说是输入x的概率的输出分数。然后，将y的标签赋给x的概率计算为

 （7）

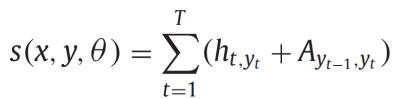
定义logadd操作

 （8）

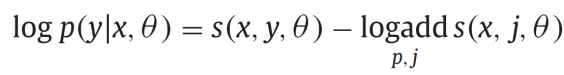
然后，对于一个训练示例，日志就变成了

 （9）

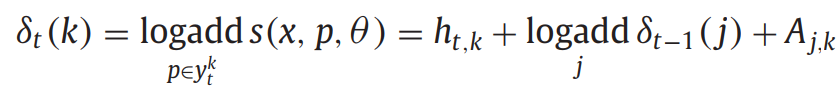
在方面术语抽取中，术语可以被组织成块，并且通常被意见术语包围。因此，为了获得额外的线索，考虑整个句子结构是至关重要的。假设一个句子中有T标记，y是标记序列，而是t-th标签的网络分数，i - th标记。我们引入一个转换分数从移动标记i到标记j。然后，这个句子的得分标签有标签路径y的定义

 （10）

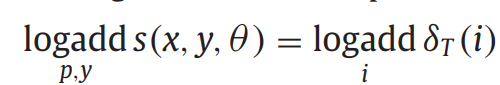
这个公式表示所有可能路径的标签路径概率。现在，从(8)我们可以写出对数的可能性

 （11）

标签路径的数量呈指数增长。然而，使用动态编程技术，可以在多项式中计算在给定的标记[9]中结束所有路径的得分。让表示所有以k为标记t的路径。然后，使用递归，得到

 （12）

为了简洁起见，我们将不会深入探讨在[9]中可以找到的recur - sive程序的细节。下一个方程给出了T的所有路径的对数相加:

 （13）

利用这些方程，我们可以最大限度地提高(11)所有训练对的可能性。对于推理，我们需要找到使用Viterbi算法的最佳标记路径;例:我们需要找到最佳的标签路径来最小化句子的得分(10)。

5网络体系结构：

一个方面的特征取决于它周围的词。因此，我们在一个句子中使用了一个五个单词的窗口，也就是。,±2字。我们形成了那个窗口的局部特征，并认为它们是中间词的特征。然后，将特征向量输入CNN。

该网络包含一个输入层，两个卷积层，两个最大池层，以及一个与软- max输出完全连通的层。第一个卷积层由100个带有过滤大小2的特征图组成。第二个褶积层有50个带滤波尺寸3的未来地图。每个卷积层的步幅是1，我们想要标记每个单词。在每一个卷积层后面都有一个混合池。我们在最大池中使用的池大小为2。我们在倒数第二层采用正则化方法，对权重向量的l2规范进行约束，具有30个时代。利用非线性函数计算了每个卷积层的输出;在我们的例子中，我们使用了双曲正切。

作为特征，我们使用了两个不同的语料库的单词嵌入。我们还使用了一些额外的特性和规则来提高准确性;见第7节。CNN在一个句子中围绕每个单词产生局部特征，然后将这些特征组合成一个全局特征向量。因为内核大小两个con -螺旋层不同,维数×第三节中提到的3×300和2×300,分别。输入层是65×65,其中300是句子中单词的最大数量,和300年的维数嵌入使用这个词,每个词。

这个过程是在一个句子中的每个单词中执行的。像传统的似是而非的学习方法，我们通过在句子中对所有的记号进行卷积后，训练了系统的传播。也就是说，我们将权重、偏差和特性分别存储到每个代币的卷积后，然后只回传播错误，以便在使用第4节中的训练方案处理所有令牌之后纠正它们。

如果一个训练实例有n个词，那么我们就表示该实例的输入向量为，

是一个k维特征向量对“我”这个词。我们发现，这个网络架构在我们的基准数据集上产生了良好的结果。添加额外的层或改变池的大小和窗口大小并不能提高精度，但只会增加计算成本。

6数据集使用

在本节中，我们将介绍我们实验中使用的数据。

6.1字嵌入

单词嵌入是文本的分布式表示，它对单词的语义和句法属性进行编码。通常他们是低维向量。在本节中，我们将描述我们在实验中使用的两个单词嵌入数据集。

6.1.1谷歌嵌入

Mikolov等[35]提出了两种不同的神经网络模型，用于创建单词嵌入。这些模型本质上是对数线性的，对大型语料库进行了训练。其中一个是基于词汇的CBOW模型;它使用单词上下文来获取单词嵌入。另一个叫做skipg模型;它预测了在“当前”这个词的基础上，单词的“嵌入”。这些作者制作了一个名为word2vec pub- licly available的数据集。这些300维的向量通过CBOW archi- tecture从谷歌News上得到一个1000亿字的语料库。

6.1.2亚马逊嵌入

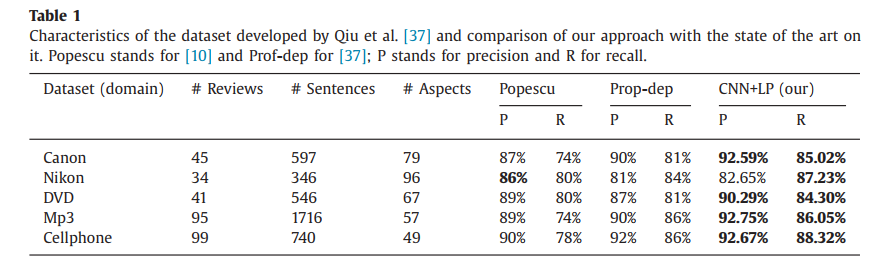
我们训练了由Mikolov等人提出的CBOW架构[35]，由McAuley和Leskovec[36]在一个大型的亚马逊产品评论数据集上进行。该数据集由1995年6月至2013年3月的23441,053个Amazon prod- ucts的34,6867770个评论(47亿个词)组成。我们保留了嵌入消息的300维。该模型可在http://sentic.net/ AmazonWE中找到。邮政编码。由于本模型所使用的文本的性质，这包含了在一般文本中不存在的，如谷歌新闻语料库中的观点/情感信息。

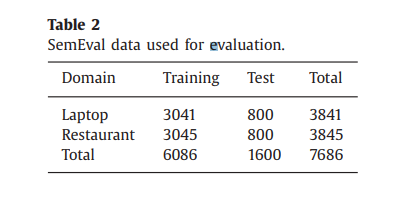
6.2评价集

为了培训和评估拟议的方法，我们使用了两个语料库:

•邱等[37]开发的基于方面的情绪分析数据集[0];参见表1和

•SemEval 2014数据集。数据集由两个领域的培训和测试集组成，包括笔记本和餐厅;见表2。





两个语料库中的注释都是按照IOB2编码的，IOB2是一种用于表示序列的广泛使用的编码方案。在这个编码中，每个块的第一个单词以一个“b - type”标记开始，“i - type”是块的延续，“O”用于标记一个词，该词不属于这个词块。在我们的例子中，我们感兴趣的是判断一个单词或块是否是一个方面，所以我们只有“b - a”、“i - a”和“O”标签。这里有一个IOB2标签的例子:

also/O excellent/O operating/B-A system/I-A ,/O size/B-A and/O weight/B-A for/O optimal/O mobility/B-A excellent/O durability/B- A of/O the/O battery/B-A the/O functions/O provided/O by/O the/O trackpad/B-A is/O unmatched/O by/O any/O other/O brand/O

7 特征与使用规则

这里我们介绍了实验中使用的特征、文本表示和语言规则

7.1我们使用以下特征

•单词嵌入我们使用了在6.1节中描述的嵌入词作为网络的特性。这样，每一个词都被编码成300维的矢量，并被送入网络。

·词性标签大部分方面术语都是名词或名词块。这证明了POS的重要性。我们用这个词的词性标记作为它额外的特征。我们用六个基本的词类(名词，动词，形容词，副词，介词，连词)编码为一个六维的二元向量。我们用斯坦福的Tagger作为一个POS标签。这两个特征向量被连接到CNN。因此，对于每个单词，最终的特征向量是306维。

7.2语言模式

在我们的一些实验中，我们使用了一组语言模式(LPs)来利用SenticNet[38]及其扩展[39,40]，一个基于感知计算的情感分析的概念级知识库[41]。所使用的5个LPs被列在下面。

规则1让一个名词h成为一个单词t的主语，它在一个大的情绪词典中有一个副词或形容词修饰词。然后把h标记为一个方面。

规则2除了当句子有助动词时，如is、was、would、should、could等，我们应用:

规则2.1如果动词t是由形容词或副词修饰的，或在状语从句中与另一个辅币进行修饰关系，那么将h标记为一个方面。例如，在“电池续航时间少”的情况下，电池是持续时间的主题，它是由一个形容词修饰修饰词修饰的，所以电池被标记为一个方面。

规则2.2如果t有一个直接宾语，名词n，在SenticNet中找不到，然后标记n个方面，例如，在“我喜欢这个相机的镜头”。

规则3如果名词h是动词的补充，那么将h作为一个明确的方面。例:“在相机是好”中，相机被标记为一个方面。

规则4:如果一个术语被CNN或其他规则标记为一个方面，则在名词-名词复合关系中与一个-另一个词，则相反，形成一个由这两个词组成的方面术语。例如，如果在“电池寿命”中，“电池”或“生命”被标记为一个方面，那么整个表达式就被标记为一个方面。

规则5以上规则1 - 4通过发现更多方面的术语来提高回忆。但是，为了提高精度，我们采用了一些启发式:例如，我们删除了诸如，a，等等的停止字，即使它们被CNN或其他规则标记为方面的术语。

我们用斯坦福解析器来确定句子中的句法关系。

我们将LPs与CNN结合在一起:LPs和基于cnn的分类器都是在文本上运行;然后，两个分类器中的任何一个都被报告为方面术语，除了那些没有被最后规则标记的术语。

8实验结果：

表1显示，我们的方法比流行的Popescu和Etzioni[10]和基于依赖的传播[37]的方法效果更好[1]，分别为5% - 10%。配对t检验表明，我们所有的改进在95%可信程度上都具有统计学意义。

表4显示了我们的方面术语提取框架的准确性——在笔记本和餐馆领域的工作。该框架提供了更好的准确性，在餐馆领域的评论，因为较低的va - riety可用的术语比在笔记本领域。然而，在这两种情况下，召回率都低于精确度。表4显示了在使用POS特性时的精确性和召回率方面的改进。

预先训练的字嵌入比随机化的特征更好(每个词的矢量初始化随机);见表3。亚马逊的嵌入比谷歌嵌入要好。这支持我们的声明，即前者包含了特定于观点的信息，这有助于它胜过谷歌在更正式的文本(谷歌新闻语料库)上训练的交流。

因此，在序列中，我们只使用亚马逊嵌入来显示性能，我们只是将其表示为我们(单词嵌入)。

在这两个领域，CNN都受到了低召回的困扰。，它漏掉了一些有效的方面。对句法结构的语法结构的分析在很大程度上帮助克服了机器学习分析的一些困难。当语言模式(第7.2节)与CNN一起使用时，我们的实验在精度和回忆方面都有了很好的改善;见表5。

至于语言模式，则最有益的是停止词、Rule 1和Rule 3。图1显示了表5的可视化。

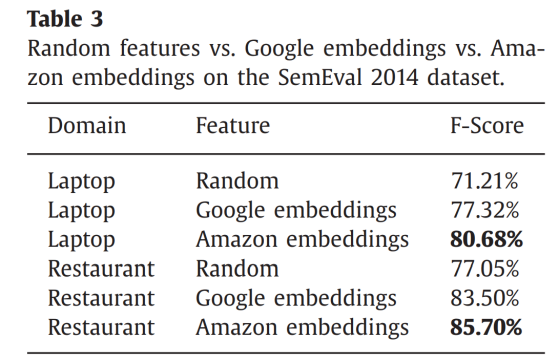
表6和图2显示了所提议的方法和在SemEval数据集上的艺术状态之间的比较。

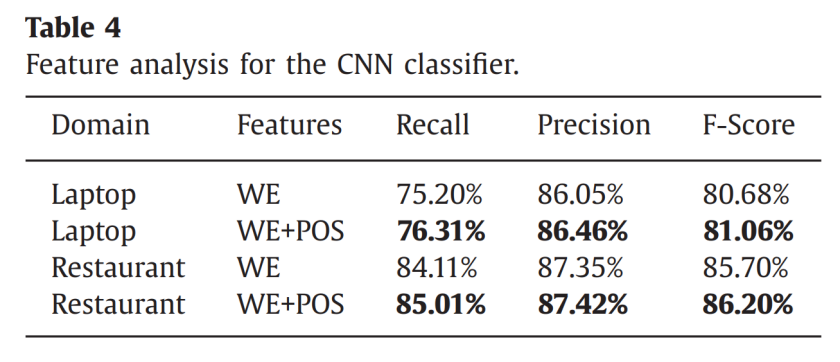
我们可以看到，在笔记本电脑的语料库中出现的大约36.55%的方面术语是短语和餐馆语料库，包括24.56%的方面术语。在这两个领域中，检测方面短语的性能都低于单个单词方面的标记。这表明顺序标记确实是一项艰巨的任务。在这种情况下，缺乏对方面短语的足够训练数据也是导致精度降低的原因之一。特别是，我们分别获得了79.20%和83.55%的f -分值来检测笔记本电脑和餐饮领域的方面。我们观察到一些情况，在一个方面短语中只有一个术语被检测为方面术语。在这些情况下，语言模式的第4条有助于正确地表达方面的短语。我们还对邱等人[37]所开发的数据集进行了实验研究。这是迄今为止最大的基于方面的情感分析数据集。表1左侧显示了该数据集的详细信息。

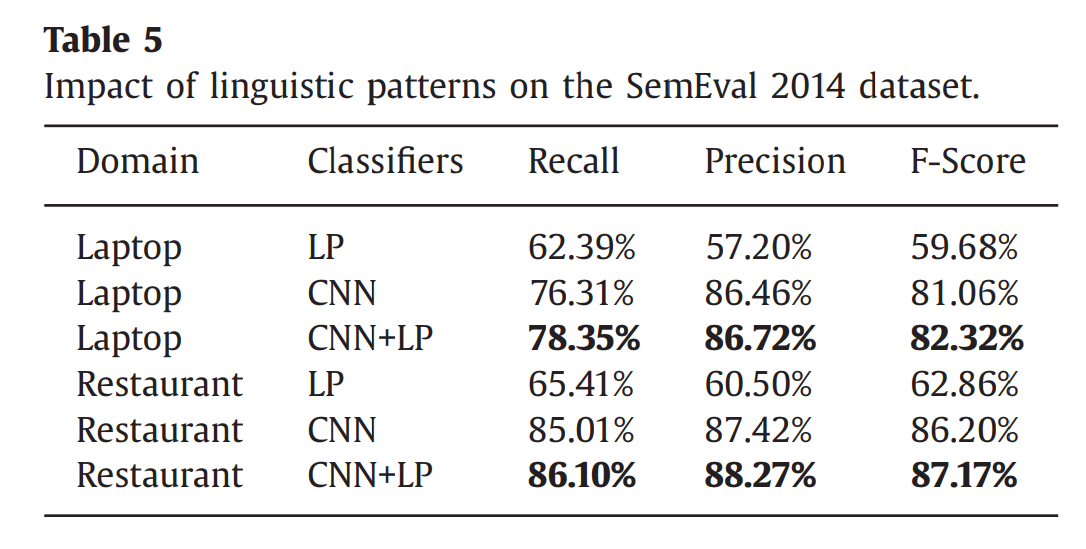
该数据集的最佳准确性是在使用床上用品与POS特性一起使用时获得的。这表明，虽然嵌入特性是最常用的，但是POS特性在方面的提取中也起着重要的作用(表7)。

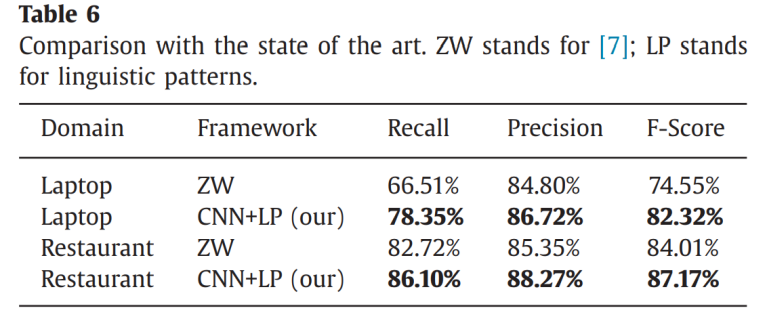
与SemEval数据集一样，语言模式与CNN一起提高了整体的准确性。然而，在这个数据集上，语言模式比在SemEval数据集中表现得要好得多。这支持了前面[37]所做的观察，即在这个数据集上，语言模式更有用。其中一个可能的原因是，这个数据集中的大部分句子都是语法正确的，而且只包含一个方面的术语。在这里，我们将语言模式和CNN结合在一起，取得了比邱等[37]所采用的仅基于语言模式的更好的结果。我们的实验结果表明，这种集成算法(CNN + LP)能比[37]的纯算法更能接受文本的语义，从而提取出更显著的方面。表8和图3展示了不同框架的性能和比较。

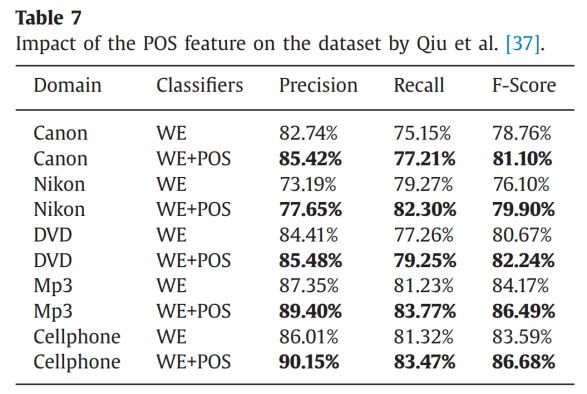
图4将所提议的方法与艺术状态进行比较。我们认为，我们的框架有两个最重要的原因来超越最先进的方法。首先，深度的CNN本质上是非线性的，它比CRF等线性模型更适合于数据。其次，预先训练的单词嵌入特性有助于我们的框架超越那些不使用字嵌入的最先进的方法。我们的框架的主要优点是它不需要任何特性工程。这样可以减少开发成本和时间。

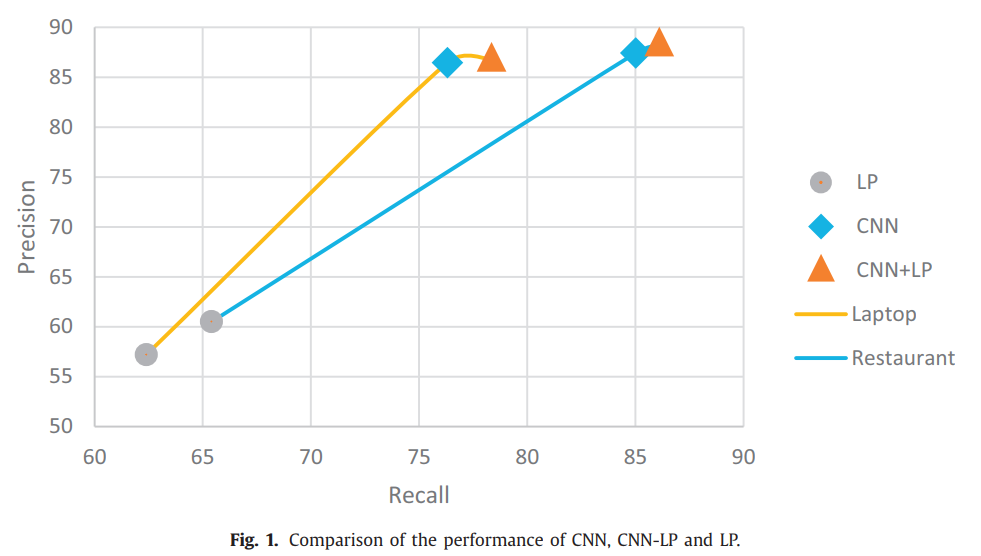


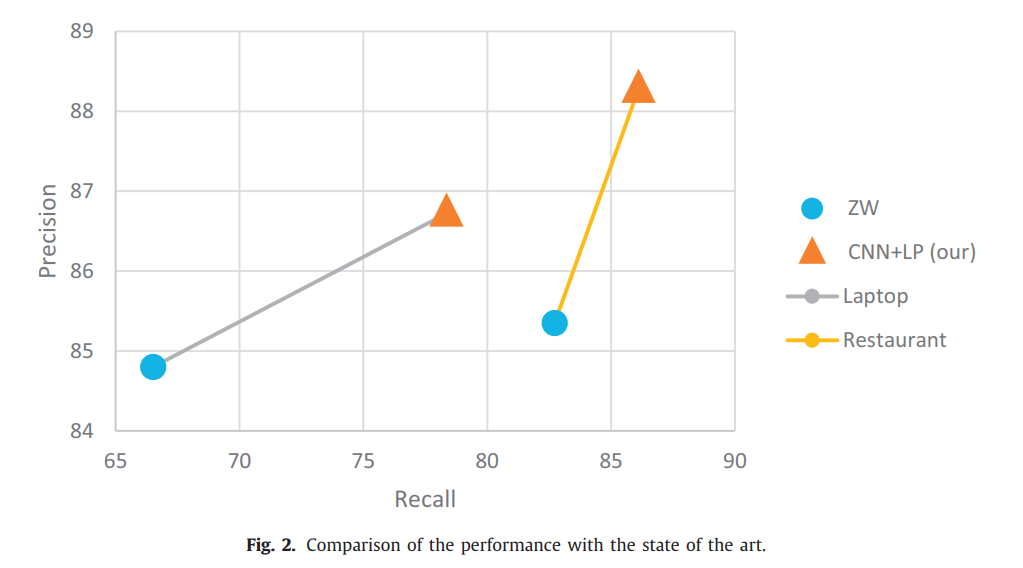


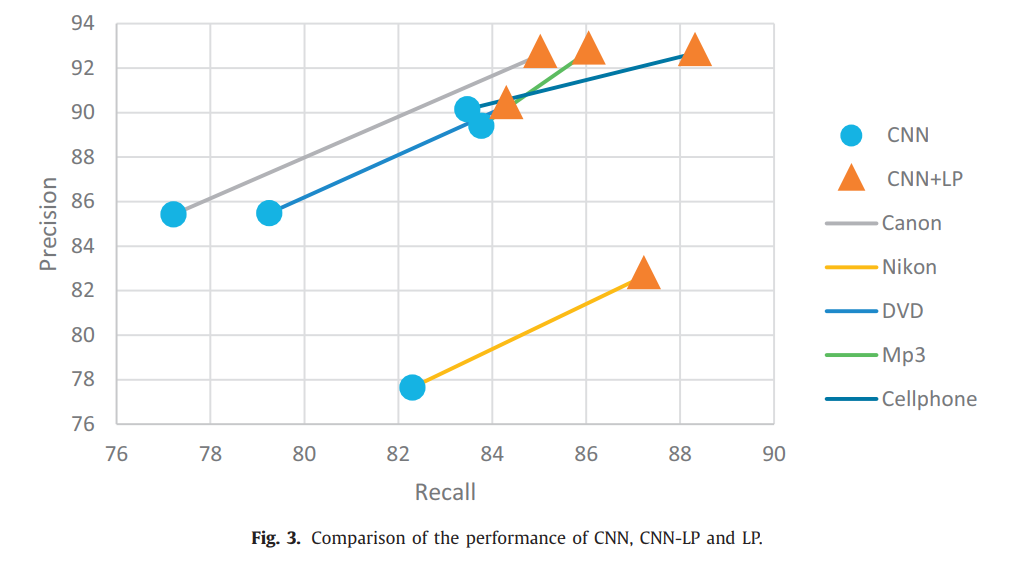
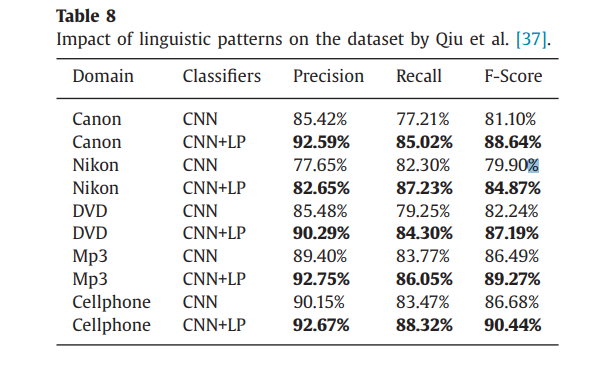


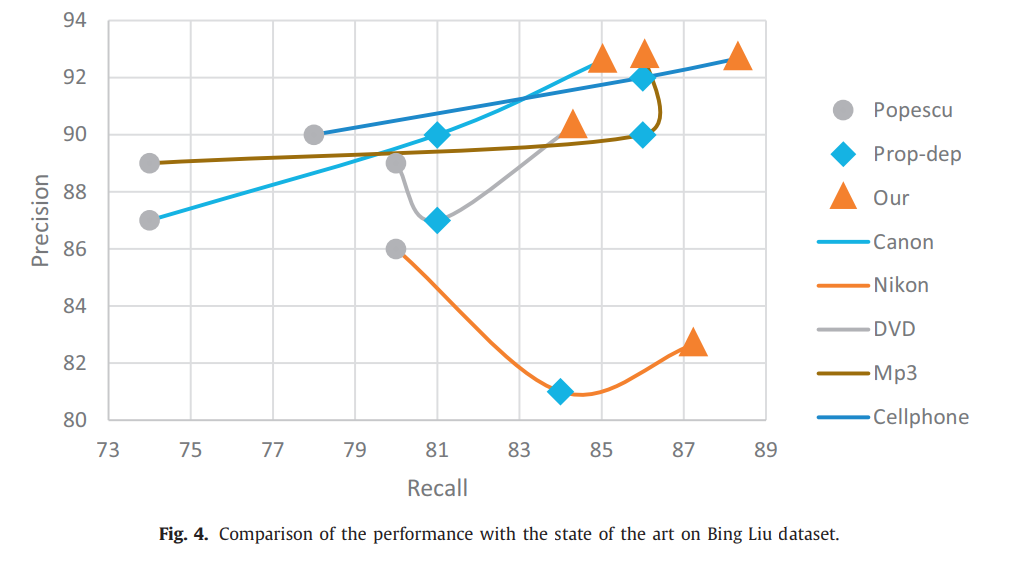












9 总结

我们已经介绍了第一个面向方面提取的基于深度学习的方法。正如预期的那样，这种方法在性能上有了显著的提高，而不是最先进的方法。我们提出了一个特定的深度CNN体系结构，包含7层:输入层，包含每个单词在句子中的每个单词的嵌入特征;两个卷积层，每个fol - lowed由一个max - pooling层;一个完全连接层;输出层，每一个字包含一个神经元。

我们还开发了一套启发式语言模式，并将其与深度学习分类器结合起来。在未来，我们计划扩展和完善这些模式。

References

1. E. Cambria , Affective computing and sentiment analysis, IEEE Intell. Syst. 31 (2) (2016) 102–107 .

[2] E. Cambria , H. Wang , B. White , Guest editorial: Big social data analysis, Knowl.-Based Syst. 69 (2014a) 1–2 .

[3] E. Cambria , B. Schuller , B. Liu , H. Wang , C. Havasi , Statistical approaches to concept-level sentiment analysis, IEEE Intell. Syst. 28 (3) (2013) 6–9 .

[4] M. Hu , B. Liu , Mining and summarizing customer reviews, in: Proceedings of ACM SIGKDD Conference on Knowledge Discovery & Data Mining. Seattle, 2004, pp. 168–177 .

[5] X. Ding , B. Liu , P.S. Yu , A holistic lexicon-based approach to opinion min- ing, in: Proceedings of First ACM International Conference on Web Search and Data Mining (WSDM-2008), Stanford University, Stanford, California, USA, 2008, pp. 231–240 .

[6] N. Jakob , I. Gurevych , Extracting opinion targets in a single- and cross-domain setting with conditional random fields, in: Proceedings of EMNLP-2010, ACL, 2010, pp. 1035–1045 .

[7] T. Zhiqiang , W. Wenting , DLIREC: Aspect term extraction and term polarity classification system, in: Proceedings of the 8th International Workshop on Se- mantic Evaluation (SemEval 2014), 2014, pp. 235–240 .

[8] S. Poria , E. Cambria , A. Gelbukh , F. Bisio , A. Hussain , Sentiment data flow anal- ysis by means of dynamic linguistic patterns, IEEE Comput. Intell. Mag. 10 (4) (2015a) 26–36 .

[9] R. Collobert , J. Weston , L. Bottou , M. Karlen , K. Kavukcuoglu , P. Kuksa ,Natu- ral language processing (almost) from scratch, J. Mach. Learn. Res. 12 (2011) 2493–2537 .

[10] A.-M. Popescu , O. Etzioni , Extracting product features and opinions from re- views, in: Proceedings of EMNLP-2005, 2005, pp. 3–28 .

[11] S. Blair-Goldensohn , K. Hannan , R. McDonald , T. Neylon , G.A. Reis , J. Reynar , Building a sentiment summarizer for local service reviews, in: Proceedings

of WWW-2008 workshop on NLP in the Information Explosion Era, 2008, pp. 14–23 .

1. C. Scaffidi , K. Bierhoff, E. Chang , M. Felker , H. Ng , C. Jin , Red Opal: Produc- t-feature scoring from reviews, in: Proceedings of the 8th ACM Conference on Electronic Commerce, ACM, 2007, pp. 182–191 .
2. Y. Hu , J. Boyd-Graber , B. Satinoff, A. Smith , Interactive topic modeling, Mach. Learn. 95 (3) (2014) 423–469 .
3. Z. Chen , B. Liu , Mining topics in documents: standing on the shoulders of big data, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2014, pp. 1116–1125 .
4. T. Hofmann , Probabilistic latent semantic indexing, in: Proceedings of 22nd ACM SIGIR Conference on Research and Development in Information Retrieval, ACM, 1999, pp. 50–57 .

[16] D.M. Blei , A.Y. Ng , M.I. Jordan , Latent Dirichlet allocation, J. Mach. Learn. Res. 3 (2003) 993–1022 .

[17] S. Rill , D. Reinel , J. Scheidt , R. Zicari , Politwi: Early detection of emerging po- litical topics on twitter and the impact on concept-level sentiment analysis, Knowl.-Based Syst. 69 (2014) 14–23 .

[18] I. Titov , R. McDonald , Modeling online reviews with multi-grain topic mod- els, in: Proceedings of 17th Conference on World Wide Web, ACM, 2008, pp. 111–120 . [19] S. Branavan , H. Chen , J. Eisenstein , R. Barzilay , Learning document-level seman- tic properties from free-text annotations, J. Artif. Intell. Res. 34 (2) (2009) 569 . [20] H. Wang , Y. Lu , C. Zhai , Latent aspect rating analysis on review text data: a rating regression approach, in: Proceedings of the 16th ACM SIGKDD Inter- national Conference on Knowledge Discovery and Data Mining, ACM, 2010, pp. 783–792 .

[21] Y. Lu , C. Zhai , N. Sundaresan , Rated aspect summarization of short comments, in: Proceedings of 18th World Wide Web Conference, ACM, 2009, pp. 131–140 .

[22] W.X. Zhao , J. Jiang , H. Yan , X. Li , Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid, in: Proceedings of the 2010 Conference

Methods in Natural Language Processing, Association for Computational Lin- guistics, 2010, pp. 56–65 .

1. J.D. Mcauliffe , D.M. Blei , Supervised topic models, in: Advances in Neural In- formation Processing Systems, 2008, pp. 121–128 .
2. Y. Lu , C. Zhai , Opinion integration through semi-supervised topic modeling, in: Proceedings of the 17th International Conference on World Wide Web, ACM, 2008, pp. 121–130 .
3. [25] D. Andrzejewski , X. Zhu , M. Craven , Incorporating domain knowledge into topic modeling via Dirichlet forest priors, in: Proceedings of the 26th Annual International Conference on Machine Learning, ACM, 2009, pp. 25–32 .
4. S. Poria , I. Chaturvedi , E. Cambria , F. Bisio , Sentic LDA: Improving on LDA with semantic similarity for aspect-based sentiment analysis, IJCNN, 2016 .
5. E. Cambria , A. Hussain , C. Havasi , C. Eckl , Common sense computing: From the society of mind to digital intuition and beyond, in: J. Fierrez, J. Ortega, A . Esposito, A . Drygajlo, M. Faundez-Zanuy (Eds.), Biometric ID Management and Multimodal Communication, Lecture Notes in Computer Science, 5707, Springer, Berlin Heidelberg, 2009, pp. 252–259 .
6. T. Wang , Y. Cai , H.-f. Leung , R.Y. Lau , Q. Li , H. Min , Product aspect extraction su- pervised with online domain knowledge, Knowl.-Based Syst. 71 (2014) 86–100 .

[29] J. Jagarlamudi , H. DauméIII , R. Udupa , Incorporating lexical priors into topic models, in: Proceedings of the 13th EACL Conference, Association for Compu- tational Linguistics, 2012, pp. 204–213 .

[30] A. Mukherjee , B. Liu , Aspect extraction through semi-supervised modeling, in: Proceedings of 50th Annual Meeting of the ACL: Long Papers, Volume 1, ACL, 2012, pp. 339–348 .

[31] S. Poria , E. Cambria , A. Gelbukh , Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal senti- ment analysis, in: EMNLP, 2015b, pp. 2539–2544 .

[32]G.E. Hinton , Training products of experts by minimizing contrastive divergence, Neural Comput. 14 (8) (2002) 1771–1800 .

[33]G.W. Taylor , G.E. Hinton , S.T. Roweis , Modeling human motion using binary latent variables, in: Advances in Neural Information Processing Systems, 19, MIT Press, 2007, pp. 1345–1352 .

[34] E.R. Fonseca , J.L.G. Rosa , A two-step convolutional neural network approach for semantic role labeling, in: The 2013 International Joint Conference on Neural Networks (IJCNN), IEEE, Dallas, TX, 2013, pp. 1–7 .

[35] T. Mikolov , W.-t. Yih , G. Zweig , Linguistic regularities in continuous space word representations., in: HLT-NAACL, 2013, pp. 746–751 .

[36] J. McAuley , J. Leskovec , Hidden factors and hidden topics: Understanding rating dimensions with review text, in: Proceedings of RecSys’13. Hong Kong, China, 2013 . [37] G. Qiu , B. Liu , J. Bu , C. Chen ,Opinion word expansion and target extraction through double propagation, Comput. Linguist. 37 (1) (2011) 9–27 .

[38] E. Cambria , D. Olsher , D. Rajagopal , SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis, in: AAAI. Quebec City, 2014b, pp. 1515–1521 .

[39] S. Poria , A. Gelbukh , E. Cambria , P. Yang , A. Hussain , T. Durrani , Merging sen- ticnet and wordnet-affect emotion lists for sentiment analysis, in: Signal Pro- cessing (ICSP), 2012 IEEE 11th International Conference on, vol. 2, IEEE, 2012a, pp. 1251–1255 .

[40] S. Poria , A. Gelbukh , E. Cambria , D. Das , S. Bandyopadhyay , Enriching Sentic- Net polarity scores through semi-supervised fuzzy clustering, in: IEEE ICDM. Brussels, 2012b, pp. 709–716 .

[41] E. Cambria , A. Hussain , Sentic Computing: A Common-Sense-Based Framework for Concept-Level Sentiment Analysis, Springer, Cham, Switzerland, 2015 .