

用于情感分析的深度学习:综述

张磊, 领英公司, lzhang32@gmail.com

王帅, 伊利诺伊大学芝加哥分校, shuaiwanghk@gmail.com

刘冰, 伊利诺伊大学芝加哥分校, liub@uic.edu

摘要

深度学习已经成为一种强大的机器学习技术, 它学习数据的多层表示或特征, 并产生最先进的预测结果。随着深度学习在许多其他应用领域的成功, 近年来深度学习也广泛用于情感分析。本文首先对深度学习进行了概述, 然后对深度学习在情感分析中的应用进行了综述。

介绍

情绪分析或意见挖掘是对人们对产品、服务、组织、个人、问题、事件、主题及其属性等实体的意见、情绪、情感、评估和态度的计算研究。¹ 该领域的出现和快速发展与网络社交媒体如评论、论坛讨论、博客、微博、Twitter 和社交网络相吻合, 因为人类历史上第一次有了大量以数字形式记录的意见数据。自 2000 年初以来, 情感分析已经成为自然语言处理中最活跃的研究领域之一。它在数据挖掘、Web 挖掘、文本挖掘和信息检索中也被广泛研究。事实上, 由于它对商业和整个社会的重要性, 它已经从计算机科学扩展到管理科学和社会科学, 如市场营销、金融、政治科学、通信、健康科学, 甚至历史。这种扩散是因为观点是几乎所有人类活动的核心, 是我们行为的关键影响者。我们对现实的信念和感知, 以及我们做出的选择, 在很大程度上取决于他人如何看待和评价这个世界。由于这个原因, 每当我们需要做决定时, 我们经常征求他人的意见。不仅对个人如此, 对组织也是如此。

如今, 如果一个人想购买一种消费品, 他不再局限于向朋友和家人征求意见, 因为在网络上的公共论坛中有许多关于该产品的用户评论和讨论。对于一个组织来说, 可能不再需要进行调查、民意测验和焦点小组来收集公众意见, 因为已经有大量这样的公开信息。近年来, 我们目睹了社交媒体上发表的观点有助于重塑企业, 影响公众情绪和情感, 这对我们的社会和政治制度产生了深远的影响。这样的帖子也动员了群众进行政治变革, 就像 2011 年在一些阿拉伯国家发生的那样。因此, 收集和研究意见就成了一种必要。

然而, 由于各种网站的激增, 在网上寻找和监控意见网站并提取其中包含的信息仍然是一项艰巨的任务。每个网站通常包含大量的意见文本, 这些文本在长篇博客和论坛帖子中并不总是容易解读。普通读者在识别相关网站、提取和总结其中的观点时会有困难。因此需要自动情感分析系统。正因为如此, 有许多初创企业专注于提供情感分析服务。许多大公司也建立了自己的内部能力。这些实际应用和工业利益为情感分析的研究提供了强大的动力。

现有的研究已经为情感分析的各种任务产生了许多技术，包括监督和非监督方法。在监督设置中，早期的论文使用了所有类型的监督机器学习方法(如支持向量机(SVM)，最大熵，朴素贝叶斯等。)和功能组合。无监督的方法包括利用情感词汇、语法分析和句法模式的各种方法。已经出版了几本调查书籍和论文，广泛涵盖了这些早期方法和应用。1, 2, 3

大约十年前，深度学习已经成为一种强大的机器学习技术⁴，并在许多应用领域产生了最先进的成果，从计算机视觉和语音识别到NLP。将深度学习应用于情感分析最近也变得非常流行。本文首先对深度学习进行了概述，然后对基于深度学习的情感分析研究进行了综述。

神经网络

深度学习是人工神经网络(简称神经网络)在使用多层网络的学习任务中的应用。它利用神经网络更多的学习(表示)能力，这种能力曾经被认为只有一两层和少量数据才是实用的。

受生物大脑结构的启发，神经网络由大量分层组织的信息处理单元(称为神经元)组成，这些单元协调工作。它可以通过调整神经元之间的连接权重来学习执行任务(如分类)，类似于生物大脑的学习过程。

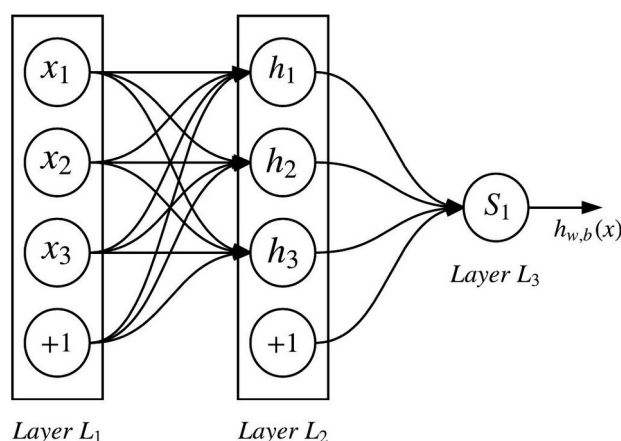


图 1:前馈神经网络

基于网络拓扑，神经网络通常可以分为前馈神经网络和递归/递归神经网络，它们也可以混合和匹配。我们将在后面描述递归神经网络。前馈神经网络的一个简单例子如图 1 所示，它由三层组成： L_1 、 L_2 和 L_3 。 L_1 是输入层，对应于输入向量(x_1 、 x_2 、 x_3)和截距项 $+1$ 。 L_3 是输出层，对应于输出向量(s_1)。 L_2 是隐藏层，其输出作为网络输出不可见。 L_1 的圆圈代表输入向量中的元素，而 L_2 或 L_3 的圆圈代表神经元，即神经网络的基本计算元素。我们也称之为激活功能。两个神经元之间的连线代表信息流的连接。每个连接都与一个权重相关联，权重是控制两个神经元之间信号的值。神经网络的学习是通过调整神经元之间的权重来实现的

信息在其中流动。神经元读取前一层神经元的输出，处理信息，然后向下一层神经元生成输出。如图 1 所示，神经网络根据训练样本 $(x(i), y(i))$ 改变权重在训练过程之后，它将获得符合数据的复杂形式的假设 $h_w, b(x)$ 。

深入到隐藏层，我们可以看到 $L2$ 的每个神经元接受输入 x_1, x_2, x_3 和截距

来自 $L1$ 的 $+1$ ，并输出一个值 $f(W^t x) = f(\sum_{i=1}^3 W_i x_i + b)$ 通过激活函数 f 。 W_i 是

连接的值； b 是截距或偏差； f 通常是非线性的。 f 的常见选择有 sigmoid 函数、双曲正切函数 (tanh) 或校正线性函数 (ReLU)。他们的方程式如下。

$$f(W^t x) = \text{sigmoid}(W^t x) = \frac{1}{1 + \exp(-W^t x)} \quad (1)$$

$$f(W^t x) = \tanh(W^t x) = \frac{e^{W^t x} - e^{-W^t x}}{e^{W^t x} + e^{-W^t x}} \quad (2)$$

$$f(W^t x) = \text{ReLU}(W^t x) = \max(0, W^t x) \quad (3)$$

sigmoid 函数接受一个实数值，并将其压缩为 0 到 1 之间的值。该函数在历史上一直被频繁使用，因为它被很好地解释为神经元的触发率：0 表示不触发，1 表示触发。但是，sigmoid 的非线性最近已经失宠，因为它的激活很容易在 0 或 1 的尾部饱和，在那里梯度几乎为零，信息流将被切断。更重要的是，它的输出不是以零为中心的，这可能在训练中的连接权重的梯度更新中引入不期望的曲折动态。因此，在实践中，通常更倾向于使用双曲正切函数，因为其输出范围以零为中心， $[-1, 1]$ 而不是 $[0, 1]$ 。ReLU 函数最近也变得流行起来。当输入小于 0 时，它的激活简单地被阈值化为零。与 sigmoid 函数和 tanh 函数相比，ReLU 易于计算，在训练中收敛速度快，并且在神经网络中产生相同或更好的性能。⁵

在 $L3$ ，我们可以使用 softmax 函数作为输出神经元，它是将任意实值的 k 维向量 X 压缩为 k 维向量的逻辑函数的推广

$\sigma(X)$ 范围 $(0, 1)$ 内加起来等于 1 的实数值。函数定义如下。

$$\sigma(X)_j = \frac{e^{x_j}}{\sum_{k=1}^K e^{x_k}} \quad \text{for } j = 1, \dots, k \quad (4)$$

通常，softmax 用于神经网络的最后一层，用于前馈神经网络中的最终分类。

通过将所有神经元连接在一起，图 1 中的神经网络具有参数 $(W, b) = (W_1, b_1, W_2, b_2)$ ，其中 $W(l)$ 表示与层中的神经元 j 和层 $l+1$ 中的神经元之间的连接相关联的权重。 $b(l)$ 是与 $l+1$ 层中的神经元 i 相关联的偏差。

为了训练神经网络，通常使用通过反向传播⁶的随机梯度下降来最小化交叉熵损失，交叉熵损失是 softmax 输出的损失函数。首先计算损失函数相对于从最后一个隐藏层到输出层的权重的梯度，然后通过以向后的方式应用链规则递归地计算表达式相对于上层网络层之间的权重的梯度。有了这些渐变，层与层之间的权重会相应地调整。这是一个迭代的细化过程，直到满足特定的停止标准。图 1 中用于训练神经网络的伪代码如下。

训练算法:通过反向传播的随机梯度下降

用随机值初始化神经网络 N 的权重 W 和偏差 b

做

对于每个培训示例 (x_i, y_i)

\hat{p}_i =神经网络预测 $(x_i N)$

计算损失函数 (\hat{p}_i, y_i) 相对于 L_3 层 w_2 的梯度

获取从隐藏层 L_2 到输出层 L_3 的所有权重的 Δw_2 在 L_2 层

通过链式法则计算相对于 w_1 的梯度获取从输入层 L_1 到隐

藏层 L_2 更新 $(1, 2)$ 的所有权重的 Δw_1

直到所有训练样本都被正确分类或者满足其他停止标准

返回训练好的神经网络

表 1:训练图 1 中的神经网络。

上述算法可以扩展到具有多个隐藏层的一般前馈神经网络训练。注意，随机梯度下降估计每个训练示例的参数，而不是批量梯度下降中的整个训练示例集。因此，参数更新具有高方差，并导致损失函数波动到不同的强度，这有助于发现新的和可能更好的局部最小值。

深度学习

研究团体在 20 世纪 90 年代后期对神经网络失去了兴趣，主要是因为它们被认为仅适用于“浅层”神经网络(具有一层或两层的神经网络),因为训练“深层”神经网络(具有更多层的神经网络)是复杂的并且在计算上非常昂贵。然而，在过去的 10 年里，深度学习取得了突破，并在许多应用领域产生了最先进的成果，从计算机视觉开始，然后是语音识别，以及最近的 NLP.^{7,8}。神经网络的复兴可以归因于许多因素。最重要的因素包括: (1) 由于硬件(例如 GPU) 的进步，计算能力的可用性，(2) 大量训练数据的可用性，以及(3) 学习中间表示的能力和灵活性。⁹

简而言之，深度学习使用多层非线性处理单元的级联进行特征提取和转换。靠近数据输入的较低层学习简单的特征，而较高层学习从较低层特征导出的更复杂的特征。该体系结构形成了一种分层的、强大的特征表示。图 2 显示了在人脸图像分类中通过深度学习从左(较低层)到右(较高层)学习到的特征层次。¹⁰ 我们可以看到，学习到的图像特征越来越复杂，从斑点/边缘开始，然后是鼻子/眼睛/脸颊，再到人脸。

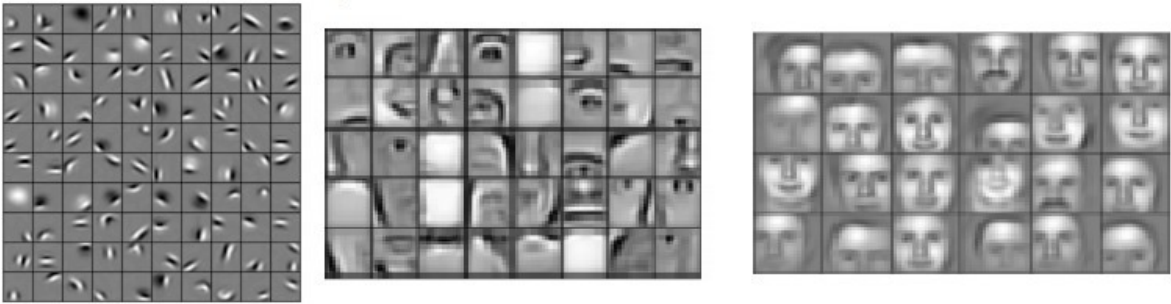


图 2:深度学习的特征层次

近年来，深度学习模型在自然语言处理领域得到了广泛的应用，并显示出巨大的潜力。在接下来的几个章节中，我们简要描述了主要的深度学习架构和已经应用于 NLP 任务的相关技术。

单词嵌入

NLP 中的许多深度学习模型需要单词嵌入结果作为输入特征。⁷ 单词嵌入是一种用于语言建模和特征学习的技术，其将词汇表中的单词转换为连续实数的向量(例如，*word "hat"* $\rightarrow (\dots, 0.15, \dots, 0.23, \dots, 0.41, \dots)$)。该技术通常涉及从高维稀疏向量空间(例如，独热编码向量空间，其中每个单词取一个维度)到低维密集向量空间的数学嵌入。嵌入向量的每个维度代表一个词的潜在特征。向量可以编码语言规则和模式。

可以使用神经网络¹¹⁻¹⁵ 或矩阵分解来学习单词嵌入。^{16, 17} 一个常用的单词嵌入系统是 Word2Vec，它本质上是一个计算高效的神经网络预测模型，从文本中学习单词嵌入。它包含连续词袋模型(CBOW)¹³ 和跳格模型(SG)¹⁴。CBOW 模型从其上下文单词(“男孩是_一顶帽子”，其中“_”表示目标单词)预测目标单词(例如，“穿着”)，而 SG 模型则相反，在给定目标单词的情况下预测上下文单词。从统计学上来说，CBOW 模型通过将整个上下文视为一个观察值来平滑大量的分布信息。这对于较小的数据集是有效的。然而，SG 模型将每个上下文-目标对视为一个新的观察，对于较大的数据集更好。另一种经常使用的学习方法是全局向量(GloVe)¹⁷，其在全局单词-单词共现矩阵的非零条目上被训练。

自动编码器和去噪自动编码器

自动编码器神经网络是一个三层神经网络，它将目标值设置为等于输入值。图 3 显示了一个自动编码器架构的示例。

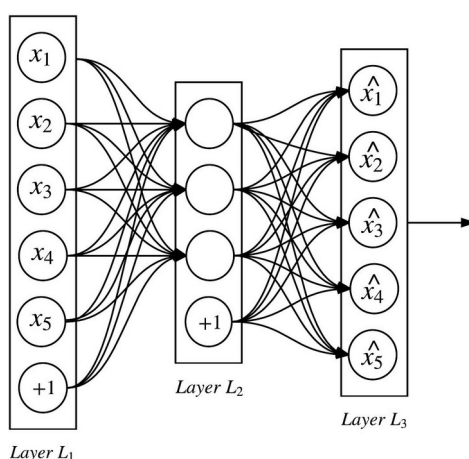


图 3: 自动编码器神经网络

ⁱ 源代码:<https://code.google.com/archive/p/word2vec/>

ⁱⁱ 源代码:<https://github.com/stanfordnlp/GloVe>

给定输入向量 $x \in [0, 1]^d$ ，自动编码器首先将其映射到隐藏表示

$y \in [0, 1]^d$ 通过编码器函数 $h(\cdot)$ (例如, sigmoid 函数)。然后, 潜在表示 y 被解码器函数 $g(\cdot)$ 映射回重建 $\hat{x} = g(h(x))$ 。自动编码器通常被训练成最小化一种形式的重建误差 ($loss(x, \hat{x})$)。自动编码器的目标是学习输入的表示, 这是隐藏层的激活。由于非线性函数 $h(\cdot)$ 和 $g(\cdot)$, 自动编码器能够学习非线性表示, 这给予它比其线性对应物更大的表达能力, 例如主成分分析 (PCA) 或潜在语义分析 (LSA)。

人们经常将自动编码器堆叠成层。较高级别的自动编码器使用较低级别的自动编码器的输出作为其训练数据。堆叠式自动编码器¹⁸和受限玻尔兹曼机器 (RBM)¹⁹是构建深度神经网络的最早方法。一旦以无监督的方式训练了一堆自动编码器, 它们的描述 x 的多级表示 (中间表示) 的参数可以用于初始化有监督的深度神经网络, 这已经在经验上显示出比随机参数初始化更好。

去噪自动编码器 (DAE)²⁰ 是自动编码器的扩展, 其中输入向量 x 被随机破坏成向量 x' 。并且该模型被训练以对其去噪, 即最小化

去噪重建误差 $loss(x, \hat{x})$ 。DAE 背后的想法是强制隐藏层

发现更强大的功能, 防止它只是学习身份。一个健壮模型应该能够很好地重建输入, 即使在存在噪声的情况下。例如, 从文档中删除或添加一些单词不应该改变文档的语义。

卷积神经网络

卷积神经网络 (CNN) 是一种特殊类型的前馈神经网络, 最初用于计算机视觉领域。它的设计灵感来自人类视觉皮层, 这是动物大脑中的一种视觉机制。视觉皮层包含许多细胞, 这些细胞负责在视野的小而重叠的子区域中检测光线, 这些子区域被称为感受野。这些单元充当输入空间的局部过滤器。CNN 由多个卷积层组成, 每一层都执行由视觉皮层中的细胞处理的功能。

图 4 示出了用于识别交通标志的 CNN。²¹ 输入是 $32 \times 32 \times 1$ 像素图像 (32×32 代表图像宽度 \times 高度; 1 表示输入通道)。在第一阶段, 使用过滤器 (尺寸为 $5 \times 5 \times 1$) 扫描图像。滤波器投射到输入图像中的每个区域都是感受野。过滤器实际上是一组数字 (称为权重或参数)。当滤波器滑动 (或卷积) 时, 它将其权重值与图像的原始像素值相乘 (逐元素乘法)。乘法都被累加成一个单一的数字, 这是感受野的代表。每个感受野都会产生一个数字。在过滤器完成对图像的扫描后, 我们可以得到一个数组 (大小为 $28 \times 28 \times 1$), 它被称为激活图或特征图。在 CNN 中, 我们需要使用不同的过滤器来扫描输入。在图 4 中, 我们应用了 108 种滤波器, 因此在第一阶段有 108 个堆叠特征图, 它由第一卷积层组成。在卷积层之后, 通常使用二次采样 (或汇集) 层来逐步减小表示的空间大小, 从而减少特征的数量和网络的计算复杂度。例如, 在第一阶段的二次采样之后, 卷积层将其维度减少到 $(14 \times 14 \times 108)$ 。注意, 虽然每个特征图的维度减少了, 但是二次采样步骤保留了最重要的信息, 通常使用的二次采样操作是最大池。然后, 第一级的输出成为第二级的输入, 并采用新的滤波器。新的过滤器大小为 $5 \times 5 \times 108$, 其中 108 是最后一个图层的要素地图大小。在第二阶段之后, CNN 使用一个完全连接的图层, 然后是一个带有输出类别的 softmax 读出图层用于分类。

CNN 中的卷积层起到特征提取器的作用，提取局部特征，因为它们将隐藏层的感受域限制为局部的。这意味着 CNN 通过在相邻层的神经元之间实施局部连接模式而具有特殊的空间局部相关性。这样的特征对于 NLP 中的分类是有用的，其中我们期望找到关于类成员的强局部线索，但是这些线索可以出现在输入中的不同位置。例如，在文档分类任务中，单个关键短语 (或 n-gram) 可以帮助确定文档的主题。我们希望了解某些单词序列是主题的良好指标，并不一定关心它们出现在文档中的什么位置。卷积层和汇集层允许 CNN 学习寻找这样的本地指示器，而不管它们的位置。⁸

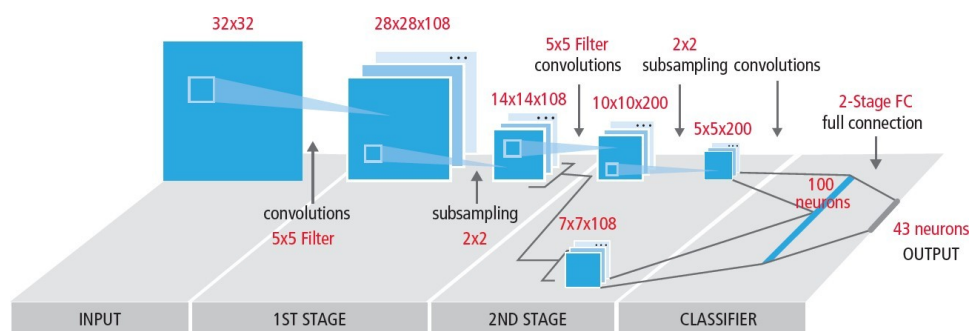


图 4: 卷积神经网络

递归神经网络

递归神经网络 (RNN)²² 是一类神经网络，其神经元之间的连接形成有向循环。与前馈神经网络不同，RNN 可以使用其内部“内存”来处理一系列输入，这使得它在处理顺序信息方面很受欢迎。“记忆”意味着 RNN 对序列的每个元素执行相同的任务，每个输出依赖于所有先前的计算，这就像“记住”到目前为止已经处理的信息。

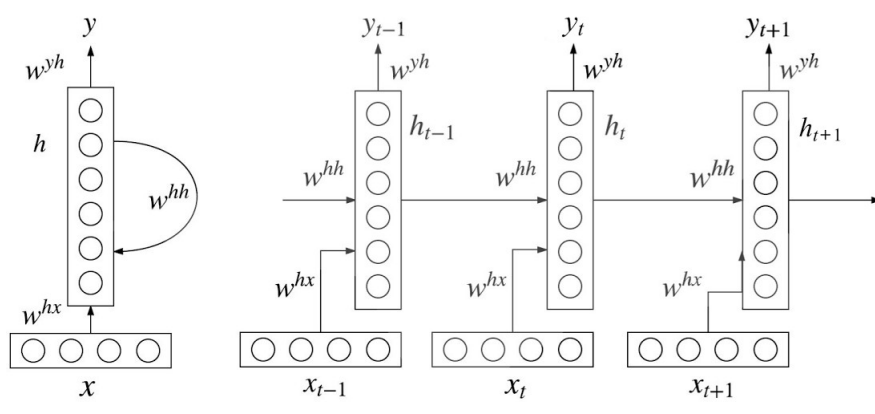


图 5: 递归神经网络

图 5 显示了一个 RNN 的例子。左图是一个具有循环的展开网络，而右图是一个具有三个时间步长的折叠序列网络。时间步长的长度由输入的长度决定。例如，如果要处理的单词序列是六个单词的句子，则 RNN 将被展开成具有六个时间步长或层的神经网络。一层对应一个词。

在图 5 中, x_t 是时间步长 t 的输入向量 h_t 是时间步长 t 的隐藏状态, 它是基于先前的隐藏状态和当前时间步长的输入计算的。

$$h_t = f(w^{hh}h_{t-1} + w^{hx}x_t) \quad (5)$$

在等式 (5) 中, 激活函数 f 通常是双曲正切函数或 ReLU 函数。 w^{hx} 是用于调节输入 x_t 的权重矩阵。 w^{hh} 是用于调节先前隐藏状态 h_t 的权重矩阵! 1.

y_t 是在步骤 t 词汇上的输出概率分布。例如, 如果我们想预测句子中的下一个单词, 它将是单词词汇上的概率向量。

$$y_t = \text{softmax}(w^{yh}h_t) \quad (6)$$

隐藏状态 h_t 被视为网络的记忆。它捕捉所有先前时间步骤中发生的信息。 y_t 仅基于 t 时间的记忆 h_t 和相应的权重矩阵 w^{yh} 来计算

与在每层使用不同参数的前馈神经网络不同, RNN 在所有步骤中共享相同的参数 (W^{hx} 、 W^{hh} 、 W^{yh})。这意味着它在每一步都执行相同的任务, 只是输入不同。这大大减少了需要学习的参数总数。

理论上, RNN 可以利用任意长序列中的信息, 但在实践中, 由于消失梯度或爆炸梯度问题, 标准 RNN 仅限于回顾几步。²³

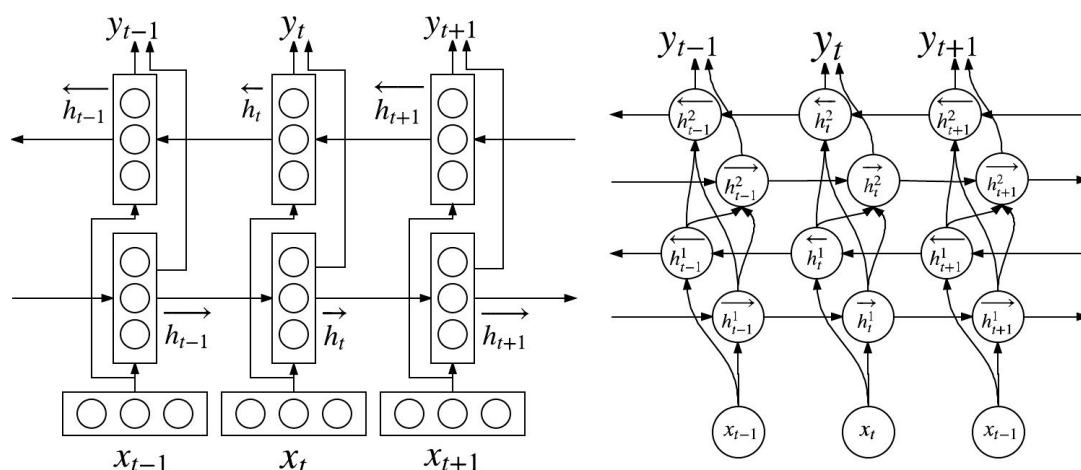


图 6: 双向 RNN (左) 和深度双向 RNN (右)

研究人员开发了更复杂类型的 RNN 来处理标准 RNN 模型的缺点: 双向 RNN、深度双向 RNN 和长短期记忆网络。双向 RNN 基于这样的思想, 即每次的输出可能不仅取决于序列中的前一个元素, 还取决于序列中的下一个元素。例如, 要预测一个序列中丢失的单词, 我们可能需要同时查看左右上下文。双向 RNN²⁴ 由两个 RNN 组成, 它们相互堆叠。一个按原始顺序处理输入, 另一个按相反的顺序处理输入。然后根据两个 rnn 的隐藏状态计算输出。深度双向 RNN 类似于双向 RNN。唯一的区别是每个时间步长有多个层,

这提供了更高的学习能力，但是需要大量的训练数据。图 6 分别显示了双向 RNN 和深度双向 RNN(有两层)的例子。

LSTM 网络

长短期记忆网络 (LSTM)²⁵ 是一种特殊类型的 RNN，能够学习长期依赖关系。

所有 rnn 都具有重复模块链的形式。在标准 RNNs 中，这种中继模块通常具有简单的结构。然而，LSTM 的重复模块更复杂。不是只有一个单一的神经网络层，而是有四层以一种特殊的方式相互作用。此外，它有两种状态：隐藏状态和单元状态。

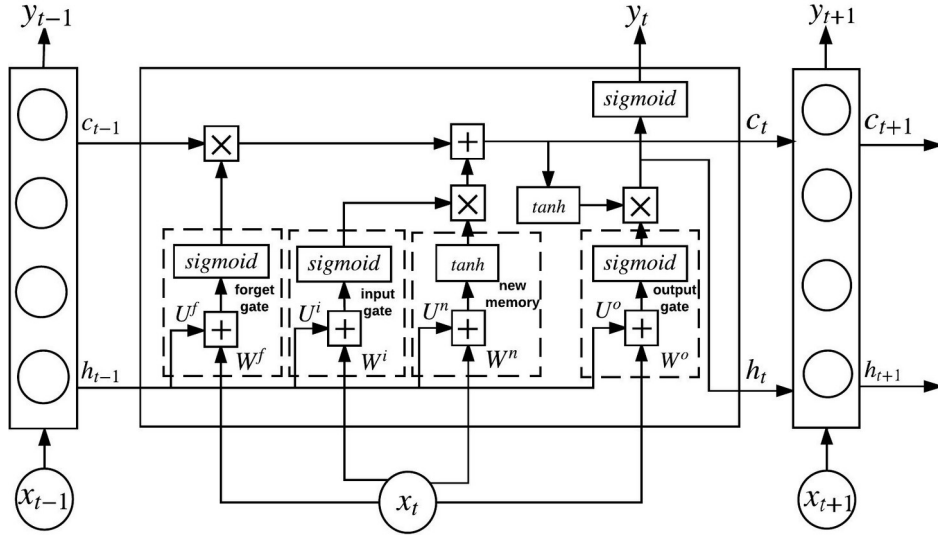


图 7:长短期记忆网络

图 7 显示了一个 LSTM 的例子。在时间步长 t ，LSTM 首先决定从单元状态转储什么信息。这个决定是由一个称为“遗忘门”的 sigmoid 函数/层 σ 做出的。该函数采用 h_{t-1} (从前一隐藏层输出) 和 x_t (当前输入)，并在 $[0, 1]$ 中输出一个数，其中 1 表示“完全保持”，0 表示等式 (7) 中的“完全转储”。

$$f_t = \sigma(W^f x_t + U^f h_{t-1}) \quad (7)$$

然后，LSTM 决定在单元状态中存储什么新信息。这有两步。首先，一个 sigmoid 函数/层，称为“输入门”，如等式 (8)，决定 LSTM 将更新哪些值。接下来，双曲正切函数/层创建新候选值 C_t 的向量，该向量将被添加到单元状态。LSTM 将这两者结合起来创建状态更新。

$$i_t = \sigma(W^i x_t + U^i h_{t-1}) \quad (8)$$

$$\tilde{C}_t = \tanh(W^n x_t + U^n h_{t-1}) \quad (9)$$

现在是时候更新旧的细胞状态 C_{t-1} 转换成新的小区状态 C_t ，如等式 (10) 所示。注意，遗忘门 f_t 可以控制通过它的梯度，并允许显式的“记忆”删除和更新，这有助于减轻标准 RNN 中的消失梯度或爆炸梯度问题。

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (10)$$

最后，LSTM 根据单元状态决定输出。LSTM 首先运行一个 sigmoid 层，该层决定在方程 (11) 中输出单元状态的哪些部分，称为“输出门”。然后，LSTM 将单元状态通过双曲正切函数，并将其乘以 sigmoid 门的输出，从而 LSTM 仅输出其决定为等式 (12) 的部分。

$$o_t = \sigma(W^o x_t + U^o h_{t-1}) \quad (11)$$

$$h_t = o_t * \tanh(C_t) \quad (12)$$

LSTM 通常应用于顺序数据，但也可用于树结构数据。Tai 等人²⁶将标准的推广到树形结构的 (Tree-LSTM)，并显示出比连续的更好的表示句子意思的性能。

LSTM 的一个微小变化是门控循环单元 (GRU)。^{27, 28} 它将“忘记”和“输入”门合并成一个更新门。它还合并了单元状态和隐藏状态，并进行一些其他更改。由此产生的模型比标准的 LSTM 模型更简单，并且越来越受欢迎。

基于递归神经网络的注意机制

据推测，双向 RNN 和 LSTM 应该能够处理数据的长期依赖性。但是在实践中，长程相关性仍然难以处理。因此，一种称为注意机制的技术被提出。

神经网络中的注意机制受到人类视觉注意机制的启发。也就是说，人类的视觉注意力能够聚焦在具有“高分辨率”的图像的某个区域上，同时以“低分辨率”感知周围的图像，然后随着时间的推移调整焦点。在 NLP 中，注意力机制允许模型根据输入文本和它迄今为止产生的内容来学习要注意什么，而不是像标准的 RNN 和 LSTM 那样将完整的源文本编码成固定长度的向量。

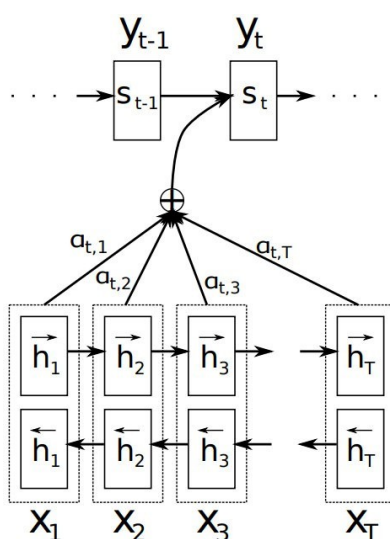


图 8: 双向递归神经网络中的注意机制

Bahdanau 等人²⁹首先在 NLP 中利用机器翻译的注意机制。他们提出了一种编码器-解码器框架，其中在翻译之前，使用注意机制来为目标语言中的单词选择原始语言中的参考单词。图 8 展示了注意机制在双向 RNN 中的应用。注意，每个解码器

输出字 y_t 取决于所有输入状态的加权组合，而不仅仅是正常情况下的最后一个状态。 a_t, T 是定义每个输入状态应该为每个输出加权多少的权重。例如，如果 $a_{2,2}$ 有一个大值，这意味着解码器在产生目标句子的第二个单词时，非常重视源句子中的第二个状态。 $a_{t,T}$ 的重量总和通常为 1。

存储网络

Weston 等人³⁰ 为问答问题引入了记忆网络 (MemNN) 的概念。它与几个推理组件和一个大的长期记忆相结合。组件可以是神经网络。记忆作为一个动态的知识库。四个可学习/推理组件的功能如下: I 组件将输入转换为内部特征表示; g 组件在给定新输入的情况下更新旧的存储器; o 组件生成输出 (也在特征表示空间中); r 组件将输出转换成响应格式。例如，给定一个句子列表和一个问答问题，MemNN 从这些句子中找到证据并生成答案。在推理过程中，I 组件一次读取一个句子，并将其编码为一个向量表示。然后 G 组件根据当前的句子表示更新一块内存。处理完所有句子后，生成一个存储矩阵 (每行代表一个句子)，存储句子的语义。对于一个问题，MemNN 将其编码成一个向量表示，然后 O 组件使用该向量从内存中选择一些相关的证据，并生成一个输出向量。最后，R 组件将输出向量作为输入，并输出最终响应。

在 MemNN 的基础上，Sukhbaatar 等人³¹ 提出了一种端到端记忆网络 (MemN2N)，这是一种在长期记忆成分上具有循环注意机制的神经网络架构，它可以通过标准的反向传播以端到端的方式进行训练。它证明了 O 组件中的多个计算层 (hops) 可以发现比单个计算层更抽象的证据，并为问题回答和语言建模产生改进的结果。值得注意的是，每个计算层都可以是基于内容的注意力模型。由此，MemN2N 在一定程度上细化了注意机制。另请注意，Graves 等人报告的神经图灵机也是一个类似的想法。³²

递归神经网络

递归神经网络 (RecNN) 是一种通常用于从数据中学习有向无环图结构 (例如，树结构) 的神经网络。递归神经网络可以被看作是递归神经网络的推广。给定句子的结构表示 (例如，解析树)，RecNN 以自下而上的方式递归生成父表示，通过组合记号来产生短语的表示，最终产生整个句子的表示。然后，句子级表示可以用于对给定的输入句子进行最终分类 (例如，情感分类)。RecNN 中矢量合成的示例过程如图 9³³ 所示。节点“非常感兴趣”的向量由节点“非常”和节点“感兴趣”的向量组成。同样，节点“非常有趣”是由短语节点“非常有趣”和单词节点“是”组成的。

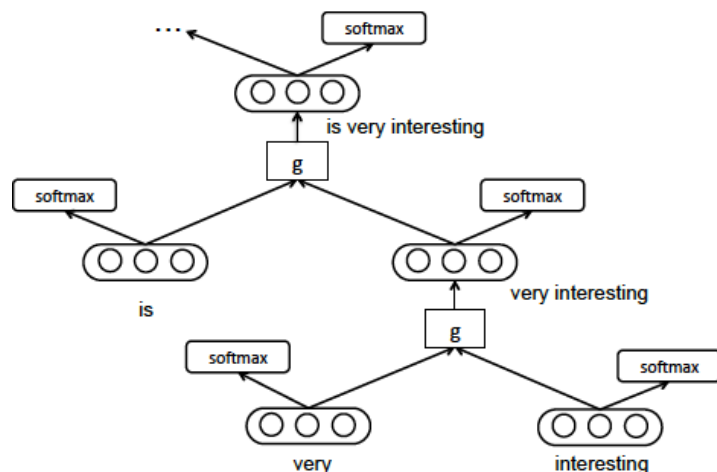


图 9:递归神经网络

情感分析任务

我们现在准备调查深度学习在情感分析中的应用。但在此之前，我们首先简要介绍一下本节中的主要情感分析任务。更多细节请参考刘的《情感分析》一书。

研究者主要在三个粒度级别上研究情感分析:文档级、句子级和方面级。文档级情感分类将有意见的文档(例如，产品评论)分类为表达总体积极或消极的意见。它将整个文档视为基本信息单元，并假设该文档已知是有观点的，并且包含关于单个实体(例如，特定电话)的观点。句子级情感分类对文档中的单个句子进行分类。但是，每一句都不能假设是自以为是。传统上，人们往往首先将一个句子分为自以为是或不自以为是，这被称为主观性分类。然后将由此产生的有意见的句子分类为表达积极或消极的意见。句子级情感分类也可以公式化为一个三类分类问题，即将一个句子分为中性、积极或消极三类。与文档级和句子级情感分析相比，方面级情感分析或者基于方面的情感分析更加细粒度。它的任务是提取和总结人们对实体和实体的方面/特征表达的意见，也称为目标。例如，在产品评论中，它旨在分别总结对产品不同方面的正面和负面意见，尽管对产品的总体看法可能是正面的或负面的。基于方面的情感分析的整个任务由几个子任务组成，例如方面提取、实体提取和方面情感分类。比如从“iPhone 的音质很棒，但它的电池很烂”这句话来看，实体提取要识别“iPhone”是实体，方面提取要识别“音质”和“电池”是两个方面。方面情感分类应该将对 iPhone 的语音质量表达的情感分类为积极的，而对 iPhone 的电池表达的情感分类为消极的。注意，为了简单起见，在大多数算法中，特征提取和实体提取被组合，并且被称为特征提取或情感/意见目标提取。

除了这些核心任务，情感分析还研究情感分析、讽刺检测、多语言情感分析等。更多细节见刘的书1。在下面的部分中，我们将调查所有这些情感分析任务中的深度学习应用。

文档级情感分类

文档级的情感分类是为意见文档分配总体情感取向/极性，即，确定该文档(例如，完整的在线评论)传达总体正面还是负面意见。在这种情况下，这是一个二元分类任务。它也可以被公式化为一个回归任务，例如，为评论推断一个从 1 星到 5 星的总体评级分数。一些研究人员也将此视为 5 级分类任务。

情感分类通常被认为是文档分类的一个特例。在这样的分类中，文档表示起着重要的作用，它应该反映文档中单词或句子所传达的原始信息。传统上，在自然语言处理和文本挖掘中，词袋模型 (BoW) 被用来生成文本表示，通过该模型，文档被视为一个词袋。基于 BoW，文档被转换成具有固定长度的数字特征向量，其每个元素可以是单词出现(不存在或存在)、词频或 TF-IDF 得分。它的维度等于词汇量的大小。来自 BoW 的文档向量通常非常稀疏，因为单个文档仅包含词汇表中的少量单词。早期的神经网络采用了这样的特征设置。

尽管 BoW 很受欢迎，但它也有一些缺点。首先，忽略了词序，这意味着两个文档可以有完全相同的表示，只要它们共享相同的单词。Bag-of-N-Grams 是 BoW 的扩展，它可以考虑短上下文 (n-gram) 中的词序，但它也存在数据稀疏和高维数的问题。其次，BoW 只能勉强编码单词的语义。比如 BoW 中的“smart”、“clever”、“book”这三个词之间的距离是相等的，但是“smart”在语义上应该比“book”更接近“clever”。

为了解决 BoW 的缺点，提出了基于神经网络的单词嵌入技术(在前述部分中介绍)，以生成用于单词表示的密集向量(或低维向量)，其在某种程度上能够编码单词的一些语义和句法属性。使用单词嵌入作为单词的输入，可以使用神经网络导出作为密集向量(或称为密集文档向量)的文档表示。

注意，除了上述两种方法，即使用 BoW 和通过单词嵌入学习文档的密集向量，还可以直接从 BoW 学习密集文档向量。我们在表 2 中区分了相关研究中使用的不同方法。

当文档被适当地表示时，可以使用遵循传统监督学习设置的各种神经网络模型来进行情感分类。在一些情况下，神经网络可以仅用于提取文本特征/文本表示，并且这些特征被馈送到一些其他非神经分类器(例如，SVM)以获得最终的全局最优分类器。神经网络和 SVM 的特性以结合其优点的方式互补。

除了复杂的文档/文本表示，研究人员还利用数据的特征——产品评论，进行情感分类。对于产品评论，一些研究人员发现联合建模情感和一些附加信息(例如，用户信息和产品信息)用于分类是有益的。此外，由于文档通常包含长的依赖关系，所以注意机制也经常用于文档级情感分类。我们在表 2 中总结了现有的技术。

调查研究工作	文档/文本表示	神经网络模型	使用注意机制	带情感的联合建模
Moraes 等人 ³⁴	弓	人工神经网络	不	-
乐和米科洛夫 ³⁵	在句子、段落、文档级别学习密集向量	段落向量	不	-
Glorot 等人 ³⁶	向密集文档向量低头	堆叠去噪自动编码器	不	来自目标域的无监督数据表示(在迁移学习设置中)
翟与章 ³⁷	向密集文档向量低头	去噪自动编码器	不	-
约翰逊和章 ³⁸	向密集文档向量低头	BoW-CNN 和 Seq-CNN	不	-
唐等 ³⁹	密集文档向量的单词嵌入	CNN/LSTM(学习句子表达)+GRU(学习文档表达)	不	-
唐等 ⁴⁰	密集文档向量的单词嵌入	基于 CNN 的用户产品神经网络	不	用户信息和产品信息
陈等 ⁴¹ 人	密集文档向量的单词嵌入	基于 LSTM 的用户产品关注度	是	用户信息和产品信息
Dou ⁴²	密集文档向量的单词嵌入	存储网络	是	用户信息和产品信息
徐等 ⁴³ 人	密集文档向量的单词嵌入	长短期记忆网络	不	-
杨等 ⁴⁴ 人	密集文档向量的单词嵌入	基于 GRU 的序列编码器	分层注意	-
尹等 ⁴⁵	密集文档向量的单词嵌入	输入编码器和 LSTM	分层注意	特征/目标信息
周等 ⁴⁶ 人	密集文档向量的单词嵌入	长短期记忆网络	分层注意	跨语言信息
李等 ⁴⁷	密集文档向量的单词嵌入	存储网络	是	跨域信息

表 2:用于文档级情感分类的深度学习方法下面,我们也对这些已有的代表性作品进行简单描述。

Moraes 等人³⁴对支持向量机(SVM)和人工神经网络(ANN)进行了文档级情感分类的实证比较,结果表明,在大多数情况下,ANN产生的结果与SVM的结果具有竞争力。

为了克服BoW的弱点,Le和Mikolov³⁵提出了段落向量,这是一种无监督的学习算法,可以学习句子、段落和文档等可变长度文本的向量表示。通过预测从段落中采样的上下文中的周围单词来学习向量表示。

Glorot 等人³⁶研究了用于情感分类的领域适应问题。他们提出了一种基于具有稀疏整流器单元的堆叠去噪自动编码器的深度学习系统，该系统可以使用标记和未标记的数据来执行无监督的文本特征/表示提取。这些特征非常有利于情感分类器的领域适应。

Zhai 和 Zhang³⁷介绍了一种半监督自动编码器，它在学习阶段进一步考虑情感信息，以便获得更好的文档向量，用于情感分类。更具体地，该模型通过将自动编码器中的损失函数放宽到 Bregman 散度并且还从标签信息中导出区别损失函数来学习文本数据的任务特定表示。

Johnson 和 Zhang³⁸提出了一个名为 BoW-CNN 的 CNN 变体，它在卷积层采用了词包转换。他们还设计了一种新的模型，称为 Seq-CNN，通过串联多个单词的一键向量来保持单词的顺序信息。

Tang 等人³⁹提出了一种学习文档表示的神经网络，其中考虑了句子关系。它首先用 CNN 或 LSTM 从单词嵌入中学习句子表征。然后利用 GRU 对文档表示中的句子语义及其内在关系进行自适应编码，用于情感分类。

唐等⁴⁰将用户表征和产品表征应用于评论分类。这个想法是，这些表示可以捕捉重要的全局线索，如用户的个人偏好和产品的整体质量，这可以提供更好的文本表示。

Chen 等人⁴¹也结合了用户信息和产品信息用于分类，但是通过单词和句子级关注，这可以在单词级和语义级都考虑全局用户偏好和产品特征。同样，Dou⁴²使用深度存储网络来捕获用户和产品信息。提出的模型可以分为两个独立的部分。第一部分，应用 LSTM 学习文档表示。在第二部分中，由多个计算层(跳)组成的深度记忆网络用于预测每个文档的评论评级。

Xu 等人⁴³提出了一种缓存的模型来捕获长文本中的整体语义信息。模型中的记忆被分成遗忘率不同的几组。直觉是让遗忘率低的记忆组捕捉全局语义特征，遗忘率高的记忆组学习局部语义特征。

Yang 等人⁴⁴提出了用于评论的文档级情感评级预测的分层注意网络。该模型包括两个级别的注意机制：一个在单词级别，另一个在句子级别，这允许模型在构建文档的表示时或多或少地注意单个单词或句子。

Yin 等人⁴⁵将文档级方面-情感评级预测任务公式化为机器理解问题，并提出了一种基于分层交互式注意的模型。具体来说，文档和伪方面问题被交错以学习方面感知的文档表示。

周等人⁴⁶设计了一个基于注意力的网络，用于在文档级别进行跨语言情感分类。该模型由两个基于注意的 LSTM 组成，用于双语表征，每个 LSTM 也是分层结构的。在这种设置下，它有效地将情感信息从资源丰富的语言(英语)适应到资源贫乏的语言(汉语)，并帮助提高情感分类性能。

李等人⁴⁷提出了一种在迁移学习环境中用于跨领域情感分类的对抗性记忆网络，其中来自源和目标领域的数据被一起建模。它联合训练两个网络用于情感分类和领域分类(即，文档是来自源领域还是目标领域)。

句子级情感分类

句子级情感分类是确定在单个给定句子中表达的情感。如前所述，句子的情感可以通过主观性分类⁴⁸和极性分类来推断，其中前者对句子是主观的还是客观的进行分类，而后者决定主观句子表达的是消极情感还是积极情感。在现有的深度学习模型中，句子情感分类通常被表述为一个联合三路分类问题，即预测一个句子是正面的、神经的还是负面的。

与文档级情感分类一样，由神经网络产生的句子表示对于句子级情感分类也很重要。此外，由于与文档相比，句子通常较短，因此一些句法和语义信息(例如，解析树、观点词典和词性标签)可能会有所帮助。还可以考虑其他信息，如评论评级、社会关系和跨领域信息。例如，社交关系已被用于发现社交媒体数据(如推文)中的情感。

在早期的研究中，分析树(提供一些语义和句法信息)与原始单词一起用作神经模型的输入，从而可以更好地推断情感成分。但是最近，CNN和RNN变得更受欢迎，它们不需要解析树来从句子中提取特征。相反，CNN和RNN使用单词嵌入作为输入，它已经编码了一些语义和句法信息。此外，CNN或RNN的模型架构也可以帮助学习句子中单词之间的内在关系。下面详细介绍相关工作。

Socher等人⁴⁹首先提出了用于句子级情感分类的半监督递归自动编码器网络(RAE)，该网络获得了句子的降维向量表示。后来，Socher等人⁵⁰提出了一种矩阵向量递归神经网络(MV-RNN)，其中每个词还与树结构中的矩阵表示(除了向量表示之外)相关联。树形结构是从外部解析器获得的。在Socher等人⁵¹中，作者进一步介绍了递归神经张量网络(RNTN)，其中基于张量的组合函数用于更好地捕捉元素之间的相互作用。钱等人³³提出了两个更先进的模型，标记引导的递归神经网络(TG-RNN)，它根据短语的词性标记选择组合函数，以及标记嵌入的递归神经网络/递归神经张力网络(TE-RNN/RNTN)，它学习标记嵌入，然后将标记和词嵌入结合在一起。

Kalchbrenner等人⁵²提出了一种用于句子语义建模的动态CNN(称为DCNN)。DCNN使用动态K-Max池操作符作为非线性子采样函数。由网络归纳出的特征图能够捕捉词汇关系。Kim⁵³还提出使用CNN进行句子级情感分类，并试验了几种变体，即CNN-rand(其中随机初始化单词嵌入)，CNN-static(其中单词嵌入是预训练和固定的)，CNN-non-static(其中单词嵌入是预训练和微调的)和CNN-multicessive(其中使用多组单词嵌入)。

dos Santos和Gatti⁵⁴提出了一个字符对句子的CNN(CharSCNN)模型。CharSCNN使用两个卷积层从任意大小的单词和句子中提取相关特征

对短文本进行情感分析。Wang 等人⁵⁵ 通过在合成过程中模拟词语的交互来利用进行推特情感分类。与简单递归神经网络中的加法运算相比,通过门结构的字嵌入之间的乘法运算用于提供更大的灵活性并产生更好的合成结果。类似于双向 RNN,通过允许隐藏层中的双向连接,单向 LSTM 可以扩展到双向 LSTM⁵⁶。

王等⁵⁷ 提出了一个区域 CNN-模型,该模型由区域 CNN 和两部分组成,用于预测文本的价唤醒等级。

Wang 等人⁵⁸ 描述了一种用于短文本情感分类的 CNN 和联合架构,其利用了 CNN 生成的粗粒度局部特征和通过学习到的远距离依赖性。

Guggilla 等人⁵⁹ 提出了一个基于 LSTM 和 CNN 的深度学习模型,该模型利用 word2vec 和语言嵌入进行主张分类(将句子分类为事实或感觉)。

Huang 等人⁶⁰ 提出将句法知识(例如,词性标签)编码在树形结构的 LSTM 中,以增强短语和句子的表示。

Akhtar 等人⁶¹ 提出了几种基于多层感知器的集成模型,用于金融微博和新闻的精细情感分类。

关等人⁶² 采用弱监督 CNN 进行句子(以及方面)级别的情感分类。它包含两步学习过程:首先学习由总体评论评级弱监督的句子表示,然后使用句子(和方面)级别标签进行微调。

Teng 等人⁶³ 提出了一种基于上下文敏感词汇的情感分类方法,该方法基于简单加权和模型,使用双向 LSTM 来学习词汇情感的情感强度、强化和否定,以构成句子的情感值。

Yu 和 Jiang⁶⁴ 研究了用于跨领域句子情感分类的学习广义句子嵌入的问题,并设计了包含两个分离的 CNN 的神经网络模型,这两个 CNN 从标记和未标记数据中联合学习两个隐藏特征表示。

赵等人⁶⁵ 介绍了一种递归随机行走网络学习方法,通过利用用户发布的推文及其社会关系的深层语义表示,对有意见的推文进行情感分类。

Mishra 等人⁶⁶ 利用 CNN 从阅读文本的人类读者的眼球运动(或凝视)数据中自动提取认知特征,并将它们作为丰富的特征与文本特征一起用于情感分类。

钱等人⁶⁷ 提出了一个语言正则化的任务。该模型将情感词汇、否定词和强度词等语言资源融入到 LSTM 中,以更准确地捕捉句子中的情感效应。

方面级情感分类

与文档级和句子级情感分类不同,方面级情感分类将情感和目标信息都视为一种情感

总是有一个目标。如前所述，目标通常是一个实体或一个实体方面。为了简单起见，实体和方面通常都被称为方面。给定句子和目标方面，方面级情感分类旨在推断句子对目标方面的情感极性/取向。比如“屏幕很清晰但是续航太短”这句话。如果目标方面是“屏幕”，情感是积极的，但是如果目标方面是“电池寿命”，情感是消极的。我们将在下一节讨论自动化方面或目标提取。

方面级情感分类是具有挑战性的，因为对目标与其周围上下文单词的语义相关度进行建模是困难的。不同的语境词对句子对目标的情感极性有不同的影响。因此，在使用神经网络建立学习模型时，有必要捕获目标词和上下文词之间的语义联系。

在使用神经网络的方面级情感分类中有三个重要的任务。第一个任务是表示目标的上下文，其中上下文意味着句子或文档中的上下文单词。这个问题可以用上面两节提到的文本表示方法来解决。第二个任务是生成一个目标表示，它可以正确地与其上下文交互。一般的解决方法是学习一个目标嵌入，类似于单词嵌入。第三个任务是为指定的目标识别重要的情感上下文(单词)。例如，在“iPhone 的屏幕清晰但电池寿命短”这句话中，“清晰”是“屏幕”的重要上下文词，“短”是“电池寿命”的重要上下文词。这个任务最近由注意机制来处理。虽然已经提出了许多深度学习技术来处理方面级情感分类，但是据我们所知，在文献中仍然没有主流的技术。相关工作及其重点介绍如下。

Dong 等人⁶⁸提出了一种用于目标相关的 twitter 情感分类的自适应递归神经网络(AdaRNN)，该网络根据上下文和句法结构学习向目标传播词语的情感。它使用根节点的表示作为特征，并将它们提供给 softmax 分类器，以预测类的分布。

Vo 和 Zhang⁶⁹利用丰富的自动特征研究了基于方面的 Twitter 情感分类，这些自动特征是使用无监督学习方法获得的附加特征。该论文表明，多重嵌入、多重池函数和情感词典可以提供丰富的特征信息来源，并有助于实现性能增益。

由于能更灵活地捕捉目标词与其语境词之间的语义关系，唐等⁷⁰提出了目标相关的(TD-LSTM)和目标关联的(TC-LSTM)来扩展 LSTM。他们将给定的目标作为一个特征，并将其与上下文特征连接起来进行体貌情感分类。

Ruder 等人⁷¹提出使用分层和双向 LSTM 模型进行方面级情感分类，该模型能够利用句子内和句子间的关系。在评论中对句子及其结构的唯一依赖使得所提出的模型与语言无关。单词嵌入被输入到句子级双向 LSTM 中。前向和后向 LSTM 的最终状态与目标嵌入连接在一起，并馈入双向审查级 LSTM。在每个时间步长，前向和后向 LSTM 的输出被连接并被馈送到最终层，该最终层输出情感的概率分布。

考虑到 Dong 等人⁶⁸以及 Vo 和 Zhang⁶⁹的工作的局限性，Zhang 等人⁷²提出了一个句子级神经网络，以解决 pooling 函数的弱点，即没有显式地对 tweet 级语义进行建模。为了实现这一点，提出了两个门控神经网络。首先，一个双-

方向选通神经网络用于连接 tweet 中的单词，以便池函数可以应用于隐藏层，而不是单词，以便更好地表示目标及其上下文。第二，使用三路门控神经网络结构来建模目标提及与其周围上下文之间的交互，通过使用门控神经网络结构来分别建模封闭 tweet 的语法和语义以及周围上下文与目标之间的交互来解决局限性。门控神经网络已经显示出通过更好地传播梯度来减少标准递归神经网络向序列末端的偏差。

Wang 等人⁷³提出了一种基于注意的目标嵌入的方法，这种方法被证明是一种有效地加强神经模型注意句子相关部分的方法。注意机制被用于强制模型关注句子的重要部分，以响应特定的方面。同样，Yang 等人⁷⁴提出了两个基于注意力的双向 LSTMs 来提高分类性能。Liu 和 Zhang⁷⁵通过区分从给定目标/方面的左上下文和右上下文获得的注意力来扩展注意力模型。他们通过增加多个门来进一步控制他们的注意力贡献。

Tang 等人⁷⁶引入了用于方面级情感分类的端到端存储网络，该网络采用具有外部存储器的注意机制来捕捉每个上下文单词相对于给定目标方面的重要性。当推断方面的情感极性时，这种方法明确地捕捉了每个上下文单词的重要性。这种重要度和文本表示用多个计算层来计算，每个计算层是外部存储器上的神经注意模型。

雷等人⁷⁷提出使用神经网络方法提取输入文本片段作为评审评级的理由。该模型由一个生成器和一个解码器组成。生成器指定在可能的基础上的分布(提取的文本)，编码器将任何这样的文本映射到特定于任务的目标向量。对于多方面情感分析，目标向量的每个坐标表示与相关方面有关的响应或评级。

李等人⁷⁸将目标识别任务集成到情感分类任务中，以更好地模拟方面-情感交互。他们表明，情感识别可以通过端到端的机器学习架构来解决，其中两个子任务通过深度记忆网络交织在一起。这样，在目标探测中产生的信号为极性分类提供线索，反过来，预测的极性为目标的识别提供反馈。

马等人⁷⁹提出了一个交互式注意网络(IAN)，它同时考虑了对目标和上下文的注意。也就是说，它使用两个注意力网络来交互式地检测目标表达/描述的重要单词及其完整上下文的重要单词。

Chen 等人⁸⁰提出利用循环注意网络来更好地捕捉复杂上下文的情绪。为了实现这一点，他们提出的模型使用循环/动态注意力结构，并在 GRUs 中学习注意力的非线性组合。

Tay 等人⁸¹设计了一个二元记忆网络(DyMemNN)，通过使用神经张量组合或全息组合进行记忆选择操作，对方面和上下文之间的二元交互进行建模。

方面提取和分类

为了执行特征级情感分类，需要具有特征(或目标)，这些特征可以手动给出或自动提取。在本节中，我们讨论使用深度学习从句子或文档中自动提取方面(或方面术语提取)的现有工作

模特。让我们用一个例子来说明这个问题。例如，在句子“图像非常清楚”中，“图像”这个词是一个体术语(或情感目标)。方面分类的相关问题是将相同的方面表达式分组到一个类别中。例如，方面术语“图像”、“照片”和“图片”可以归入一个名为图像的方面类别。在下面的回顾中，我们包括了与观点相关的方面和实体的提取。

深度学习模型对这项任务有帮助的一个原因是，深度学习本质上擅长学习(复杂的)特征表示。当一个方面在某个特征空间中被适当地特征化时，例如，在一个或某个隐藏层中，一个方面及其上下文之间的语义或相关性可以通过它们对应的特征表示之间的相互作用来捕获。换句话说，深度学习提供了一种无需人工参与的自动化特征工程的可能方法。

Katihar 和 Cardie⁸² 研究了使用深度双向 LSTMs 联合提取意见实体以及连接实体的 IS-FORM 和 IS-ABOUT 关系。Wang 等人⁸³ 进一步提出了一种结合和条件随机场(CRF)的联合模型来共同提取方面和意见术语或表达。该模型能够学习高层次的区别特征，同时在方面和观点项之间双向传播信息。Wang 等人⁸⁴ 进一步提出了一个耦合的多层注意模型(CMLA)，用于方面和观点术语的共同提取。该模型由一个方面注意和一个意见注意组成，使用 GRU 单位。李和 Lam⁸⁵ 报道了一种改进的基于的方法，专门用于特征项提取。它由三个 lstm 组成，其中两个 lstm 用于捕获方面和情感交互。第三个 LSTM 是使用情感极性信息作为附加指导。

何等人⁸⁶ 提出了一种基于注意力的无监督特征提取模型。主要的直觉是利用注意机制来更多地关注与方面相关的单词，同时在方面嵌入的学习期间去强调与方面无关的单词，类似于自动编码器框架。

Zhang 等人⁸⁷ 使用神经网络扩展了 CRF 模型，以联合提取方面和相应的情感。所提出的 CRF 变体用连续的单词嵌入代替了 CRF 中原来的离散特征，并且在输入和输出节点之间增加了一个神经层。

周等人⁸⁸ 提出了一种半监督的词嵌入学习方法，用于在大量带有噪声标签的评论集上获得连续的词表示。通过学习单词向量，通过神经网络叠加单词向量来学习更深层次的混合特征。最后，用混合特征训练的逻辑回归分类器用于预测方面类别。

Yin 等人⁸⁹ 首先通过考虑连接单词的依存路径来学习单词嵌入。然后，他们设计了一些嵌入特性，这些特性考虑了基于 CRF 的方面术语抽取的线性上下文和依赖上下文信息。

熊等人⁹⁰ 提出了一种基于注意的深度距离度量学习模型来分组体短语。基于注意的模型是学习上下文的特征表示。方面短语嵌入和上下文嵌入都用于学习 K-means 聚类的深度特征子空间度量。

茯苓等⁹¹ 提出用 CNN 进行方面提取。他们开发了一个七层深度卷积神经网络，将固执己见的句子中的每个单词标记为体或非体单词。一些语言模式也被集成到模型中以进一步改进。

Ying 等人⁹²提出了两个基于 RNN 的跨领域方面提取模型。他们首先使用基于规则的方法为每个句子生成一个辅助标签序列。然后，他们使用真实标签和辅助标签来训练模型，这显示了有希望的结果。

意见表达抽取

在本节和接下来的几节中，我们将讨论深度学习在一些其他情感分析相关任务中的应用。这一部分关注观点表达提取(或观点术语提取，或观点识别)的问题，其目的是识别句子或文档中的情感表达。

类似于方面提取，使用深度学习模型的意见表达提取是可行的，因为它们的特征也可以在一些特征空间中被识别。

Irsoy 和 Cardie⁹³探索了深层双向 RNN 在该任务中的应用，其性能优于具有相同参数数量的传统浅层 RNNs 以及之前的 CRF 方法。⁹⁴

Li 等人⁹⁵提出了基于架构和单词嵌入的一般类别的判别模型。作者在不同的 RNN 架构中使用了来自三个外部来源的预训练单词嵌入，这些架构包括 Elman 型、Jordan 型、LSTM 型及其变体。

Wang 等人⁸³提出了一个集成递归神经网络和 CRF 的模型来共同提取方面和观点术语。前面提到的 CMLA 也被提议用于方面和观点术语的共同提取。⁸⁴

情感构成

情感合成主张观点表达的情感取向由其成分的意义以及语法结构决定。由于其特殊的树结构设计，RecNN 自然适合于这项任务。⁵¹ Irsoy 和 Cardie⁹⁶报告说，具有深层架构的 RecNN 可以更准确地捕捉语言中复合性的不同方面，这有利于情感复合。朱等⁹⁷提出了一种神经网络，用于在情感合成过程中融合合成情感和非合成情感。

意见持有人抽取

意见持有人(或来源)提取是识别谁持有意见(或意见来自谁/哪里)的任务。例如，在句子“约翰讨厌他的汽车”中，意见持有人是“约翰”。这个问题通常被公式化为一个序列标记问题，如意见表达抽取或方面抽取。请注意，意见持有人可以是明确的(来自句子中的名词短语)或隐含的(来自作者)，如杨和卡迪⁹⁸所示。Deng 和 Wiebe⁹⁹提出使用意见表达的词嵌入作为识别参与者意见和非参与者意见的来源的特征，其中来源可以是名词短语或作者。

时态观点挖掘

在情绪分析的问题定义中，时间也是一个重要的维度(见刘的书 1)。通过久而久之，人们可以保持或改变他们的想法，甚至给出新的观点。

因此，预测未来观点在情感分析中很重要。最近报道了一些使用神经网络来解决这个问题的研究。

Chen 等人¹⁰⁰提出了一个基于内容的社会影响模型(CIM)来预测 twitter 用户的意见行为。也就是说，它使用过去的推文来预测用户未来的观点。它基于神经网络框架来编码用户内容和社会关系因素(一个人对目标的看法受到他的朋友的影响)。

Rashkin 等人¹⁰¹在社交媒体环境中使用 LSTMs 进行有针对性的情绪预测。他们引入了多语言内涵框架，旨在预测参与框架的世界事件参与者的隐含情绪。

嵌入单词的情感分析

很明显，单词嵌入在基于深度学习的情感分析模型中起着重要的作用。它还表明，即使不使用深度学习模型，单词嵌入也可以用作各种任务的非神经学习模型的特征。因此，该部分特别强调了单词嵌入对情感分析的贡献。

我们首先介绍了情绪编码的单词嵌入的工作。对于情感分析，直接应用常规单词方法如 CBOW 或 Skip-gram 来从上下文学习单词嵌入会遇到问题，因为具有相似上下文但相反情感极性(例如，“好”或“坏”)的单词可能被映射到嵌入空间中的附近向量。因此，已经提出了情感编码的单词嵌入方法。Mass el al.¹⁰²学习了可以捕获语义和情感信息的单词嵌入。Bespalov 等人¹⁰³表明，与潜在表示相结合的 n-gram 模型将产生更适合情感分类的嵌入。Labutov 和 Lipson¹⁰⁴通过将句子的情感监督视为正则化项，用逻辑回归重新嵌入现有的单词嵌入。

Le 和 Mikolov³⁵提出了段落向量的概念，首先学习可变长度文本片段的固定长度表示，包括句子、段落和文档。他们对句子和文档级别的情感分类任务进行了实验，并获得了性能增益，这证明了段落向量在捕捉语义以帮助情感分类方面的优点。Tang 等人^{105, 106}提出了用于学习特定于情感的词嵌入的模型(SSWE)，其中不仅语义而且情感信息都被嵌入到所学习的词向量中。Wang 和 Xia¹⁰⁷开发了一种神经架构，通过在文档和单词级别上集成情感监督来训练带有情感的单词嵌入。Yu 等人¹⁰⁸采用了一种细化策略来获得联合语义-情感承载词向量。

特征丰富和多义词嵌入也被用于情感分析。Vo 和 Zhang⁶⁹利用丰富的自动特征研究了基于方面的 Twitter 情感分类，这些自动特征是使用无监督学习技术获得的附加特征。李和 Jurafsky¹⁰⁹实验了在各种 NLP 任务中使用多义词嵌入。实验结果表明，虽然这种嵌入确实提高了一些任务的性能，但它们对情感分类任务几乎没有帮助。Ren 等人¹¹⁰提出了学习主题丰富的多原型词嵌入的方法，用于 Twitter 情感分类。

多语言单词嵌入也已经被应用于情感分析。周等人¹¹¹报道了用于跨语言情感分类的双语情感词嵌入(BSWE)模型。它通过使用标注语料库及其翻译，而不是大规模的平行语料库，将情感信息融入英汉双语嵌入中。巴恩斯等人

a1.¹¹² 比较了几种类型的双语单词嵌入和神经机器翻译技术，用于跨语言的基于方面的情感分类。

Zhang 等人¹¹³ 将单词嵌入与矩阵分解相结合，用于基于个性化评论的评级预测。具体而言，作者使用情感词典来细化现有的面向语义的词向量（例如，word2vec 和 GloVe）。Sharma 等人¹¹⁴ 提出了一种半监督技术，使用带有情感的单词嵌入来对形容词的情感强度进行排序。在最近的许多其他研究中，还利用或改进了单词嵌入技术来帮助解决各种情感分析任务。^{55, 62, 87, 89, 95}

讽刺分析

讽刺是言语讽刺的一种形式，也是与情感分析密切相关的概念。最近，自然语言处理社区对讽刺检测的兴趣越来越大。由于在许多其他 NLP 问题上取得了令人印象深刻的成功，研究人员试图使用深度学习技术来解决它。

张等¹¹⁵ 构建了一个深度神经网络模型，用于推文挖苦检测。他们的网络首先使用双向 GRU 模型在本地捕获推文的句法和语义信息，然后使用池神经网络从历史推文中自动提取上下文特征，以检测讽刺性的推文。

Joshi 等人¹¹⁶ 研究了用于讽刺检测的基于单词嵌入的特征。他们用增强的单词嵌入特征对四种过去的讽刺检测算法进行了实验，并显示了有希望的结果。

茯苓等人¹¹⁷ 开发了一个基于 CNN 的讽刺检测模型（讽刺或非讽刺推文分类），通过联合建模预先训练的情绪、情感和个性特征，以及推文中的文本信息。

Peled 和 Reichart¹¹⁸ 提议基于 RNN 神经机器翻译模型来翻译讽刺性的推文。

Ghosh 和 Veale¹¹⁹ 提出了一种 CNN 和双向 LSTM 混合技术，用于推特中的讽刺检测，该技术模拟了语言和心理背景。

Mishra 等人⁶⁶ 利用 CNN 从眼球运动（或凝视）数据中自动提取认知特征，以丰富讽刺检测的信息。单词嵌入还用于英语推文中的反语识别¹²⁰ 以及辩论中有争议的单词识别。¹²¹

情感分析

情绪是人类的主观感受和思想。主要的情绪包括爱、喜悦、惊讶、愤怒、悲伤和恐惧。情绪的概念与情操密切相关。例如，情绪的强度可以与某种情绪的强度联系起来，如喜悦和愤怒。因此，许多深度学习模型也按照情感分析的方式应用于情感分析。

王等人¹²² 为语码转换情绪预测建立了一个双语注意网络模型。使用 LSTM 模型来构建每个帖子的文档级表示，并使用注意机制来捕获单语和双语上下文中的信息词。

周等人¹²³提出了一种情感聊天机，基于 GRU 对大规模会话生成中的情感影响进行建模，该技术在其他论文中也有应用。

39, 72, 115

Abdul-Mageed 和 Ungar¹²⁴ 首先通过使用远程监督自动建立了一个大型情感检测数据集，然后使用 GRU 网络进行细粒度的情感检测。

Felbo 等人¹²⁵ 使用社交媒体中出现的数百万个表情符号对神经模型进行预训练，以学习更好的情绪背景表示。

提出了一种使用深度记忆网络进行情感原因提取的问答方法。¹²⁶ 情感原因提取旨在识别文本中表达的某种情感背后的原因。

用于情感分析的多模态数据

多模态数据，例如携带文本、视觉和听觉信息的数据，已经被用于帮助情感分析，因为它向传统的文本特征提供了附加的情感信号。由于深度学习模型可以将输入映射到用于特征表示的一些潜在空间，因此来自多模态数据的输入也可以被同时投影以学习多模态数据融合，例如，通过使用特征串联、联合潜在空间或其他更复杂的融合方法。现在越来越多的趋势是将多模态数据与深度学习技术结合使用。

茯苓等人¹²⁷ 提出了一种基于 CNN 内层的激活值从短文本中提取特征的方法。该论文的主要创新点是使用深度 CNN 从文本中提取特征，并使用多核学习 (MKL) 对异构多模态融合特征向量进行分类。

Bertero 等人¹²⁸ 描述了一个 CNN 模型，用于在交互式对话系统的声音数据中识别情感和情绪。

Fung 等人¹²⁹ 演示了一个虚拟交互对话系统，该系统整合了由深度学习模型训练的情绪、情感和个性识别能力。

Wang 等人¹³⁰ 报道了一种 CNN 结构化深度网络，命名为深度耦合形容词和名词 (DCAN) 神经网络，用于视觉情感分类。DCAN 的关键思想是利用形容词和名词文本描述，将它们视为两个 (弱) 监督信号来学习两个中间情绪表征。这些学习到的表示然后被连接并用于情感分类。

Yang 等人¹³¹ 开发了两种基于条件概率神经网络的算法来分析图像中的视觉情感。

朱等¹³² 提出了一个统一的 CNN-RNN 模型的视觉情感识别。该架构利用具有多个层的 CNN 来提取不同级别的特征 (例如，颜色、纹理、对象等。) 在多任务学习框架内。并且提出了双向 RNN 来整合 CNN 模型中不同层的学习特征。

尤等¹³³ 采用了视觉情感分析的注意机制，可以共同发现相关的局部图像区域，并在这些局部区域之上构建情感分类器。

茯苓等人¹³⁴ 提出了一些用于对视频数据进行多模态情感分析和情感识别的深度学习模型。特别地，基于 LSTM 的模型被提出用于话语级

情感分析，可以在同一视频中从他们的周围环境中捕捉上下文信息。

Tripathi 等人¹³⁵ 使用深度和基于 CNN 的模型对多模态数据集 DEAP 进行情绪分类，该数据集包含脑电图和外周生理和视频信号。

扎德等人¹³⁶ 将多通道情感分析的问题公式化为对通道内和通道间动态的建模，并引入了一种称为张量融合网络的新神经模型来解决该问题。

Long 等人¹³⁷ 提出了一种用基于认知的眼动跟踪数据训练的注意力神经模型，用于句子级情感分类。为神经情感分析建立了基于认知的注意层。

Wang 等人¹³⁸ 提出了一种选择相加学习 (SAL) 方法来解决多模态情感分析中的混杂因素问题，该方法去除了由神经网络 (例如，CNN) 学习的个体特定潜在表示。为了实现它，涉及两个学习阶段，即用于混杂因素识别的选择阶段和用于混杂因素去除的去除阶段。

资源贫乏的语言和多语言情感分析

最近，由于深度学习模型的使用，资源贫乏语言 (与英语相比) 的情感分析也取得了重大进展。此外，多语言功能也可以像多模态数据一样帮助进行情感分析。同样的，深度学习已经被应用到多语言情感分析设置中。

Akhtar 等人¹³⁹ 报告了一种基于 CNN 的混合架构，用于在资源贫乏的语言印地语中进行句子和方面级别的情感分类。

Dahou 等人¹⁴⁰ 使用单词嵌入和基于 CNN 的模型在句子级别进行阿拉伯语情感分类。

Singhal 和 Bhattacharyya¹⁴¹ 设计了一种在评论/句子级别进行多语言情感分类的解决方案，并尝试了多种语言，包括印地语、马拉地语、俄语、荷兰语、法语、西班牙语、意大利语、德语和葡萄牙语。作者应用机器翻译工具将这些语言翻译成英语，然后使用英语单词嵌入、情感词典的极性和 CNN 模型进行分类。

Joshi 等人¹⁴² 在 LSTM 体系结构中引入了子词级表示，用于印地语-英语代码混合句子的情感分类。

其他相关任务

深度学习在其他一些与情感分析相关的任务中也有应用。

情感主体间性:Gui 等人¹⁴³ 解决了情感分析中的主体间性问题，该问题是研究语言的表面形式和相应的抽象概念之间的差距，并将主体间性的建模纳入到提议的 CNN 中。

词库扩充:Wang 等人¹⁴⁴ 提出了一种基于 PU 学习的神经方法用于意见词库扩充。

金融波动性预测:Rekabsaz 等人¹⁴⁵ 使用金融披露情绪和基于单词嵌入的信息检索模型进行波动性预测, 其中单词嵌入用于相似单词集扩展。

意见推荐:Wang 和 Zhang¹⁴⁶ 介绍了意见推荐的任务, 其目的是生成特定用户可能给出的产品的定制评论分数, 以及如果用户已经评论了该产品, 用户将为目标产品撰写的定制评论。为了解决这个问题, 提出了一种多重注意记忆网络, 它考虑了用户评论、产品评论和用户邻居(相似用户)。

立场检测:Augenstein 等人¹⁴⁷ 提出了一种双向 LSTMs, 具有条件编码机制, 用于政治 twitter 数据中的立场检测。杜等¹⁴⁸ 为站姿分类设计了目标特定的神经注意模型。

结论

将深度学习应用于情感分析已经成为一个热门的研究课题。本文介绍了各种深度学习架构及其在情感分析中的应用。这些深度学习技术中的许多已经显示了各种情绪分析任务的最先进的结果。随着深度学习研究和应用的进展, 我们相信在不久的将来会有更多令人兴奋的深度学习情感分析研究。

感谢

刘冰和王帅的工作得到了美国国家科学基金会(NSF)的部分资助, 资助号为 IIS1407927 和 IIS-1650900, 并得到了华为技术有限公司的研究资助。

参考

- [1] 刘情感分析:挖掘观点、情感和情绪。剑桥大学出版社, 2015 年。
- [2] 刘。情感分析和观点挖掘(介绍和调查), 摩根和克莱普, 2012 年 5 月。
- [3] 意见挖掘和情感分析。信息检索的基础和趋势, 2008。2(1-2):第 1-135 页。
- [4] Goodfellow I, Bengio Y, 库维尔 a .深度学习。麻省理工学院出版社。2016.
- [5] 深度稀疏整流器神经网络。《人工智能与统计国际会议论文集》(AISTATS 2011), 2011 年。
- [6] 通过反向传播误差学习表征。认知模型, 1988 年。
- [7] Collobert R、Weston J、Bottou L、Karlen M、Kavukcuoglu K 和 Kuksa P .自然语言处理(几乎)从零开始。机器学习研究杂志, 2011 年。
- [8] 自然语言处理的神经网络模型入门。人工智能研究杂志, 2016。
- [9] 本吉奥 Y, 库维尔 A, 文森特 p, 《表征学习:回顾与新视角》。2013 年 IEEE 模式分析与机器智能汇刊。

- [10] Lee H, Grosse R, Ranganath R 和 Ng A.Y, 《用于分级表示的可扩展无监督学习的卷积深度信念网络》。《机器学习国际会议论文集》(2009 年, ICML), 2009 年。
- [11] 神经概率语言模型。机器学习研究杂志, 2003。
- [12] 分层概率神经网络语言模型。2005 年人工智能与统计国际研讨会论文集。
- [13] Mikolov T, Chen K, Corrado G, 和 Dean J. 《向量空间中单词表示的有效估计》. 在...里
学习表征国际会议论文集 (ICLR, 2013 年), 2013 年。
- [14] Mikolov T, Sutskever I, Chen K, Corrado G 和 Dean J . 单词和短语的分布式表示及其组合性。
《神经信息处理系统进展年会论文集》(NIPS 2013), 2013 年。
- [15] 用噪声对比估计有效地学习单词嵌入。《神经信息处理系统进展年会论文集》(NIPS 2013), 2013 年。
- [16] 通过全球语境和多词原型改进词的代表。《计算语言学协会年会论文集》(ACL 2012), 2012 年。
- [17] Pennington J, Socher R, Manning C.D. GloVe: 单词表示的全局向量。《自然语言处理经验方法会议论文集》(EMNLP 2014), 2014 年。
- [18] 深度网络的逐层贪婪训练。《神经信息处理系统进展年会论文集》(NIPS 2006), 2006 年。
- [19] 用神经网络降低数据的维数。科学, 2006 年 7 月。
- [20] Vincent P, Larochelle H, Bengio Y 和 Manzagol P-A, 《用去噪自动编码器提取和合成鲁棒特征》。《机器学习国际会议论文集》(2008 年, ICML), 2008 年。
- [21] 多尺度卷积网络的交通标志识别。《神经网络国际联合会议论文集》(IJCNN 2011), 2011 年。
- [22] Elman J.L. 《及时发现结构》。认知科学, 1990 年。
- [23] Bengio Y, Simard P, Frasconi P . 学习具有梯度下降的长期依赖性困难的。1994 年 IEEE 神经网络汇刊。
- [24] 双向递归神经网络。IEEE 信号处理汇刊, 1997。
- [25] Hochreiter S, Schmidhuber J . 长短期记忆。神经计算, 9(8): 1735-1780, 1997。
- [26] Tai K.S, Socher R, Manning C. D . 改进了树形结构长短期记忆网络的语义表示。《计算语言学协会年会论文集》(ACL 2015), 2015 年。
- [27] 使用统计机器翻译的 RNN 编码器-解码器学习短语表示。《自然语言处理经验方法会议论文集》(EMNLP 2014), 2014 年。

- [28] 门控循环神经网络对序列建模的经验评估。arXiv 预印本 arXiv:1412.3555, 2014。
- [29] Bahdanau D, Cho K, Bengio Y. 通过联合学习对齐和翻译的神经机器翻译。arXiv 预印本 arXiv:1409.0473, 2014。
- [30] 《记忆网络》。arXiv 预印本 arXiv:1410.3916。2014。
- [31] 端到端存储网络。《第 29 届神经信息处理系统会议论文集》(NIPS 2015), 2015 年。
- [32] 神经图灵机。预印本 arXiv:1410.5401。2014。
- [33] 在递归神经网络中学习标签嵌入和标签特定组合函数。《计算语言学协会年会论文集》(ACL 2015), 2015 年。
- [34] 文档级情感分类:SVM 和人工神经网络的实证比较。专家系统及其应用。2013。
- [35] 句子和文件的分布式表示。《机器学习国际会议论文集》(2014 年, ICML), 2014 年。
- [36] Glorot X, Bordes A, Bengio Y, 《大规模情感分类的领域适应:一种深度学习方法》。《机器学习国际会议论文集》(ICML, 2011 年), 2011 年。
- [37] 翟(马克)张。用于情感分析的半监督自动编码器。《AAAI 人工智能会议论文集》(AAAI 2016), 2016 年。
- [38] 卷积神经网络文本分类中词序的有效使用。《计算语言学协会北美分会会议记录:人类语言技术》(NAACL-HLT 2015), 2015 年。
- [39] 唐 D, 秦 B, 刘 t。用于情感分类的门控递归神经网络文档建模。《自然语言处理经验方法会议论文集》(EMNLP 2015), 2015 年。
- [40] 面向文档级情感分类的用户和产品语义表示学习。《计算语言学协会年会论文集》(ACL 2015), 2015 年。
- [41] 陈 H, 孙 M, 涂 C, 林 Y, 刘 z。神经情感分类与用户和产品的关注。在...里
《自然语言处理中的经验方法会议录》(EMNLP 2016), 2016 年。
- [42] 窦 ZY。利用深度记忆网络捕获用户和产品信息, 进行文档级情感分析。《自然语言处理经验方法会议论文集》(EMNLP 2017), 2017。
- [43] 用于文档级情感分类的高速缓存长短期记忆神经网络。《自然语言处理经验方法会议论文集》(EMNLP 2016), 2016 年。
- [44] 杨 z、杨 d、代尔 C、何 X、斯莫拉 AJ、霍维 EH。用于文档分类的分层注意网络。《计算语言学协会北美分会会议记录:人类语言技术》(NAACL-HLT 2016), 2016 年。
- [45] 作为机器理解的文档级多方面情感分类。《自然语言处理经验方法会议论文集》(EMNLP 2017), 2017。
- [46] 周 X, 万 X, 肖 j。基于注意力的跨语言情感分类的网络。在...里
《自然语言处理中的经验方法会议录》(EMNLP 2016), 2016 年。

[47] 用于跨领域情感分类的端到端对抗性记忆网络。《人工智能国际联合会议论文集》(IJCAI 2017), 2017。

[48] 主观分类黄金标准数据集的开发和使用。《计算语言学协会年会论文集》(ACL 1999), 1999 年。

[49] 预测情感分布的半监督递归自动编码器。《自然语言处理经验方法会议论文集》(EMNLP 2011), 2011 年。

[50] 通过递归矩阵向量空间的语义合成。《自然语言处理经验方法会议论文集》(EMNLP 2012), 2012 年。

[51] 情感树库语义合成的递归深度模型。《自然语言处理经验方法会议论文集》(EMNLP 2013), 2013 年。

[52] 用于句子建模的卷积神经网络。在...里
计算语言学协会年会论文集(ACL 2014), 2014。

[53] 用于句子分类的卷积神经网络。《计算语言学协会年会论文集》(ACL 2014), 2014 年。

[54] 《短文本情感分析的深度卷积神经网络》。在...里
计算语言学国际会议论文集(COLING 2014), 2014。

[55] 用长时短时记忆合成词嵌入预测微博的极性。《计算语言学协会年会论文集》(ACL 2015), 2015 年。

[56] 使用双向 LSTM 和其他神经网络架构的逐帧音素分类。神经网络, 2005。

[57] 使用区域 CNN-LSTM 模型进行维度情感分析。《计算语言学协会年会论文集》(ACL 2016), 2016 年。

[58] 结合卷积和递归神经网络的短文本情感分析。《计算语言学国际会议论文集》(COLING 2016), 2016 年。

[59] 基于 CNN 和 LSTM 的在线用户评论中的索赔分类。在...里
计算语言学国际会议论文集(COLING 2016), 2016。

[60] 黄敏, 钱青, 朱翔。情感分类的神经网络句法知识编码。
2017 年美国计算机学会信息系统汇刊

[61] Akhtar MS、Kumar A、Ghosal D、Ekbali A 和 Bhattacharyya P . 一种基于多层感知器的集成技术, 用于细粒度的金融情绪分析。《自然语言处理经验方法会议论文集》(EMNLP 2017), 2017。

[62] 关 Z, 陈 L, 赵 W, 郑 Y, 谭 S, 蔡 d. 用于顾客评论情感分类的弱监督深度学习。《人工智能国际联合会议论文集》(IJCAI 2016), 2016。

[63] 用于神经情感分析的上下文敏感词汇特征。在...里
《自然语言处理中的经验方法会议录》(EMNLP 2016), 2016 年。

- [64] 基于辅助任务的跨领域情感分类句子嵌入学习。《自然语言处理经验方法会议论文集》(EMNLP 2016), 2016 年。
- [65] 赵, 陆, 蔡, 何, 庄。基于递归随机游走网络学习的微博情感分类。《人工智能内部联合会议论文集》(IJCAI 2017), 2017。
- [66] 使用卷积神经网络从凝视数据中学习认知特征用于情感和讽刺分类。《计算语言学协会年会论文集》(ACL 2017), 2017 年。
- [67] 钱清, 黄敏, 雷军, 和朱。《从语言上规范 LSTM 的情感分类》。在…里
计算语言学协会 (ACL 2017) 年会论文集, 2017。
- [68] 董, 魏, 谭, 唐, 周, 徐。自适应递归神经网络用于目标相关的推特情感分类。《计算语言学协会年会论文集》(ACL 2014), 2014 年。
- [69] 具有丰富自动特征的目标相关推特情感分类。在…里
人工智能内部联席会议 (IJCAI 2015) 会议录, 2015。
- [70] 唐 D, 秦 B, 冯 X, 刘 t。用于目标相关情感分类的有效 LSTMs。在…里
计算语言学国际会议论文集 (COLING 2016), 2016。
- [71] 基于方面的情感分析评论的层次模型。在…里
自然语言处理经验方法会议录 (EMNLP 2016), 2016。
- [72] 面向情感分析的门控神经网络。《AAAI 人工智能会议论文集》(AAAI 2016), 2016 年。
- [73] 王 Y, 黄 M, 朱 X, 赵 l。基于注意的方面级情感分类的 LSTM。在…里
《自然语言处理中的经验方法会议录》(EMNLP 2016), 2016 年。
- [74] 杨敏, 涂伟, 王军, 徐峰, 陈曦。基于注意的目标相关情感分类的 LSTM。《AAAI 人工智能会议论文集》(AAAI 2017), 2017 年。
- [75] 面向目标情感的注意建模。《计算语言学协会欧洲分会会议论文集》(EACL 2017), 2017 年。
- [76] 唐 D, 秦 B, 刘 t。基于深记忆网络的体貌层次情感分类。arXiv 预印本 arXiv:1605.08900, 2016。
- [77] 合理化的神经预测。《自然语言处理经验方法会议论文集》(EMNLP 2016), 2016。
- [78] 用于态度识别的深度记忆网络。《美国计算机学会网络搜索和数据挖掘国际会议论文集》(WSDM, 2017), 2017 年。
- [79] 马 D, 李松, 张 X, 王 h。面向方面级情感分类的交互式注意网络。在…里
《人工智能内部联席会议论文集》(IJCAI 2017), 2017。
- [80] 陈平, 孙志军, 李冰, 杨威。用于情状体分析的记忆循环注意网络。《自然语言处理经验方法会议论文集》(EMNLP 2017), 2017。
- [81] 泰伊, 团拉, 回族。用于基于方面的情感分析的三元记忆网络。《信息与知识管理国际会议论文集》(CIKM 2017), 2017 年。
- [82] 研究意见实体和关系的联合提取。在…里
计算语言学协会 (ACL 2016) 年会论文集, 2016。

- [83]用于基于方面的情感分析的递归神经条件随机场。《自然语言处理经验方法会议论文集》(EMNLP 2016), 2016年。
- [84] Wang W, Pan SJ, Dahlmeier D, 和 Xiao .耦合多层注意, 用于方面和观点术语的共同提取。《AAAI 人工智能会议论文集》(AAAI 2017), 2017年。
- [85] 具有记忆交互的体貌术语提取的深度多任务学习。《自然语言处理经验方法会议论文集》(EMNLP 2017), 2017。
- [86]用于特征提取的无监督神经注意模型。在…里
计算语言学协会 (ACL 2017) 年会论文集, 2017。
- [87]面向开放领域情感的神经网络。《自然语言处理经验方法会议论文集》(EMNLP 2015), 2015年。
- [88] 周 X, 万 X, 肖 j. 面向在线评论体貌类别检测的表征学习。在…里
2015 年 AAAI 人工智能会议论文集 (AAAI 2015)。
- [89]尹 Y, 魏 F, 董 L, 徐 K, 张 M, 和周 M. 无监督的词和依存路径嵌入的方面术语提取。《人工智能国际联合会会议论文集》(IJCAI 2016), 2016。
- [90]熊松, 张艳, 纪丹, 楼艳。体短语分组的距离度量学习。《计算语言学国际会议论文集》(COLING 2016), 2016年。
- [91]用深度卷积神经网络进行观点挖掘的方面提取。基于知识的系统杂志。2016。
- [92] 英 D, 于 J, 蒋 J. 用于跨领域意见目标抽取的带辅助标签的递归神经网络。《AAAI 人工智能会议论文集》(AAAI 2017), 2017年
- [93]用深度递归神经网络进行观点挖掘。《自然语言处理经验方法会议论文集》(EMNLP 2014), 2014年。
- [94] 用半马尔可夫条件随机场抽取意见表达。在…里
自然语言处理经验方法会议录 (EMNLP 2012), 2012。
- [95] 基于递归神经网络和词嵌入的细粒度观点挖掘。《自然语言处理经验方法会议论文集》(EMNLP 2015), 2015年。
- [96]语言合成的深度递归神经网络。《神经信息处理系统进展年会论文集》(NIPS 2014), 2014年。
- [97] 用于在情感合成中整合合成和非合成情感的神经网络。《计算语言学协会北美分会会议记录: 人类语言技术》(NAACL-HLT 2015), 2015年。
- [98] 杨 B, Cardie C . 用于细粒度意见抽取的联合推理。《计算语言学协会年会论文集》(ACL 2013), 2013年。
- [99] 基于意见类型的新分类识别意见来源。在…里
人工智能国际联合会会议论文集 (IJCAI 2016), 2016。
- [100]基于内容的影响模型在意见行为预测中的应用。在…里
计算语言学国际会议论文集 (COLING 2016), 2016。
- [101]拉什金 H, 贝尔 E, 崔 Y 和沃尔科娃 S, 《多语言内涵框架: 用于目标情感分析和预测的社交媒体案例研究》。《计算语言学协会年会论文集》(ACL 2017), 2017年。

- [102] 学习用于情感分析的词向量。《计算语言学协会年会论文集》(ACL 2011), 2011 年。
- [103] Bespalov D, Bai B, Qi Y 和 Shokoufandeh A . 基于监督潜在 n 元语法分析的情感分类。《信息与知识管理国际会议论文集》(CIKM 2011), 2011 年。
- [104] 拉布托夫 I, 利普森 h. 重新嵌入单词。《计算语言学协会年会论文集》(ACL 2013), 2013 年。
- [105] 唐 D, 魏 F, 杨 N, 周 M, 刘 T, 秦 b. 《计算语言学协会年会论文集》(ACL 2014), 2014 年。
- [106] 情感嵌入及其在情感分析中的应用。2016 年 IEEE 知识与数据工程汇刊。
- [107] 基于分层情感监督的表征学习情感词典构建。《自然语言处理经验方法会议论文集》(EMNLP 2017), 2017。
- [108] 于立春, 王军, 赖克瑞, 张. 情感分析中词嵌入的改进。《自然语言处理经验方法会议论文集》(EMNLP 2017), 2017。
- [109] Li J, Jurafsky D . 多义项嵌入提高自然语言理解吗? 《自然语言处理经验方法会议论文集》(EMNLP 2015), 2015 年。
- [110] 使用主题丰富的多原型词嵌入改进推特情感分类。《AAAI 人工智能会议论文集》(2016 年, AAAI), 2016 年。
- [111] 学习双语情感词嵌入用于跨语言情感分类。《计算语言学协会年会论文集》(ACL 2015), 2015 年。
- [112] 探索基于方面的跨语言情感分类的分布表示和机器翻译。《第 27 届计算语言学国际会议论文集》(COLING 2016), 2016 年。
- [113] 基于评论的多层嵌入学习在评分预测中的应用。《人工智能国际联合会议论文集》(IJCAI 2016), 2016。
- [114] Sharma R, Somani A, Kumar L, 和 Bhattacharyya P . 使用情感词嵌入的形容词中的情感强度排序。《自然语言处理经验方法会议论文集》(EMNLP 2017), 2017。
- [115] 基于深度神经网络的微博挖苦检测。《计算语言学国际会议论文集》(COLING 2016), 2016 年。
- [116] Joshi A, Tripathi V, Patel K, Bhattacharyya P 和 Carman M . 基于单词嵌入的特征对讽刺检测有用吗? 《自然语言处理经验方法会议论文集》(EMNLP 2016), 2016。
- [117] 使用深度卷积神经网络深入观察讽刺性的推文。《计算语言学国际会议论文集》(COLING 2016), 2016 年。

[118] 讽刺符号:用基于情感的单语机器翻译解释讽刺。《计算语言学协会年会论文集》(ACL 2017), 2017 年。

[119] 《讽刺的磁铁:使讽刺检测及时、有背景且非常个人化》。《自然语言处理经验方法会议论文集》(EMNLP 2017), 2017。

[120] 周一早晨是我的最爱:)#不探索英语推文中讽刺的自动识别。《计算语言学国际会议论文集》(COLING 2016), 2016 年。

[121] 陈, 林风云, 卢伟。WordForce:可视化辩论中有争议的词语。《自然语言处理经验方法会议论文集》(EMNLP 2016), 2016 年。

[122] 用于语码转换情绪预测的双语注意网络。《计算语言学国际会议论文集》(COLING 2016), 2016 年。

[123] 周 H, 黄 M, 张 T, 朱 X, 刘 b。情绪聊天机:具有内外记忆的情绪对话生成。arXiv 预印本。arXiv:1704.01074, 2017。

[124] Abdul-Mageed M, Ungar L. EmoNet:用门控递归神经网络进行细粒度情感检测。《计算语言学协会年会论文集》(ACL 2017), 2017 年。

[125] Felbo B, Mislove A, gaard A, Rahwan I 和 Lehmann S。使用数百万个表情符号来学习任何领域的表示, 以检测情绪、情感和讽刺。《自然语言处理经验方法会议论文集》(EMNLP 2017), 2017。

[126] 桂琳, 胡军, 何燕, 徐瑞, 陆青, 杜军。一种情感原因抽取的问答方法。在...里
自然语言处理经验方法会议录 (EMNLP 2017), 2017。

[127] 用于话语级多模态情感分析的深度卷积神经文本特征和多核学习。《自然语言处理经验方法会议论文集》(EMNLP 2015), 2015 年。

[128] 用于交互式对话系统的实时语音情感和情感识别。《自然语言处理经验方法会议论文集》(EMNLP 2016), 2016 年。

[129] 冯平、戴、西迪基、林、杨、贝泰罗、万、陈瑞和吴。Zara:一个融合了情感、情绪和个性识别的虚拟互动对话系统。《计算语言学国际会议论文集》(COLING 2016), 2016 年。

[130] 超越物体识别:用深度耦合的形容词和名词神经网络进行视觉情感分析。《人工智能内部联合会议论文集》(IJCAI 2016), 2016。

[131] 基于增广条件概率神经网络的视觉情感分布学习。《AAAI 人工智能会议论文集》(AAAI 2017), 2017 年。

[132] 朱 X, 李 L, 张 W, 饶 T, 徐 M, 黄 Q, 徐 d。依赖开发:一种统一的 CNN-RNN 视觉情感识别方法。《人工智能内部联合会议论文集》(IJCAI 2017), 2017。

[133] 基于局部图像区域的视觉情感分析。《AAAI 人工智能会议论文集》(AAAI 2017), 2017 年。

[134] 茯苓、柬埔寨、哈扎里卡、马朱曼德、扎德和莫伦西。用户生成视频中的上下文相关情感分析。《计算语言学协会年会论文集》(ACL 2017), 2017 年。

- [135] 使用深度和卷积神经网络对 DEAP 数据集进行精确的情感分类。《AAAI 人工智能会议论文集》(AAAI 2017), 2017 年。
- [136] 扎德 A, 陈 M, 茯苓 S, 柬埔寨 E 和 Morency LP。用于多模态情感分析的张量融合网络。《自然语言处理经验方法会议论文集》(EMNLP 2017), 2017。
- [137] 龙宇, 秦, 向, 李, 黄。基于认知的情感分析注意模型。在...里
自然语言处理经验方法会议录 (EMNLP 2017), 2017。
- [138] 选择-加法学习: 提高多模态情感分析的泛化能力。《多媒体与博览会国际会议论文集》(ICME 2017), 2017 年。
- [139] 用于情感分析的混合深度学习架构。《计算语言学国际会议论文集》(COLING 2016), 2016 年。
- [140] 用于阿拉伯语情感分类的词嵌入和卷积神经网络。《计算语言学国际会议论文集》(COLING 2016), 2016 年。
- [141] 从你富有的表亲那里借用一点: 使用英语单词的嵌入和极性进行多语言情感分类。《计算语言学国际会议论文集》(COLING 2016), 2016 年。
- [142] Joshi A、Prabhu A、Shrivastava M 和 Varma V。对印地语-英语代码混合文本进行情感分析的子词级组合。《计算语言学国际会议论文集》(COLING 2016), 2016 年。
- [143] 桂 L, 徐 R, 何 Y, 陆 Q, 魏 z。主体间性与情感: 从语言到知识。在...里
人工智能国际联合会议论文集 (IJCAI 2016), 2016。
- [144] 基于神经 PU 学习、双词典查找和极性联想的情感词典扩充。《自然语言处理经验方法会议论文集》(EMNLP 2017), 2017。
- [145] 雷卡巴萨兹 N, 卢普 M, 巴克拉诺夫 A, 汉伯里 A, 杜尔 A 和安德森 I。使用基于单词嵌入的 IR 模型的财务披露情绪进行波动性预测。《计算语言学协会年会论文集》(ACL 2017), 2017 年。
- [146] 基于神经网络模型的意见推荐。《自然语言处理经验方法会议论文集》(EMNLP 2017), 2017。
- [147] Augenstein I、rocktschel T、Vlachos A、Bontcheva K。利用双向条件编码进行姿态检测。
《自然语言处理经验方法会议论文集》(EMNLP 2016), 2016 年。
- [148] 杜军, 徐瑞, 何燕, 桂莉。基于目标特定神经注意网络的姿态分类。《人工智能内部联合会议论文集》(IJCAI 2017), 2017。