

方面级情感分类的深度学习:调查、愿景和挑战

Jie Zhou, Jimmy Xiangji Huang, Qin Chen, Qinmin Vivian Hu,
Tingting Wang, Liang He

华东师范大学上海市多维信息处理重点实验室

摘要: 本调查聚焦于基于深度学习的方面级情感分类(ASC), 旨在确定文档中提到的某个方面的情感极性。随着深度学习在许多应用中的成功应用, 基于深度学习的 ASC 近年来引起了学术界和工业界的极大兴趣。然而, 仍然缺乏对现有方法的系统分类和对其性能的比较, 这是我们的调查旨在填补的空白。此外, 为了定量评估各种方法的性能, 评估方法和共享数据集的标准化是必要的。本文对当前最先进的基于深度学习的方法进行了深入的概述, 展示了 ASC 已经取得的巨大进步。特别是, 首先, 提供了对基于深度学习的 ASC 的最新研究成果的全面回顾。更具体地说, 我们设计了一个基于深度学习的 ASC 分类法, 并提供了最先进方法的全面总结。然后, 我们收集了所有的基准 ASC 数据集, 供研究人员使用各种常用的评估方法, 在五个公共标准数据集上进行广泛的研究和实验。最后, 我们讨论了一些最具挑战性的开放问题, 并指出了该领域未来的研究方向。

关键词: 基于方面的情感分析; 方面级情感分类; 注意力; 卷积神经网络; 深度学习; 记忆网络; 神经网络; 循环神经网络

1 相关工作

情感分析可以分为三个层次, 即文档级、句子级和方面级[1]。文档级情感分析假设整个文档只包含对一个主题的观点。显然, 这在很多情况下是不合理的。句子级情感分析同样假设一个句子中只表达一个主题。然而, 通常情况下, 一句话包含多个主题(即方面)或同一句子中的观点相反。对于文档级和句子级情感分析, 所决定的情感极性基于整个文档/句子, 而不是文档/句子中给出的主题。相比之下, 方面级别的情感分析旨在判断为正在讨论的每个方面表达的情感极性。这允许利用评论/推文提供的更多信息进行更详细的分析。

作为情感分析的一个基本子任务[2], 方面级情感分析受到了业界和学术界的广泛关注。

图 1 一个方面级情感分类的例子

最近,方面级情感分析是语义网和计算语言学研究领域关注的焦点[3]-[5]。方面级情感分析的目标是识别方面(方面提取)并推断为每个方面表达的情感(也称为方面级情感分类)。本文主要研究基于深度学习的方面级情感分类问题。这使我们能够涵盖更多最近的发展,而不是重复其他调查[1],[6]-[8]中提供的既定见解。显而易见,ASC 的深度学习领域正在蓬勃发展。ASC 的目标是确定用户在评论/推文中对给定方面(方面类别或方面术语)表达的意见是积极的、消极的还是中立的[1]。图中的一个例子 1 呈现取自 SemEval 2014 [3]数据集(例如 Restaurants14)的示例句子。方面类别隐含地描述了实体的一般类别。例如,在“三文鱼好吃,服务员很粗鲁”的帖子中,用户分别对“食物”和“服务”两个方面表达了积极和消极的情绪。一个方面术语表征一个明确出现在句子中的特定实体。对于同一个句子,方面术语是“鲑鱼”和“服务员”,其中用户分别对它们表达积极和消极的情绪。就方面粒度而言,方面类别是粗粒度的,而方面术语是细粒度的[9]。

近年来,在传统情绪分析领域已经发表了几项调查。例如,在 2008 年,Pang 等人[6]对感觉分析进行了很好的回顾。讨论了各种技术和应用,涵盖了伦理、理论和实践方面的考虑。然而,他们主要关注用于文档级情感分析的传统机器学习方法。在 2009 年,Tang 等人[8]介绍了一项调查,该调查也主要关注用于消费者评论领域的文档级情感分析的机器学习方法。2011 年,Tsytsarau 和 Palpanas [7]发布了一项调查,也侧重于文档级别的情感分析,并讨论了预测情感极性的四种不同方法,即分别基于机器学习、基于词典、基于语义和基于统计的方法。2012 年,Liu[1]提出了一项调查,该调查介绍了情绪分析的整个领域。在关于方面级情感分析的部分中给出了实现实际解决方案时的一系列子问题:1)方面提取的定义,包括各种挑战,如解决隐性和显性的实体和意见;2)如何识别方面和情感极性并相互联系(又名 ASC)。与我们的调查工作最相关的是[10]侧重于方面层面的感觉分析。对传统的用于方面级情感分析的机器学习方法进行了深入的综述,包括方面抽取和 ASC。然而,尽管近年来基于深度学习的自主学习取得了很大的成功,但现有的工作都没有集中在这方面。此外,缺乏对基于深度学习的 ASC 方法的系统分类及其在基准数据集上的性能报告,这是本次调查旨在填补的空白。

本调查旨在全面回顾基于深度学习的自主学习的研究进展。它提供了一个概述,读者可以通过它快速理解和进入基于深度学习的 ASC 领域。这项调查服务于对 ASC 感兴趣的研究人员、实践者和教育工作者,希望他们在选择深度学习模型来解决手头的 ASC 任务时有一个大致的指导方针。综上所述,本次调查与以往调查的区别包括:1)据我们所知,这是第一次很好地总结了基于 ASC 的深度学习领域,并整理了现有的工作和当前

的进展。2)我们收集和分析了几乎所有的 ASC 基准数据集;3)我们实现了经典的最先进的模型,并用广泛使用的评估方法在五个经典数据集上对它们进行评估。

本文的主要贡献可以总结如下。

- 1) 鉴于目前关于高级计算机深度学习的研究日益增多,有必要对现有的文献进行全面的总结。为此,我们首先对有代表性的方法进行了详细的回顾,并用分类方案总结了当前的工作。现有的方法分为五大类:用于 ASC 的 RecNN,用于 ASC 的 RNN,用于 ASC 的基于注意力的 RNN,用于 ASC 的 CNN 和用于 ASC 的记忆网络。
- 2) 这项调查为 ASC 提供了最全面的现代深度学习技术概述。对于每一种 ASC 模型,我们都详细描述了有代表性的算法,并对相应的算法进行了必要的比较和总结。
- 3) 我们收集了几乎所有的标准 ASC 数据集,包括 SemEval 2014、SemEval 2015、SemEval 2016、Twitter 等。我们还详细探索了这些数据集,并将其翻译成统一的 XML/JSON 格式,供研究人员研究。
- 4) 请注意,现有模型在不同数据集上用不同的指标验证有效性这些模型的实验设置是不同的。我们重现了经典的先进的深度学习方法,并在超过公共基准的 ASC 数据集上用常用的度量标准(如准确度、查准率、召回率、宏观 F1 分数等)来评估它们的性能。
- 5) 我们讨论了基于深度学习的自主学习的重要挑战和开放问题,这为进一步的研究提供了一条充满希望的途径和启发。

2 基于深度学习的 ASC 的综述

在我们深入研究这项调查的细节之前,我们首先讨论一下将深度神经网络引入 ASC 的原因和动机。我们还介绍了基于深度学习的 ASC 技术的基本术语和概念。

基于方面的情感分析是情感分析研究领域的一项基本任务[3],[11],包括几个关键子任务:方面提取[12]-[14]、观点识别[15],[16]和 ASC[17]-[19]。先前的一些研究试图联合解决这些子任务[20],[21],将大部分研究工作用于处理单个子任务。在本研究中,我们着重于解决 ASC 问题的深度学习方法。与文档级和句子级情感分类不同,情感分类同时考虑情感和目标信息,因为情感总是有目标的。如上所述,目标通常是实体或实体的一个方面。为简单起见,实体和方面通常都称为方面。给定一个句子和一个方面,ASC 旨在推断句子对给定方面的情感极性/取向。

ASC 的传统方法大多是基于词汇和句法特征的传统机器学习模型[17],[18],[22]。这类模型的性能高度依赖于人工特征工程的质量。因此,最近的研究已经将注意力转向开发端到端的深度神经网络模型。为了深入了解大量提出的基于深度学习的 ASC 方法,根据所采用的深度学习技术的类型进行分类,将所有方法分为以下五类:用于 ASC 的递归神经网络(RecNN)、用于 ASC 的循环神经网络(RNN)、用于 ASC 的基于注意力的 RNN、用于 ASC 的卷积神经网络(CNN) 和用于 ASC 的记忆网络。

图 2 ASC 深度学习方法的分类

图 2 总结了分类方案。此外,表 1 列出了所有经过审查的方法,这些方法按照上述分类方案进行组织。我们在下面的章节中对每个类别进行了简单的介绍,这些方法的细节将在章节 III 中提供。

A. 用于 ASC 的 RECNN

递归神经网络(RecNN) [56]是一种应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。它可以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,通过组合标记来获得短语的表示,最终是整个句子。然后,句子的表示被用于对给定的输入句子进行最终预测(例如,情感分类)。Dong 等人[23]和 Nguyen 和 Shirai[24]将基于树的 RecNN 引入 ASC。

B.用于 ASC 的 RNN

循环神经网络(RNN)已被证明在许多(语言)序列学习问题中是强大的。此外,大多数先进的 ASC 方法都基于 RNN[25]-[28]。在这个类别中,模型可以分为三个子类别:RNN、双向 RNN(Bi-RNN)和分级 RNN (HRNN)。为了更灵活地描述体与其上下文词之间的语义关系,Tang 等人[26]提出了一个目标依赖的(TD-LSTM)和一个目标连接的(TC-LSTM)来扩展,将应用到 ASC 中。此外,情感相关概念的常识知识被纳入 ASC 的 LSTM 模型的端到端训练中[27]。Zhang 等人[25]使用门控神经网络结构通过 Bi-RNN 对句子中的句法和语义以及方面与其周围上下文词之间的交互进行建模。HRNN 被 Ruder 等人采用。[28],他提出对 ASC 使用分层双向 LSTM 模型,该模型能够学习句内和句间关系。

表 1 ASC 现有公共方法的统计

C. 用于 ASC 的基于注意力的 RNN

注意力机制[57]已成功应用于许多自然语言处理(NLP)任务[58],如神经机器翻译[57], [59], 问答[60],[61], 以及机器理解[62],[63]。最近,各种基于注意力的 RNN 模型

被引入到 ASC 中, 它可以有效地关注句子中针对给定方面的重要部分。用于 ASC 的基于注意力的 RNN 模型可以分为基于基本注意力的 RNN 模型和基于交互注意力的 RNN 模型。特别是, 有大量的研究集中在优化基于基本注意力的 RNN 模型[27],[31]-[33],[35]-[37],[41],[55]。Wang 等人[29]提出了一种基于注意力的方面嵌入的 LSTM 方法, 该方法被证明是一种有效的增强模型以捕获感兴趣部分的方法。此外, 基于交互注意力的模型被广泛用于 ASC[38]-[40],[42],[43]。例如, Ma 等人[38]提出了一种交互注意机制, 从特定的方面和语境中交互学习注意。

D. 用于 ASC 的 CNN

卷积神经网络(CNN)[64]擅长捕获局部模式, 在自然语言处理[65]中起着重要作用。CNN 能够从一个句子中提取本地和全局的代表量。有些工作采用了用于 ASC 的适应性的 CNN[45]-[47]。具体来说, Huang 和 Carley [45]利用参数化滤波器和参数化门将方面信息并入 CNN。Li 等人[47]采用邻近策略来利用单词和方面之间的位置相关性来缩放卷积层的输入。Fan 等人[46]提出了一种卷积存储网络, 该网络结合了注意力机制来捕获句子中的单词和多单词表达。此外, Xue 和 Li[44]提出了一个基于卷积神经网络和门控机制的模型。

E. 用于 ASC 的记忆网络

记忆网络[61]在 ASC [19],[48]-[55],[66]中获得了巨大的成功。Tang 等人[19]首先介绍了一种用于 ASC 的端到端记忆网络, 该网络采用了一种带有外部存储器的注意机制来捕捉句子中与给定方面相关的重要信息。Chen 等人[49]提出了一种基于记忆网络的各方面重复注意机制, 以捕捉相隔很远的感知信息。在[55]中, 提出了句子级内容注意机制来克服记忆模型的近视问题。

3 基于深度学习的 ASC

在本节中, 首先给出问题定义和符号。然后, 我们重点介绍了基于深度学习的 ASC 的最新研究成果, 以确定近年来最显著和最有希望的进展。

A. 问题定义与符号

给定一个句子-方面对 (S, A) , 其中方面 $A = \{w_{start}, t_{start+1}, \dots, w_{end-1}, w_{end}\}$ 是由 n 个字组成的句子 $S = \{w_1, w_2, \dots, w_n\}$, 开始和结束是由 $m=end-start+1$ 个字组成的方面 A 的开始和结束的索引。ASC 的目标是预测句子 S 朝向方面 A 的情感极性 $c \in C = \{N, O, P\}$, 其中 N, O 和 P 分别表示“消极”、“中性”和“积极”的情感极性。

对于句子 $S = \{w_1, w_2, \dots, w_n\}$ 和方面单词 $A = \{w_{start}, t_{start+1}, \dots, w_{end-1}, w_{end}\}$, 我们将每个单词映射到它的嵌入向量 $X = \{x_1, x_2, \dots, x_n\}$ 和 $V = \{v_{start}, v_{start+1}, \dots, v_{end-1}, v_{end}\}$ 。它将

单词表示从高维稀疏向量空间(例如一维编码向量空间)映射到低维密集向量空间。一种常用的单词嵌入方法是 Word2Vec,² 其中包含连续单词包模型(CBOW) [67]和跳过语法模型(SG) [68]。另一种常用的学习方法是全局向量(GloVe)³ [69], 它是在全局词-词共现矩阵的非零条目上训练的。我们在表 2 中总结了表中常用的符号。

B. 用于 ASC 的 RECNN

在这一节中, 我们首先介绍了基本的递归神经网络模型, 然后详细描述了基于树的递归神经网络在自动驾驶系统中的研究。

表 2 常用符号

1) RECNN

我们首先简要描述递归神经网络(RecNN) [56]。RecNN [56]是一类可以学习有向无环图结构化输入(例如树结构)的架构。作为循环神经网络[70]的推广, RecNN 具有一种特定的树结构。通过基于解析树的结构表示, 如句子级情感分析[71], [72]和对位短语检测[73], RecNN 已被成功地用于在 NLP 中建模组合性。基于 RecNN 和解析树, Socher 等人[56]提出了一种短语级情感分析方法, 其中解析树中的每个节点都被分配了一个情感标签。给定句子的结构有向无环图(例如, 解析树), RecNN 以拓扑顺序访问节点, 并以自下而上的策略递归生成父表示, 该策略结合标记来生成短语表示, 最终生成整个输入句子。然后, 句子表示被用于进行最终分类(例如, 情感分类)。图 3 给出了一个 RecNN 中矢量合成的实例过程。节点“非常好”的向量表示是从节点“非常”和节点“好”的向量表示生成的。同样, 节点“不太好”是由短语“节点很好”和单词“节点不太好”生成的。

2) 用于 ASC 的基于树的 RECNN

Dong 等人[23]首次将 RecNN 应用于自动控制系统, 提出了一种自适应递归神经网络(AdaRNN)用于目标相关的推特情感分类。

图 3 RecNN 模型的框架

AdaRNN 学习根据上下文和句法结构预测单词对方面的情感极性。根节点的表示被输入到 softmax 分类器中, 以预测情感极性的分布。对于包含目标方面的给定句子, 从句子的依赖树构建二元依赖树。直观地说, 它表示与方面相关联的句法关系。二元依存关系树中的每个单词(叶子)或短语(内部节点)都表示为一个 d 维向量。父节点 v 的表示

是通过经由 RecNN 中的全局函数 g 自下而上地组合左子向量表示 v_l 和右子向量表示 v_r 来计算的:

公式 (1)

其中 v_l 、 v_r 是其左右子的向量表示, g 和 f 是合成函数和非线性函数(例如, \tanh 、 sigmoid 、 softsign)分别。 $W \in \mathbb{R}^{\text{dim}_w \times 2\text{dim}_w}$ 为参数矩阵, b 表示偏差向量。

并非仅仅使用全局函数 G , AdaRNN 选择 n 个组成函数 $G=\{g_1, \dots, g_n\}$ 基于语言标签和组合向量, 如下所示:

公式 (2)

其中 $P(g_i|v_l, v_r, e)$ 表示给定外部特征向量 e 和孩子 v_l 、 v_r 的向量表示的函数 g_i 的概率。概率计算如下:

公式 (3)

其中 $\beta \in \mathbb{R}$ 是超参数, R 表示参数矩阵。

二叉依赖树的根节点的向量表示(作为目标方面的表示)被馈送到 softmax 函数以推断给定方面的情感极性。

此外, Nguyen 和 Shirai [24] 提出了一个 PhraseRNN 模型来判断句子对给定方面的情绪, 这表明 RecNN 可以有效地从递归结构中获得句子表示。该模型从“目标依赖二元短语依赖树”中获得了一个方面的表示, 该树由依赖树和构成树的组合构成。与 AdaRNN 不同的是, PhraseRNN 没有使用全局函数 G 的列表, 而是在内部短语中使用了 $G=\{g_1, \dots, g_n\}$ 和在外部分短语中使用了 $H=\{h_1, \dots, h_m\}$ 这两种组合函数, 其中 n 和 m 分别是 G 和 H 中的函数数。具体来说, 首先将依赖树转化为短语依赖树。然后将短语依赖树转化为目标依赖二元短语依赖树。

图 4 RNN 模型的框架

然而, 这些现有的用于 ASC 的 RecNN 模型可能遭受在实践中常见的语法分析错误[22], [25]。

C. 用于 ASC 的 RNN

在本节中, 我们描述了 RNN 模型的细节, 包括标准的 RNN、LSTM 和 GRU。我们还回顾了用于 ASC 的最原始的 RNN 方法, 这些方法可以分为基于 RNN 的 ASC、基于 Bi-RNN 的 ASC 和基于 HRNN 的 ASC。请注意, 我们在这里只介绍最经典的用于 ASC 的 RNN 方法, 因为 RNN 被广泛应用于 ASC。为了更好地理解现有工作中使用的 RNN 方法, 我们在表 1 中列出了每个工作的模型类型(例如 LSTM、GRU、Bi-GRU 和 Bi-LSTM)。

1) RNN

我们首先提供一个对循环神经网络(RNN) [70]模型的简要描述。RNN 对序列化输入(例如, 句子中的单词序列)进行建模, 一个 RNN 的基本工作框架如图 4 所示。RNN 是一种典型的具有循环连接的神经网络, 它允许某种形式的记忆。这也使得它适用于具有任意时空维度的序列预测问题。因此, 许多 NLP 任务采用 RNN 的结构, 将句子的解释视为分析一系列表征。给定一个句子 S , 通过 RNN 的馈入输入 $X = [x_1, x_2, \dots, x_n] \in \mathbb{R}^{n \times \dim_w}$, 我们可以得到一个连续的隐藏状态 $H = [h_1, h_2, \dots, h_n] \in \mathbb{R}^{n \times \dim_h}$, 其中 \dim_w 和 \dim_h 分别表示单词嵌入的维数和隐藏状态。特别地, RNN 可以分为以下三类。

- 1) 标准 RNN 标准 RNN [70]是 RNN 的基本工作框架。标准 RNN 的过渡函数是一个线性层, 后跟一个非线性层(例如 \tanh)。网络在时刻 t 的输入为 x_t , h_t 代表同一时间步长的隐藏状态。 h_t 的计算如下:

公式 (4)

因此, h_t 通过当前输入 x_t 和上一时刻 h_{t-1} 的隐藏状态计算。函数 f 表示非线性变换函数(如 \tanh 、ReLU)。U、V、W 是跨时间共享的标准 RNN 权重。 x_t 是 NLP 中典型的词的向量表示。如前所述, 它可以被认为是网络的存储元件, 用于积累来自其他时间步长的信息。然而, 这个标准的 RNN 在实践中很难学习和调整参数, 因为它受到臭名昭著的消失梯度问题的困扰。

- 2) LSTM 长期短期记忆网络(LSTM) [74]是一种特殊类型的 RNN, 它能够学习长期依赖。类似地, 从当前输入 x_t 和先前隐藏状态 h_{t-1} 的非线性变换函数计算时刻 t 处的隐藏层 h_t 。然后使用隐藏状态 h_t 计算输出 y_t 。 h_t 可以看作是总结过去的一种表示, 用于对当前输入做出最终决定。除了隐藏状态向量, LSTM 还有一个存储单元结构, 它由三个门组成: 一个输入门、一个遗忘门和一个输出门。输入门用于指示存储单元受新输入影响的程度; 遗忘门控制存储单元中先前信息被遗忘的程度; 并且输出门控制存储单元影响当前隐藏状态的程度。所有这三个门都取决于明显的隐藏状态和当前输入。具体来说, LSTM 单元计算如下:

公式 (5)

公式 (6)

公式 (7)

公式 (8)

公式 (9)

公式(10)

其中 \odot 代表逐元素乘法, σ 是 sigmoid 函数。 W_i 、 b_i 是输入门的参数, W_f 、 b_f 是遗忘门的参数, W_o 、 b_o 是输出门的参数。更多 LSTM 的细节详见 Graves[75]和 Greff 等人[76]。

- 3) GRU 门控循环单元(GRU) [59]是一个较新的框架, 它类似于 LSTM 模型, 但更简单, 参数更少。从经验上看, GRU 的表现与 LSTM 相当, 尽管它相对简单[77]。与 LSTM 不同的是, 代替存储单元, GRU 使用更新门来控制隐藏门将被更新多少, 并使用复位门来控制信息如何被更新到隐藏状态, 以及控制先前的隐藏状态将影响当前隐藏状态多少。GRU 状态可以计算如下:

公式 (11)

公式 (12)

公式 (13)

公式 (14)

其中 W_z 、 U_z 、 b_z 、 W_r 、 U_r 、 b_r 是更新门和复位门的参数。

2) 基于 RNN 的 ASC

RNN 在 ASC 中扮演着重要的角色。Tang 等人[26]首先将 LSTM 引入到 ASC 中, 因为它可以更灵活地捕捉体与其语境词之间的语义关系。通过考虑方面目标, 他们提出了依赖目标的 LSTM (TD- LSTM)和目标连接的 LSTM (TC-LSTM)来扩展 LSTM。如图 5 所示, TD-LSTM 分别利用两个 LSTM 网络, 即 $LSTM_L$ 和 $LSTM_R$, 从给定方面的左右上下文中学习表示。之后, 他们将 $LSTM_L$ 和 $LSTM_R$ 的最后一个隐藏向量连接起来, 并将其馈送到 softmax 层, 以预测句子对方面的情感极性。为了捕捉方面与其背景之间的相互作用, 提出了 TC-LSTM。它扩展了 TD-LSTM, 加入了一个方面连接组件, 在构建句子的表示时, 它明确地利用了方面和每个上下文单词之间的连接。给定的方面被视为一个特性, 并与 ASC 的上下文特性连接在一起。

图 5 TD-LSTM 模型的框架(来自[26])

Ma 等人[27]将情感相关概念的常识知识纳入 ASC 模型的端到端训练。LSTM 模型通过将常识知识整合到门机制中得到扩展。他们假设情感概念对于控制词级信息的信息是有意义的。例如, 多词方面“腐烂的鱼”可能暗示“腐烂”一词是“鱼”一词的情感相关限定词, 因此在下一个时间步骤中需要过滤掉的信息较少。因此, 为了过滤信息, 知识概念被纳入标准 LSTM 的遗忘、输入和输出门。输入门使用情感概念来防止存储单

元受到与知识冲突的输入标记的影响。同样,输出门利用这样的知识过滤掉存储在存储器中的无关信息。

3) 基于 BI-RNN 的 ASC

双向 RNN (Bi-RNN)是基于时间刻 t 的输出应该依赖于句子中先前和未来的内容的思想。如图 6 所示,Bi-RNN 由两部分 RNN 组成:一个从 w_1 到 w_n 读句子 S 的正向 \overrightarrow{RNN} , 和一个从 w_n 到 w_1 的反向 \overleftarrow{RNN} 。

公式 (15)

公式 (16)

其中 θ_{RNN} 代表 RNN 模型的参数。单词的最终上下文感知表示是通过连接两个隐藏状态向量,即 $h=[\vec{h}_i, \overleftarrow{h}_i]$ 。作为 Bi-RNN 的两个典型类别,双向 GRU (Bi-GRU)和双向 LSTM (Bi-LSTM) [75]在 ASC 中得到了广泛应用并取得了巨大成功。

图 6 Bi-RNN 模型的框架

Zhang 等人[25]提出了一种门控递归神经网络(GRNN)来建模句子中的句法和语义以及体与其周围环境之间的相互作用。广义回归神经网络的框架如图 7 所示。该模型采用 Bi-RNN(如 Bi-GRU)来克服池函数的弱点。为此,提出了两种门控神经网络。首先,它利用 Bi-GRU 来连接句子中的单词,以便在隐藏状态上应用汇集函数,而不是单词嵌入,从而更好地表示方面及其上下文。第二,使用三路门控神经网络结构来模拟句子中所提到的方面与其周围环境之间的相互作用。门控神经网络已被证明通过更好地传播梯度来减少标准 Bi-GRU 朝向句子结尾的偏差。

4) 基于 HRNN 的 ASC

分层 RNN (HRNN)模型主要用于段落和文档的表示学习。

图 7 广义回归神经网络模型的框架(来自[25])

图 8 LSTM 模型的框架

Ruder 等人[28]还提出对 ASC 使用分层双向 LSTM (H-LSTM)模型,该模型能够利用句内和句间关系。如图 8 所示,单词嵌入被输入句子级的 Bi-LSTM。前向 LSTM 和后向 LSTM 的最终隐藏状态与方面嵌入连接在一起,并馈入综合级 Bi-LSTM。在每个时刻,前向 LSTM 和后向 LSTM 的输出被连接并馈送到 softmax 层,该层生成情感极性的概率分布。

D. 用于 ASC 的基于注意力的 RNN

在本节中，我们首先简要介绍基于注意力的标准 RNN。然后，我们将用于 ASC 的基于注意力的 RNN 的工作分为用于 ASC 的基于注意力的 RNN 和用于 ASC 的基于交互注意力的 RNN，并对它们进行了详细的描述。

图 9 基于注意力的 RNN 模型的框架

1) 基于注意力的 RNN

Bahdanau 等人[59]在神经机器翻译任务中首次提出了注意机制。由于不是序列中的所有信息都是重要的，因此提出了注意力机制来加强 RNN 模型以集中在序列的重要部分。注意力的概念最近吸引了神经网络研究人员的大量兴趣，因为它能够捕捉文本的重要部分(相比之下，例如，取决于最终的隐藏状态向量)。注意力机制已成功应用于许多 NLP 任务[58]，如神经机器翻译[57],[59]、问题回答[60],[61]和机器理解[62],[63]。图 9 展示了基于注意力的 RNN 框架。具体地，上下文向量被计算为这些注释 h_i 的加权和：

公式 (17)

每个注释 h_i 的权重 α_i 由下式计算：

公式 (18)

这里， $\text{score}()$ 被称为方面感知函数，我们考虑了三种不同的选择[57]：

公式 (19)

公式 (20)

公式 (21)

2) 用于 ASC 的基于基本注意力的 RNN

如前所述，大多数用于 ASC 的神经网络模型没有考虑特定方面与其上下文词之间的关系。因此，这样的模型容易遭受语义不匹配的问题。为了解决这个问题，最近提出了一系列基于注意力的神经网络模型，因为它们可以自动识别关于句子中特定方面的相关信息，这可以直接用于提高由神经网络模型提取的特征的质量[57]。下面讨论为 ASC 任务提出的一些代表性的基于基本注意力的 RNN 模型。

Wang 等人[29]提出了一种基于单跳注意的 LSTM 体嵌入模型(名为 ATAE-LSTM)，该模型以体表示和词嵌入的连接作为输入，利用的隐藏状态进行注意计算。图 10 展示了 ATAE-LSTM 的框架。在这个模型中，“concat 方式的注意力”用于捕捉句子中针对给定方面的重要部分。它被证明是一种有效的方法来加强神经模型，以关注句子的相关部分，以响应特定的方面。同样，Yang 等人[30]提出了两种基于注意的双向 LSTM 模型(AB-LSTM)来提高分类性能。Zeng 等人[35]提出了一个 PosATT-LSTM 模型，该模型考

虑了上下文词的重要性，并引入了表示一个方面与其上下文词之间的显式位置上下文的位置感知向量。

图 10 ATAE-LSTM 模型的框架(来自[29])

此外，He 等人[36]基于注意力的 LSTM 模型，通过预训练和多任务学习(PRET+MULT)将文档级情感分类数据集的知识转移到 ASC。现有的 ASC 基准数据集相对较小，很大程度上限制了神经网络模型的性能。尽管缺乏有标签的 ASC 数据，但大规模的文档级情绪分类标签数据很容易在网上获得(例如亚马逊和 Yelp 的评论)。这些评论自然带有评级标签，并包含大量的语言模式。利用从文档级情感分类数据集获得的知识边缘，提高了自动分类器的性能。Li 等人[9]提出了一种新的框架，称为多粒度对齐网络(MGAN)，以跨域同时对齐方面粒度和方面特定的特征表示。

He 等人[31]提出了一种通过捕获给定方面的语义信息来获得更好的方面表示的方法。然后，他们将句法信息纳入注意机制，以获得更好的句子表达。提出的“LSTM+SynATT+TarRep”模型的框架如图 11 所示。每个方面的表示是通过方面项的 m 个嵌入的混合获得的，因此每个嵌入的方面可以表示密切相关的方面项的组合。采用自动编码器结构来学习方面嵌入和方面的表示，这是方面嵌入的加权和。其次，将句法信息整合到注意中，即基于句法的注意模型。在以前的工作中，一个句子中的所有单词对于注意力模型都是同等重要的。因此，注意权重完全取决于特定体与其语境词之间的语义关系。然而，对于不同的方面，捕捉相关的意见词可能是不够的。因此，一个依赖解析器被应用在评论句子上以获得句法路径，然后基于句法的注意机制被设计来选择性地捕获句法路径上最接近方面的最相关的情感词。

图 10 LSTM+SynATT+TarRep 模型的框架(来自[31])

Tay 等人[33]提出了一个方面融合 LSTM 模型(AF-LSTM)，通过对单词-方面的关系建模，将方面信息集成到神经网络模型中。为了自适应地捕捉一个给定方面的正确单词，AF-LSTM 学习根据方面和句子单词之间的关联关系来参与。这解决了通过简单连接来模拟单词方面相似性的其他最先进方法的局限性。相反，为了模拟方面与其内容词之间的相似性，该模型发展了循环卷积和循环相关，并将它们合并到基于不同注意的神经网络模型中。

Hazarika 等人[37]利用 RNN 模型预测了同一句子中所有方面的情感极性，以捕捉它们相应句子表示的方面间依赖性和学习的时间依赖性。具体来说，所提出的模型首先输

入一个句子及其所有方面，然后生成与每个方面相关的句子表示，以获得更好的方面感知表示[26]。基于注意力的 LSTM 网络被用于注意力机制，使得该模型能够捕捉与给定方面相关的内容词的关键部分。与[29]相同，方面表示与每个单词嵌入连接在一起，因此注意机制可以使模型捕获方面信息。最后，为了捕捉方面间的依赖关系，方面感知的句子表示被排序为一个序列，并被馈送到另一个 LSTM 中以建模时间依赖关系。这个 LSTM 的每一步都对应一个特定的方面。然后将每个方面的隐藏状态输出馈送到一个密集层和一个软最大层，以判断每个给定方面的情感极性。

Wang 等人[32]对 ASC 采用了分层注意网络模型[78]。他们提出了一个具有词级和子句级关注的层次网络(即词&子句级 ATT)，用于方面情感分类，以考虑句子中词和子句的重要程度。提议的模型的整体架构如图 12 所示。具体来说，他们首先利用句子层次的话语分割将一个句子分成几个小句。然后，他们利用一个双 LSTM 模型来模拟句子中的所有分句，并设计了一个词级注意机制来捕捉每个分句中的重要词，因为分句中的词并不都是有意义的。最后，他们采用了另一个双 LSTM 模型来模拟每个分句的注意表示，并设计了一个分句级的注意机制来捕捉句子中的重要分句，因为并非句子中的所有分句都是有意义的。

3) 用于 ASC 的基于交互注意力的 RNN

对于这类用于 ASC 的基于注意力的 RN 方法，给定方面与其内容词之间的交互作用被考虑在内。Ma 等人[38]提出了一种交互注意网络(IAN)，它同时考虑了体和全语境的注意机制。如图 13 所示，它使用两个基于注意力的 LSTM 交互捕获方面术语的关键词和其上下文的重要单词。给定方面及其上下文的单词嵌入被输入到两个 LSTM 中，以分别获得单词的隐藏状态。

图 12 词和子句层面的 ATT 框架(来自[78]) (a)单词级注意 (b)子句级注意

图 13 IAN 模型的框架(来自[38])

为关注上下文中的重要信息方面，采用注意机制，并采用方面的隐藏状态和其上下文的隐藏状态的平均值来指导注意权重的生成。因此，方面和完整的上下文可以交互地影响它们的表示的生成。最后，句子的最终表示通过将方面的表示与其上下文联系起来而获得，并输入到用于推断情感类的 softmax 层。为了有效地获得体的表征，特别是当体是多词时，并利用体、其左语境和右语境之间的相互作用来聚焦其中的关键词，Zheng 和 Xia[43]提出了一种具有旋转注意机制的左-中-右分离神经网络(LCR-Rot)。具体来说，他们开发了一个左-中-右分离的 LSTM，它由三个 LSTM(即左、中-和右- LSTM)组成，以

模拟句子的左上下文、方面和右上下文。此外，引入了旋转注意机制，以考虑体与其左/右上下文之间的相互作用来更好地表示给定的方面及其上下文。

图 14 PBAN 模型的框架(来自[41])

他们采用了 `target2context` 注意力来关注左/右上下文中最相关的情感词汇。同时，设计了一个 `context2target` 注意力来聚焦体中的重要词，从而获得体的一个双面表示，即左感知体和右感知体。最后，通过连接组件表示获得的对给定方面的感知的最终表示被馈送到 `softmax` 函数中以预测情感极性。

Gu 等人[41]提出了一种基于 **Bi-GRU** 的位置感知双向注意网络(PBAN)，考虑了体项的位置信息，并利用双向注意机制对体项与上下文的相关性进行了多模型建模。具体来说，如图 14 所示，提出的模型包括三个步骤。

图 15 MGAN 模型的框架(来自[40])

首先，获得句子中单词相对于给定方面项的位置信息，并将其转换为位置嵌入。然后，采用两个 **Bi-GRU** 网络分别提取特定方面及其上下文的特征。最后，使用双向注意机制来建模方面术语与其内容词之间的相关性。受[79]，[80]的启发，他们将位置表示附加到单词嵌入中，以获得方面特定的嵌入。

Huang 等人[42]提出了一种用于 **ASC** 的注意-过度注意()神经网络，该网络同时对体和句子进行建模，以明确地捕捉给定体与其上下文词之间的交互作用。此外，从 **LSTMs** 生成的方面及其上下文的表示通过 **AOA** 模块相互作用。据观察，并不是所有的词都在句子中对某一特定方面起重要作用。句子中的意见词与具体方面高度相关。就拿“开胃菜还行，但是服务很慢”这句话来说。例如，有两个方面“上诉人”和“服务”。根据语言经验，阳性词“ok”描述的是“开胃菜”，而不是“服务”。引入 **AOA** 模块是为了从方面到上下文和上下文到方面产生相互关注，并捕获特定方面及其对应上下文的最重要部分。

Fa 等人[40]提出了一种细粒度的注意机制，在词的层面上模拟了体与其语境之间的相互作用。如图 15 所示，**MGAN** 框架由两部分组成，分别是细粒度注意机制和粗粒度注意机制。特别地，引入了细粒度的注意机制(即 **F-Aspect2Context** 和 **F-context 2 Aspect**)来建模给定方面与其对应的上下文词在词级上的交互，并减少由粗粒度的注意机制导致的信息缺失。

图 16 CNN 模型的框架(来自[65])

此外,双向粗粒度注意(即 C-Aspect2Context 和 C-context2Aspect)被开发并与细粒度注意向量相结合,以构建用于预测句子相对于给定方面的情感极性的 MGAN 框架。此外,为了利用体层面的交互信息,在损失函数中采用了体对齐损失,以增强同一句子中具有不同情感极性的体的注意权重的差异。

Liu 和 Zhang[39]利用注意机制来计算每个词对于给定方面的情感极性的重要程度。他们通过增强从特定方面的左右语境中获得的注意权重的差异来扩展注意机制。引入了多个门来进一步控制注意力的贡献。具体来说,采用双 LSTM 模型对句子中的单词嵌入进行建模,然后对隐藏状态采用注意机制来计算句子中每个单词对特定方面的贡献。

E. 用于 ASC 的 CNN

在这一部分,我们提供了 CNN 的简要介绍。然后,我们详细回顾了基于 CNN 方法的 ASC。

1) CNN

卷积神经网络(CNN) [81]在用卷积和池运算处理非结构化多媒体数据方面是强大的。CNN 可用于特征表示学习。它利用词嵌入将句子映射成低维语义表示,并保持词的序列信息。然后,提取的句子表示连续通过带有多个滤波器的卷积层、最大池层和全连接层。图 16 呈现了 CNN 的框架。

具体来说,令 $x_{i:i+j}$ 表示对向量组 x_i, x_{i+1}, \dots, x_j 的串接。卷积运算在该输入嵌入层上执行。为了生成新的特征,对 h 字窗口采用了 $h \in \mathbb{R}^{h \cdot \dim w}$ 滤波器。例如,在 h 字窗口上单词 $x_{i:i+h-1}$ 计算的一个特征 c_i 如下:

$$\text{公式 (22)}$$

这里 $b \in \mathbb{R}$ 是偏置项, f 是非线性激活函数(如 \tanh , ReLU)。利用相同的权重将滤波器 k 用于 h 个字的所有可能窗口,以产生特征图。

$$\text{公式 (23)}$$

在一个 CNN 中,使用几个不同宽度的核(也称为卷积滤波器)在整个单词嵌入矩阵 x 上滑动。每个核提取一个特定的 n -gram 模式。在卷积层之后,通常在特征图上采用最大池化策略,并且最大值 $\hat{c} = \max\{c\}$ 被作为对应于该特定核的特征。通常采用每个内核上的 \max 操作对输入进行二次采样。这个想法是捕捉最重要的 n -gram 特征——一个具有最高值的特征。这种汇集策略通过将输入映射到固定大小的输出,自然地解决了可变句子长度的问题。

2) 基于 CNN 的 ASC

CNN 因其能够从文本中提取局部和全局表示而被采纳[45]-[47]。Huang 等[45]通过应用参数化过滤器和参数化门,将方面信息并入 CNN。特别地,提出了两个简单的基于 CNN 的模型,其中包含了方面信息。他们引入了两个考虑因素的神经单元,即参数化滤波器 and 参数化门。这些单元被设计用于学习特定方面的特性。然后,介绍了两种模型变量:参数化滤波器(PF-CNN)和参数化选通滤波器(PG-CNN)。

Xue 和 Li[44]提出了一个基于 CNN 和门机制的模型。该模型在输入嵌入层上包括两个独立的卷积层,其输出由选通单元组合。使用具有多个滤波器的卷积层来有效地生成 n-gram 特征。设计了两个非线性门,分别连接到两个卷积层。给定方面信息,他们选择地为 ASC 捕获了方面感知的情感信息。由于所提出的模型可以很容易地进行比较,因此比基于 LSTM 和注意力机制的模型花费的训练时间少得多。当方面由多个字组成时,采用另一个卷积层来获得方面表示。

Fan 等人[46]提出了一种用于 ASC 的卷积存储网络,其灵感来源于卷积运算并基于存储网络。该模型结合了注意机制来学习句子中的单词和多个单词信息。通过将上下文信息同时存储到一个固定大小的窗口中,他们所提出的内存网络能够捕获长距离依赖。

Li 等人[47]采用了一种方法,利用特定方面及其上下文词之间的位置信息来缩放卷积层的输入。在重新检查注意机制的缺点和阻碍 CNN 良好表现的障碍之后,为 ASC 开发了一个 TNet 模型。代替注意机制, CNN 层用于从 Bi-LSTM 层获得的隐藏状态生成重要特征。具体来说,为了更好地将方面信息融入到单词的表示中,引入了一个方面感知转换组件。此外,采用 CNN 作为特征提取工具,并利用上下文保护和位置信息来克服 CNN 模型的不足。

F. 用于 ASC 的记忆网络

在这一部分中,我们首先介绍了在 NLP 中广泛使用的记忆网络的细节。然后,我们概述了用于 ASC 的深度记忆网络。

1) 记忆网络

记忆网络在 NLP 中取得了巨大的成功。具体地说,给定一个句子 $S = \{w_1, w_2, \dots, w_n\}$, 和方面词 $S = \{w_{start}, w_{start+1}, \dots, w_{end-1}, w_{end}\}$, 每个单词都被映射到其嵌入向量中。这些词嵌入向量分为两部分,即体表示和上下文表示。当方面是单个词时(如“食物”和“服务”),方面表示是嵌入方面词。当方面是多词短语时(例如“电池寿命”),方面表示是嵌入词向量的平均值。上下文单词向量 $\{m_1, m_2, \dots, m_{start-1}, \dots, m_{end+1}, \dots, m_n\}$ 被堆叠并转换为记忆 m 。内部状态 u 被设置为方面表示。 u 和每个记忆 m_i 之间的匹配是通过 softmax 层取内积来计算的:

$$\text{公式 (24)}$$

其中 \mathbf{p} 是输入上的概率向量。

然后，来自记忆 \mathbf{o} 的输出向量是变换后的输入 \mathbf{m}_i 的和，通过来自输入的概率向量加权求和。

公式 (25)

然后将该模型扩展来处理多跳操作。记忆层按如下方式堆叠：第一个 \mathbf{u}_{k+1} 上方的层的输入计算为层 k 的输出 \mathbf{o}_k 和输入 \mathbf{u}_k 之和：

公式 (26)

2) 基于存记忆网络的 ASC

Tang 等人[19]首先为 ASC 开发了基于多跳注意机制的深度记忆网络，它不仅高效而且计算成本低。

图 17 记忆网络模型的框架(来自[19])

图 18 RAM 模型的框架(来自[49])

如图 17 所示，它在外存储器上采用了多跳注意机制来关注给定方面的上下文词的重要性水平。该方法明确捕获了上下文词的重要信息，用于推断特定方面的情感极性。这种重要度和文本表示由多个计算层计算，这些计算层是具有外部记忆的基于注意力的神经模型。

Tay 等人[48]引入了并矢记忆网络(DyMemNN)，通过将参数化的神经张量成分和全息成分结合到记忆选择操作中来捕捉给定方面与其上下文词之间丰富的并元相互作用。两种并矢记忆网络，即张量 DyMemNN 和全息 DyMemNN，被发展用于研究丰富的方面与句的并矢交互。

Chen 等人[49]提出了一种基于记忆网络用于每个方面的循环注意机制(RAM)，以提取相隔很远的情感信息。如图 18 所示，为了实现这一点，所提出的模型利用了一个循环/动态的注意结构，并学习了一个非线性的注意组合。然而，由于所提出模型中使用的 LSTM 特征，注意力权重仅基于局部句子信息来计算。

在[55]中，他们为 ASC 提出了一种上下文注意机制，该机制明确考虑了给定方面和每个上下文单词之间的关联。该模型包括两种注意力增强机制，即句子层面的内容注意机制和内容注意机制。句子级内容注意机制能够从全局的角度捕捉句子对给定方面的重要信息，克服了深度记忆网络模型的短视问题。

为了更好地模拟方面和情感之间的相互作用，Li 和 Lin[66]将方面检测任务结合到情感预测任务中。他们通过端到端的方法实现了情感识别，在这种方法中，两个任务通

过深度记忆网络同时学习。这样,在方面检测中产生的信号为情感分类提供反馈,相反,预测的极性为方面识别提供线索。

此外, Wang 等[53]提出了用于 ASC 的目标敏感记忆网络(TMNs)。TMNs 可以捕捉体与其上下文词之间的情感交互。此外,还介绍了构建 TMNs 的六种技术。Majumder 等人[52]提出了一种方法,该方法通过记忆网络将相邻方面相关信息集成到预测该方面的情感极性中。

4 数据集

已略(见原文)

5 评估方法

已略(见原文)

6 实验实现

已略(见原文)

7 未来方向和挑战

已略(见原文)

8 总结

已略(见原文)

致谢:致谢已略(见原文)

参考文献:参考文献已略(见原文)