

文章编号: 2095-2163(2020)06-0285-05

中图分类号: TP391.4

文献标志码: A

基于局部注意力的方面级别情感分析算法

张 硕¹, 卜 巍², 邬向前¹

(1 哈尔滨工业大学 计算机科学与技术学院, 哈尔滨 150001; 2 哈尔滨工业大学 媒体技术与艺术学院, 哈尔滨 150001)

摘要: 考虑到对方面进行情感分析时,距方面词越近的上下文内容对方面的情感倾向影响可能越大,因此将结合距离因素,提出了基于局部自注意力机制的算法用于方面级别的情感分类。将输入分为全局文本序列、方面序列以及局部文本序列,通过自注意力分别对三个序列进行特征提取,对全局文本和方面序列进行交互注意力计算,更新全局文本表示,对局部文本和方面序列进行交互注意力计算,更新局部文本表示。最终将全局文本和局部文本表示连接起来输入分类器。实验结果显示本文的结果优于近两年提出的方面级别情感分析算法。

关键词: 情感分析; 方面级别; 局部文本; 注意力

A local attention network for aspect-level sentiment analysis

ZHANG Shuo¹, BU Wei², WU Xiangqian¹

(1 School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China;

2 School of media Technology and Art, Harbin Institute of Technology, Harbin 15001, China)

[Abstract] Considering that the context content closer to the aspect words may have a greater influence on the sentiment of the aspect when analyzing the emotion of the aspect, an algorithm based on the local attention mechanism is proposed to classify the emotion of the aspect level based on the distance factor. Input can be divided into global text, aspect, and the local text. Extract feature of the three sequences by self-attention, and then apply interactive attention on global text and aspect to update global text representation, apply interactive attention on local context and aspect to update the local text representation. Finally, the global text and the local text representation are joined into the input classifier. Experimental results show that our results are superior to the aspect-level sentiment analysis algorithm proposed in the past two years.

[Key words] sentiment analysis; aspect-level; local context; attention

0 引言

随着互联网的快速发展,可以轻松地得到海量带有观点的数据。各种购物平台、点餐平台以及社交平台的兴起,人们可以在购物平台上对所购买的商品进行评价,对消费过的餐厅进行点评,尽情地在社交平台上表达自己的观点,对事件的情感态度。对这些饱含大众情感的内容进行研究,发掘这些文本背后隐藏的大众的情感态度,将帮助购物平台和点餐平台等更好地了解购买者的喜好,同时也是对商家的监督。对社交媒体上大众对于某些事件的情感分析,可以帮助了解大众,引导正确的社会舆论。因此,近年来情感分析在自然语言处理领域的研究非常广泛。研究表明:40%的情感分类错误是由于在情感分类中没有考虑目标造成的^[1]。近年来,研究者们开始将研究方向从单纯的文本情感分

析转向方面级别的文本情感分析。

本文针对方面级别的情感分析,考虑上下文单词与方面序列之间的距离,提出了局部文本序列的定义,并提出了一种基于局部注意力的网络框架。

1 相关工作

研究者们研究了各种基于神经网络的方法来解决方面级别的情感分类问题。一些研究者设计了有效的神经网络,从目标及其上下文中自动生成有用的低维表示,在方面级别情感分类任务中取得了很好的效果,典型的方法是基于 LSTM 的神经网络。Dong 等人通过开发两个 LSTM 网络来模拟方面目标的左右上下文来解决这个问题,该方法利用这两个 LSTMs 的最后隐藏状态来预测情绪。为了更好地抓住句子中的重要部分,Wang 等人使用一个方面词嵌入,来生成一个注意力向量,集中在句子的不同

基金项目: 国家自然科学基金(61672194); 国家重点研究与发展计划(2018YFC0832304); 中国黑龙江省杰出青年科学基金(JC2018021); 国家机器人与系统国家重点实验室项目(SKLR5-2019-2019-KF-14); 中兴通讯产学研合作论坛合作项目。

作者简介: 张 硕(1995-),女,硕士研究生,主要研究方向:自然语言处理、机器学习、情感分析等; 卜 巍(1977-),女,博士,副教授,主要研究方向:数字媒体技术、数字图像处理、医学图像分析等; 邬向前(1973-),男,博士,教授,博士生导师,主要研究方向:数字图像处理、模式识别、生物特征识别等。

收稿日期: 2020-04-05

部分。Vo 和 Zhang 将整个上下文分为目标、左上下文和右上下文三个部分,使用情感词典和神经池函数生成目标相关特征。Tang 等人将上下文分为带目标的左半部分和带目标的右半部分,分别使用两个 LSTM 模型对这两个部分建模,利用这两部分合成的目标特定表示进行情感分类。

2 算法详述

2.1 局部文本序列

局部文本序列:对于给定一个句子 $s = [w_1, w_2, \dots, w_i, \dots, w_j, \dots, w_n]$, 和方面序列 $t = [w_i, w_{i+1}, \dots, w_{i+m-1}]$, 其中方面序列可能包含一个或多个词。如图 1 所示, 设局部文本序列为整个文本中, 距方面词中心距离为 k 的序列片段。当方面词为一个单词时, 局部文本序列 $l = [w_{i-k}, w_{i-k+1}, \dots, w_i, \dots, w_{i+k-1}, \dots, w_{i+k}]$, 其中 w_i 为方面词。当方面词为多个单词时, 局部文本序列 $l = [w_{i-k}, \dots, w_i, \dots, w_j, \dots, w_{j+k}]$, 其中 w_i 为方面词的第一个词, w_j 为方面词的最后一个词。

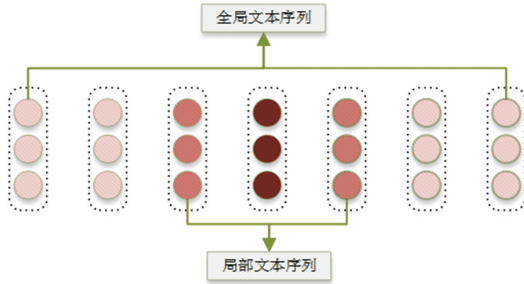


图 1 局部文本序列示意图

Fig. 1 The chart of local context

2.2 网络框架介绍

网络整体结构如图 2 所示(与图的名字不一致), 包括词嵌入层, 特征提取层, 注意力层以及分类层。

词嵌入层。(1)Glove 词嵌入。设 $L \in R^{d_w \times |V|}$ 为预先训练的 Glove (Pennington 等人, 2014) 嵌入矩阵, 其中 d_w 为词向量的维数, $|V|$ 为词汇的数量。将上下文和方面的每个词 $w_i \in R^{|V|}$ 映射到对应的嵌入向量 $e_i \in R^{d_w \times 1}$, 它是嵌入矩阵 L 中的一列。(2)BERT 词嵌入。使用预先训练的 BERT 生成序列的词向量。为了便于 BERT 模型的训练和微调, 将给定的上下文和目标分别转换为“[CLS]+上下文文本+SEP”和“[CLS]+方面序列+SEP”, 从而作为文本序列的词嵌入进行模型训练。

通过词嵌入模块, 将全局文本序列映射为 $S = \{v_{s_1}, v_{s_2}, \dots, v_{s_n}\}$, 将方面序列映射为 $T = \{v_{t_1}, v_{t_2}, \dots,$

$v_{t_m}\}$, 将局部文本序列映射为 $L = \{v_{l_1}, v_{l_2}, \dots, v_{l_k}\}$ 。其中 $v_{s_i} \in R^{d_w}, v_{s_i} \in R^{d_w}, v_{l_i} \in R^{d_w}, d_w$ 为词嵌入维度。

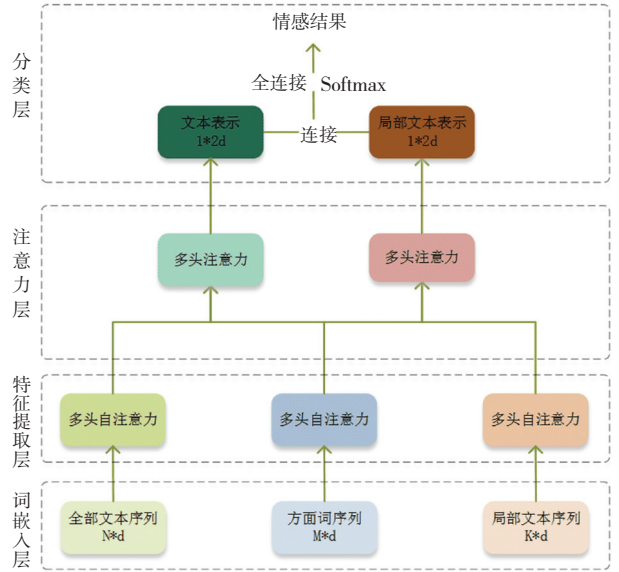


图 2 网络框架图

Fig. 2 The framework chart of our algorithm

特征提取层。多头自注意力是一种同时执行多次自注意力的操作。为了对文本进行特征提取, 利用多头自注意力机制分别对全本文本序列、方面词序列以及局部文本序列进行建模。注意函数将 Key 序列 $K = \{k_1, k_2, \dots, k_n\}$ 和 Query 序列 $Q = \{q_1, q_2, \dots, q_m\}$ 映射到输出序列, 其中对全局文本自注意力计算方式如公式(1), 局部文本的自注意力计算方式如公式(2), 方面序列的自注意力计算方式如公式(3)。

$$h_s = Attention(S, S, S) = softmax\left(\frac{S S^T}{\sqrt{d_s}}\right) S, \quad (1)$$

$$h_l = Attention(L, L, L) = softmax\left(\frac{L L^T}{\sqrt{d_l}}\right) L, \quad (2)$$

$$h_t = Attention(T, T, T) = softmax\left(\frac{T T^T}{\sqrt{d_t}}\right) T. \quad (3)$$

多头自注意力就是平行地进行多词注意力计算, 将结果拼接到一起作为最终的结果, 其中三个序列的计算方式是一样的, 这里只以全局文本为例子, 具体计算如公式(4)和公式(5):

$$\tilde{h}_s = MultiHead(S, S, S) = Concat(head_1, \dots, head_n) W^0, \quad (4)$$

$$head_i = Attention(S W_i^Q, S W_i^K, S W_i^V). \quad (5)$$

除了多头注意力之外, 每一次注意力计算后都包含一个全连接的逐点前馈网络(Point-wise Feed Forward, PFF)。共包括两个线性转换, 以及一个

relu 激活。公式如(6)和(7)。

$$PFF(x) = \text{relu}(x W_1 + b_1) W_2 + b_2, \quad (6)$$

$$\tilde{h}_s = PFF(\tilde{h}_s). \quad (7)$$

其中: $W_1 \in R^{d_u \times d_w}$, $W_2 \in R^{d_w \times d_w}$ 是两个可学习权重, $b_1 \in R^{d_w}$, $b_2 \in R^{d_w}$ 是两个偏置参数。*relu* 为激活函数。

注意力层。在这一层中,将利用特征提取层得到的结果,以方面序列的输出作为新的注意力中的 Query,全局文本序列的输出作为新的注意力计算中的 Key 以及 Value,计算方面序列对全局文本的注意力分数,如公式(8)。同样地,以方面序列作为 Query,以上一层中局部文本序列的输出作为 Key 和 Value,采用多头注意力的方式计算方面序列对局部文本的注意力分数,并更新局部文本表示,如公式(9)。之后将更新后的全局文本表示 r_s 以及局部文本表示 r_l 分别作为输入,通入到 PFF 函数中,如公式(10)和公式(11),从而得到最终的全局文本表示以及局部文本序列的表示。

$$r_s = \text{softmax}\left(\frac{\tilde{h}_l \tilde{h}_s^T}{\sqrt{d_s}}\right) \tilde{h}_s, \quad (8)$$

$$\tilde{r}_s = PFF(r_s), \quad (9)$$

$$r_l = \text{softmax}\left(\frac{\tilde{h}_l \tilde{h}_l^T}{\sqrt{d_l}}\right) \tilde{h}_l, \quad (10)$$

$$\tilde{r}_l = PFF(r_l). \quad (11)$$

分类层。将注意力层输出的全局文本表示和局部文本表示连接到一起,如公式(11),由于分类结果分为积极、消极和中性三个类别,所以将连接后的向量输入一个三个神经元的全连接层,如公式(12),然后送入 SoftMax 函数得到和为 1 的三个概率值,如公式(13),概率最大的即为最终的情感类别。

$$r = [\tilde{r}_s, \tilde{r}_l], \quad (11)$$

$$x = W_l \cdot r + b_l, \quad (12)$$

$$P(y = c) = \text{softmax}(x) = \frac{\exp(x_c)}{\sum_{i \neq c} \exp(x_i)}. \quad (13)$$

其中: $W_l \in R^{3 \times 2d_w}$ 和 $b_l \in R^{3 \times 1}$ 分别表示权重矩阵和偏差。

3 实验

3.1 数据集介绍

实验的数据集来自 Semeval2014 语义评测任务

的任务 4 的第四个子任务:方面级别情感分类。该任务给定一个句子以及句子中包含的一个或多个方面类别,输出每个方面类别对应的情感分类。分类分为积极、消极和中性。数据集一共包括餐厅评论数据集以及笔记本电脑数据集。本文用到的数据集还包括 ACL 14 Twitter 数据集。

“All the money went into the interior decoration, none of it went to the chefs”,对于这个句子,当方面序列为“interior decoration”时,情感分类为积极;当方面序列为“chefs”时,情感分类为消极。这些数据集被大多数提出的模型所采用,是方面级别情感分析任务中最流行的数据集,表 1 展示了 3 个数据集中训练集和测试集的细节。

表 1 数据集统计表

Tab. 1 The statistics of dataset

数据集	积极	中性	消极
Laptop-Train	994	464	870
Laptop-Test	341	169	128
Restaurant-Train	2 164	637	807
Restaurant-Test	728	196	196
Twitter-Train	1 561	1 560	3 127
Twitter-Test	173	173	346

3.2 实验评价指标

为了能够评估不同的情感分析算法的行能,对算法的泛化性能进行评估,本文采用该领域常用的评价指标:准确率(Accuracy)和综合指标(F1),其中准确率计算方式如公式(14),F1 计算方式如公式(15):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (14)$$

$$F_1 = \frac{2 \times TP}{N + TP + FN}. \quad (15)$$

其中: TP 表示被模型预测为正的正样本; FP 表示被模型预测为正的负样本; FN 表示被模型预测为负的正样本; TN 表示被模型预测为负的负样本。N 表示总的样本数量。

3.3 模型参数

词嵌入中 Glove 的嵌入方法的词嵌入维数设置为 300。BERT 词嵌入方法,隐藏层维度设置为 768。LSTM 网络隐藏状态的维度同样设置为 300,并在训练期间固定。为了确保在训练过程可以学到有用的信息,参数梯度不为 0,模型的权重是用 Glorot 随机初始化的^[2]。L2 正则化项的系数 λ 为 10^{-3} , dropout 设置为 0.1, Batchsize 设置为 32, Adam 优化器用

于更新所有参数。

3.4 损失函数

本文在损失函数中使用了标签平滑正则化 (Label smoothing Regularization, LSR) 的。它惩罚了低熵输出分布^[3]。LSR 可以通过防止网络在训练期间为每个训练示例分配全 0 或全 1 概率来减少过度拟合,用平滑的值替换分类器的 0 和 1 目标,如用 0.1 或者 0.9 替代 0 和 1。利用公式 (17) 中的 $q'(k|x)$ 代替原本的标签分布 $q(k|x)$ 。令 $u(k)$ 服从简单的均匀分布,如公式 (16)。

$$u(k) = 1/c. \quad (16)$$

其中, c 为情感分类的类别数 3,得到了新的标签分布,如公式 (17)。

$$q'(k|x) = (1 - \varepsilon)q(k|x) + \varepsilon/c. \quad (17)$$

最终的损失函数依旧采用交叉熵损失函数,但是将真实标签换成平滑后的标签,如公式 (18)。

$$L(\theta) = -q' \log(p) + \lambda \|\theta\|^2 = (1 - \varepsilon)q \log(p) + \frac{\varepsilon \log(p)}{c} + \lambda \|\theta\|^2. \quad (18)$$

其中, q' 为平滑后的标签, p 为预测标签, q 为真实标签。 λ 为 L2 正则化项系数, θ 为参数。

3.5 实验结果

首先针对局部文本序列的长度进行了实验,分别取 $k = 1, 2, 3, 4, 5$ 。对应的局部文本序列长度位: 3, 5, 7, 9, 11。如表 2 所示,可以发现当 $k = 2$ 或 3 时,结果最好,也就是局部文本序列的长度为 5 或者 7 时,模型的效果最佳。当序列长度过长时,局部文本序列的长度趋近于全局文本序列长度,本文的模型失去了意义;当序列长度过短时,局部文本序列几乎等于方面序列。因此,在后续的实验,都采用 $k = 2$ 来训练。

表 2 局部文本序列长度 K 不同取值的对比

Tab. 2 Comparison of K in local context length

数据库 K 取值	Restaurant		Laptop		Twitter	
	acc	F1	acc	F1	acc	F1
1	80.36	69.68	71.79	65.17	70.95	69.32
2	80.45	70.00	72.88	71.85	72.25	70.66
3	80.89	71.74	71.00	65.20	71.32	66.18
4	79.34	68.11	71.32	65.55	70.09	66.84
5	79.64	70.88	70.38	65.04	69.51	66.75

为了证明本文提出的基于局部注意力的方面级情感分析网络 (Local Attention Network, LAN) 的优越性,将其与以下模型进行了比较,结果如表 3 所

示。

表 3 模型对比结果

Tab. 3 Model comparison results

数据库 对比方法	Restaurant		Laptop		Twitter	
	acc	F1	acc	F1	acc	F1
LSTM	76.43	61.61	68.65	63.22	67.63	65.39
TD-LSTM	77.77	64.93	68.97	62.24	70.95	68.50
ATAE-LSTM	77.86	64.88	69.91	64.18	68.93	66.46
IAN	78.71	67.71	72.10	63.72	69.22	66.90
MGAN	81.25	71.94	75.39	72.47	72.54	70.81
TNet	80.79	71.27	76.54	71.75	74.97	73.60
TransCap	79.29	70.85	73.87	70.10	-	-
IACapsNet	81.79	73.40	76.80	73.29	75.01	73.81
LAN	82.77	74.02	78.06	73.46	73.41	71.63

LSTM。使用一个 LSTM 网络对句子进行建模,最后一个隐藏状态作为句子的表示形式进行最终分类;TD-LSTM:使用两个 LSTM 网络来模拟,围绕方面术语的前后上下文,这两个 LSTM 网络的最后隐藏状态被连接起来,以预测情感极性。ATAE-LSTM:通过将方面嵌入到每个词向量中来进一步扩展 T-LSTM;IAN:使用两个 LSTM 网络分别对句子和方面项进行建模,利用句子中隐藏的状态为方面生成一个注意向量,反之亦然。基于这两个注意向量,输出用于分类的句子表示和方面表示;MGAN:提出了一种细粒度的注意机制,捕获方面和上下文之间的词级交互;PBAN:关注方面术语的位置信息,并通过双向注意相互模拟方面术语和句子之间的关系;TNet:提出了一个特定目标的转换组件,用于生成基于特定目标的词向量,并利用上下文信息保留机制维持原有信息;TransCap:提出了一种将文档级知识转移到方面级情感分类的转移胶囊网络模型;IACapsNet:提出利用胶囊网络构造基于向量的特征表示,并通过 EM 寻路算法对特征进行聚类。此外,在胶囊寻路过程中,引入了交互注意机制,以建立方面和上下文之间的语义关系。结果显示本文的模型在推特数据集上略低于 IACapsNet 和 TNet,餐厅数据集以及笔记本电脑数据集上准确率和 F1 都取得了最好的结果。

4 结束语

本文提出了一种基于局部注意力的网络框架,由于距离越近,上下文对方面序列情感倾向影响可能越大,考虑到上下文与方面序列之间的位置关系,提出了局部序列的定义,局部文本序列的构建取决 (下转第 293 页)