

Fine-Grained Sentiment Analysis Based on Hierarchical Attention Networks

Xinglin Shao, Shaozhang Niu

Beijing Key Laboratory of Intelligent Communication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing
Email: 1464872331@qq.com

Received: Nov. 6th, 2019; accepted: Nov. 19th, 2019; published: Nov. 26th, 2019

Abstract

Fine-grained sentiment analysis analyzes the emotional polarity of text from multiple angles, and it has become a hot issue in the field of sentiment analysis. Different from previous LSTM networks in which the algorithm uses attribute information as the embedded vector and single-layer attention mechanism, this paper proposes a multi-layer network based on hierarchical attention mechanism, which can give different attention weights to words and sentences. To help the model increase the attention to important parts, on the other hand use the entity information as the embedded vector, which is more representative of the meaning of the target phrase than the attribute information. The experimental results show that the model has achieved excellent results on the SemEval 2014 dataset and is superior to the existing algorithms.

Keywords

Attention Mechanism, LSTM, Fine-Grained Sentiment Analysis

基于分层注意力机制的细粒度情感分析

邵兴林, 牛少彰

北京邮电大学智能通信软件与多媒体北京重点实验室, 北京
Email: 1464872331@qq.com

收稿日期: 2019年11月6日; 录用日期: 2019年11月19日; 发布日期: 2019年11月26日

摘要

细粒度情感分析从多个角度对文本情感极性进行分析, 越来越成为情感分析领域的热点问题。不同于以

往算法使用属性信息作为嵌入向量和单层注意力机制的LSTM网络, 本文提出了一种基于分层注意力机制的多层网络, 一方面可以对单词和句子分别给予不同的注意力权重, 帮助模型增加对重要部分的注意力, 另一方面使用实体信息作为嵌入向量, 比属性信息更能表示目标短语的含义。实验结果表明, 该模型在SemEval 2014数据集上取得了出色的效果, 优于现有的算法。

关键词

注意力机制, LSTM, 细粒度情感分析

Copyright © 2019 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 背景与简介

情感分析[1]是自然语言处理领域的基本任务之一, 特别是 aspect-level 级别的细粒度情感分析越来越成为研究热点, 对于商品评论、舆情分析等领域有重要的现实意义。一般来说情感分析只能对一整段文本进行情感极性判断, 而细粒度情感分析针对文本的不同属性做出具体的情感分析。例如, “苹果派很美味, 但服务员真的太没礼貌了”。对于这段评论, 针对实体“食物(苹果派)”, 情感极性判断为“积极”, 对于实体“服务(服务员)”, 情感极性判断为“消极”。针对一段评论的不同方面, 情感极性的判断有可能完全相反。

近来深度学习方法对于自然语言处理领域的各种任务都取得了很好的效果。比如机器翻译[2], 文本分类[3], 问答[4], 文本摘要[5]等等。情感分析作为文本分类的子领域之一, 也出现了很多优秀成果。Yoon kim [6]提出基于卷积神经网络(CNN)的文本分类模型, 成为文本分类问题的一个基准; D Tang [7]提出用分层注意力机制分别对句子和文档建模, 对文档级别文本进行情感分析; Zhang [8]等采用了新的思路, 采用字符级别词嵌入实现文本分类, 也取得了很好的效果; D Tang [9]提出用记忆网络实现细粒度情感分析。文档和句子级别的情感分析取得了不错的效果, 但难度更高、更有实际价值的 aspect-level 级别的情感分析的研究还处于初级阶段。D Tang [10]提出的 Target-Dependent LSTM (TD-LSTM)和 Target-Connection LSTM (TC-LSTM)实现了不错的效果, 通过加入属性嵌入向量, 考虑具体的属性信息, 实现细粒度情感分析。这些模型仅仅考虑了属性信息, 但在 aspect-level 级别的情感分析中, 实体信息更为重要, 同时由于一段文本通常会有多个句子, 选出那些对情感极性判断有更重要作用的句子也有很重要的意义。

本文提出了一种基于分层注意力机制的多层网络, 采取实体信息作为嵌入向量, 实现细粒度情感分析。首先, 提取实体信息作为嵌入向量, 比如“苹果派很美味, 但服务员真的太没礼貌了”。这段评论, 属性“苹果派”属于实体“食物”, 则把“食物”对应的向量表示作为嵌入向量。aspect-level 级别的情感分析中, 实体信息的提取对结果有重要影响。另外, 模型引入了多级注意力机制, 注意力机制最先开始被应用于计算机视觉领域, 后来被证实在包括机器翻译[11], 文本摘要, 文本分类等多个自然语言处理领域也有重要的作用。由于文本通常包含多个句子, 构成了单词 - 句子的层次结构, 因此相应地可以构建句子 - 文档分层表示, 而且由于上下文语境的不同, 不同的单词和句子会对语义表达有不同的影响, 而分层注意力机制则是很好的解决办法。通过加入多级注意力机制, 帮助模型对不同单词、不同句

子赋予不同的注意力权重, 从而对情感极性做出更为准确的判断。

本文的主要贡献有以下两点: 一是采取实体信息而不是属性信息作为嵌入向量, 二是使用了分层注意力机制, 对不同的单词和句子分别赋予不同的注意力权重。

2. 相关工作

2.1. 注意力机制

Volodymyr Mnih 等人首先将注意力机制应用于图像领域, 他们研究的动机也是受到人类注意力机制的启发。该模型是在传统的 RNN 上加入了 attention 机制, 通过 attention 去学习一幅图像要处理的部分, 每次当前状态, 都会根据前一个状态学习得到的要关注的位置和当前输入的图像, 去处理注意力部分像素, 而不是图像的全部像素。这样的好处就是更少的像素需要处理, 减少了任务的复杂度。随后注意力机制也被应用于 NLP 领域。Bahdanau 等人在 2014 年第一次将注意力机制用于机器翻译, 使用 encoder-decoder 框架, 并且在翻译之前使用注意力机制来选择对应外语单词的原始语言的参考单词, 基于 attention 的机器翻译系统把源语言端的每个词学到的表达和当前要预测翻译的词联系起来。T Luong 等人的工作[12]将 attention 在 RNN 中进行扩展, 这篇论文对后续各种基于 attention 的模型在 NLP 应用起到了很大的促进作用。在论文中他们提出了两种 attention 机制, 一种是全局(global)机制, 一种是局部(local)机制。Z. Yang 等人提出一种分层注意力机制, 对句子和单词级别分别使用注意力机制, 用于分别对句子和文档中的单词、句子的重要性进行建模。更多注意力机制的应用还包括自然语言问答[13], 图像问答[14]等等。注意力机制在图像和文本领域都取得了显而易见的成效。本文将使用一种分层注意力机制, 对不同的单词和句子分别赋予不同的注意力权重, 提高分类正确率。

2.2. Aspect-Level 级别情感分析

aspect-level 级别的情感分析是典型的分类问题, 正如我们之前提到的, aspect-level 级别的情感分析作为一种细粒度的情感分析, 能够从多个角度对文本情感极性进行分析, 越来越成为情感分析领域的热点问题。目前的大多数方法都是对一个句子整体进行情感极性的分析, 而很少针对句子某个方面进行分析。传统的解决这些问题的方法是手动设置一系列特征。Ravichandran 等人基于情感词汇的丰富[15]、Mohammad 等人基于词典的特征[16]来进行细粒度的情感分析, 这些研究都是构建具有特征的情感分类器。但是这些方法大都依赖于提取特征的质量, 而特征工程本身就是一项繁重的工程。

LSTM 网络已经被证明在 NLP 各个领域都取得了很好的效果。Tang 等人提出的 TD-LSTM 和 TC-LSTM 在判断句子情感极性时考虑了属性信息, 在细粒度情感分析上取得了不错的效果。TC-LSTM 通过取目标短语包含的单词的向量平均值来获得嵌入向量, 然而, 简单地平均目标短语的单词嵌入不足以表示目标短语的语义, 效果并不理想。

尽管这些方法取得了一定的效果, 但是细粒度的情感分析仍然具有很大的挑战性。本文基于之前的工作提出了一种改进的基于分层注意力机制的网络, 可以充分利用实体信息进行细粒度情感分析。

3. 用于细粒度情感分析的分层注意力机制网络

3.1. 长短期记忆网络

循环神经网络(RNN)能够把以前的信息联系到现在, 从而解决现在的问题。这种记忆单元使其成为自然语言处理领域最常用的网络架构, 然而标准的 RNN 会出现梯度爆炸和梯度消失问题, 为了解决这些问题, LSTM 网络于 1997 年由 Hochreiter 等人提出, 被广泛地应用于解决各类问题, 并取得了非常棒的效果。LSTM 的结构主要包括三个门和一个记忆单元。图 1 展示了标准 LSTM 的架构。

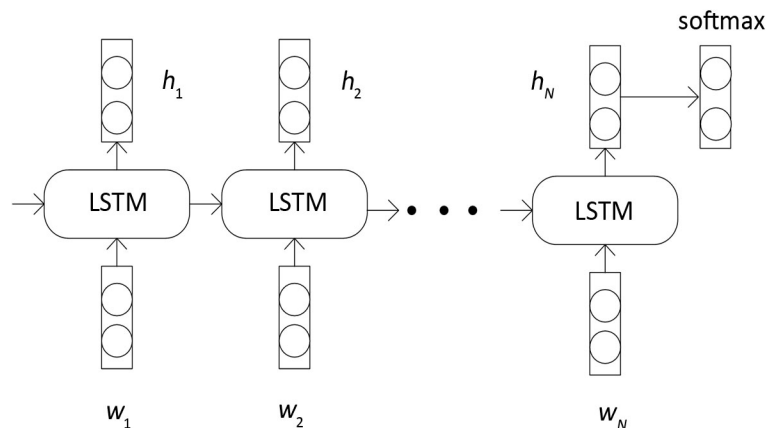


Figure 1. Standard LSTM network architecture
图 1. 标准 LSTM 网络架构

每一个记忆单元的计算如下:

遗忘门:

$$f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \quad (1)$$

传入门:

$$i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \quad (2)$$

$$\Delta C_t = \tanh(W_c * [h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * \Delta C_t \quad (4)$$

输出门:

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

其中, W_f , W_i , W_o 是权重矩阵, b_f , b_i , b_o 是偏置。 σ 是激活函数, 此处是 sigmoid 函数。 x_t 代表记忆单元的输入, 具体是指单词的词嵌入向量。 h_t 代表隐层的输出向量。实验中将隐层的最后一个输出向量作为句子的向量表示并且将其经过 softmax 层处理最终输出每个类别的预测概率。

3.2. 分层注意力机制网络

实体信息对于 aspect-level 级别的细粒度情感分析有着重要的作用, 不同于 Tang 等人将作为属性信息的单词词向量进行平均求值, 本文采用实体信息作为嵌入向量。对于一段评论的情感极性, 如果考虑的实体信息不同, 情感极性可能会有相反的结果。

标准的 LSTM 只能判断句子的情感极性, 但不能检测句子的哪一部分、哪个单词对句子情感极性的判断具有重要作用。因此本文提出一种分层注意力机制网络, 在给定具体的实体信息后, 通过赋予不同单词、不同句子不同的注意力权重, 让网络更注意那些对句子情感极性判断有更重要作用的部分。分层即在句子级别, 单词级别分别加入注意力机制, 在句子级别, 注意力机制可以给予重要的单词更大的注意力权重, 在单词级别, 注意力机制可以给予重要的句子更大的注意力权重。这样不仅每一个单词有其向量表示, 每一个句子也有其向量表示, 并且单词和句子都有自己的注意力权重。图 2 代表了新提出的的分层注意力机制网络。

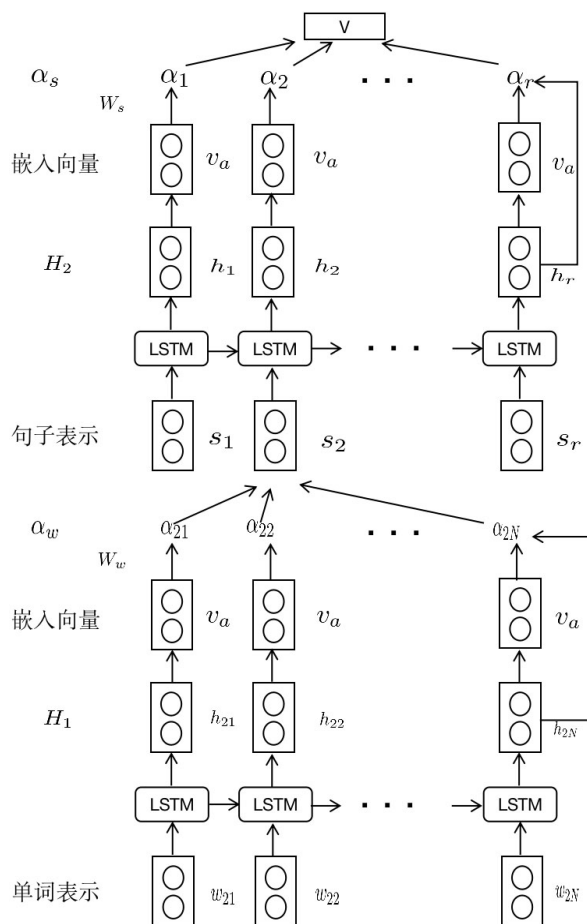


Figure 2. Layered attention network
图 2. 分层注意力网络

单词级别:

给定一个包含 N 个单词的句子, 表示为 $[w_{i1}, \dots, w_{iN}]$, 通过一个词向量矩阵 W_e , 得到每个单词的词向量表示 $x_{ij} = W_e w_{ij}$, 作为输入。然后使用 LSTM 网络得到单词级别的隐层表示, 即将输入的单词向量抽象成隐层表示, 隐层表示包含了输入的信息。

$$x_{ij} = W_e w_{ij}, j \in [1, N] \quad (7)$$

$$h_{ij} = \text{LSTM}(x_{ij}), j \in [1, N] \quad (8)$$

然后加入实体信息作为嵌入向量, 相比于 TC-LSTM 通过取目标短语包含的单词的向量平均值来获得嵌入向量, 实体信息更能表示目标短语的含义。其中嵌入向量 $v_a \in R^{N \times d}$, N 表示句子长度, d 表示隐藏层神经元个数。

并不是所有单词对句子的意义表示都有相同重要的作用, 因此引入注意力机制来提取那些对句子意义有更重要作用的单词, 通过得到的注意力权重 α 与隐层表示 h_{ij} 加权求和就可以得到句子的向量表示。

$$M_1 = \tanh(W_w [H_1, v_a]) \quad (9)$$

$$\alpha_w = \text{softmax}(u_w^T M_1) \quad (10)$$

$$s_i = H_1 \alpha_w^T \quad (11)$$

首先将嵌入向量 v_a 与 LSTM 输出的隐层表示 h_{ij} 进行拼接作为输入, 通过一层 MLP 网络得到 M_{ij} , 然后用单词级别的向量 u_w 来衡量单词的重要程度, 通过 softmax 激活函数得到归一化的注意力权重 α_{ij} , 接下来将注意力权重 α_{ij} 与 LSTM 的隐层表示进行加权求和, 就可以得到一个句子的向量表示 s_i 。其中向量 u_w 随机初始化, 在训练过程与其他参数一起参与训练。

句子级别:

Y. Wang 等人提出的 Attention-based LSTM 网络加入嵌入向量后考虑了单词级别的注意力机制, 本文提出的分层注意力机制网络则新加入一层句子级别的注意力机制, 除了对单词的不同重要程度进行标明, 对于一段评论, 可能包含多个句子, 通过对句子加入注意力机制, 可以使模型提取那些跟嵌入向量更符合的句子作为情感极性分析的依据。

将输出的句子表示 s_i 作为 LSTM 网络新的输入, 即 $[s_1, \dots, s_T]$, T 表示句子长度, 经过 LSTM 网络输出句子级别的隐层表示 h_i , 此时隐层表示抽象后的句子信息。

$$h_i = \text{LSTM}(s_i), i \in [1, T] \quad (12)$$

然后再一次加入实体信息作为嵌入向量。然后在句子级别引入注意力机制, 来提取那些对评论的情感极性有更重要作用的句子, 通过得到的注意力权重 α 与隐层表示 h_i 的输入加权求和就可以得到整个评论的向量表示。

$$M_2 = \tanh(W_s [H_2, v_a]) \quad (13)$$

$$\alpha_s = \text{softmax}(u_s^T M_2) \quad (14)$$

$$v = H_2 \alpha_s^T \quad (15)$$

这里引入句子级别的向量 u_s 来衡量句子的重要程度, 最终得到的向量 v 就是整个评论的向量表示, 类似的, 向量 u_s 也是随机初始化, 在训练过程与其他参数一起参与训练。本文参考 Rocktaschel 的想法, 将句子级别的隐层表示 h_T 加入最后整个评论的表示。

$$h^* = \tanh(W_1 v + W_2 h_T) \quad (16)$$

$$y = \text{softmax}(W h^* + b) \quad (17)$$

h^* 就是一个高层次的最终的整个评论的向量表示, 将 h^* 经过 softmax 层就可以得到最终的分类结果, 其中 W , b 分别是 softmax 层的权重与偏置。实验结果标明加入隐层表示 h_T 后取得了更好的实验结果。

3.3. 模型训练

本文模型使用交叉熵损失作为损失函数, y 代表目标评论对应的真实分类结果, \hat{y} 代表预测的分类结果, 训练目标则是最小化交叉熵损失。

$$\text{loss} = -\sum_i \sum_j y_i^j \log \hat{y}_i^j + \lambda \|\theta\| \quad (18)$$

其中, i 代表评论的下标, j 代表类别的下标。本文是一个三分类问题。 λ 代表 L_2 正则系数, θ 则代表参数集合。与标准的 LSTM 网络类似, 参数集合包括 $\{W_f, b_f, W_i, b_i, W_c, b_c, W_o, b_o, W, b\}$, 除此之外, 网络加入了嵌入向量 v_a , MLP 网络的参数 W_w , W_s , 衡量单词和句子重要程度的参数 u_w , u_s , 以及 W_1 , W_2 。因此除了包含标准 LSTM 网络的参数外, 分层注意力机制网络还包括以下参数 $\{v_a, W_w, W_s, u_w, u_s, W_1, W_2\}$ 。其中词向量和嵌入向量也是在训练过程不断优化。

训练集中大约有 5% 的单词不在已有的词向量矩阵中, 本文将这些单词随机初始化。最后, 本文实验采用 Adam 算法作为我们的优化算法。

4. 实验

本文将提出的分层注意力模型应用于细粒度的情感分析。实验中训练集的所有词向量都通过 Glove 初始化。词嵌入向量是在未标记的语料库上预先训练的,其大小约为 8400 亿,其他参数都是随机初始化得到。词嵌入向量的维度大小为 300,方面嵌入向量和隐藏层神经元的数目也为 300。单词级别的注意力机制权重 α_w 长度与每个句子包含的单词数相同,句子级别的注意力机制权重长度 α_s 与每篇评论包含的句子数相同。本文采用 tensorflow 来进行实验,其中 batch size 大小为 64, L_2 正则化系数为 0.5,学习率为 0.001。

4.1. 数据集

本文实验采用的数据集是 SemEval 2014 Task 4。数据集由顾客评论组成,每一篇评论对应一个实体信息以及相应的情感极性。我们的目标就是给定特定实体判断相应的情感极性。具体的数据统计见表 1。

Table 1. Data set emotional polarity classification statistics table

表 1. 数据集情感极性分类统计表

Asp	Positive		Negative		Neural	
	Train	Test	Train	Test	Train	Test
Fo	867	302	209	69	90	31
Pr	179	51	115	28	10	1
Se	324	101	218	63	20	3
Am	263	76	98	21	23	8
An	546	127	199	41	357	51
Total	2179	657	839	222	500	94

4.2. 实验结果

4.2.1. 模型效果对比

本文首先将改进的模型与 TC-LSTM, TD-LSTM 等进行了对比。采用的是 SemEval 2014 Task 4 下的“restaurants”数据集,该数据集包含 aspect 与 aspect-term 两个嵌入信息,对分别应“实体”,“属性”信息。对于给定的句子,采用实体信息作为嵌入信息,来判断给定实体对应的句子情感极性。比如,“这家餐厅实在是太贵了”,实体信息对应的是“价格”,数据集总共包含了五种实体信息,分别为食物,价格,服务,环境,其他。表 2 是实验结果。

Table 2. Entity information as the classification accuracy of each model of embedded information

表 2. 实体信息作为嵌入信息的各模型分类准确率

Models	Three-way	Two-way
LSTM	82.0	88.3
TD-LSTM	82.6	89.1
TC-LSTM	81.9	89.2
AE-LSTM	82.5	88.9
AT-LSTM	83.1	89.6
ATAE-LSTM	84.0	89.9
HAT-LSTM	85.1	91.3

在将实体信息作为嵌入信息的前提下, 可以看出新改进的模型在实验结果上有了一定的提升, 特别是在二分类问题上已经能够达到 90% 的准确率。

4.2.2. 嵌入信息效果对比

实验还是采用 SemEval 2014 Task 4 下的 “restaurants” 数据集, 模型也跟之前的基本相同, 不同的是于给定的句子, 采用属性信息作为嵌入信息, 来判断给定实体对应的句子情感极性。比如, “这家餐厅实在是太贵了”, 将 “餐厅” 作为嵌入信息, 来判断对应的情感极性。表 3 是实验结果。

Table 3. Attribute information as the classification accuracy of each model of embedded information

表 3. 属性信息作为嵌入信息的各模型分类准确率

Models	Three-way	Two-way
TD-LSTM	75.6	-
TC-LSTM	74.3	-
AE-LSTM	76.6	89.6
ATAE-LSTM	77.2	90.9
HAT-LSTM	78.3	90.3

可以看出相对于实体信息, 属性信息作为嵌入信息实验效果会有明显的下滑, 说明在细粒度情感分析中实体信息比属性信息更重要。

4.3. 注意力机制的有效性

本文采用了注意力机制来赋予不同的单词和句子不同的权重, 来更好的进行情感极性的判断。通过公式 (10), (14) 可以得到注意力权重的具体数值, 下面通过可视化的注意力机制的图来展示注意力机制的有效性。

对于例句, “The noodles tastes good, and the price is reasonable, but the service is not good”, 图 3 展示了在给定实体为 “food” 的时候, 句子中每个单词被赋予的注意力权重, 可以看出 “noodles”, “good” 被赋予了更高的权重, 说明在句子情感极性的判断中二者起到了重要的作用, 这也与我们的预期相符。

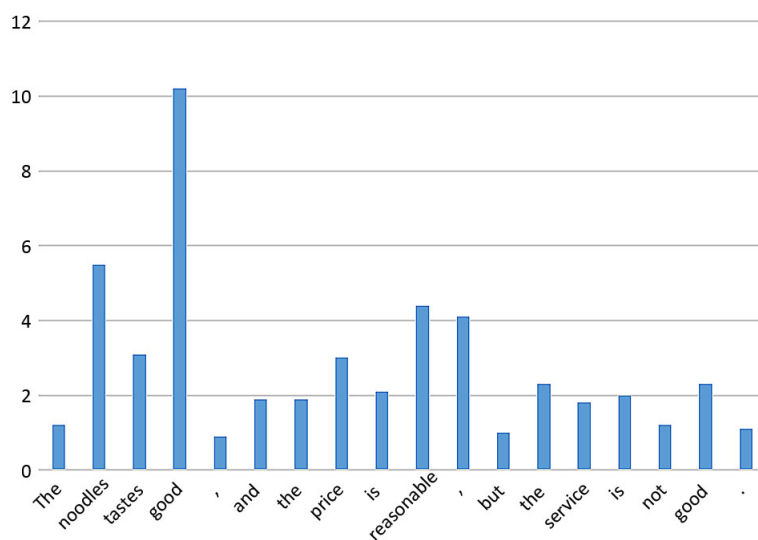


Figure 3. Word level attention weight distribution diagram

图 3. 单词级别注意力权重分布示意图

更进一步, 对于该评论中的三个句子, “The noodles tastes good”, “and the price is reasonable”, “but the service is not good”, 得到每个句子的注意力权重并且可视化, 结果如图 4, 可以看出第一个句子在情感极性的判断中起到了重要作用, 而这一句子正是实体 “food” 所对应的句子。

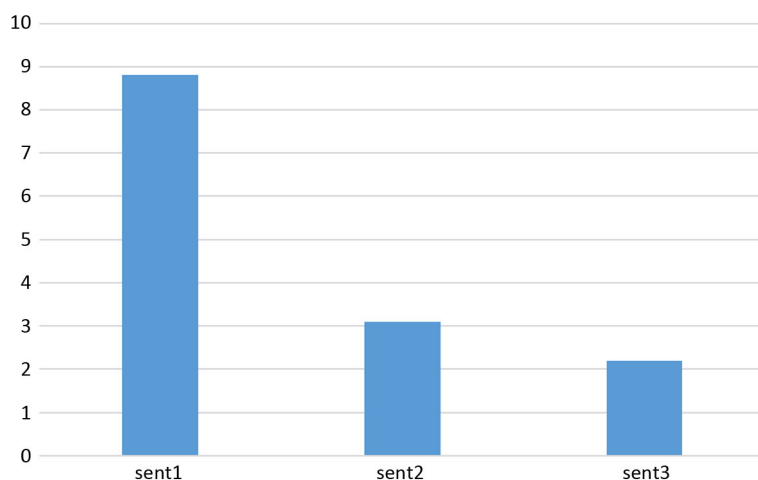


Figure 4. Sentence level attention weight distribution diagram

图 4. 句子级别注意力权重分布示意图

类似的, 将实体 “service” 作为嵌入信息时, 每个单词及句子的注意力权重可视化如图 5 和图 6。

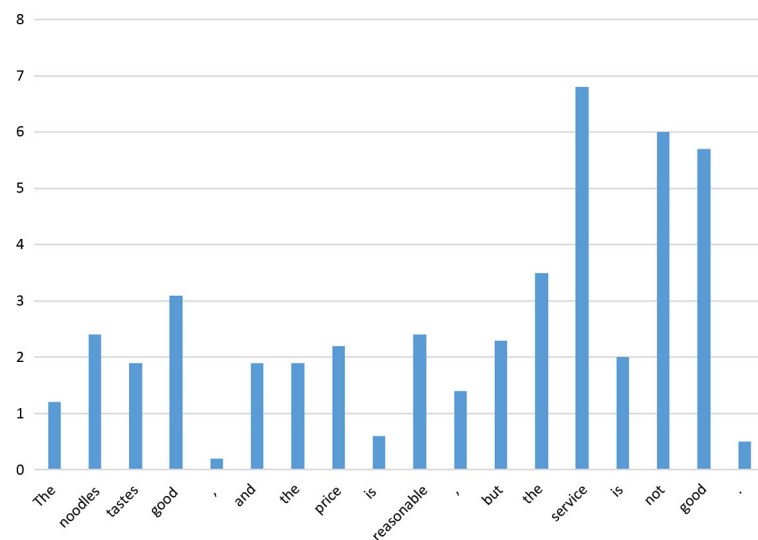


Figure 5. Word level attention weight distribution diagram

图 5. 单词级别注意力权重分布示意图

与预期的结果相同, 在单次级别的注意力权重计算中, “service”, “not”, “good” 占有比较大的权重, 而在句子级别的注意力权重计算中, “but the service is not good” 占有的权重则最大。

5. 结论

本文提出了一种用于细粒度情感分析的分层注意力机制网络, 一方面加入实体信息作为嵌入信息, 生成包含实体信息的句子表示, 使模型能够在给定实体下更好地判定相应情感极性, 另一方面采用分层注意力机制, 在单词, 句子两个层面分别加入注意力机制, 使模型能赋予那些更重要的单词, 句子更大

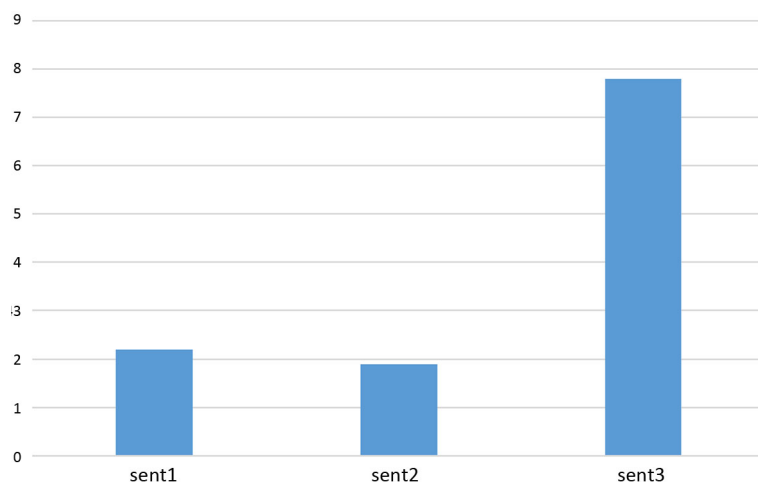


Figure 6. Sentence level attention weight distribution diagram

图 6. 句子级别注意力权重分布示意图

的权重。实验结果表明新改进的 HAT-LSTM 比之前的模型取得了更好的效果。对注意力权重的可视化也说明了模型可以有效地挑选出那些更重要的单词和句子。

参考文献

- [1] Nasukawa, T. and Yi, J. (2003) Sentiment Analysis: Capturing Favorability Using Natural Language Processing. In: *Proceedings of the 2nd International Conference on Knowledge Capture*, ACM, New York, 70-77. <https://doi.org/10.1145/945649.945658>
- [2] Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K. and Dyer, C. (2016) Neural Architectures for Named Entity Recognition. *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, San Diego, June 2016, 260-270. <https://doi.org/10.18653/v1/N16-1030>
- [3] Yin, W.P., Schutze, H., Xiang, B. and Zhou, B. (2015) ABCNN: Attention-Based Convolutional Neural Network for Modeling Sentence Pairs.
- [4] Golub, D. and He, X.D. (2016) Character-Level Question Answering with Attention.
- [5] Rush, A.M., Chopra, S. and Weston, J. (2015) A Neural Attention Model for Abstractive Sentence Summarization.
- [6] Kim, Y. (2014) Convolutional Neural Networks for Sentence Classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, October 2014, 1746-1751. <https://doi.org/10.3115/v1/D14-1181>
- [7] Tang, D., Qin, B. and Liu, T. (2015) Document Modeling with Gated Recurrent Neural Network for Sentiment Classification. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, Lisbon, 17-21 September 2015, 1422-1432. <https://doi.org/10.18653/v1/D15-1167>
- [8] Zhang, X., Zhao, J.B. and Le Cun, Y. (2015) Character-Level Convolutional Networks for Text Classification.
- [9] Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T. and Qin, B. (2014) Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Volume 1, 1555-1565. <https://doi.org/10.3115/v1/P14-1146>
- [10] Tang, D., Qin, B., Feng, X.C. and Liu, T. (2015) Target-Dependent Sentiment Classification with Long Short Term Memory.
- [11] Bahdanau, D., Cho, K. and Bengio, Y. (2014) Neural Machine Translation by Jointly Learning to Align and Translate.
- [12] Li, J.W., Luong, M.-T. and Jurafsky, D. (2015) A Hierarchical Neural Autoencoder for Paragraphs and Documents.
- [13] Sukhbaatar, S., Szlam, A., Weston, J. and Fergus, R. (2015) End-to-End Memory Networks.
- [14] Yang, Z.C., He, X.D., Gao, J.F., Deng, L. and Smola, A. (2015) Stacked Attention Networks for Image Question Answering. *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 8-10. <https://doi.org/10.1109/CVPR.2016.10>

- [15] Rao, D. and Ravichandran, D. (2009) Semisupervised Polarity Lexicon Induction. In: *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, Stroudsburg, 675-682. <https://doi.org/10.3115/1609067.1609142>
- [16] Mohammad, S.M., Kiritchenko, S. and Zhu, X.D. (2013) NRC-Canada: Building the State-of-the-Art in Sentiment Analysis of Tweets.