



一种注意力增强的自然语言推理模型

李冠宇^{a,b}, 张鹏飞^{a,b}, 贾彩燕^{a,b}

(北京交通大学 a. 计算机与信息技术学院; b. 交通数据分析与挖掘北京市重点实验室, 北京 100044)

摘 要: 在自然语言处理任务中使用注意力机制可准确衡量单词重要度。为此, 提出一种注意力增强的自然语言推理模型 aESIM。将词注意力层以及自适应方向权重层添加到 ESIM 模型的双向 LSTM 网络中, 从而更有效地学习单词与句子表示, 同时提高前提与假设文本之间局部推理的建模效率。在 SNLI、MultiNLI 及 Quora 数据集上的实验结果表明, 与 ESIM、HBMP、SSE 等模型相比, aESIM 模型的准确率能够提升 0.5% ~ 1%。

关键词: 自然语言处理; 自然语言推理; ESIM 模型; 注意力机制; 双向 LSTM 网络

开放科学(资源服务)标志码(OSID):



中文引用格式: 李冠宇, 张鹏飞, 贾彩燕. 一种注意力增强的自然语言推理模型[J]. 计算机工程, 2020, 46(7): 91-97.

英文引用格式: LI Guanyu, ZHANG Pengfei, JIA Caiyan. An attention-enhanced natural language reasoning model[J]. Computer Engineering, 2020, 46(7): 91-97.

An Attention-enhanced Natural Language Reasoning Model

LI Guanyu^{a,b}, ZHANG Pengfei^{a,b}, JIA Caiyan^{a,b}

(a. School of Computer and Information Technology; b. Beijing Key Lab of Traffic Data Analysis and Mining, Beijing Jiaotong University, Beijing 100044, China)

[Abstract] In natural language processing tasks, the attention mechanism can be used to evaluate the importance of a word. On this basis, this paper proposes an attention-enhanced natural language reasoning model, aESIM. The model adds the word attention layer and the adaptive direction weight layer to the bidirectional LSTM network of the ESIM model, so as to learn the representation of words and sentences more effectively, and increase the modelling efficiency of local inference between premises and hypothetical texts. Experimental results on datasets of SNLI, MultiNLI and Quora show that, compared with ESIM, HBMP, SSE and other models, aESIM increases the accuracy rate by 0.5% ~ 1%.

[Key words] natural language processing; natural language reasoning; ESIM model; attention mechanism; bidirectional LSTM network

DOI: 10.19678/j.issn.1000-3428.0054953

0 概述

自然语言推理也称为文本蕴涵识别^[1], 是自然语言处理中的一项重要任务, 其主要关注的问题是由前提文本推断出假设文本。此项任务需要理解前提和假设文本之间的语义相似性来判断它们之间的关系, 包括蕴涵、矛盾和中立 3 种关系^[2-3]。

自然语言推理任务通常被视为一种关系分类任务, 其在较大的训练集中学习前提和假设文本之间的关系, 在此基础上预测新的前提和假设文本对之间的关系。现有自然语言推理模型大致分为基于特

征的推理模型^[4]和基于神经网络的推理模型^[5-6]两类。基于特征的推理模型主要通过句子的非词汇特征(两个句子长度差的绝对值)和词汇化特征(如 n-gram 特征)的学习来表示前提与假设文本之间的关系类别。近年来, 端到端神经网络模型在机器翻译、情感分类等自然语言处理任务中表现较好, 受到大量学者的关注。

虽然将深度学习引入自然语言推理任务中, 在现阶段取得了较好的效果, 但仍存在不足, 如传统的双向 LSTM 网络^[7]忽略了重要单词对于句子向量的贡献, 每个单词的相同权重造成句子表征不够准确,

基金项目: 国家自然科学基金(61876016); 中央高校基本科研业务费专项资金(2017JBM023)。

作者简介: 李冠宇(1993—), 男, 硕士研究生, 主研方向为自然语言处理、机器学习; 张鹏飞, 硕士研究生; 贾彩燕, 教授、博士。

收稿日期: 2019-05-20 修回日期: 2019-08-09 E-mail: 17120379@bjtu.edu.cn

以及前向向量与后向向量被简单地连接到一起,未考虑到每个方向对于句子语义表示的重要程度,而造成句意缺失。

针对以上问题,本文在 ESIM 模型双向 LSTM 网络中引入注意力机制与自适应的权重计算层,分别刻画不同单词和不同方向对于句子向量的贡献度,构建一种注意力增强的自然语言推理模型。

1 相关工作

对于自然语言推理,基于特征的推理模型多为基于统计的方法,对文本的语义刻画不够深入。将深度学习引入自然语言推理领域,端到端深度学习模型取得了较好的效果。

将基于神经网络的自然语言推理模型分为句子编码模型和句子交互模型两类,两种模型的架构如图 1 所示。

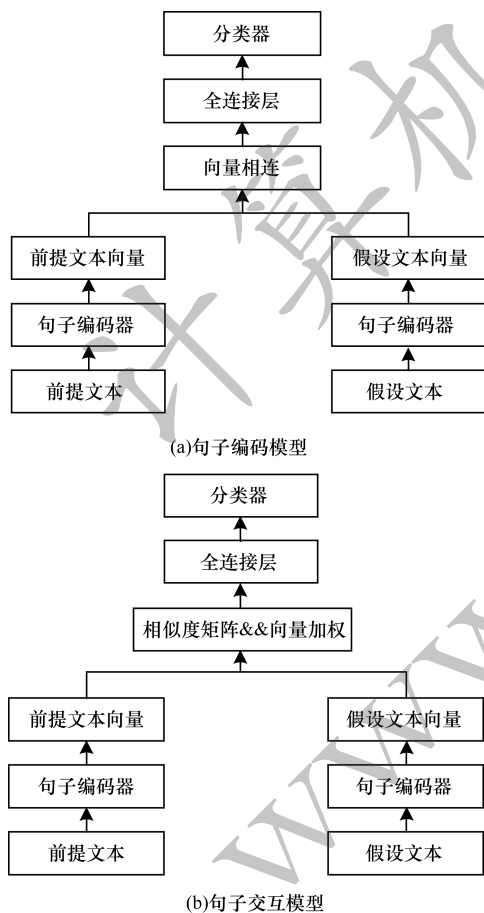


图 1 两种自然语言推理模型架构

Fig. 1 Architectures of two natural language reasoning models

句子编码模型^[8-9]利用预先训练好的词向量来进行编码,将前提文本与假设文本分别编码为文本向量^[10-11],两个文本向量相连之后再利用全连接层来学习两个文本之间的语义关系,主要架构如

图 1(a) 所示。在此类模型中, LSTM^[12] (Long Short-Term Memory) 网络及其变体 GRU^[13] (Gated Recurrent Units) 以及双向 LSTM 能够学习句子内部的长距离依赖关系,因此常被用于句子编码。文献[8]提出了一种通用的自然语言推理任务的训练方案,并对比了 LSTM、GRU、最大(平均)池化双向 LSTM、自注意力机制网络以及分层的卷积神经网络等编码架构。实验结果表明,采用最大池化层的双向 LSTM 网络具有最佳的性能。文献[9]设计了一个分层的双向 LSTM 模型(HBMP)进行句子编码,该模型采用上一层双向 LSTM 的参数初始化下一层双向 LSTM 的参数进行信息传递,比使用单一的双向 LSTM 层效果更佳。文献[7]则将单层双向 LSTM 网络替换成多层双向 LSTM 网络来加强模型的句子编码能力。

句子交互模型^[14-16]与句子编码模型架构相似,首先学习句子对的向量表示,使用更新的单词向量表示计算两个句子之间的相似度矩阵,然后利用相似度矩阵进行句子向量的增强表示,最后做出分类判断,主要架构如图 1(b) 所示。与句子编码模型相比,句子交互模型能够利用句子对之间的文本相似性,并且能够捕获前提文本与假设文本之间的潜在语义信息。文献[17]将注意力机制引入机器翻译任务中,同步完成翻译和文本对齐任务。文献[18]设计了一个成对词交互模型(PWIM),充分利用了词级细粒度信息,取得了较好的效果。文献[15]提出了双向多视角匹配模型(BiMPM),对比了多种不同注意力机制的匹配策略。文献[12]提出了 ESIM 模型,利用两层不同的双向 LSTM 分别进行文本语义编码。实验结果表明,句子交互模型,特别是 ESIM 模型,优于上述的句子编码模型。

尽管 ESIM 模型表现优异,但该模型并没有考虑双向 LSTM 层中每个单词的不同贡献。因此,将注意力机制引入双向 LSTM 层将有利于刻画词的重要性。已有的双向 LSTM 层对于前向和后向向量的重要度不加区分,传统的双向 LSTM 层中只是将它们简单地连接到一起,但前向与后向向量都代表着信息的流动方向,因此,有必要考虑每个方向(向前或向后)对字向量编码是否具有不同的重要性,从而自适应性地将两个方向向量以不同的权重联合在一起。本文使用 ESIM 模型作为基线,在每个双向 LSTM 层后添加一个词注意力层,并使用自适应方向权重层来学习前向和后向向量对于词向量的贡献度,将此注意力增强的双向 LSTM 模型命名为双向 aLSTM 模型,并将修改后的 ESIM 模型命名为 aESIM 模型。

2 aESIM 模型框架

假定有 2 个句子,前提文本 $p = (p_1, p_2, \dots, p_{l_p})$ 与假设文本 $q = (q_1, q_2, \dots, q_{l_q})$,任务目标是预测它们的关系标签 y 。

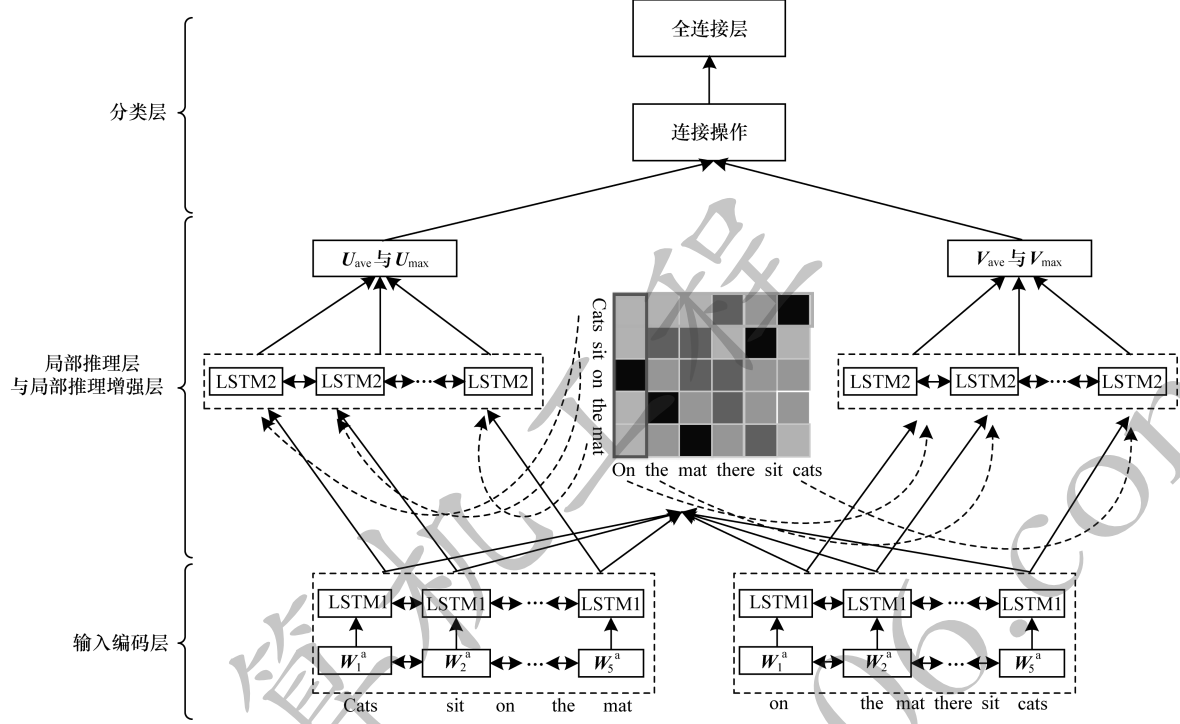


图 2 ESIM 模型架构

Fig.2 Architecture of ESIM model

在输入编码层中,ESIM 模型利用预训练的词向量(如 GloVe 向量^[20])在双向 LSTM 层中对输入的句子对进行编码:

$$\bar{p}_i = \text{Bi-LSTM}(p, i), \forall i \in [1, 2, \dots, l_p] \quad (1)$$

$$\bar{q}_j = \text{Bi-LSTM}(q, j), \forall j \in [1, 2, \dots, l_q] \quad (2)$$

其中, (p, i) 代表句子 p 中的第 i 个词, (q, j) 代表句子 q 中的第 j 个词。

$$M = \bar{p}^T \times \bar{q} \quad (3)$$

$$\tilde{p}_i = \frac{\sum_{j=1}^{l_q} \exp(M_{ij})}{\sum_{k=1}^{l_q} \exp(M_{ik})} \bar{q}_j, \forall i \in [1, 2, \dots, l_p] \quad (4)$$

$$\tilde{q}_j = \frac{\sum_{i=1}^{l_p} \exp(M_{ij})}{\sum_{k=1}^{l_p} \exp(M_{kj})} \bar{p}_i, \forall j \in [1, 2, \dots, l_q] \quad (5)$$

局部推理层用于增强句向量表示,将输入编码层输出的前提文本向量 \bar{p} 与假设文本向量 \bar{q} 计算相似度矩阵 M (见式(3)),再通过矩阵 M 计算前提文本向量与假设文本向量的新表示 \tilde{p} 和 \tilde{q} (见式(4)、式(5))。

如式(6)、式(7)所示, \tilde{p} 和 \tilde{q} 表示的是前提文

2.1 ESIM 模型

ESIM 模型主要由输入编码层、局部推理层、局部推理增强层以及分类层 4 个部分组成,模型架构如图 2 所示。

本向量 \bar{p} 与假设文本向量 \bar{q} 的加权求和。通过局部推理增强层,可以得到前提文本与假设文本的加权表示向量,并输入到下一个 LSTM 层中进行局部信息提取。

$$m_p = [\bar{p}; \tilde{p}; \bar{p} - \tilde{p}; \bar{p} \odot \tilde{p}] \quad (6)$$

$$m_q = [\bar{q}; \tilde{q}; \bar{q} - \tilde{q}; \bar{q} \odot \tilde{q}] \quad (7)$$

在分类层中,ESIM 模型并未用到文献[14]提出的向量求和的方法,而是提出同时使用最大池化与平均池化的方法,并将拼接的结果送入最后的多层全连接层进行最终分类。

2.2 aESIM 模型

本文基于 ESIM 模型提出了注意力增强的自然语言推理模型 aESIM。类似于 ESIM 模型,aESIM 模型仍包含编码层、局部推理层、局部推理增强层以及分类层 4 个主要部分。ESIM 模型与 aESIM 模型的区别在于将 ESIM 模型中的两层双向 LSTM 层替换为两层双向 aLSTM 层。如图 2 所示,ESIM 模型中的虚线框将被替换为双向 aLSTM 层,在右上角用实线框标出,双向 aLSTM 层的细节展示如图 3 所示。

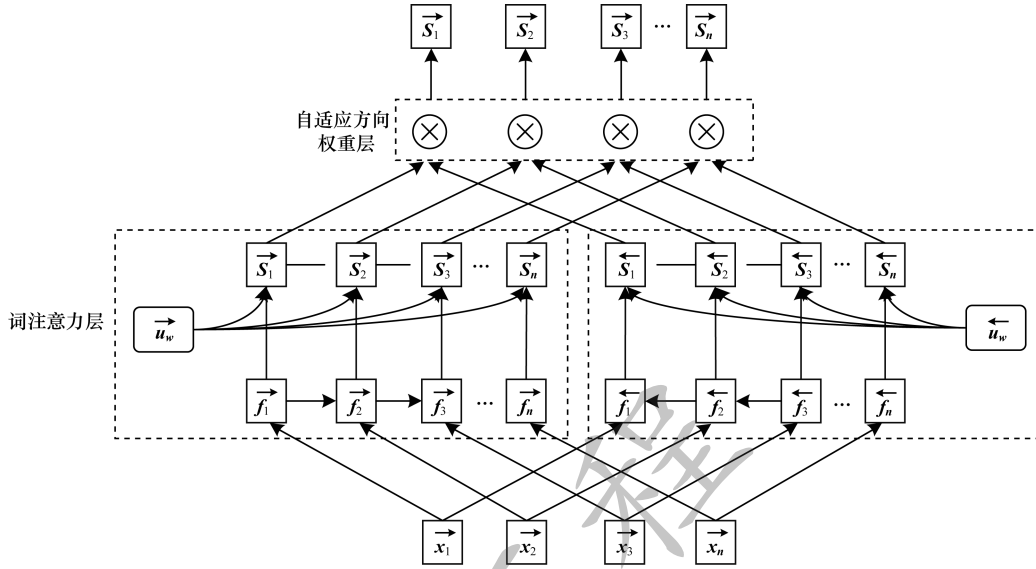


图3 基于方向注意力机制的双向 LSTM 结构

Fig.3 Architecture of bidirectional LSTM based on directional attention mechanism

在 aESIM 模型编码层中, 给定一个词向量 x_{il} , $l \in [1, T]$ 代表的是语料库中第 i 个句子中第 l 个单词, 可利用一个前向 LSTM 层和一个后向 LSTM 层来进行初步编码得到 \vec{f} 和 \overleftarrow{f} :

$$\vec{f}_{il} = \text{LSTM}(\vec{x}_{il}), l \in [1, T] \quad (8)$$

$$\overleftarrow{f}_{il} = \text{LSTM}(\overleftarrow{x}_{il}), l \in [1, T] \quad (9)$$

aLSTM 模型在传统的双向 LSTM 模型中加入了词注意力层以及自适应方向权重层来增强模型句子表征能力。

2.2.1 词注意力层

在一个句子中, 并非所有单词对与句向量的贡献都是相同的。类似于文献[5]的方法, 本文引入了相似的注意力机制, 它是一项衡量单词重要度且十分有效的工具。在词向量经过双向 LSTM 层之后, 通过添加词注意力层来计算词的重要度以增强模型的语义表示能力。

类似于文献[5]的方法, 假设 $f_{il} \in \{\vec{f}_{il}, \overleftarrow{f}_{il}\}$, 利用式(10)、式(11)计算句子中每个单词的重要度:

$$u_{il} = \tanh(Wf_{il} + b) \quad (10)$$

$$\alpha_{il} = \frac{\exp(u_{il}^T u_w)}{\sum_l \exp(u_{il}^T u_w)} \quad (11)$$

$$s_{il} = \alpha_{il} * f_{il} \quad (12)$$

其中, u_{il} 是 f_{il} 通过一个单层的感知机得到的, u_w 是一个随机初始化的文本向量, 并且可以在训练过程中进行更新, 通过一个 SoftMax 单元计算每个单词对于句向量的权重, 通过将权重 α_{il} 与原始向量 f_{il} 相

乘可以得到注意力增强的向量表示 s_{il} , $s_{il} \in \{\vec{s}_{il}, \overleftarrow{s}_{il}\}$, 包含前向文本向量与后向文本向量。

2.2.2 自适应方向权重层

在传统的双向 LSTM 层中, 每个单词的前向与后向向量被认为在生成最终词向量时有着相同的权重, 因此, 前向和后向向量被简单地拼接在一起, 但这并不能体现信息流动方向的差异性。对一个单词, 前面的单词与后面的单词在不同的方向上是相反的, 而不同方向的词向量对于词向量的贡献度可能是不同的。因此, 本文设计一个新的自适应方向权重层来学习不同方向的词向量对于最终的词向量的贡献度。

给定 2 个方向的词向量 \vec{s}_{il} 和 \overleftarrow{s}_{il} , 最终的词向量可以通过式(13)计算:

$$\overline{s}_{il} = g[(W_F \times \vec{s}_{il} + b_F) \cdot (W_B \times \overleftarrow{s}_{il} + b_B)] \quad (13)$$

其中, W_* 与 b_* 表示权重矩阵以及偏置项, g 表示非线性变换, $[\cdot]$ 表示连接操作。在训练过程中, 所有的参数都可以进行更新。利用式(14)和式(15)可以得到整个句子的向量表示:

$$\overline{p}_i = \text{Bi-aLSTM}(\overline{s}_{il}), \forall [1, 2, \dots, l_p] \quad (14)$$

$$\overline{q}_j = \text{Bi-aLSTM}(\overline{s}_{jl}), j \in [1, 2, \dots, l_q] \quad (15)$$

本文将添加了词注意力层以及自适应权重方向层增强的双向 LSTM 层的网络称作双向 aLSTM。双向 aLSTM 层被应用到 ESIM 模型中替换其中的双向 LSTM 层。

3 实验与结果分析

3.1 数据集

本文在 SNLI、MultiNLI^[19] 和 Quora 3 个公开自然语言推理数据集上进行实验。

1) SNLI 数据集共包含了共计 570 152 个句子对,其中训练集包含 5.49×10^5 个句子对,验证集与测试集各包含 1×10^5 个句子对。每一个句子对都有唯一的一个类标签(“蕴涵”“中立”“矛盾”以及“-”)。标签为“-”表示句子对之间没有明确的可分类结果,因此,在训练,验证以及测试中移除了这些关系不明确的句子对。

2) MultiNLI 语料库是由源文本蕴含信息注释的 4.33×10^5 句子对组成的众包集合。该语料库以 SNLI 语料库为模型,但不同之处在于涵盖了一系列口语和书面文本,并支持独特的跨类型生成评估。

3) Quora 数据集包含 4×10^5 个问题对。这个语料库旨在判断两个句子是否具有相同的语义关系来判断两个文本是否要表达相同的意图。

3.2 实验参数设置

本文使用验证集进行模型选择与参数调试。aESIM 模型的超参数设置如下:实验使用 Adam 算法作为模型的优化方法,第 1 个动量设定为 0.9,第 2 个动量设定为 0.999。初始化学率设置为 0.000 5,训练批量大小设置为 128。所有双向 aLSTM 层以及单词嵌入的隐藏状态的维度设置为 300。实验使用了非线性函数 $f = \text{selu}$,因为它具有比 Relu 更快的收敛速度。失活率在训练过程中设置为 0.2,实验使用预训练的 300D 的 GloVe 向量来初始化词嵌入。未登录词用高斯随机函数来随机初始化。该算法完全使用 Keras 框架实现,并在 TITAN Xp 显卡上进行训练。

3.3 实验结果

本文在 SNLI 数据集上比较多个方法性能,如表 1 所示。其中,第 1 行方法是传统特征工程模型,第 2 行~第 6 行方法是句子向量编码模型,第 7 行、第 8 行方法是句子交互模型,第 9 行、第 10 行方法是基准模型 ESIM 以及本文提出的 aESIM 模型。ESIM 以及 aESIM 模型是基于 Keras 实现的。对于 MultiNLI 数据集,主要用到的基准模型有 CBOW、双向 LSTM、PWIM 模型以及 ESIM 模型,结果如表 2 所示。Quora 数据集的结果如表 3 所示。对于每一种方法,本文均使用准确率来进行评价。

表 1 不同模型在 SNLI 数据集上的准确率

Table 1 Accuracies of different models on SNLI dataset %

模型	准确率
非词法特征 + 单字与双字节特征模型 ^[4]	78.2
LSTM 编码器 ^[8]	80.7
Inner-attention 编码器 ^[8]	82.5
双向 LSTM 拼接最大池化层模型 ^[8]	84.5
HBMP 模型 ^[9]	86.0
SSE 模型 ^[7]	86.0
PWIM 模型 ^[18]	82.2
BiMPM 模型 ^[15]	87.5
ESIM 模型	87.3
aESIM 模型	88.1

表 2 不同模型在 MultiNLI 数据集上的准确率

Table 2 Accuracies of different models on MultiNL dataset %

模型	匹配集	非匹配集
CBOW 模型	64.8	64.5
双向 LSTM 模型	66.9	72.2
PWIM 模型	66.9	71.6
ESIM 模型	73.4	73.5
aESIM 模型	73.9	73.9

表 3 不同模型在 Quora 数据集上的准确率

Table 3 Accuracies of different models on Quora dataset %

模型	准确率
Siamese-CNN 模型	79.60
Siamese-LSTM 模型	82.58
PWIM 模型 ^[18]	83.40
SSE 模型 ^[7]	87.80
ESIM 模型	86.98
aESIM 模型	88.01

从表 1~表 3 可以看出,aESIM 模型在 SNLI 数据集上达到了 88.1% 的准确率,相比于 ESIM 模型上提升了 0.8 个百分点,在 MultiNLI 数据集上相比于 ESIM 提升了 0.5 个百分点,并且表现出了最好的性能。在 Quora 数据集上,达到了 88.01% 的准确率,对比于 ESIM 模型提升了 1 个百分点。由此可以得出结论,本文提出的 aESIM 在 3 个数据集上都表现出优于 ESIM 的性能,表明引入该注意力机制与自适应方向权重的双向 LSTM 模型具备更强的句子编码能力。

3.4 注意力机制可视化

为了更直观地展示双向 aLSTM 的提升效果,本文进行了可视化展示,如图 4~图 6 所示。从 SNLI 测试集中抽取了一个前提文本以及 3 个相对应的假设文本。其中,前提文本是“A woman with a green headscarf, blue shirt and a very big grin”,3 个假设文本分别为“the woman has been shot”“the woman is very happy”“the woman is young”,对应的标签分别是“矛盾”“蕴涵”以及“无关”。

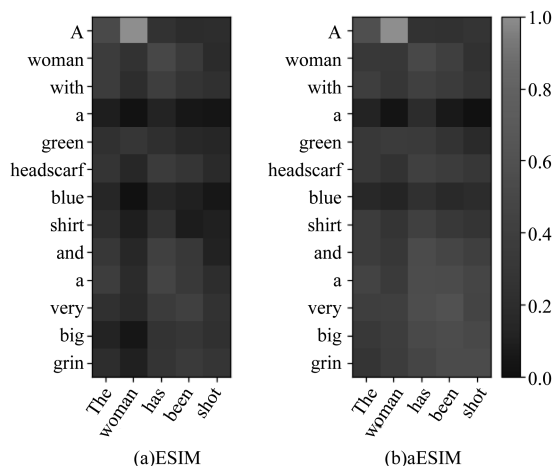


图 4 矛盾句可视化展示

Fig. 4 Visual presentation of contradictory sentence

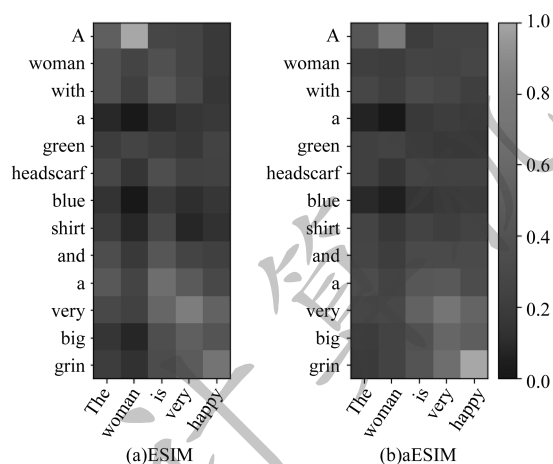


图 5 蕴涵句可视化展示

Fig. 5 Visual presentation of implication sentence

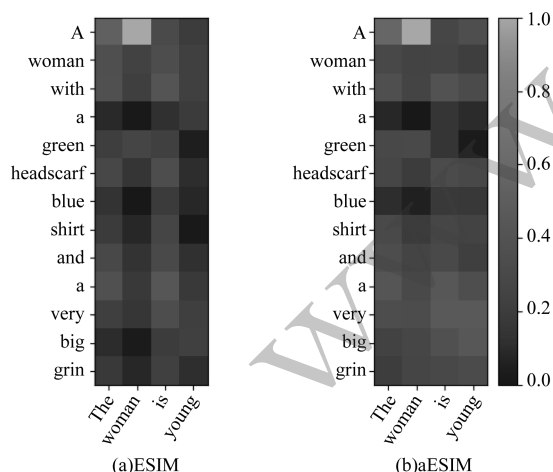


图 6 中立句可视化展示

Fig. 6 Visual presentation of neutral sentence

从图 4 ~ 图 6 可以看出,每对前提假设句子对都会有相应的关键词对,如“grin-shot”“grin-happy”以及“grin-young”,这些关键词对于前提文本能否推断

出假设文本具有决定性的作用。图 4 ~ 图 6 分别表示 ESIM 模型与 aESIM 模型中句向量分别经过双向 LSTM 层以及双向 aLSTM 层之后的可视化结果,这样可以直接观察到两个词之间的相似度。在每幅图中,颜色越浅,表明相似度越大。可以看出,aESIM 模型对于关键词对的相似度计算结果均高于 ESIM 模型的计算结果。尤其是在图 5 中,“happy”与“grin”在 aESIM 模型中相似度明显高于在 ESIM 模型中的相似度。因此,本文提出的 aESIM 模型相比于 ESIM 模型有着更强的信息提取以及句子表征能力,在计算词相似度方面也有更好的表现。

4 结束语

本文在 ESIM 模型的基础上提出注意力增强的自然语言推理模型 aESIM。在双向 LSTM 层中引入注意力机制与自适应的权重计算层,将前向与后向向量加权相连,分别刻画不同单词和不同方向对于句子向量的贡献度。在 SNLI、MultiNLI 以及 Quora 3 个自然语言推理公开数据集上的实验结果表明,与 ESIM 模型相比,aESIM 模型表现出更优的性能,具备较强的信息提取能力。下一步将对注意力机制的应用方法进行深入研究,同时构建基于字符以及单词联合表示的词向量模型。

参考文献

- [1] ZHANG Zhichang, YAO Dongren, LIU Xie, et al. Textual entailment recognition fused with syntactic structure transformation and lexical semantic features[J]. Computer Engineering, 2015, 41(9): 199-204. (in Chinese)
张志昌, 姚东任, 刘霞, 等. 融合句法结构变换与词汇语义特征的文本蕴涵识别[J]. 计算机工程, 2015, 41(9): 199-204.
- [2] LAN Wuwei, XU Wei. Neural network models for paraphrase identification, semantic textual similarity, natural language inference, and question answering[C]// Proceedings of the 27th International Conference on Computational Linguistics. Santa Fe, USA: [s. n.], 2018: 3890-3902.
- [3] GUO Maosheng, ZHANG Yu, LIU Ting. Research advances and prospect of recognizing textual entailment and knowledge acquisition[J]. Chinese Journal of Computers, 2017, 40(4): 119-140. (in Chinese)
郭茂盛, 张宇, 刘挺. 文本蕴含关系识别与知识获取研究进展及展望[J]. 计算机学报, 2017, 40(4): 119-140.
- [4] BOWMAN S R, ANGELI G, POTTS C, et al. A large annotated corpus for learning natural language inference[J]. Empirical Methods in Natural Language Processing, 2015, 41: 632-642.
- [5] YANG Z, YANG D, DYER C, et al. Hierarchical attention networks for document classification [EB/OL]. [2019-04-10]. <https://arxiv.org/abs/1707.00896v1>.

- [6] CHEN Qian, ZHU Xiaodan, LING Zhenhua, et al. Enhanced LSTM for natural language inference [C]// Proceedings of IEEE Meeting on Association for Computational Linguistics. Washington D. C., USA: IEEE Press, 2017: 1657-1668.
- [7] NIE Y, BANSAL M. Shortcut-stacked sentence encoders for multi-domain inference [C]// Proceedings of the 2nd Workshop on Evaluating Vector Space Representations for NLP. Washington D. C., USA: IEEE Press, 2017: 165-178.
- [8] CONNEAU A, KIELA D, SCHWENK H, et al. Supervised learning of universal sentence representations from natural language inference data [EB/OL]. [2019-04-10]. <https://arxiv.org/abs/1705.02364v5>.
- [9] TALMAN A, YLIJYRA A, TIEDEMANN J, et al. Natural language inference with hierarchical BiLSTM max pooling architecture [EB/OL]. [2019-04-10]. <https://arxiv.org/abs/1808.08762v1>.
- [10] IM J, CHO S. Distance-based self-attention network for natural language inference [EB/OL]. [2019-04-10]. <https://arxiv.org/abs/1712.02047v1>.
- [11] SHEN Tao, ZHOU Tianyi, LONG Guodong, et al. Reinforced self-attention network: a hybrid of hard and soft attention for sequence modeling [EB/OL]. [2019-04-10]. <https://arxiv.org/abs/1801.10296>.
- [12] CHENG J, DONG L, LAPATA M, et al. Long short-term memory-networks for machine reading [EB/OL]. [2019-04-10]. <https://arxiv.org/abs/1601.06733>.
- [13] CHUNG J, GULCEHRE C, CHO K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [EB/OL]. [2019-04-10]. <https://arxiv.org/abs/1412.3555>.
- [14] PARIKH A P, TACKSTROM O, DAS D, et al. A decomposable attention model for natural language inference [EB/OL]. [2019-04-10]. <https://arxiv.org/abs/1606.01933v2>.
- [15] WANG Z, HAMZA W, FLORIAN R, et al. Bilateral multi-perspective matching for natural language sentences [C]// Proceedings of IEEE International Joint Conference on Artificial Intelligence. Washington D. C., USA: IEEE Press, 2017: 4144-4150.
- [16] KIM S, KANG I, KWAK N. Semantic sentence matching with densely-connected recurrent and co-attentive information [EB/OL]. [2019-04-10]. <https://arxiv.org/abs/1805.11360>.
- [17] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate [EB/OL]. [2019-04-10]. <https://arxiv.org/abs/1409.0473>.
- [18] HE H, LIN J J. Pairwise word interaction modeling with deep neural networks for semantic similarity measurement [C]// Proceedings of 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, USA: [s. n.], 2016: 937-948.
- [19] WILLIAMS A, NANGIA N, BOEMAN S R, et al. A broad-coverage challenge corpus for sentence understanding through inference [EB/OL]. [2019-04-10]. <https://arxiv.org/abs/1704.05426v2>.
- [20] PENNINGTON J, SOCHER R, MANNING C D, et al. GloVe: global vectors for word representation [EB/OL]. [2019-04-10]. <https://www.aclweb.org/anthology/D14-1162>.

编辑 金胡考