

面向问答文本的属性级情感分类研究

江明奇¹ 李逸薇² 刘欢¹ 李寿山¹

(苏州大学计算机科学与技术学院 江苏 苏州 215006)¹

(香港理工大学中文及双语学系 香港 999077)²

摘要 传统情感分析任务的目的是分析整个文本的情感极性,这是一种粗粒度的任务。近年来,随着技术的革新,情感分析任务也在不断细化,研究者们希望能获取关于文本中具体对象的情感极性。文中的研究任务是获取问答文本中关于产品属性的情感极性。针对问答文本的属性级情感分析问题,提出了一种基于注意力机制的方法。首先,将属性信息拼接到答案词向量上;其次,对答案文本和问题文本学习一个 LSTM 模型;然后,通过注意力机制获得问题文本和答案文本的相关性,并根据相关性的重要程度获取答案文本的整体特征;最后,通过分类器输出最终的整体特征结果。实验结果表明,所提方法优于传统的属性级情感分析方法。

关键词 问答文本,情感分析,注意力机制

中图分类号 TP391 文献标识码 A

Attribute Sentiment Classification Towards Question-answering Text

JIANG Ming-qi¹ LEE Sophia Yat Mei² LIU Huan¹ LI Shou-shan¹

(School of Computer Science & Technology, Soochow University, Suzhou 215006, China)¹

(Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, Hong Kong 999077, China)²

Abstract The goal of conditional sentiment analysis is getting the sentiment polarity of whole text, which is a coarse task. Recently, with the improved technology, the sentiment analysis task is also refined, and the researchers hope to get sentiment polarity of given target of the text. This paper's purpose is getting the sentiment polarity of product attribute on question-answering text. To perform attribute sentiment classification towards QA text pair, this paper proposed a novel approach based on attention mechanism. Firstly, this paper concatenated the attribute information on answer words' vectors. Secondly, this paper leveraged LSTM models to encode the question text and answer text. Thirdly, this paper got the relation of question and answer by using attention mechanism and got the whole feature of answer. Finally, this paper got the result of whole feature by using classifier. Empirical studies demonstrate the effectiveness of the proposed approach to attribute sentiment classification towards question-answering text.

Keywords Question-answering text, Sentiment analysis, Attention mechanism

1 引言

随着互联网时代的到来,用户在互联网中发布以及收集信息,众多的用户成为了互联网中庞大信息的提供者。其中,淘宝、京东、亚马逊等电子商务平台上存在关于商品的各种信息。这些电子商务平台提供了客户问答平台,用户可以在平台上提问和回答有关商品的信息。由于该方式难以大规模造假,因此问答评论的方式比传统评论方式更为可靠。

情感分类是情感分析中最基础的任务,该任务的目标是判断给定文本的情感极性(正面、中性、负面),针对文本的情感极性进行分类^[1]。研究者们使用许多方法(如 SVM、最大熵、贝叶斯等分类器)来进行情感分类^[2-3]。近年来,研究者们不断细化情感分类任务,有关细粒度情感分类的文章不断增加。细粒度情感分类不再判定整个句子或者篇章的情感极性,而是针对评价对象及其属性进行情感极性的分类^[4]。针对指定的评价对象和属性的情感分类任务,研究者们提出了

将评价对象和属性信息加入到学习模型中的方法^[5]。本文面向问答文本开展有关商品的属性级情感分类任务研究,目的在于在给定属性的前提下,获得文本中有关属性的情感极性。据我们所知,此前还未有相关学者进行面向问答文本的属性级情感分类方法的研究。表 1 给出了一些问答样例及其相关属性和情感类别。从表中可以看出,问答文本中问题和答案都有关于属性的描述,情感信息一般存在于答案中。如表 1 中的问题“这款包包的具体尺寸是多少”涉及到了尺寸属性,答案中的“可以装 6p”则客观地回答到了尺寸属性的问题,因此我们把它标为中性。

问答文本的属性级情感分类任务通常是直接将问答文本和其他文本同样对待,使用已有的属性级情感分类方法进行。目前,属性级情感分类方法中性能较好的方法是基于深度学习模型的方法。这些模型使用词向量作为词的表示,并使用词向量的序列来代表句子或段落^[6]。然而,问答文本与普通文本之间有一些差异。经过分析,这些差异体现在

本文受国家自然科学基金(61672366)资助。

江明奇(1994—),男,硕士生,主要研究方向为自然语言处理、情感分析,E-mail:mqjiang@stu.suda.edu.cn;李寿山(1980—),男,教授,主要研究方向为自然语言处理、情感分析,E-mail:lishoushan@suda.edu.cn(通信作者)。

有关属性的描述在问题文本中,而属性的情感极性往往在答案文本中。因此,本文提出了一种基于注意力机制的属性级情感分类方法,通过使用长短时记忆神经网络,结合问题文本、属性信息和答案文本的关联性进行分类。具体而言,首先,把属性信息拼接到答案文本的词向量上;其次,对问题文本和拼接后的答案文本各学习一个 LSTM 模型;然后,通过注意力机制学习问题关于答案各词的注意力权重,并和答案的 LSTM 模型进行点乘以获得答案的整体特征;最后,使用分类器对整体特征进行分类并输出。实验结果表明,本文方法能有效地提高面向问答文本的属性级情感分类性能。

表 1 淘宝“问大家”中的样例及其属性信息和情感信息

属性	情感	问答文本
掉色	负面	问题:请问大家,持久度怎么样?容易掉色吗? 答案:我个人觉得也是会容易掉。
味道	正面	问题:什么价买的,味道真的很大吗? 答案:没有什么味道啊。
尺寸	中性	问题:这款包包的具体尺寸是多少? 答案:很好看,我很满意!可以装 6p。

本文第 2 节介绍与本文相关的一些工作;第 3 节介绍本文提出的基于注意力机制的属性级情感分类方法;第 4 节展示实验设置、实验结果及相关分析;最后总结全文,并对下一步工作进行展望。

2 相关工作

属性级情感分类是一个细粒度层次的情感分析,旨在获得给定的对象属性的情感极性。有许多学者对其展开了研究,以各样的方式去融合文本信息和属性信息,以获得属性对应的情感极性。计算语言学领域的语义评估会议 Semantic Evaluation(SemEval)^[1]在 2014 年组织开展了一项属性级情感分析评测任务。该任务针对评论文本进行属性级情感分类和属性分类等。早期,研究者们使用传统的机器学习方法,并设计各种特征来进行实验,如 Jiang 等^[7]考虑了对象的直接相关性,并与对象独立的特征结合来获得相应的情感信息。近年来,基于神经网络的方法在属性级情感分类上取得了优异的成绩。Liu 等^[8]使用不同的递归神经网络和词向量的组合,实现了基于属性的情感分析。Tang 等^[9]使用深度记忆网络方法来获取属性方面的情感,该方法的每一层记忆网络是由之前的记忆网络加上文本的注意力模型生成的。Tang 等^[10]考虑了属性信息的加入方法,使用了前向和后向 LSTM 模型指向属性信息,以获得属性的情感信息。Lambert 等^[11]使用了基于注意力机制的 LSTM 模型,并把属性信息的向量拼接到文本中每个词的词向量上。Ma 等^[12]针对属性级情感分析设计了一个交互式学习的模型。具体而言通过 LSTM 模型来获得文本和属性信息的隐层信息,互相学习并计算文本和属性信息的注意力权重。

3 基于注意力机制的属性级情感分类方法

3.1 语料收集

本文语料的原始来源为淘宝“问大家”^[2]平台上箱包领域中的问答语料,我们对其进行标注。为了保证较高的标注一

致性,在进行语料分析以及预标注后,我们定义了一系列关于问答文本的属性级情感的数据标注规范。

(1)对于问答文本,需要答案回答到问题文本相关对象的提问,才可进行关于对象的情感标注。

(2)对于问答文本中的情感评价对象,在问题文本中标记出对应的词或词组。

(3)对于问答文本中关于评价对象的情感信息,根据答案来判断其情感极性。

根据以上的标注规范,标注者分为两人一组对原始问答语料进行标注,并进行了一致性的校对。

语料的属性级情感极性分布如表 2 所列。从表中可以看出,箱包领域语料的情感极性分布较为不平衡,其中正面情感的样本占据了整体样本的大半,负面样本数少于正面,中性样本数最少,且远远少于正面样本数。

表 2 问答文本中属性的情感分布

情感	负面	中性	正面	合计
样本	424	98	3 003	3 525

3.2 LSTM 网络

LSTM^[13]神经网络避免了传统神经网络在反向传播过程中的梯度消失和梯度爆炸问题,并且可以使用记忆单元来学习长期依赖关系,充分利用历史信息。Graves^[14]于 2013 年对 LSTM 进行了改良和推广,使其被广泛应用于自然语言处理、语音识别等领域中。

LSTM 在 t 时刻的神经单元输出由以下公式计算:

$$i_t = \sigma(W_i X_t + U_i h_{t-1} + V_i C_{t-1}) \quad (1)$$

$$\tilde{c}_t = \tanh(W_c x_t + U_c h_{t-1}) \quad (2)$$

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + V_f C_{t-1}) \quad (3)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (4)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + V_o C_t) \quad (5)$$

$$h_t = o_t \odot \tanh(c_t) \quad (6)$$

其中, σ 为 logistic sigmoid 函数; i_t, f_t, c_t, o_t 分别为输入门、遗忘门、记忆单元和输出门在 t 时刻的表示; $\tilde{c}_t$ 为记忆单元的候选记忆的状态值; h_t 为 t 时刻 LSTM 单元的输出。LSTM 层的输入为样本的词向量表示,词向量在自然语言处理任务中有良好的语义特征,因此常被用来表示词语特征;并且,经过 LSTM 的向量会生成高维向量,学到更深层次的特征。

3.3 基于注意力机制的属性级情感分类方法

通过观察语料发现问题和答案文本具有相关性,并且指定对象属性对于提取答案文本属性相关情感信息有一定的辅助作用。为了有效利用问答文本的相关性和对象属性信息,本文提出了一种基于注意力机制的属性级情感分类方法,具体而言,该方法把对象属性对应的向量拼接到答案文本的词向量上,使用 LSTM 模型获得问题文本和答案文本的隐藏特征,利用问题文本的隐藏特征对答案文本进行相关度计算,并获得对应的答案文本的整体特征,最后通过分类器输出分类结果。

图 1 给出了关于问题文本处理的框架, Q_{w1} 到 Q_{wM} 为问题文本的词向量,经过 LSTM 网络后,取最后一个时间步的输出 h_{qM} 为该问题文本的整体隐层表示。

^[1] <http://alt.qcri.org/semeval2014/task4/>

^[2] <https://www.taobao.com/>

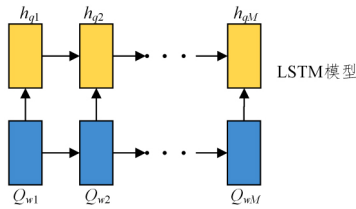


图 1 问题文本框架

图 2 给出了基于注意力机制的属性级情感分析模型的框架。首先,在答案文本的词 A_{w1} 到 A_{wN} 上拼接了属性对象的信息向量 i_a ,其中 i_a 由属性中的各词向量的平均值确定。

$$i_a = \frac{1}{L} \sum_{i=1}^L x_i \quad (7)$$

其中, L 为属性对象的词的数目, x_i 为词 i 的词向量。拼接后的答案文本经过一个 LSTM 层生成答案文本的隐层表示 h_{a1} 到 h_{aN} , 公式如下:

$$\vec{h}_{ai} = \text{LSTM}([A_{wi}; h_{qM}]) \quad (8)$$

$$\overleftarrow{h}_{ai} = \overleftarrow{\text{LSTM}}([A_{wi}; h_{qM}]) \quad (9)$$

$$h_{ai} = \vec{h}_{ai} \parallel \overleftarrow{h}_{ai} \quad (10)$$

之后,在答案文本的隐层表示 h_{a1} 到 h_{aN} 上拼接问题文本的隐层表示 h_{qM} ,并用全连接层和 softmax 层来获得答案文本的注意力权重 a_1 到 a_N , 公式如下:

$$u_i = \tanh(W_u[h_{ai}; h_{qM}] + b_u) \quad (11)$$

$$a_i = \text{softmax}(u_i) = \frac{\exp(u_i)}{\sum_{j=1}^N \exp(u_j)} \quad (12)$$

其中, $[h_{ai}; h_{qM}]$ 表示 h_{ai} 和 h_{qM} 向量进行拼接, $1 \leq i \leq N$, W_u 为对应的权重矩阵, b_u 为偏置项。 $a = [a_1, a_2, \dots, a_N]$ 为答案文本中各词的注意力权重。

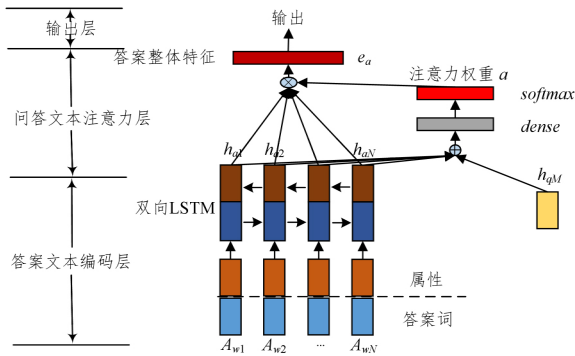


图 2 基于注意力机制的属性级情感分类模型

之后我们通过使用注意力权重对答案文本的隐层向量 h_{a1} 到 h_{aN} 进行加权求和,来计算答案文本的整体特征 t_a :

$$t_a = \sum_{i=1}^N a_i h_{ai} \quad (13)$$

最后使用 softmax 输出层对答案的整体特征进行提取:

$$y = \text{softmax}(W_o \cdot t_a + b_o) \quad (14)$$

其中, W_o 为 softmax 层中的权重矩阵, b_o 为偏置项, y 为各情感极性的概率,它的长度等于情感极性的类别数。

在模型训练的过程中,我们选取交叉熵误差作为损失函数,并最小化损失。损失函数如下:

$$\text{Loss}(\hat{y}, y) = - \sum_{i=1}^S \sum_{c=1}^C y_i^c \cdot \log \hat{y}_i^c \quad (15)$$

其中, S 为样本总数, C 为类别数。

4 实验

4.1 实验设置

本文选用淘宝“问大家”中的问答语料作为实验语料。该语料一共分为 3 个情感极性:负面、中性、正面。对于所有基准实验以及本文提出的方法实验,各个类别的数据均按比例切分,并保证各实验训练集与测试集的一致性。

首先使用 Jieba¹⁾ 工具进行文本中文分词。使用深度学习框架 Keras²⁾ 搭建本文方法。在各个模型中,词语的特征用词向量来表示。具体地,利用 python 的 gensim³⁾ 库学习淘宝的所有问答文本中的语言模型,并进行词向量的生成,之后使用该词向量进行各个模型的实验。考虑到性能等因素,本文的词向量维度设置为 200。实验中各参数的设置如表 3 所列。

表 3 神经网络中的参数设置

参数表述	参数值
输入向量长度	40
LSTM 层输出维度	64
问题输入向量长度	20
答案输入向量长度	30
迭代次数	30
Batch_size	32

采用正确率和 Macro-F1 值作为衡量各个模型属性级情感分类效果的标准。单类别 F_1 值的计算方法如下:

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} \quad (16)$$

其中, precision 代表准确率, recall 代表召回率, F_1 值为准确率和召回率的调和平均数。

Macro-F1 值的定义如下:

$$\text{Macro-F1} = \frac{1}{C} \sum_{c=1}^C F_1^c \quad (17)$$

其中, F_1^c 为第 c 个类别的 F_1 值^[15]。

4.2 实验结果

我们比较了以下几种属性级情感分类方法,除本文方法,其他方法均处理问题文本与答案文本直接拼接后的问答文本。

1) LSTM: 对问答文本和属性信息进行拼接,拼接后的文本使用 LSTM 分类器进行分类。

2) TC-LSTM: 在使用 LSTM 模型时考虑到属性信息,其中包含了到达属性信息的 LSTM 前向模型和后向模型。本文实现了该模型的方法,并使用与其他文本相同的处理方式处理整个问答文本。

3) ATAE-LSTM: 使用了基于注意力机制的 LSTM 模型,并把属性信息的向量拼接到文本中每个词的词向量上。该模型使用与其他文本相同的处理方式处理整个问答文本。

4) IAN: 该模型开始通过 LSTM 模型来获得文本和属性信息的隐层信息,互相学习并计算文本和属性信息的注意力权重,最终获得文本的属性情感信息。该模型使用与其他文本相同的处理方式处理整个问答文本。

¹⁾ <https://pypi.python.org/pypi/jieba/>

²⁾ <http://keras.io/>

³⁾ <https://radimrehurek.com/gensim/>

5)QA-AE:使用本文提出的模型,针对问答文本所构建的模型。

表 4 列出了在箱包领域的问答语料上,各个属性级情感分类方法的正确率和 *Macro-F1* 值的比较。从表中可以看出,不同方法在不同评判标准上的结果有些许不同。LSTM 模型直接在问答文本上拼接属性信息的方法并不能很好地把属性信息传递给模型,因此该模型的效果不尽理想。TC-LSTM 方法和 ATAE 均使用属性信息来提升获取文本中对应情感信息的效果,但都不能很好地提取文本中对应的属性情感。IAN 方法在两个评价标准上都是基准方法中表现最好的,表明了 IAN 使用注意力机制使属性信息和文本信息相互影响学习的方法相较于其他基准方法的确可以更好地抓住文本中的属性情感分析信息。QA-AE 方法在各个评价标准上的数值都达到了最高,证明了本文方法有效地使用了属性信息和问答相关性,获得了更为准确的属性级情感信息。

表 4 各个属性级情感分类方法的正确率和 *F1* 值
(单位:%)

模型	Accuracy	Macro-F1
LSTM	84.99	57.1
TC-LSTM	86.96	59.14
ATAE	86.54	59.96
IAN	87.96	60.5
QA-AE	89.09	63.6

在每个类别的 F_1 值上,表 5 展示了在箱包领域的问答语料上各个属性级情感分类方法的比较。

表 5 各情感类别上各个方法的 F_1 值
(单位:%)

模型	负面	中性	正面
LSTM	52.79	26.67	91.81
TC-LSTM	54.75	29.63	93.04
ATAE	53.09	34.29	92.51
IAN	56.38	31.58	93.55
QA-AE	62.42	34.29	94.10

由于箱包领域存在较为严重的不平衡现象,因此各个方法在各个类别的判断上各有优劣。中性类别在实验中效果最差,这可能是由于中性类别的问答文本远少于负面和正面样本,各个学习方法并不能很好地通过少量样本学习到样本中的中性类别信息。IAN 方法的 F_1 值在负面和正面类别中均为基准方法中最好的,这从侧面证明了该方法能较好地识别这两个类别的属性情感信息,但不能很好地识别中性类别。ATAE 方法使用注意力机制获取文本中与属性信息关系较为紧密的部分,可以更准确地识别中性样本,与 QA-AE 一样获得了所有方法中最高的中性类别 F_1 值。本文提出的方法有效地利用了问答文本之间的相关性和属性信息,在 3 个类别上均获得了最高的 F_1 值,尤其适合识别负面样本,在负面样本上的效果远高于其他方法。

结束语 面向问答文本这种新颖的文本方式,在研究属性级情感分析任务时,本文提出了一种基于注意力机制的属性级情感分析方法来联系问答之间的关系。该方法首先将属性信息拼接到答案词向量上;之后使用 LSTM 模型获取问题文本与答案文本的隐层信息,通过注意力机制获取这些隐层信息之间的相关性,并通过注意力权重获得答案文本的整体表示;最后使用 softmax 层进行属性级情感分类。该方法的分类效果在各领域上均优于传统的分类方法。

在未来的工作中,我们将继续验证本方法在其他领域的有效性。同时,我们将继续探索问答文本的特性,学习如何获得更好的问答对关联信息,从而有效地利用问答文本的特性获得更好的分类性能。

参 考 文 献

- [1] GO A, BHAYANI R, HUANG L. Twitter sentiment classification using distant supervision; CS224N Project Report[R]. Stanford, 2009.
- [2] MULLEN T, COLLIER N. Sentiment analysis using support vector machines with diverse information sources[C]// Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, 2004.
- [3] KHAIRNAR J, KINIKAR M. Machine learning algorithms for opinion mining and sentiment classification[J]. International Journal of Scientific and Research Publications, 2013, 3(6): 1-6.
- [4] 唐晓波, 刘广超. 细粒度情感分析研究综述[J]. 图书情报工作, 2017, 61(5): 132-140.
- [5] BOIY E, MOENS M F. A machine learning approach to sentiment analysis in multilingual Web texts[J]. Information retrieval, 2009, 12(5): 526-558.
- [6] JO Y, OH A H. Aspect and sentiment unification model for online review analysis[C]// ACM International Conference on Web Search and Data Mining. ACM, 2011: 815-824.
- [7] JIANG L, YU M, ZHOU M, et al. Target-dependent twitter sentiment classification[C]// Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1. Association for Computational Linguistics, 2011: 151-160.
- [8] LIU P, JOTY S, MENG H. Fine-grained opinion mining with recurrent neural networks and word embeddings[C]// Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 2015: 1433-1443.
- [9] TANG D, QIN B, LIU T. Aspect level sentiment classification with deep memory network[C]// Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016: 214-224.
- [10] TANG D, QIN B, FENG X, et al. Effective LSTMs for target-dependent sentiment classification[C]// Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics, 2016.
- [11] LAMBERT P. Aspect-level cross-lingual sentiment classification with constrained SMT[C]// Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 2015: 781-787.
- [12] MA D, LI S, ZHANG X, et al. Interactive attention networks for aspect-level sentiment classification[C]// Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence Main track, 2017: 4068-4074.
- [13] HOCHREITER S, SCHMIDHUBER J. Flat minima[J]. Neural Computation, 1997, 9(1): 1-42.
- [14] GRAVES A. Generating Sequences With Recurrent Neural Networks [J]. arXiv:1308.0850, 2013.
- [15] SOKOLOVA M, LAPALME G. A systematic analysis of performance measures for classification tasks[J]. Information Processing & Management, 2009, 45(4): 427-437.