面向问答文本的属性分类方法*

江明奇，沈忱林，李寿山

（苏州大学 计算机科学与技术学院，江苏省 苏州市 215006）

摘要：属性分类是属性级情感分析中的一个重要任务。该任务旨在对文本包含的某些具体属性进行自动分类。已有的属性分类方法研究基本都是面向新闻、评论等文本类型。与已有研究不同的是，本文的研究主要面向问答文本的属性分类任务。针对问答文本的属性分类问题，本文提出了一种多维文本表示的方法。首先，该方法进行中文句子切分；其次，对每个子问题和答案学习一个 LSTM 模型；再其次，通过融合多个 LSTM 模型，形成多维文本表示；最后，使用卷积层处理多维文本表示，获得最终分类结果。实验结果表明该方法明显优于传统的属性分类方法。

关键词：属性分类；问答文本；多维文本表示

中图分类号：TP391　　**文献标识码：**A

Attribute Classification towards Question-Answer TextName[1,2]

Mingqi Jiang, Chenlin Shen, Shoushan Li

(School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China)

Abstract: Attribute classification is an important research task in aspect-based sentiment classification. It aims at classifying the category of attribute automatically. Most existing studies for attribute classification mainly focus on text styles like news and reviews. Unlike these studies, this paper focuses on a totally different kind of text, i.e., question-answer (QA) text pair. To perform attribute classification towards QA text pair, we propose a novel approach called multi-dimension textual representation. Firstly, we segment the question text of a QA text pair into sentences. Then, we leverage LSTM models to encode each sentence in question text and the whole answer text. Finally, we leverage a CNN layer to extract important information in all sentences of question text and the whole answer text. Empirical studies demonstrate the effectiveness of our proposed approach to attribute classification towards question-answer text.

Key words: Attribute classification; Question-Answer text; multi-dimension textual representation

1. 引言

随着互联网行业的兴起，亚马逊、京东和淘宝等众多电子商务平台提供了客户问答平台，用户在这些平台上可以对感兴趣的商品的属性进行提问，购买过相应商品的用户可以给出相应的回答。由于该方式的评价信息难以大规模造假，从而使得问答评论的方式比传统评论方式更为可靠。近些年来，针对属性级情感分析任务的研究越来越多，该任务旨在做出更细粒度方面的意见挖掘[1]。属性分类任务则是属性级情感分析任务的基础与前提，旨在获得文中所提到的相关属性[2]。本文面向问答文本开展属性分类任务研究。据我们所知，此前还未有相关论文进行面向问答文本的属性分类方法研究。表 1 给出了一些问答样例及其相关的属性类别。从表 1 可以看出，问答文本中的属性分类需要答案回答到问题提到的相关属性才可计入，如果答案没有回答到问题的相关属性，则该问答文本不带有属性信息。

针对问答文本的属性分类，一种直接的方法是把问答文本和其他文本同样对待，利用已有的机器学习方法来进行属性分类。目前，属性分类研究中性能表现较好的方法是基于深度

* **收稿日期：**　　**定稿日期：**

基金项目：国家自然科学基金（No.61672366）

作者简介：江明奇（1994——），男，硕士研究生，主要研究方向为情感分析；沈忱林（1993——），男，学生，主要研究方向为情感分析；李寿山（1980——），男，教授，主要研究方向为情感分析。

表 1 淘宝“问大家”中样例及其所属的属性类别

Tab.1 Some examples in *Taobao "asking people"* with their attribute categories

属性类别	问答文本
重量	问题：镜头重吗？ 答案：800 多克，比较重，女生慎入
系统性能	问题：怎么样？打王者荣耀卡不卡 答案：不卡反应快
电池	问题：电池正常使用可以用多久？快充多久可以充满？ 答案：反正觉得挺耐用的

模型。学习的分类模型。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者段落[3]。然而，本文研究的对象是问答文本，问答文本同普通文本存在较大差异。例如，通过分析语料，我们发现属性的描述一般出现在问题文本上，而答案文本较少的体现属性类别。因此，本文提出了一种基于多维文本表示的属性分类方法，具体使用长短时记忆（Long Short-Term Memory, LSTM）神经网络[4]，并结合答案和问题文本进行属性分类。具体而言，首先，我们把问题切分为多个子句，并和答案文本一起作为多文本输入；其次，针对每一个文本学习一个 LSTM 模型表示；再其次，通过 Merge 层来融合多个 LSTM 模型生成多维文本表示；最后，使用卷积层来对这个表示进行特征提取，并获得最终分类结果。实验结果表明该方法能有效地提高面向问答文本的属性分类性能。

本文结构安排如下：第二节介绍了与本文相关的一些工作进展；第三节介绍本文提出的基于多维文本的属性分类方法；第四节给出实验结果及相关分析；第五节对本文做出总结，并对下一步工作进行展望。

2. 相关工作

2.1 属性抽取、聚类

近些年来，许多研究把属性抽取任务建模为聚类、抽取问题，并在产品服务评论领域和新闻领域取得了一定的效果。

杨等[5]人用属性归类方式来进行属性抽取。具体而言，杨等人使用了改进的 EM 算法，并利用了两条自然语言知识进行属性抽取，（1）含有相同字的两个属性有可能是同类属性；（2）同义词的两个属性有可能是同类属性。李等[6]使用聚类方法进行属性抽取。具体而言，该方法首先对各分类属性按照取值进行划分，得到最初的各聚类成员；然后，通过不断调整寻找最优的融合结果。Xiong 等[7]提出一种基于注意力机制的神经网络模型，考虑了属性短语与上下文之间的语义关系，来进行属性短语的聚类。

2.2 属性分类

目前，文本情感分析方法的研究主要是针对极性情感类别（如正面、中性、负面），而针对细粒度的情绪分类方法的研究比较缺乏。

计算语言学领域著名的语义评估会议 Semantic Evaluation(SemEval)[1]在 2014 年组织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、属性分类等[8]。该评测任务提供了餐厅和笔记本电脑这两个领域的语料，共包含 531 个评论文本、3054 个句子的数据集。Toh 等[9]把属性分类任务作为多标签分类任务并对每个属性做一个二分类任务。针对每一个二元分类任务，作者使用了含有一层隐藏层的神经网络模型，并结合了多元词特征。Khalil 等[10]分别使用了预训练了词模型的 CNN 分类模型和词袋模型的 SVM 分类模型，并把这两个分类模型融合，获得了较好的分类效果。Kalchbrenner 等[11]提出了一种 CNN 改

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

进模型，具体在模型中引入广范围 CNN 层和动态最大池化层。实验结果表明，使用该方法获得了较好的分类结果。

已有的属性分类研究基本都是面向评论文本、新闻文本等较为普遍的文本的，面向问答文本的属性分类方法研究还很缺乏。据我们所知，本文是首次面向问答文本进行属性分类方法研究。

3. 基于多维文本的属性分类方法

3.1 语料收集

本文在淘宝的“问大家”中收集了 2296 条关于数码领域的问答语料。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，我们定义了一系列的关于数码领域的属性并设计了相应的标注规范。规范如下：

1) 经过语料的预标注和统计，数码领域相关属性类别：质量、系统、存储、系统性能、电池等。
2) 问题中提到上述所提出的类别相关的问题，并且答案明确回答了类别相关的问题时，计入。

根据以上标注规范标注者分为两人一组对这些数码领域的问答语料进行了标注，并进行了一致性校对。

表 2 问答文本中各属性分布情况

Tab.2 Attribute distribute in QA text

属性类别	IO	质量	电池	系统性能	正品	计算	功能	合计
样本数量	820	264	227	525	357	93	110	2296

语料的属性分类分布情况如表 2 所示。从表中可以看出，问答文本的各属性类别样本分布不平衡，大部分的问答语料都和 IO 和系统性能有关，而计算、功能类别的样本数量较少。

3.2 问答文本的多维文本表示

与新闻文本、评论文本等文本类型不同，问答文本是由问题文本和答案文本两部分组成，

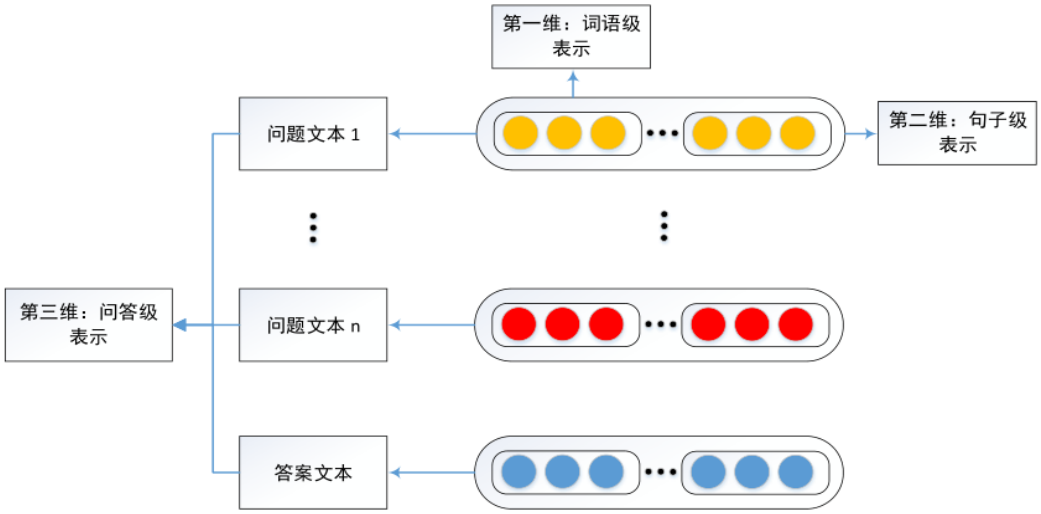

图 1 问答对的三维表示

Fig.1 3-Dimension representation of question-answer context

因此有一定的特殊性。在做属性分类任务时，属性的判断主要依据问题，然而答案也可以进行一定的辅助。问题文本一般有多个子句，其中一到多句描述了属性相关的内容，没有描述属性相关内容的子句就会有一定的噪音干扰。因此切分问题来获得更准确的信息是一个很自然的想法。文本表示是该学习方法中的关键部分。我们使用多维文本表示来进行面向问答文本的属性分类方法的实验。

如图 1 所示，图中每组带颜色的圆形为词语级维度的表示，包含一系列词向量的表示称为句子级维度的表示，所有子问题句子与答案句子一起形成了问答对的表示，称为问答级维度的表示。

问答级文本包含多个问题文本和一个答案文本，即用工具把问题文本拆分成若干个子问题，并根据语料固定问题的子问题个数为 N 。

具体而言，使 $T=(x_1,x_2,\mathrm{L})$ 为词的序列，其中每一个 x_i 表示一个词向量。之后，使 T_{Q_1} 到 T_{Q_N} ， T_A 来分别表示第一个子问题到第 N 个子问题和答案文本。因此，整个 QA 对可以被表示为 $(T_{Q_1}\ \mathrm{L}\ T_{Q_N},T_A)$ ，这是一个三维的矩阵。

3.3 基础的 LSTM 和 CNN 网络

LSTM 神经网络避免了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。Alex Graves[12]于 2013 年对 LSTM 进行了改良和推广，使得 LSTM 被广泛应用于自然语言处理、语音识别等领域中。

LSTM 的 t 时刻的神经单元由以下公式计算：

$$i_t=\sigma(W_i x_t+U_i h_{t-1}+V_i C_{t-1}) \tag{3.1}$$

$$\%_t=tanh(W_c x_t+U_c h_{t-1}) \tag{3.2}$$

$$f_t=\sigma(W_f x_t+U_f h_{t-1}+V_f C_{t-1}) \tag{3.3}$$

$$c_t=f_t\ \mathrm{e}\ c_{t-1}+i_t\ \mathrm{e}\ \%_t \tag{3.4}$$

$$o_t=\sigma(W_o x_t+U_o h_{t-1}+V_o c_t) \tag{3.5}$$

$$h_t=o_t\ \mathrm{e}\ tanh(c_t) \tag{3.6}$$

其中 σ 为 logistic sigmoid 函数， i_t ， f_t ， o_t ， c_t 分别为输入门、输出门、遗忘门和记忆单元在 t 时刻的值。 $\%_t$ 为记忆单元的候选记忆状态值， 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，词向量有良好的语义特征，因此常用词向量来表示词语特征[13]。并且，经过 LSTM 的向量会生成高维向量，学到更深层次的特征。

卷积神经网络（CNN）是前向反馈的神经网络[14]，由一个或多个卷积层和池化层组成。CNN 的特点是针对输入的局部进行感知和权值的共享。CNN 广泛应用于图像识别任务。本研究使用 CNN 网络来对问答级文本表示进行建模。值得注意的是，卷积层中包含了一些过滤器，这些过滤器融合了 h 个词的信息并产生了新的特征。对于 h 个词 $X_{i:i+h-1}$ ，过滤器 F_l 生成特征 y_i^l 的方式如下：

$$y_i^l=\sigma(W\cdot X_{i:i+h-1}+b) \tag{3.7}$$

其中 σ 是非线性激活函数， b 是偏置。

3.4 基于 Multi-LSTM+CNN 模型的属性分类方法

为了有效的利用子问题与答案文本，本文提出了基于 Multi-LSTM+CNN 模型的属性分类方法，即使用 LSTM 生成每个词序的隐藏特征，并融合它们生成文本的多维表示，并用 CNN 来提取这个多维表示中的信息。

图 2 给出了 Multi-LSTM+CNN 模型的框架，每个子问题文本和答案文本分别经过一个 LSTM 层生成该文本的隐层表示 $h_{Q_1}-h_{Q_N}$ ，以及 h_A 。

Merge 层使用拼接的方法把这些隐层表示结合起来。在拼接的位置上，本文采用了时间步的位置。卷积层接收 Merge 层的输出来提取特征，并用最大池化层进行进一步的子采样。

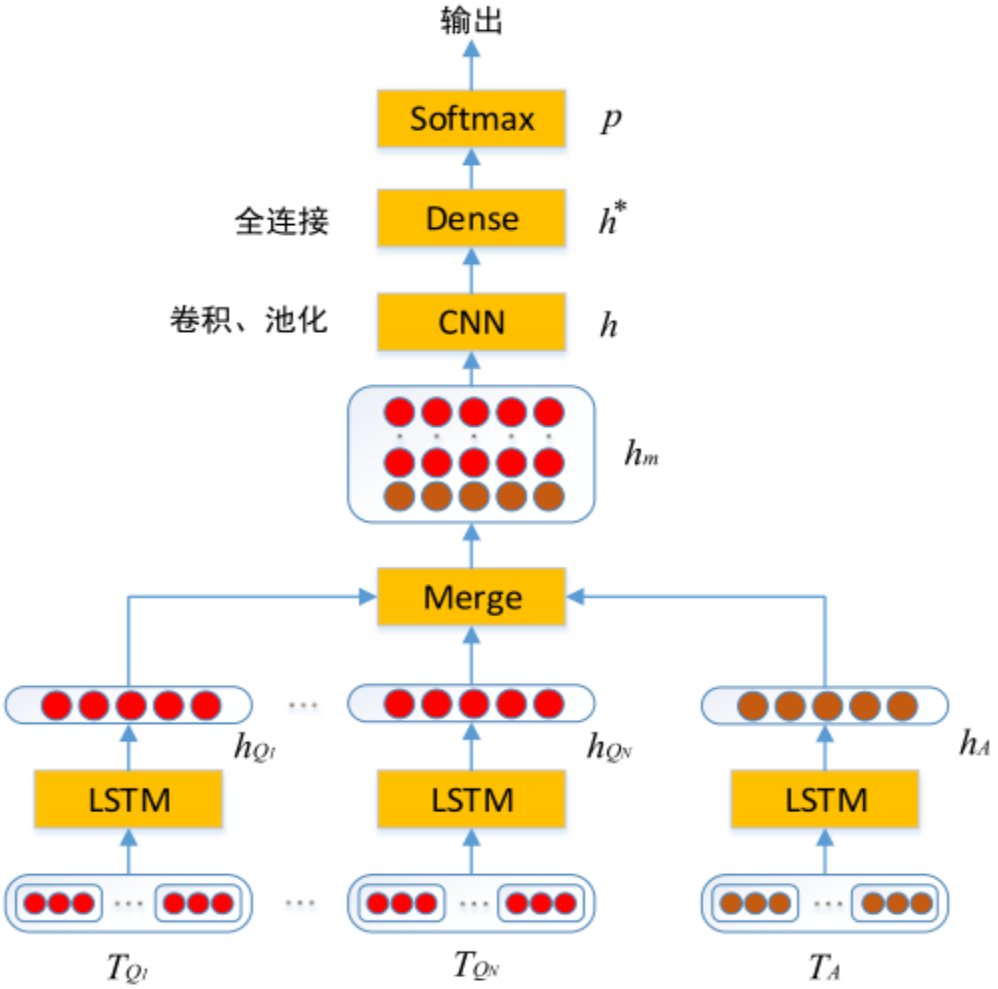

图 2 Multi-LSTM+CNN 模型

Fig.2 The overview of our Multi-LSTM+CNN model

全连接层用来把池化层的输出进行加权求和，并用激励函数包装，激励函数如下：

$$h^* = \varphi\left(\theta^T h + b\right) \tag{3.8}$$

φ 是非线性的激励函数，本文选择“Relu”作为全连接层的激励函数，h 表示全连接层的输出。

Dropout 层在训练和预测时会随机让网络中的某些隐藏层节点不工作，可以获得更少相互依赖特征，有效防止过拟合，用该层来接收全连接层的输出：

$$h^d = h^* * D(p) \tag{3.9}$$

其中，D 为 Dropout 操作，p 表示可调超参（在网络中保留隐藏单元的概率），h^d 为 Dropout 层输出。

最后，使用 softmax 输出层来对样本进行分类，获得的预测概率值如下：

$$P = softmax\left(W^d h^d + b^d\right) \tag{3.10}$$

其中 P 为预测标签的概率集，选出最大值即为预测标签，W^d 为需要学习的权重，b^d 为偏置。

在模型训练的过程中，我们选择最小化交叉熵误差作为损失函数：

$$J = -\sum_{i=1}^{N}\sum_{l=1}^{m}\left[1\{t_i = l\} log\left(y_l\right)\right] + \frac{\lambda}{2N}\sum_{k=1}^{n}\left(\sum_{\varepsilon\in\omega}\left\|W_\varepsilon^k\right\|_F^2 + \sum_{\varepsilon\in\mu}\left\|U_\varepsilon^k\right\|_F^2 + \sum_{\varepsilon\in\nu}\left\|V_\varepsilon^k\right\|_F^2\right) \tag{3.11}$$

其中，N 是训练样本的个数，m 是目标类别的数量，y 是 softmax 层输出的每个类别的预测

概率，t_i 是第 i 个训练样本的真实标签。$\|\cdot\|_F$ 表示 Frobeniu 范数，n 是通道的个数，$\omega=\{i,f,o,c\}$，$\mu=\{i,f,o,c\}$和$v=\{i,f,o\}$表示不同门的集合（分别为W，U，V），λ 是用来指定惩罚权重的超参。

4. 实验设计与分析

4.1 实验设置

本文选用淘宝[2]“问大家”中关于数码领域的问答语料作为实验语料。该语料共有七个类别，具体类别级样本个数可以参考 3.1 节的表 2。在本节中，对于所有实验，我们将每个类别的 70%的数据作为训练集，10%的数据作为验证集，20%的数据作为测试集来进行实验，测试样本具体分布如表 3。

表 3 各属性类别测试样本数和剩余样本数

Tab.3 The number of test samples and remaining samples

情绪类别	IO	质量	电池	系统性能	正品	计算	功能	合计
样本总量	820	264	227	525	357	93	110	2296
测试样本	163	32	45	104	71	18	21	454

在实验中，首先，使用 CoreNLP[3]工具对问题文本和答案文本进行句子切分；其次，使用 Jieba[4]工具来进行文本中文分词。使用深度学习框架 Keras[5]进行 LSTM 网络和 CNN 网络的搭建。在使用基于 LSTM 网络与 CNN 网络的分类器时，词语特征用词向量来表示。具体而言，利用 python 的 gensim[6]库对淘宝的问答文本进行词向量模型的生成，并使用该词向量进行所有神经网络模型的实验。考虑到实验性能等因素，词向量维度设为 200。实验模型各参数如表 4 所示。

表 4 神经网络中的参数设置

Tab.4 Parameters setting in Neutral network

参数表述	**参数值**
输入向量长度	60
LSTM 层输出维度	64
问题数目设置	3
问题输入向量长度	15
答案输入向量长度	30
卷积层过滤器数目	50
卷积层过滤器长度	3
Dropout rate	0.2
迭代次数	60

在实验中，采用正确率（Accuracy）和 Macro-F1 值作为衡量分类效果的标准。单类别 F_1 值具体的计算方法为式（4.1）：

$$F_1=2\cdot\frac{precision\cdot recall}{precision+recall} \tag{4.1}$$

其中 $Precision$ 表示精确率，$Recall$ 表示召回率。Macro-F1 值为各类别 F_1 值的平均值。

[2] https://www.taobao.com/
[3] https://stanfordnlp.github.io/CoreNLP/
[4] https://pypi.python.org/pypi/jieba/
[5] http://keras.io/
[6] https://radimrehurek.com/gensim/

4.2 实验结果

在实验中，我们比较了以下几种属性分类的方法。这些方法如下

1) **LSTM（Q）：**对问题文本使用 LSTM 分类器进行分类。
2) **LSTM（A）：**对答案文本使用 LSTM 分类器进行分类。
3) **LSTM（Q+A）：**对问题答案拼接的文本使用 LSTM 分类器进行分类。
4) **CNN（Q+A）：**对问题文本和答案文本拼接，并使用 CNN 神经网络方式进行属性分类。
5) **Multi-LSTM+CNN（Sub-A）：**使用本文提出的模型，但针对答案文本进行切分，问题文本不切分进行实验。
6) **Multi-LSTM+CNN（Sub-Q）：**使用本文提出的模型，针对问题文本进行切分，答案文本不切分进行实验。
7) **Multi-LSTM+CNN（Sub-Q and Sub-A）：**使用本文提出的模型，但针对问题文本与答案文本均进行切分。

图 3 展示了各属性分类方法在数码领域问答语料上正确率的比较。可以看出，同样使用了 LSTM 模型的答案文本在分类正确率上明显低于问题文本，这证明了问题文本的词语特征在属性分类上的重要性，而答案文本的词语特征并不能很好的帮助属性分类。在正确率上，LSTM（Q+A）则对于 LSTM（Q）方法没有明显的提升，可以推测直接拼接方式并不能有效利用答案文本的信息。本文同时实现 Multi-LSTM+CNN（Sub-A）方法，结果表明切分答案并没有显著效果，而 Multi-LSTM+CNN（Sub-Q）方法较 LSTM（Q+A）效果提升显著，证明了面向问答文本时切分问题的理论可以使属性分类效果更好。而 Multi-LSTM+CNN（Sub-Q and sub-A）方法效果仅得到了 92.7%的正确率，并不如只切分问题，这是因为答案文本中属性相关信息较少，在切分答案文本并融合所有文本时会加入许多噪音。

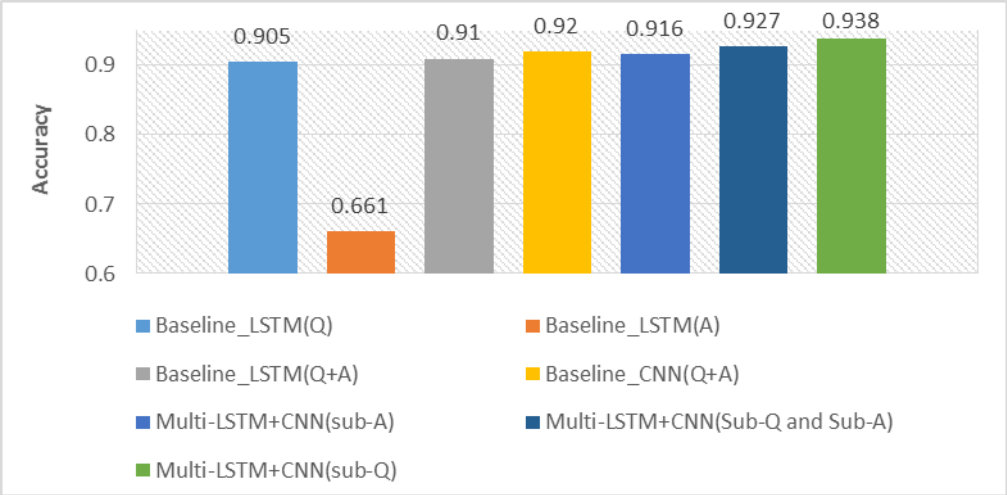

图 3 属性分类方法分类正确率比较

Fig.3 Accuracies of attribute classification methods

图 4 为各属性分类方法在测试集上的 Macro-F1 值结果。从图中可以发现，基准方法中效果最好的是 CNN 神经网络（Q+A）方法，该方法达到了 89.7%的 Macro-F1 值。与在正确率上一样，LSTM 神经网络（Q+A）比 LSTM 神经网络（Q）在 Macro-F1 值上也没有较好的提升。该图结果同时表明了本文提出的基于多维文本表示的方法在 F_1 值上相较于其他方法也有明显的优势，Macro-F1 值比其他方法中最高的 Multi-LSTM+CNN（Sub-Q and sub-A）高 1.5 个百分点。这些结果说明了我们的基于多维文本表示的方法抓住了问题文本和答案文

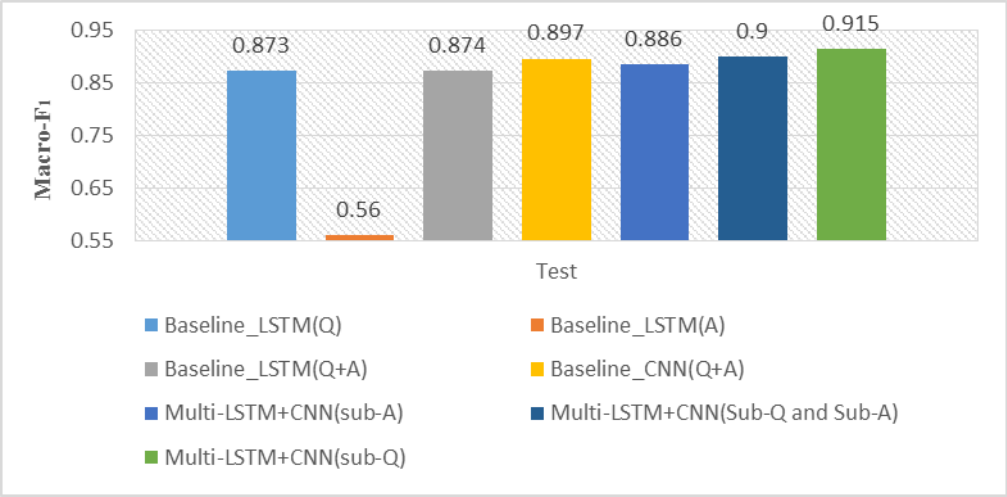

图 4 属性分类方法分类 Macro-F1 值比较

Fig.4 Macro-F1 of attribute classification methods

本之间的关系，并考虑了属性任务在问题文本中一般在某些子句中这个特性，该方法可以有效利用答案文本来对问题文本进行辅助的属性分类。

5. 结语

问答文本作为一种新颖的文本方式，本文对这种类型文本的处理进行了探索，提出了进行多维文本来表示整个问答对的方法，以得到更好的属性分类效果。该方法首先把问题切分为子句级的问题，之后每个子问题和答案均训练一个 LSTM 模型形成，通过 Merge 层融合这些模型，最后使用 CNN 网络进行特征提取并获得最终结果。实验证明，该方法能有效利用问答文本的逻辑关系，其分类效果优于传统的分类方法。

在下一步的工作中，我们将收集其他领域的问答语料并验证我们方法的有效性。此外，我们将继续探索问答文本的特性，来获得更好的问答对表示，以更有效的利用文本信息来获得更好的分类性能。

参考文献

[1] 施寒潇. 细粒度情感分析研究[D]. 苏州大学, 2013.

[2] Wu Y, Oard D W. Bilingual topic aspect classification with a few training examples[C]// International Acm Sigir Conference on Research & Development in Information Retrieval. DBLP, 2008:203-210.

[3] 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.

[4] Hochreiter S, Schmidhuber J. Flat minima.[J]. Neural Computation, 1997, 9(1):1 - 42.

[5] 杨源，马云龙，林鸿飞. 评论挖掘中产品属性归类问题研究[J]. 中文信息学报, 2012, 26(3):104-109.

[6] 李桃迎，陈燕，张金松,等. 一种面向分类属性数据的聚类融合算法研究[J]. 计算机应用研究，2011, 28(5):1671-1673.

[7] Xiong S, Zhang Y, Ji D, et al. Distance Metric Learning for Aspect Phrase Grouping[J]. 2016.

[8] Kiritchenko S, Zhu X, Cherry C, et al. NRC-Canada-2014: Detecting Aspects and Sentiment in Customer Reviews[C]// International Workshop on Semantic Evaluation. 2014:437-442.

[9] Toh Z, Su J. NLANGP: Supervised Machine Learning System for Aspect Category Classification and Opinion Target Extraction[C]// International Workshop on Semantic Evaluation. 2015:496-501.

[10] Khalil T, El-Beltagy S R. NileTMRG at SemEval-2016 Task 5: Deep Convolutional Neural Networks for

Aspect Category and Sentiment Extraction[C]// International Workshop on Semantic Evaluation. 2016:271-276.

[11] Kalchbrenner N, Grefenstette E, Blunsom P. A Convolutional Neural Network for Modelling Sentences[J]. Eprint Arxiv, 2014, 1.

[12] Graves A. Supervised Sequence Labelling with Recurrent Neural Networks[J]. Studies in Computational Intelligence, 2012, 385.

[13] Turian J, Ratinov L, Bengio Y. Word representations: a simple and general method for semi-supervised learning[C]// Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, 2010:780-1.

[14] Bouvrie J. Notes on Convolutional Neural Networks[J]. Neural Nets, 2006.

作者联系方式：江明奇　江苏省苏州市苏州大学本部理工楼412实验室 215006 15839867554 mqjiang_learning@163.com