

文章编号: 1003-0077 (2011) 00-0000-00

基于问题与答案共同表示学习的半监督问题分类方法*

张栋, 李寿山, 王晶晶

(苏州大学 计算机科学与技术学院, 江苏 苏州 215006)

摘要: 问题分类旨在对问题的类型进行自动分类, 该任务是问答系统研究的一项基本任务。本文提出了一种基于问题和答案共同表示学习的问题分类方法。该方法的特色在于, 利用问题及其答案作为共同的上下文环境, 学习词的分布式表示, 从而充分利用未标注样本中问题和答案隐含的分类信息。具体而言, 首先, 我们引入神经网络语言模型, 利用问题与答案共同学习词向量表示, 增加问题词向量的信息量; 其次, 加入大量未标注的问题与答案样本参与词向量学习, 进一步增强问题词向量表示能力; 最后, 将已标注的问题样本以词向量形式表示作为训练样本, 采用卷积神经网络建立问题分类模型。实验结果表明, 本文提出的基于半监督问题分类方法能够充分利用词向量表示和大量未标注样本来提升性能, 明显优于其他基准半监督分类方法。

关键词: 问题分类; 共同表示; 半监督

中图分类号: TP391

文献标识码: A

Semi-supervised Question Classification with Jointly Learning Question and Answer Representations

Dong Zhang, Shoushan Li, Jingjing Wang

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

Abstract: Question classification aims at classifying the types of questions automatically, and this task is a basic task of question answering system. This paper proposes a method of semi-supervised question classification with jointly learning question and answer representations. The characteristic is that we consider the question and its corresponding answer as conjunct context to learn the word distributed representation. Specifically, neural network language model is introduced to learn question and answer representations jointly, so that the word vectors of question are added more information. Secondly, large numbers of unlabeled questions and answers participate in word vectors learning, which could strengthen the representation capacity of question word vectors. Finally, we represent the questions of word vectors as training samples, adopting the convolutional neural network to construct the question classifier. The experimental results demonstrate that the method of semi-supervised question classification with synergetic representations learning in this paper can make full use of word vectors and the unlabeled samples to improve the performance, and is better than other strong semi-supervised methods.

Key words: Question classification; Joint representations; Semi-supervised classification

1 引言

问答系统能够为用户提出的自然语言问题提供一个简明、准确的答案。随着网络中用户交互信息的迅猛发展, 问答系统越来越受到人们的关注。问题分类是其中最基础的任务, 可以有效地减小答案的候选空间, 并影响答案的抽取策略, 根据问题的不同类别采用不同的答案选择策略和知识库^[1]。

问题分类可以看作一种特殊的文本分类任务。然而, 问题分类与一般的文本分类任务存在明显的差别。首先, 问题文本通常比较短, 问题中包含的数目非常有限^[2]。仅利用问题文本进行分类, 会存在信息量非常缺乏的问题。此外, 传统的词袋模型并不能捕捉文本中词的语义关系, 丢失了大量有用信息; 其次, 已标注的问题资源比较匮乏, 语料标注需要大量的

* 收稿日期: 2016-07-24

定稿日期: 2016-08-05

基金项目: 国家自然科学基金重点项目 (61331011); 国家自然科学基金项目 (61375073, 61273320)

作者简介: 张栋 (1991—), 男, 硕士, 自然语言处理; 李寿山 (1980—), 男, 教授, 自然语言处理; 王晶晶 (1990—), 男, 硕士, 自然语言处理。

时间、人力和物力^[3,4]。因此，如何加入更多的辅助特征扩充问题信息并充分利用未标注样本信息提升分类性能^[5,6]，是问题分类任务迫切需要解决的重要问题。

针对以上问题，本文提出了基于问题与答案共同表示学习的半监督问题分类方法：

一方面，本文的方法引入神经网络语言模型，学习分布式词向量来表示每个词特征，该表示考虑了词语的语义信息和词语在上下文中的依赖关系并且所得特征维数可控，有效缓解词袋模型存在的数据稀疏性和语义敏感性问题。此外，该方法充分利用问题已有的答案来扩充分类信息，较好地解决了问题包含的词汇信息量少的难点。例如表 1 所示，在未利用答案文本的情况下，仅利用问题文本学习到的词向量作为特征表示，“360 问答开放平台是做什么的？”很难判断其所属类别。其原因可能是该问题中没有包含任何明显的主题信息，例如“电脑”、“互联网”等。然而，加入了答案文本后，由于答案文本中出现“互联网”，问题与答案共同学习词向量，问题特征所学得的多维度词向量也包含“互联网”等信息。因此，该问题可以被准确地识别为“电脑”类问题。

表 1 利用答案文本扩充问题信息实例
Table 1 Examples of question classification with answers

	正确类别——电脑	仅利用问题特征分类为	利用问题与答案特征分类为
问题	360 问答开放平台是做什么的?		
答案	360 问答企业平台是开放、简单、高效、精准的互联网企业服务平台.....	无法确定	电脑类

另一方面，本文的方法是一种半监督学习方法，能够充分利用大量未标注的问题与对应答案，通过互相帮助地分布式词向量学习，增加问题文本中词表示的内在信息，解决标注语料匮乏的问题^[7]。

具体而言，首先，本文引入神经网络语言模型，由于答案与问题保持语义一致性，可以利用答案扩充问题信息，共同学习词向量表示，如图 1 所示。其次，充分利用大量未标注的问题与对应答案参与词向量学习，增加问题词向量的表示能力。最后，将已标注的问题样本以词向量形式表示作为训练集，采用卷积神经网络（Convolutional Neural Network, CNN）训练问题文本分类器。实验结果表明，本文提出的基于问题与答案共同表示学习的半监督分类方法有助于大幅提升问题分类准确率。

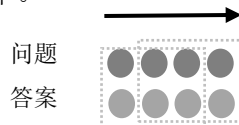


图 1 问题与答案共同学习词向量

Figure 1 Questions and answers learning word vectors jointly

本文其他部分组织如下：第二节介绍问题分类的相关工作；第三节描述问答语料的收集和构成；第四节介绍本文提出的基于问题与答案共同表示学习的半监督问题分类方法；第五节给出实验设置与结果分析；第六节简述结论及下一步工作展望。

2 相关工作

目前，问题分类研究主要采用基于特征统计的机器学习分类方法，且大部分的研究集中在全监督方式的问题分类上。Aijawa^[8]等人根据提问者期望主观或者客观的答案，将问题定义为主客观问题两种类型，利用一元和二元词特征采用平滑的朴素贝叶斯方法进行问题分类；Hui^[9]等人考虑了问题文本中词序和词间距对问题分类的影响，提出一种扩展类顺序规则模型；Mishra^[10]等人从问题文本中抽取出词特征、句法特征、语义特征，融合这些特征训练三

种分类器：最近邻、朴素贝叶斯、支持向量机，进行问题分类。

杨思春^[11]等人针对当前问题分类研究中特征提取的处理开销较大，提出一种融合基本特征和词袋绑定特征的问句特征模型。在分别提取问题中的词袋、词性等基本特征及其对应的词袋绑定特征的基础上，通过将基本特征与词袋绑定特征进行融合，以获取更加高效的问句特征集合。Liu^[12]等人认为标准核函数的SVM方法忽视了中文问题的结构信息，因而提出一种问题文本属性核函数的SMO方法，该方法利用了句法依赖关系和词性特征。

对于半监督方式的问题分类工作相对比较少。Chen^[13]等人利用基于图概念的标签传播算法实现问题分类，在百度知道语料上，评价结果远远高于KNN和SVM方法。与之不同的是，张栋^[14]等人提出了一种基于答案辅助的半监督问题分类方法。首先，将答案特征结合问题特征使用词袋模型一起来表示样本，然后利用标签传播方法（LP）自动标注未标注样本的类别，接着剔除答案特征，将初始标注的问题和自动标注的问题合并作为训练样本，利用最大熵模型对问题的测试样本进行分类，实验结果明显优于其他基准方法。

与以上研究不同的是，本文提出了基于问题与答案共同表示学习的半监督问题分类方法。一方面，引入神经网络语言模型，利用问题和对应答案共同学习词向量，有效地补充了问题文本的词向量信息。另一方面，充分利用大量未标注的问题文本以及对应的答案文本参与词向量学习，进一步增强问题词向量表示能力，从而提升分类准确率，减小人工标注工作量。

3 语料收集与描述

表 2 问题与答案实例
Table 2 Question and answer examples

类别	问题	答案
电脑/网络	没有 dns 网关 可以做桥接吗？	不可以的，需要提供 dns 网关。
文化/艺术	《当祖国召唤的时候》写于什么年代？	上世纪八十年代中期. 战争年代
体育/运动	冬奥会是是不是指冬天的奥运会？	正规点来说是 冬季的奥运会

本文语料来自360问答社区¹，其大类别总共有15个。本文随机抓取了其中五个类别的问答数据，分别是：电脑/网络、文化/艺术、健康/医疗、生活、体育/运动。每个类别包含2000条问答（每个问题对应一个答案），共10000条问答。表2给出几组问答实例。

4 基于问题与答案共同表示学习的半监督问题分类方法

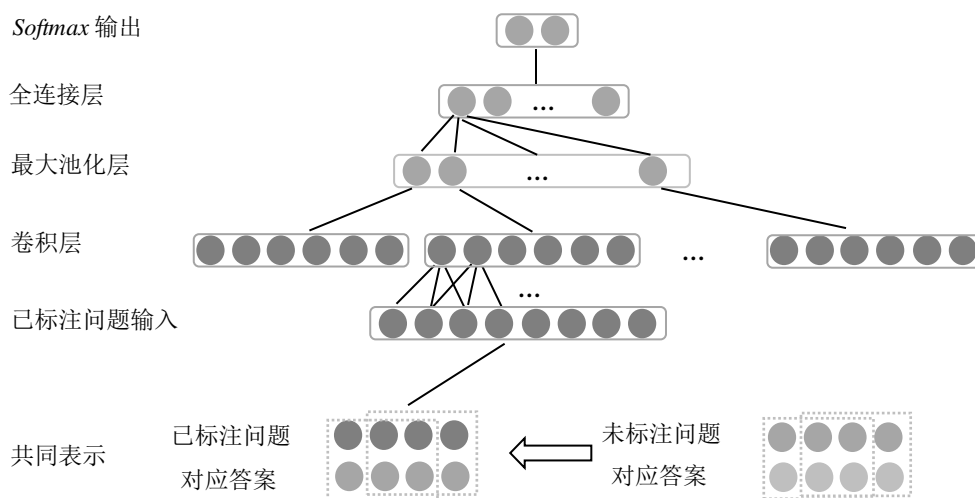


图 2 基于问题与答案共同表示学习的半监督问题分类方法架构图

Figure 2 Semi-supervised question classification with jointly learning question and answer representations

¹ <http://wenda.so.com/>

如图 2 所示, 是本文所提出的问题分类方法的完整架构图。首先, 我们引入神经网络语言模型, 利用问题与答案共同学习词向量表示, 增加问题词向量的信息量; 其次, 加入大量未标注的问题与答案样本参与词向量学习, 进一步增强问题词向量表示能力; 最后, 将已标注的问题样本以词向量形式表示作为训练样本, 采用卷积神经网络建立问题分类模型。

4. 1 词向量表示学习

本文引入神经网络语言模型, 把每个词表示成词向量的形式, 目的是将最小的语义单位表示成一个 n 维的向量, 向量中的每一维表示某种隐含的句法或者语义信息。该过程是一种无监督的过程, 因此, 可以充分利用大量未标注样本中词的共现信息、语义信息以及依赖关系。本文学习词向量的基础模型采用基于层次化 *Softmax* 的 *Skip-gram* 模型^[15]。该模型输出层对应一棵二叉树, 它是以语料中出现过的词当叶子节点, 以各词在语料中出现的次数当权重构造出来的 *Huffman* 树。在这棵 *Huffman* 树中, 叶子节点共 N 个, 对应词典 D 中的词。本文在此基础上提出一种问题与答案共同学习词向量的方式, 具体计算如下:

<p>算法流程:</p> <p>输入: 当前词 w; 当前词上下文词集合 $Context_q(w) \cup Context_a(w)$;</p> <p>输出: 窗口中所有词的向量表示 $v(\cdot)$;</p> <p>程序: FOR $u \in Context_q(w) \cup Context_a(w)$ DO { $e = 0$ FOR $j = 2:l^u$ DO { 1. $q = \sigma(v(u)^T \theta_{j-1}^u)$ 2. $g = \eta(1 - d_j^u - q)$ 3. $e := e + g\theta_{j-1}^u$ 4. $\theta_{j-1}^u := \theta_{j-1}^u + gv(w)$ } $v(u) := v(u) + e$ } }</p>
--

图 3 问题与答案共同学习词向量的算法流程

Figure 3 The algorithm flow of question and answer learning word vectors jointly

首先将问题和答案平行拼接, 补齐至等长。已知问题或者答案中当前词 w (此处假设是问题中的一个词), 需要对其上下文 $Context(w)$ 中 (该上下文既包含问题词也包含答案词) 的词进行预测, 因此目标函数表示为:

$$L = \sum_{w \in C_q \cup C_a} \log p(Context_q(w) \cup Context_a(w) | w) \quad (1)$$

这里, C_q 、 C_a 分别表示问题与答案语料, $Context_q(w)$ 表示 w 上下文中属于问题词的集合, $Context_a(w)$ 表示 w 上下文中属于答案词的集合。对数似然函数 (1) 的具体表达式为:

$$L = \sum_{w \in C_q \cup C_a} \sum_{u \in Context_q(w) \cup Context_a(w)} \sum_{j=2}^{l^u} \left\{ (1 - d_j^u) \cdot \log \left[\sigma(v(w)^T \theta_{j-1}^u) \right] + d_j^u \cdot \log \left[1 - \sigma(v(w)^T \theta_{j-1}^u) \right] \right\} \quad (2)$$

这里, 问题与答案两组窗口词组成 *Huffman* 树, $d_j^u \in \{0,1\}$ 表示路径 p^u (从根节点出发到达

词 u 对应叶子节点的路径) 中第 j 个节点对应的编码。 θ_{j-1}^u 表示路径 p^u 中第 j 个非叶子节点对应的向量。 $v(w)$ 表示当前词的词向量表示。 利用随机梯度法对其进行优化， 得到两类关键梯度：

$$\theta_{j-1}^u := \theta_{j-1}^u + \eta \left[1 - d_j^u - \sigma(v(w)^T \theta_{j-1}^u) \right] + v(w) \quad (3)$$

$$v(w) := v(w) + \eta \sum_{\substack{u \in \text{Context}_q(w) \\ \cup \text{Context}_a(w)}} \sum_{j=2}^u \left[1 - d_j^u - \sigma(v(w)^T \theta_{j-1}^u) \right] \theta_{j-1}^u \quad (4)$$

以样本 $(w, \text{Context}_q(w) \cup \text{Context}_a(w))$ 为例， 如图 3 所示， 给出了问题与答案共同学习词向量的算法流程。

4. 2 基于问题与答案共同表示学习的半监督问题分类方法

卷积神经网络模型最早被应用在图像处理中， 随后被证明在自然语言处理领域也是十分有效的^[16]， 可以充分利用数据内部结构信息^[17]。 图2描述了一种经典的处理文本的卷积神经网络模型^[18]。 在对卷积神经网络进行训练时， 本文采用的目标函数是最小化目标概率分布和实际概率分布的交叉熵， 具体定义为：

$$J(\theta) = -\frac{1}{t} \sum_{i=1}^t \log p(z^* | X_i, \theta) + \alpha \|\theta\|^2 \quad (5)$$

其中， t 是训练样本数目， α 是正则化因子， z^* 是样本 X_i 的正确类别， θ 是卷积神经网络中需要学习的超参数的集合。

根据训练目标函数， 通过梯度下降算法计算样本的误差， 并使用反向传播 (Back Propagation, BP) 的方式更新网络结构的参数集合 θ ， 具体的更新公式为：

$$\theta \leftarrow \theta + \lambda \frac{\partial J(\theta)}{\partial \theta} \quad (6)$$

算法流程：
输入： 已标注问题样本集合 L_q ， 对应的答案样本集合 L_a ， 分别包含 n^+ 个正类样本和 n^- 个负类样本； 未标注问题样本集合 U_q ， 对应的答案样本集合 U_a ； 问题测试样本集合 T ， 对应真实类别集合 C ； 输出： 问题测试样本分类正确率 Acc ； 程序： (1) 已标注问题样本与对应答案样本平行拼接为样本集合 L ； 未标注问题样本与对应答案样本平行拼接为样本集合 U ； $L+U$ 共同表示学习， 获得 L_q 词向量表示； (2) 将 L_q 分为训练集 $Trans$ 和验证集 $Vals$ ： a) 利用 $Trans$ 训练卷积神经网络分类器 CNN ， 并保存 CNN 模型； b) 利用 (a) 中保存的 CNN 预测 $Vals$ 类别， 并记录对应的分类正确率 $valAcc$ ； c) 以上两步迭代 m 次， 在这 m 次中选择 $valAcc$ 最高时对应的 CNN 模型作为最优分类模型； (3) 利用最优分类模型预测 T ， 获得测试样本预测类别集合 $predC$ (4) 比对测试样本真实类别， 获得分类正确的个数 num ， 则 $Acc = num/ C $

图 4 基于问题与答案共同表示学习的半监督分类方法流程

Figure 4 The algorithm flow of semi-supervised approach with jointly learning question and answer

如图 4 所示， 给出了基于问题与答案共同表示学习的半监督分类方法的流程。 首先， 我

们引入神经网络语言模型,利用问题与答案共同学习词向量表示,增加问题词向量的信息量;其次,加入大量未标注的问题与答案样本参与词向量学习,进一步增强问题词向量表示能力;最后,将已标注的问题样本以词向量形式表示作为训练样本,采用卷积神经网络建立问题分类模型,对测试问题样本进行类别预测。

5 实验

5.1 实验设置

数据集: 实验使用五个主题的问答语料,每个主题设计为一个二元分类问题,即“该主题”与“非该主题”分类。例如:“电脑”类 2000 条问答对(一个问题对应一个答案)，“非电脑”类是从其他四个主题中分别随机选取 500 条问答对,构成 2000 条问答语料。因此,实验中一共包括 5 个二元分类问题。5 个二元主题分类任务,分别随机选取每个主题 400 条问答(320 条问题作为训练集,80 问题条作为验证集)作为已标注样本,2800 条问答作为未标注样本,800 条问题作为测试样本(不包含答案)。语料分词采用复旦大学自然语言处理实验室开发的分词软件 FudanNLP²。

特征表示: (1) Baseline、LP: 选取词的一元特征 (*Unigram*), 采用词袋 (*Bag of Words*) 模型表示 (2) CNN: 采用分布式词向量表示 (*Word Embedding*)

分类器: (1) 采用 *MALLET* 机器学习工具包中的最大熵分类器³ (2) 采用 *Keras* 框架中卷积神经网络分类器⁴

评价标准: 使用准确率 (*Accuracy*) 作为结果的评价标准。

模型参数设置: (1) 最大熵模型采用默认参数 (2) 卷积神经网络模型参数也是采用默认参数, 设置如表 3 所示:

表 3 卷积神经网络分类模型中的参数设置

Table 3 Parameters setting in convolutional neural network

参数描述	值
词向量维度	100
词向量学习迭代次数	35
词向量学习窗口大小	5
问题输入最大长度	100
卷积核个数	50
卷积核宽度	3
全连接层维度	100
卷积神经网络分类模型迭代次数	3

5.2 实验结果与分析

实验比较的方法详细描述如下:

- **Baseline:** 只利用初始标注样本训练问题分类器(没有利用任何非标注样本)
- **LP(问题):** 利用已标注样本通过标签传播方法预测未标注样本的类别,将这些已确定类别的问题全部作为训练样本,训练问题分类器
- **LP(问题+答案):** 将对应的答案特征加入到已标注和未标注问题中,利用已标注样本通过 LP 方法预测未标注样本的类别,再剔除答案特征,将这些已确定类别的问题全部作为训练样本,训练问题分类器
- **CNN(问题):** 只利用已标注问题文本与未标注问题文本学习词向量,并使用卷积神经网络训练问题分类器

² <https://code.google.com/p/fudannlp/>

³ <http://mallet.cs.umass.edu/>

⁴ <https://github.com/fchollet/keras>

- **CNN (问题+答案)**: 使用标注、未标注的问题与对应答案共同学习词向量，并使用卷积神经网络训练问题分类器

图 5 给出了五种方法的分类性能比较。从图中结果可以看出，我们提出的问题与答案共同表示学习方法获得的分类效果明显优于其他方法，其分类准确率比使用问题与答案的 LP 方法提高了 2.9%。具体比较结果如下：

- (1) 从图中可以明显看出四种半监督方法相对于 **Baseline** 均有明显的性能提升，说明利用未标注样本对问题分类任务是十分有效的。
- (2) 对比 CNN (问题) 与 LP (问题)、CNN (问题+答案) 与 LP (问题+答案) 方法，可以发现 CNN 相关的两种方法要明显优于 LP 相关的两种方法。说明在问题分类任务中，基于词向量表示的半监督方法挖掘了词之间的语义关系，相比词袋模型表示的半监督方法表现出更加优越的性能。
- (3) 对比 LP (问题+答案) 与 LP (问题)、CNN (问题+答案) 与 CNN (问题) 方法，可以发现利用答案文本扩充问题文本信息，确实可以提升问题分类性能。
- (4) 本文提出的问题与答案共同学习词向量的半监督问题分类方法，在五个主题实验中分类准确率均远远高于其他四种方法。该结果表明在问题与答案共同学习词向量时，每个窗口中既包含问题词又包含答案词，可以更加有效地利用答案信息，从而提升问题词向量的表示能力，最终用来提升问题分类性能。

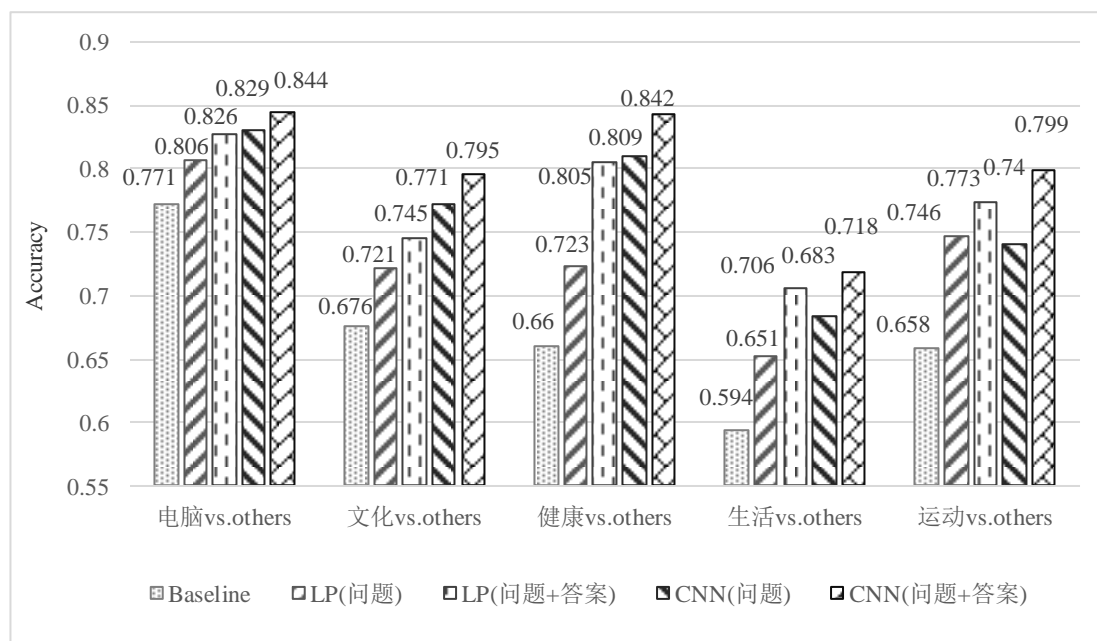


图 5 不同半监督分类方法性能比较

Figure 5 Comparison of different semi-supervised methods

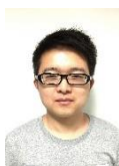
6 总结

本文提出了一种基于问题与答案共同表示学习的半监督问题分类方法。该方法的特色在于：一方面，利用问题文本与答案文本共同学习词向量，增强问题词向量的表示能力；另一方面，充分利用大量的未标注样本，参与已标注样本共同学习词向量，进一步捕捉数据中的隐藏信息，减少了大量的人工标注工作。实验结果表明该方法在不同主题的问题分类任务中都表现得非常优秀，分类性能明显优于传统的半监督学习方法。

下一步工作中，我们将考虑使用更多的表示学习方法（如主题词向量模型）进一步增加词向量所包含的信息。我们也将考虑利用更多的神经网络分类模型（如：循环、递归神经网络），考察这些方法是否可以进一步捕捉词向量之间的内在信息。

参考文献

- [1] 李鑫, 黄萱菁, 吴立德. 基于错误驱动算法组合分类器及其在问题分类中的应用[J]. 计算机研究与发展, 2008, 45(3):535-541.
- [2] 高超. 中文问题分类中特征选择研究[D]. 安徽: 安徽工业大学, 2011.
- [3] Li S, Huang C R, Zhou G, et al. Employing Personal/Impersonal Views in Supervised and Semi-supervised Sentiment Classification[C]. In Proceedings of ACL, 2010: 414-423.
- [4] Li S, Lee S Y M, Gao W, et al. Semi-supervised Text Categorization by Considering Sufficiency and Diversity[M]. Natural Language Processing and Chinese Computing. Springer Berlin Heidelberg, 2013: 105-115.
- [5] 高伟, 王中卿, 李寿山. 基于集成学习的半监督情感分类方法研究[J]. 中文信息学报, 2013, 27(3):120-126.
- [6] Wang J, Xue Y, Li S, et al. Leveraging Interactive Knowledge and Unlabeled Data in Gender Classification with Co-training[C]. International Conference on Database Systems for Advanced Applications. Springer International Publishing, 2015: 246-251.
- [7] Li S, Huang L, Wang J, et al. Semi-Stacking for Semi-supervised Sentiment Classification[C]. In Proceedings of ACL, 2015.
- [8] Aikawa N, Sakai T, Yamana H. Community qa question classification: Is the asker looking for subjective answers or not?[J]. IPSJ Online Transactions, 2011, 4: 160-168.
- [9] Hui Z, Liu J, Ouyang L. Question Classification Based on an Extended Class Sequential Rule Model[C]. In Proceedings of IJCNLP, 2011: 938-946.
- [10] Mishra M, Kumar Mishra V, Sharma H R. Question Classification Using Semantic, Syntactic and Lexical features[J]. International Journal of Web & Semantic Technology, 2013, 4(3).
- [11] 杨思春, 高超, 秦锋, 等. 融合基本特征和词袋绑定特征的问句特征模型[J]. 中文信息学报, 2012, 26(5):46-52.
- [12] Liu L, Yu Z, Guo J, et al. Chinese Question Classification Based on Question Property Kernel[J]. International Journal of Machine Learning & Cybernetics, 2014, 5(5):713-720.
- [13] Chen J, Su L, Li Y, et al. Label Propagation for Question Classification in CQA[M]. Advances in Swarm and Computational Intelligence. Springer International Publishing, 2015: 333-340.
- [14] 张栋, 李寿山, 周国栋. 基于答案辅助的半监督问题分类方法[J]. 计算机工程与科学, 2015, 37(12):2352-2357.
- [15] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[J]. arXiv preprint arXiv:1301.3781, 2013.
- [16] Kalchbrenner N, Grefenstette E, Blunsom P. A Convolutional Neural Network for Modelling Sentences[J]. arXiv preprint arXiv:1404.2188, 2014.
- [17] Collobert R, Weston J, Bottou L, et al. Natural language processing (almost) from scratch[J]. The Journal of Machine Learning Research, 2011, 12: 2493-2537.
- [18] Kim Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.



张栋 (1991-), 男, 通讯作者, 硕士研究生, 主要研究领域为自然语言处理。
电话: 15295676710 地址: 苏州市干将东路 333 号 邮编: 215006
E-mail: dzhangsuda@qq.com



李寿山 (1980-), 男, 教授, 硕士生导师, 主要研究领域为自然语言处理。

电话: 15862405154 地址: 苏州市干将东路 333 号 邮编: 215006

E-mail: lishoushan@suda.edu.cn



王晶晶 (1990-), 男, 硕士研究生, 主要研究领域为自然语言处理。

电话: 18896957736 地址: 苏州市干将东路 333 号 邮编: 215006

E-mail: djingwang@gmail.com