

结合 RNN 和 CNN 层次化网络的中文文本情感分类

罗帆 王厚峰[†]

北京大学计算语言学研究所, 北京 100871; [†] 通信作者, E-mail: wanghf@pku.edu.cn

摘要 提出一种多层网络 H-RNN-CNN, 用于处理中文文本情感分类任务。将文本按句子进行划分, 引入句子层作为中间层, 以改善文本过长带来的信息丢失等问题。模型中使用循环神经网络建模词语序列和句子序列, 并通过卷积神经网络识别跨语句的信息。探讨循环神经网络变种和不同输入向量对模型的影响。实验结果表明, 所提方法在多数数据集上都取得良好的效果。

关键词 中文情感分类; 深度学习; 卷积神经网络; 循环神经网络

中图分类号 TP312

Chinese Text Sentiment Classification by H-RNN-CNN

LUO Fan, WANG Houfeng[†]

Institute of Computational Linguistics, Peking University, Beijing 100871; [†] Corresponding author, E-mail: wanghf@pku.edu.cn

Abstract The authors present a hierarchical neural network H-RNN-CNN as a general model to represent text in sentiment analysis. Firstly, since information may lose in long text, the authors divide the text by sentence and use them as middle layer. Secondly, recurrent neural network is used to process sequence and relationship across sentences is captured by convolutional neural network. Moreover, the effectiveness of the variants of recurrent neural network and the pre-trained embedding are discussed. Experiment results demonstrate that the approach works well on several datasets.

Key words sentiment classification; deep learning; convolutional neural network; recurrent neural network

文本情感分类是指判断文本的情感极性。在互联网时代, 大量带有情感倾向的文本数据对人们的决策具有重要的作用。例如, 消费者在购物时, 希望根据商品已有的评论来选择。商家也可以通过消费者的评价信息, 对商品或相关的服务进行改进。

文本情感分类是自然语言处理中最常见的任务之一, 早期的研究方法有规则+词典的方法和基于统计的机器学习方法。其中, 规则+词典方法一般基于情感资源和句式模板。Kim 等^[1]利用情感词典, 通过每个词汇的情感得分来衡量文本的情感倾向。Turney^[2]通过构造固定的句式模板, 来分析文本的情感。基于统计的学习方法利用大量标注的数据, 来训练情感分析的分类器。Wang 等^[3]以二元语法

(bigram)为基本特征, 设计 NBSVM 分类器, 综合了朴素贝叶斯(Naïve Bayes, NB)和支持向量机(support vector machine, SVM)两种分类器, 在多个数据集上取得不错的效果。

精心设计的机器学习方法已经取得很好的效果, 但是特征和模板均需要人工进行设计, 所设计的特征往往没有很好的适应性, 当领域发生变化时, 需要重新设计。为了避免过多的人工设计特征, 研究者开始使用可以自动学习有用表示的深度学习方法。深度学习基于神经网络, 是一种端到端(end-to-end)的解决方案, 以词向量(word embedding)作为输入, 自动抽象出高级特征并学习到对应的参数, 最后得到输出。

对于小窗口内的情感信息,可以利用卷积神经网络(convolutional neural network, CNN)抽取特征。Kim^[4]使用单层的 CNN 对文本建模, Kalchbrenner 等^[5]也使用了 CNN 结构。两者的区别在于多层的卷积(convolution)和更复杂的池化(pooling)操作。对于长距离的情感信息,则采用循环神经网络(recurrent neural network, RNN)模型。Longpre 等^[6]研究长、短期记忆网络(long-short term memory, LSTM)及其变体在文本情感建模中的作用。Socher 等^[7]提出使用递归神经网络(recursive neural network, RecNN),对文本进行更细致的分析。RecNN 依赖于树的结构,一般使用短语结构树(constituency tree)或句法依存树(dependency tree),关注如何通过子节点的信息得到父节点的信息。Socher 等^[8-9]经过改进,提出矩阵-向量递归神经网络(matrix-vector recursive neural network, MV-RNN)和张量递归神经网络(recursive neural tensor network, RNTN)模型,并通过实验证明了模型的有效性。此外,也有研究关注如何使用更好的词向量来改进效果。Tang 等^[10]在 Collobert 等^[11]模型的基础上添加情感信息,使得学习到的词向量包含一定的情感知识,并通过实验证明了这种词向量在情感分类任务上的有效性。杜慧等^[12]在 Word2vec 的基础上,利用情感标签来进一步调整中文语料中的词向量。

除以词作为主要单位外,也有研究者使用字符甚至字节作为处理单位。Radford 等^[13]以 utf-8 编码的字节为单位,使用 Krause 等^[14]提出的乘法 LSTM (multiplicative-LSTM, M-LSTM)模型进行训练,利用在无监督语料上预训练得到的字节语言模型,可以在小规模短文本上取得更突出的效果。

此外,由于仅仅使用循环神经网络的隐藏状态

无法表达足够的信息,可能会丢失一些关键知识,所以还有使用记忆网络(memory network)的相关研究, Munkhdalai 等^[15]提出神经语义解编码器(neural semantic encoders),模拟图灵机的操作,通过读写等操作维护记忆,在需要使用时响应输出。

现有的方法在短文本情感分类上已经能够取得很好的效果,但在长文本上的效果稍差一些。主要原因是长文本带有更多的信息,一些与文本情感无关的信息可能成为噪声。通常,长文本的结构也比较复杂,存在一些跨句的情感信息,使用以上结构很难学习到这些信息。

1 Hierarchical RNN-CNN 模型

为了更好地对文本(特别是包含多句的长文本)建模,本文提出一种结合 RNN 和 CNN 的层次化网络(hierarchical RNN-CNN, H-RNN-CNN)模型,整体框架如图1所示,其中包括双层的 RNN 和顶层的 CNN。前两层都属于 RNN 的变种,第一层在词向量上使用 DVA(图2)对句子进行建模,第二层将得到的句子用 RNN 计算得到对应的隐藏层来表示,第三层对句子的隐藏层表示使用 CNN 建模,得到整体文本的表示。

1.1 DAN 和 DVA

DAN (deep averaging network)网络由 Iyyer 等^[16]提出,是一种带噪声求平均的多层网络,其定义如式(1)所示:

$$\text{DAN}(X) = \text{MLP}(\text{Average}(\text{Dropout}(X))), \quad (1)$$

即图 2 右侧部分。DAN 先用 Dropout 处理序列,之后求平均得到一个表示,再利用 MLP 实现更复杂的非线性变换。Longpre 等^[6]进一步提出 DVA (deep

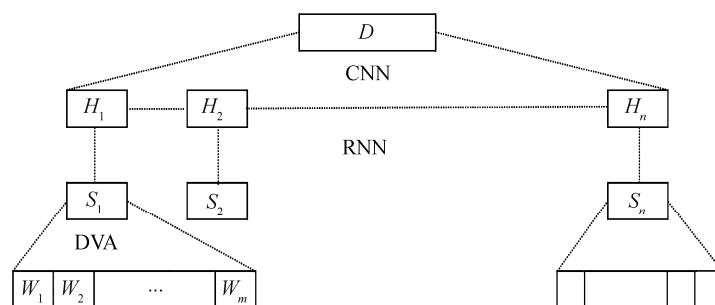


图 1 H-RNN-CNN 结构

Fig. 1 H-RNN-CNN model architecture

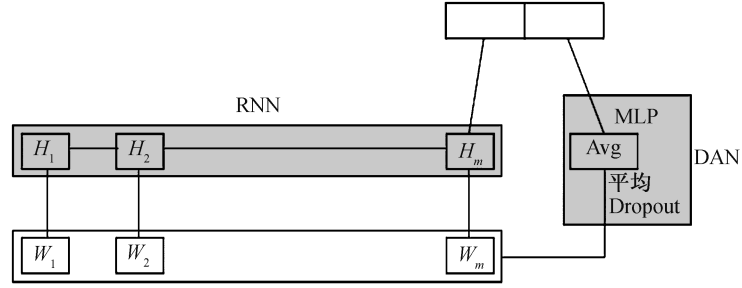


图 2 DVA 结构

Fig. 2 DVA model architecture

vector averaging)网络,由有序RNN和无序DAN两部分组成,其定义如式(2)所示:

$$\text{DVA}(X)=[\text{RNN}(X),\text{DAN}(X)], \quad (2)$$

其中, $[*,*]$ 表示行向量的拼接。

图2中, $W_1W_2...W_m$ 为词向量序列,对应一个句子。一方面,对词序列使用LSTM得到最后一个隐藏状态 H_m ; 另一方面,先计算经过Dropout处理的序列平均向量 Avg, 再利用MLP得到更抽象的表示。将两方面的表示拼接,最终得到句子的表示。

在传统机器学习方法中, unigram 和 bigram 是最主要的特征,然而 RNN 结构却没有能力学习这些 n-gram 的信息。DAN 结构中的求平均是一种最简单的利用 unigram 信息的方法,能够对 RNN 结构进行补充。

1.2 H-RNN-CNN

H-RNN-CNN 是一种对文档建模的多层网络。对于给定的文本,将其划分为多个句子。每个句子使用 DVA 进行建模,得到句子的表示 S 。使用 LSTM 得到每个句子对应的隐藏层表示 H 。利用这些表示,通过 CNN 进行建模,得到文本整体的表示 D 。

本文使用最常见的 LSTM 结构,共包括 3 种门 (Gate): 遗忘门 f 、输入门 i 和输出门 o 。在每一个时刻,3 种门都由上一时刻的隐藏层 h^{t-1} 和当前层 x^t 通过非线性变换得到。根据遗忘门 f^t 和输入门 i^t ,可以确定当前的细胞状态 C^t ,进而利用输出门 o^t 来更新隐藏层 h^t 。

$$f^t = \sigma(W_f \cdot [h^{t-1}, x^t] + b_f), \quad (3)$$

$$i^t = \sigma(W_i \cdot [h^{t-1}, x^t] + b_i), \quad (4)$$

$$o^t = \sigma(W_o \cdot [h^{t-1}, x^t] + b_o), \quad (5)$$

$$\tilde{C}^t = \tanh(W_c \cdot [h^{t-1}, x^t] + b_c), \quad (6)$$

$$C^t = f^t \times C^{t-1} + i^t \times \tilde{C}^t, \quad (7)$$

$$h^t = o^t \times \tanh(C^t). \quad (8)$$

顶层的 CNN 建模采用单层的卷积和 1-max 的池化操作。本文使用 n 个不完全相同的卷积核。以第 j 个核 $f_j (\in R_{kd})$ 为例, d 表示隐藏层的维数。式(9)定义 k 个句子的窗口表示 $h_{i:i+k} (\in R_{kd})$:

$$h_{i:i+k} = [h_i; h_{i+1}; \dots; h_{i+k-1}], \quad (9)$$

其中,行向量 h_i 表示第 i 个句子对应的隐藏层表示,“ $[:,]$ ”表示将行向量纵向拼接。对于文本中所有这样的窗口,保留经过卷积处理后得到的最大值。

$$d_j = \max(\sigma(f_j \circ h_{i:i+k} + b_j)), \quad (10)$$

式(10)中, \circ , b_j 和 σ 分别表示卷积运算、偏移和非线性函数,一个 f_j 可以得到一个最大值 d_j 。

$$d = [d_1, d_2, \dots, d_n], \quad (11)$$

式(11)中,整体文本的表示 $d (\in R_n)$ 由所有卷积核得到的最大值拼接得到,这样的表示可以理解为文本针对该任务的显著特征表示。

得到文本整体的表示 d 后,使用 softmax 进行分类。网络使用交叉熵作为损失函数,用 Adadelta 方式更新参数。

为防止过拟合,除使用 L2 正则和 Dropout 技术外,还使用 max-norm 和蒙特卡洛采样技术。max-norm 指在更新参数时,通过限制其矩阵范数来防止过大的更新幅度。蒙特卡洛采样由 Srivastava 等^[17]提出,由 Dropout 演变而来。在 Dropout 中,训练时按一定比例去掉输入,但在测试时则使用不同的处理方式,将所有输入按比例进行缩放。在蒙特卡洛采

样中,测试时也使用相同的处理方式,区别在于需要对输入进行多次采样。使用蒙特卡洛采样的原因如下:一是与像素值不同,按比例缩放词向量没有直观的解释;二是多份采样可以减少误差。

一些参数的取值范围如表1所示,在不同数据集上的取值略有不同。

表 1 一些参数的取值范围
Table 1 Hyper-parameters setting

词向量维度	批处理个数	隐藏层向量维度	Dropout
200~300	8~128	100~200	0.2~0.5
蒙特卡洛 采样个数	max-norm	卷积核窗口 长度	每种卷积核 个数
20~80	1~3	1, 2, 3, 4	20~120

2 实验

本文的实验评测以中文数据为主,同时也在英文数据上做了实验,以说明方法的适用性。

2.1 数据集和评价指标

本实验使用的数据集如表2所示,所有数据集都平衡地分为正负两类,基本情况如下。

1) PR(NLPCC 2014^①)。中文产品评论数据,包括书籍音像制品的评论。数据集中共有 10000 条训练样例和 5000 条测试样例。

2) ChnSentiCorp(Songbo Tan^②)。中文情感语料,包括书籍(Book)、酒店(Hotel)和笔记本电脑(Notebook)的评论。数据集中有去重的 4000 条样例,划分 15%为测试集。

3) MR(ACL 2004^③)。英文电影评论数据,共有 10662 条样例,划分 15%为测试集。

表 2 数据集概况
Table 2 Summary statistics for datasets

数据集	类别数	训练/开发/ 测试集	最长长度/ 词数	平均长度/ 词数
PR	2	8500/1500/5000	900	40
Book	2	2800/600/600	1500	80
Hotel	2	2800/600/600	1700	90
Notebook	2	2800/600/600	150	40
MR	2	7462/1600/1600	70	20

表 2 中,长度表示词的个数(近似)。本文使用的中文分词工具为 jieba^④。在评测中,我们使用 F1 值和准确率(Acc)两种指标。

2.2 结果与分析

为了进一步测试方法的有效性,我们选择多个 Baseline 进行比较。其中, CNN, LSTM 和 Bi-LSTM 分别为单层卷积网络、单层循环网络和单层双向循环网络。H-RNN 去掉了 H-RNN-CNN 顶层的 CNN 结构。同时引用在相同数据集上已有的代表性工作进行比较。其中, Wang 等^[18]在 NLPCC 2014 的该项评测上取得最好的效果,他们在 Mikolov 等^[19]的 Skip-Gram 模型中,将情感输出也作为目标函数的一部分,从而得到包含情感信息的词向量,最后将其用于 SVM。ChnSentiCorp 数据集有多个版本(从领域、规模到是否去重都不同),我们整理了 2012 年以来最好的一些使用单个分类器的方法。Zhang 等^[20]利用句子的子串特征,通过 KNN 和 Rocchio 进行分类。Luo 等^[21]提出一种结合 DF 和 CHI 的特征选择技术 CDPNC,使用 SVM 进行分类。Zhang 等^[22]使用 self-training 弱监督方法,通过迭代,不断更新整体效果。

我们在 5 组数据上做了实验,结果如表 3 所示。在实验中,我们实现的基于神经网络的方法均以预训练好的词向量作为输入。从表 3 可以看出, H-RNN-CNN 在多个数据集上都取得最好的效果。

表 3 实验结果比较 (%)
Table 3 Results of our model against other methods (%)

模型	F1/Acc				
	PR	Book	Hotel	Notebook	MR
CNN	74.3/72.9	91.4/91.2	89.1/89.3	91.8/ 92.2	-/81.1
LSTM	75.4/73.7	90.2/90.4	85.9/85.6	89.8/89.1	-/77.3
Bi-LSTM	75.8/74.3	92.2/92.5	87.7/87.0	91.4/90.6	-/79.5
H-RNN-CNN	77.9/75.5	92.9/92.6	90.3/89.4	92.1/92.1	-/81.9
H-RNN	77.1/75.2	92.8/ 92.6	90.1/89.2	92.0/91.1	-/81.5
Wang 等 ^[18]	76.9/-				
Zhang 等 ^[20]		-/91.0	-/87.1	-/87.7	
Luo 等 ^[21]			90.1*/-		
Zhang 等 ^[22]				90.5*/-	

说明: *表示使用未去重的数据集,加粗数字表示效果最好。

① http://tcci.ccf.org.cn/conference/2014/pages/page04_sam.html
 ② <http://www.nlp.ir.org/?action-viewnews-itemid-77>
 ③ <http://www.cs.cornell.edu/people/pabo/movie-review-data/>
 ④ <https://github.com/fxsjy/jieba>

实验数据集有两个明显的特性：一是很多文本的情感信息集中在小范围的上下文中；二是中文数据集文本的长度较长，相比之下，英文文本要短很多。CNN 中卷积权值共享和池化采样的特性有利于局部信息的识别，实验结果证明了这一点。特别是在较短的英文语料中，没有长文本带来的过多噪声干扰，CNN 模型的效果比 LSTM 和 Bi-LSTM 更好。LSTM 的问题在于容易忽视处于中间位置的词语，较难将情感信息完整地传递下来。双向 LSTM 注重同时参考句首和句尾的信息，因而能够纠正单向 LSTM 导致的一些错误。面对长文本，与单层 LSTM 相比，多层的网络取得更好的效果。多层的网络按照句子对文本进行划分，对于单个句子而言，较短的文本更容易提取到关键的情感信息；对于整体文本而言，比在词语和在少量的句子的建模容易得多。在语料中，有些文本的情感信息对相邻句子有一定的依赖性，句子级 CNN 结构的引入正是为了能够学习到跨句子的信息。实验结果表明，多加一层 CNN 结构，效果有不到一个百分点的提升。

我们还发现，在情感分析任务上，基于神经网络的方法并不依赖于大量的标注文本。在 Book, Hotel 和 Notebook 的训练集中，样本总数不到 3000 条，使用神经网络的效果却非常好。相比之下，基于统计的机器学习方法需要更大量的标注文本。

本文还对 RNN 的变种做了实验，结果如表所示。可以看出，我们采用的 DVA+MC (蒙特卡洛采样)取得更好的效果。使用 Bi-GRU 替换 LSTM，效果略微差一些，不过，由于门的数量减少，速度有一些提升。单独使用 DVA 不一定更有效，需要结合蒙特卡洛采样技术。模型采用的噪声方式在一定程度上相当于增加训练文本，整体而言，增强了模型的泛化能力。

本文以 PR 数据集为例，考察词向量、字向量和随机生成词向量等不同输入对 H-RNN-CNN 模型的影响，结果如表 5 所示。可以看出，虽然字向量的效果不大好，但可以作为词向量的补充，在词向

表 4 RNN 变种的影响

Table 4 Results with different RNN variants

数据集	Acc/%		
	Bi-GRU	DVA	DVA+MC
PR	74.8	74.2	75.5

说明：Bi-GRU 为双向 GRU，DVA+MC 和 DVA 的区别在于使用蒙特卡洛采样。

表 5 不同输入向量的影响

Table 5 Results with different input embeddings

数据集	Acc/%			
	词向量	字向量	字+词向量	随机词向量
PR	75.5	72.1	75.8	74.6

说明：词向量和字向量均使用 Mikolov 等^[19]提出的 Word2vec，在更大的语料(包括新闻和评论数据)上预先训练得到。

量中增加字词信息取得更好的效果。此外，与微调预训练好的词向量方法相比，使用随机生成的词向量方法效果稍差一些。原因可能是：1) 数据集较小，很难直接训练出足够好的向量，而预训练的向量使用了额外的非标注文本，包含更多共现、相关的信息；2) H-RNN-CNN 模型比较复杂，RNN 的结构也比较深，加大了训练随机生成向量的难度。字向量提升效果的原因之一在于，在中文处理中，字也表达了相当的含义；其次，在分词过程中，错误难以避免。单独使用字向量效果不够理想的原因可能是部分词(如外来词)以整体为单位，不可划分，并且词的歧义比字的歧义更少，一个字向量难以表达复杂的多义信息。本文将字和词的向量表示经过 RNN 结构处理，再将特征表示进行拼接，相当于补充了特征输入。

3 样例分析

样例 1 酒店设备陈旧，服务差劲，还不如二星级水准，更令人郁闷的是我早晨七点半下火车赶到酒店，却被告知要到九点方能入住，否则就要加半天房租。结果为了避免挨冻，多付了 250 元。结果携程还不算我的结分。真是冤透了。

经过分词、分句等预处理，样例 1 可以划分为以下 4 个句子。

1) 酒店 设备陈旧，服务 差劲，还 不如 二星级 水准，更 令人 郁闷 的 是 我 早晨 七点 半 下火车 赶到 酒店，却 被 告知 要 到 九点 方能 入住，否则 就要 加 半天 房租。

2) 结果 为了 避免 挨冻，多付 了 250 元。

3) 结果 携程 还 不算 我 的 结分。

4) 真是 冤透 了。

然后，对文本对进行补齐等处理，使得不同样例的文本都有相同的句子数，并且每个句子的词语数也一致。尤其是需要使用 CNN 结构，在文本前后都添加空句子，以便句首和句尾的卷积操作。这些操作使得并行运算变得更容易。本实验使用 GPU

进行并行运算。由于 GPU 显存的限制,需要调整批处理的大小和文本的大小(句子长度和句子数)。对词语进行编号,利用预训练好的词向量进行映射,即可使用 H-RNN-CNN 模型。

不同窗口大小的卷积核能够学习到不同的特征。窗口大小为 1 的卷积核提取最多的是第 4 句,其次为第 1 句,共占 60%。通过分析可知,第 4 句“真是冤透了。”和第 1 句“酒店设备陈旧,服务差劲,还不如二星级水准,更令人郁闷的是我早晨七点半下火车赶到酒店,却被告知要到九点方能入住,否则就要加半天房租。”确实表达了更突出的情感倾向。窗口大小为 2 的卷积核提取最多的则是前两句(也占 60%),也的确对文本整体的负面情感倾向起着重要的作用。说明顶层 CNN 结构的意义:窗口大小为 1 的卷积核可以捕获重要的句子,无论句子位于文本的任何位置,窗口更大的卷积核也能提取出相对重要的上下文窗口。如果仅仅使用 RNN 结构,这些重要特征相对不容易提取。

我们还分析了主要错误,其中多数同时包含褒贬的文本,例如样例 2 和样例 3。

样例 2 酒店门口时条单行道,外出有些不方便! 其他还行

样例 2 可以划分为以下两个句子。

1) 酒店 门口 时 条 单行道, 外出 有些 不方便!

2) 其他 还行

样例 3 因为工作关系,这里位置还是可以的,周围环境呢就不是很好,房间还行,干净就好。

样例 3 可以划分为: 因为 工作 关系, 这里 位置 还是 可以 的, 周围环境 呢 就 不是 很好, 房间 还 行, 干净 就 好。

样例 2 中,文本划分后得到两个句子,分别表达贬义和褒义的情感信息。在标注数据中的答案应为褒义,但模型将文本划分成贬义。尽管样例 1 显示了 CNN 结构提取特征的能力,但仍然有一定的噪声干扰,特别是在窗口较小时噪声更多。难点在于候选过多时噪声也增加。样例 3 是更为常见的情况,在一个句子而不是多个句子中表达了两种情感,我们认为这些错误不一定在于 CNN 结构,底层的 RNN 结构可能就很难对句子的情感做出判断,从而影响最终的预测。对于这种错误,我们使用的长、短期记忆网络(LSTM)在理论上是能够处理的,

可以利用遗忘门(forget gate)等机制,跟踪情感信息的变化。但是,在实际上不容易训练而导致过拟合。经过实验测试,可以通过对门(gate)等进行正则化运算,提高模型的泛化能力。

4 结语

文本情感分类是自然语言处理的重要内容,有广泛的应用需求。本文联合循环神经网络和卷积神经网络,提出多层网络模型 H-RNN-CNN。在模型中引入句子层,并使用两层的 RNN 对文本建模,有助于长文本的情感倾向判断。在此基础上,通过 CNN 来协助判断跨语句的信息。在多个中文数据集上的实验表明,本文提出的模型取得较好的结果,只在 Notebook 上的 Acc 略低于 CNN 模型。在英文数据 MR 上的实验结果也超过所有的基线系统。

在未来的工作中,我们将研究如何在本框架下进一步引入注意力机制,考察文本中不同句子对文本情感分类的影响。此外,针对小规模语料,可能需要更多外部知识的引入,以无监督方式学习到的词语、字符甚至字符知识也可以成为研究的方向。

参考文献

- [1] Kim S M, Hovy E. extracting opinions, opinion holders, and topics expressed in online news media text // Proceedings of the Workshop on Sentiment and Subjectivity in Text. Sydney, 2006: 1-8
- [2] Turney P D. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews // Proceedings of the 40th annual meeting on association for computational linguistics. Philadelphia: Association for Computational Linguistics, 2002: 417-424
- [3] Wang S, Manning C. Baselines and bigrams: simple, good sentiment and topic classification // Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Jeju Island, 2012: 90-94
- [4] Kim Y. Convolutional neural networks for sentence classification // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, 2014: 1746-1751
- [5] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences // Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1:

- Long Papers). Baltimore, 2014: 655–665
- [6] Longpre S, Pradhan S, Xiong C, et al. A way out of the odyssey: analyzing and combining recent insights for LSTMs [EB/OL]. (2016–11–16) [2017–06–01]. <https://arXiv.org/abs/1611.05104>
- [7] Socher R, Pennington J, Huang E H, et al. Semi-supervised recursive autoencoders for predicting sentiment distributions // Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Edinburgh, 2011: 151–161
- [8] Socher R, Huval B, Manning C D, et al. Semantic compositionality through recursive matrix-vector spaces // Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing. Jeju Island, 2012: 1201–1211
- [9] Socher R, Perelygin A, Wu J, et al. Recursive deep models for semantic compositionality over a sentiment treebank // Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle, 2013: 1631–1642
- [10] Tang D, Wei F, Yang N, et al. Learning sentiment-specific word embedding for twitter sentiment classification // Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Baltimore, 2014: 1555–1565
- [11] Collobert R, Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning // Proceedings of the 25th International Conference on Machine Learning. Helsinki, 2008: 160–167
- [12] 杜慧, 徐学可, 伍大勇, 等. 基于情感词向量的微博情感分类. 中文信息学报, 2017, 31(3): 170–176
- [13] Radford A, Jozefowicz R, Sutskever I. Learning to generate reviews and discovering sentiment [EB/OL]. (2017–04–05) [2017–06–01]. <https://arXiv.org/abs/1704.01444>
- [14] Krause B, Lu L, Murray I, et al. Multiplicative LSTM for sequence modelling [EB/OL]. (2016–09–26) [2017–06–01]. <https://arXiv.org/abs/1609.07959>
- [15] Munkhdalai T, Yu H. Neural semantic encoders [EB/OL]. (2016–07–14) [2017–06–01]. <https://arXiv.org/abs/1607.04315>
- [16] Iyyer M, Manjunatha V, Boyd-Graber J, et al. Deep unordered composition rivals syntactic methods for text classification // Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Beijing, 2015: 1681–1691
- [17] Srivastava N, Hinton G E, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 2014, 15(1): 1929–1958
- [18] Wang Y, Li Z, Liu J, et al. Word vector modeling for sentiment analysis of product reviews // Natural Language Processing and Chinese Computing. Springer. Shenzhen, 2014: 168–180
- [19] Mikolov T, Chen K, Corrado G, et al. Efficient estimation of word representations in vector space [EB/OL]. (2013–01–16) [2017–06–01]. <https://arXiv.org/abs/1301.3781>
- [20] Zhang Y, Xiang X, Yin C, et al. Parallel sentiment polarity classification method with substring feature reduction // Pacific-Asia Conference on Knowledge Discovery and Data Mining. Gold Coast, 2013: 121–132
- [21] Luo F, Li C, Cao Z. Affective-feature-based sentiment analysis using SVM classifier // 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD). Nanchang, 2016: 276–281
- [22] Zhang P, He Z. A weakly supervised approach to Chinese sentiment classification using partitioned self-training. Journal of Information Science, 2013, 39(6): 815–831