

doi: 10.11830/ISSN.1000-5013.201606077



多尺度卷积循环神经网络的情感分类技术

吴琼, 陈锻生

(华侨大学 计算机科学与技术学院, 福建 厦门 361021)

摘要: 结合卷积神经网络对于特征提取的优势和循环神经网络的长短期记忆算法的优势, 提出一种新的基于多尺度的卷积循环神经网络模型, 利用卷积神经网络中的多尺寸滤波器提取出具有丰富上下文关系的词特征, 循环神经网络中的长短期记忆算法将提取到的词特征与句子的结构联系起来, 从而完成文本情感分类任务. 实验结果表明: 与多种文本情感分类方法相比, 文中算法具有较高的精度.

关键词: 文本情感分类; 卷积神经网络; 循环神经网络; 长短期记忆; 多尺度

中图分类号: TP 391.4 **文献标志码:** A **文章编号:** 1000-5013(2017)06-0875-05

Sentiment Classification With Multiscale Convolutional Recurrent Neural Network

WU Qiong, CHEN Duansheng

(College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China)

Abstract: Combining the advantages of convolution neural network (CNN) for feature extraction and recurrent neural network (RNN) for long short-term memory, a new model based on multiscale convolutional recurrent neural network is proposed. This model utilize multi-size filter of CNN to extract word feature which contain a rich context information and use the long short-term memory algorithm of RNN to reflect the grammatical relations about the word and the sentence, and then completing the sentiment classification task. The experimental results show that: through comparing with many other sentiment classification, this new model has a high accuracy.

Keywords: text sentiment classification; convolutional neural network; recurrent neural network; long short-term memory; multiscale

从海量而庞杂的网络评论信息中分析和挖掘用户的兴趣取向或公众态度, 已成为政府和业界关注的问题. 网络舆情文本的情感分类技术已经成为自然语言处理的研究热点. 目前, 情感分类方法有多种. 基于深度学习的分类方法, 传统的支持向量机、朴素贝叶斯的分类方法和基于句法分析的方法都有不错的效果. Wang 等^[1] 提出多项朴素贝叶斯 (MNB) 模型和使用朴素贝叶斯特征的支持向量机模型 (NBSVM). 为了防止过拟合, Hinton 等^[2] 提出 dropout 方法; Wang 等^[3] 通过对 dropout 方法进行改进, 提出高斯 dropout (G-dropout) 和快 dropout (F-dropout). Dong 等^[4] 从另外一个角度, 根据情感表达的方式, 构建统计分析器, 得出句子的情感极性. 卷积神经网络 (convolutional neural network, CNN) 通过卷积滤波器提取特征^[5]. Kalchbrenner 等^[6] 利用卷积神经网络对句子进行建模. Kim^[7] 利用一个简单

收稿日期: 2016-06-28

通信作者: 陈锻生 (1959-), 男, 教授, 博士, 主要从事计算机视觉与多媒体技术的研究. E-mail: dschen@hqu.edu.cn.

基金项目: 国家自然科学基金资助项目 (61370006); 福建省科技计划 (工业引导性) 重点项目 (2015H0025)

的单层卷积神经网络,通过多种输入特征与参数设置方式进行对比实验. Severyn 等^[8]使用与 Kim 相似的卷积网络结构,但是参数初始化方式不同,完成对 twitter 文本的情感分析. Zhang 等^[9]针对 Kim 提出的卷积神经网络,从多个角度对实验结果的影响进行讨论. 另一种网络结构是循环神经网络(recurrent neural network,RNN),传统的循环神经网络在梯度反向传播过程中,可能会产生梯度消失现象. 为了解决这个问题,Hochreiter 等^[10]提出了长短时记忆模型. 可以看出,卷积神经网络可以方便地利用滤波器的尺寸,提取句子中每个词与其上文和下文中的关系,而通过使用长短时记忆模型可以处理任意句子长度序列,还可以更好地体现句子语法规范. 因此,为了更加方便灵活地提取词的上下文特征,充分利用语言特性,本文提出了基于多尺度的卷积循环神经网络模型.

1 多尺度的卷积循环神经网络模型

提出的多尺度卷积循环神经网络模型,如图 1 所示. 它包含两级结构:在卷积神经网络部分,使用 Mikolov 等^[11]从谷歌新闻中训练出来的 300 维的词向量作为每个词对应的特征,通过多尺度的卷积滤波器,提取多种具有丰富上下文的信息的特征;循环神经网络通过将得到的词的上下文信息特征进行组合,输入到网络中,最终的到情感分类结果.

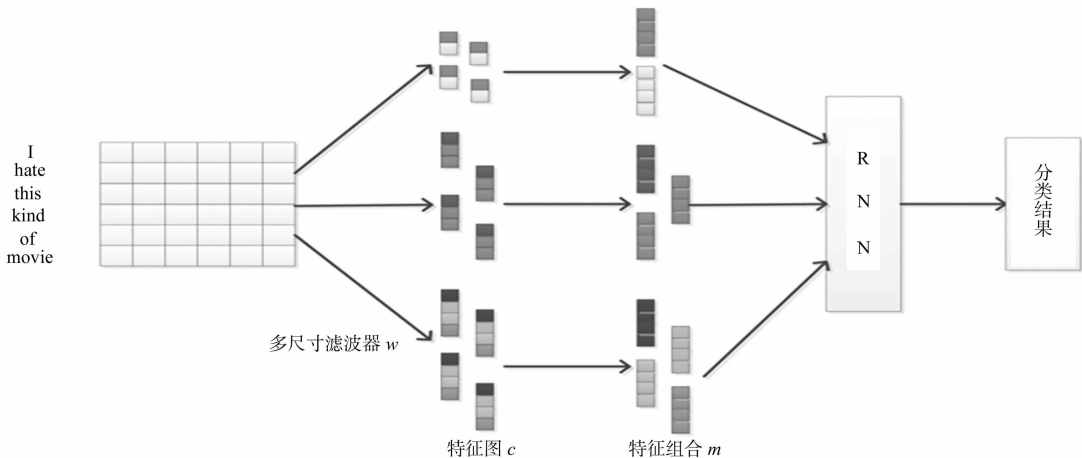


图 1 多尺度卷积循环神经网络模型

Fig. 1 Model architecture with multiscale convolution recurrence neural network

1.1 句子矩阵

x_i 表示一个句子的第 i 个词所对应的词向量,每个词向量的维度为 300 维. 由于句子包含的单词数量不等,通过补 0 的方式,将句子全部扩充成相同的长度. 那么,一个长度为 n 的句子可以表示为

$$\mathbf{X}_{1:n} = \mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_n. \tag{1}$$

式(1)中: $+$ 代表词向量的纵向连接操作. 那么,利用谷歌的词向量,就可以将所有句子都转换成大小相同的句子矩阵 $\mathbf{X}_{1:n} \in \mathbf{R}^{n \times 300}$,作为模型的输入.

1.2 卷积提取特征

对句子矩阵卷积操作时,会涉及滤波器的选择及初始化. 一个滤波器 $\mathbf{W} \in \mathbf{R}^{h \times k}$,其中, h 代表每次卷积参与到的词的数量,也就是滤波器的尺寸; k 代表词向量的维度,这样一个滤波器通过与一个包含 h 个词的字符串进行卷积运算后,就得到了一个标量特征. 如当 \mathbf{W} 滤波器卷积某一个字符串 $\mathbf{X}_{i:i+h-1}$ 时,特征 c_i 就产生了,其表达式为

$$c_i = f(\mathbf{W} \cdot \mathbf{X}_{i:i+h-1} + b). \tag{2}$$

式(2)中: $b \in \mathbf{R}$ 是一个偏置项; f 是一个非线性激活函数. 那么,当这个滤波器对整个句子矩阵进行逐窗口 $\{\mathbf{X}_{1:h}, \mathbf{X}_{2:h+1}, \cdots, \mathbf{X}_{n-h+1:n}\}$ 计算时,就会产生一个特征图 $\mathbf{C} \in \mathbf{R}^{n-h+1}$,表示为

$$\mathbf{C} = [c_1, c_2, \cdots, c_{n-h+1}]. \tag{3}$$

在特征图产生后,并不对它进行池化操作,因为得到的特征顺序对情感分类有很大的作用. 不难理解,句子之所以构成句子,是因为它是词的有序组合,所以会产生语法和句法. 因此,特征的顺序对于句

子结构的表示具有重要性.

以上描述了一种尺寸的滤波器对一个句子矩阵进行操作的过程. 文中模型使用多尺寸的滤波器, 每种尺寸包含多个滤波器对输入矩阵进行操作. 所以, 在对句子矩阵进行多尺寸的多个滤波器滤波后, 每种尺寸的多个滤波器产生出多个特征图. 那么, 通过某一尺寸的某个滤波器得到的特征图 C , 变换成

$$C(i, j) = [C(i, j)_1, C(i, j)_2, \dots, C(i, j)_{n-h+1}]. \quad (4)$$

式(4)中: i 表示第 i 个尺寸的滤波器, 实验采用 3 种尺寸的滤波器; j 表示同一尺寸的第 j 个滤波器, 由于卷积神经网络中的滤波器的参数是随机生成的, 同一尺寸包含的多个滤波器可以提取多种不同的特征, 从而获得同一尺寸下更丰富的特征, 更精准地分析句子情感. 对得到的多个特征图进行组合, 即

$$M_i^k = [C(i, 1)_k, C(i, 2)_k, \dots, C(i, m)_k], \quad (5)$$

$$M^i = [M_1^i, M_2^i, \dots, M_{n-h+1}^i]. \quad (6)$$

式(5), (6)中: i 代表第 i 种滤波器尺寸; m 代表每个尺寸滤波器的个数. 这样, 就得到输入到循环神经网络的输入特征.

1.3 循环神经网络

循环神经网络模型是长短时记忆的循环神经网络^[10], 其模型结构如图 2 所示.

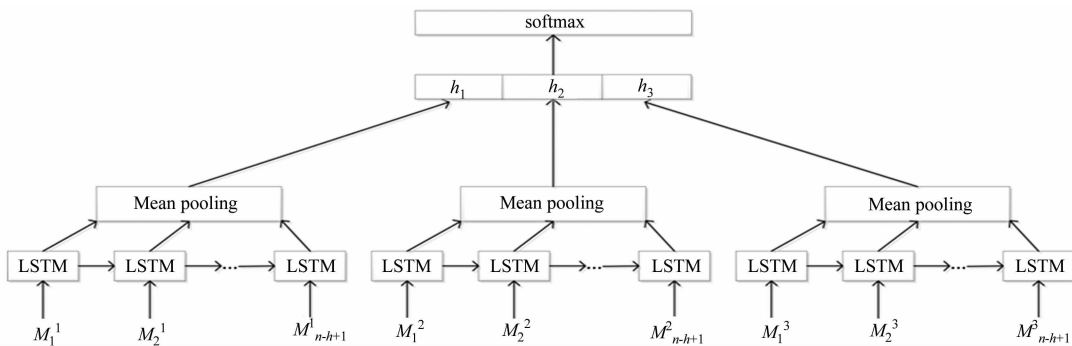


图 2 RNN 结构图

Fig. 2 Architecture of RNN

长短时记忆(long short-term memory, LSTM)算法引入了一个新的内存单元的结构, 它主要包含 4 个主要元素: 输入门、自连接的神经元、忘记门和输出门. 输入门控制输入信号的多少; 输出门控制内存单元输出对其他神经元的影响; 忘记门控制自连接神经元受多少之前状态的影响. 它的每一个内存单元标识为一个 LSTM 单元, 每个 LSTM 单元按照下式, 逐个时刻 t 进行计算, 即

$$i_t = \sigma(W_i M_t^i + U_i + b_i), \quad (7)$$

$$V = \tanh(W_c \cdot M_t^i + U_c \cdot h_{t-1} + b_c), \quad (8)$$

$$f_t = \sigma(W_f \cdot M_t^i + U_f \cdot h_{t-1} + b_f), \quad (9)$$

$$C_t = i_t \times V + f_t \times C_{t-1}, \quad (10)$$

$$o_t = \sigma(W_o \cdot M_t^i + U_o \cdot h_{t-1}), \quad (11)$$

$$h_t = o_t \times \tanh(C_t), \quad (12)$$

式(7)~(12)中: σ 表示 sigmoid 函数; W_* , U_* 表示随机初始化的参数, 下标 i, c, f, o 分别代表输入门、自连接、忘记门和输出门.

那么, 通过将每种尺寸滤波器滤波得到的特征 M^i , 逐一输入到循环神经网络中的对应位置, 从而得到对应尺寸的状态向量 h_i , 最后, 将每种尺寸滤波器得到的状态向量拼接成一个状态向量, 送入 softmax 分类器进行 $[0, 1]$ 分类, 便可得到句子所对应的情感分类.

1.4 正则化

在训练模型的过程中, 使用 dropout 方法和权重向量 L2 范数约束^[2]这两种方法防止过拟合现象的发生. dropout 方法是通过人为设置的随机概率, 将一些单元设置为 0, 从而让网络中的有些节点不工作. 对于不工作的那些节点, 可以暂时认为不是网络结构的一部分, 但是它的权重也要保留, 因为这些节点的权重只是暂时不更新. 当下一批样本输入时, 它就有可能又要工作, 那么当训练模型的时候, 相当于

每一次都在训练不同的网络,所以通过这样的方法可以有效地防止过拟合现象的发生.而权重向量的 L2 范数约束是在梯度更新过程中,将权重的 L2 范数强制的约束在某一范围中,使权重参数的元素值都很小,避免出现个别元素的值较大,对分类结果产生较大影响,从而有效地防止过拟合现象,提高模型的泛化能力.

2 实验与分析

为了测试文中模型,将提出的多尺度卷积循环神经网络模型与其他模型的情感分类实验结果进行对比分析.实验使用的是预训练的词向量,它来源于谷歌开源的 word2vector 工具,连续词袋结构从谷歌新闻中训练而来,向量的维度是 300 维^[11].使用 3 种数据集进行对比分析,一个是电影评论(MR)的数据集,每条数据是由一句话组成,总共包含 5 331 个消极评论和 5 331 个积极评论;一个是多种产品的顾客评论(CR),总共包含 1 367 条消极评论和 2 406 条积极评论;一个是观点极性判断的数据集(MRQA),一共包含 7 293 条消极观点和 3 311 条积极评论.其中,MR 数据集与 CR 数据集是评论数据集,它们的语句长度更长,而 MRQA 数据集主要是判断观点极性,所以数据集中的句子长度相对较短,甚至包含部分的单个单词和单词短语.3 个数据集的标签都是采用 0 和 1 对情感极性进行标注,其中,0 代表消极,1 代表积极.

对于所有数据集,采用 5 折交叉验证;迭代次数为 30 次;L2 范数的约束系数设置为 3;每批训练的大小为 50 个句子;dropout 设置为 0.5;选择 3 种滤波器尺寸,分别为 5,7,9;特征图选择的数量为 200.卷积前句子长度是对应数据集中句子的最大长度,而当输入到 LSTM 模型时,句子长度变为卷积前数据集中最大句子长度减去滤波器大小的长度,对应每个单词向量的维度变为特征图的数量(即 200 维),LSTM 模型最终的输出为二维向量,分别代表其情感极性.对于句子长度较长的数据集,滤波器的尺寸选择在一定范围内的增加可以提高准确率;但是当尺寸选择过大、甚至超越句子的平均长度时,准确率就会降低很多.由于特征图是用来尽可能提取丰富的特征,所以特征图选的数量越多,会增加实验的准确率;当增加到一定程度时,准确率基本不变化;再继续增大时,反而会降低准确率.随着滤波器的尺寸和特征图数量的增加,训练过程会变得相当耗时,且需要较大的内存空间存储.因此,通过对比不同数据集的实验结果,选择对各个数据集都相对合适的参数.

采用随机梯度下降算法的 Adadelta 更新规则^[12],该更新规则可以自适应地调整学习率,减少人为指定学习率给更新带来的影响.为了保证算法的鲁棒性,避免随机扰动带来的影响,实验结果是经过多次实验求平均值的结果,从而使该方法有更好的鲁棒性.情感分类结果准确率(η),如表 1 所示.表 1 中: M-CNN-RNN 代表文中提出的模型.

表 1 情感分类结果

Tab. 1 Sentiment classification results			
模型	$\eta/\%$		
	MR	CR	MPQA
NBSVM ^[1]	79.4	81.8	86.3
MNB ^[1]	79.0	80.0	86.3
G-Dropout ^[3]	79.0	82.1	86.1
F-Dropout ^[3]	79.1	81.9	86.3
SentParser ^[4]	79.5	—	86.3
LSTM ^[10]	80.2	81.9	88.2
CNN-non-static ^[7]	81.5	84.3	89.5
CNN-multichannel ^[7]	81.1	85.0	89.4
M-CNN-RNN	81.2	82.5	89.7

选择对比的模型都是近几年比较经典的模型,并且都在数据集中取得了不错的效果.通过对比实验结果发现,提出的模型获得了很好的结果.通过卷积神经网络多种尺寸的滤波器,可以提取比一般三元语法特征更多更广泛的上下文特征.在卷积神经网络后面连接循环神经网络分类器可以将词与句子的语法关系更好地体现出来,而且可以充分地利用卷积滤波器所提取出来的特征.所以,文中模型可以超越几乎所有的模型,也可以看出,提出的模型很适合情感分类任务.但是在 MR 和 CR 数据集中,情感分类结果略低于 CNN 两种模型结构的准确率.因为在评论数据集中,通常句子主要是表达用户及顾客的情感,然而却并不注意句子的语法规则,所以语句中的重点很明显,其他字词起的作用不大.而恰好在 CNN-not-static 和 CNN-multichannel 两种模型中,它们在进行卷积滤波操作后,对得到的特征进行池化操作,会将卷积滤波后所提取到特征中的最显著的特征保存下来,然后进行分类.而文中模型是利用卷积滤波器得到的全部特征进行情感分类,为了保证特征的完整性,并没有

进行池化操作.因此,CNN两种模型的准确率会略高于文中模型.

3 结束语

通过结合深度学习领域中常用的两种主流方法卷积神经网络和循环神经网络,提出基于多尺度的卷积循环神经网络模型.该模型利用卷积神经网络中的多尺寸滤波器,提取出具有丰富上下文关系的词特征,循环神经网络中的长短时记忆算法将提取到的词特征与句子的结构联系起来,从而将句子的结构与词的相互依赖关系尽可能好地体现出来.通过实验对比分析,文中模型获得了很好的实验结果,体现出该模型对情感分类的适用性.接下来,需要探索如何将注意力模型应用到文中模型中,改进网络的结构,使它更适应现实生活中的语言模型,提升文本情感分析的效果.

参考文献:

- [1] WANG Sida, MANNING C D. Baselines and bigrams: Simple, good sentiment and topic classification[C]// Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics. Stroudsburg: ACM, 2012: 90-94.
- [2] HINTON G, SRIVASTAVA N, KRIZHEVSKY A, *et al.* Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Science, 2012, 3(4): 212-223.
- [3] WANG Sida, MANNING C D. Fast dropout training[C]// Proceedings of the 30th International Conference on Machine Learning. Atlanta: JMLR, 2013: 118-126.
- [4] LI Dong, WEI Furu, LIU Shujie, *et al.* A statistical parsing framework for sentiment classification[J]. Computational Linguistics, 2014, 41(2): 293-336. DOI: 10.1162/COLI_a_00221.
- [5] LECUN Y, BOTTOU L, BENGIO Y, *et al.* Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324. DOI: 10.1109/5.726791.
- [6] KALCHBRENNER N, GREFFENSTETTE E, BLUNSOM P. A convolutional neural network for modelling sentences[C]// Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics. Baltimore: Eprint Arxiv, 2014: 655-665. DOI: 10.3115/v1/P14-1062.
- [7] KIM Y. Convolutional neural networks for sentence classification[C]// Proceedings of Conference on Empirical Methods in Natural Language Processing. Doha: [s. n.], 2014: 1746-1751. DOI: 10.3115/v1/d14-1181.
- [8] SEVERYN A, MOSCHITTI A. Twitter sentiment analysis with deep convolutional neural networks[C]// Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM, 2015: 959-962. DOI: 10.1145/2766462.2767830.
- [9] ZHANG Ye, WALLACE B. A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification[EB/OL]. (2016-04-06)[2016-06-15]. <http://arxiv.org/pdf/1510.03820v4.pdf>.
- [10] HOCHREITER S, SCHMIDHUBER J. Long short-term memory neural computation[J]. Neural Computation, 1997, 9(8): 1735-1780. DOI: 10.1162/neco.1997.9.8.1735.
- [11] MIKOLOV T, SUTSKEVER I, CHEN Kai, *et al.* Distributed representations of words and phrases and their compositionality[C]// Proceedings of Neural Information Processing Systems. South Lake Tahoe: Advances in Neural Information Processing Systems, 2013: 3111-3119.
- [12] ZEILER M. Adadelta: An adaptive learning rate method[EB/OL]. (2012-12-22)[2016-06-15]. <http://arxiv.org/pdf/1212.5701v1.pdf>.

(责任编辑: 黄晓楠 英文审校: 吴逢铁)