

DOI: 10.11830/ISSN.1000-5013.201606123



采用循环神经网络的情感 分析注意力模型

李松如, 陈锻生

(华侨大学 计算机科学与技术学院, 福建 厦门 361021)

摘要: 针对目前情感分析中的循环神经网络模型缺乏对情感词的关注的问题, 提出一种基于循环神经网络的情感词注意力模型, 通过引入注意力机制, 在情感分类时着重考虑文本中的情感词的影响. 在 NLPCCC 2014 情感分析数据集及 IMDB 影评数据集上进行试验, 结果表明: 该模型能够提高情感分析的效果.

关键词: 情感分析; 循环神经网络; 注意力; 长短时记忆

中图分类号: TP 311 **文献标志码:** A **文章编号:** 1000-5013(2018)02-0252-04

Recurrent Neural Network Using Attention Model for Sentiment Analysis

LI Songru, CHEN Duansheng

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

Abstract: Aim at the overlook of emotional words in the present recurrent neural network model used for sentiment analysis, we propose an emotional word attention model based on recurrent neural network. By introducing the attention mechanism, the model can pay more attention to the emotional words in the text sentiment classification. Experiments are conducted on the NLPCCC 2014 sentiment analysis dataset and IMDB movie review dataset, the results show that our model can improve the sentiment analysis effect.

Keywords: sentiment analysis; recurrent neural network; attention; long short term memory

情感分析是自然语言理解的一个基本任务, 其目标是对一段给定的文本分析其情感倾向^[1], 按分析结果可分为二元分类和多元分类^[1-2], 文中主要考虑二元情感分类. 自 Pang 等^[3]使用机器学习方法对文本进行情感分类之后, 很多学者都致力于设计更有效的情感分类器, 而机器学习方法的性能很大程度上依赖于所选取或提取的特征. 传统的文本特征表示方法, 如词袋模型特征、 n -gram 模型特征都面临数据稀疏及维数灾难等问题. 为了解决这个问题, 越来越多学者采用深度学习的方法进行情感分析^[4-9], 而在自然语言处理(NLP)领域使用深度学习模型时, 往往使用词向量^[4]作为词语的特征. 本文在率先情感分析中引入注意力机制^[8], 提出基于循环神经网络的情感词注意力模型.

1 模型的提出

1.1 双向长短时记忆

理论上, 使用传统的循环神经网络模型能够处理任意长度的序列数据, 但在实践中, 训练模型时, 梯

收稿日期: 2016-07-01

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)

度在反向传播过程中会逐渐减小直到消失, 形成梯度消失的问题. 因此, 序列数据的长度被限制在一个较短的范围. 而长短时记忆^[10] (LSTM) 模型通过引入门限机制, 在一定程度上克服了梯度消失问题. LSTM 的网络结构, 如图 1 所示.

长短时记忆模型的每一个单元由输入门 i 、输出门 o 、遗忘门 f 和内部记忆单元 c 组成, 通过 i, o, f 等 3 个门限对内部记忆进行选择性的输入、通过和遗忘操作, 能有效地克服梯度消失问题. 长短时记忆模型在 t 时刻隐藏状态为

$$\left. \begin{aligned} i_t &= \sigma(x_t U^i + h_{t-1} W^i + b^i), & f_t &= \sigma(x_t U^f + h_{t-1} W^f + b^f), \\ o_t &= \sigma(x_t U^o + h_{t-1} W^o + b^o), & g_t &= \tanh(x_t U^g + h_{t-1} W^g + b^g), \\ c_t &= c_{t-1} \otimes f_t + g_t \otimes i_t, & h_t &= \tanh(c_t) \otimes o_t. \end{aligned} \right\} \quad (1)$$

式(1)中: $\sigma(\cdot)$ 是 Sigmoid 激活函数; \otimes 为逐元素乘积运算; W^*, U^* 为网络中的权重参数矩阵; b^* 为各个偏置项; i_t, o_t, f_t 分别为 t 时刻输入门、输出门和遗忘门的值.

在原始的循环神经网络(RNN)模型中, 由于信息只按照从前向后的顺序传播, 所以 t 时刻的状态仅仅依赖于 t 时刻之前的信息, 为了使每个时刻都能包含它前后的上下文信息, 使用双向循环神经网络结构^[11]. 双向循环神经网络由前向 RNN 和反向 RNN 组成. 前向 RNN 按照正常的顺序读取输入序列 $(x_1 \sim x_t)$, 并计算出每个时刻的前向隐藏状态 (h_1, h_2, \dots, h_t) ; 反向 RNN 则按逆序读取输入序列 $(x_t \sim x_1)$, 并计算出对应时刻的反向隐藏状态 $(h_1^*, h_2^*, \dots, h_t^*)$. 最后, 将各个时刻的前向隐藏状态 h_j 与其对应时刻的后向隐藏状态 h_j^* 串联, 得到该时刻的隐藏状态, 即

$$h_j = [h_j \quad h_j^*]^T. \quad (2)$$

通过这种方式, 使 j 时刻的最终隐藏状态 h_j 包含该时刻前后两个方向上的上下文信息.

1.2 基于 RNN 的情感分析基础模型

由于循环神经网络的特殊结构, 将其用于情感分类时, 需考虑如何将各个时刻的隐藏状态用于最后的情感分类. 一种直观的方法是使用 RNN 最后一个时刻的隐藏状态作为情感分类的特征. 最后一个时刻的隐层状态 h_t 理论上把整个输入序列的语义都编码在其中, 后续情感分类则使用该语义编码作为 Softmax 分类器的输入, 计算得到情感极性属于第 c 类的概率为

$$p(y = c | h_t, U, b) = \exp(h_t U_c + b_c) / \sum_{l=1}^n \exp(h_t U_l + b_l). \quad (3)$$

式(3)中: U, b 为 Softmax 分类器的参数. 该模型称为 RNN-last, 其结构如图 2(a)所示.

对于长度较长的文本, 将所有语义信息编码为一个定长向量一定程度上会有信息的损失, 从而影响分类的效果. 为了利用每个时刻的语义信息, 在 RNN 与 Softmax 层之间加入一个 Mean pooling 层, 对每个时刻的隐藏状态取平均值, 即

$$\bar{h} = \sum_{k=1}^t h_k / t. \quad (4)$$

然后, 将 \bar{h} 作为 Softmax 分类器的输入. 与 RNN-last 模型仅使用最后一个时刻隐藏状态不同, RNN-mean 通过加入 Mean pooling 层, 将每个时刻的隐藏状态都用于最终的情感分类, 相当于每个时刻对最后的分类进行一个投票, 且投票的权重相同. RNN-mean 模型结构, 如图 2(b)所示.

1.3 基于 RNN 的情感词注意力模型

在上文的两个模型中, RNN-mean 的思想是每个时刻所得的隐藏状态对最终的情感分类的贡献程度都相同, 而 RNN-last 则是将整个词语序列的信息都压缩为一个固定长度的语义编码用于分类. 对于情感分析而言, 句子中的情感词对整个句子的情感倾向起着关键性的作用. 因此, 提出基于 RNN 的情感词注意力模型(RNN-attention), 通过引入注意力机制计算得到文本中每个词对情感分类的权重, 使情感词所在时刻的隐藏状态对情感分类的贡献程度更大. RNN-attention 模型的结构, 如图 3 所示.

在该模型中, 每个时刻将会学习到一个表示该时刻隐藏状态所占的权重 α_t , 最终用于分类的隐藏状态 \tilde{h} 通过加权求和的形式计算得到, 即

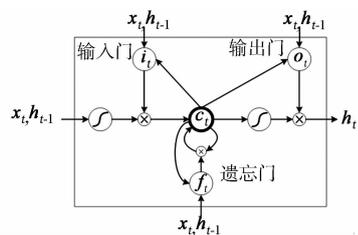
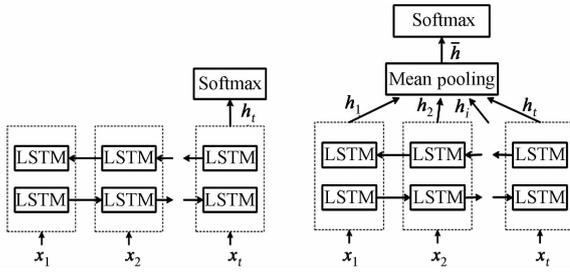


图 1 LSTM 单元结构

Fig. 1 Structure of LSTM cell



(a) RNN-last

(b) RNN-mean

图2 RNN-last和RNN-mean模型结构图

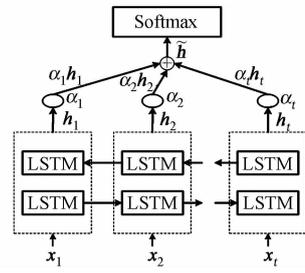


图3 RNN-attention模型结构图

Fig. 2 Architectures of RNN-last and RNN-mean model

Fig. 3 Architectures of RNN-attention model

$$\tilde{h} = \sum_{i=1}^t h_i \alpha_i. \quad (5)$$

权重 α_i 通过一个单层的神经网络计算得到,该网络的输入为该时刻原始的隐藏状态 h_i ,即

$$\alpha_i = \exp(\beta_i) / \sum_{k=1}^t \exp(\beta_k), \quad \beta_i = \mathbf{V}_a \tanh(\mathbf{h}_i \mathbf{W}_a + \mathbf{b}_a). \quad (6)$$

式(6)中: $\mathbf{W}_a, \mathbf{V}_a$ 为注意力模型的网络参数; \mathbf{b}_a 为偏置单元,通过对所有时刻的权重使用Softmax函数进行归一化,使所有权重之和为1.

权重 α_i 反映该时刻隐藏状态对最终情感分类的重要程度,通过有监督的学习方式,模型将较大的权重赋予情感词所在时刻的隐藏状态,实现了注意力机制,使分类器对每个时刻的隐藏状态进行有区别的对待,更加关注权重大的时刻,而不是对每个时刻平等对待或者仅使用末尾时刻的隐藏状态.最后,将 \tilde{h} 作为 Softmax 分类器的输入特征进行情感分类.

RNN-attention 模型提供了一种直观的方式检查模型对该句子的注意力所在,如图4所示.对选定句子样本中每个词语的权重进行可视化,每一个像素代表对应词语的权重(0表示黑色,1表示白色).

由图4给出的例子可知:模型正确识别出句子中的情感倾向较强的词(如 terrible, brilliant, 好听, 物超所值),并对其赋予了较大的权值,而后在分类时更多地考虑了这些词语的影响.

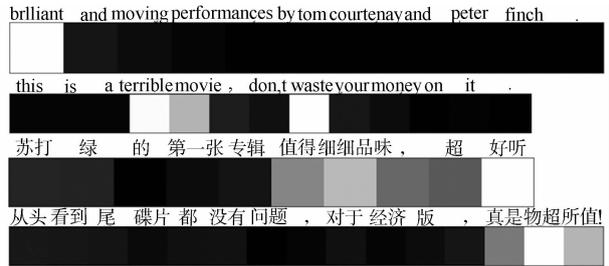


图4 样本的注意力权重可视化示例

Fig. 4 Examples of sample's attention weight visualization

2 实验及结果分析

2.1 实验设置

为了验证 RNN-attention 模型的有效性,在

NLPCC 2014 的“Sentiment Classification with Deep Learning Technology”评测数据及 IMDB 影评数据集上进行对比实验.其中,NLPCC 2014 的情感分析任务数据集包括训练集的 10 000 条中文产品评论,正负样本各 5 000 条,并包含 2 500 条测试样本,正负样本各 1 250 条;IMDB 影评数据的训练集和测试集各包含 25 000 条样本,正负样本各 12 500 条.实验的对比模型有:两个基本模型 RNN-mean 和 RNN-last、CNN 模型^[5]、Ensemble-SA 模型^[12].其中,Ensemble-SA 模型结合了 3 种不同的情感分析机器学习模型,该模型是目前 IMDB 数据集上的最新水平.实验的评价指标为宏平均 F1 值.

各模型所使用的词向量均采用连续词袋模型(CBOW)^[4]预训练得到,词向量的维度统一为 300 维,训练的语料来自搜狐新闻数据集,训练的工具包为 word2vec. RNN-last, RNN-mean, RNN-attention 模型的隐藏单元统一为 500 维, RNN-attention 的 attention 网络的隐藏层也设置为 500 维,使用批量梯度下降(SGD)训练 RNN-mean, RNN-last, RNN-attention 模型, Adadelta 算法^[13]用于更新学习率. CNN 模型和 Ensemble-SA 模型的参数与原文中的设置保持一致.

2.2 实验结果

不同模型在实验设置下的宏平均 F1 值,如表 1 所示.由表 1 可知:基于循环神经网络的模型较卷

神经网络模型有一定优势,这验证了 RNN 的结构更适合捕捉文本的语义信息,处理序列数据比 CNN 更加自然. RNN-attention 模型的效果优于 RNN-last 与 RNN-mean,接近 Ensemble-SA 模型的准确率,这说明 RNN-attention 通过实现情感词注意力机制能更加注重文本中情感词的影响. 各模型在中文数据上的效果较英文数据相差较大,这是由于两种语言内部差异所造成,不在讨论范围之内.

表 1 各模型实验的宏平均 F_1 值对比Tab. 1 Classification Macro- F_1 comparison of different models experiments

数据	宏平均 F_1 值				
	Ensemble-SA	CNN	RNN-last	RNN-mean	RNN-attention
NLPCC	79.23	75.76	76.14	77.81	78.59
IMDB	92.35	89.03	89.85	90.59	91.20

3 结束语

提出一种在传统循环神经网络模型中引入注意力机制的新模型,通过神经网络学习得到文本中各个词所对应的权重,并通过监督学习使情感词的权重大于一般词,克服了传统循环神经网络模型对情感词缺乏足够关注的缺点,通过实验在公开的情感分析数据集上获得比传统循环神经网络模型更优的效果. 在单一模型中逼近综合模型的效果,验证了文中模型的有效性. 该模型还存在一些问题需要进一步探讨,下一步工作将研究如何增强模型的鲁棒性,以适应各种不同长度和语言的文本.

参考文献:

- [1] LIU Bing. Sentiment analysis and opinion mining[J]. Synthesis Lectures on Human Language Technologies, 2012, 5(1): 1-167.
- [2] 赵妍妍, 秦兵, 刘挺. 文本情感分析[J]. 软件学报, 2010, 21(8): 1834-1848. DOI: 10. 3724/SP. J. 1001. 2010. 03832.
- [3] PANG Bo, LEE L, VAITHYANATHAN S. Thumbs up? Sentiment classification using machine learning techniques [C]// Proceedings of Conference on Empirical Methods in Natural Language Processing. Philadelphia: [s. n.], 2002: 79-86.
- [4] MIKOLOV T, SUTSKEVER I, CHEN Kai, *et al.* Distributed representations of words and phrases and their compositionality[J]. Advances in Neural Information Processing Systems, 2013, 26: 3111-3119.
- [5] KIM Y. Convolutional neural networks for sentence classification[C]// Conference on Empirical Methods in Natural Language Processing. Doha: [s. n.], 2014: 1746-1751. DOI: 10. 3115/v1/D14-1181.
- [6] IRSOY O, CARDIE C. Opinion mining with deep recurrent neural networks[C]// Conference on Empirical Methods in Natural Language Processing. Doha: [s. n.], 2014: 720-728. DOI: 10. 3115/v1/D14-1080.
- [7] BAHDANAU D, CHO K, BENGIO Y. Neural machine translation by jointly learning to align and translate[EB/OL]. (2015-04-24)[2016-05-12]. <https://arxiv.org/pdf/1409.0473v6>.
- [8] RUSH A M, CHOPRA S, WESTON J. A neural attention model for abstractive sentence summarization[C]// Conference on Empirical Methods in Natural Language Processing. Lisbon: [s. n.], 2015: 379-389. DOI: 10. 18653/v1/D15-1044.
- [9] XU K, BA J, KIROS R, *et al.* Show, attend and tell: Neural image caption generation with visual attention[C]// Proceedings of the 32nd International Conference on Machine Learning. Lille: ACM, 2015: 2048-2057.
- [10] GRAVES A. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [11] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks[J]. IEEE Transactions on Signal Processing, 1997, 45(11): 2673-2681. DOI: 10. 1109/78. 650093.
- [12] MESNIL G, MIKOLOV T, RANZATO M A, *et al.* Ensemble of generative and discriminative techniques for sentiment analysis of movie reviews[J]. Journal of Lightwave Technology, 2014, 32(17): 3043-3060.
- [13] ZEILER M D. ADADELTA: An adaptive learning rate method[EB/OL]. (2012-12-22)[2016-05-12]. <http://arxiv.org/pdf/1212.5701v1.pdf>.

(责任编辑: 钱筠 英文审校: 吴逢铁)