

基于双向 LSTM 模型的文本情感分类

任 勉, 甘 刚⁺

(成都信息工程大学 网络空间安全学院, 四川 成都 610225)

摘 要: 为解决文本情感分类研究中传统循环神经网络模型存在梯度消失和爆炸问题, 提出一种基于双向长短期记忆循环神经网络模型 (Bi-LSTM)。通过双向传播机制获取文本中完整的上下文信息, 采用 CBOW 模型训练词向量, 减小词向量间的稀疏度, 结合栈式自编码深度神经网络作为分类器。实验结果表明, Bi-LSTM 模型比传统循环神经网络 LSTM 模型分类效果更好, 对比实验中 Bi-LSTM2 能达到更优的召回率和准确率。

关键词: 双向长短期记忆循环神经网络; 词向量; 长短期记忆网络; 循环神经网络; 文本情感倾向性分析

中图法分类号: TP391.1 **文献标识号:** A **文章编号:** 1000-7024 (2018) 07-2064-05

doi: 10.16208/j.issn1000-7024.2018.07.044

Sentiment analysis of text based on bi-directional long short-term memory model

REN Mian, GAN Gang⁺

(College of Cybersecurity, Chengdu University of Information Technology, Chengdu 610225, China)

Abstract: To solve the problems of gradient disappearance and explosion in the traditional cyclic neural network model of text sentiment analysis, a neural network model based on bi-directional long short-term memory (Bi-LSTM) loop was proposed. The bi-directional mechanism was used to obtain the complete context information in the text, and the continuous bag of words model was used to train the word vector to reduce the sparseness between the word vectors, and the stack self-coding depth neural network was used as the classifier. Experimental results show that the Bi-LSTM model is better than the traditional cyclic neural network LSTM model, and the Bi-LSTM2 can achieve better recall rate and accuracy.

Key words: Bi-LSTM; word vector; long short-term memory; recurrent neural networks; sentiment analysis of text

0 引 言

通常对文本情感倾向性的传统方法有基于词典、基于语义分析等^[1-6]。在之前的自然语言处理中, 通常针对文本分类部分采用的是基于机器学习^[7]的分类算法。发展至今, 深度学习在自然语言处理中应用广泛。其中 Yong Zhang 等提出基于卷积神经网络的语句建模情感分析方法^[8], Bolan su 等提出基于循环神经网络进行对文本的分词处理^[9]。但这些方法存在训练过程中会丢失单词在上下文中关系的问题, 因此在后续的情感倾向性分析任务中引入了深度循环神经网络。通过对比以上信息, 循环神经网络中的长短期记忆网络 (long short-term memory, LSTM) 作为语言模型在解决文本分类问题上有很大的进步^[10], 但也存在梯度消失和爆炸等问题, 在上下文的记忆关联上也容易丢

失信息。

针对以上问题, 本文提出基于双向长短期记忆的循环神经网络模型 (bi-directional long short-term memory, Bi-LSTM)。通过该模型中的双向传播机制能降低上下文丢失信息的情况, 在文本情感分析任务中获得更好的分类效果。

1 相关工作

1.1 CBOW 模型

本文采用 CBOW (continuous bag of words) 来训练词向量^[11]。CBOW 是一种预测模型, 是从原始语句来推测目标字词, 在 CBOW 模型中, 已知词 w 中上下文 $context(w)$, 同时预测 w , 因此指定 $context(w)$, 词 w 就是一个正样本, 其它词就是负样本, 可以通过采样负样本得到一个关于 w 的负样本子集 $NEG(w)$, 因此词的标签可以表示如下

收稿日期: 2017-06-08; 修订日期: 2017-08-05

基金项目: 国家重大科技专项基金项目 (2014ZX01032401-001)

作者简介: 任勉 (1992-), 女, 四川广元人, 硕士研究生, 研究方向为数据挖掘、网络舆情、信息安全; +通讯作者: 甘刚 (1974-), 男, 四川茂县人, 硕士, 教授, 授研究方向为网络与信息系统安全、云计算与大数据安全、网络舆情。E-mail: test_me@qq.com

$$L^w(\tilde{w}) = \begin{cases} 1, \tilde{w} = w \\ 0, \tilde{w} \neq w \end{cases} \quad (1)$$

CBOW 模型目标函数表示如下

$$g(w) = \prod_{u \in (w) \cup NEG(w)} p(u | Context(w)) \quad (2)$$

其中

$$p(u | Context(w)) = \begin{cases} \sigma(X_w^T \theta^u), L^w(u) = 1 \\ 1 - \sigma(X_w^T \theta^u), L^w(u) = 0 \end{cases} \quad (3)$$

1.2 LSTM

长短时记忆网络 (LSTM) 解决了长程依赖的问题, 不需要特别复杂地调试超参数, 默认地记住了长期信息。而且解决了传统 RNN 模型存在梯度的消失和爆炸的问题^[12]。

图 1 为一个 LSTM 网络单元^[13], 使用门控单元来控制信息的流向。

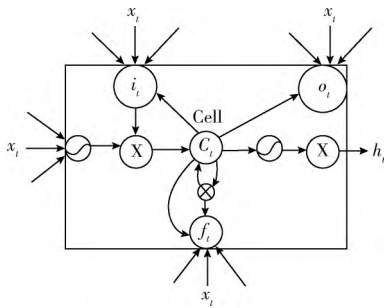


图 1 LSTM 单元网络结构

图 1 中的 i_t 代表输入门、 o_t 代表输出门、 f_t 代表遗忘门、 C_t 代表细胞控制单元结构。

LSTM 单元结构中 W_{ij} 是指神经元 i 到 j 的连接权重, 神经元输入是用 a 表示, 输出用 b 表示。

从细胞初始状态值到输入门状态值, 遗忘门状态值和输出门状态值的权重分别记做 $W_{i\alpha}$, $W_{f\varphi}$ 以及 $W_{o\omega}$, S_c 表示 Cell 的状态。

控制门的激活函数用 f 表示, g , h 分别表示 cell 的输入输出激活函数。

首先计算时间内存单元状态 i 的输入 i_t 和 \tilde{C} 的值

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (4)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (5)$$

其次计算当时存储单元的遗忘门的激活值

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (6)$$

给定输入门激活值 i_t , 遗忘门激活值 f_t 和候选状态值 \tilde{C}_t , 计算 C_t 存储单元在时间上的新状态 t

$$C_t = i_t * \tilde{C}_t + f_t * C_{t-1} \quad (7)$$

随着存储单元的新状态, 我们可以计算其输出门激活值, 然后计算其输出值

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + V_o C_t + b_o) \quad (8)$$

2 BI-LSTM 模型设计

2.1 两种 LSTM 变型

(1) lstm-peephole 连接模型:

lstm-peephole 连接模型 (如图 2 所示) 在 LSTM 单元网络结构上增加了 peephole connection, 控制门层接收输入门的激活值, 同时接收细胞状态的输入^[14,15]。

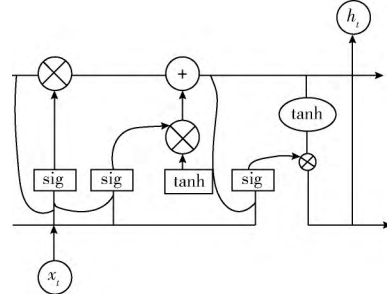


图 2 lstm-peephole 连接模型

改进的 peephole 设计门结构用来除去或增加信息细胞状态值, 包含一个 sigmoid 神经网络层和一个逐点乘法操作。Sigmoid 层输出 $[0, 1]$ 间的数据, 叙述每个连接之间有多少流量可以通过, 0 表示“禁止任何量通过”, 1 代表“允许任意量通过”。

通常在语言模型中, 我们可以基于已经看到的词来预测下一个词, 比如说“中国”-“北京”这对词组。一般细胞状态可能包含当前语句中主语的类别, 选择出正确的代词, 在此出现新的代词时, 该结构会忘记较为久的代词。在式 (9) 中, C_{t-1} 代表前一个时序的向量, 首先读取 C_{t-1} 和 x_t , 输出一个在 0 到 1 之间的数值给每个细胞状态 C_{t-1} 的数字

$$f_t = \sigma(W_f \cdot [C_{t-1}, h_{t-1}, x_t] + b_f) \quad (9)$$

式 (10) 是确定更新的信息

$$i_t = \sigma(W_i \cdot [C_{t-1}, h_{t-1}, x_t] + b_i) \quad (10)$$

更新细胞状态在 C_{t-1} 状态的值更新为 C_t , 把旧状态 C_{t-1} 与 f_t 相乘, 丢弃将要确定丢弃的信息, 加上 i_t 与 C_t 就是候选的新值, 最终根据每个状态的跟新程度进行变化

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (11)$$

在语言模型的例子中, 如果首先得到一个代词, 可能需要输出与一个与动词相关的信息

$$o_t = \sigma(W_o \cdot [C_t, h_{t-1}, x_t] + b_o) \quad (12)$$

通过确定需要输出的值来决定当前 cell 的状态, 也是一个过滤后的状态。首先运行一个 sigmoid 层来确定细胞状态的哪个部分将输出出去。把细胞状态通过 tanh 进行处理, 得到一个在 -1 到 1 之间的值并将它和 sigmoid 门的输出相乘, 最终输出确定输出的那部分。

(2) coupled 遗忘门和输入门

双门结构^[16]通过使用 coupled 遗忘门和输入门如图

3 所示。

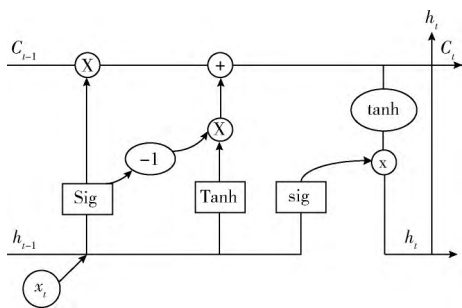


图3 coupled 遗忘门和输入门

不同于其它结构中把遗忘门的更新信息与其它位置信息分开,此结构可以同时完成把遗忘门和输入门值信息更新,在未来时间段输入在当前位置时忘记。输入新的值到已经忘记的旧信息的状态位置处

$$C_t = f_t * C_{t-1} + (1 - f_t) * \tilde{C}_t \quad (13)$$

2.2 基于 Bi-LSTM 模型的情感倾向性分类模型设计

通过改进两种 LSTM 模型后,结合分类器实现基于双向循环网络的情感倾向性分类模型。

该模型主要分为以下 3 个部分:词向量化、通过 Bi-LSTM 特征提取、深度神经网络分类器,如图 4 所示。

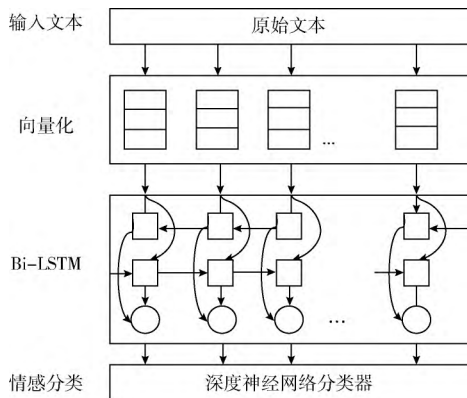


图4 基于 Bi-LSTB 语言模型情感倾向性分类模型结构

词向量化部分:词向量通常用于分类的文本数据量比较大,需要从大量未标注的普通文本数据中无监督地学习出词向量。在词向量训练之前,通常将字词转换成离散单独的符号,一个词对应一个向量,接着在文本分类模型中,将文章对应的稀疏矩阵合并成一个向量,即把每一个词对应的向量加到一起,此部分采用 CBOW 模型。

特征抽取部分:采用 Bi-LSTM 作为特征提取部分模型,同时在该部分的矩形框单元格代表一个 LSTM 单元,在后续的训练模型中,采用两种变型的 LSTM 模型。

Bi-LSTM 训练层:通过构造两个循环神经网络实现以两个不同的方向获取信息,同时这两层都连接相同的输入层。这个结构能够提供给上一层中每个单元结构完整的上

下文信息。

其中一层信息在同一时步向前传递,更新所有隐含层的信息

$$\tilde{h}_t = H(W_{\tilde{h}}x_t + W_{\tilde{h}h}h_{t-1} + b_{\tilde{h}}) \quad (14)$$

另一层信息的传播是与上一层相反,通过先计算输出层然后得到不同方向的隐含层值

$$y_t = w_{\tilde{h}y}\tilde{h}_t + W_{hy}h_t + b_y \quad (15)$$

情感分类层:此处采用栈式自编码式神经网络作为分类器^[17]。栈式自编码神经网络参数是通过逐层贪婪训练获得。以包含 2 个隐藏层,输出层为 softmax 的神经网络。

3 实验

实验工具采用的是 TensorFlow, TensorFlow 使用数据流式图来规划计算流程,它可以将计算映射到不同的硬件和操作系统平台。本文使用 TensorFlow 工具结合 Word2vec 生成词向量并实现 Bi-LSTM 模型的训练。

本次实验设计两组对比实验,第一组是利用不同的 LSTM 单元结构进行对比,其中一个 LSTM 变型结构是 LSTM-peephole 连接模型,另一个 LSTM 变型结构是 coupled 遗忘门和输入门构成。第二组对比实验是基于 Bi-LSTM 和 LSTM 结构,测试双向循环网络结构对分类的影响。

3.1 实验数据

实验语料来自 Github 上开源两万条中文标注预料,涉及到书籍、酒店、计算机、牛奶、热水器、手机 6 个方面的购物评论数。其中 16 000 条评论作为训练数据、4000 条作为测试数据。另一部分语料来自已经标注分类的两万条微博数据,主要是针对网络上某些热点事件的评论,其中 10 000 条为积极评价,另 10 000 条为消极评价。其中以 15 000 条作为训练数据,5000 条作为测试数据。

3.2 参数调整

首先是利用语料文本生成词向量,首先使用 CBOW 模型训练用少量噪声词汇来估计。当模型预测到真正的目标词汇为最高概率,同时预测其它噪声词汇为低概率。用编造的噪声词汇训练的方法被称为 Negative Sampling。这种方法计算 loss function 有非常高的效率,方法是选择 K 个随机词汇而非词汇表中的词汇。

在训练模型时通过调整词向量的维度、词频阈值、窗口大小来调整参数。

首先词向量的维度是从 50 到 200 之间进行测试,通过测试发现词向量维度在 120 左右时,测试数据 F 值最优,如图 5 所示。

由于词频阈值在小于 5 时不能生成词向量同时也会造成不能产生索引,因此词频阈值选择为 5。

窗口大小在训练过程中接近 20 时为准确率最高,如图 6 所示。

在训练 Bi-LSTM 时,通过损失值来观察迭代次数。在

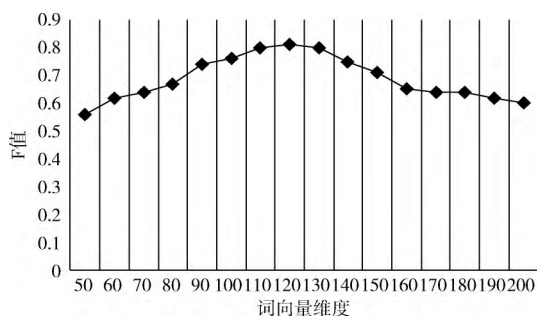


图 5 基于词向量维度增长的 F 值变化

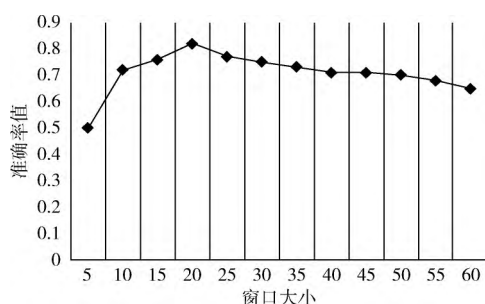


图 6 基于窗口大小增长的准确率值变化

实验过程中,发现使用 LSTM 模型时,迭代次数到 5 之后损失值维持不变,而 Bi-LSTM 在迭代次数到 10 时损失值维持不变,如图 7 所示。

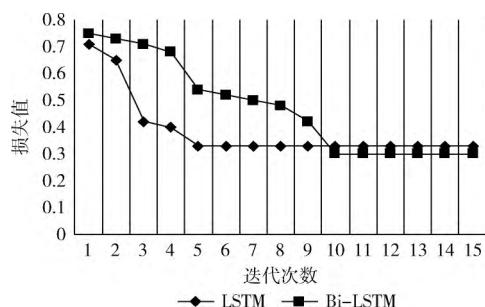


图 7 基于迭代次数增加损失值变化

最后分类器训练结果是一个 $[0, 1]$ 区间的连续的实数,而程序的在默认的情况下将 0.5 设置为阈值,因此最后将大于 0.5 的值判断为正,小于 0.5 的结果判断为负。

3.3 实验结果分析

本实验通过两组对比实验,分析 LSTM 的两种变型结构与 Bi-LSTM 模型之间的评价指标数据。

首先第一组是利用不同的 LSTM 单元结构进行对比,第一组 LSTM 变型结构是 LSTM-peephole 连接模型,另一个是 coupled 遗忘门和输入门在下表中分别称为 Bi-LSTM-1 与 Bi-LSTM-2,见表 1。

第二组对比实验是通过分析标准 LSTM 与 Bi-LSTM 之间的差异,见表 2。

表 1 两种 LSTM 结构对比

模型结构	准确率	召回率
Bi-LSTM-1	0.7521	0.7415
Bi-LSTM-2	0.8236	0.8125

表 2 标准 LSTM 与 Bi-LSTM 对比

模型结构	准确率	召回率
LSTM	0.7859	0.7654
Bi-LSTM	0.8421	0.8658

下面对实验结果数据进行分析:

(1) 根据第一组的对比实验看,LSTM 结构的两种变型运用在情感分类的语言模型中,第二种模型使用 coupled 遗忘门和输入门的结构有较高的准确率。

(2) 从第二组对比实验来看,使用双向循环神经网络对只使用单层 LSTM 结构准确率要高,说明使用上下文信息之间关联同时考虑时序问题的这种方式能够更好地解决文本情感倾向性分类问题。

4 结束语

本文在总结之间的文本情感分类基础上,提出一种 Bi-LSTM 语言模型来完成对文本倾向性的分类,通过训练序列向前和向后的两个循环神经网络,得到完整的过去和未来的上下文信息。通过实验对比不同的 LSTM 变型结构,结合深度神经网络分类器更好地完成分类任务,得到双向循环网络结构能更好地解决上下文联系的问题。本次实验中同时也有待完善的部分,如在语料的选择上,需要对比不同语料的分类情况来完善模型,在后续的工作中将会改进这一部分并研究不同的分类器构造方法。

参考文献:

- [1] XING Yunfei. Research on the characteristics and path of network public opinion information transmission in mobile environment [D]. Changchun: Jilin University, 2016 (in Chinese). [邢云菲. 移动环境下网络舆情信息传播特征及路径研究 [D]. 长春: 吉林大学, 2016.]
- [2] HU Gaili, Chen Ting, CHEN Fuji. A literature review on the network of public opinion in China [J]. Information Science, 2016, 34 (1): 160-166 (in Chinese). [胡改丽, 陈婷, 陈福集. 我国网络舆情热度分析文献综述 [J]. 情报科学, 2016, 34 (1): 160-166.]
- [3] Chun Liao, Chong Feng, Sen Yang. Topic-related Chinese message sentiment analysis [J]. Neurocomputing, 2016, 210 (1): 237-246
- [4] Chen Tao, Xu Ruifeng, Yang Sen. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN [J]. Expert Systems with Applications, 2016, 72

- (10): 221-230.
- [5] Fu Xinghua, Liu Wangwang, Xu Yingying. Combine HowNet lexicon to train phrase recursive autoencoder for sentence-level sentiment analysis [J]. Neurocomputing, 2017, 241 (79): 18-27.
- [6] Meng JE, Yong Z, Ning W. Attention pooling-based convolutional neural network for sentence modelling [J]. Information Sciences, 2016, 373 (84): 388-403.
- [7] Su B, Lu Shijian, Chu Chingyun. Accurate recognition of words in scenes without character segmentation using recurrent neural network [J]. Pattern Recognition, 2017, 34 (1): 397-405.
- [8] Taft M, Xu J, Chowdhury V. Letter coding in visual word recognition: The impact of embedded words [J]. Journal of Memory and Language, 2017, 25 (2): 14-25.
- [9] Liao Chun, Yang Sen, Huang Heyan. Topic-related Chinese message sentiment analysis [J]. Neurocomputing, 2016, 14 (10): 237-246.
- [10] Ren Yafeng, Wang Ruimin, Ji Donghong. A topic-enhanced word embedding for Twitter sentiment classification [J]. Electronic Letters, 2016, 10 (11): 188-198.
- [11] Lohar P, Ganguly D, Afli H, et al. FaDA: Fast document aligner using word embedding [J]. Prague Bulletin of Mathematical Linguistics, 2016, 106 (1): 169-179.
- [12] Yousfi S, Berrani S, Garcia C. Contribution of recurrent connectionist language models in improving LSTM-based Arabic text recognition in videos [J]. Pattern Recognition, 2017, 41 (5): 245-254.
- [13] Chen Tao, Rui Feng, He Yulan. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN [J]. Expert Systems with Application, 2017, 72 (15): 221-231.
- [14] Greff K, Srivastava K, Koutnik J. LSTM: A search space odyssey [J]. IEEE Transactions on Neural Networks and Learning Systems, 2017 (99): 1-11.
- [15] Nogueira C, Stanos D, Gatti R. Deep convolutional neural networks for sentiment analysis [C] //25th International Conference on Computational. Ireland: Technical Papers, 2014: 23-29.
- [16] Wei Jiamei, Pei E, Jiang Dongmei. Multimodal continuous affect recognition based on LSTM and multiple kernel learning [C] //Signal and Information Processing Association Annual Summit and Conference. Piscataway, NJ: IEEE, 2014: 1-4.
- [17] Severyn A, Moschitti A, Chu Chingyun. Learning to rank short text pairs with convolutional deep neural network [C] //38th International ACM SIGIR Conference on Research and Development in Information Retrieval Symposium on Circuits and Systems. Newyork: ACM, 2015: 373-382.

(上接第 2063 页)

- [11] LI Qingwu, QIU Chunchun, YU Kai, et al. Real time pedestrian detection based on multi-scale aggregation channel features [J]. Journal of Electronic Measurement and Instrumentation, 2015, 29 (11): 1691-1697 (in Chinese). [李庆武, 仇春春, 俞楷, 等. 基于多尺度聚合通道特征的实时行人检测 [J]. 电子测量与仪器学报, 2015, 29 (11): 1691-1697.]
- [12] Rothe R, Guillaumin M, Van Gool L. Non-maximum suppression for object detection by passing messages between windows [C] //Asian Conference on Computer Vision. Springer, Cham, 2014: 290-306.
- [13] Ba LJ, Frey B. Adaptive dropout for training deep neural networks [J]. Advances in Neural Information Processing Systems, 2013: 3084-3092.
- [14] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding [C] //Proceedings of the 22nd ACM International Conference on Multimedia. ACM, 2014: 675-678.
- [15] Dollár P, Wojek C, Schiele B, et al. Pedestrian detection: An evaluation of the state of the art [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2012, 34 (4): 743-761.