

多主题下基于 LSTM 语义关联的长文本过滤研究

曹春萍, 武 婷

(上海理工大学 光电信息与计算机工程学院, 上海 200082)

摘要:现如今互联网上出现了很多评论性文章, 这些文章字符数多, 且包含较多与主题无关的信息, 会影响后续的文本分析任务的性能。因此, 针对传统的解决方案不能够对多主题长文本进行建模, 以及现有的神经网络无法从相对较长的时间步长中捕获语义关联等问题, 文中提出了一种结合单层神经网络和分层长短记忆网络的深度网络模型, 并在长文本过滤任务中进行应用。该模型通过词语层 LSTM 网络获得句子内部词语之间的关系并得到具有语义的句向量, 然后将句向量输入主题依赖度计算模型和句子层 LSTM 网络模型, 进而得到句子与各主题类别的依赖度以及待过滤句子与其他句子之间的关联。最后, 在从马蜂窝获取的游记数据集上进行的实验表明, 该模型相比 SVM、朴素贝叶斯、LSTM、Bi-LSTM 等效果更好。

关键词:长文本过滤; 多主题; 语义关联; LSTM; 分层模型

中图分类号: TP31

文献标识码: A

文章编号: 1673-629X(2019)11-0001-06

doi: 10.3969/j.issn.1673-629X.2019.11.001

Research on LSTM Semantic Correlation Long Text Filtering Based on Subject Dependence

CAO Chun-ping, WU Ting

(School of Optical-electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200082, China)

Abstract: Nowadays, there are a lot of critical articles on the Internet. These articles have more characters and contain more information irrelevant to the topic, which will affect the performance of subsequent text analysis tasks. The traditional solution cannot model multi-theme long text, and the existing neural network cannot capture the semantic association from a relatively long time step. Therefore, we propose a deep network model combining single-layer neural networks and layered long and short memory networks. The model obtains the relationship between the internal words of the sentence through the word layer LSTM network and obtains the semantic vector. Then the sentence vector is input into the subject dependence calculation model and the sentence layer LSTM network model, in turn, the dependence of the sentence and each topic category and the relationship between the sentence to be filtered and other sentences are obtained. Finally, experiments on the travel data set acquired from the Mafengwo shows that this model is superior to SVM, Naive Bayes, LSTM, Bi-LSTM.

Key words: long text filtering; multi-topic; semantic association; LSTM; hierarchical model

0 引言

近年来, 随着社交网络愈来愈繁荣, 如何从海量文本中进行信息提取受到越来越多的关注。文本过滤作为更多文本分析的前序工作, 成为了自然语言处理工作的研究热点之一^[1]。随着互联网进入 Web 2.0 时代, 互联网用户使用网络分享知识、经验、意见、感受等, 越来越多的主观性评论文章充斥网络。与科学性

文章不同, 评论性文章不仅包含多个主题而且含有很多与主题无关的句子, 过滤掉这些句子可以提高后续文本分析的效率, 所以对评论性文章内容进行文本过滤是进行文本分析中很重要的一步。目前对文本过滤的研究已经有很多, 但是在长文本过滤方面仍存在一些亟待解决的问题, 如多主题、句子间语义关联等问题造成长文本过滤困难。因此, 对评论性文章进行文本

收稿日期: 2018-12-24

修回日期: 2019-04-25

网络出版时间: 2019-06-27

基金项目: 国家自然科学基金(61402288); 上海市自然科学基金(15ZR1429100)

作者简介: 曹春萍(1968-), 女, 硕士, 副教授, 硕导, 研究方向为智能数据处理、个性化服务; 武 婷(1993-), 女, 硕士研究生, 研究方向为数据挖掘、自然语言处理。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190627.1105.052.html>

过滤依然具有很大的研究价值。

为了从文本中筛选出有用的信息,越来越多的学者采用自然语言处理技术进行文本过滤。传统的基于规则的过滤方法,性能好坏依赖于复杂的人工规则和特征工程^[2]。评论性文章内容冗长杂乱,领域性不强等特点使得设计规则和提取特征都非常困难。此外,每个文本都有对应的主题,利用主题模型可以提取文本主题,根据文本主题进行过滤,虽然采用机器学习的方法能在主题分类达到很好的效果,但是仍然依赖人工抽取特征,对专家经验要求较高^[3]。深度神经网络模型在自然语言处理方面的巨大成功,使得研究人员将其应用于文本过滤。深度神经网络具有强大的特征学习能力,能够克服人工特征抽取的困难^[4]。但是对于评论性长文本过滤,不同的模型和数据处理方法会对过滤效果产生不同的影响。

文中主要针对的是评论性文章多主题且句子语义关联造成的长文本过滤困难的问题。多主题是指一篇文章中涉及到多个主题,如游记中通常会有景色、住宿、美食和交通等多个主题;如果使用传统主题模型进行过滤,可能会把具有语义关联的句子筛选掉,这样会影响后续的方面级情感分析工作。因此,需要采用新方法来提高长文本过滤的准确率。对此,文中提出结合单层神经网络和具有两个隐藏层的长短记忆网络的深度网络模型(A-HLSTM)用于长文本过滤任务中,利用其可以进行多主题分类及上下文语义分析的能力,完成评论性文章的过滤。

1 相关研究

传统的文本过滤方法主要分为两大类,基于规则的过滤方法和基于统计的过滤方法^[5]。其中基于规则的过滤方法是专家根据需求设置匹配规则,然后通过实验反馈的信息完善规则^[6]。基于统计的方法利用人工标注的语料库,通过统计经验的理论进行文本过滤^[7]。早期的方法存在很多缺点,例如规则的制定需要依靠很多专家的经验,且有的特征可能考虑的并不全面,另外这些模型计算和语料库的规模成正比,计算量很大,效率很低。

随着机器学习在自然语言中的广泛应用,许多学者将机器学习技术应用于文本过滤。一些学者将此问题看作多分类问题,通常使用支持向量机(SVM)、k近邻分类算法等。文献[8]在涉恐信息文本的研究中将文本过滤视为分类问题,采用k近邻分类算法进行过滤;文献[9]提出基于主题分类的文本过滤方法,筛选出文本内容的最优特征项集合,利用SVM分类技术过滤。文献[1]利用文本分类系统创建特定领域过滤器,减少了手动注释的训练数据量。以上几种方法

都不能体现语义,导致文本相似度计算一直很低,因此,一些学者又提出了基于语义的文本过滤方法。文献[10]提出了多词-贝叶斯分类算法,将词与词之间的关系作为重要参考项,克服了传统分类器对语义分析的忽视;文献[11]提出基于多谓词语义框架文本过滤算法,利用文本依存句法分析进行语义分析,提高了文本过滤的准确率。这些方法虽然对文本过滤的准确率有所提高,但不能发现深层次特征并且算法计算量较大,计算时间很长,导致算法的计算效率很低。

最近,深度学习在自然语言处理方面的巨大成功使得研究人员将其应用于文本过滤,因为它可以从大量数据训练中学习并判别特征,并且可以考虑到整体上下文信息。然而,对于深度学习在评论性文章这类长文本过滤的研究很少,但深度学习在其他自然语言处理中的应用给了笔者很大的启发。众所周知,好的词向量作为输入可以改善神经网络模型,Pennington等提出的GloVe词向量构造了一个全局的词共现矩阵,能够融合文本的全局信息和局部上下文信息^[12]。目前,用于自然语言处理的神经网络模型主要是RNN和LSTM,RNN可以考虑序列的上下文信息,但RNN在训练过程中梯度向量的分量可能会在长序列上指数增长或消失^[13]。LSTM可以解决RNN的梯度问题,但仍然会忘记距离当前序列较远的信息,尤其在处理长文本任务中这个问题更加明显^[14]。为了能存储更远距离的信息,各种模型被用来提高LSTM存储远程信息的能力。例如,文献[15]提出将外部存储器加入LSTM中,但外部存储器矩阵庞大,性能不佳;文献[16]提出基于注意力机制的双向LSTM来处理文档级情感分析任务;文献[17]提出了缓存长短记忆神经网络(CLSTM)模型,引入缓存机制来捕获长文本中的整体语义信息。

综上所述,这些都是基于一层LSTM并对其结构稍加改变。受这些研究的启发,文中提出结合单层神经网络和分层长短记忆网络的深度网络模型用于长文本过滤任务中。利用词语层LSTM网络模型可以得到具有语义的句向量,第二层主题依赖度计算模型、句子层LSTM网络以第一层得到的句向量作为输入,既可以获得句子与各主题类别的依赖度,同时也通过句子层LSTM网络挖掘整个文章中的长距离依赖关系,有效提高长文本过滤的性能。

2 模型

2.1 模型整体框架

评论性长文本过滤问题主要是在评论性文章中将无主题且与其他句子语义关联低的句子过滤掉,文中旨在研究如何能更好地对长文本进行过滤,主要用

LSTM 模型提取词语间的语义形成句向量,然后进行主题判断和语义关联。模型的整体框架如图 1 所示。

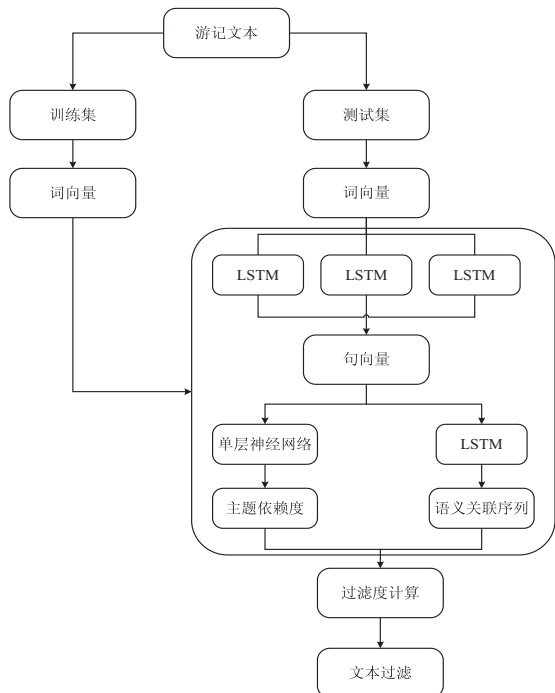


图 1 长文本过滤结构

数据处理的第一步是将预处理后的全部数据进行分词,然后用 GloVe 进行词向量训练,将训练好的词向量作为第一层 LSTM 模块的输入;接着将词向量经过 LSTM 模型进行训练,得到具有语义的句向量,并将此句向量分别作为主题依赖度计算模型和句子层 LSTM 的输入;然后通过主题依赖度模型计算,得到句子于主题类别的概率;经过句子层 LSTM 进一步得到完整的句子语义关系表示;最后综合考虑主题依赖和语义关联两个因素实现长文本过滤。

2.2 词语层 LSTM 层

模型的第一层是 LSTM 模块,该层主要用来接收

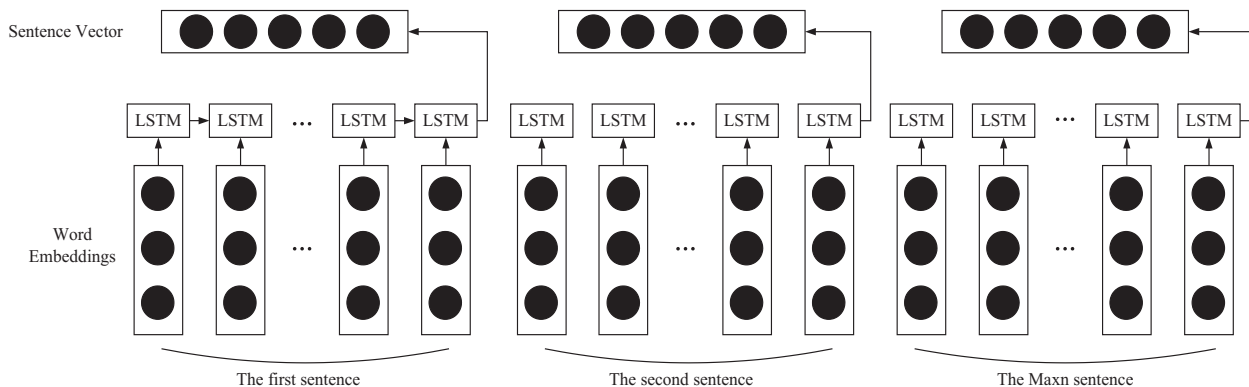


图 3 词语层 LSTM 结构

2.3 主题依赖度计算模块

模型的第二层的第一部分是主题依赖度计算层,用来将句子与其所属主题类别以加权的形式联接,再通过 softmax 函数得到句子对于类别的概率分布,即

最初的词向量数据,将游记文本语料库中的词语用 GloVe 训练得到词向量表示,通过 LSTM 模型训练后得到连续的句向量。如图 2 所示,LSTM 主要包含三个门单元(输入门、输出门、遗忘门)和一个记忆单元。

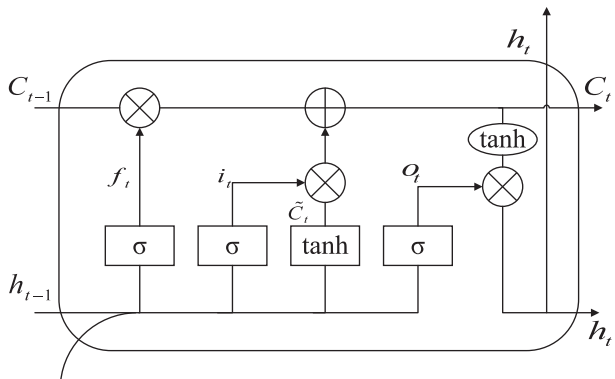


图 2 LSTM 标准结构

一般地,每个 LSTM 单元的计算公式如下:

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

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (2)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (3)$$

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

$$c_t = f_t c_{t-1} + i_t \hat{c}_t \quad (5)$$

$$h_t = o_t \tanh(c_t) \quad (6)$$

其中, σ 表示 sigmoid 激活函数; \tanh 表示双曲正切激活函数; W_f 、 W_i 、 W_o 、 U_f 、 U_i 、 U_o 分别表示输入门、忘记门、输出门的权重矩阵; b_f 、 b_i 、 b_o 表示输入门、忘记门、输出门的偏置向量; h_t 表示 t 时刻的输出。

词语层 LSTM 接收一个以词语为单位的句子作为网络的序列化输入,每个 LSTM 单元的输入由上个单元隐藏层的输出和本次输入的词向量组成,词语层 LSTM 可以得到句子内部词语之间的相互关系。文中提出的词语层 LSTM 网络结构如图 3 所示。

句子的主题依赖度向量。

图 4 所示的主题依赖度计算模型实际上是基于单层神经网络的 softmax 分类器,输入为句向量,输出是句向量对于主题类别的概率。图模型的输入为词语层

LSTM 训练得到的句向量,输出 Y 是一维实向量, Y 的计算公式为

$$Y = W \cdot s_i + b \quad (7)$$

其中, W 是权重矩阵, b 为偏置项。

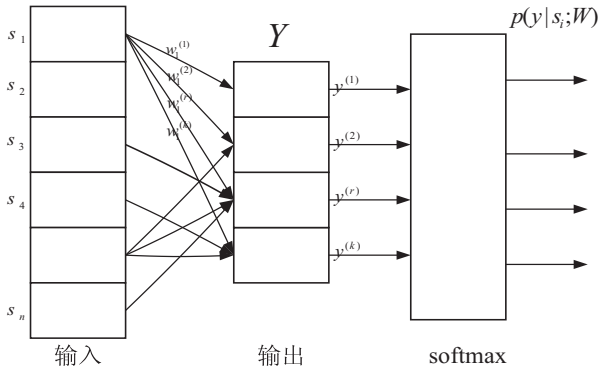


图 4 主题依赖度计算模型

输出 Y 经过 sigmoid 及 softmax 函数,得到属于各类别的概率。softmax 的输出公式为:

$$p(y = r | s_i; W) = \frac{\exp(y^{(r)})}{\sum_{j=1}^k \exp(y^{(j)})} \quad (8)$$

2.4 句子层 LSTM 网络

模型第二层的另一部分是句子层 LSTM 网络,用来对句子关系进行编码。词语层 LSTM 网络可以有效获取句子内部词语之间的关系,但对于评论性文档长距离语义关联的问题,仅仅依靠词语层 LSTM 网络难以正确识别句子间的语义关系,因此,文中使用句子层 LSTM 来进一步挖掘句子之间的依赖关系。将词语层 LSTM 输出的句向量作为句子层 LSTM 的输入,得到的隐藏层输出矩阵作为文档表示,文档表示被用作文档级语义关联的特征,将其馈送到输出长度为关联等级的线性层,并添加 softmax 层输出语义强关联、弱关联、不关联的概率。softmax 函数计算如式 9 所示,其中 C 是语义关联程度划分。

$$\text{softmax}_i = \frac{\exp(x_i)}{\sum_{i=1}^C \exp(x_i)} \quad (9)$$

2.5 过滤度计算

句子的过滤度值综合考虑句子属于主题类别的概率值和语义关联度,过滤度越低,这类句子对后续文本分析任务的贡献度也较低。为了提高后续文本分析任务的效率,因此需要过滤掉过滤度值低的句子。经过前两个模块的训练,已经得到了每个句子主题依赖度和语义关联度,文中提出将主题依赖度向量的均方差与语义关联度之和作为过滤度,对于句子 s ,其过滤度计算如下:

$$g_s = \lambda_1 \frac{1}{k} \sum_{r=1}^k (\text{att}_i[r] - a)^2 + \lambda_2 d_i \quad (10)$$

其中, k 为主题类别数; $\text{att}_i[r]$ 为 s_i 对主题类别 r 的依赖度; a 为主题依赖度均值,为 $1/k$; d_i 为句子语义关联度。

文中设置超参数 threshold 来控制过滤度,在实验中使用交叉验证法来更新 threshold 对比其对过滤效果的影响。

2.6 模型训练

文中通过随机梯度下降进行模型训练,其中损失函数是监督交叉熵误差。要避免出现过度拟合,过度拟合意味着模型将训练数据(包括噪声数据)进行超分割,从而获得最低成本。但是,总体规律会被忽略,对于未知数据,如测试数据,该模型不能很好地执行。为了克服这一问题,文中在所有参数中加入 L_2 正则化,用于限制权重的大小,使得模型不能随机拟合训练数据中的随机噪声。设 y 是待过滤句子的预测类别, z 是待过滤句子的实际类别。训练目标是尽量减少所有训练文本中 y 和 z 之间的交叉熵误差。

$$\text{loss} = - \sum_i \sum_j y_i^j \log z_i^j + \frac{\lambda \|\theta\|}{2} \quad (11)$$

3 实验

3.1 数据处理

通过使用网络爬虫软件在马蜂窝上采集了关于上海的游记 2 000 篇,并使用 Stanford CoreNLP 进行标记和分词,并将数据集分为 80/10/10 用于训练、验证和测试。训练集主要用于训练模型、避免过度拟合,使用验证数据集来进一步确定模型的参数并在不同的参数下评估模型过滤效果,不能根据测试集的结果调整参数。

3.2 实验设置

对于参数配置,使用 GloVe 词向量来初始化实验数据中的词向量,其中每个词向量为 300 维的连续值。对于模型初始化,从均匀分布 $[-0.1, 0.1]$ 之间随机采样初始化所有矩阵,并使用随机梯度下降法来更新所有参数。文中使用 Adagrad 作为优化器,其初始学习率设置为 0.01。

3.3 对比实验

将文中方法和机器学习算法以及典型的机器学习和神经网络模型进行了对比实验:

(1)NB(Naïve Bayesian,朴素贝叶斯):朴素贝叶斯是一种常见的机器学习分类算法,使用词袋模型收集特征。

(2)SVM(Support Vector Machine,支持向量机):文中对文献[9]提出的 SVM 算法加以修改,使用词袋模型收集特征并且使用 LibLinear 训练 SVM 分类器。

(3)RNN:文献[13]提出的 RNN 是对连续文本进

行建模的基本方法。

(4)LSTM:文献[14]提出的 LSTM 是具有存储单元和三个门机制的递归神经网络。

(5)2-layer LSTM:文献[16]提出的 2-layer LSTM 模型中,为了让第二层 LSTM 捕获输入序列的长期依赖关系,第一层 LSTM 单元的隐藏层输出矩阵在同一时间步骤中输入第二层 LSTM 中。

(6)CLSTM:文献[17]提出的 CLSTM 旨在通过缓存机制捕获远程信息,它将存储器分为若干组,并且不同的遗忘率(过滤器)分成不同的组。

3.4 实验结果与分析

3.4.1 不同模型对比实验结果分析

文中使用准确度和 MSE(mean square error,均方误差)来评估模型,其中准确度是衡量文本过滤的标准指标。MSE 是一种测量平均误差的便捷方法。由此,通过评估数据的变化度,MSE 值越小,表明实验模型的可靠性越高。

$$MSE = \frac{\sum_{j=1}^N (\text{standard}_j - \text{predicted}_j)^2}{N} \quad (12)$$

文中对比了不同模型下长文本过滤的准确度和 MSE,结果如表 1 所示。

表 1 不同模型下长文本过滤的准确度和 MSE

模型	Acc/%	MSE
NB	35.3	4.36
SVM	40.4	3.54
RNN	23.2	5.82
LSTM	39.8	3.21
2-layer LSTM	40.1	3.18
A-HLSTM	46.3	2.13

从表 1 可以发现:

(1)对比了两种机器学习算法(NB 和 SVM),可以发现 SVM 比 NB 有更好的过滤效果。机器学习方法几乎能达到 LSTM 相同的效果,但它需要大量的特征工程。标记有效特征是一项非常基础的工作,机器学习分类器的性能很大程度上取决于数据表示和特征的选择,但神经网络模型可以根据数据的特征自动学习,这是它被广泛应用的原因。

(2)在循环神经网络中,由于梯度消失问题,RNN 在长文本建模方面表现最差。相比而言,LSTM 有更好的性能,这表明内部存储器和三个门的结构在长文本建模中的作用很关键。

(3)提出的 A-HLSTM 深度分层网络模型具有最佳性能,比 Bi-LSTM 提升了 1.4%。

(4)在双向体系结构中,长文本模型可以向前和向后捕获特征,因此,Bi-LSTM 比单向模型具有更好

的性能。在双向模型中,文中模型具有良好的性能,准确度达到 46.3%。

(5)在时间复杂度和参数数量方面,A-HLSTM 和 2-layer LSTM 都有两个隐藏层,但 A-HLSTM 比 2-layer LSTM 需要更少的计算资源,却达到了更高的准确率。与完全连通层比,该模型仅使用第一层输出的句子向量作为第二层的输入,因此该模型具有较少的参数和计算时间。

3.4.2 threshold 值对文本过滤的影响

图 5 是用文中算法进行过滤后的文本长度、句子数占过滤前文本的比例随 threshold 的变化情况。实验结果表明,当 threshold 为 3.1×10^{-4} 时,没有任何句子被过滤;当 threshold 为 5.0×10^{-4} 时,过滤后的文本长度是过滤前的 42.13%,过滤后的句子数是过滤前的 42.49%,并且文本长度变化和句子数变化趋势基本一致,说明文中算法过滤质量较好,没有出现集中过滤短句或者长句的现象。

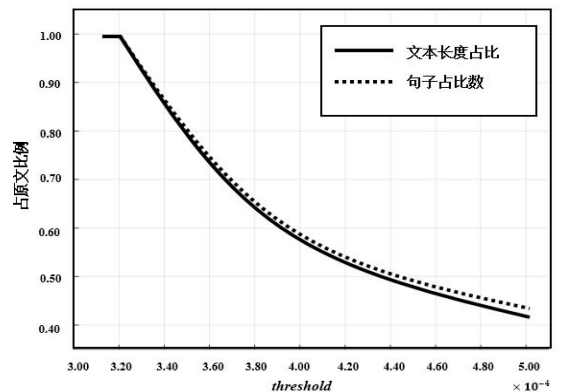


图 5 threshold 值对文本过滤的影响

3.4.3 词向量的影响

众所周知,神经网络的输入是词向量,词向量的选择对优秀的文档表示至关重要。为了了解不同词向量对模型的影响,文中选择随机初始化向量,word2vec 模型(CBOW 和 Skipgram)和 GloVe 在两种模型(LSTM 和 A-HLSTM)上做了对比实验。所有词向量都是 300 维,结果见表 2

表 2 不同词向量下 LSTM,A-HLSTM 的文本过滤准确率

模型	Random	CBOW	Skip-gram	GloVe
LSTM	0.343	0.382	0.375	0.398
A-HLSTM	0.398	0.432	0.438	0.440

从表 2 中可以发现,word2vec 和 GloVe 比随机初始化向量表现更优。这表明上下文信息对词向量学习的重要性。此外,还可以看出 GloVe 在这两个模型上的准确度略有提高,这充分说明一个好的词向量需要考虑全局上下文信息。

文中还对比了不同维度的 GloVe 向量 (50/100/200/300)。表 3 和表 4 分别给出文本过滤准确度和时间成本,可以发现 200 维词向量比 50 和 100 维表现更好,而 300 维词向量没有显著改进。此外,A-HLSTM 比 LSTM 花费更多的时间,因为 A-HLSTM 的参数数量更多,但它们有更高的准确率。

表 3 不同维度 GloVe 词向量下 LSTM、A-HLSTM 的文本过滤准确率

模型	50 d	100 d	200 d	300 d
LSTM	0.358	0.377	0.391	0.398
A-HLSTM	0.405	0.422	0.436	0.440

表 4 不同维度 GloVe 词向量下 LSTM、A-HLSTM 的训练时间成本(单位:分钟)

模型	50 d	100 d	200 d	300 d
LSTM	7.8	23.2	48.9	113.1
A-HLSTM	12.5	37.2	72.0	124.2

4 结束语

社交网络上存在大量的分享个人经验的长文本,如游记等,这些长文本与专业文献不同,主题类别多且语义间的关联性强。因此,为了更好地进行下一步的方面级情感分析工作,先对其进行文本过滤。首先通过词语层 LSTM 网络获得句子内部词语之间的关系并得到具有语义的句向量,然后将句向量输入主题依赖度计算模型和句子层 LSTM 网络模型,进而得到句子与各主题类别的依赖度以及待过滤句子与其他句子之间的关联。最后在游记数据集上进行了实验,实验结果验证了文中模型的有效性。

参考文献:

[1] 邓一贵,伍玉英. 基于文本内容的敏感词决策树信息过滤算法[J]. 计算机工程,2014,40(9):300-304.

[2] 黄萱菁,夏迎炬,吴立德. 基于向量空间模型的文本过滤系统[J]. 软件学报,2003,14(3):435-442.

[3] 薛 玥. 基于相关主题建模的短文本过滤引擎研究与应用[D]. 北京:北京邮电大学,2018.

[4] ZHOU P,SHI W,TIAN J,et al. Attention-based bidirectional long short-term memory networks for relation classification[C]//Proceedings of 54th annual meeting of the association for computational linguistics. Berlin, Germany:[s. n.],

2016:207-212.

[5] 刘淑梅,桑书娟,山 岚. 文本过滤研究概述[J]. 信息技术,2009,33(9):187-189.

[6] 林鸿飞,战学刚,姚天顺. 文本结构分析与基于示例的文本过滤[J]. 小型微型计算机系统,2000,21(4):422-425.

[7] 杨彬彬. 基于多谓词语义框架的网络文本过滤技术研究[D]. 北京:北方工业大学,2018.

[8] 唐 华,杨解君,王 俊,等. 基于 K 近邻分类算法的涉恐信息过滤模型研究[J]. 情报杂志,2018,37(3):64-70.

[9] 刘梅彦,黄改娟. 面向信息内容安全的文本过滤模型研究[J]. 中文信息学报,2017,31(2):126-131.

[10] SCHMIDT S,SCHNITZER S,RENSING C. Text classification based filters for a domain-specific search engine[J]. Computers in Industry,2016,78(C):70-79.

[11] 王景中,杨彬彬,何云华. 基于多谓词语义框架的长短视频文本相似度计算[J]. 计算机工程与设计,2018,39(4):1022-1028.

[12] PENNINGTON J,SOCHER R,MANNING C. Glove: global vectors for word representation [C]//Proceedings of the 2014 conference on empirical methods in natural language processing. Stroudsburg, PA: ACL,2014:1532-1543.

[13] BENGIO Y,SIMARD P,FRASCONI P. Learning long-term dependencies with gradient descent is difficult [J]. IEEE Transactions on Neural Networks,1994,5(2):157-166.

[14] BHATIA P,JI Y,EISENSTEIN J. Better document-level sentiment analysis from RST discourse parsing [C]//Proceedings of conference on empirical methods in natural language processing. Lisbon, Portugal: [s. n.], 2015: 2212 - 2218.

[15] TRAN K,BISAZZA A,MONZ C. Recurrent memory networks for language modeling [C]//Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics; human language technologies. [s. l.]:[s. n.],2016:321-331.

[16] TANG D,QIN B,FENG X,et al. Effective LSTMs for target-dependent sentiment classification [C]//Proceedings of the 26th international conference on computational linguistics. Madison, Wisconsin, USA:[s. n.],2016:3298-3307.

[17] XU Jiacheng,CHEN Danlu, QIU Xipeng, et al. Cached long short-term memory neural networks for document-level sentiment classification [C]//Proceedings of the 2016 conference on empirical methods in natural language processing. Austin, USA:[s. n.],2016:1660-1669.