

用于情感分类的双向深度 LSTM

曾 蒸¹ 李 莉² 陈 晶³

(重庆师范大学新闻与传媒学院 重庆 401331)¹ (西南大学计算机与信息科学学院 重庆 400715)²

(中冶赛迪工程技术股份有限公司 BIM 中心 重庆 401122)³

摘 要 对商品、电影等的评论的体现人们对商品的喜好程度,从而为意向购买该商品的人提供参考,也有助于商家调整橱窗货品以取得最大利润。近年来,深度学习在文本上强大的表示和学习能力为理解文本语义、抓取文本所蕴含的情感倾向提供了极好的支持,特别是深度学习中的长短记忆模型(Long Short-Term Memory, LSTM)。评论是一种时序数据形式,通过单词前向排列来表达语义信息。而 LSTM 恰好是时序模型,可以前向读取评论,并把它编码到一个实数向量中,该向量隐含了评论的潜在语义,可以被计算机存储和处理。利用两个 LSTM 模型分别从前、后两个方向读取评论,从而获取评论的双向语义信息;再通过层叠多层双向 LSTM 来达到获取评论深层特征的目的;最后把这个模型放到一个情感分类模型中,以实现情感分类任务。实验证明,该模型相对基准 LSTM 取得了更好的实验效果,这表示双向深度 LSTM 能抓取更准确的文本信息。将双向深度 LSTM 模型和卷积神经网络(Convolutional Neural Network, CNN)进行实验对比,结果表明双向深度 LSTM 模型同样取得了更好的效果。

关键词 LSTM, 深度学习, 情感分类

中图法分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.08.038

Deeply Hierarchical Bi-directional LSTM for Sentiment Classification

ZENG Zheng¹ LI Li² CHEN Jing³

(College of Journalism and Communication, Chongqing Normal University, Chongqing 401331, China)¹

(College of Computer and Information Science, Southwest University, Chongqing 400715, China)²

(BIM Center, CSDI Engineering Co., LTD., Chongqing 401122, China)³

Abstract The comments on goods, films and others contribute to assess people's preference degree for goods, which provides reference for the people who intend to buy the goods, and can help businesses adjust shelves to maximize profits. In recent years, the powerful representation and learning ability in deep learning technologies provides a good support for understanding text semantics and grasping the emotional tendency of texts, especially the long short-term memory (LSTM) model in deep learning. The comment is a form of temporal data, which expresses semantic information through the forward arrangement of words. LSTM is a sequential model that reads the comment forward and encodes it into a real vector, and this vector implies the potential semantics of the comment and can be stored and processed by the computer. In this paper, two LSTM models are utilized to read comments from forward and backward directions respectively, and thus the two-way semantic information of the review can be obtained. Then the purpose of obtaining the deep features of comments is achieved by stacking the multilayer bidirectional LSTM. Finally, the model is put into a sentiment classification model to implement the sentiment classification. Experimental results show that the proposed method outperforms baseline LSTM, which means that deeply hierarchical bi-directional LSTM (DHBL) can capture more accurate text information. Compared with the convolutional neural network (CNN) model, the proposed model also achieves better effect.

Keywords LSTM, Deep learning, Sentiment classification

1 引言

当前,各种用户对商品的评论或者对事务的喜好评论盛行,例如博客论坛、tweets、微博、互联网广告、电子商务网站

上的商品评论等。阅读这些评论并且从中获取有用信息已经成为人们日常生活中不可或缺的一部分。这些评论表达了用户对商品的喜好和关注程度。挖掘这些评论所表达的用户情感可以实现对用户与商品关系的建模,帮助用户获得个性化

收稿日期:2018-05-02 返修日期:2018-06-09 本文受国家自然科学基金项目(61170192),重庆市科委重点项目(cstc2017zdcy-zdyf0366),重庆市教委项目(113143)资助。

曾 蒸(1972—),女,硕士,工程师,主要研究方向为大数据与深度学习、智能教育技术、计算机网络及应用、新媒体运营等,E-mail:akk310@qq.com(通信作者);李 莉(1967—),女,博士,教授,博士生导师,主要研究方向为机器学习、数据挖掘与分析、服务计算、智能教育,E-mail:lily@swu.edu.cn。

的服务,也有助于商家及时调整货架商品以提升销售额。近年来,情感分类研究已经得到了越来越多研究者的关注,并且取得了突出成绩^[1-2],例如潜在购买用户会在购买前查阅评论信息,并将其作为是否购买对应商品的重要参考依据^[3]。

传统的情感分类模型基本采用词袋模型结构来抓取文本中符号(单词、字符等)的统计信息,然后通过计算机容易处理的向量形式表示出来,最后通过回归模型实现情感分类。这类模型实现简单,效果显著。LDA 主题模型是一种常用于情感分类的词袋模型,它依据指定的情感单元搭配模式提取情感信息,包括情感词和上下文;使用主题模型发掘情感信息中的关键特征,并将其融入到情感向量空间中,最后利用机器学习分类算法实现中文评论文本的情感分类^[4-6]。在文献^[7-8]中,主题特征被谨慎选择以提升实验效果。此外,还有 n -gram 模型,将 n 个连续的单词构成统计信息的最小单元,并统计这些单元以实现文本表示^[9]。

文本中隐含的信息除了单词的统计信息外,还有很多其他信息,特别是单词之间的词条依存关系(Term Dependencies),然而词袋模型只能抓取到有限且单薄的信息来源,或者仅仅考虑了临近单词之间的关联关系^[10-11]。单词之间的词条依存关系(词间关系)是文本表示中的一项重要特征,极大地影响了文本所隐含的语义信息。例如,文本“小明喜欢穿新衣服的小丽”和“穿新衣服的小明喜欢小丽”由相同的单词词汇表构成并且单词的统计信息一致,单词统计信息都为{小明:1,喜欢:1,穿:1,新衣服:1,的:1,小丽:1}。如果使用词袋模型来处理这两段文本,则表示它们的语义向量相同,并且会在计算机中把它们识别为相同的语义。事实上,这两段文本的语义完全不一致,其原因就是单词之间的顺序发生了变化,词间关系发生了偏移。尽管通过引入人工先验知识(例如语法和句法分析规则)可以区分这两段文本的不同语义,但是仍然可以用更好的方法来解决这一问题,即通过引入词间关系的处理,使计算机自动抽取词间关系而无需先验知识。

近年来,深度学习强大的表示学习能力为文本表示学习注入了强大的活力。深度学习中的卷积神经网络(Convolutional Neural Network, CNN)可以学习文本的静态特征,获取文本中的局部特征信息(Regional Information),也即临近单词之间的关联信息。通过这些信息来表示文本特征,并实现情感分类^[12]。深度学习中的循环神经网络(Recurrent Neural Network, RNN)在局部特征信息抓取的基础上还可以抓取到更久远的特征信息,从而实现全局特征信息(Global Information)的抓取。RNN 也是一种序列模型,可以顺序“读取”文本中的单词,抓取单词之间的“顺序”特征,也就是词间关系。RNN 可以“记忆”并存储文本中的上下文全局信息,这种“记忆”并不是“记忆”临近几个单词的局部信息,而是“记忆”整个文本的长时全局信息。

本文将利用深度学习中的 LSTM 模型(RNN 模型的升级,被广泛应用于各种文本表示中),从文本的两个方向来抓取词间关系,也即单词之间的顺序和反序特征关系,从而丰富文本的特征表示。其次,引入多层 LSTM,实现文本深层特征的挖掘。最后,把提出的模型引入到一个情感分类回归模型中,最终完成情感分类任务。实验表明,本文方法不仅优于传统的词袋模型,而且优于基准 LSTM 模型和 CNN 模型。在

实验中,我们使用了两个情感分类数据集:1)电影评论 IM-DB,有 5 万条二分类电影评论;2)SemEval2016 任务 4 的数据集,拥有 6 千条 tweets 评论,是一个三分类数据集。

2 相关工作

本文模型是基于 LSTM 进行的改进。LSTM 是 RNN 的变种,拥有比 RNN 更强的文本表示能力。

RNN 模型由 3 层神经网络构成,分别是输入层、隐藏层和输出层。其中,隐藏层是一个循环神经网络层,其按顺序依次读取并处理文本中的单词,然后把所有单词的信息合并输出。这个过程隐含地保留了单词之间的顺序关系。隐藏层类似一个记忆单元,可以“记住”文本中的历史信息,使上下文单词时序地通过并被保存下来。RNN 已经被广泛应用于各类自然语言任务,并在其中进行文本表示。例如,文献^[13-14]通过读取文本中的历史上下文来预测下一个单词的具体内容。与前馈神经网络相比,RNN 拥有编码和记忆历史信息的能力,因此更适用于文本表示类的序列化数据处理(时序处理)。在前馈神经网络中较为成功的 BP 训练算法在 RNN 中已经不太适用。通过改进 BP 得到了时序 BP(Backpropagation Through Time Algorithm, BPTT)和实时循环学习算法(Real-Time Recurrent Learning algorithm, RTRL)^[15]。BPTT 和 RTRL 更适用于训练 RNN 一类的时序模型。然而,RNN 也存在一些问题,例如它对长距离历史信息的把控不足。在利用 BPTT 和 RTRL 训练 RNN 模型时,经过多层循环之后,错误信号发生了很大的衰减,从而无法产生有效的错误信号来更新模型的权重参数,这就是梯度消失^[16]。梯度消失将导致长距离的词间关系无法被学习到^[17]。

下面给出梯度消失的数学抽象过程。隐藏层的第 j 个神经元在 $t-1$ 时刻的错误信号可以公式化为:

$$\vartheta_j(t-1) = \delta_j'(net_j(t-1)) \sum_i w_{ij} \vartheta_i(t)$$

其中, ϑ 是错误信号, $\delta'(\cdot)$ 激活函数的偏导数, $net(\cdot)$ 是净输入, w_{ij} 是隐藏层中上一时刻的第 i 个神经元到当前时刻的第 j 个神经元的权值参数。因此,第 $t-q$ 时刻的错误信号可以公式化为:

$$\frac{\partial \vartheta_j(t-q)}{\partial \vartheta_i(t)} \begin{cases} \delta_j'(net_j(t-1)) w_{ij}, & q=1 \\ \delta_j'(net_j(t-q)) \sum_{l=1}^q \frac{\partial \vartheta_l(t-q+l)}{\partial \vartheta_i(t)} w_{lj}, & q>1 \end{cases}$$

将上述公式展开可得:

$$\frac{\partial \vartheta_j(t-q)}{\partial \vartheta_i(t)} = \sum_{l_1} \dots \sum_{l_{q-1}} \prod_{m=1}^q \delta'_{l_m}(net_{l_m}(t-m)) w_{l_m l_{m-1}}$$

其中:

$$T = \prod_{m=1}^q \delta'_{l_m}(net_{l_m}(t-m)) w_{l_m l_{m-1}}$$

如果 $|T| > 1$,则随着 q 的增长,误差会呈指数级增长;如果 $|T| < 1$,则误差随着 q 的增长会急剧衰减。在工程实践中, $|T| < 1$ 的情况较普遍,也就是误差梯度消失现象较普遍。

为了解决这一问题,Hochreiter^[18]提出一种 RNN 的改进模型——LSTM。该模型额外引入了一个记忆单元(Constant Error Carrousel, CEC)。CEC 是一个自我连接的单元,能够保留长久的错误信号,这使得 LSTM 天生就可以编码距离久远的上下文历史信息。图 1 是 LSTM 的结构图^[19],其中各种符号的详细描述见表 1,其中下标 t 和 $t-1$ 表示当前时刻和

上一时刻(例如, f_t 是当前时刻的遗忘门, f_{t-1} 是上一时刻的遗忘门)。在 LSTM 模型中, C 就好比是一个历史信息传送带。文本中的历史上下文信息沿着这个传送带流动,不断地将当前读取到的单词补充进来。 f, i 和 o 分别是遗忘门、输入门和输出门。历史信息通过遗忘门后发生衰减,从而给新的信息(当前输入的单词)腾出存储空间;而新的信息(当前输入的单词)通过输入门抑制后进入历史信息传送带中;通过输出门,模型输出当前的历史信息编码。3 个门电路均采用压缩输出,均在 $[0, 1]$ 范围内取值。 C 是模型的核心,它保存了长久的历史信息,从而保存了长距离词间的关系。以下公式从数学上将这一过程进行抽象^[19]。

$$\begin{aligned} f_t &= \delta(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t &= \delta(W_i \cdot [h_{t-1}, x_t] + b_i) \\ C'_t &= \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\ C_t &= f_t * C_{t-1} + i_t * C'_t \\ h_t &= o_t * \tanh(C_t) \end{aligned}$$

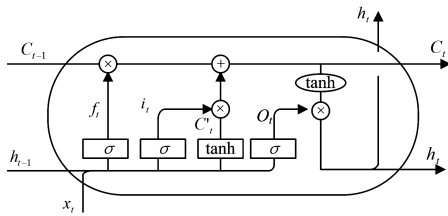


图 1 LSTM 模型
Fig. 1 LSTM model

表 1 LSTM 的符号说明

Table 1 Symbol description of LSTM

符号	说明
w_f, w_i, w_o	遗忘门、输入门和输出门的权值参数矩阵
w_f, w_h	输入的权值参数矩阵
b_f, b_i, b_o, b_c, b_h	各权值矩阵的偏移值
h	输出
δ	sigmoid 激活函数
$\tanh$	反正切激活函数
x	输入
f, i, o	遗忘门、输入门和输出门
c	记忆单元 (CEC)
c', h'	更新后的目标值
$\phi_j(t)$	第 j 个神经元在 t 时刻的参数值
$net_j(t)$	在输入层, 第 j 个神经元在 t 时刻的净输入

3 双向深度 LSTM 模型

为了抓取文本的语义信息,实现评论的情感语义表示,提出了一种基于 LSTM 的双向深度文本表示模型。在该模型的基础上,构建了一个文本分类模型以实现情感分类任务。

3.1 双向深度文本表示模型

本文提出了一种双向深度文本表示模型(Deeply Hierarchical Bi-directional LSTM, DHBL)。

传统的序列模型只能朝一个方向读取文本,一般是从前往后读取文本中的单词。这种前向读取方式可以获取文本中单词之间的前向关联关系,而且这种方式也符合人们阅读文本的习惯。但是,自然语言经过几千年的发展已变得十分复杂,除了前面的文本为后面的单词提供支撑以外,后面的文本也对前面的单词提供解读,即文本中单词之间的关联关系从前后两个方向都有。本文使用两个 LSTM 分别从文本的两

个方向读取单词,然后把各自编码得到的文本表示合并起来,如此便得到一个经过双向编码文本的表示。双向 LSTM 结构如图 2 所示,其中 $\overleftarrow{\text{LSTM}}$ 是从前往后读取文本的 LSTM 模型, $\overrightarrow{\text{LSTM}}$ 是从后往前读取文本的 LSTM 模型,将两个模型各自得到的前向和后向文本表示联合起来,就得到了最终的文本表示。双向信息流模型也即双向 LSTM (bi-directional LSTM, biLSTM),在文本表示领域取得了重大成功^[20]。

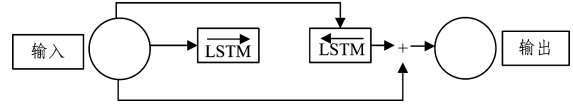


图 2 双向 LSTM

Fig. 2 Bi-directional LSTM

为了实现文本的深层次特征挖掘,我们把多个 LSTM 层叠起来,通过多层神经网络结构来挖掘文本的深层特征。但当神经网络的层数过多时,会出现梯度误差消失的问题,从而导致模型无法训练。第 2 节已经给出了梯度误差消失的原因,对此本文引入 CNN 模型中使用的网中网概念(Network In Network, NIN)^[21]和防止梯度误差消失的残差网络^[22]。具体来说,首先构建有限的三层 biLSTM,这种三层 biLSTM 是一个 NIN 模块。由于神经网络结构的层数不高,因此可以避免梯度误差消失的问题;为了挖掘深层文本特征模型,需要加深神经网络层数,因此我们把多个 NIN 模块层叠起来,以解决由此带来的梯度误差消失问题;另外,在 NIN 模块之间引入残差网络连接,模型结构如图 3 所示。

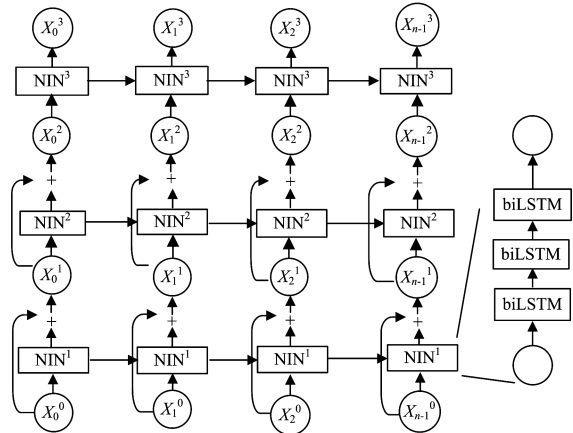


图 3 双向深度文本表示模型 DHBL

Fig. 3 Deeply hierarchical bi-directional LSTM(DHBL)

图 3 右侧部分展示了一个 NIN 模块,该模块由三层 biLSTM 层叠构成。图 3 中的弧线是残差网络连接,误差可以通过弧线反向无损失地传回上一层 NIN 模块中,从而避免产生误差梯度消失的问题。 $x^0 = \{x_0^0, x_1^0, x_2^0, \dots, x_{n-1}^0\}$ 是由 n 个单词构成的一段文本,其作为输入进入模型中,其中 $x_i^0 (i=0, 1, 2, \dots, n-1)$ 是这段文本的第 i 个单词。 x^1, x^2 和 x^3 分别是第一层、第二层和第三层 NIN 模块的输出,并分别作为下一层 NIN 模块的输入。模型中,残差网络在 LSTM 中的作用可以通过下面的数学公式抽象出来:

$$\begin{aligned} c_t^i, h_t^i &= biKWAM^i(c_{t-1}^i, h_{t-1}^i, x_t^{i-1}; \theta^i) \\ x_t^i &= x_t^{i-1} + h_t^i \\ c_t^{i+1}, h_t^{i+1} &= biKWAM^{i+1}(c_t^{i+1}, h_t^{i+1}, x_t^i; \theta^{i+1}) \end{aligned}$$

其中, c_t^i 和 h_t^i 分别是记忆单元和输出单元,来自于 NIN 模块

中的最后一层 biLSTM,其上标 i 表示第 i 层 NIN 模块,下标 t 表示时刻 t ,也就是第 t 个单词输入。 x_t^i 是第 i 层 NIN 在第 t 时刻的输入, θ^i 是第 i 层 NIN 的参数集合。残差网络可以帮助训练层数较多的神经网络,但是过深的神经网络不仅会耗费较长的训练时间,而且其效果随着层数的不断增加而停滞。因此,本文采用 5 层 NIN 模块。

3.2 情感分类模型

为了完成情感分类任务,本文将 3.1 节中提出的双向深度文本表示模型 DHBL 引入到一个情感分类模型中,如图 4 所示。 $x^0 = \{x_0^0, x_1^0, x_2^0, \dots, x_{n-1}^0\}$ 是由 n 个单词构成的一段文本,其作为输入进入模型中,其中 $x_i^0 (i=0, 1, 2, \dots, n-1)$ 是这段文本的第 i 个单词。各单词采用热点表示方式,单词表示的长度即词汇表的大小。每个热点表示的单词通过 Embedding 层之后编码为词嵌入表示,这是一种低维的词向量表示方式,本文采用 128 维低维词向量。DHBL 将文本中的当前单词与历史单词编码到向量 $h_i (i=0, 1, \dots, n-1)$ 中,然后得到一个历史文本序列 $\{h_0, h_1, \dots, h_{n-1}\}$ 。Max Pooling 层抓取所有历史文本向量中每一维度的最高值作为显著特征向量,并以此构成文本的显著特征表示向量 h_0 。最后,通过一个逻辑回归层对显著特征表示向量进行回归分类。

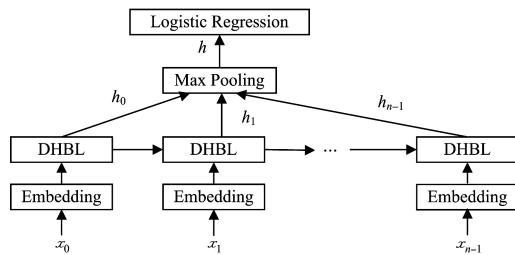


图 4 基于 DHBL 的情感分类模型

Fig. 4 Emotional classification model based on DHBL

为了使模型更加健壮,防止模型训练时出现过拟合的问题^[23],我们在情感分类模型中引入 Dropout 方法^[24],使 Embedding 层中随机的 50% 的神经元不发挥作用。具体来说,模型训练时,当误差反向传递到 Embedding 层时,其中随机的 50% 的神经元的权值参数会随着误差发生变化,而另外 50% 的神经元的权值参数不变化。也就是说,Embedding 层中总会有两个神经元不会在同一个训练循环(epoch)中同时出现。因此,任意两个神经元之间的固有关系不会对参数更新造成约束。Dropout 方法增强了模型的可扩展能力,避免了训练集和测试集中数据分布不完全一致所带来的过拟合问题。

4 实验与分析

4.1 实验设置

本文使用两个情感分类数据集来评估模型:1)拥有 5 万数据量的大型数据集,二分类(积极和消极);2)拥有 6 千数据量的小型数据集,三分类(积极、中性和消极)。

1)IMDB^[25]:一个大型的电影评论数据集,每条评论有积极和消极两方面的情感标签,共有 5 万条数据。本文根据评论文本的长度将其划分为 4 个子数据集,从而可以评价模型在短文本(少于 140 个单词的文本)和长文本上的效果。4 个子数据集如下:①少于(包含)140 个单词的评论数据集;②141~200 个单词之间的评论数据集;③201~400 个单词之间的评论数据集;④多于 400 个单词的评论数据集。此外,我

们还在所有长度的文本数据集上进行实验。

2)SemEval-2016 Task 4^[26]:该数据集包含 6 千条 tweets,每条 tweet 有积极、中性和消极三分类标签。所有数据均是长度短于 140 个单词的短文本。

为了对比不同模型的效果,在以下不同种类的模型中进行实验。

1)lda1-svd;Latent Dirichlet Allocation model (LDA)^[27]模型将单词编码到 30 维的向量空间中,并提取其中的主题词,然后用 support vector machine (SVM)进行分类。

2)lda2-svm;LDA 模型将单词编码到 100 维的向量空间中,并提取其中的主题词,然后使用 SVM 进行分类。

3)lda3-svm;LDA 模型将单词编码到 200 维的向量空间中,并提取其中的主题词,然后使用 SVM 进行分类。

4)word2vec-svm;Word2vec^[28]使用 100 维的实数向量来编码所有单词并计算单词的词向量,然后使用 SVM 进行分类。

5)cLSTM;使用第 3 节所介绍的情感分类模型,其中文本的语义表示采用基准 LSTM 模型来进行处理,模型中采用 128 维向量来表示单词。

6)cDHBL;使用第 3 节所介绍的情感分类模型,其中文本的语义表示采用 DHBL 模型来进行处理,模型中采用 128 维向量来表示单词。

7)Kim-CNN^[12]:使用 Kim 等^[12]提出的 CNN 变种模型来进行文本表示处理。CNN 变种模型有一层卷积层,128 组变长且步长为 1 的卷积核,每组变长卷积核有 3 个长度分别为 3,4 和 5 的卷积核。然后把这个 CNN 变种模型引入第 3.2 节中的情感分类模型中来实现情感分类任务。

4.2 实验结果与分析

首先,分别在两个数据集上对这 7 个模型进行实验,得到的分类准确率如表 2 所列。为了评价训练误差,展现所提模型相比基准 LSTM 模型的改进效果,我们计算基于 LSTM 的模型(cLSTM 和 cDHBL)的误差,结果如表 3 所列。最后,将本文模型和基准 LSTM 模型分别在 5 种优化训练算法上进行实验对比(结果见图 5 和图 6),以展现本文改进的通用性。5 种优化算法为:Adagrad^[29],Adadelta^[30],Gradient (GD) Descent, Momentum 和 Adam^[31]。

如表 2 所列,在情感分类的准确度上,深度学习模型(cLSTM, cDHBL 和 Kim-CNN)远远优于基于 SVM 的模型;并且深度学习模型在短文本(少于 140 个单词的数据)上的效果稳定,而基于 SVM 的模型在短文本上的效果急剧下降。在不同数据集上,本文模型比基准模型效果更好,表明了本文的改进模型具有显著效果。表 3 的实验结果也验证了这一点。

表 2 不同长度文本的情感分类准确度

Table 2 Emotional classification accuracy of texts with different lengths

模型	数据集					SemEval-2016
	≤140	≤200	≤400	>400	所有	
lda1-svm	0.7857	0.8049	0.8169	0.8168	0.8008	—
lda2-svm	0.7783	0.8098	0.8252	0.8191	0.8088	—
lda3-svm	0.7739	0.8074	0.8161	0.8084	0.8089	—
word2vec-svm	0.8020	0.8260	0.8470	0.8220	0.8280	—
cLSTM	0.8653	0.8D647	0.8643	0.8648	0.8649	0.7085
cDHBL	0.8895	0.8914	0.8899	0.8911	0.8895	0.7312
Kim-CNN	0.8868	0.8873	0.8854	0.8868	0.8868	—

表 3 不同长度文本的情感分类误差

Table 3 Emotional classification errors of texts with different lengths

模 型	数 据 集					SemEval-2016
	IMDB					
	≤140	≤200	≤400	>400	所有	
cLSTM	0.4572	0.4563	0.4567	0.4566	0.4566	0.5911
cDHBL	0.4439	0.4433	0.4441	0.4437	0.4435	0.5785

图 5 和图 6 中,Adadelta 和 Adam 的效果最好,在所有优化算法中,本文的改进模型的实验效果都显著优于基准 LSTM 模型。

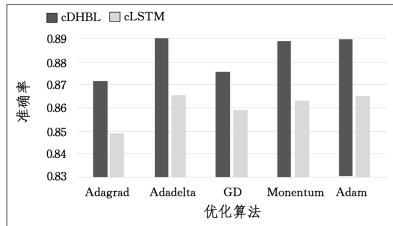


图 5 不同优化算法在 IMDB 数据集上的准确率

Fig. 5 Accuracy of different optimization algorithms on IMDB dataset

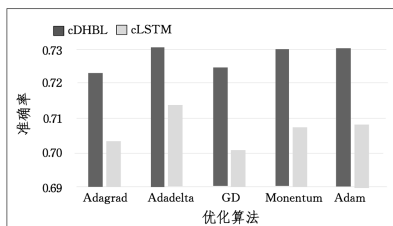


图 6 不同优化算法在 SemEval-2016 数据集上的准确率

Fig. 6 Accuracy of different optimization algorithms on SemEval-2016 dataset

结束语 本文提出了一种双向深度文本表示模型,并将其用于情感分类任务中。该模型解决了从前、后两个方向同时抓取文本中单词间关联关系的问题,并将深层神经网络用于文本深度特征的抓取。实验结果证明了本文方法在情感分类任务中优于基准模型。

参 考 文 献

- [1] IRSOY O, CARDIE C. Opinion Mining with Deep Recurrent Neural Networks[C]// Conference on Empirical Methods in Natural Language Processing, 2014:720-728.
- [2] ZHOU X, WAN X, XIAO J. CMiner: Opinion Extraction and Summarization for Chinese Microblogs[J]. IEEE Transactions on Knowledge and Data Engineering, 2016, 28(7):1650-1663.
- [3] WEI J, RUAN H, LI Z. Analysis of economic impact of online reviews: an approach for market-driven requirements evolution [M]// Requirements. Springer Berlin Heidelberg: Engineering, 2014:45-59.
- [4] SONG G, YE Y, DU X, et al. Short text classification: A survey [J]. Journal of Multimedia, 2014, 9(5):635-643.
- [5] WANG B K, HUANG Y F, YANG W X, et al. Short text classification based on strong feature thesaurus[J]. Frontiers of Information Technology & Electronic Engineering, 2012, 13(9):649-659.
- [6] KIM K, CHUNG B S, CHOI Y, et al. Language independent semantic kernels for short-text classification[J]. Expert Systems with Applications, 2014, 41(2):735-743.
- [7] WANG M, LIN L, WANG F. Improving Short Text Classification through Better Feature Space Selection[C]// International Conference on Computational Intelligence and Security. IEEE, 2014:120-124.
- [8] FAN X, HU H. Construction of High-quality Feature Extension Mode Library for Chinese Short-text Classification[C]// WASE International Conference on Information Engineering. IEEE, 2010:87-90.
- [9] ZHANG X, WU B. ShortText Classification based on feature extension using The N-Gram model[C]// International Conference on Fuzzy Systems & Knowledge Discovery. IEEE, 2016:710-716.
- [10] HUANG P S, HE X, GAO J, et al. Learning deep structured semantic models for web search using clickthrough data[C]// Acm International Conference on Conference on Information & Knowledge Management. ACM, 2013:2333-2338.
- [11] SHEN Y, HE X, GAO J, et al. Alateral semantic model with convolutional-pooling structure for information retrieval[C]// International Conference on Conference on Information and Knowledge Management. ACM, 2014:101-110.
- [12] KIM Y. Convolutional neural networks for sentence classification[J]. arXiv preprint arXiv:1408.5882, 2014.
- [13] MIKOLOV T, KARAFIÁT M, BURGET L, et al. Recurrent neural network based language model[C]// INTERSPEECH 2010, Conference of the International Speech Communication Association. DBLP, 2010:1045-1048.
- [14] MIKOLOV T. Statistical language models based on neural networks[OL]. <https://pdfs.semanticscholar.org/e753/714f98099e3da1e96c652d34cc45e315ad23.pdf>. Accessed on Jan. 2018.
- [15] WILLIAMS R J, ZIPSER D. Gradient-based learning algorithms for recurrent networks and their computational complexity[J]. Backpropagation: Theory, Architectures, And applications, 1995, 1:433-486.
- [16] HOCHREITER S. The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 1998, 6(2):107-116.
- [17] GUSTAVSSON A, MAGNUSON A, BLOMBERG B, et al. On the difficulty of training recurrent neural networks[J]. Computer Science, 2013, 52(3):337-345.
- [18] HOCHREITER S, SCHMIDHUBER J. Long short-term memory[J]. Neural Computation, 1997, 9(8):1735-1780.
- [19] OLAH C. Understanding LSTM Networks[OL]. <http://colah.github.io/posts/2015-08-Understanding-LSTMs>.
- [20] GRAVES A, SCHMIDHUBER J. Framewise phoneme classification with bidirectional LSTM networks[C]// IEEE International Joint Conference on Neural Networks, 2005(IJCNN'05). IEEE, 2005:2047-2052.
- [21] LIN M, CHEN Q, YAN S. Network In Network[J]. arXiv preprint arXiv:1312.0400, 2013.
- [22] WU Y, SCHUSTER M, CHEN Z, et al. Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation[J]. arXiv preprint arXiv:1609.08144, 2016.

- thresholding methods in recent 20 years(1994—2014)[J]. Journal of Data Acquisition and Processing, 2015, 30(1): 1-23. (in Chinese)
- 吴一全, 孟天亮, 吴诗嫻. 图像阈值分割方法研究进展 20 年(1994—2014)[J]. 数据采集与处理, 2015, 30(1): 1-23.
- [2] ZHANG P F, LU S F, LI J Q. Multi-component segmentation of X-ray computed tomography (CT) image using multi-Otsu thresholding algorithm and scanning electron microscopy[J]. Energy Exploration & Exploitation, 2017, 35(3): 281-294.
- [3] YIN P Y, WU T H. Multi-objective and multi-level image thresholding based on dominance and diversity criteria[J]. Applied Soft Computing, 2017, 54: 62-73.
- [4] OTSU N. A threshold selection method from gray-level histogram[J]. IEEE Transactions on Systems, 1979, 9(1): 62-66.
- [5] JYOTIKA P, GAURAV G. Image Segmentation Using Genetic Algorithm OTSU[C] // 5th International Conference on Soft Computing for Problem Solving(SocProS). 2016: 473-480.
- [6] LIU J Z, LI W Q. The Automatic Thresholding of Gray-Level Pictures Via Two-Dimensional OTSU Method[J]. Acta Automatica Sinica, 1993, 91(1): 101-105. (in Chinese)
- 刘健庄, 栗文青. 灰度图像的二维 Otsu 自动阈值分割法[J]. 自动化学报, 1993, 19(1): 101-105.
- [7] JING X J, LI J F, LIU Y L. Image Segmentation Based on 3-D Maximum Between-Cluster Variance[J]. Acta Electronica Sinica, 2003, 31(9): 1281-1285. (in Chinese)
- 景晓军, 李剑峰, 刘郁林. 一种基于三维最大类间方差的图像分割算法[J]. 电子学报, 2003, 31(9): 1281-1285.
- [8] FAN J L, ZHAO F, ZHANG X F. Recursive Algorithm for Three-Dimensional Otsu's Thresholding Segmentation Method[J]. Acta Electronica Sinica, 2007, 35(7): 1398-1402. (in Chinese)
- 范九伦, 赵凤, 张雪峰. 三维 Otsu 阈值分割方法的递推算法[J]. 电子学报, 2007, 35(7): 1398-1402.
- [9] ZHAO F, FAN J L. Three-dimensional Otsu's method with non local spatial gray information[J]. Computer Engineering and Applications, 2013, 49(3): 30-33. (in Chinese)
- 赵凤, 范九伦. 融合非局部空间灰度信息的三维 Otsu 法[J]. 计算机工程与应用, 2013, 49(3): 30-33.
- [10] ZENG Y Z, WANG R M. A method for three-dimension OTSU image segmentation based on adaptive particle swarm optimization[J]. Electronic Design Engineering, 2011, 19(13): 173-175. (in Chinese)
- 曾业战, 王润民. 基于自适应粒子群优化的三维 OTSU 图像分割算法[J]. 电子设计工程, 2011, 19(13): 173-175.
- [11] SHILPA S, SHYAM L. Multilevel thresholding based on Chaotic Darwinian Particle Swarm Optimization for segmentation of satellite images[J]. Applied Soft Computing, 2017(55): 503-522.
- [12] HE L F, HUANG S W. Modified firefly algorithm based multi-level thresholding for color image segmentation[J]. Neurocomputing, 2017, 240: 152-174.
- [13] FAN J W, WANG Q P, LUO H, et al. Fast Iterative Algorithm for Segmentation Based on an Improved Three-Dimensional Otsu[C] // 2015 National Microwave and Millimeter Wave Conference. 2015: 5. (in Chinese)
- 范加武, 王青平, 罗慧, 等. 基于改进的三维 Otsu 分割快速迭代算法[C] // 2015 年全国微波毫米波会议. 2015: 5.
- [14] PARE S, KUMAR A, BAJAJ V. A multilevel color image segmentation technique based on cuckoo search algorithm and energy curve[J]. Applied Soft Computing, 2016, 47(C): 76-102.
- [15] SUDARSHAN N, YANG X S, PRATIM S P. Color Image Segmentation By Cuckoo Search[J]. Intelligent Automation and Soft Computing, 2015, 21(4): 673-685.
- [16] HE X S, LI N, YANG X S, et al. Multi-objective Cuckoo Search Algorithm[J]. Journal of System Simulation, 2015, 27(4): 731-737. (in Chinese)
- 贺兴时, 李娜, 杨新社, 等. 多目标布谷鸟搜索算法[J]. 系统仿真学报, 2015, 27(4): 731-737.
- [17] YANG X S, DEB S. Cuckoo Search via Levy Flights[C] // Proc. of World Congress on Nature & Biologically Inspired Computing, India, USA: IEEE Publications, 2009: 210-214.
- [18] LIU X N, MA M. Application of Cuckoo Algorithm in Multi-threshold Image Segmentation[J]. Computer Engineering, 2013, 39(7): 274-278. (in Chinese)
- 柳新妮, 马苗. 布谷鸟搜索算法在多阈值图像分割中的应用[J]. 计算机工程, 2013, 39(7): 274-278.
- (上接第 217 页)
- [23] HAWKINS D M. The Problem of Overfitting[J]. Cheminform, 2004, 35(19): 1-12.
- [24] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Science, 2012, 3(4): 212-223.
- [25] MAAS A L, DALY R E, PHAM P T, et al. Learning word vectors for sentiment analysis[C] // Meeting of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, 2011: 142-150.
- [26] NAKOV P, RITTER A, Rosenthal S, et al. SemEval-2016 Task 4: Sentiment Analysis in Twitter[C] // International Workshop on Semantic Evaluation. 2016: 1-18.
- [27] ANOOP V S, PREM S C. Generating and visualizing topic hierarchies from microblogs: An iterative latent dirichlet allocation approach[C] // 2015 International Conference on Advances in Computing Communications and Informatics(ICACCI). 2015: 824-828.
- [28] LE Q V, MIKOLOV T. Distributed Representations of Sentences and Documents[J]. ICML, 2014, 4(2): 1188-1196.
- [29] DUCHI J, HAZAN E, SINGER Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of Machine Learning Research, 2011, 12(7): 257-269.
- [30] ZEILER M D. Adadelta: an adaptive learning rate method[J]. arXiv preprint arXiv:1212. 5701, 2012.
- [31] KINGMA D, BA J. Adam: a method for stochastic optimization[J]. arXiv preprint arXiv:1412. 6980, 2014.