(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 110046223 B
(45)授权公告日 2021.05.18

(21)申请号 201910191148.0

(22)申请日 2019.03.13

(65)同一申请的已公布的文献号
申请公布号 CN 110046223 A

(43)申请公布日 2019.07.23

(73)专利权人 重庆邮电大学
地址 400065 重庆市南岸区南山街道崇文路2号
专利权人 重庆信科设计有限公司

(72)发明人 李俭兵 刘栗材 张功国

(74)专利代理机构 重庆市恒信知识产权代理有限公司 50102
代理人 刘小红 陈栋梁

(51)Int.Cl.
G06F 16/33(2019.01)
G06F 40/289(2020.01)
G06K 9/62(2006.01)
G06N 3/04(2006.01)

(56)对比文件
CN 107506722 A,2017.12.22
CN 108427670 A,2018.08.21
CN 108446271 A,2018.08.24
CN 106096664 A,2016.11.09
CN 109213868 A,2019.01.15
CN 105740349 A,2016.07.06
US 2019050875 A1,2019.02.14
关鹏飞 等."注意力增强的双向LSTM情感分析".《中文信息学报》.2019,第33卷(第2期),第105-111页.
周敬一 等."基于深度学习的中文影评情感分析".《上海大学学报》.2018,第25卷(第5期),第703-712页.
Xi Quyang et al.."Sentiment Analysis Using Convolutional Neural Network".《2015 IEEE International Conference on Computer and Information Technology Ubiquitous Computing and Communications》.2015,第2359-2364页.

审查员 崔倩倩

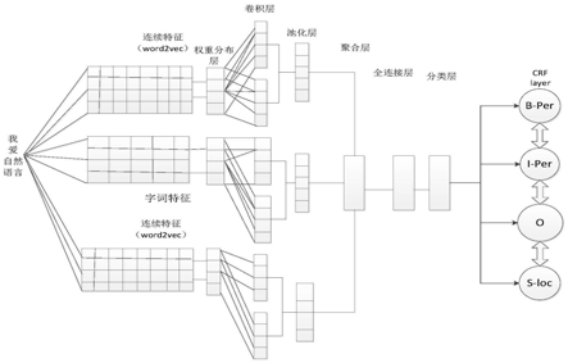
权利要求书2页 说明书8页 附图1页

(54)发明名称

基于改进型卷积神经网络模型的影评情感分析方法

(57)摘要

本发明请求保护一种基于改进型卷积神经网络模型的影评情感分析方法。在输入层和卷积层之间引入权重分布层,可以对文本中重要部分分析,减少噪音,使处理的特征得到提升。利用卷积建立模型,卷积方法是在字词的周围产生局部特征,然后使用局部最大值的方式组合,以创建固定大小的特征。在卷积层中使用的是梯度下降法来计算,可能会出现梯度弥散,引入门控机制降低弥散;其次,在新模型中取消了softmax层,加入支持向量机层;最后,使用条件随机场不仅处理传统模型在第i个标签上的特征函数也处理其前后位置的信息特征函数。本发明在传统卷积神经网络的基础上进行改进,添加了条件随机场层,从而可以提取高级抽象的特征,具有更好的分类能力。



CN 110046223 B

1.一种基于改进型卷积神经网络模型的影评情感分析方法，其特征在于，包括以下步骤：

步骤1、文本预处理步骤：输入原始的中文影评文本，并通过文本预处理过程来转化成便于计算机处理的数字形式，输入步骤2改进的卷积神经网络模型；

步骤2、在输入层和卷积层之间引入了权重分布层，权重分布层用于自动提取出非连续词语的前后文信息间的关系，权重分布层对文本中重要部分进行分析；所述卷积层采用梯度下降法来计算，梯度下降法中加入门控机制来降低弥散，还保留了模型的非线性；

步骤3、把影评文本特征经过线性变化和支持向量机层后，得到的概率传送给条件随机场，条件随机场模型选择概率最大的标注序列为最终的序列标注结果，将传统条件随机场的线性特征函数转化为该模型输出的非线性特征函数，拟合数据，条件随机场层是基于训练的词性知识信息帮助模型更好的理解了文本的语义，同时和神经网络学习的语义特征进行整体的优化求解，最后用条件随机场层获取全局最优的输出序列，即文本情感分析概率值；

所述步骤2的权重分布层自动提取出非连续词语的前后文信息间的关系，具体包括：权重分布层首先为每个字词建立上下文向量，使字词向量与其进行拼接，从而作为该字词的新表示方式，另外，根据汉语的表达习惯，距离远的词汇往往联系较少，权重分布层里考虑到距离衰减度的影响；

权重分布层是在输入层输出句子X_i后根据重要性将不同权重赋予在上下文向量Z_i上，再分别对字词打分再进行加权计算；

$$X_i=\sum_{j\neq i} a_{i,j} * Z_i$$

权重$a\geqslant 0$且$\sum_{j\neq i} a_{i,j}=1$，其中

$score(x_i,x_j)=x_i^T w_a x_j$

w_a是一个词向量，通过加大权重分布的数量，增加不同$score(x_i,x_j)$的个数，即词向量w_a变成对应词向量矩阵W_a；

$score(x_i,x_j)'=x_i^T W_a x_j$

通过使用欧式距离计算两个字词间距离，在权重计算里面加入距离衰减度，由于欧式距离值较大，为了保证在同一级别中，再对其归一化，使$sim(x_i,x_j)\in[0,1]$，最终可用下式来表示$sim(x_i,x_j)$，其中x_j'是x_j的扩展词向量；

$$dist(x_i,x_j)=\sqrt{\sum_{i,j=1}^{n}(x_i-x_j)^2}$$

$$sim(x_i,x_j)=\frac{1}{1+dist(x_i,x_j)}$$

从而得到下面式子：

$$a_{i,j}=\frac{\exp(score(x_i,x_j')')'}{\sum_{j'}\exp(score(x_i,x_j')')'} * sim(x_i,x_j)$$

；

使score(x_i,x_j)'值大的在上下文向量Z_i中的权重更大,随着句子长度增加会产生一定的噪声,为了避免这些影响,增加了衰减因子γ∈[0,1]作为惩罚;

score(x_i,x_j)'=(1-γ)^k$x_i^T$$W_a$$x_j$

k=|j-i|-1,当γ趋于1时,代表考虑的只有局部范围上下文,趋于0时,考虑更广的范围,得到最新评分后带入$a_{i,j}$计算式计算权重,从而得到Z_i;将权重分布层获得的向量与单词向量串联,得到更新的X_i,再传入卷积层,使其在宽度为n的滑窗上进行卷积;

所述卷积层具体包括:卷积方法是在字词的周围产生局部特征,然后使用局部最大值的方式来组合,以创建固定大小的特征,为了提取不同的局部特征,使用3层卷积层,使其卷积上下文窗口n的大小依次为2,3,4倍的字词粒度向量维度;

在卷积层中使用的是梯度下降法来确定模型中的参数值,使用梯度下降法的过程中可能会出现梯度弥散或爆炸,所以加入门控机制来解决这个问题,门控机制的梯度如下式:

$$\nabla(X+X\otimes\delta(X))=\nabla X+\nabla X\otimes\delta(X)+X\otimes\delta'(X)\nabla X$$

上式中衰减项只有一个δ'(X),证明加入门控机制可以用来降低弥散,使在卷积层中衰减的速度降慢;

在卷积层后再用池化层连接,在池化层采用局部最大值的方法进行特征的采样,得到宽度大小一样的特征值$\overline{C}$,则有:

$$\overline{C}=\max[C]$$。

2.根据权利要求1所述的一种基于改进型卷积神经网络模型的影评情感分析方法,其特征在于,所述步骤1文本预处理步骤具体包括:先将输入样本的词语序列分别转换成对应的预训练字和词粒度词向量序列,词粒度嵌入,包含了嵌入矩阵的列向量,每一方框包含了一句话中的某个词,每一列表示了这一句话,通过用矩阵向量表示词粒度嵌入;字粒度嵌入,从字中提取信息,考虑语句中所有的包括哈希标签在内的字符,并选择重要的特征;字粒度嵌入由嵌入矩阵中的列向量进行编码,给定一个字符,其嵌入有矩阵向量乘积获得。

3.根据权利要求2所述的一种基于改进型卷积神经网络模型的影评情感分析方法,其特征在于,所述步骤1文本预处理步骤还包括:

首先要对中文文本原始语料使用jieba软件进行分词,为了充分保留文本信息,分词后的语料依然保留标点与所有字、词;使用word2vec算法对分词后的语料先单独进行预训练生成词向量词典;其中的字词是不重复的,word2vec会对文本中的字、词和标点等基本元素的出现频率进行统计,通过无监督训练,获得作为语料基础构成元素的字词对应的指定维度的向量表征。

4.根据权利要求1所述的一种基于改进型卷积神经网络模型的影评情感分析方法,其特征在于,所述步骤3将传统条件随机场的线性特征函数转化为CNN-SVM模型输出的非线性特征函数,从而更好的拟合数据,如下式:

$$P_{tb}=\prod_{t=1}^{n}\Gamma\left(y_{b_{t-1}},y_{b_t}\right)\bullet P_{bt}\left(y|h_t\right)$$

$\Gamma\left(y_{b_{t-1}},y_{b_t}\right)$表示转移概率,$b_{t-1}$是前一词语的情感标识,$b_t$表示当前字词的情感标识,$P_{bt}$(y|$h_t$)表示发射概率,表示先位置字词归属每一类情感分类的概率值,P_{tb}表示序列结果的概率值。

基于改进型卷积神经网络模型的影评情感分析方法

技术领域

[0001] 本发明属于中文文本情感分析，尤其涉及一种基于改进型卷积神经网络模型的影评情感分析方法。

背景技术

[0002] 近年来很多人开始在网络上表达自己的想法和意见。在观看了电影后，在豆瓣等地方留下自己的影评，来表达自己对这电影的一些观点。将这些海量的情感文本进行数据分析，有助于用户在网络上有更好的体验感以及帮助网站更好的运营。传统的电影推荐方法主要是利用目标客户的打分记录来找到和他相似的用户，或者利用用户的历史偏好找到前n个和以往观看过相似的电影来做推荐，这类推荐方法的操作简便，精度较高。缺点也很明显，有些观众可能会随某时刻心情随意打分导致推荐的结果并不可信。这时候观影用户对电影主观的评价内容可以更好的发挥作用，通过评论的分析可更为准确的获取用户对电影的评价。

[0003] 目前，文本情感分析主要方法有利用机器学习方法和基于词典方法。在现在的研究中，基于词典的情感分析最为基础的还是进行情感词典的构建，而中文的情感词典资源过少且不够完善，加上中文语言的“一义多词”和“网络化”的影响，一部情感词典往往很难去解决情感分析中的问题。

[0004] 由于计算能力受到内存和带宽等因素的约束，循环神经网络(RNN)和长短期记忆网络(LSTM)的计算是非常困难。RNN的缺点在于没有办法使各部分平行化处理，导致训练时间长，有较高的时间复杂度，LSTM也没有办法达到并行化，它当前时刻的输出信息依赖前一个时刻隐藏层的状态。反而，CNN算法可以提升计算的速度，并且避免了RNN经常出现的梯度消失及梯度爆炸问题。但是，CNN也有很大的缺点，对于长文本上下文信息的获取和非连续词汇间的相关性计算较困难。

发明内容

[0005] 本发明旨在解决以上现有技术电影影评文本短，新鲜词多，语序不合理等特点，传统的情感分析方法已无法体现句子的正确语义的问题，提出了一种基于改进型卷积神经网络模型的影评情感分析方法。本发明的技术方案如下：

[0006] 一种基于改进型卷积神经网络模型的影评情感分析方法，其包括以下步骤：

[0007] 步骤1、文本预处理步骤：输入原始的中文影评文本，并通过文本预处理过程来转化成便于计算机处理的数字形式，输入步骤2改进的卷积神经网络模型；

[0008] 步骤2、对传统卷积神经网络模型进行改进，改进主要在于：在输入层和卷积层之间引入了权重分布层，可以对影评文本重要部分进行分析，减少噪音，使处理的特征得到提升；所述卷积层采用梯度下降法来计算，会出现梯度弥散，梯度下降法中加入门控机制来降低弥散，还保留了模型的非线性；

[0009] 步骤3、把影评文本中特征经过线性变化和支持向量机层后，得到的概率传送给条

件随机场，条件随机场模型选择概率最大的标注序列为最终的序列标注结果，将传统条件随机场的线性特征函数转化为该模型输出的非线性特征函数，拟合数据。条件随机场层是基于训练的词性知识信息帮助模型更好的理解了文本的语义，同时和神经网络学习的语义特征进行整体的优化求解。最后用条件随机场层获取全局最优的输出序列，即文本情感分析概率值。

[0010] 进一步的，所述步骤1文本预处理步骤具体包括：先将输入样本的词语序列分别转换成对应的预训练字和词粒度词向量序列，词粒度嵌入，包含了嵌入矩阵的列向量，每一方框包含了一句话中的某个词，每一列表示了这一句话，通过用矩阵向量表示词粒度嵌入；字粒度嵌入，从字中提取信息，考虑语句中所有的包括哈希标签在内的字符，并选择重要的特征；字粒度嵌入由嵌入矩阵中的列向量进行编码，给定一个字符，其嵌入有矩阵向量乘积获得。

[0011] 进一步的，所述步骤1文本预处理步骤还包括：

[0012] 首先要对中文文本原始语料使用jieba软件进行分词，为了充分保留文本信息，分词后的语料依然保留标点与所有字、词；使用word2vec算法对分词后的语料先单独进行预训练生成词向量词典；其中的字词是不重复的，word2vec 会对文本中的字、词和标点等基本元素的出现频率进行统计，通过无监督训练，获得作为语料基础构成元素的字词对应的指定维度的向量表征。

[0013] 进一步的，所述步骤2的权重分布层自动提取出非连续词语的前后文信息间的关系，具体包括：权重分布层首先为每个字词建立上下文向量，使字词向量与其进行拼接，从而作为该字词的新表示方式，另外，根据汉语的表达习惯，距离远的词汇往往联系较少，权重分布层里考虑到距离衰减度的影响；

[0014] 权重分布层是在输入层输出句子X后根据重要性将不同权重赋予在上下文向量Z_i上，再分别对字词打分再进行加权计算；

[0015] $$Z_i=\sum_{j\neq i} a_{i,j} * x_j$$

[0016] 权重$a\geq 0$且$\sum_{j\neq i} a_{i,j} * x_j=1$，其中

[0017] $score(x_i,x_j)=x_i^T w_a x_j$

[0018] w_a是一个词向量，通过加大权重分布的数量，增加不同$score(x_i,x_j)$的个数，即词向量w_a变成对应词向量矩阵W_a；

[0019] $score(x_i,x_j)'=x_i^T W_a x_j$

[0020] 通过使用欧式距离计算两个字词间距离，在权重计算里面加入距离衰减度，由于欧式距离值较大，为了保证在同一级别中，再对其归一化，使$sim(x_i,x_j)\in[0,1]$；

[0021] $$dist(x_i,x_j)=\sqrt{\sum_{i,j=1}^{n}(x_i-x_j)^2}$$

[0022] $$sim(x_i,x_j)=\frac{1}{1+dist(x_i,x_j)}$$

[0023] 从而得到下面式子：

[0024] $$a_{i,j}=\frac{\exp(score(x_i,x_j')')'}{\sum_{j'}\exp(score(x_i,x_j')')'}*sim(x_i,x_j)$$

[0025] 进一步的，使score$(x_i,x_j)'$值大的在上下文向量Z_i中的权重更大，随着句子长度增加会产生一定的噪声，为了避免这些影响，增加了衰减因子$\gamma\in[0,1]$来作为惩罚；

[0026] $score(x_i,x_j)'=(1-\gamma)^k x_i^T W_a x_j$

[0027] $k=|j-i|-1$，当γ趋于1时，代表考虑的只有局部范围上下文，趋于0时，考虑更广的范围；然后把权重分布层获得的向量与单词向量串联，得到更新的x_i，再传入卷积层，使其在宽度为n的滑窗上进行卷积。

[0028] 进一步的，所述卷积层具体包括：卷积方法是在字词的周围产生局部特征，然后使用局部最大值的方式来组合，以创建固定大小的特征，为了提取不同的局部特征，使用3层卷积层，使其卷积上下文窗口n的大小依次为2,3,4倍的字词粒度向量维度；

[0029] 一个句子$\{r_1,r_2,r_3,\ldots,r_m\}$，定义向量$z_m$为词向量的连接，可表示为：

[0030] $z_m=(r_{m-(k-1)/2},\cdots,r_{m+(k-1)/2})^T$

[0031] 在卷积层中使用的是梯度下降法来确定模型中的参数值，使用梯度下降法的过程中可能会出现梯度弥散或爆炸，所以引入门控机制来解决这个问题。门控机制的梯度如下式：

[0032] $$\nabla(X+X\otimes\delta(X))=\nabla X+\nabla X\otimes\delta(X)+X\otimes\delta'(X)\nabla X$$

[0033] 上式中衰减项只有一个，证明加入门控机制可以用来降低弥散，使在卷积层中衰减的速度降慢；

[0034] 在卷积层后再用池化层连接，在池化层采用局部最大值的方法进行特征的采样，得到宽度大小一样的特征值$\overline{C}$，则有：

[0035] $$\overline{C}=\max[C]$$

[0036] 进一步的，所述步骤3将传统CRF的线性特征函数转化为CNN-SVM模型输出的非线性特征函数，从而更好的拟合数据，如下式：

[0037] $$P_{tb}=\prod_{t=1}^{n}\Gamma\left(y_{b_{t-1}},y_{b_t}\right)\bullet P_{bt}\left(y|h_t\right)$$

[0038] $\Gamma\left(y_{b_{t-1}},y_{b_t}\right)$表示转移概率，$b_{t-1}$是前一词语的情感标识，$b_t$表示当前字词的情感标识，$P_{bt}(y|h_t)$表示发射概率，表示先位置字词归属每一类情感分类的概率值。P_{tb}表示序列结果的概率值。

[0039] 本发明的优点及有益效果如下：

[0040] 本发明通过一种基于改进型卷积神经网络模型的影评情感分析方法对电影影评进行情感分析。传统的情感分析模型在处理电影影评文本短，新鲜词多，语序不合理等特点时十分困难。首先对卷积神经网络做了改进，传统卷积神经网络存在对长文本上下文信息的获取和非连续词汇间的相关性计算上困难的问题。本方法的权重分布层可以对重要部分进行分析，减少噪音，使处理的特征得到提升，弥补卷积神经网络的长语句中上下文信息关联上的缺陷问题。再因为卷积层中使用的是梯度下降法来计算，会出现梯度弥散，加入门控机制来降低弥散，并且保留了模型的非线性。另外考虑到上下文信息带有自身固有的属性

特征和语序不合理性。并且利用字粒度词向量为特征，从而解决了歧义词的切分问题，并且能学习到更加具体的特征。

[0041] 为了对重要部分进行分析，减少噪音，使处理的特征得到提升，在输入层和卷积层中加入权重分布层。权重分布层是在输入层输出句子X后根据重要性将不同权重赋予在上下文向量Z_i上，再分别对字词打分再进行加权计算，引入的Z_i计算式如下，其中x_j'是x_j的扩展词向量。

[0042] $$Z_i=\sum_{j\neq i}a_{i,j}*x_j$$

[0043] $$score(x_i,x_j)'=(1-\gamma)^k x_i^T W_a x_j$$

[0044] $$a_{i,j}=\frac{\exp(score(x_i,x_j')')'}{\sum_{j'}\exp(score(x_i,x_j')')'}*sim(x_i,x_j)$$

[0045] 另外在卷积层中使用的是梯度下降法来计算，会出现梯度弥散，引入门控机制来降低梯度弥散，并且保留了模型的非线性。因此针对中文长文本局部和上下文信息传递，从两个方向来卷积，使影评上下文和局部有更多联系以达到进一步提高特征学习和特征提取能力。

[0046] $$h_i^1=(X_i^0\times W_i^1+b_i^1)+(X_{i-1}^1\times W_{i-1,i}^1+b_i^1)\otimes\delta(X_{i-1}^1\times V_{i-1}^1+b_i^1)$$

[0047] $$h_i^2=(X_i^0\times W_i^2+b_i^2)+(X_{i-1}^2\times W_{i,i+1}^2+b_i^2)\otimes\delta(X_{i+1}^2\times V_{i+1}^2+b_i^2)$$

[0048] $$h_i^3=((h_i^1,h_i^2)\times W_i^3+b_i^3)$$

[0049] $M(X,V,b)=\delta(X\times V,b)$

[0050] W和V分别为不一致的卷积核，输出通道数为n，核宽度为k，b为偏置参数，M(X,V,b)是门函数。对输入的传递信息进行卷积就是进行门控，将信息传递的速度进行控制，使其始终在(0,1)间。

[0051] 最后的条件随机场进一步弥补了卷积神经网络不能正确的对上下文信息的获取和非连续词汇间的相关性计算问题。我们考虑利用词语的词性对句子中知识信息进行序列标注。条件随机场层是基于我们前期训练的词性知识信息帮助模型更好的理解文本的语义，同时和神经网络学习的语义特征进行整体的优化求解。最后用条件随机场获取全局最优的输出序列。

附图说明

[0052] 图1是本发明提供优选实施例流程示意图。

具体实施方式

[0053] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、详细地描述。所描述的实施例仅仅是本发明的一部分实施例。

[0054] 本发明解决上述技术问题的技术方案是：

[0055] 如图1所示，先将输入样本的词语序列分别转换成对应的预训练字和词粒度词向量序列。词粒度嵌入，包含了嵌入矩阵的列向量，每一方框包含了一句话中的某个词，每一

列表示了这一句话。通过用矩阵向量表示词粒度嵌入。字粒度嵌入，从字中提取信息，考虑语句中所有的字符(包括哈希标签等)，并选择重要的特征。字粒度嵌入由嵌入矩阵中的列向量进行编码，给定一个字符，其嵌入有矩阵向量乘积获得。

[0056] 权重分布层首先为每个字词建立上下文向量，使字词向量与其进行拼接，从而作为该字词的新表示方式。另外，根据汉语的表达习惯，距离远的词汇往往联系较少，权重分布层里考虑到距离衰减度的影响。

[0057] 权重分布层是在输入层输出句子X后根据重要性将不同权重赋予在上下文向量Z_i上，在影响语句情感分析时，通过这种方式可以知道哪些词语更重要，对句意影响更大，再分别对字词打分再进行加权计算。

[0058] $$Z_i=\sum_{j\neq i} a_{i,j} * x_j$$

[0059] 权重$a\geq 0$且$\sum_{j\neq i} a_{i,j} * x_j=1$，其中

[0060] $score(x_i,x_j)=x_i^T W_a x_j$

[0061] W_a是一个词向量。由于汉语中一词多义等因素，一句话在不同环境中语义不一样，一种意思的权重分布只能在对应的语义上面合理，在其他语义上效果就不明显，所以通过加大权重分布的数量，增加不同$score(x_i,x_j)$的个数，即词向量 W_a变成对应词向量矩阵W_a。

[0062] $score(x_i,x_j)'=x_i^T W_a x_j$

[0063] 另外考虑到距离远的词汇往往联系较少的原因，通过使用欧式距离计算两个字词间距离，在权重计算里面加入距离衰减度。由于欧式距离值较大，为了保证在同一级别中，再对其归一化，使$sim(x_i,x_j)\in[0,1]$。

[0064] $$dist(x_i,x_j)=\sqrt{\sum_{i,j=1}^{n}(x_i-x_j)^2}$$

[0065] $$sim(x_i,x_j)=\frac{1}{1+dist(x_i,x_j)}$$

[0066] 从而得到下面式子：

[0067] $$a_{i,j}=\frac{\exp(score(x_i,x_j')')'}{\sum_{j'}\exp(score(x_i,x_j')')'} * sim(x_i,x_j)$$

[0068] 通过计算，使$score(x_i,x_j)'$值大的在上下文向量Z_i中的权重更大。随着句子长度增加会产生一定的噪声，为了避免这些影响，增加了衰减因子$\gamma\in[0,1]$来作为惩罚。

[0069] $score(x_i,x_j)'=(1-\gamma)^k x_i^T W_a x_j$

[0070] $k=|j-i|-1$，当γ趋于1时，代表考虑的只有局部范围上下文，趋于0时，考虑更广的范围。

[0071] 然后把权重分布层获得的向量与单词向量串联，得到更新的x_i，再传入卷积层，使其在宽度为n的滑窗上进行卷积。

[0072] 然后利用卷积建立模型，卷积方法是在字词的周围产生局部特征，然后使用局部最大值的方式来组合，以创建固定大小的特征。为了提取不同的局部特征，使用3层卷积层，使其卷积上下文窗口n的大小依次为2,3,4倍的字词粒度向量维度。

[0073] 一个句子$\{r_1,r_2,r_3,\cdots,r_m\}$,定义向量$z_m$为词向量的连接:

[0074] $z_m=(r_{m-(k-1)/2},\cdots,r_{m+(k-1)/2})^T$

[0075] 在卷积层进行最大化操作后,特征向量X^{wch}中的第j个元素如下:

[0076] $$[X^{wch}]_j=\max_{1<m<M}[W^0z_m+b^0]_j$$

[0077] 该卷积层的权重矩阵W^0,用这权重矩阵提取给定字词的窗口周围的局部特征。为了增加神经网络模型的非线性,在上面的卷积层后再加ReLu作为激活函数,但是使用sigmod函数会导致将近一半的神经元被激活。ReLu会使一部分神经元的输出为0,自动引入稀疏性,相当于无监督预练习。并且减少了参数的相互依存关系,缓解了过拟合问题的发生。句子1中的特征矩阵包含n个词表示为X[1:n],则有:

[0078] $X[1:n]=x_1+x_2+x_3+\cdots+x_n$

[0079] “+”表示串接操作。然后利用大小为h×k的滤波器对输入特征矩阵进行卷积操作,提取次序列特征,计算公式如下:

[0080] $C_i=f(w.x_{i:(i+h-1)}+b)$

[0081] C_i代表特征图中第i个特征值,h表示在窗口大小k下的长度,w为滤波器参数,b为偏置量,而f(.)表示该卷积核函数。因此可以得出特征C表示为:

[0082] $C=[c_1,c_2,c_3,\cdots,c_{n-h+1}]$

[0083] 在卷积层中使用的是梯度下降法来计算,会出现梯度弥散,加入门控机制来降低弥散,并且保留了模型的非线性。因为分析的字词对前后的字词有依赖性,如果一句话的开头是积极正向的,但是结尾是负面,最后实际情感分类也就是负面。例如“这部电影的导演和剧本都是非常不错的,但主角那糟糕的演技把这全毁了”。因此针对中文长文本局部和上下文信息传递,从两个方向来卷积,使上下文和局部有更多联系以达到进一步提高特征学习和特征提取能力。

[0084] $$h_i^1=(X_i^0\times W_i^1+b_i^1)+(X_{i-1}^{\ 1}\times W_{i-1,i}^{\ 1}+b_i^1)\otimes\delta(X_{i-1}^{\ 1}\times V_{i-1}^{\ 1}+b_i^1)$$

[0085] $$h_i^2=(X_i^0\times W_i^2+b_i^2)+(X_{i-1}^{\ 2}\times W_{i,i+1}^{\ 2}+b_i^2)\otimes\delta(X_{i+1}^{\ 2}\times V_{i+1}^{\ 2}+b_i^2)$$

[0086] $$h_i^3=((h_i^1,h_i^2)\times W_i^3+b_i^3)$$

[0087] $M(X,V,b)=\delta(X\times V,b)$

[0088] W和V分别为不一致的卷积核,输出通道数为n,核宽度为k,b为偏置参数,M(X,V,b)是门函数。对输入的传递信息进行卷积就是进行门控,将信息传递的速度进行控制,使其始终在(0,1)间。

[0089] 门控机制的梯度如下式:

[0090] $$\nabla(X+X\otimes\delta(X))=\nabla X+\nabla X\otimes\delta(X)+X\otimes\delta'(X)\nabla X$$

[0091] 上式中衰减项只有δ’(X)一个,证明加入门控机制可以用来降低弥散,使在卷积层中衰减的速度降慢。

[0092] 在卷积层后再用池化层连接,在池化层采用局部最大值的方法进行特征的采样,得到宽度大小一样的特征值$\overline{C}$,则有:

[0093] $$\overline{C}=\max[C]$$

[0094] 池化层(Max-over-time pooling)解决了句子长度不一的问题,保证全连接输入神经元数目一定。池化方式有降维处理,从而降低计算复杂度,只需通过提取其中的最大值,池化层的输出为各个特征图的最大值,即一个一维向量。卷积层和池化层为特征提取层,模型经过三次特征提取层可得到全局特征值V 如下:

[0095] $$V=[\overline{C_{1,h_1}},\cdots\overline{C_{m,h_1}},\cdots,\overline{C_{1,h_k}},\cdots,\overline{C_{L,h_k}},\cdots]$$

[0096] 其中$\overline{C_{l,h_k}}$表示第k种类型的滤波器产生的第L个特征值。将句子的全局特征值给两个全连接层进行随机参数更新的方法处理,在每次前向传播进行参数学习的时候,随机参数选取指定的比例学习特征,在反向传播进行参数的梯度下降更新时,更新在前向传播中选定的特征。从而得出句子x的每个情感标签的得分,如下式:

[0097] $S=W^3h(W^2X_{wch}+b^2)+b^3$

[0098] W^3,W^2表示权重矩阵,b^2,b^3表示需要学习的超参数,h(.)表示正切函数。为了把情感标签的得分转换成条件概率分布,引入了Softmax,通过比较预测的标签值和真实的标签值来调整CNN模型。

[0099] $$P(y=j\mid x,B)=\frac{e^{x_{(w_j+b_j)}}}{\sum_{k=1}^{k}e^{x_{(w_j+b_k)}}}$$

[0100] 表示每个情感标签,B表示参数集合。再对这个式子取对数,可得:

[0101] $$\log P(y=j\mid x,B)=s(x)-\log(\sum_{k=1}^{k}e^{x_{(w_j+b_k)}})$$

[0102] 在神经网络中,经常要计算按照正向传播计算的分数S1,和按照正确标注计算的分数S2的差距,来计算出Loss,才能应用反向传播。在下式中计算出占得比重越大,这样本的Loss就越小。在训练集c中,我们采用随机梯度下降法 (SGD)来进行最优化训练,每一次迭代计算mini-batch的梯度,然后对参数进行更新。

[0103] $$\theta\rightarrow\frac{1}{|C|}\sum_{(x,y)\in C}-\log P(y=j\mid x,B)$$

[0104] 卷积层执行dropout,并根据训练集的规模选择适合的mini-batch。在卷积层加dropout提高了模型泛化能力。dropout是指网络中隐藏层节点会随机的暂时被选择隐藏而不工作,其权重会保留下来。dropout的作用是防止隐含层神经元之间的自适应性。

[0105] 由于传统的卷积神经网络的softmax层在执行分类的时候,容易过拟合。所以在CRCNN-SVM模型中取消了softmax层。当训练集的精确率在CNN上表现稳定时,保持训练好的模型参数,将经过采样层获取的特征向量S_{train}导出。再把句子放进模型,从而获取特征向量S_{test}。

[0106] 另外再添加一个SVM层,从而能更有效进行二分类。

[0107] SVM是一种有监督的学习模型。通过上述方法,我们可以获得数据特征向量,然后选择模型适用的核函数,通过核函数巧妙地将数据映射到更高维度,从而利用一个超平面来对非线性数据进行分类。核函数事先在低维上进行运算,而分类效果表现在高维上,因此不会增加计算复杂度。该模型的优化目标,是最大化分类的超平面和两类数据的间距,最后得到类别标签。核方法是一种很巧妙的方法,既可以将特征映射到较高的维度,又可以地利

用了SVM的内积运算避免了维度计算量的复杂。最后的最优化问题如下式：

[0108] $$\max[\sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i,j=1}^{n} a_i a_j y_i y_j K(x_i, x_j)]$$

[0109] $$s.t., \sum_{i=1}^{n} a_i y_i = 0$$

[0110] $a_i \geq 0, i = 1, \cdots, n$

[0111] 最后用条件随机场获取全局最优的输出序列，进一步弥补了CNN不能正确的对上下文信息的获取和非连续词汇间的相关性计算问题。我们考虑利用词语的词性对句子中知识信息进行序列标注。条件随机场层是基于我们前期训练的词性知识信息帮助模型更好的理解文本的语义，同时和神经网络学习的语义特征进行整体的优化求解。最后用条件随机场层获取全局最优的输出序列。条件随机场模型选择概率最大的标注序列为最终的序列标注结果，此处的概率是指转移概率和发射概率，发射概率是指序列中的词语或字符属于每一标签类的概率，转移概率是指该标签类到下一个标签类的概率。特征通过线性变化和SVM层后输出的概率是发射概率。

[0112] CNN-SVM模型在第i个位置上的标签的输出，可以看作条件随机场里当前字词的标签概率，另外条件随机场还考虑了当前位置的上下文标签的概率。通过计算所有可能的序列标注结果中序列概率值最大的标注序列作为模型最终的预测结果。通过这种方式，将传统条件随机场的线性特征函数转化为CNN-SVM模型输出的非线性特征函数，从而更好的拟合数据。

[0113] $$P_{tb} = \prod_{t=1}^{n} \Gamma\left(y_{b_{t-1}}, y_{b_t}\right) \bullet P_{bt}\left(y \middle| h_t\right)$$

[0114] $\Gamma\left(y_{b_{t-1}}, y_{b_t}\right)$表示转移概率，$b_{t-1}$是前一词语的情感标识，$b_t$表示当前字词的情感标识。$P_{bt}(y|h_t)$表示发射概率，表示先位置字词归属每一类情感分类的概率值。P_{tb}表示序列结果的概率值。

[0115] 以上这些实施例应理解为仅用于说明本发明而不用于限制本发明的保护范围。在阅读了本发明的记载的内容之后，技术人员可以对本发明作各种改动或修改，这些等效变化和修饰同样落入本发明权利要求所限定的范围。

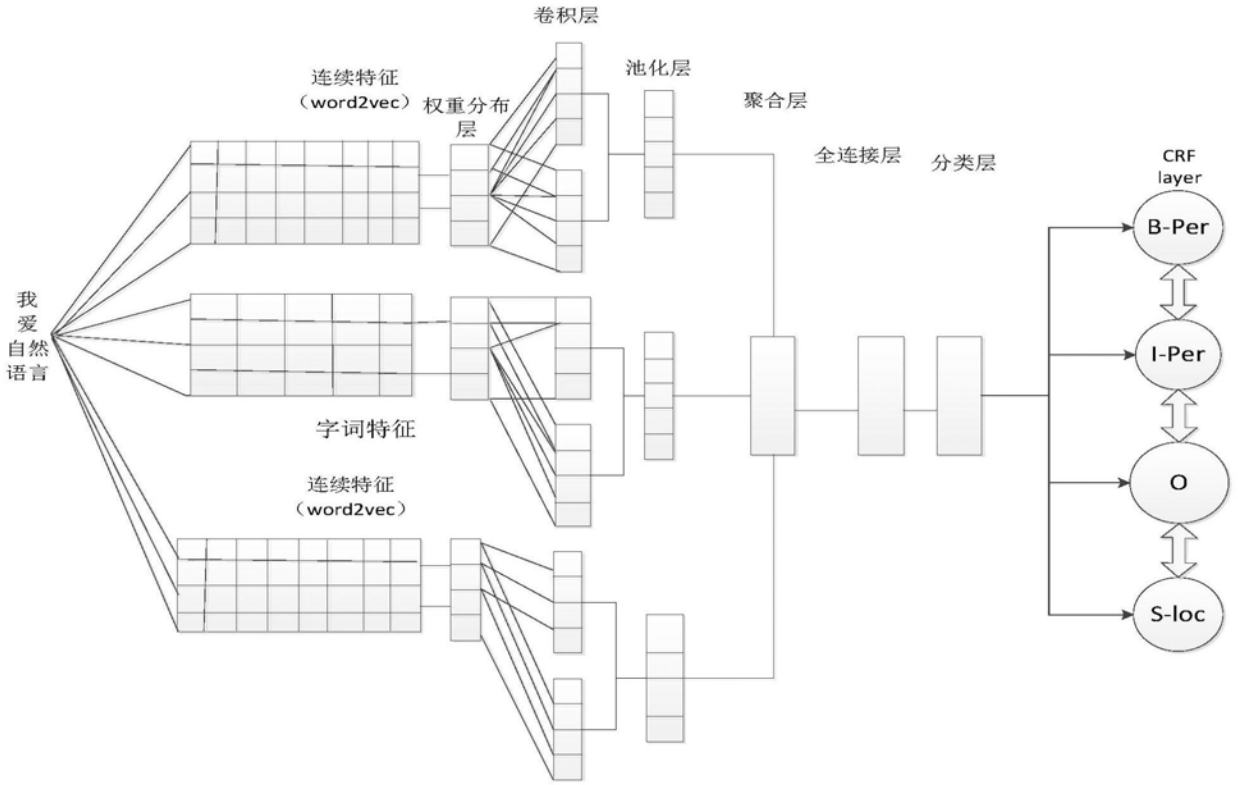

卷积层
连续特征
（word2vec）
权重分布
层
池化层
聚合层
全连接层
分类层
CRF
layer
B-Per
I-Per
O
S-loc
我
爱
自然
语言
字词特征
连续特征
（word2vec）

图1