



(12)发明专利申请

(10)申请公布号 CN 110288081 A

(43)申请公布日 2019.09.27

(21)申请号 201910476156.X

(22)申请日 2019.06.03

(71)申请人 北京信息科技大学

地址 100192 北京市海淀区清河小营东路12号

申请人 北京理工大学

(72)发明人 王军茹 卢继华 易军凯 徐懿

李梦泽 何天恺

(74)专利代理机构 北京理工正阳知识产权代理

事务所(普通合伙) 11639

代理人 王民盛

(51)Int.Cl.

G06N 3/04(2006.01)

G06N 3/08(2006.01)

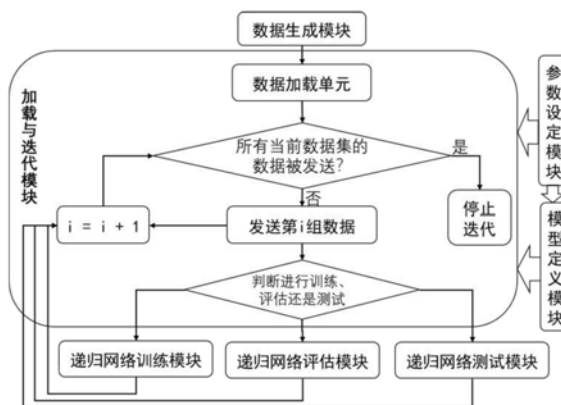
权利要求书3页 说明书9页 附图5页

(54)发明名称

一种基于FW机制及LSTM的递归网络模型及学习方法

(57)摘要

本发明涉及一种基于FW机制及LSTM的递归网络模型及学习方法,属于递归神经网络以及自然语言处理技术领域。包括基于FW机制及LSTM的递归网络模型及依托的学习方法;前者包括数据导入模块、数据生成模块、加载与迭代模块、参数设定模块、模型定义模块、递归网络训练、评估以及测试模块;学习方法包括:1导入数据;2将导入数据进行拆分得到训练数据、评估数据和测试数据;3根据导入数据,获取预先设定好的配置参数;4完成权重参数的初始化;5将训练、评估和测试数据送入LSTM单元计算输出向量;6计算loss函数,对网络参数进行优化,输出复杂度。所述网络模型及学习方法进一步提升了LSTM模型处理的准确度及收敛速度。



1. 基于FW机制及LSTM的递归网络模型,其特征在於:包括数据导入模块、数据生成模块、加载与迭代模块、参数设定模块、模型定义模块、递归网络训练模块、递归网络评估模块以及递归网络测试模块;

其中,数据生成模块又包括数据拆分单元;加载与迭代模块包括数据加载单元和迭代单元;

数据拆分单元包括训练数据生成单元、评估数据生成单元以及测试数据生成单元;

递归网络训练模块包括dropout单元、更新单元和结果储存单元;递归网络评估模块以及递归网络测试模块仅包括更新单元和结果储存单元;

其中,更新单元包括长短时记忆单元和快速权重单元;

所述基于FW机制及LSTM的递归网络模型中各模块的连接关系如下:

数据导入模块与数据生成模块相连,数据生成模块和加载与迭代模块相连,参数设定模块和加载与迭代模块以及模型定义模块相连,递归网络训练模块和加载与迭代模块、递归网络评估模块以及模型定义模块相连,递归网络评估模块与加载与迭代模块、递归网络训练模块、递归网络测试模块以及模型定义模块相连;递归网络测试模块与加载与迭代模块、递归网络评估模块和模型定义模块相连;

数据生成模块中各单元的连接关系如下:数据拆分单元中训练数据、评估数据和测试数据分别与训练标签生成单元、评估标签生成单元以及测试标签生成单元相连;

加载与迭代模块中各单元的连接关系如下:数据加载单元和迭代单元相连;

所述基于FW机制及LSTM的递归网络模型中各模块的信号产生及输出关系如下:

数据导入模块的输出接入数据生成模块;数据生成模块处理后接入加载与迭代模块;参数设定模块为加载与迭代模块和模型定义模块提供输入参数及FW模型参数;加载与迭代模块分别为递归网络训练模块、递归网络评估模块和递归网络测试模块提供训练数据和训练标签、评估数据和评估标签以及测试数据和测试标签;模型定义模块将FW模型参数分别输入递归网络训练模块、递归网络评估模块以及递归网络测试模块;递归网络训练模块将训练好的网络参数送入递归网络评估模块;递归网络评估模块将评估后的网络参数送入递归网络测试模块;

递归网络训练模块、评估模块以及测试模块中的各单元连接关系如下:

dropout单元接收数据并与长短时记忆单元相连,长短时记忆单元与数据输入和快速权重单元相连,结果储存单元与快速权重单元和结果相连。

2. 如权利要求1所述的基于FW机制及LSTM的递归网络模型依托的学习方法,其特征在於:包括如下步骤:

步骤一、待训练和测试的数据经数据导入模块导入,具体为:

步骤二、数据生成模块对经数据导入模块导入的数据经数据拆分单元进行拆分,分别得到训练数据、评估数据和测试数据;

步骤三、训练数据生成单元随机选取x%比例经数据拆分单元拆分后的数据生成训练集;评估数据生成单元随机选取y%比例经数据拆分单元拆分后的数据生成评估集;测试数据生成单元随机选取z%比例经数据拆分单元拆分后的数据生成测试集;

步骤四、训练标签生成单元将训练数据生成单元生成的训练集中每一个数据后移一位得到训练标签;评估标签生成单元将评估数据生成单元生成的评估集中每一个数据后移一

位得到评估标签;测试标签生成单元将测试集中每一个数据后移一位得到测试标签;

步骤五、参数设定模块根据数据导入模块导入文本的模型规模,获取配置参数,再将获取的配置参数输入参数设定模块;

其中,配置参数包括初始规模、学习率、最大梯度正则值、层数、步数、隐藏层大小、最大epoch数、极大epoch值、dropout率、衰减率、批大小以及vocab大小;

步骤六、加载与迭代模块中的数据加载单元按照参数设定模块中获取的配置参数加载训练集、评估集以及测试集中的数据,并设定初始化数据序号*i*为1;

步骤七、模型定义模块根据参数设定模块中的配置参数,使用伪随机函数在配置范围内生成随机值作为权重矩阵参数,完成权重参数的初始化;

步骤八、加载与迭代模块中迭代模块判断当前数据集中的数据是否发送完毕,并依据判断结果进行操作,具体为:

若当前数据集中的数据没发送完,则发送第*i*组数据,判断进行训练、评估还是测试,并跳至步骤九,跳至步骤八;否则停止迭代;

步骤九、判断当前数据是否是训练数据,若是则依据dropout率对输入数据进行抽取,抽取后数据,跳至步骤十;否则,跳至步骤十;

步骤十、将步骤九输入的数据送入更新单元中的长短时记忆单元以及快速权重单元计算得到输出向量,同时利用梯度下降法对网络进行优化,具体为:

步骤10.1更新单元基于输入层权重 $W_x$ 、标准权重 $W_h$ 计算起始隐藏层状态,通过公式(4)计算当前*t*时刻的初始隐藏状态:

$$h_t^0 = f(LN(W_x * x_t + W_h * h_{t-1})) \quad (4)$$

其中,输入层权重记为 $W_x$ 、标准权重记为 $W_h$ ;  $h_0$ 为起始隐藏层状态,  $LN$ 为层正则化函数;  $f$ 为激活函数;  $x_t$ 为当前*t*时刻的输入层数据;  $h_{t-1}$ 为当前时刻的前一时刻,即*t*-1时刻,隐藏层状态对应的数据,简称隐藏层状态;

标准权重 $W_h$ 是RNN网络中隐藏层向下一个时间步进传播的权重;输入层权重 $W_x$ 是输入层到隐藏层传播的权重;

步骤10.2、快速权重单元计算快速权重,具体通过公式(5)计算:

$$W_A(t) = \lambda W_A(t-1) + \eta h_{t-1} h_{t-1}^T \quad (5)$$

其中,  $W_A(t)$  是第*t*时刻的快速权重,是仅作用在隐藏层每个时间步进内的权重;一个时间步进更新的总次数,记为*s*+1次;  $\lambda$ 是衰减率、 $\eta$ 是学习率、 $h_{t-1}$ 为*t*-1时刻对应的隐藏层状态;  $h_{t-1}^T$ 是 $h_{t-1}$ 即*t*-1时刻对应隐藏层状态的转置;

其中,时间步进更新的总次数*s*+1中的*s*即步数;

步骤10.3、快速权重单元计算隐藏层状态并更新*s*次隐藏层状态;

步骤10.4、慢速权重单元计算归一化输出;

其中,网络的归一化输出通过Softmax或sigmoid函数两者之一实现;

步骤10.5、结果存储单元计算基于步骤10.4计算出的归一化输出计算损失loss和复杂度perplexity;

步骤10.6、慢速权重单元判断是否达到最后一个Epoch,如果没有达到,则更新单元则更新隐藏层状态以及训练参数或测试参数,将当前*i*加1,跳至步骤八。

3. 如权利要求2所述的基于FW机制及LSTM的递归网络模型依托的学习方法,其特征在

于:步骤一通过读取文本路径,获取文本数据。

4.如权利要求2所述的基于FW机制及LSTM的递归网络模型依托的学习方法,其特征在于:步骤二中,拆分具体为:将步骤一导入的文本数据按照每j个字符为一句话进行拆分。

5.如权利要求4所述的基于FW机制及LSTM的递归网络模型依托的学习方法,其特征在于:j的取值范围为5到50。

6.如权利要求2所述的基于FW机制及LSTM的递归网络模型依托的学习方法,其特征在于:步骤三中, $x\%+y\%+z\%=1$ 。

7.如权利要求2所述的基于FW机制及LSTM的递归网络模型依托的学习方法,其特征在于:步骤10.1中激活函数f为SeLU函数、Leaky Relu函数以及Swish函数中的一种。

8.如权利要求2所述的基于FW机制及LSTM的递归网络模型依托的学习方法,其特征在于:步骤10.2中,衰减率的取值范围为0.9到0.995。

9.如权利要求2所述的基于FW机制及LSTM的递归网络模型依托的学习方法,其特征在于:步骤10.2中,学习率的取值范围为0.3到0.8。

## 一种基于FW机制及LSTM的递归网络模型及学习方法

### 技术领域

[0001] 本发明涉及一种基于FW机制及LSTM的递归网络模型及学习方法,属于递归神经网络以及自然语言处理技术领域。

### 背景技术

[0002] 自然语言处理模型通常采用递归神经网络(Recurrent Neural Network,RNN)结构。RNN由隐藏层状态以及权重这两种时间规模的变量组成。隐藏层状态在每个时间步进都会更新一次;而权重则在序列所有信息全部输入网络之后再更新。因此,代表着网络中各层间连接关系的权重往往对应着网络的“长期记忆”。但是,实际网络各层关系变化与递进,往往与输入序列长度相关,可能是3,5个时间步进,也可能是30,50个时间步进需要更新。

[0003] 基于LSTM单元的语言模型是RNN应用较为广泛的改进网络之一,该模型通过对文本数据的训练,根据输入的文本预测该段文本中即将出现的下一个单词。网络参数初始状态用零向量进行初始化,并在读取每个单词后得到更新。模型在处理输入数据时使用反向传播法进行网络参数的优化。把输入数据,即若干句子组成的段落划分为固定长度的输入块,每个输入块还有固定长度个单词,每当处理完一个输入块后执行反向传播对网络参数进行更新。

[0004] Jimmy Ba等人提出了快速权重(Fast Weights,FW)机制,即引入更新周期处于隐藏层状态以及权重这两种时间规模之间的新变量来存储快速更新的隐藏层状态,对于序列到序列模型的学习已被证明十分有效。出于上述考虑,在保留现有隐藏层状态和标准权重的同时引入新的变量,这种变量的更新周期要比隐藏层更新周期更长,但是比标准权重更新周期更短,也称为快速权重。

[0005] 在神经网络训练方面,一般需经过复杂而耗时的处理,才能获得较好的学习性能,即需要较高的时间和计算成本。因此,研究者们为降低此时间及计算成本,往往选择批量处理。

[0006] 其中,批量正则化为其中的一个很典型的技术,然而其在递归神经网络的作用并不明显。因此,G.Hinton等人提出了层正则化(layer normalization,LN),具体实现为计算递归神经网络中某一个训练样本中在一个隐藏层上所有隐藏单元的状态的均值和标准差。LN用于解决快速权重机制中随着训练变多,解决隐藏层更新值期间的溢出问题。

[0007] 衡量语言模型性能的评价指标参数是复杂度perplexity和loss。其中,perplexity表示语言模型在学习文本数据后根据句子前面的单词预测下一单词的平均可选数量。例如,一个序列是由A、B、C、D、E五个字母无规律随机组成,那么预测下一个字母时,有5个等概率的选项,那么复杂度的值即为5。因此,若一个语言模型的复杂度为K,就说明语言模型在预测即将出现的单词时,平均有K个单词拥有相同的概率作为合理预测的选择。其中,K为整数,为目标单词的总数。以PTB模型为例,评价PTB模型性能指标的复杂度perplexity值的计算公式为(1):

$$[0008] \quad perplexity = e^{-\frac{1}{K} \sum_{i=1}^K \ln P_{target_i}} \quad (1)$$

[0009] 其中,  $P_{target_i}$  表示第  $i$  个目标单词,  $\ln$  为对数函数;

[0010] 另一衡量语言模型性能的评价指标参数  $loss$  定义为目标单词出现概率的平均负对数, 表达式如 (2):

$$[0011] \quad loss = -\frac{1}{N} \sum_{i=1}^N \ln P_{target_i} \quad (2)$$

[0012]  $perplexity$  值与  $loss$  的关系为 (3):

$$[0013] \quad perplexity = e^{loss} \quad (3)$$

[0014] 当语言模型学习句子中单词与单词之间相互关联的逻辑关系时, 模型的学习能力越强, 根据之前出现的单词预测下一个单词时, 备选的单词数量就越少, 相应的复杂度  $perplexity$  就越低。所以复杂度  $perplexity$  能够很好地反映网络的学习性能。复杂度  $perplexity$  越低, 代表网络预测句子中下一个单词的能力越强, 效果也就越好。

## 发明内容

[0015] 本发明的目的在于进一步提升现有基于 LSTM 递归神经网络在处理时间关联度强的自然语言时存在复杂度性能有待进一步提升的技术现状提出了一种基于 FW 机制及 LSTM 的递归网络模型及学习方法。

[0016] 所述基于 FW 机制及 LSTM 的递归网络模型及学习方法包括基于 FW 机制及 LSTM 的递归网络模型以及所依托的学习方法;

[0017] 其中, 所述基于 FW 机制及 LSTM 的递归网络模型包括数据导入模块、数据生成模块、加载与迭代模块、参数设定模块、模型定义模块、递归网络训练模块、递归网络评估模块以及递归网络测试模块;

[0018] 其中, 数据生成模块又包括数据拆分单元; 加载与迭代模块包括数据加载单元和迭代单元;

[0019] 数据拆分单元包括训练数据生成单元、评估数据生成单元以及测试数据生成单元;

[0020] 递归网络训练模块包括 dropout 单元、更新单元和结果储存单元; 递归网络评估模块以及递归网络测试模块仅包括更新单元和结果储存单元;

[0021] 其中, 更新单元包括长短时记忆单元和快速权重单元;

[0022] 所述基于 FW 机制及 LSTM 的递归网络模型中各模块的连接关系如下:

[0023] 数据导入模块与数据生成模块相连, 数据生成模块和加载与迭代模块相连, 参数设定模块和加载与迭代模块以及模型定义模块相连, 递归网络训练模块和加载与迭代模块、递归网络评估模块以及模型定义模块相连, 递归网络评估模块与加载与迭代模块、递归网络训练模块、递归网络测试模块以及模型定义模块相连; 递归网络测试模块与加载与迭代模块、递归网络评估模块和模型定义模块相连;

[0024] 数据生成模块中各单元的连接关系如下: 数据拆分单元中训练数据、评估数据和测试数据分别与训练标签生成单元、评估标签生成单元以及测试标签生成单元相连;

[0025] 加载与迭代模块中各单元的连接关系如下: 数据加载单元和迭代单元相连;

[0026] 所述基于FW机制及LSTM的递归网络模型中各模块的信号产生及输出关系如下：

[0027] 数据导入模块的输出接入数据生成模块；数据生成模块处理后接入加载与迭代模块；参数设定模块为加载与迭代模块和模型定义模块提供输入参数及FW模型参数；加载与迭代模块分别为递归网络训练模块、递归网络评估模块和递归网络测试模块提供训练数据和训练标签、评估数据和评估标签以及测试数据和测试标签；模型定义模块将FW模型参数分别输入递归网络训练模块、递归网络评估模块以及递归网络测试模块；递归网络训练模块将训练好的网络参数送入递归网络评估模块；递归网络评估模块将评估后的网络参数送入递归网络测试模块；

[0028] 递归网络训练模块、评估模块以及测试模块中的各单元连接关系如下：

[0029] dropout单元接收数据并与长短时记忆单元相连，长短时记忆单元与数据输入和快速权重单元相连，结果储存单元与快速权重单元和结果相连。

[0030] 所述基于FW机制及LSTM的递归网络模型以及所依托的学习方法，包括如下步骤：

[0031] 步骤一、待训练和测试的数据经数据导入模块导入，具体为：

[0032] 通过读取文本路径，获取文本数据；

[0033] 步骤二、数据生成模块对经数据导入模块导入的数据经数据拆分单元进行拆分，分别得到训练数据、评估数据和测试数据；

[0034] 其中，拆分具体为：将步骤一导入的文本数据按照每j个字符为一句话进行拆分；

[0035] 其中，j的取值范围为5到50；

[0036] 步骤三、训练数据生成单元随机选取x%比例经数据拆分单元拆分后的数据生成训练集；评估数据生成单元随机选取y%比例经数据拆分单元拆分后的数据生成评估集；测试数据生成单元随机选取z%比例经数据拆分单元拆分后的数据生成测试集；

[0037] 其中， $x\%+y\%+z\%=1$ ；

[0038] 步骤四、训练标签生成单元将训练数据生成单元生成的训练集中每一个数据后移一位得到训练标签；评估标签生成单元将评估数据生成单元生成的评估集中每一个数据后移一位得到评估标签；测试标签生成单元将测试集中每一个数据后移一位得到测试标签；

[0039] 步骤五、参数设定模块根据数据导入模块导入文本的模型规模，获取配置参数，再将获取的配置参数输入参数设定模块；

[0040] 其中，配置参数包括初始规模、学习率、最大梯度正则值、层数、步数、隐藏层大小、最大epoch数、极大epoch值、dropout率、衰减率、批大小以及vocab大小；

[0041] 步骤六、加载与迭代模块中的数据加载单元按照参数设定模块中获取的配置参数加载训练集、评估集以及测试集中的数据，并设定初始化数据序号i为1；

[0042] 步骤七、模型定义模块根据参数设定模块中的配置参数，使用伪随机函数在配置范围内生成随机值作为权重矩阵参数，完成权重参数的初始化；

[0043] 步骤八、加载与迭代模块中迭代模块判断当前数据集中的数据是否发送完毕，并依据判断结果进行操作，具体为：

[0044] 若当前数据集中的数据没发送完，则发送第i组数据，判断进行训练、评估还是测试，并跳至步骤九，跳至步骤八；否则停止迭代；

[0045] 步骤九、判断当前数据是否是训练数据，若是则依据dropout率对输入数据进行抽取，抽取后数据，跳至步骤十；否则，跳至步骤十；

[0046] 步骤十、将步骤九输入的数据送入更新单元中的长短时记忆单元以及快速权重单元计算得到输出向量,同时利用梯度下降法对网络进行优化,具体为:

[0047] 步骤10.1更新单元基于输入层权重 $W_x$ 、标准权重 $W_h$ 计算起始隐藏层状态,通过公式(4)计算当前 $t$ 时刻的初始隐藏状态:

$$[0048] \quad h_t^0 = f(LN(W_x * x_t + W_h * h_{t-1})) \quad (4)$$

[0049] 其中,输入层权重记为 $W_x$ 、标准权重记为 $W_h$ ;  $h_0$ 为起始隐藏层状态,  $LN$ 为层正则化函数;  $f$ 为激活函数;  $x_t$ 为当前 $t$ 时刻的输入层数据;  $h_{t-1}$ 为当前时刻的前一时刻,即 $t-1$ 时刻,隐藏层状态对应的数据,简称隐藏层状态;

[0050] 优选的,激活函数 $f$ 为SeLU函数、Leaky Relu函数以及Swish函数中的一种;

[0051] 标准权重 $W_h$ 是RNN网络中隐藏层向下一个时间步进传播的权重;输入层权重 $W_x$ 是输入层到隐藏层传播的权重;

[0052] 步骤10.2、快速权重单元计算快速权重,具体通过公式(5)计算:

$$[0053] \quad W_A(t) = \lambda W_A(t-1) + \eta h_{t-1} h_{t-1}^T \quad (5)$$

[0054] 其中,  $W_A(t)$  是第 $t$ 时刻的快速权重,是仅作用在隐藏层每个时间步进内的权重;一个时间步进更新的总次数,记为 $s+1$ 次;  $\lambda$ 是衰减率、 $\eta$ 是学习率、 $h_{t-1}$ 为 $t-1$ 时刻对应的隐藏层状态;  $h_{t-1}^T$ 是 $h_{t-1}$ 即 $t-1$ 时刻对应隐藏层状态的转置;

[0055] 其中,时间步进更新的总次数 $s+1$ 中的 $s$ 即步数;

[0056] 其中,衰减率的取值范围为0.9到0.995,学习率的取值范围为0.3到0.8;

[0057] 步骤10.3、快速权重单元计算隐藏层状态并更新 $s$ 次隐藏层状态;

[0058] 步骤10.4、慢速权重单元计算归一化输出;

[0059] 其中,网络的归一化输出通过Softmax或sigmoid函数两者之一实现;

[0060] 步骤10.5、结果存储单元计算基于步骤10.4计算出的归一化输出计算损失loss和复杂度perplexity;

[0061] 步骤10.6、慢速权重单元判断是否达到最后一个Epoch,如果没有达到,则更新单元则更新隐藏层状态以及训练参数或测试参数,将当前 $i$ 加1,跳至步骤八。

[0062] 有益效果

[0063] 本发明一种基于FW机制及LSTM的递归网络模型及学习方法,与现有技术相比,具有如下有益效果:

[0064] 1. 所述递归网络模型引入快速权重以及LSTM机制,通过衰减系数及学习率的参数优化,使得以储存短期记忆信息的网络模型的学习准确度得到了大幅度提高;

[0065] 2. 本发明所述方法与现有LSTM模型以及引入快速权重的RNN模型相比,模型的训练所述方法采用LSTM结合SeLU激活函数以及层正则化使得训练、评估及测试的收敛速度大大提高。

## 附图说明

[0066] 图1是本发明基于FW机制及LSTM的递归网络模型的组成及各模块的连接示意图;

[0067] 图2是本发明基于FW机制及LSTM的递归网络模型中数据生成模块的组成及连接示意图;

[0068] 图3是本发明基于FW机制及LSTM的递归网络模型中加载与迭代模块的组成示意图



以及与数据生成模块、参数设定模块、模型定义模块、递归网络训练模块、递归网络评估模块和递归网络测试模块的连接关系；

[0069] 图4是本发明基于FW机制及LSTM的递归网络模型中递归网络训练模块、递归网络评估模块以及递归网络测试模块三者的关系与组成示意图；

[0070] 图5是本发明基于FW机制及LSTM的递归网络模型中长短时记忆单元和快速权重单元的组成示意图；

[0071] 图6是本发明基于FW机制及LSTM的递归网络模型依托的方法处理关联度大的短句文本数据集不同batch size的学习效果对比；

[0072] 图7是本发明基于FW机制及LSTM的递归网络模型依托的方法处理关联度大的短句文本数据集不同模型的log(perplexity)对比。

### 具体实施方式

[0073] 下面结合附图和实施例对本发明基于FW机制及LSTM的递归网络模型及学习方法做进一步说明和详细描述。

[0074] 实施例1

[0075] 本实施例阐述了基于本发明所述的基于FW机制及LSTM的递归网络模型的组成及工作流程。

[0076] 具体实施时,语料采用流行应用于自然语言处理的NLTK文本语料库中的富有代表性的短句库——欧盟国家会议语料europarl\_raw进行试验。

[0077] europarl\_raw语料库文本数据来源于会议对话,句子大多数为中短句,长度大约为十个单词左右,句式较为简单,大多为主谓宾结构。具体到本实施例,采用图1中各模块对该数据集进行处理。

[0078] 图1示意了基于FW机制及LSTM的递归网络模型的组成及各模块的连接,从图1中可以看出,数据导入模块导入的数据送入数据生成模块中;数据生成模块生成训练数据、评估数据以及测试数据及其标签,输入加载与迭代模块中;加载与迭代模块与模型定义模块接收参数设定模块的参数,并分别连入递归网络训练模块、评估模块以及测试模块,进行训练、评估及测试。

[0079] 。首先通过数据导入模块通过读取文本路径将文本数据导入;导入后输出至数据生成模块,数据生成模块进一步将原始数据拆分成训练数据、评估数据和测试数据,再经过训练标签生成单元、评估标签生成单元以及测试标签生成单元生成各数据集的标签,其结构如图2中数据生成模块的连接示意图所示。

[0080] 其中,参数设定模块中预先设置的配置参数有如下表1中所述的4种:

[0081] 表1各配置具体参数设置

[0082]

	SmallConfig	MediumConfig	LargeConfig	TestConfig
初始规模	0.1	0.05	0.04	0.1
学习率	1.0	1.0	1.0	1.0
最大梯度正则值	5	5	10	1
层数	2	2	2	1

[0083]

步数	20	35	35	2
隐藏层大小	50	650	1500	2
最大 epoch 数	4	6	14	1
极大 epoch 值	13	39	55	1
dropout 率	1.0	0.5	0.35	1.0
衰减率	0.5	0.8	1/1.15	0.5
批大小	20	20	20	20
vocab 大小	10000	10000	10000	10000

[0084] 参数设定模块依据数据导入模块导入的文本的模型规模,获取由表1所示的合适配置参数,将其输入参数设定模块,尔后送往模型定义模块和加载与迭代模块。模型定义模块根据参数设定模块中的配置参数,使用伪随机函数在配置范围内生成随机值作为权重矩阵参数,完成权重参数的初始化。

[0085] 加载与迭代模块中的数据加载单元按照参数设定模块中获取的配置参数加载训练集、评估集以及测试集中的数据;迭代模块判断当前数据集中的数据是否发送完毕,并依据判断结果进行操作。若当前数据集为训练数据,则输出至递归网络训练模块;若为评估数据,则输出至递归网络评估模块;若为测试数据,则输出至递归网络测试模块。由图3可看出加载与迭代模块的工作示意图以及与数据生成模块、参数设定模块、模型定义模块、递归网络训练模块、递归网络评估模块和递归网络测试模块的连接关系。

[0086] 图4示意了基于FW机制及LSTM的递归网络模型中递归网络训练模块、递归网络评

估模块以及递归网络测试模块三者的关系与组成。递归网络评估模块以及递归网络测试模块与递归网络训练模块差异是不包括dropout单元而仅包含更新单元和结果储存单元。递归网络训练模块将训练好的网络参数送入递归网络评估模块；递归网络评估模块将评估后的网络参数送入递归网络测试模块。

[0087] 从图4可以看出,更新单元包括长短时记忆单元和快速权重单元;图4可以看出,递归网络评估模块以及递归网络测试模块与递归网络训练模块差异是不包括dropout单元;仅包含更新单元和结果储存单元。

[0088] 图5示意了本模型中长短时记忆单元和快速权重单元的组成。图5中, $X_t$ 对应t时刻的输入层数据; $C(t-1)$ 以及 $C'(t)$ 分别对应LSTM在t时刻记忆单元C的输入和输出; $C'(t)$ 再经过快速权重进行更新,生成 $C(t)$ ;作为下一时刻LSTM记忆单元C的输入; $h_{t-1}$ 以及 $h_t$ 分别为t-1时刻以及t时刻的LSTMcell的输出。图5中的 $\sigma$ 为激活函数sigmoid; $\tanh$ 为tanh激活函数。

[0089] 图5中, $C'(t) = h_0(t)$ 以及 $C(t) h_s(t)$ 分别对应初始快速权重更新前的t时刻记忆单元的输入,以及更新后t时刻记忆单元的输入。

[0090] 实施例2

[0091] 本实例阐述了基于本发明递归网络模型所依托的方法,处理关联度大的断句文本数据集的学习效果对比。

[0092] 我们将目光转移到对由句子之间关联性较强,且句子长度较短的文本数据的处理,由于句子较短,更加注重短期内输入单词与单词之间的联系。我们使用流行应用于自然语言处理的NLTK文本语料库中的富有代表性的短句库——欧盟国家会议语料europarl\_raw进行试验。

[0093] 在使用europarl\_raw语料库时,将num\_steps统一设置为10,代表网络按照每输入十个单词为一句完整的句子处理。

[0094] 首先需确定合适的更新次数s。

[0095] 当快速权重在当前的时刻得到更新后,将对隐藏状态进行循环s次的更新,相比于toy game场景的样本数据,文本数据的前后单词关联性较为复杂,我们需要加快更新频率,即加大s的数值以更大的发挥快速权重处理短期记忆的功能。我们调整一个时间步骤内隐藏状态的更新次数,固定隐藏单元数为50, batch\_size为20,改变S=5,6,7,8,记录模型训练效果,如下表2所示:

[0096] 表2不同更新次数下模型分别在训练到第5,10,13个epoch时的复杂度对比

[0097]

更新次数s	复杂度-5	复杂度-10	复杂度-13
5	189.380	108.083	105.231
6	145.939	73.875	71.331
7	138.889	68.323	65.946
8	139.400	70.049	67.642

[0098] 如表2所示,当更新次数s=7时,快速权重模型在训练到第5个epoch时复杂度为138.889,第10个epoch下降为68.323,第13个epoch时收敛于65.946。

[0099] 此后我们将确定合适的batch size。

[0100] 合适的batch size对于一个网络的学习性能至关重要,batch size过大,会导致模型在进行梯度下降法寻找最优解时找到的是局部最小值而不是全局最小值,而batch size过小则会导致收敛速度慢,模型学习效果差。所以为了提升引入了快速权重的新模型的性能,我们固定隐藏单元数量为50,更新次数s设置为前文中验证过的最优值7,组成句子的单词数num\_steps=10,改变batch size等于10,20,30,50,记录模型训练效果,如下表3所示:

[0101] 表3不同batch size下模型在第10个epoch时的复杂度对比

[0102]

Batch size	复杂度-5	复杂度-10	复杂度-13
10	98.214	53.852	51.430
<b>20</b>	<b>90.945</b>	<b>45.139</b>	<b>43.344</b>

[0103]

30	108.475	53.599	51.604
50	138.889	68.323	65.946

[0104] 从表3可以看到,batch\_size=20时模型收敛后的复杂度最低,在训练到第10个epoch时复杂度为45.139,在训练到第13个epoch时复杂度低至43.344,batch size等于10和30时,模型训练到第13个epoch的复杂度约为51。为了更直观地表示出不同batch size下复杂度的差异,我们对复杂度取以10为底的对数log(perplexity),对比不同batch size下快速权重模型的log(perplexity)差异,如图6所示。

[0105] 图6中,横坐标为训练epoch数,纵坐标为以10为底,复杂度的对数log(perplexity)。可以看到以每20个epoch为一批数据进行训练时模型的复杂度最低,学习效果最佳,并进行语言模型的对比。

[0106] 固定隐藏单元数量为50,组成句子的单词数num\_steps=10,使用SeLU函数作为激活函数。对比LSTM模型,RNN模型,快速权重与LSTM网络结合的模型和快速权重与RNN结合的模型共四个模型的训练效果。模型训练复杂度如表4所示:

[0107] 表4:不同模型基于europarl\_raw数据库的训练复杂度对比

[0108]

模型名称	复杂度-5	复杂度-10	复杂度-15	复杂度-20
LSTM	267.602	178.175	174.935	174.824
LSTM+FW	90.945	45.139	43.280	43.208
RNN	1037.719	421.531	412.841	412.510
RNN+FW	533.806	378.564	369.842	369.474

[0109] 从表4能够看到,引入快速权重的LSTM模型在训练到第5个epoch时复杂度为90.945,在第10个epoch时进一步降低至45.139.训练到第15个epoch时模型复杂度为

43.280,模型达到收敛。同样训练到第15个epoch时,LSTM模型的复杂度收敛至174.824,比引入快速权重的LSTM模型高出131,其次是引入快速权重的RNN网络,复杂度收敛于369.474,效果最差的是RNN模型,复杂度收敛于412.510。

[0110] 为了更直观地表示不同模型的复杂度差异,将复杂度取以10为底的对数,对比不同模型的 $\log(\text{perplexity})$ 差异,如图7所示。

[0111] 从图7能够看出,引入了快速权重的LSTM模型收敛后的复杂度最低,模型学习效果最好,且与未引入快速权重的LSTM模型差异很大,说明在LSTM网络中引入快速权重,模型训练效果提升明显。RNN模型收敛后的复杂度最高,加入快速权重后的RNN模型训练效果略有提升,但效果不明显。

[0112] 以上所述为本发明的较佳实施例而已,本发明不应该局限于该实施例和附图所公开的内容。凡是不脱离本发明所公开的精神下完成的等效或修改,都落入本发明保护的范围。

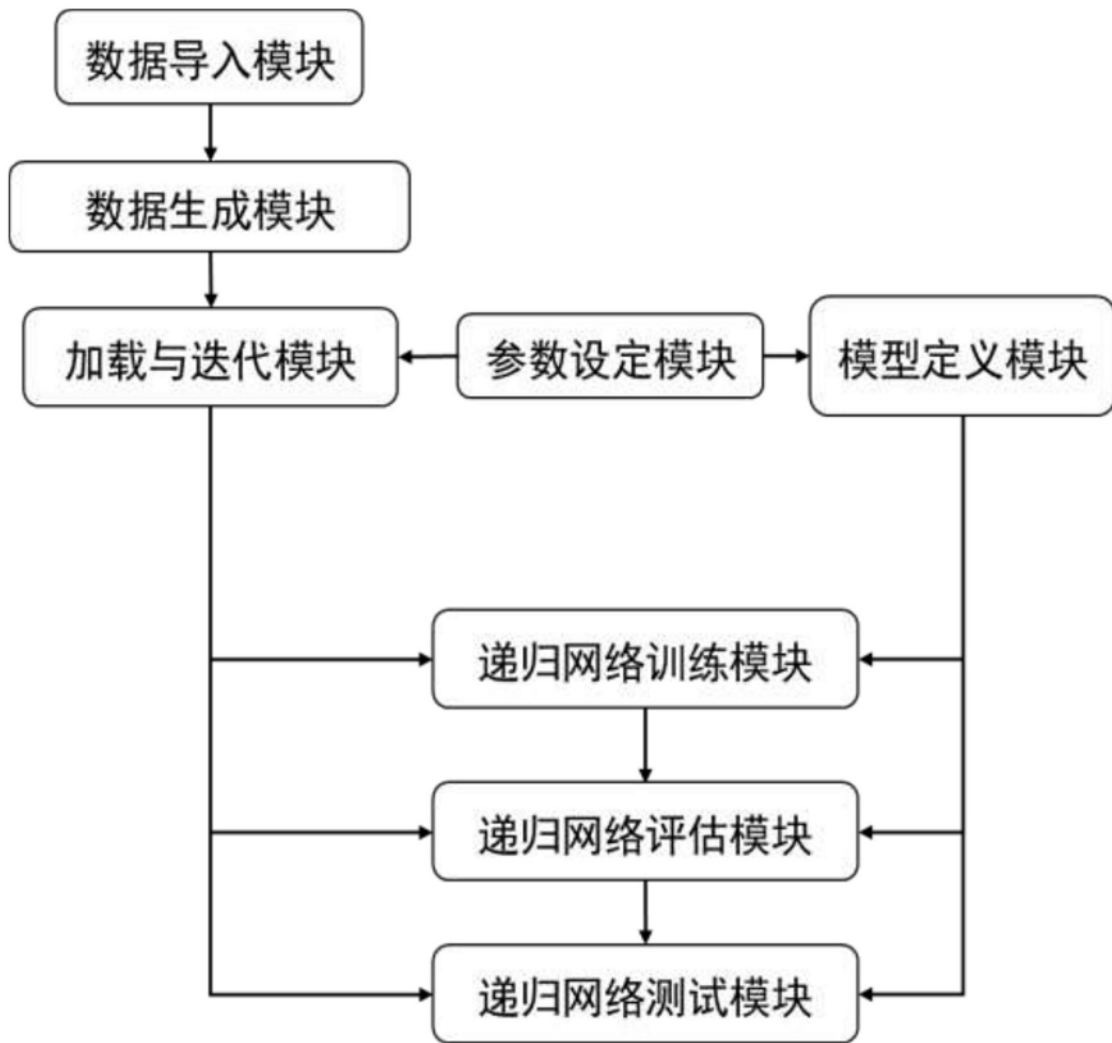


图1

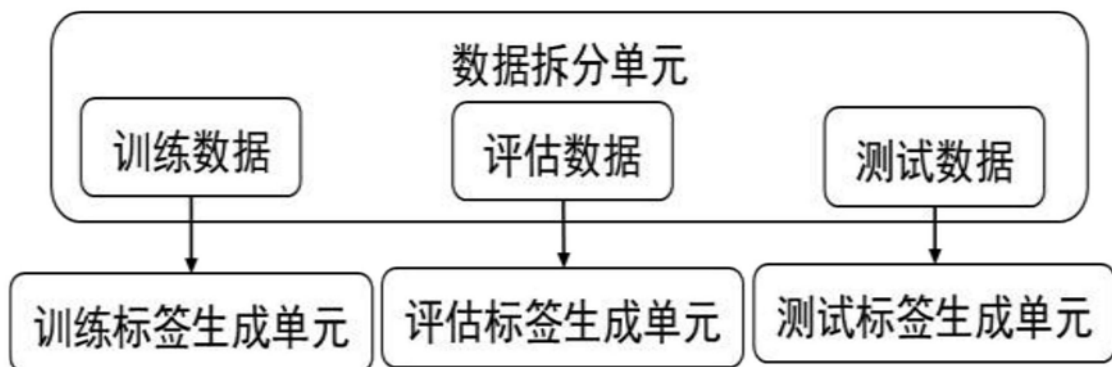


图2



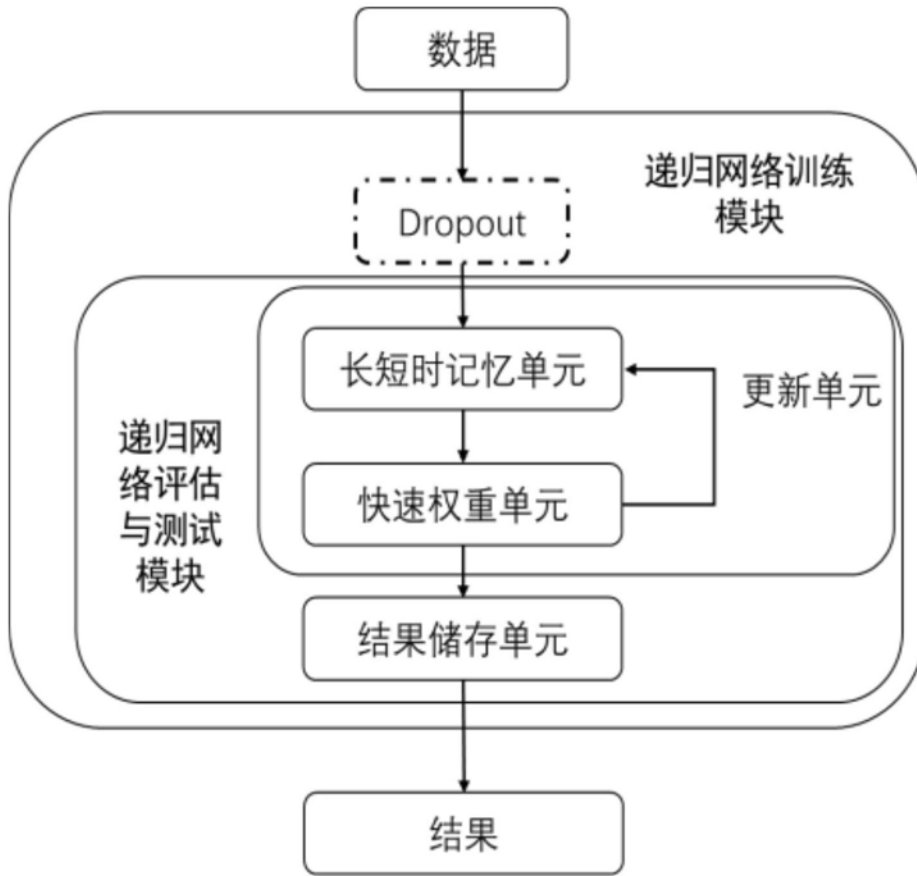


图4



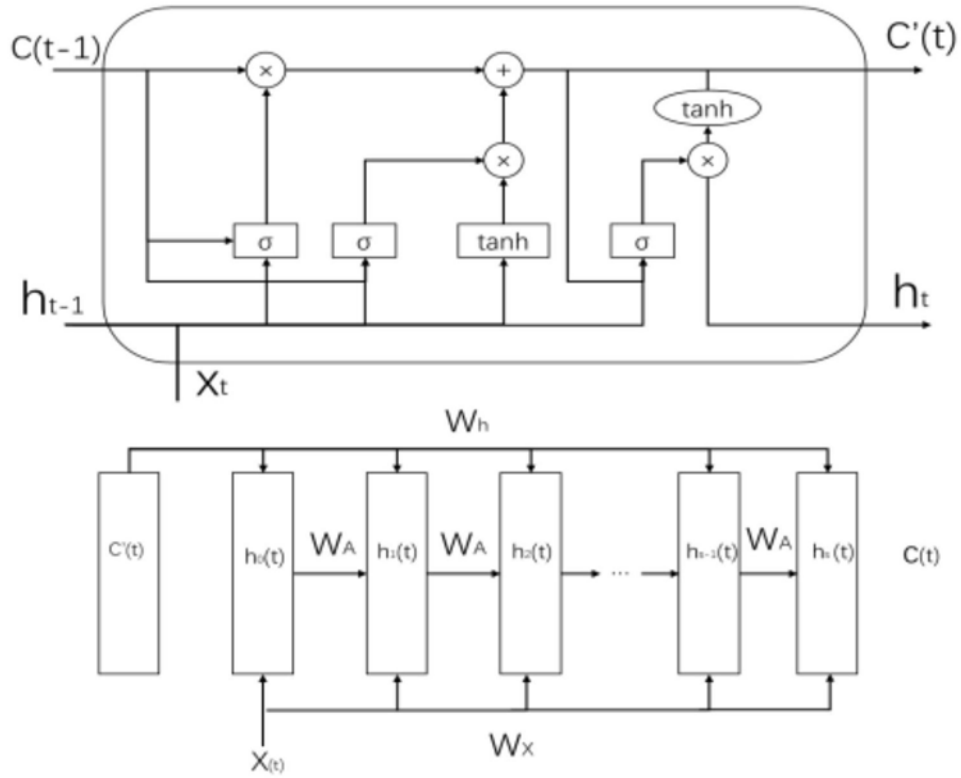


图5

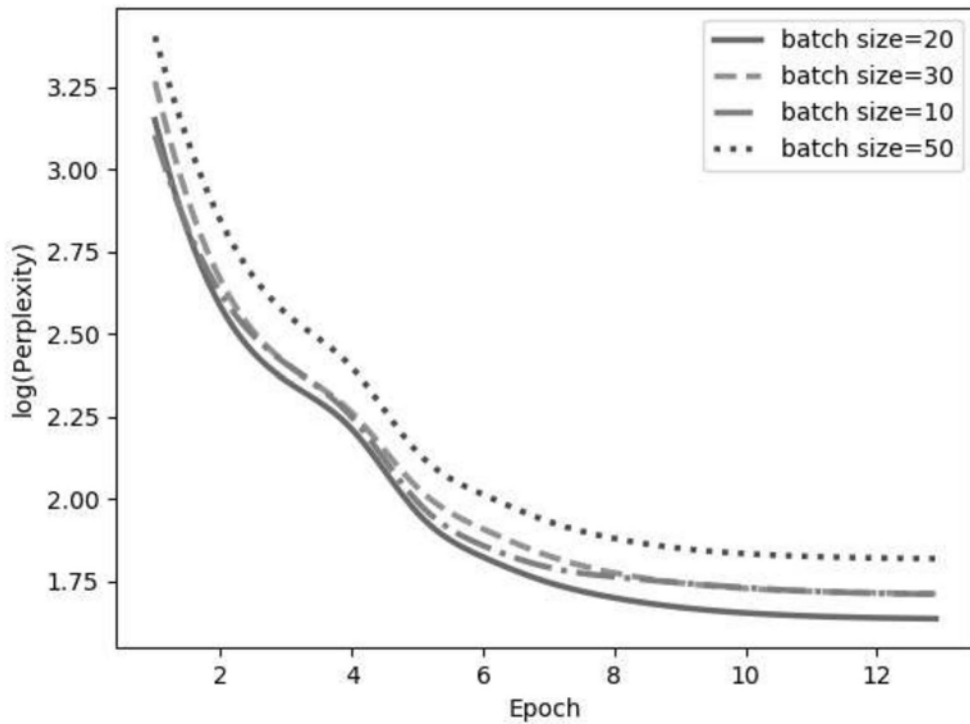


图6

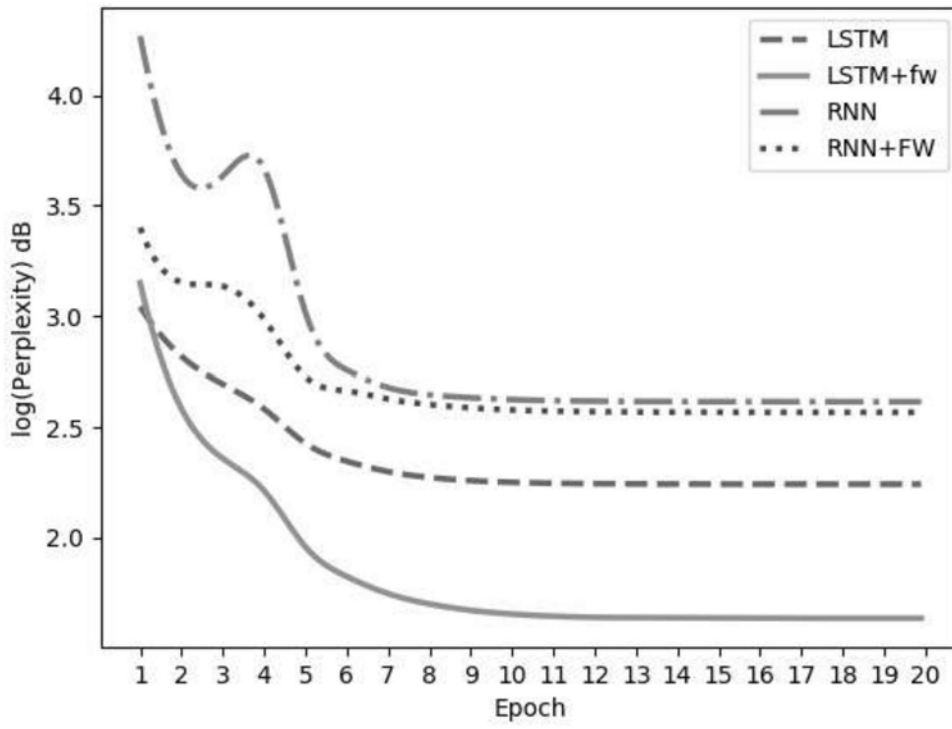


图7