(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 116090010 A
(43)申请公布日 2023.05.09

(21)申请号 202310121433.1

(22)申请日 2023.02.01

(71)申请人 北京邮电大学
地址 100876 北京市海淀区西土城路10号

(72)发明人 杨震 徐子涵 张茹 张鹏

(74)专利代理机构 北京永创新实专利事务所 11121
专利代理师 周长琪

(51)Int.Cl.
G06F 21/62(2013.01)
G06F 40/211(2020.01)
G06F 40/30(2020.01)
G06F 40/289(2020.01)
G06N 3/084(2023.01)
G06N 3/044(2023.01)

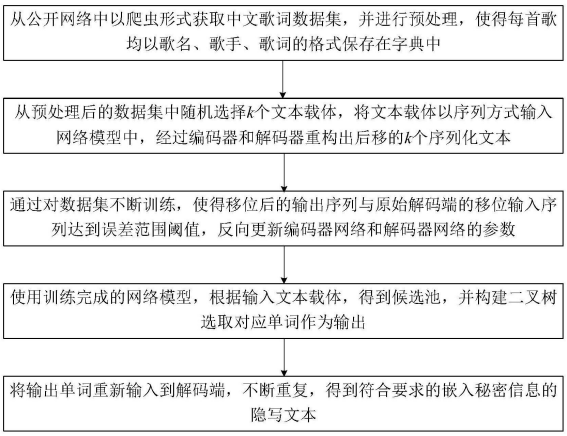
权利要求书3页 说明书7页 附图3页

(54)发明名称

基于上下文语境联系的文本生成式隐写方法

(57)摘要

本发明公开了一种基于上下文语境联系的文本生成式隐写方法，属于信息隐藏技术领域。本发明方法包括：从公开网络中获取中文流行歌词数据集，将每首歌以约定格式保存，从中随机选择k个文本载体进行数字序列化后，按序输入文本隐写网络模型，该模型包括编码端和解码端，使用了门限递归单元和注意力模块，输出重构的隐写文本；通过不断训练，使得重构的隐写文本与原始文本之间的误差不断缩小，反向更新编码器和解码器的参数；使得训练后的模型生成嵌入有秘密信息的隐写文本。本发明生成隐写文本时考虑了文本的序列化特点、文本间上下文语境联系、文本载体的统计特性和语义特性等，使得隐写文本隐蔽性更强，实现了高质量的文本隐写和秘密信息提取。



CN 116090010 A

1.一种基于上下文语境联系的文本生成式隐写方法,其特征在于,包括如下步骤:

步骤一、获取中文歌词数据集,将每首歌以歌名、歌手和歌词的格式保存;

步骤二、从中文歌词数据集中随机选取k个文本载体,每个文本载体为一句歌词或者一首歌的歌名;将文本载体转换为序列信号输入文本隐写网络模型中,经过编码器和解码器重构出后移的k个序列化文本;k为正整数;

所述文本隐写网络模型包括编码器和解码器,编码器网络包括嵌入层和门限递归单元GRU层,解码器网络包括嵌入层、GRU层、注意力模块、线性拼接层、线性层和Softmax输出层;将第i个文本载体的序列信号输入编码器获得编码器输出向量$E_{i,o}$和解码器GRU层初始隐状态,将第i+1个文本载体的序列信号输入解码器,将$E_{i,o}$和解码器的GRU层的输出向量$D_{i,o}$输入到注意力模块,得到考虑到句内不同词权重影响的修正输出向量$D'_{i,o}$;将修正输出向量$D'_{i,o}$与原有输出向量$D_{i,o}$经过线性拼接层、线性层和Softmax输出层,得到移位输出序列;

步骤三、对文本隐写网络模型进行训练,使得移位输出序列与解码器的原始移位输入序列达到误差范围阈值,更新编码器网络和解码器网络的参数;

步骤四、使用训练后的文本隐写网络模型生成载秘文本;

随机获取一个内容为歌名的文本载体作为编码端输入,解码端输入句首标识符,经文本隐写网络模型获取一个d_o维输出向量;该输出向量中元素按概率从高到低排序,每一维度元素对应语料库中的一个单词,根据预设嵌入比特数b,选取输出向量中前$K=2^b$维度元素所对应的单词构建候选池;根据秘密信息二进制比特流从候选池中选取单词输出;将输出的单词添加到解码端输入,继续利用文本隐写网络模型生成输出向量,重复上述单词选取过程,不断生成隐写文本。

2.根据权利要求1所述的方法,其特征在于,所述的步骤二中,将文本载体转换为序列信号,包括:按顺序取k个文本载体,对每个文本载体进行分词处理,获取每个单词在语料库词汇表中对应的ID,然后将文本载体用固定长度为L的序列信号表示,序列信号中元素为单词对应的ID;当文本载体包含的单词数量小于L时,序列信号中不足位置填充0,否则截取文本载体的前L个单词。

3.根据权利要求1或2所述的方法,其特征在于,所述的步骤二中,文本隐写网络模型对输入的k个序列信号执行如下处理:

(1)将k个序列信号分别经过编码器的word2vec词嵌入层后,得到k个文本载体的序列信号矩阵;

获取预训练完的word2vec词嵌入矩阵C,设第i个文本载体的序列信号为f_i,对f_i中每个元素在词嵌入矩阵C中寻找对应的d维词向量,得到第i个文本载体的序列信号矩阵c_i;

(2)将第i个文本载体的序列信号矩阵输入到编码器的GRU层,得到编码器输出向量和解码器的GRU层的初始隐状态;

设第i个文本载体的序列信号矩阵$c_i=(x_1,\cdots,x_t,\cdots,x_L)$,L为序列信号长度,对应L个时刻;GRU层的计算公式如下:

$z_t=\sigma(W_z[h_{t-1},x_t])$

$r_t=\sigma(W_r[h_{t-1},x_t])$

$\widetilde{h_t}=\tanh(W[r_t\odot h_{t-1},x_t])$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \widetilde{h_t}$$

其中，h_{t-1}和h_t分别为t-1时刻和t时刻更新门计算的隐状态向量，$\widetilde{h_t}$表示基于重置门计算的隐状态向量；z_t为更新门输出，r_t为重置门输出，W_z、W_r、W分别为更新门、重置门、向量$\widetilde{h_t}$的训练参数；初始h_0置为0；

得到编码器输出向量$E_{i,o}=(h_1,\cdots,h_t,\cdots,h_L)$以及解码器GRU层初始隐状态$E_{i,h}=h_L$，$h_L$为编码器GRU层中L时刻的隐状态输出；

(3)将第i+1个文本载体的序列信号输入解码器，经解码器嵌入层得到序列信号矩阵c_{i+1}并输入GRU层，解码器GRU层的初始隐状态为$E_{i,h}$，经GRU层后得到输出向量$D_{i,o}$；

(4)将编码端输出向量$E_{i,o}$和解码器的GRU层的输出向量$D_{i,o}$输入到注意力模块，得到考虑到句内不同词权重影响的修正输出向量，包括：

设第i+1个文本载体经过解码端GRU层后的输出向量$D_{i,o}=(h'_1,\cdots,h'_t,\cdots,h'_L)$，计算$D_{i,o}$中的第t个信号与编码器输出向量$E_{i,o}$的相关系数$e_t$，如下：

$e_t=[a(h'_t,h_1),a(h'_t,h_z),..,a(h'_t,h_L)]$

$a(h'_t,h_i)=\tanh(W_1h_i+W_2h'_t)$

其中，a为一个线性层函数，W_1，W_2为权重参数矩阵；经过softmax激活函数后，得到相关系数e_t中每一个维度对应的权重系数α_{t_p}，如下：

$$\alpha_{t_p}=\frac{\exp\left(e_{t_p}\right)}{\sum_{j=1}^{N}\exp\left(e_{t_j}\right)}$$

融合$E_{i,o}$及权重系数α_{t_p}，得到新向量c_t如下：

$$c_t=\sum_{p=1}^{N}\alpha_{t_p}h_p$$

最后，得到与$D_{i,o}$维数相同的带有影响权重的矩阵向量$D'_{i,o}=(c_1,\cdots,c_t,\cdots,c_L)$。

(5)将新的输出向量$D'_{i,o}$与原有输出向量$D_{i,o}$经过线性拼接层、线性层以及Softmax输出层，得到重构后的移位输出序列，生成隐写文本序列信号；

在线性拼接层将$D'_{i,o}$和$D_{i,o}$直接拼接，得到重构向量$D_{i,concat}$；在全连接线性层对$D_{i,concat}$降维处理；在Softmax输出层经softmax激活函数得到重构后的移位输出序列。

4.根据权利要求3所述的方法，其特征在于，所述的步骤三中，设文本隐写网络模型输出的移位输出序列为$f'_i=(f_i(2),\cdots,f_i(t),\cdots,f_i(L))$，i=1,…,k；解码器的原始移位输入序列表示为$f'_{i+1}=(f_{i+1}(2),\cdots,f_{i+1}(t),\cdots,f_{i+1}(L))$，i=1,…,k；利用交叉熵损失函数计算每一单个样本的损失值，再进行叠加求和得到总的损失值。

5.根据权利要求1所述的方法，其特征在于，所述的步骤四中，对候选池中的单词构建完全二叉树或考虑概率差异的哈夫曼树，每个叶子节点对应一个单词，根据秘密信息二进制比特流从二叉树或哈夫曼树中选取对应的单词输出。

6.根据权利要求1或5所述的方法，其特征在于，所述的步骤四中，包括：

(1)将输出的单词添加到解码端输入序列中，重新输入文本隐写网络模型，编码端输入不变，利用文本隐写网络模型生成输出向量，根据秘密信息二进制比特流选取下一个单词，

重复该过程,直到选取的单词为句尾标识符,得到一句完整的隐写文本;

(2)将整句隐写文本作为编码端的输入,句首标识符作为解码端输入,重新输入文本隐写网络模型,重复上述(1)的过程,得到新的隐写文本,并不断重复,直到自动得到一首完整的歌曲或达到预先设定好的歌词总句数。

7.根据权利要求1所述的方法,其特征在于,所述的方法,在对生成的载秘文本进行秘密信息提取时,将生成载密文本的关键句以及句首标识符输入到文本隐写网络模型中,重构出秘密信息二进制比特流。

8.根据权利要求1所述的方法,其特征在于,所述的步骤一中,从公开网络中搜集获取中文歌词数据集,并对每首歌进行预处理以使得数据集中文本格式一致。

基于上下文语境联系的文本生成式隐写方法

技术领域

[0001] 本发明属于信息隐藏技术领域,涉及文本隐写,具体是一种基于上下文语境联系的文本生成式隐写方法。

背景技术

[0002] 信息隐藏可以在不改变公开媒体中信息感知性的前提下将秘密信息嵌入其中,然后通过在公开信道上传递嵌有秘密信息的载体来完成秘密信息的传输。信息隐藏除却隐藏秘密信息和通信行为外,还可以用于数字水印,来解决诸如版权保护和篡改鉴定之类的需求。密码学只隐藏信息的内容,使得加密后的信息晦涩难懂,与此不同的是,信息隐藏不仅隐藏信息的内容,还隐藏信息的存在,一定程度上提高了秘密信息的安全性。信息隐藏有着图像、视频、语音和文本等多种数字媒体形式的载体,但目前大多数的信息隐藏研究与图像和视频相关。

[0003] 在计算机时代,信息隐藏不仅要让人的肉眼无法辨别,还要骗过计算机的识别。信息隐藏的一个分支——隐写术,通过改变载体的空间信息或变换域系数来嵌入信息,这样可能会造成统计指标上的显著差异。而隐写分析方法主要是拟合统计特征来进行分析。随之,隐写方法通过不断改变嵌入方式来尽量减少修改带来的统计特征的变化,使得对载体的修改不易被发现;但同时隐写分析方法要不断构造更有效的统计特征来对载体进行分析。两者之间不断促进前行,随着时间的推移,两者的发展逐渐来到瓶颈期,隐写方法很难再利用传统的方式找到更加优秀的嵌入方法以保证不会被隐写分析方法检测出来。

[0004] 关于衡量隐写方法的性能指标,现有的工作主要考虑隐藏效率、不可感知性、隐藏容量、鲁棒性和复杂性等方面。隐藏效率指生成一定数量的嵌入秘密信息的载体所需要的时间;不可感知性指嵌入秘密信息的载体与原始载体的差异性,差异性越大,不可感知性越差,可以采用perplexity(PPL,困惑度)、Earth's Mover's Distance(EMD,陆地移动距离)等不同指标来衡量;隐藏容量指嵌入到载体中的秘密信息量,采用嵌入率(Embedding Rate,ER)进行评价;鲁棒性则是指嵌入秘密信息的载体受到干扰时,其中的秘密信息是否还能够恢复出来;复杂性则是指运行该隐写模型所需要的资源,该指标对数据量大、实时性高的一些场景很重要。

[0005] 深度学习的崛起,助力了各行各业的快速发展,尤其是卷积神经网络(CNN)与隐写术及隐写分析的简单结合尝试取得了成功,让专家们思考能否将更多的深度学习技术应用于隐写术当中。通过深度学习,隐写算法可以摆脱很多原始数据预处理时的专家知识,并将隐写分析算法的对抗加入到隐写模型的训练过程中,使得隐写算法具有更高的安全性。

[0006] 但就目前而言,基于深度学习的信息隐藏模型大都是围绕图像展开的,一方面是由于图像具有较大的冗余空间来隐藏信息,另一方面图像处理工具较多且隐藏效果很直观,而且卷积神经网络初始即为图像而设计的。而文本作为最广泛使用的媒体,是一种很有发展潜力的载体对象,但文本信息冗余量较少的特点使得文本隐写算法研究目前仍有很多空白部分,因此研究基于深度学习,尤其是自然语言处理的文本隐写方法也是很有必要的。

发明内容

[0007] 针对上述问题，本发明以RNN(循环神经网络)为基础，基于编码器-解码器架构，考虑到Transformer中的注意力机制，提出了一种基于上下文语境联系的文本生成式隐写方法，实现对文本进行隐写，自动生成嵌入有秘密信息二进制比特流的隐写文本，并能高质量地提取秘密信息。

[0008] 本发明提供的基于上下文语境联系的文本生成式隐写方法，包括如下步骤：

[0009] 步骤一、获取中文歌词数据集，将每首歌以歌名、歌手和歌词的格式保存；

[0010] 步骤二、从中文歌词数据集中随机选取k个文本载体，每个文本载体为一句歌词或者一首歌的歌名；将文本载体转换为序列信号输入文本隐写网络模型中，经过编码器和解码器重构出后移的k个序列化文本；k为正整数；

[0011] 文本隐写网络模型包括编码器和解码器，编码器网络包括嵌入层和门限递归单元GRU层，解码器网络包括嵌入层、GRU层、注意力模块、线性拼接层、线性层和Softmax输出层；将第i个文本载体的序列信号输入编码器获得编码器输出向量$E_{i,o}$和解码器GRU层初始隐状态，将第i+1个文本载体的序列信号输入解码器，将$E_{i,o}$和解码器的GRU层的输出向量$D_{i,o}$输入到注意力模块，得到考虑到句内不同词权重影响的修正输出向量$D'_{i,o}$；将修正输出向量$D'_{i,o}$与原有输出向量$D_{i,o}$经过线性拼接层、线性层和Softmax输出层，得到移位输出序列；

[0012] 步骤三、对文本隐写网络模型进行训练，使得移位输出序列与解码器的原始移位输入序列达到误差范围阈值，更新编码器网络和解码器网络的参数；

[0013] 步骤四、使用训练后的文本隐写网络模型生成载秘文本；

[0014] 随机获取一个内容为歌名的文本载体作为编码端输入，解码端输入句首标识符，经文本隐写网络模型获取一个d_o维输出向量；该输出向量中元素按概率从高到低排序，每一维度元素对应语料库中的一个单词，根据预设嵌入比特数b，选取输出向量中前$K=2^b$维度元素所对应的单词构建候选池；根据秘密信息二进制比特流从候选池中选取单词输出；将输出的单词添加到解码端输入，继续利用文本隐写网络模型生成输出向量，重复上述单词选取过程，不断生成隐写文本。

[0015] 所述步骤二中，将文本载体转换为序列信号，包括：按顺序取k个文本载体，对每个文本载体进行分词处理，获取每个单词在语料库词汇表中对应的ID，然后将文本载体用固定长度为L的序列信号表示，序列信号中元素为单词对应的ID；当文本载体包含的单词数量小于L时，序列信号中不足位置填充0，否则截取文本载体的前L个单词。

[0016] 所述步骤二中，文本隐写网络模型对输入的k个序列信号执行如下处理：(1)将k个序列信号分别经过编码器的word2vec词嵌入层后，得到k个文本载体的序列信号矩阵；(2)将第i个文本载体的序列信号矩阵输入到编码器的GRU层，得到编码器输出向量和解码器的GRU层的初始隐状态；(3)将第i+1个文本载体的序列信号输入解码器，经解码器嵌入层得到序列信号矩阵c_{i+1}并输入GRU层，解码器GRU层的初始隐状态为$E_{i,h}$，经GRU层后得到输出向量$D_{i,o}$；(4)将编码端输出向量$E_{i,o}$和解码器的GRU层的输出向量$D_{i,o}$输入到注意力模块，得到考虑到句内不同词权重影响的修正输出向量；(5)将新的输出向量$D'_{i,o}$与原有输出向量$D_{i,o}$经过线性拼接层、线性层以及Softmax输出层，得到重构后的移位输出序列，生成隐写文本序列信号。

[0017] 本发明的优点与积极效果在于：本发明基于上下文语境联系的文本生成式隐写方

法,从人类感觉系统来看,考虑到了文本的序列化特点、文本间的上下文语境联系、文本载体的统计特性和语义特性等,使得生成的嵌入秘密信息的隐写文本与原始文本之间的统计差异性和语义差异性都较小,对于人类来说实际上也是区分性不大,实现了高质量的文本隐写。本发明方法使用交叉熵损失函数训练文本隐写网络模型,使得移位后的输出序列与原始解码端的移位输入序列达到误差范围阈值,从而实现有效的秘密信息提取。

附图说明

[0018] 图1为本发明基于上下文语境联系的文本生成式隐写方法的流程图;

[0019] 图2为本发明基于上下文语境联系的文本生成式隐写网络的训练和测试过程示意图;

[0020] 图3为本发明的网络模型内部结构示意图;

[0021] 图4为本发明的网络模型内部参数示意图;

[0022] 图5为采用本发明文本生成式隐写方法嵌入秘密信息后生成的隐写文本示意图;

[0023] 图6为本发明模型在文本隐写上的表现示意图。

具体实施方式

[0024] 下面将结合附图和实施例对本发明作进一步的详细说明。

[0025] 为了详细说明本发明的特点和优越之处,下面将从训练到应用的全流程对本发明作实际应用说明。

[0026] 现有循环神经网络可以利用自身良好的序列化建模特点,基于编码器-解码器架构,在大量文本载体上进行训练,完成后自动生成嵌入秘密信息二进制比特流的隐写文本。本发明以此为基础,考虑到Transformer中的注意力机制,设计了基于上下文语境联系的文本生成式隐写方法,可以对文本进行隐写,自动生成嵌入有秘密信息二进制比特流的隐写文本,并能高质量地提取。

[0027] 如图1所示,本发明基于上下文语境联系的文本生成式隐写方法,分如下五个步骤说明。

[0028] 步骤一、从公开网络中搜集得到中文歌词数据集,并进行预处理,使得每首歌均以歌名、歌手、歌词的格式保存在字典中。

[0029] 本发明实施例中,中文歌词数据集从网络上选取公开的流行华语乐曲,为便于隐写模型的训练,需要确保数据集中文本格式一致;如果不满足上述要求,则需要进行数据预处理以达到要求。

[0030] 步骤二、从预处理后的中文歌词数据集中随机选择多首歌曲中的包含歌名和歌词在内的共k个文本载体,其中每个文本载体为一句歌词或者一首歌的歌名,将文本载体以序列方式输入文本隐写网络模型中,经过编码器和解码器重构出后移的k个序列化文本。

[0031] 如图2和图3所示,本发明的文本隐写网络模型,可称为T-GRU网络模型,包括编码端网络E和解码端网络D。在训练过程中,将第i个文本载体输入编码端网络E,将第i+1个文本载体输入解码端网络D,生成对应于编码端输入文本的重构后的移位输出文本,通过比较重构文本与解码端输入文本的差异,使用交叉熵损失函数计算损失值进而反向传播梯度,不断训练,减小二者直接的差异。在测试过程中,除了梯度不反向传播外,其他步骤均与训

练过程保持一致，最后经过损失函数计算得到损失值。

[0032] 如图3所示，本发明的文本隐写网络模型的编码端网络E包括嵌入层和门限递归单元(Gated Recurrent Unit，GRU)，解码端网络D包括嵌入层、GRU、注意力模块、线性拼接层、线性层和Softmax输出。本发明的创新之处在于将RNN变体GRU与Seq2Seq序列模型进行融合，构建了新的模型，同时还在模型中提出了注意力模块，在生成隐写文本时不仅考虑句内单词对后续生成的影响，更考虑了上下文句子间的联系，使得隐写文本隐蔽性更强。

[0033] 如图4所示，为本发明网络模型的网络参数。

[0034] 步骤201，按顺序取k个文本载体，并分别进行分词处理，形成序列信号；

[0035] 第i个文本载体的序列信号如下：

[0036] $f_i=(f_i(1),\cdots,f_i(t),\cdots,f_i(L)),i=1,\cdots,k$；

[0037] L表示序列化后预先设定的固定长度，N表示编码端中该段文本所包含的单词数量。若N<L，则后续L-N个位置处填充为0；若N≥L，则取前L个单词。$f_i(t)$表示第t个单词在语料库词汇表中对应的ID。

[0038] 本发明方法将输入的中文文本载体进行分词处理，对于分词后的每个单词在语料库词汇表中查询其对应的ID，将单词替换为数字ID，得到文本载体对应的数字化序列信号。

[0039] 步骤202，将k个序列信号分别经过编码端的word2vec词嵌入层后，得到k个文本载体的序列信号矩阵；

[0040] 第i个文本载体的序列信号矩阵c_i，获取过程如下：

[0041] 首先，加载在大规模语料库上预训练完的word2vec词嵌入矩阵$\boldsymbol{C}\in\mathbb{R}^{n\times d}$，其中n为语料库的单词数量，d为每个单词对应的向量化维度，即嵌入维度。

[0042] 然后，对于序列信号f_i中的每一个单词$f_i(t)$，在词嵌入矩阵C中寻找对应的d维词向量，得到序列信号f_i对应的序列信号矩阵$\boldsymbol{c}_i\in\mathbb{R}^{L\times d}$。

[0043] 步骤203，将序列信号矩阵输入到编码端的门限递归单元GRU层，得到各自对应的编码端输出向量和解码端GRU层的初始隐状态。

[0044] 第i个文本载体的编码端输出向量$E_{i,o}$以及解码端GRU层的初始隐状态$E_{i,h}$，获取过程如下：

[0045] 首先，根据GRU的内部构造得到理论计算公式：

[0046] $z_t=\sigma(W_z[h_{t-1},x_t])$，

[0047] $r_t=\sigma(W_r[h_{t-1},x_t])$，

[0048] $\widetilde{h_t}=\tanh(W[r_t\odot h_{t-1},x_t])$，

[0049] $h_t=(1-z_t)\odot h_{t-1}+z_t\odot\widetilde{h_t}$.

[0050] 其中，x_t为t时刻的输入向量，h_{t-1}和h_t分别为t-1时刻和t时刻更新门计算的隐状态向量，$\widetilde{h_t}$表示基于重置门计算的隐状态向量。z_t为更新门输出，r_t为重置门输出，$W_z\in\mathbb{R}^{d\times d}$、$W_r\in\mathbb{R}^{d\times d}$、$W\in\mathbb{R}^{d\times d}$分别为更新门$z_t$、重置门$r_t$、向量$\widetilde{h_t}$训练时的习得参数。长期依赖记忆能力通过重置门$r_t$、更新门$z_t$来保证。

[0051] 令$c_i=(x_1,\cdots,x_t,\cdots,x_L)$，对应包含L个时刻的输入，初始$h_0$置为0，据此，可得到编码端输出向量$\boldsymbol{E}_{i,o}\in\mathbb{R}^{L\times d_h}$以及解码端GRU层的初始隐状态$E_{i,h}=h_L$；$E_{i,o}=(h_1,\cdots,h_t,\cdots,$

h_L)；其中，d_h为经过GRU层后的输出维度，h_L为GRU层中L时刻的隐状态输出。

[0052] 步骤204，与上述步骤类似，在解码端输入第i+1个文本载体的序列信号，经过GRU层后得到输出向量$D_{i,o}$；

[0053] 获取过程如下：

[0054] 首先，对解码端输入的第i+1个文本载体分词处理，与步骤201相同，得到序列化输入信号：

[0055] $f_{i+1}=(f_{i+1}(1),\cdots,f_{i+1}(t),\cdots,f_{i+1}(L)),i=1,\cdots,k$；

[0056] 经过word2vec词嵌入层，与步骤202相同，得到第i+1个文本载体的序列信号矩阵c_{i+1}。

[0057] 然后，在经过解码端的GRU层时，此时的初始隐状态h_0并非置为0，而是置为$E_{i,h}$，据此，可得到解码端GRU层的输出向量$\boldsymbol{D}_{i,o}\in\mathbb{R}^{L\times d_h}$。

[0058] 步骤205，将编码端输出向量和解码端经过GRU层后的输出向量输入到注意力模块，得到考虑到句内不同词权重影响的修正输出向量。

[0059] 获取过程如下：

[0060] 已知第i个文本载体经过编码端后的输出向量可表示为$E_{i,o}=(h_1,\cdots,h_t,\cdots,h_L)$，第i+1个文本载体经过解码端GRU层后的输出向量可表示为$D_{i,o}=(h'_1,\cdots,h'_t,\cdots,h'_L)$。因此，对于$D_{i,o}$中的第t个信号，可以得到其与编码端输出向量的相关系数e_t：

[0061] $e_t=[a(h'_t,h_1),a(h'_t,h_2),..,a(h'_t,h_L)]$

[0062] $a(h'_t,h_i)=\tanh(W_1h_i+W_2h'_t)$

[0063] 其中，a为一个线性层函数，权重参数矩阵$W_1,W_2\in\mathbb{R}^{d_h\times d_h}$。

[0064] $$\alpha_{t_p}=\frac{\exp\left(e_{t_p}\right)}{\sum_{j=1}^{N}\exp\left(e_{t_j}\right)}$$

[0065] $$c_t=\sum_{p=1}^{N}\alpha_{t_p}h_p$$

[0066] 经过softmax激活函数后，得到相关系数e_t中每一个维度对应的权重系数α_{t_p}，然后，通过融合$E_{i,o}$及其权重系数α_{t_p}，得到新向量c_t。

[0067] 最后，得到与$D_{i,o}$维数相同的带有影响权重的矩阵向量$D'_{i,o}=(c_1,\cdots,c_t,\cdots,c_L)$。

[0068] 在图4里面，注意力模块中包括分数计算层、分数融合层和Dropout层。

[0069] 步骤206，新的输出向量与原有输出向量经过拼接层、线性层以及softmax激活函数后，得到隐写文本序列信号。

[0070] 第i+1个输入文本载体的序列信号对应的$D'_{i,o}$和$D_{i,o}$直接拼接，得到重构向量$D_{i,concat}$：

[0071] $$\boldsymbol{D}_{i,concat}=[\boldsymbol{D}'_{i,o},\boldsymbol{D}_{i,o}]\in\mathbb{R}^{L\times 2d_h}$$

[0072] 接着，将$D_{i,concat}$经过全连接线性层来实现降维处理，再经过softmax激活函数得到重构后的第i+1个文本载体的移位输出序列$f'_i=(f_i(2),\cdots,f_i(t),\cdots,f_i(L)),i=1,\cdots,k$。

[0073] 对k个文本载体顺序输入文本隐写网络模型，经过编码器和解码器处理，输出移位

的序列信号，进一步可转换获取对应的文本。

[0074] 步骤三、通过利用数据集不断训练网络模型，使得移位后的输出序列f'_i与原始解码端的移位输入序列$f'_{i+1}=(f_{i+1}(2),\cdots,f_{i+1}(t),\cdots,f_{i+1}(L))$，i＝1，…，k达到误差范围阈值，反向更新编码器网络和解码器网络的参数。

[0075] 训练所用的损失函数为交叉熵损失函数：

[0076] 单个样本的交叉熵损失函数为：

[0077] $$Loss=-[y log\hat{y}+(1-y)\log(1-\hat{y})]$$

[0078] 其中，y为真实标签，$\hat{y}$为预测标签。

[0079] 在此，需要对原始解码端的移位输入序列$f'_{i+1}=(f_{i+1}(2),\cdots,f_{i+1}(t),\cdots,f_{i+1}(L))$，i＝1，…，k与经过网络模型后的输出序列$f'_i=(f_i(2),\cdots,f_i(t),\cdots,f_i(L))$，i＝1，…，k进行交叉熵损失函数计算，计算得到每一单个样本的损失值后进行叠加求和处理得到总的损失值。

[0080] 步骤四、使用训练完成的网络模型，根据输入文本载体、预设嵌入比特数以及秘密信息二进制比特流B＝"01110000101001…"，得到候选池，并构建二叉树选取对应单词作为输出。

[0081] 随机选取一首歌曲，初始在编码端输入内容为歌名的文本载体，解码端输入句首标识符，经过网络模型后，得到一个d_o维输出向量，按照一定的编码规则选取符合b中比特串所对应的单词，作为嵌入秘密信息的隐写文本。

[0082] 使用训练完成的网络模型得到输出单词的过程包括如下：

[0083] (4.1)首先，得到的d_o维输出向量中元素为概率从高到低排序完成的，每一维度元素代表语料库中的一个单词，根据预先设定的嵌入比特数b，选取前$K=2^b$维向量所对应的单词构建候选池$CP=[m_1,\cdots,m_K]$；

[0084] (4.2)然后，对候选池中的单词构建完全二叉树或考虑概率差异的哈夫曼树，每个叶子节点对应一个单词；从候选池中选取对应的单词作为输出，是指根据秘密信息比特流中的比特串选定二叉树中对应的单词作为最终输出。

[0085] 步骤五、将输出单词重新输入到解码端，不断重复，得到符合要求的嵌入秘密信息的隐写文本。

[0086] 将步骤四选取得到的单词添加到解码端的输入序列中，重新输入网络模型，编码端输入不变，重复上面过程，经网络模型选取下一个单词，直到选取的单词为句尾标识符，则得到一句完整的隐写文本。

[0087] 将整句隐写文本作为编码端的输入部分，句首标识符作为解码端输入，重新输入网络模型，重复上述步骤四和五，得到新的隐写文本，并不断重复，直到自动得到一首完整的歌曲或达到预先设定好的歌词总句数。

[0088] 进行秘密信息提取时，将生成载密文本的关键句以及句首标识符输入到文本隐写网络模型中，重构出秘密信息二进制比特流，完成提取过程。

[0089] 实施例：

[0090] 本发明实验时使用了一个中文歌词数据集。该数据集由从公开网络中搜集到的流行华语乐曲组成，主要用来研究自然语言处理领域的文本生成问题；该数据集包含有代表性的40位歌手(男女歌手各20位)的7259首歌曲作品，共计223042句歌词，1189292个歌词。

将所有歌曲按照9:1划分构建训练集和测试集的文本数据。

[0091] 在本实验中,对于该数据集中的中文歌词,首先采用jieba分词进行处理,采用基于Pytorch的Python实现,同时,用GeForce RTX 3090GPU和CUDA11.2加速训练过程。模型的编码端和解码端均将每个词映射到一个200维向量,GRU层均为两层,每层包含200个GRU单元,GRU单元内部采用tanh作为非线性激活函数σ。在模型训练时,为加强正则化,避免过拟合,加入dropout机制及采用Adam优化方法。

[0092] 一轮训练流程:从训练集中随机选择多首歌曲中的包含歌名和歌词在内的共k个文本载体,其中每个文本载体为一句歌词或者一首歌的歌名,将文本载体以序列方式输入网络模型中,经过编码器和解码器重构出后移的k个序列化文本。

[0093] 完成重构后,对于第i个文本载体,用交叉熵损失函数计算移位后的输出序列f'_i与原始解码端的移位输入序列f'_{i+1}之间的误差。

[0094] 最后根据误差计算编码器网络和解码器网络的参数梯度,根据Adam优化器和学习率更新参数值。同样,在每轮训练过程中,用测试集对模型进行测试,在测试过程中,无需更新网络参数梯度。

[0095] 训练多个epoch,直到测试集的损失值不再降低,则完成训练过程,导出模型,可以进行秘密信息嵌入与提取操作。

[0096] 在嵌入过程中,对于每次输出的d_o维向量,构建候选池,并根据构建的二叉树选择秘密信息二进制比特流b=”01110000101001…”对应的单词输出,不断重复生成隐写文本,实现秘密信息的嵌入。

[0097] 在进行提取过程时,将生成载密文本的关键句以及句首标识符输入到网络模型中,重构出秘密信息二进制比特流。

[0098] 如图5所示,为利用本发明方法所生成的隐写文本,从人类感觉系统来看,本发明考虑到了文本的序列化特点、文本间的上下文语境联系、文本载体的统计特性和语义特性等,使得生成的嵌入秘密信息的隐写文本与原始文本之间的统计差异性和语义差异性都较小,对于人类来说实际上也是区分性不大。

[0099] 如图6所示,为本发明的文本隐写方法与RNN方法的对比效果。本发明采用的基于上下文语境联系的文本生成式隐写方法生成隐写文本时,隐藏效率要低于RNN,但这是由于网络模型更为复杂,因此训练和嵌入过程需要花费更多的时间。但本发明使用的网络模型在不可感知性和隐藏容量方面均要优于RNN的性能,可知,本发明具有一定的创新性,在实际中存在一定的使用价值。

[0100] 除说明书所述的技术特征外,均为本专业技术人员的已知技术。本发明省略了对公知组件和公知技术的描述,以避免赘述和不必要地限制本发明。上述实施例中所描述的实施方式也并不代表与本申请相一致的所有实施方式,在本发明技术方案的基础上,本领域技术人员不需要付出创造性的劳动即可做出的各种修改或变形仍在本发明的保护范围内。

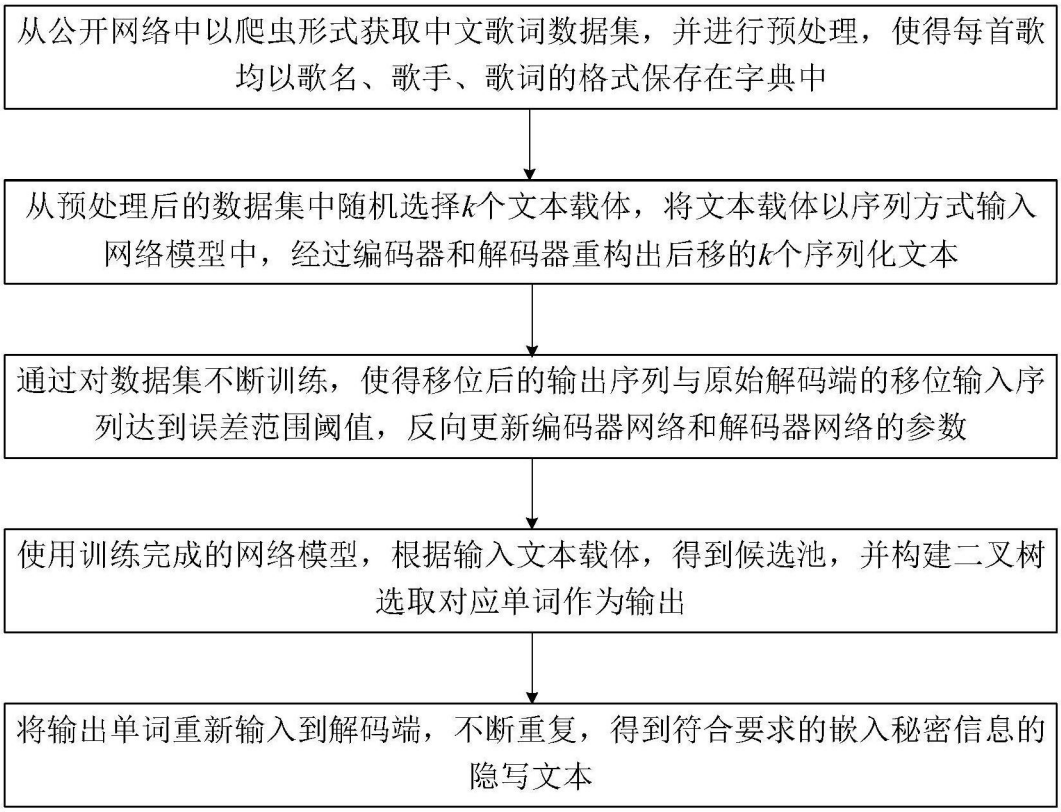
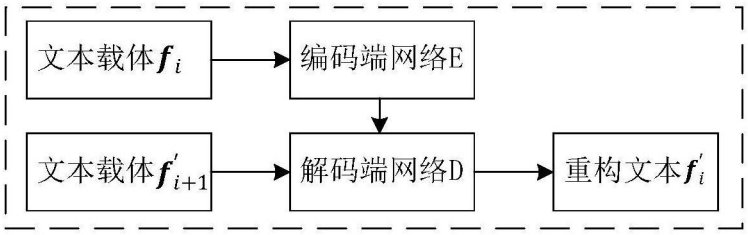


图1



训练阶段：梯度更新
测试阶段：梯度不更新

图2

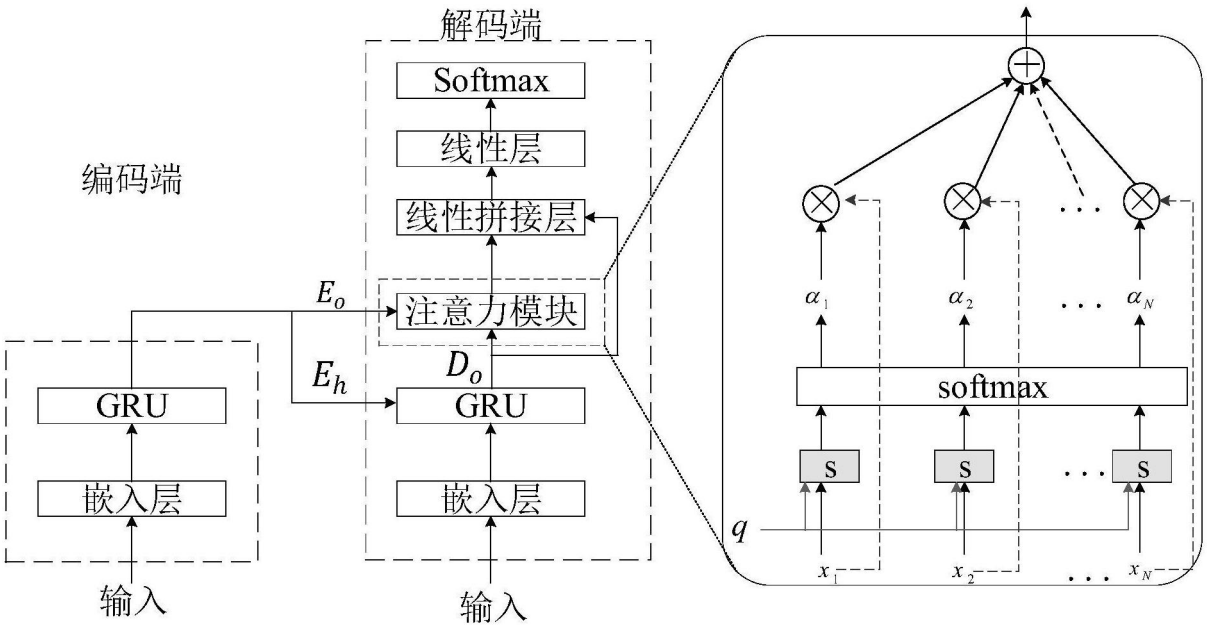


图3

模块	网络层名称	网络层参数
嵌入层模块	嵌入层	嵌入层(输入维度=33974,输出维度=200)
编码层模块	GRU	GRU(输入维度=200,输出维度=200,GRU 层数=2,批次优先=True,dropout=0.1)
解码端模块	GRU	GRU(输入维度=200,输出维度=200,GRU 层数=2,批次优先=True,dropout=0.1)
	注意力模块	
	线性拼接层	线性拼接层(输入维度=400,输出维度=200,偏置参数=True)
	线性层	线性层(输入维度=200,输出维度=10004,偏置参数=True)
注意力模块	分数计算层	分数计算层(输入维度=400,输出维度=200,偏置参数=True)
	分数融合层	分数融合层(输入维度=200,输出维度=1,偏置参数=True)
	Dropout	Dropout(被丢弃的概率=0.1, 覆盖原有位置数值=False)

图4

鱼
对生活很好步伐
面对一切都铺满更
你和你的故事
曾经是曾经在梦
我相信我们能给
去爱从此就来
风雨在何方我的
只为你我的
我劝了你忘记

图5

模型	本发明	RNN
嵌入 1 比特时隐藏效率(s/word)	8.96	1.84
嵌入 1 比特时 PPL 指标	4.155203	8.862709
嵌入 1 比特时 EMD 均值	0.684512	0.79861
嵌入 1 比特时嵌入容量 ER(%)	4.9320543	4.4827009

图6