基于序列到隐写序列的约束型自然语言信息隐藏方法

向凌云[1),2)] 杨双辉[2)] 王 蓉[2)] 刘宇航[2)] 章登勇[1),2)]

[1)](长沙理工大学综合交通运输大数据智能处理湖南省重点实验室 长沙 410114)

[2)](长沙理工大学计算机与通信工程学院 长沙 410114)

摘 要 已有的基于文本生成的无约束型自然语言信息隐藏方法主要利用不同的文本生成模型在秘密信息的控制下实现隐写文本的生成,它们生成的隐写文本质量较好且嵌入容量高.但这些方法大都局限于生成短隐写文本,整体的文本质量和句间语义相关性会随着句子长度增加而急剧下降.与无约束型方法不同,已有的约束型自然语言信息隐藏方法能针对特定场景实现长文本生成任务下的信息隐藏,具有更高的语言隐蔽性和安全性.为提高约束型方法面对各类应用场景的普适性,本文提出了一种通用的序列到隐写序列模型框架,该框架包含语言编码器和隐写器两部分,能实现从一种约束信息序列到另一种隐写文本序列的变换.以摘要生成为例,本文以序列到隐写序列模型为基本框架,提出了一种新颖的约束型自然语言信息隐藏方法.该方法在语言编码器中引入注意力优化单元以提升特征学习性能,在隐写器中融合复制机制和新设计的基于多候选优化的自适应隐写编码方法,使得隐写器可以根据候选单词序列的概率分布情况和待嵌入的秘密信息自适应地选择不同的输出优化策略,通过输出多个候选序列以及仅在嵌入时刻选择合适位置嵌入信息的方式来提高隐写文本质量.实验结果表明,本文提出的方法能够通过优化语言编码器和隐写器的设计,提高隐写摘要文本与原始摘要文本的语义相似度以及隐写摘要文本的质量,其质量甚至还优于基于基础语言编码器生成的正常摘要文本的质量,具有极强的隐蔽性.此外,本文方法生成的隐写摘要文本安全性高,利用多种隐写分析方法进行检测的各项指标基本低于60%,难以从正常文本中识别出本文方法生成的隐写摘要文本.

关键词 自然语言信息隐藏;文本生成;约束型自然语言隐藏;序列到隐写序列;隐写编码

中图法分类号 TP309 **DOI号** 10.11897/SP.J.1016.2023.01650

Constrained Linguistic Steganography Based on Sequence to Steganographic Sequence Model

XIANG Ling-Yun[1),2)] YANG Shuang-Hui[2)] WANG Rong[2)] LIU Yu-Hang[2)] ZHANG Deng-Yong[1),2)]

[1)](*Hunan Provincial Key Laboratory of Intelligent Processing of Big Data on Transportation, Changsha University of Science and Technology, Changsha* 410114)

[2)](*School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha* 410114)

Abstract Up to now, generative linguistic steganographic methods have become a research hotspot in the fields of information security and information hiding, which can be mainly divided into two categories: constrained and unconstrained generative linguistic steganography. The prevailing unconstrained generative linguistic steganography mainly employs diverse text generation language models to generate steganographic texts (stego text) under the control of the secret messages, which can produce high-quality stego texts and ensure high hidden capacity. Nevertheless, most methods are confined to automatically generating short stego texts, and the overall text quality

收稿日期:2022-04-15;在线发布日期:2023-04-10.本课题得到国家自然科学基金(61972057)、湖南省自然科学基金项目(2022JJ30623)、湖南省教育厅科研基金(21A0211)、河南省网络空间态势感知重点实验室开放课题基金(HNT2022018)资助.**向凌云**,博士,副教授,中国计算机学会(CCF)会员,主要研究方向为信息安全、信息隐藏、数字水印、隐写分析和自然语言处理.E-mail:xiangly210@163.com.**杨双辉**,硕士研究生,主要研究方向为自然语言处理和信息隐藏.**王 蓉**(通信作者),硕士研究生,中国计算机学会(CCF)会员,主要研究方向为自然语言处理和信息隐藏.E-mail:joywangrong@163.com.**刘宇航**,硕士研究生,主要研究方向为自然语言处理和隐写分析.**章登勇**,博士,副教授,中国计算机学会(CCF)会员,主要研究方向为多媒体信息安全、图像处理与模式识别、图像取证和信息隐藏.

and semantic relevance between sentences sharply decline as the length of the text increases. Different from the unconstrained linguistic steganographic methods, the existing constrained linguistic steganographic methods have been developed to achieve information hiding under long text generation tasks for specific scenarios with higher linguistic imperceptibility and security, but they lack generalization and universality. In order to improve the universality of constrained linguistic steganographic methods for various application scenarios, this paper first proposes a general sequence to steganographic sequence framework composed of a Linguistic-Encoder and a Steganographic-Encoder. In this framework, the Linguistic-Encoder encodes linguistic features of the input original constrained text, while the Steganographic-Encoder mainly contains linguistic decoding and steganographic encoding modules, which are used to decode the encoding results obtained by the Linguistic-Encoder, and then hide or extract messages. This framework can implement the transformation from one kind of constrained sequence to another kind of steganographic text sequence, and can serve as a general paradigm for various generative linguistic steganographic methods. Afterward, by applying the sequence to steganographic sequence framework into the applicable scenario of the text summarization, this paper presents a novel constraint generative linguistic steganography. The proposed method introduces an attentional refinement unit at the Linguistic-Encoder to enhance the feature learning performance. Besides, an adaptive steganographic coding algorithm based on multi-candidate optimization (MOASC) is proposed at the Steganographic-Encoder. It enables the Steganographic-Encoder to adaptively select different output optimization strategies, according to the probability distribution of candidate word sequences and the secret information to be embedded. MOASC improves the quality of steganographic summarized texts by outputting multiple candidate sequences and selecting only the right place to embed information at the moment of embedding. Meanwhile, some generation moments, which are evaluated to be not suitable for embedding secret messages, directly generate multiple candidate word sequences. For the proposed steganographic encoding method, this paper comprehensively analyzed and compared the impact of dual thresholds on the quality of generated steganographic text. Experimental results show that the proposed method can improve the quality of the steganographic summarized text, and enhance the semantic similarity between steganographic summarized text and normal one by optimizing the Linguistic-Encoder and Steganographic-Encoder. The quality of the generated steganographic summarized text even surpass that of the summarized text generated based on a basic Linguistic-Encoder, without embedding secret messages. In addition, this paper also utilized various linguistic steganalysis methods to detect the generated steganographic summarized text for simulating eavesdroppers in covert communication. The detection metrics were mostly lower than 60%, proving that the generated steganographic summarized text has strong security.

Keywords　linguistic steganography; text generation; constrained linguistic steganography; sequence to steganographic sequence; steganographic encoding

1　引　言

自然语言信息隐藏是一种以文本为载体的信息隐藏技术，已经逐渐成为信息安全领域中一个富有挑战和前景的研究热点. 已有的自然语言信息隐藏方法主要分为三类：修改式[1]、无载体式[2]和生成式自然语言信息隐藏方法[3].

修改式自然语言信息隐藏方法主要利用语义的等价变换对文本内容进行修改，在保持原始载体文本语义不变的同时隐藏秘密信息，如主被动句法变换[4]、同义词替换[5]等. 无载体式方法不对原始文本载体进行任何修改[6]，而是直接在选定的、真实、未经修改的自然文本中嵌入或提取秘密信息，如通过

设计不同的标签定位秘密信息对应的关键词[7]、检索大规模的文本数据集[8-9]等方式得到一个或多个符合要求的隐写文本.尽管上述两类方法在提高隐蔽性和安全性等方面取得了不小的进展,但普遍存在嵌入容量偏低的问题.而生成式自然语言信息隐藏方法通过控制文本自动生成的过程实现秘密信息的嵌入[10].由于可选择的生成项数量多且生成的文本长度可变,这类方法在文本生成过程中冗余空间大,大大提升了信息隐藏容量.特别是,得益于深度学习的迅猛发展,生成式隐写文本的质量也得到了很大程度的改善.因此,生成式方法成为了目前自然语言信息隐藏领域的研究重点.

根据是否输入初始信息规划隐写文本内容的生成,生成式自然语言信息隐藏方法分为无约束型和约束型两类[11],其中无约束型方法在秘密信息的控制下自由生成任意文本内容,而约束型方法在文本生成和秘密信息编码过程中还受其他输入信息的控制和约束以生成指定类型的隐写文本.在前期,研究者们提出了一系列无约束型自然语言信息隐藏方法,如基于语法规则[12]、马尔可夫模型[13]等来控制隐写文本的生成.但这些方法受限于不成熟的文本生成技术[4],生成的隐写文本可读性差,且易于被文本隐写分析技术检测到秘密信息的存在[14-15].近年来,随着基于深度神经网络的自然语言生成技术的高速发展,越来越多无约束型自然语言信息隐藏方法通过训练 Recurrent Neural Network(RNN)[16]、Long Short-Term Memory(LSTM)[17]等单个深度神经网络来构建性能良好的神经网络语言模型,生成统计分布更接近自然语言文本特性的隐写文本.为了能更好地关注输入文本内容的前后关系和顺序信息,研究者们陆续提出了基于编码器-解码器结构的自然语言信息隐藏方法[18-19].这些方法先在编码器端将输入文本序列编码成上下文向量,再在解码器端解码序列生成文本的同时嵌入秘密信息,进一步提高了隐写文本的质量.

尽管无约束型的自然语言信息隐藏方法理论上能够生成看起来足够自然的隐写文本,但仍然存在长文本上下文语义不连贯和场景适用性不足等问题.因此,研究者们尝试提出多种约束型自然语言信息隐藏方法来提高隐写文本质量或实际可安全嵌入的信息隐藏容量,如以关键词为约束条件的笑话生成[20]、指定主题的故事生成[21]、以图像为约束条件的标题生成[22]、以知识图为语义约束信息的长文本生成[23]等.但上述约束型自然语言信息隐藏方法都是针对于单一特定文本生成场景提出的信息隐藏方法,针对性极强.

为了提高约束型自然语言信息隐藏方法的普适性,本文提出了一种可适用于各类隐写文本生成的通用型序列到隐写序列框架.该框架由语言编码器和隐写器两部分组成,它先将约束信息输入到语言编码器提取特征,在秘密信息的控制下,经由隐写器解码信息生成不同形式的隐写文本.不同的场景任务可通过优化语言编码器或隐写器实现更高质量的隐写文本生成.本文以提出的通用框架为基础,以原始文本为约束信息,以生成隐写摘要文本为目标,提出了一种新的约束型自然语言信息隐藏方法.该方法一方面在语言编码器端设计了优化注意力单元以提升特征学习能力,在隐写器端引入了复制机制,充分考虑了原始文本中的单词出现率;另一方面考虑到隐写器在生成摘要过程中候选词概率分布的差异,设计了一种基于多候选优化的自适应隐写编码方法,根据候选嵌入位置间概率分布情况来自适应地选择多候选优化策略,从而进一步降低文本质量损耗以提高所生成隐写摘要文本的整体质量.本文基于不同的文本生成条件设计了多组对比实验,实验结果表明,本文提出的基于序列到隐写序列模型的约束型自然语言信息隐藏方法能生成与原始约束文本语义高度相关且隐蔽性强的隐写摘要文本.

本文工作的主要贡献可以总结为以下三点:

(1)提出了一种通用的约束型自然语言信息隐藏框架——序列到隐写序列模型,可适用于多场景的约束型隐写文本生成任务,提高了自然语言信息隐藏研究的应用价值;

(2)以序列到隐写序列模型为基本框架,针对摘要文本生成,提出了一种新的约束型自然语言信息隐藏方法.该方法在语言编码器端融合了注意力优化单元,在隐写器端引入了复制机制并设计了一种基于多候选优化的自适应隐写编码方法,充分考虑候选词预测概率之间的差异,自适应地选择多候选优化策略,从而提高所生成摘要文本的质量;

(3)实验证明,本文提出的方法生成的隐写摘要文本与原始文本语义相关性极高,其质量甚至高于基于基础编码器生成的正常摘要文本,且具有极强的隐蔽性和抗隐写分析方法.

本文第 2 节简要介绍生成式自然语言信息隐藏的相关工作;第 3 节将详细阐述本文提出的序列到隐写序列通用隐写框架;第 4 节将详细阐述基于通用隐写框架提出的一种新颖的约束型自然语言信息隐藏方法;第 5 节将阐述对比实验结果及分析;最后是本文总结和未来工作展望.

2　相关工作

在生成式自然语言信息隐藏方法涌现初期，研究者们主要沿用文本生成相关技术设计无约束型信息隐藏方法，如基于上下文无关文法规则来构建句法模板[24]、基于诗词语料库来设计特定格律语调模板[25]、使用马尔可夫模型近似统计语言模型[26-31]等.但这些方法生成的隐写文本可读性较差，且易于被文本隐写分析技术识别[14-15].随着自然语言生成技术的飞速发展，基于各种深度神经网络的无约束型自然语言信息隐藏方法被陆续提出，如通过训练单层 LSTM[32]、双层 LSTM[14]、多层 LSTM(不同数据集下结构不同)[17]、RNN[16]、预训练 Generative Pre-training Transformer(GPT-2)[33-35]等神经网络模型来获得良好的神经语言模型，并对不同时刻的词概率分布进行编码，最后通过对预测词的选择控制来嵌入秘密信息.但实际上，单个神经网络无法完成序列到序列数据的映射，无法满足高质量文本生成任务的需求.因此，研究者们陆续通过基于编码器-解码器框架、借用特殊神经网络结构提升自然语言信息隐藏性能，如利用带有注意力机制的 LSTM-LSTM 模型[19]生成隐写诗词、用基于变分自动编码器获取大量正常文本的统计分布特性来控制隐写文本的生成[36]、用生成对抗网络的判别器不断优化文本生成模型来提高隐写文本质量[15,37]等.

上述无约束型自然语言信息隐藏方法用大规模文本语料库训练神经网络模型以获得良好的语言模型，尽管能生成看起来足够自然的隐写句子，但难以生成具有连贯上下文语义的长文本，且场景适用性差.因此，研究者们开始陆续研究各种不同场景下的约束型自然语言信息隐藏方法，尝试生成具有更高应用价值的长隐写文本.如在特定的名单或棋盘步骤的约束下生成训练分析文档来隐藏秘密信息的嵌入[38]；基于特定关键词和自动笑话生成技术生成隐写笑话文本，利用笑话之间的共同变化来掩盖数据差异实现信息隐藏[20]；利用自动生成技术生成的隐写文本和人类笔记之间的数据差异存在的冗余来隐藏秘密信息[39]；通过控制考试题目和答案的自动生成来隐藏信息[40]；将待嵌入的秘密信息转换为随机系列值(字母、数字的组合序列)，再以特定主题为约束生成隐写文本[41]；以指定信息提示为约束条件控制隐写故事生成[25]；基于知识子图的语义信息为约束控制隐写文本的语义生成[23,42]；基于图像特征信息生成与其语义相关的描述性隐写标题[22,43]等.

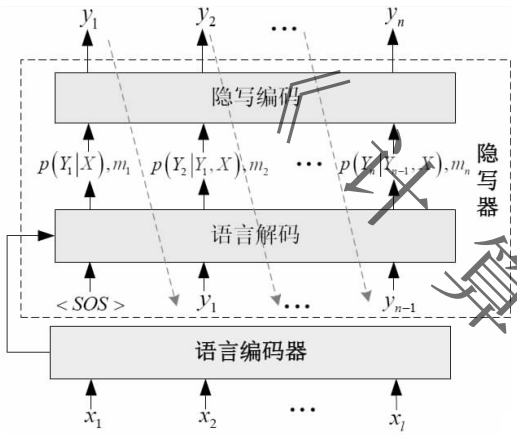
已有无约束型和约束型的生成式自然语言信息隐藏方法通过文本生成网络模型来提高生成文本的质量，同时还通过优化隐写编码策略提高隐写文本的安全性.对于无约束型自然语言信息隐藏方法，文献[17]最早提出直接对已有词汇表进行分组并进行唯一编码，在文本生成过程中通过生成词组别的控制以实现信息嵌入.为了提高编码效率并实现单词的动态隐写嵌入，文献[16]提出了可变长度编码和固定长度编码两种动态编码方式.之后，为了提高隐写文本的不可察觉性，文献[36]利用每个词的概率分布的 Kullback-Leibler divergence(KL 散度)动态调整单词嵌入率，改进了可变长度编码方式.而为了优化隐写编码引起的失真，文献[14]设计了一种自适应动态编码方法，通过每时刻自适应地将词汇表分成多个分组来优化生成词的选择.此外，文献[33]引入算术编码来动态控制秘密信息编码过程.之后，文献[35]考虑每时刻单词的概率分布情况，提出了一种自适应算术编码方法.为了动态保持概率分布多样性且有效缓解嵌入概率损失问题，文献[15]利用相似度函数和伪随机函数动态地构建候选池，再采用固定长度编码对候选池进行编码.文献[22]则在图像隐写标题生成中分别借用了逐字编码、逐句编码和基于哈希隐藏三种隐写编码方式实现信息隐藏.基于上述方法的同样框架下，文献[43]设计了一种基于动态同义词替换的隐写编码方法，以降低由同义词替换引发的统计特性差异.

尽管已有的生成式自然语言信息隐藏方法在文本生成网络模型和隐写编码策略两方面均获得不错的成果.但无约束型方法存在长文本上下文语义不连贯、场景适用性不足等问题，而约束型方法都是基于特定场景而提出的针对性较强的自然语言信息隐藏方法.因此，为了增强约束型方法的普适性，本文设计了一种序列到隐写序列的通用隐写框架.而已有约束型方法重点考虑的是如何成功编码秘密信息，对秘密信息嵌入引起隐写文本在语言上和统计上的失真考虑较少.因此，本文在具体设计约束型自然语言信息隐藏方法时，通过约束序列来生成长隐写序列，且动态地优化隐写编码过程以减少信息嵌入引起的失真，提高隐写序列的质量和安全性.

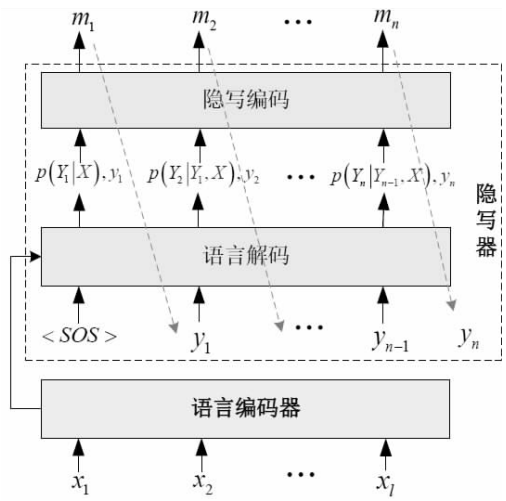
3　序列到隐写序列模型

现有约束型的生成式自然语言信息隐藏方法均为面向某种特定场景而设计的针对性方法.为了提高约束型方法的普适性，本文提出了一种通用的序

列到隐写序列框架，实现了从一种约束信息序列到另一种隐写文本序列的转换. 受序列到序列模型结构[44]的启发，该序列到隐写序列框架沿用了编码器-解码器结构，由语言编码器和隐写器两部分构成，通用框架图如图 1 所示. 其中，语言编码器用于编码输入原始约束文本的语言特性；隐写器包含语言解码和隐写编码两个模块，语言解码模块用于解码由语言编码器获得的编码结果，隐写编码模块则在隐写文本序列生成过程中实现信息嵌入(发送端)或信息提取(接收端). 发送端的隐写器在秘密信息的控制下生成隐写文本序列，而接收端的隐写器则是在隐写文本序列的协助下实现秘密信息提取.



(a) 信息嵌入(发送端)



(b) 信息提取(接收端)

图 1 序列到隐写序列通用框架图

3.1 语言编码器

语言编码器通过编码输入的约束文本序列，得到编码器隐藏状态来计算上下文状态信息，它可接收任意可变长度的文本序列 $X=\{x_1,x_2,\cdots,x_l\}$ 作为输入(其中 x_l 表示输入的第 l 个词)，并将其转换为具有固定长度的编码状态，即上下文状态信息. 在语言编码器中，t 时刻语言编码器的隐藏状态 h_t 由上一时刻的隐藏状态 h_{t-1} 和当前时刻对应的词 x_t 所决定，用式(1)表示如下：

$$h_t=f(h_{t-1},\boldsymbol{e}(x_t)) \tag{1}$$

其中，函数 $f(\cdot)$ 表示 RNN、LSTM 和 Glavnoe Razvedivatelnoe Upravlenie(GRU)等神经网络中一系列的线性变化和非线性变换；$\boldsymbol{e}(x_t)$ 指词 x_t 对应的词向量.

通过自定义函数，如 RNN 输出层的输出函数，将根据前 t 时刻的隐藏状态来获取当前 t 时刻语言编码器的上下文状态向量 $\boldsymbol{h}_t$，用式(2)表示如下：

$$\boldsymbol{h}_t=q(h_1,\cdots,h_t) \tag{2}$$

其中，$q(\cdot)$ 表示自定义函数，用于获取上下文状态信息.

3.2 隐写器

接收到语言编码器的上下文状态信息后，隐写器将对其进行解码并得到各时刻的词概率分布情况，并通过控制词的采样或者根据隐写文本中词的采样结果来实现秘密信息的嵌入或提取. 这个过程主要包括两个模块：语言解码和隐写编码模块，如图 1 所示. 考虑到信息隐藏方法涉及到发送端嵌入信息和接收端提取信息两个过程，隐写器的隐写编码模块在发送端和接收端的工作模式有差异. 如图 1(a)所示，发送端的隐写器主要在目标文本 Y 的生成过程中，实现秘密信息 $M=\{m_1,m_2,\cdots,m_n\}$ 的嵌入，此时隐写编码模块以 M 为输入控制 Y 的生成. 如图 1(b)所示，接收端的隐写器主要在文本生成过程中，解码接收到的隐写文本 Y 的隐写编码值，提取秘密信息 M，此时隐写编码模块以 Y 为输入解码其隐写编码值以提取 M.

具体的，在 t 时刻，隐写器先通过语言解码模块对约束文本序列 X 在语言编码器所获得的编码信息进行解码，并以模型已生成的前缀 Y_{t-1} 为条件预测下一个时刻单词的概率分布情况 $P(y_t|Y_{t-1},X)$；隐写编码模块则根据概率分布来选择符合当前语境的预测单词集合，并给集合中每个单词匹配唯一的编码值. 在发送端，隐写编码模块在待嵌入秘密信息 $m_t\in M$ 的约束下，选择与其匹配的单词 y_t. 在接收端，隐写编码模块则根据接收到的隐写文本 Y 在当前时刻的词 y_t，查找 y_t 的隐写编码值作为当前时刻提取的秘密信息 $m_t\in M$.

一般情况下，隐写器的语言解码模块采用与语言编码器相同的神经网络模型来进行语言解码，再通过隐写编码模块来嵌入或提取秘密信息. 在 t 时

刻，隐写器的语言解码模块接收语言编码器产生的上下文向量 $\boldsymbol{h}_t$ 以及 t 时刻前已生成的前馈序列 Y_{t-1} 后，语言编码模块通过与语言编码器一致的 RNN、LSTM 或 GRU 等神经网络模型将上下文向量 $\boldsymbol{h}_t$ 中的信息解码，得到 t 时刻的词概率分布 $P(y_t)$，如式(3)所示：

$$P(y_t)=\sigma(V\cdot(\sigma(U\cdot\boldsymbol{h}_t+W\cdot s_{t-1}+b))+c) \quad (3)$$

其中，$\boldsymbol{h}_t$ 表示 t 时刻语言编码器的上下文向量；s_t 表示 t 时刻隐写器的隐藏状态；$\sigma(\cdot)$ 为非线性函数；V,U,W,b,c 为神经网络模型学习到的参数.

对于 $P(y_t)$，隐写器的隐写编码模块将对概率值高的选定词集合进行隐写编码，每一个词将编码成唯一不重复的隐写编码值. 发送端的隐写编码模块将受秘密信息 $M=\{m_1,m_2,\cdots,m_n\}$ 的控制，根据当前待嵌入的秘密信息 m_t 对 $P(y_t)$ 进行的采样，选择隐写编码值为 m_t 的词 y_t 作为当前的输出，生成隐写文本序列 Y，如图 1(a)所示. 接收端的隐写编码模块则根据隐写文本序列 Y 的当前词 y_t 进行隐写解码，将其与预测的单词集合进行对比，一旦匹配则输出该词对应的隐写编码值，即提取当前时刻的秘密信息 $m_t\in M$.

以最简单的隐写编码方式为例，即每个嵌入时刻仅嵌入 1 比特秘密信息. 具体编码过程如下：将语言解码模块输出的词概率分布降序排列，选择概率最高和次高的词分别编码成 0 和 1. 在发送端，若待嵌入的秘密信息为“0”，则选择概率最高的词作为当前生成词；否则选择次高概率的词. 在接收端，若隐写文本中出现的是概率最高的词，则提取的秘密信息为“0”，否则提取的秘密信息为“1”. 隐写器和语言编码器一样，需要经过大规模的文本语料库的训练与优化，才能得到最优性能的隐写器. 考虑到已有大规模语料库中的训练文本，其包含的词与秘密信息的关系是不确定的，因此，在训练过程中，隐写器端假设输入的是随机秘密信息，即不输入确定的秘密信息来约束隐写文本的生成，仅以 X 和 Y_{t-1} 为条件预测 t 时刻的词概率分布情况. 除此之外，隐写器利用反向传播算法不断优化模型的参数和性能，再用最大似然估计优化模型，最大化 $P(y_t|Y_{t-1},X)$，以获得最优的语言模型，从而有效提升隐写器端生成的文本质量.

4　方法描述

上述提出的序列到隐写序列框架适用于各种约束型文本生成场景，可通过设计并优化语言编码器或隐写器提升自然语言信息隐藏方法的性能，实现高质量隐写文本的生成，进一步保障信息隐藏方法的隐蔽性和安全性. 本文以隐写摘要文本生成任务为例，设计了一种新的约束型自然语言信息隐藏方法. 该方法在发送端信息嵌入过程中的整体结构图如图 2 所示. 该方法的语言编码器(命名为 ARU-

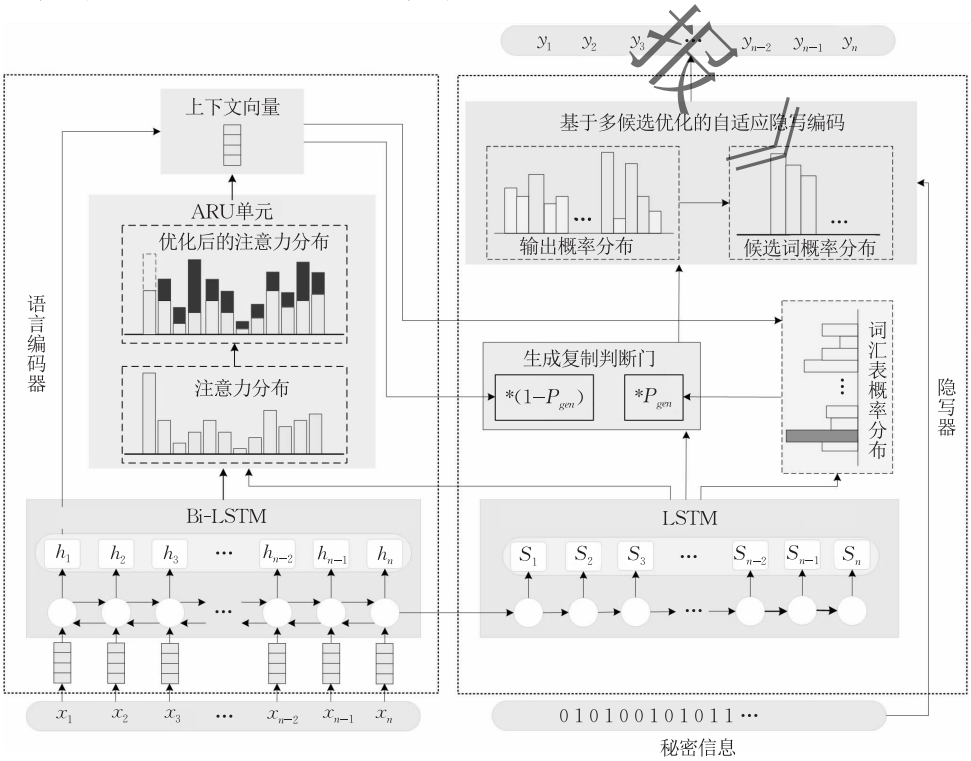


图 2　基于序列到隐写序列的自然语言信息隐藏方法结构图

LiEncoder)融合了注意力优化单元以提升语言编码器对原始文本的特征提取能力;隐写器(命名为Copy-SteEncoder)的语言解码模块引入了复制机制以提高解码性能,在隐写编码模块则设计了一种基于多候选优化的自适应隐写编码方法(简写为MOASC)以降低秘密信息嵌入引起的文本质量损失.接下来,本文将从以下三个部分具体阐述所提出约束型自然语言信息方法的原理:(1)基于注意力优化的语言编码器;(2)基于复制机制的隐写摘要文本生成的隐写器;(3)基于多候选优化的自适应隐写编码方法.

4.1 基于注意力优化的语言编码器

基于序列到隐写序列的自然语言信息隐藏方法的语言编码器 ARU-LiEncoder 要先将输入的原始文本数据表示为向量形式,再对其进行自动学习与编码,以辅助隐写器生成质量高的隐写摘要文本.如图 2 所示,ARU-LiEncoder 将输入的约束信息序列(即原始文本)经由嵌入层的数据预处理后转换成词向量形式,如通过分布式词向量模型 word2vec 完成词的向量化,将向量化的原始文本信息输入神经网络进行编码,再经由注意力优化单元 ARU(Attention Refinement Unit)优化原始文本信息编码过程中的注意力分布情况,以提升对约束文本信息特征的学习效果.

ARU-LiEncoder 采用的基础模型与指针生成网络模型[45]的编码结构类似,使用双向长短期记忆模型 Bi-LSTM(Bidirectional LSTM)来学习输入的原始文本的特征,获得每个输入词所对应的固定长度隐藏状态 h_i.为了提高语言编码器的性能并满足在隐写器端生成与输入约束序列不同形式的隐写序列的需求,ARU-LiEncoder 结合隐写器的隐藏状态 s_t,并引入了注意力机制来优化语言编码器.注意力分布的计算如式(4)所示:

$$\begin{cases} e_i^t = v^T \tanh(W_h h_i + W_s s_t + b_{attn}) \\ a_i^t = \mathrm{Softmax}(e^t) \end{cases} \tag{4}$$

其中,v,W_h,W_s,b_{atten} 均为神经网络模型学习到的参数,e_i^t 为摘要文本中第 t 个词与输入的约束原始文本中第 i 个词的相关性.注意力分布 a^t 是输入约束文本中每一个词的重要程度,隐写器可依据注意力分布值更好地生成隐写摘要序列的下一个词.

为更好地聚合当前时刻的约束文本信息,ARU-LiEncoder 根据注意力分布和隐藏状态获取上下文向量 $\boldsymbol{h}_t$,用式(5)表示如下:

$$\boldsymbol{h}_t = \sum_i a_i^t h_i \tag{5}$$

为更好地识别重要内容并解决生成的文本摘要序列中存在的内容重复问题,ARU-LiEncoder 对注意力机制进行了优化,设计了 ARU 模块,用于增强输入文本中重要部分的注意力且减弱不相关部分的注意力.ARU 模块,结合注意力分布 a^t 和当前隐写器的隐藏状态 s_t 计算相关性 r^t,计算过程如式(6)所示:

$$r^t = \sigma(W_s^r s_t + W_a^r a^t + b_r) \tag{6}$$

其中,σ 表示激活函数,W_s^r,W_a^r,b_r 均为神经网络模型学习到的参数.r^t 用来度量当前注意力分布应该更新的程度.r_i^t 越小表示第 i 个位置的内容与隐写器的隐藏状态 s_t 越不相关,则应减弱第 i 个位置的注意力分布值.之后,注意力分布值 a^t 将被更新为 $a^{t'}$,从而更进一步更新上下文向量 $\boldsymbol{h}_t$,如式(7)所示:

$$\begin{cases} a^{t'} = r^t \odot a^t \\ \boldsymbol{h}_t = \sum_i a_i^{t'} h_i \end{cases} \tag{7}$$

4.2 基于复制机制的隐写摘要文本生成的隐写器

本文提出的自然语言信息隐藏方法的隐写器 Copy-SteEncoder 接收 ARU-LiEncoder 的状态信息后,在复制机制的辅助下,综合考虑生成的词汇表概率分布情况和原始文本的注意力分布情况,预测每时刻候选词的概率分布值,再采用基于多候选优化的自适应隐写编码方法,在秘密信息的控制下对候选词进行选择来生成隐写摘要文本,如图 2 右半部分所示.

Copy-SteEncoder 的语言解码模块接收到 ARU-LiEncoder 输出的上下文向量 $\boldsymbol{h}_t$ 后,与其神经网络解码的隐藏状态 s_t 结合,再经由两个线性层后获得词汇表中所有单词的概率分布 P_{vocab},如式(8)所示:

$$P_{vocab} = \mathrm{Softmax}(V'(V[s_t, \boldsymbol{h}_t] + b) + b') \tag{8}$$

其中,V,V',b,b' 均为模型学习到的参数,$V[s_t, \boldsymbol{h}_t] + b$ 和 $V'(V[s_t, \boldsymbol{h}_t] + b) + b'$ 分别表示两个线性层.

同时,Copy-SteEncoder 利用当前时刻的上下文向量 $\boldsymbol{h}_t$、隐写解码模块的隐藏状态 s_t 和隐写器当前时刻的输入 $\boldsymbol{O}_{t-1}$ 共同计算生成预测权重 $P_{gen} \in [0,1]$,如式(9)所示:

$$P_{gen} = \sigma(W_h^T \boldsymbol{h}_t + W_s^T s_t + W_x^T \boldsymbol{O}_{t-1} + b_{ptr}) \tag{9}$$

其中 W_h^T,W_s^T,W_x^T,b_{ptr} 均为神经网络模型学习到的参数,$\sigma(\cdot)$ 为 Sigmoid 函数,$\boldsymbol{O}_{t-1}$ 为前一时刻隐写器输出的摘要单词对应的词向量,$t=1$ 时,$\boldsymbol{O}_{t-1}$ 为预定义的起始词向量.

在摘要文本生成过程中,t 时刻预测的词 W(生成的隐写摘要文本中的候选词 y_t)对应的概率分

布值 $P(W)$由词汇表概率分布和输入序列的注意力分布 $a^{t'}$综合计算获得，如式(10)所示. 若 W 未在原始文本中出现，即原始文本中任意词 $W_i \neq W$，则注意力分布 $\sum_{i:W_i=W} a_i^{t'}=0$；若 W 未在词汇表中出现，则 $P_{vocab}(W)$为 0.

$$P(W)=P_{gen}P_{vocab}(W)+(1-P_{gen})\sum_{i:W_i=W} a_i^{t'} \tag{10}$$

在上述计算词 W 的概率分布值 $P(W)$的公示中，生成预测权重 $P_{gen}\in[0,1]$，它被视为摘要文本生成过程中的软开关，称为生成复制门，用来判断该时刻生成的单词是直接复制的还是根据概率分布值预测生成的. 当 $P_{gen}=0$ 时，Copy-SteEncoder 将根据注意力分布 $a^{t'}$从输入序列(原始文本)中复制单词；当 $P_{gen}=1$ 时，将从词汇表中取样单词；当 P_{gen}为其他值时，将根据概率分布值 $P(W)$选择生成单词.

隐写器 Copy-SteEncoder 计算得到每时刻摘要单词的概率分布值后，将选择合适的候选单词用来嵌入秘密信息进行隐写摘要文本的生成. 由于对摘要单词的选择是在秘密信息的控制下强制完成的，生成的隐写文本质量将受到一定程度的影响，即难以总选择最优的单词输出，导致生成的隐写摘要文本质量可能会差于未受秘密信息控制生成的正常摘要文本. 为了生成高质量隐写摘要文本，优化候选摘要单词的隐写编码与选择，本文设计了一种基于多候选优化的自适应隐写编码方法 MOASC，实现候选摘要单词的自适应生成.

在秘密信息的嵌入过程中，自适应隐写编码方法 MOASC 将首先根据预测的词概率分布情况判断某生成时刻是否适合嵌入秘密信息，再自适应选择多候选优化策略. 若适合嵌入，则该生成时刻为嵌入时刻，将根据多候选隐写优化规则对最优的两项候选嵌入位置进行编码，再根据秘密信息，选择匹配的候选嵌入位置，并输出该位置的 k 个候选序列；若不适合嵌入，则该生成时刻为非嵌入时刻，将不嵌入秘密信息，直接选择联合概率分布值最高的 k 个候选序列输出. 一直重复上述操作，直至将秘密信息全部嵌入完成，后续生成时刻将均按非嵌入时刻选择摘要词，直至遇到结束符或达到最长长度为止. 由于每个时刻将生成 k 个候选隐写序列，因此，最终会选择概率值最高的一个序列作为最终的隐写摘要文本，进一步优化隐写摘要文本的质量.

接收端接收到发送端通过公开渠道传输的源文本、隐写摘要文本和共享的序列到隐写序列模型后，将与发送端采用相同方式和参数来生成摘要文本，不同之处只在于基于多候选优化的自适应隐写编码方法 MOASC 对嵌入时刻的处理. MOASC 在完成嵌入时刻的候选嵌入位置的编码后，将检索接收到的隐写摘要文本中当前时刻的单词所在的候选嵌入位置，获得其编码值，即为提取到的秘密信息. 同时，MOASC 将选择该候选嵌入位置的 k 个候选序列进行输出.

下面详细介绍基于多候选优化的自适应隐写编码方法(MOASC)的原理.

4.3　基于多候选优化的自适应隐写编码方法

生成式自然语言信息隐藏方法通常是在秘密信息的约束控制下从各生成时刻的候选单词集合中进行采样来生成隐写文本. 由于候选单词的预测概率分布存在差异，受秘密信息强制控制，易采样到不合适的低质量词，导致生成的隐写文本的质量随秘密信息的不同而动态变化，且通常要差于不受控制自由生成的文本的质量. 同时，现有的方法在隐写编码时通常只考虑了当前时刻的候选词预测概率分布，而忽略了词间的长距离依赖关系，严重影响了长文本质量.

为了尽可能地减少由秘密信息嵌入引起的摘要文本质量下降并提高隐写摘要文本的多样性，本文提出了一种基于多候选优化的自适应隐写编码方法 MOASC. 考虑到词间的长距离影响，因此，MOASC 不单一依赖当前时刻的预测词概率分布情况来确定嵌入位置，将根据前馈生成序列和当前时刻的预测词概率分布情况来衡量候选序列间的差异，过滤掉不适合的嵌入位置，并选择合适的候选序列进行编码. 其次，每个时刻生成多个候选序列以提高隐写文本的多样性，并获得全局最优的隐写摘要文本.

基于多候选优化的自适应隐写编码方法MOASC的具体过程，主要分为以下三个步骤：(1) 嵌入条件判断. 根据候选嵌入位置各候选序列间的联合概率分布差异程度，衡量候选嵌入位置是否适合嵌入秘密信息，再自适应地选择多候选优化策略；(2) 多候选概率优化. 对于不适合嵌入秘密信息的生成时刻，选择联合概率分布值最高的固定个候选序列作为当前时刻的多候选概率优化输出，从而最小化文本质量损失；(3) 多候选隐写优化. 对于适合嵌入秘密信息的生成时刻，先对所有候选嵌入位置求平均联合概率分布值，再根据概率差异评估值筛选出两个最优的候选序列嵌入位置进行隐写编码，减少由秘密信息控制引起的隐写失真，即生成质量的下降.

4.3.1 嵌入条件判断

在隐写摘要文本生成过程中,MOASC 每个时刻均输出 k 个序列作为候选嵌入位置.在 $t-1$ 时刻自适应优化输出的 k 个序列将作为 t 时刻的前馈序列,参与 t 时刻的词概率分布预测.对于每个候选嵌入位置,MOASC 将根据其生成的 k 个候选序列的情况来评估候选嵌入位置是否满足嵌入条件.

MOASC 在衡量候选嵌入位置是否满足秘密信息嵌入条件时,不仅考虑了与已生成的前馈隐写摘要文本序列的联合概率,还考虑了该候选嵌入位置上各候选序列之间的概率差异.MOASC 先根据隐写器的语言解码模块所预测的当前候选嵌入位置所有词的概率分布,再选择 k 个概率分布最高的作为候选词;再将候选词分别与每个候选嵌入位置的前馈序列组合成候选序列,并计算各候选序列的联合概率来评估候选嵌入位置是否适合嵌入秘密信息.设 t 时刻的 k 个候选嵌入位置分别为 $Text_1^{t-1},\cdots,Text_k^{t-1}$.对于第 i 个候选嵌入位置 $Text_i^{t-1}$,语言解码模块预测获得的词概率分布值最高的 k 个候选词,记为 $W_{i1},W_{i2},\cdots,W_{ik}$,其中 W_{ij} 经过式(11)计算得到的预测概率值为 $P(W_{ij})$.k 个候选词与 $Text_i^{t-1}$ 全排列组合成得到当前候选嵌入位置的 k 个候选序列,记为 $ST_i=\{st_{i1}=Text_i^{t-1}+W_{i1},\cdots,st_{ik}=Text_i^{t-1}+W_{ik}\}$.设前馈序列 $Text_i^{t-1}$ 的概率值为 $P(Text_i^{t-1})$,则候选序列 st_{ij} 的联合概率值 $P(st_{ij})$ 计算如下:

$$P(st_{ij})=P(Text_i^{t-1})+P(W_{ij}) \tag{11}$$

为了便于设置阈值,MOASC 利用式(12)对候选序列的联合概率 $P(st_{ij})$ 进行归一化,再利用式(13)判断候选嵌入位置是否符合嵌入条件:

$$P(st_{ij})=\frac{P(st_{ij})}{\sum_{j=1}^{k}P(st_{ij})} \tag{12}$$

$$\begin{cases}P(st_{i1})\geq\cdots\geq P(st_{ik})\geq\alpha\\P(st_{i1})-P(st_{ik})\leq\beta\end{cases} \tag{13}$$

其中 α、β 为阈值.

若每一种候选嵌入位置编码秘密信息的一种取值状态,满足嵌入条件的候选嵌入位置数越多,则可编码的秘密信息越多.由于编码 1 比特秘密信息的两种取值状态"0"和"1",至少需要 2 个可供编码的候选嵌入位置,因此,以是否含有 2 个满足嵌入条件的候选嵌入位置为标准,MOASC 引入了嵌入时刻和非嵌入时刻的概念.具体定义如下:

定义 1. 嵌入时刻指满足秘密信息嵌入条件的候选嵌入位置数大于等于 2 的生成时刻;

定义 2. 非嵌入时刻指满足秘密信息嵌入条件的候选嵌入位置数小于 2 的生成时刻.

MOASC 根据嵌入条件即式(13)所示判断每个时刻是否为适合嵌入秘密信息的嵌入时刻,再在文本生成过程中自适应地选择多候选优化策略.在嵌入时刻嵌入秘密信息,利用多候选隐写优化策略编码秘密信息并输出 k 个序列;在非嵌入时刻则不嵌入秘密信息,利用多候选概率优化策略输出 k 个序列.如图 3 所示,在隐写文本生成的最初时刻,先初始化 k 个单词,分别作为 $t=1$ 时刻的候选嵌入位置,记为 $Text_1^0,\cdots,Text_k^0$.$t=1$ 时刻的第 1 个候选嵌入位置为 $Text_1^0$,预测的多个候选词为 $W_{i1},W_{i2},\cdots,W_{ik}$.对每个候选嵌入位置 $Text_i^0$,需要组合所有候选词 $W_{i1},W_{i2},\cdots,W_{ik}$ 来判断该候选嵌入位置是否满足嵌入条件(13).$t=1$ 时刻,设 k 个候选嵌入位置 $Text_1^0,\cdots,Text_k^0$ 中只有一个满足嵌入条件(13),因此,$t=1$ 时刻为非嵌入时刻.$t=2$ 时刻,设 k 个候选嵌入位置中至少有两个满足嵌入条件(13),因此,$t=2$ 时刻为嵌入时刻.

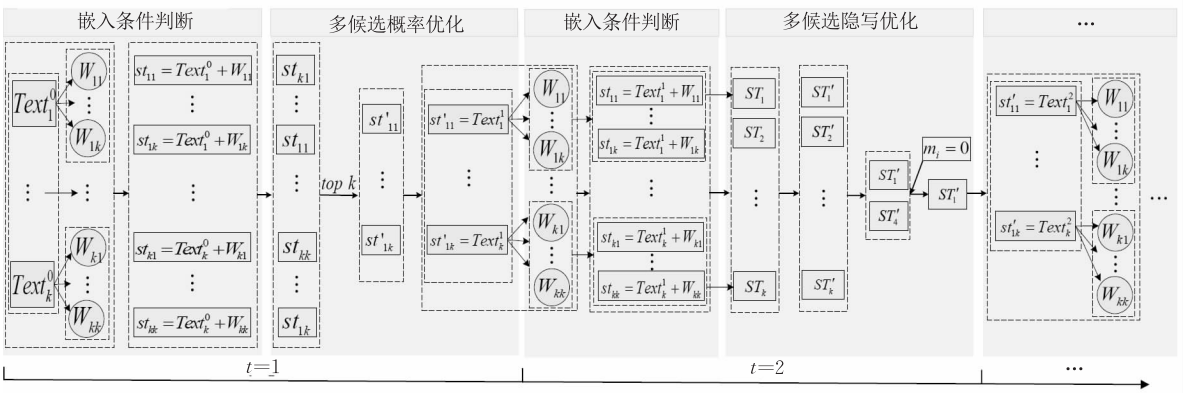


图 3 基于多候选优化的自适应隐写编码示例图

4.3.2　多候选概率优化输出

对嵌入条件判定为非嵌入时刻的 t 时刻，MOASC 将自适应地选择最优的 k 个候选序列作为该时刻的输出且不嵌入秘密信息. 在 t 时刻，每个候选嵌入位置将生成 k 个候选序列，k 个候选嵌入位置将生成 $k\times k$ 个候选序列. 设第 i 个候选嵌入位置的第 j 个候选序列为 st_{ij}，其联合概率为 $P(st_{ij})$. MOASC 根据候选序列的联合概率，对 $k\times k$ 个候选序列 st_{ij} 降序排序，记为 $st'_{11},\cdots,st'_{1k},\cdots,st'_{k1},\cdots,st'_{kk}$，再选择联合概率值最高的 k 个候选序列 $st'_{11},\cdots,st'_{1k}$ 作为多候选概率优化输出，作为下一时刻的候选嵌入位置，记为 $Text^t_1,\cdots,Text^t_k$，其中 $Text^t_i=st'_{i1}$. 如图 3 所示，MOASC 先根据嵌入条件(13)判断 $t=1$ 时刻为非嵌入时刻，再根据 $P(st_{ij})$ 对 st_{ij} 进行降序排列，再取联合概率值最高的 k 个候选序列 $st'_{11},\cdots,st'_{1k}$ 作为该时刻的优化输出，以及作为 $t=2$ 时刻的 k 个前馈序列并设为该时刻的候选嵌入位置.

4.3.3　多候选隐写优化输出

若根据嵌入条件判定 t 时刻为嵌入时刻，MOASC 将自适应地选择满足嵌入条件且与编码值匹配的候选嵌入位置输出.

记 t 时刻 k 个候选嵌入位置联合该时刻词概率值最高的 k 个词为 $W_{i1},W_{i2},\cdots,W_{ik}$，更新候选嵌入位置为 $ST_1,\cdots,ST_i,\cdots,ST_k$，其中 $ST_i=\{st_{i1}=Text_i^{t-1}+W_{i1},\cdots,st_{ik}=Text_i^{t-1}+W_{ik}\}$. 对于满足嵌入条件的候选嵌入位置 ST_i，计算其平均概率值 $\bar{P}_{ST_i}=\frac{1}{k}\sum_{j=1}^{k}P(st_{ij})$ 并根据 $\bar{P}_{ST_i}$ 降序排列并记为 $ST'_1,ST'_2,\cdots,ST'_k$. 为了减少采样到低质量词的概率，MOASC 将选择各候选序列间差异最小的候选嵌入位置来编码秘密信息，因此，MOASC 根据如下式(14)计算每个满足嵌入条件的候选嵌入位置的概率差异评估值 $D_{ST'_i}$：

$$D_{ST'_i}=\frac{\bar{P}_{\max}-\bar{P}_{ST'_i}}{\bar{P}_{\max}}+\frac{P(st'_{i1})-P(st'_{ik})}{P_{\min}} \tag{14}$$

其中，$\bar{P}_{\max}$ 为所有满足嵌入条件候选嵌入位置 ST'_1，$ST'_2,\cdots,ST'_k$ 中平均概率最大值，$P_{\min}$ 为候选最大最小概率差最小值. 根据概率差异评估值，选择 $D_{ST'_i}$ 最小的两个候选嵌入位置编码秘密信息，如式(15)所示：

$$\begin{cases}a=\underset{1\le i\le k}{\arg\min}D_{ST'_i}\\ b=\underset{1\le i\le k,\,i\ne a}{\arg\min}D_{ST'_i}\end{cases} \tag{15}$$

其中，a、b 分别表示概率差异评估值最小、次小的候选嵌入位置在已排序 k 个候选嵌入位置 $ST'_1,ST'_2,\cdots,ST'_k$ 中的序号. 第 a 和 b 个候选嵌入位置将分别被编码为“0”和“1”，具体编码规则如式(16)所示：

$$\begin{cases}C(ST'_a)=0\\ C(ST'_b)=1\end{cases} \tag{16}$$

其中 $C(\cdot)$ 表示编码值.

在信息嵌入过程中，MOASC 将根据待嵌入秘密信息选择对应编码值的候选嵌入位置，并输出该候选嵌入位置中的 k 个候选序列 $st'_{i1},st'_{i2},\cdots,st'_{ik}$ 作为多候选分组隐写优化输出 $Text^t_1,\cdots,Text^t_k$，其中 $Text^t_j=st'_{ij}$.

如图 3 所示，MOASC 先根据嵌入条件判定 $t=2$ 时刻为嵌入时刻. 之后，根据平均概率值 $\bar{P}_{ST_i}$ 对该时刻所有满足嵌入条件的 k 个候选嵌入位置降序排列，排序后的序列被记为 $ST'_1,ST'_2,\cdots,ST'_k$. MOASC 评估所有候选嵌入位置的概率差异值 $D_{ST'_i}$，选择差异值最小的两个候选嵌入位置，并依次被编码为“0”和“1”，即 ST'_1 和 ST'_4. 根据需要嵌入的秘密信息比特位 $m_i=0$，最终选择候选嵌入位置 ST'_1 作为该时刻的多候选隐写优化输出，且 ST'_1 中包含的 k 个候选序列将作为下一时刻的前馈序列.

5　实验结果与分析

5.1　实验数据集及参数设置

本文在 CNN/Daily Mail 英文数据集[46] 上进行实验，该数据集包含了大量新闻文本-摘要文本对，每个新闻文本平均长度为 781 个单词，每个摘要文本平均长度为 56 个单词. 在模型训练前，本文先使用 Nallapati 等人[44] 提供的脚本预处理数据集，处理后的数据集包含 287 226 对训练文本，13 368 对测试文本. 在隐写摘要文本生成过程中，本文采用随机生成的二进制比特序列作为秘密信息，共生成 13 368 个隐写摘要文本.

本文实验在 NVIDIA GeForce GTX TITAN 环境下进行，采用 Tensorflow 框架. word2vec 输出 128 维的词嵌入向量，语言编码器和隐写器均采用 256 维的隐藏状态向量. 在训练过程中，本文构建了 50 k 大小的词汇表，模型训练 epoch 数为 35，batchsize 大小为 16，学习率为 0.15，多候选优化输出策略的输出序列个数 k 为 2.

5.2　评价指标

5.2.1　文本质量评价指标

ROUGE 是文本摘要生成领域中衡量文本质量

的重要指标之一.本文使用 ROUGE 系列评价指标[47]来评估参考摘要文本与生成摘要之间的相关性,以此评估本文方法生成的隐写摘要文本的质量.ROUGE 系列评价指标通过统计参考摘要文本与生成摘要文本间共现的内容衡量两者的语义相关性.共现内容越多,ROUGE 得分将越高,说明生成隐写摘要文本与参考摘要文本语义相似度越高,隐写摘要文本质量越好.

ROUGE 系列指标的统一计算方法如式(17)所示:

$$\begin{cases} P=\dfrac{c(X,Y)}{c(X)} \\ R=\dfrac{c(X,Y)}{c(Y)} \\ F=\dfrac{(1+\gamma^2)P\times R}{P+\gamma^2\times R} \end{cases} \tag{17}$$

其中,P,R,F 分别表示准确率、召回率和 $F1$ 值,X 表示生成的摘要文本,Y 表示参考摘要文本.$c(X,Y)$ 表示统计生成的摘要文本和参考摘要文本间内容共现情况的特定函数.$c(X)$和 $c(Y)$分别表示生成摘要文本和参考摘要文本的特定序列统计数.γ 为设置的常数.

考虑从不同角度统计生成的摘要文本和参考摘要文本的共现内容,ROUGE 指标有多种不同变型,本文主要使用了 6 种 ROUGE 指标,分别是 ROUGE-N($N=1,2$)、ROUGE-L、ROUGE-W、ROUGE-S、ROUGE-SU.ROUGE-N 主要考虑两种文本之间无序重叠共现的 n-gram 单词词组情况.ROUGE-L 则利用有序重叠共现的最长公共子序列来评估两种文本的相似度.在此基础上,ROUGE-W 考虑了加权最长公共子序列,使连续匹配的单词序列比不连续匹配被赋予更大的权重.ROUGE-S 使用跳跃二元组捕获长距离的句子结构来有效衡量两种摘要文本之间的内容相似度.ROUGE-SU[47]则融合 ROUGE-S 和 ROUGE-L 来设计评估函数.

5.2.2 抗隐写分析能力评价指标

抗隐写分析能力是评测自然语言信息隐藏方法安全性的有效指标,通常使用隐写分析方法对隐写文本进行检测,根据检测结果的性能指标来进行有效评估.对于各类方法,本文使用分类任务中常用的测试指标来评估本文提出方法的性能,即准确率 Acc、精确率 P 和召回率 R.值越小表示该方法的抗隐写分析能力越强,即隐藏了秘密信息的隐写摘要文本越难被识别.具体计算如式(18)所示:

$$\begin{cases} Acc=\dfrac{TP+TN}{TP+FN+FP+TN} \\ P=\dfrac{TP}{TP+FP} \\ R=\dfrac{TP}{TP+FN} \\ F_1=\dfrac{2\cdot P\cdot R}{P+R} \end{cases} \tag{18}$$

其中,TP 表示被正确检测为隐写摘要文本的数量;FP 表示正常摘要文本被错误检测为隐写摘要文本的数量;TN 表示被正确检测为正常摘要文本的数量;FN 表示隐写摘要文本错误检测为正常摘要文本的数量.

本文使用了 TS-GCN[48]、TS-RNN[49]、TS-CNN[50]三种典型的隐写分析方法来衡量隐写文本的抗隐写分析能力.这些隐写分析方法利用不同的神经网络模型自动学习文本特征来区分正常文本和隐写文本.其中,TS-GCN[48]方法用图卷积神经网络 GCN 来收集上下文语义信息以更新节点表示,并进一步采用全局共享矩阵来获得更好的文本表示;TS-RNN[49]用循环神经网络 RNN 分析并提取待检测文本的条件概率分布特性差异;TS-CNN[50]用卷积神经网络 CNN 从语义和概率分布差异两方面同时捕获文本特征.

5.3 文本质量分析

为了评估本文提出方法所生成隐写文本的质量,衡量本文方法的隐蔽性,本文从嵌入条件判断时阈值、语言编码器、隐写器三个方面着手设计了多组实验,对不同条件下的隐写文本的质量进行了评估和讨论分析.

5.3.1 嵌入条件判断时阈值的影响

基于多候选优化的自适应隐写编码方法在嵌入条件判断时使用了两个阈值 α 和 β,这两个阈值的取值情况将影响满足嵌入条件的候选序列和嵌入时刻的范围.当阈值 α 越大,候选位置中各候选序列的词概率分布将整体更高,即候选序列的词概率分布最低值将越大;当阈值 β 越小,最高与最低的词概率分布的差值越小,即各候选序列间的词概率分布差异越不显著,在秘密信息控制下选择不同的候选序列时文本质量差异越小,即嵌入引起的文本质量下降越不明显.

为评估阈值 α 和 β 对应生成的隐写摘要文本质量的影响,本文先将 α 固定为具体值,再动态设置 β 为不同的值来生成隐写摘要文本,并对文本质量分

别进行评价. 根据对实验数据的观察，α 被固定为 0.1 和 0.3 进行实验. 当 $\alpha=0.1$ 时，β 值取值 0.2，0.4，0.6，0.8 时，利用 ROUGE 系列指标对所生成隐写摘要文本的质量进行评价的结果见表 1，当 $\alpha=0.3$ 时对应的结果见表 2.

表 1　$\alpha=0.1$ 时 β 在不同取值条件下隐写摘要文本的质量评价结果

评价指标		$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=0.8$
ROUGE-1	*P*	**0.3675**	0.3666	0.3660	0.3661
	R	**0.3823**	0.3792	0.3761	0.3695
	*F*1	**0.3610**	0.3592	0.3575	0.3545
ROUGE-2	*P*	**0.1582**	0.1566	0.1544	0.1496
	R	**0.1652**	0.1627	0.1594	0.1522
	*F*1	**0.1554**	0.1534	0.1508	0.1452
ROUGE-L	*P*	0.3369	0.3366	0.3368	**0.3383**
	R	**0.3499**	0.3477	0.3458	0.3410
	*F*1	**0.3306**	0.3295	0.3288	0.3273
ROUGE-W	*P*	**0.2485**	0.2480	0.2477	0.2470
	R	**0.1553**	0.1542	0.1531	0.1501
	*F*1	**0.1839**	0.1829	0.1820	0.1797
ROUGE-S	*P*	**0.1339**	0.1329	0.1318	0.1294
	R	**0.1431**	0.1403	0.1375	0.1299
	*F*1	**0.1202**	0.1186	0.1169	0.1139
ROUGE-SU	*P*	**0.1425**	0.1415	0.1405	0.1383
	R	**0.1527**	0.1499	0.1470	0.1395
	*F*1	**0.1286**	0.1271	0.1255	0.1215

表 2　$\alpha=0.3$ 时 β 在不同取值条件下隐写摘要文本的质量评价结果

评价指标		$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=0.8$
ROUGE-1	*P*	0.3669	0.3679	0.3669	**0.3680**
	R	**0.3823**	0.3792	0.3800	0.3769
	*F*1	**0.3607**	0.3598	0.3597	0.3601
ROUGE-2	*P*	**0.1576**	0.1574	0.1570	0.1574
	R	**0.1649**	0.1627	0.1632	0.1630
	*F*1	**0.1550**	0.1538	0.1538	0.1540
ROUGE-L	*P*	0.3363	0.3378	0.3369	**0.3379**
	R	**0.3499**	0.3477	0.3485	0.3482
	*F*1	0.3303	0.3300	0.3301	**0.3304**
ROUGE-W	*P*	0.2481	**0.2491**	0.2483	0.2490
	R	**0.1553**	0.1543	0.1545	0.1544
	*F*1	**0.1837**	0.1833	0.1832	0.1834
ROUGE-S	*P*	0.1333	0.1337	0.1330	**0.1338**
	R	**0.1428**	0.1401	0.1408	0.1405
	*F*1	**0.1199**	0.1189	0.1190	0.1192
ROUGE-SU	*P*	0.1419	**0.1424**	0.1416	**0.1424**
	R	**0.1525**	0.1497	0.1504	0.1500
	*F*1	**0.1283**	0.1274	0.1274	0.1277

如表 1 所示，当 α 固定时，隐写摘要文本的整体质量随着 β 值的增大而降低. 这主要因为当 β 值越大时，满足嵌入条件的候选嵌入位置越多，新增的候选嵌入位置所包含的候选序列间的概率分布差异越显著，因此，将导致生成的隐写摘要文本的质量下降. 表 2 也呈现了同样的变化规律. 对比表 1 和表 2 中相同 β 时的评价结果，可发现隐写摘要文本的整体质量随着 α 值的增大而提升，特别是当 β 值越大，提升的效果越明显. 这主要是因为当 α 值越大时，候选序列的概率分布值普遍更高，即隐写摘要文本质量越好，与原始摘要文本具有更强的语义相关性.

由于在每个候选嵌入位置选择候选序列时，已限定选择概率分布值最高的前 k 个序列，这些候选序列的概率分布值通常会比较高. 因此，对进一步筛选候选序列时阈值 α 的影响没有阈值 β 的影响大，特别是当 β 值越小，β 的作用越明显，使嵌入条件的要求越高；当 β 值越大，β 限定的条件越容易满足，此时阈值 α 的影响才越明显. 如表 2 和表 3 所示，当 $\beta=0.2,0.4$ 时，相同 β 不同 α 对应隐写摘要文本的质量差异不大；但当 $\beta=0.6,0.8$ 时，相同 β 但 α 不同时，$\alpha=0.3$ 对应隐写摘要文本的质量明显要好于 $\alpha=0.1$ 对应文本的质量，且 β 越大质量提高越多.

阈值 α 和阈值 β 对隐写摘要文本质量的影响主要在于它们的值将改变嵌入条件的判定，即本文方法自适应选择的候选序列和嵌入时刻将随之发生改变，从而导致隐写摘要文本中实际可嵌入秘密信息量的改变. 隐写文本中可嵌入秘密信息量，又称隐藏容量，是评估隐写编码方法的常见指标. 隐藏容量可以利用嵌入率 *ER*(Embedding Rate)[16] 与隐写文本长度的乘积来度量. 嵌入率固定，隐写文本长度越长，隐藏容量越大. 本文通过实际嵌入秘密信息的比特长度在对应隐写摘要文本的比特长度中所占比例的平均值来计算嵌入率，具体计算如式(19)所示：

$$ER=\frac{1}{N}\sum_{i=1}^{N}\frac{S_i}{L_i} \tag{19}$$

其中，S_i 表示第 i 个隐写摘要文本中嵌入的秘密信息比特长度，L_i 表示第 i 个隐写摘要文本内容的比特长度，N 表示数据集中隐写摘要文本总数. 嵌入率越高说明相同长度的隐写文本可嵌入秘密信息越多. 本文对不同 α 和 β 条件下的隐写摘要文本的嵌入率进行了统计，实验结果如表 3 所示.

表 3　不同阈值条件下生成的隐写摘要文本的嵌入率情况

	$\beta=0.2$	$\beta=0.4$	$\beta=0.6$	$\beta=0.8$
$\alpha=0.1$	0.000324	0.001067	0.002453	**0.006048**
$\alpha=0.3$	0.000323	0.001073	0.001071	0.001079

由于本文提出的自然语言信息隐藏方法主要根据是否为嵌入时刻对多候选结果进行优化，当某时刻为嵌入时刻时，该方法将对嵌入位置进行筛选再获得对应的多候选隐写优化结果. 因此，嵌入率相对

不高.从表 3 可以发现,当 β 相同时,α 越大,嵌入率越小,且随着 β 的增大,α 越大,嵌入率下降越多;当 α 相同时,β 越小,嵌入率越低,β 越大,嵌入率越高.其主要原因是因为阈值的变化将影响嵌入条件,当 α 越大或 β 越小时,满足嵌入条件的候选嵌入位置越少,从而嵌入时刻越少,导致嵌入率也越小,由表 3 所示,当 $\alpha=0.1$,$\beta=0.8$ 时,本文方法的嵌入率最高.

综合考虑隐写摘要文本的质量和嵌入率情况,后续实验中,本文设定阈值 α 为 0.1,阈值 β 为 0.2.

5.3.2 语言编码器的影响

序列到隐写序列模型为基本框架,不同的语言编码器对原始约束文本的特征信息学习能力不同,而导致生成的摘要文本的质量存在一定的差异.为了分析语言编码器对隐写摘要文本质量的影响,本文利用文献[44]所提出基于序列到序列的摘要生成模型记作基础编码器来生成摘要文本的方法与本文方法进行对比.基础编码器生成摘要文本时没有隐写编码模块,生成的摘要文本不受秘密信息控制,是未含秘密信息的正常摘要文本.为了方便对比,本文去掉隐写编码模块,基于 ARU-LiEncoder 语言编码器生成了未含秘密信息的正常摘要文本.对这两类正常摘要文本和本文方法所生成的隐写摘要文本(记为 ARU-LiEncoder-隐写摘要文本)的质量,利用 ROUGE 系列指标进行评价,评价结果见表 4.

表 4 不同语言编码器条件下对摘要文本质量的评估结果

评价指标		模型		
		基础编码器-摘要文本	ARU-LiEncoder-摘要文本	ARU-LiEncoder-隐写摘要文本
ROUGE-1	*P*	0.3336	**0.3679**	**0.3675**
	R	0.3111	**0.3883**	**0.3823**
	*F*1	0.3116	**0.3649**	**0.3610**
ROUGE-2	*P*	0.1427	**0.1604**	**0.1582**
	R	0.1373	**0.1695**	**0.1652**
	*F*1	0.1355	**0.1588**	**0.1554**
ROUGE-L	*P*	0.3169	**0.3360**	**0.3369**
	R	0.2959	**0.3543**	**0.3499**
	*F*1	0.2962	**0.3331**	**0.3306**
ROUGE-W	*P*	0.2380	**0.2481**	**0.2485**
	R	0.1342	**0.1574**	**0.1553**
	*F*1	0.1660	**0.1857**	**0.1839**
ROUGE-S	*P*	0.1255	**0.1345**	**0.1339**
	R	0.1125	**0.1470**	**0.1431**
	*F*1	0.1060	**0.1230**	**0.1202**
ROUGE-SU	*P*	0.1342	**0.1429**	**0.1425**
	R	0.1206	**0.1566**	**0.1527**
	*F*1	0.1139	**0.1315**	**0.1286**

从表 4 所示的实验数据可知,ARU-LiEncoder 生成的正常摘要文本在各项评价指标上都远远优于基础编码器,ARU-LiEncoder 通过注意力优化单元充分学习和考虑原始文本特征和原始文本对各单词的注意力分布情况,因此生成了与参考摘要文本语义更相关、质量更优异的摘要文本.同时,从实验结果可发现基于 ARU-LiEncoder 生成的隐写摘要文本质量要略差于 ARU-LiEncoder-摘要文本.这是因为隐写编码模块是在秘密信息的控制下生成的隐写摘要文本,即在秘密信息嵌入时将会导致不理想的低质量词被强制选中,从而影响最终隐写摘要文本的质量.对比基础编码器-摘要文本和 ARU-LiEncoder-隐写摘要文本的质量可发现,本文方法生成的隐写摘要文本质量显著优于基于基础编码器生成的正常摘要文本,这说明提升语言编码器的性能在一定程度上能提升本文方法生成的隐写文本质量.因此,为生成质量更优的隐写摘要文本,可设计和提升基于序列到隐写序列模型的语言编码器的性能.

5.3.3 隐写器的影响

以序列到隐写序列模型为基本框架,相同的语言编码器,隐写器使用不同的隐写编码方式将影响摘要文本生成过程中词的选择,在隐写摘要文本中引入不同程度的嵌入失真.引入的隐写嵌入失真越小,生成的隐写摘要文本质量越好.

为衡量提出的基于序列到隐写序列的自然语言信息隐藏方法的性能,本文设计了一组对比实验分析不同编码方式下的隐写器对生成的隐写摘要文本质量影响.所对比的编码方式包括三种:本文提出的基于多候选优化的自适应隐写编码方式 MOASC、分组编码[17]和动态分组编码.分组编码是一种将所有候选词分为 $2^{|B|}$ 个分组并分别编码每个分组的隐写编码方式,其中 B 指每个词嵌入的比特数.该编码方式与本文提出的 MOASC 具有类似的编码思想,都是根据模型预测的词概率分布情况对候选词分组进行编码,因此本文选择分组编码与 MOASC 进行对比.动态分组编码则是 MOASC 一种变型,该编码方式不判断某时刻是否为嵌入时刻,而是对每个时刻候选嵌入位置的各候选词序列求平均概率值,降序排列后进行隐写编码,最后选择与待嵌入秘密信息相同的平均概率值最高的 k 组输出.

对采用不同编码方式的隐写器所生成隐写摘要文本的质量进行评价的结果见表 5.从表 5 中的数据可发现,本文提出的基于多候选优化的自适应隐写编码方法 MOASC 的各项评价指标均高于其他两类隐写编码方法,基于 MOASC 的隐写器能生成

与参考摘要文本语义相似度更高的隐写摘要文本.而本文对比的分组编码方法是基于词的局部概率分布情况直接分组并进行编码,没有考虑候选词间的概率差异,而是对所有候选词进行统一长度的二进制字符串的编码,也未考虑长距离词间的影响从而输出多个候选序列供后续选择;动态分组编码方法尽管考虑了各候选嵌入位置中各候选序列的平均概率值来衡量各候选嵌入位置是否适合嵌入信息,但如若具有平均概率值高的各候选序列间的概率的差异较大,即概率分布不平均,则嵌入信息时可能选中低概率的候选序列,影响后续的文本质量.因此,考虑了多个长候选序列之间的概率差异,自适应选择不同多候选优化策略的 MOASC 方法比分组编码和动态分组编码具有更优的性能.采用更优编码方法的隐写器能使本文提出方法生成更高质量的隐写摘要文本.

表 5　不同隐写器条件下对隐写摘要文本质量的评价结果

评价指标		编码方式		
		分组编码[17]	动态分组编码	MOASC
ROUGE-1	*P*	0.3618	0.3651	**0.3675**
	R	0.3477	0.3514	**0.3823**
	*F*1	0.3424	0.3458	**0.3610**
ROUGE-2	*P*	0.1177	0.1235	**0.1582**
	R	0.1143	0.1200	**0.1652**
	*F*1	0.1118	0.1173	**0.1554**
ROUGE-L	*P*	0.3347	0.3386	**0.3369**
	R	0.3214	0.3258	**0.3499**
	*F*1	0.3167	0.3206	**0.3306**
ROUGE-W	*P*	0.2370	0.2406	**0.2485**
	R	0.1377	0.1400	**0.1553**
	*F*1	0.1680	0.1707	**0.1839**
ROUGE-S	*P*	0.1201	0.1233	**0.1339**
	R	0.1072	0.1110	**0.1431**
	*F*1	0.0993	0.1023	**0.1202**
ROUGE-SU	*P*	0.1294	0.1326	**0.1425**
	R	0.1168	0.1206	**0.1527**
	*F*1	0.1081	0.1111	**0.1286**

5.4　安全性分析

安全性是自然语言信息隐藏方法的一个重要指标.在理论上,本文提出的方法具有高安全性.攻击者若想要破译本文方法生成的隐写摘要文本中隐藏的秘密信息,那么需要满足以下条件:(1)在根据原始文本生成摘要的特定场景下,需要能察觉和分析出摘要文本中含有秘密信息,即检测出隐写摘要文本的存在;(2)获取序列到隐写序列模型的语言编码器和隐写器的具体结构以及参数配置情况,并保证训练出能使用与隐写摘要文本生成者一致的模型和参数;(3)通过对实验数据的分析,估算并确定正确的阈值 α 和 β.

对于条件(2)和(3),攻击者基本不可能通过穷举法或理论推导来达到条件,所需信息的一点细微的差别均会导致条件不满足,破译失败.对于条件(1),攻击者可利用隐写分析技术来检测隐写摘要文本.检测结果的各项指标越低说明隐写摘要文本抗隐写分析的能力越强,安全性越高.

为分析本文方法的抗隐写分析能力,本文利用 TS-GCN[48]、TS-RNN[49]、TS-CNN[50] 三种典型的隐写分析方法对本文方法生成的隐写摘要文本和生成的正常摘要文本进行分类,从而识别隐写摘要文本.隐写分析检测结果见表 6～表 8.

表 6　基于 TS-GCN 隐写分析方法下的安全性评估结果

不同阈值	*Acc*/%	*P*/%	*R*/%	*F*1/%
$\alpha=0.1,\beta=0.2$	54.04	53.81	57.07	55.39
$\alpha=0.1,\beta=0.4$	55.72	56.15	52.21	54.11
$\alpha=0.1,\beta=0.6$	56.06	56.08	50.41	53.10
$\alpha=0.1,\beta=0.8$	62.49	63.67	56.68	59.97
$\alpha=0.3,\beta=0.2$	53.93	53.12	51.90	52.51
$\alpha=0.3,\beta=0.4$	54.74	54.20	52.22	53.19
$\alpha=0.3,\beta=0.6$	54.77	54.48	48.46	51.30
$\alpha=0.3,\beta=0.8$	55.47	54.92	52.84	53.86

表 7　基于 TS-RNN 隐写分析方法下的安全性评估结果

不同阈值	*Acc*/%	*P*/%	*R*/%	*F*1/%
$\alpha=0.1,\beta=0.2$	48.98	46.90	51.27	51.16
$\alpha=0.1,\beta=0.4$	52.13	52.44	45.77	48.88
$\alpha=0.1,\beta=0.6$	50.67	50.65	52.51	51.56
$\alpha=0.1,\beta=0.8$	54.94	55.90	46.75	50.92
$\alpha=0.3,\beta=0.2$	51.35	51.25	55.27	53.18
$\alpha=0.3,\beta=0.4$	51.23	51.59	40.09	45.12
$\alpha=0.3,\beta=0.6$	51.08	51.22	45.47	48.18
$\alpha=0.3,\beta=0.8$	50.71	50.59	61.33	55.44

表 8　基于 TS-CNN 隐写分析方法下的安全性评估结果

不同阈值	*Acc*/%	*P*/%	*R*/%	*F*1/%
$\alpha=0.1,\beta=0.2$	51.38	51.60	44.50	47.79
$\alpha=0.1,\beta=0.4$	50.56	50.71	39.79	44.59
$\alpha=0.1,\beta=0.6$	51.42	51.66	44.20	47.64
$\alpha=0.1,\beta=0.8$	50.60	50.62	48.69	49.64
$\alpha=0.3,\beta=0.2$	52.02	52.37	44.58	48.16
$\alpha=0.3,\beta=0.4$	49.81	49.77	40.46	44.64
$\alpha=0.3,\beta=0.6$	49.14	48.98	41.44	44.89
$\alpha=0.3,\beta=0.8$	51.46	51.64	45.85	48.57

通过对比三种隐写分析方法下的检测数据,可发现 TS-GCN 隐写分析方法的各项检测指标均高于 TS-RNN 和 TS-CNN 隐写分析方法,说明本文方法在完全基于语义特征检测的隐写分析方法 GCN 下的抗隐写分析能力相对更差,其主要原因在于,GCN 隐写分析方法的检测能力最强.但即使是

GCN 隐写分析方法,各项检测指标基本低于 60%;其余两种基于统计分布特性的隐写分析方法的各项检测指标更低,均在 50%左右;这些检测结果说明已有隐写分析方法很难从正常摘要文本中识别出本文方法所生成的隐写摘要文本,本文方法的抗隐写分析能力较强,安全性较高.

综上所述,本文提出的自然语言信息隐藏方法的安全性极高,隐藏在隐写摘要文本中的秘密信息很难被发现,攻击者也难以实现对秘密信息的成功提取.

6 总 结

为了提高约束型自然语言信息隐藏方法的普适性,本文设计了一种基于序列到隐写序列的通用隐写框架,该框架由语言编码器和隐写器两部分组成,其中语言编码器主要是对原始输入的约束文本建模,而隐写器则是紧接着对语言编码器的输出进行解码并嵌入秘密信息.基于文本摘要生成任务和上述通用隐写框架,本文在语言编码器中融入了注意力优化单元,并在隐写器中设计了一种基于多候选优化的自适应隐写编码方法,实现了约束型的自然语言信息隐藏.我们从文本质量和安全性两个方面对本文提出的约束型自然语言信息隐藏方法的性能进行了分析和讨论.实验结果表明,该方法能生成与原始参考摘要文本语义相关性更强的隐写摘要文本,同时具有较强的抗隐写分析能力.甚至,生成的隐写摘要文本的质量要优于基于基础编码器的文本生成模型所生成的正常摘要文本的质量.

在今后的研究工作中,我们将对基于序列到隐写序列框架中的语言编码器和隐写器进行进一步的优化,而设计和实现更优秀的约束型自然语言信息隐藏方法,进一步提升词间、句间的语义相关性等来提升隐写摘要文本的质量.同时,我们也将探索更多其他形式的信息隐藏方法,如文本到文本的生成、数据到文本的生成、意义到文本的生成、图像到文本的生成以及多模态语言生成等,以生成更有意义、更自然、更安全的隐写文本.

致 谢 在此,我们向对论文提出宝贵意见的审稿专家们表示衷心的感谢!

参 考 文 献

[1] Liu T Y, Tsai W H. A new steganographic method for data hiding in Microsoft word documents by a change tracking technique. IEEE Transactions on Information Forensics and Security, 2007, 2(1): 24-30

[2] Chen X, Sun H, Tobe Y, et al. Coverless information hiding method based on the Chinese mathematical expression//Proceedings of the International Conference on Cloud Computing and Security. Nanjing, China, 2015: 133-143

[3] Yang Z, Jin S, Huang Y, et al. Automatically generate steganographic text based on Markov model and Huffman coding. arXiv preprint arXiv:1811.04720, 2018

[4] Meral H M, Sankur B, Özsoy A S, et al. Natural language watermarking via morphosyntactic alterations. Computer Speech & Language, 2009, 23(1): 107-125

[5] Muhammad H Z, Rahman S M S A A, Shakil A. Synonym based Malay linguistic text steganography//Proceedings of the 2009 Innovative Technologies in Intelligent Systems and Industrial Applications. Kuala Lumpur, Malaysia, 2009: 423-427

[6] Luo Y, Qin J, Xiang X, et al. Coverless real-time image information hiding based on image block matching and dense convolutional network. Journal of Real-Time Image Processing, 2020, 17(1): 125-135

[7] Zhang J, Shen J, Wang L, et al. Coverless text information hiding method based on the word rank map//Proceedings of the International Conference on Cloud Computing and Security. Nanjing, China, 2016: 145-155

[8] Chen X, Chen S, Wu Y. Coverless information hiding method based on the Chinese character encoding. Journal of Internet Technology, 2017, 18(2): 313-320

[9] Zhou X, Chen X, Zhang F, et al. A novel coverless text information hiding method based on double-tags and twice-send. International Journal of Computational Science and Engineering, 2020, 21(1): 116-124

[10] Grosvald M, Orgun C O. Free from the cover text: A human-generated natural language approach to text-based steganography. Journal of Information Hiding and Multimedia, 2011, 2(2): 133-141

[11] Xiang L Y, Wang R, Yang Z L, Liu Y L. Generative linguistic steganography: A comprehensive review. KSII Transactions on Internet and Information Systems, 2022, 16(3): 986-1005

[12] Wayner P. Mimic functions. Cryptologia, 1992, 16(3): 193-214

[13] Shniperov A N, Nikitina K A. A text steganography method based on Markov chains. Automatic Control and Computer Sciences, 2016, 50(8): 802-808

[14] Zhang S Y, Yang Z L, Yang J S, et al. Provably secure generative linguistic steganography//Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (ACL-IJCNLP 2021). Bangkok, Thailand, 2021: 3046-3055

[15] Zhou X J, Peng W L, Yang B Y, et al. Linguistic steganography based on adaptive probability distribution. IEEE Transactions on Dependable and Secure Computing, 2021, 19(5): 2982-2997

[16] Yang Z L, Guo X Q, Chen Z M, et al. RNN-stega: Linguistic steganography based on recurrent neural networks. IEEE Transactions on Information Forensics and Security, 2018, 14(5): 1280-1295

[17] Fang T, Jaggi M, Argyraki K. Generating steganographic text with LSTMs//Proceedings of the Association for Computational Linguistics 2017, Student Research Workshop. Vancouver, Canada, 2017: 100-106

[18] Yang Z L, Xiang L Y, Zhang S Y, et al. Linguistic generative steganography with enhanced cognitive-imperceptibility. IEEE Signal Processing Letters, 2021, 28: 409-419

[19] Luo Y B, Huang Y F. Text steganography with high embedding rate: Using recurrent neural networks to generate Chinese classic poetry//Proceedings of the ACM Workshop on Information Hiding and Multimedia Security. USA, 2017: 99-104

[20] Desoky A. Jokestega: Automatic joke generation-based steganography methodology. International Journal of Security & Networks, 2012, 7(3): 148-160

[21] Yang R, Ling Z H. Linguistic steganography by sampling-based language generation//Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). Lanzhou, China, 2019: 1014-1019

[22] Wen J, Zhou X J, Li M D, et al. A novel natural language steganographic framework based on image description neural network. Journal of Visual Communication and Image Representation, 2019, 61: 157-169

[23] Yang Z L, Gong B T, Li Y M, et al. Graph-stega: Semantic controllable steganographic text generation guided by knowledge graph. arXiv preprint arXiv:2006.08339, 2020

[24] Chapman M, Davida G. Hiding the hidden: A software system for concealing ciphertext as innocuous text//Proceedings of the International Conference on Information and Communications Security. Beijing, China, 2005: 335-345

[25] Yu Zhen-Shan, Huang Liu-Sheng, Chen Zhi-Li, et al. High embedding ratio text steganography by ci-poetry of the Song dynasty. Journal of Chinese Information Processing, 2009, 23(4): 55-63(in Chinese)
(余振山，黄刘生，陈志立等. 用宋词实现高嵌入率文本信息隐藏. 中文信息学报，2009，23(4)：55-62)

[26] Luo Y B, Huang Y F, Li F F, et al. Text steganography based on ci-poetry generation using Markov chain model. KSII Transactions on Internet and Information Systems, 2016, 10(9): 4568-4584

[27] Wu N, Yang Z L, Yang Y, et al. STBS-stega: Coverless text steganography based on state transition-binary sequence. International Journal of Distributed Sensor Networks, 2020, 16(3): 155014772091425

[28] Dai W H, Yu Y, Deng B. Bintext steganography based on Markov state transferring probability//Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human, New York, USA, 2009: 1306-1311

[29] Dai W H, Yu Y, Deng B. Text steganography system using Markov chain source model and DES algorithm. Journal of Software, 2010, 5(7): 785-792

[30] Wu N, Ma W B, Liu Z R, et al. Coverless text steganography based on half frequency crossover rule//Proceedings of the 2019 4th International Conference on Mechanical, Control and Computer Engineering(ICMCCE 2019). Hohhot, China, 2019: 726-729

[31] Wu N, Liu Z, Ma W, et al. Research on coverless text steganography based on multi-rule language models alternation//Proceedings of the 2019 4th International Conference on Mechanical, Control and Computer Engineering (ICMCCE 2019). Hohhot, China, 2019: 803-806

[32] Xiang L Y, Yang S H, Liu Y H, et al. Novel linguistic steganography based on character-level text generation. Mathematics, 2020, 8(5): 1558

[33] Ziegler Z M, Deng Y T, Rush A M. Neural linguistic steganography//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China, 2019: 1210-1215

[34] Dai F, Cai Z. Towards near-imperceptible steganographic text //Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. Florence, Italy, 2019: 4303-4308

[35] Shen J, Ji H, Han J. Near-imperceptible neural linguistic steganography via self-adjusting arithmetic coding//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing(EMNLP). Online, 2020: 303-313

[36] Yang Z L, Zhang S Y, Hu Y T, et al. VAE-Stega: Linguistic steganography based on variational auto-encoder. IEEE Transactions on Information Forensics and Security, 2021, 16: 880-895

[37] Yang Z L, Wei N, Liu Q H, et al. GAN-TStega: Text steganography based on generative adversarial networks//Proceedings of the 18th International Workshop on Digital Forensics and Watermarking (IWDW). Chengdu, China, 2020, 12022: 18-31

[38] Desoky A, Younis M. Chestega: Chess steganography methodology. Security and Communication Networks, 2009, 2(6): 555-566

[39] Desoky A. Notestega: Notes-based steganography methodology. Information Security, 2009, 18(4): 178-193

[40] Desoky A. Edustega: An education-centric steganography methodology. International Journal of Security & Networks, 2011, 6(2/3): 153-173

[41] Desoky A. Matlist: Mature linguistic steganography methodology. Security and Communication Networks, 2011, 4(6): 697-718

[42] Li Y M, Zhang J, Yang Z L, et al. Topic-aware neural linguistic steganography based on knowledge graphs. ACM/IMS Transactions on Data Science, 2021, 2(2): 1-13

[43] Li M, Wu K, Zhong P, et al. Generating steganographic image description by dynamic synonym substitution. Signal Processing, 2019, 164: 193-201

[44] Nallapati R, Zhou B, Santos C D, et al. Abstractive text summarization using sequence-to-sequence RNNs and beyond //Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL). Berlin, Germany, 2016: 280-290

[45] Gui M, Tian J, Wang R, et al. Attention optimization for abstractive document summarization//Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Hong Kong, China, 2019, 1: 1222-1228

[46] Hermann K M, Kočiský T, Grefenstette E, et al. Teaching machines to read and comprehend//Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge, USA, 2015, 1: 1693-1701

[47] Lin C Y. ROUGE: A package for automatic evaluation of summaries//Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004). Barcelona, Spain, 2004: 74-81

[48] Wu H Z, Yi B, Ding F, et al. Linguistic steganalysis with graph neural networks. IEEE Signal Processing Letters, 2021, 28: 558-562

[49] Yang Z L, Wang K, Li J, et al. TS-RNN: Text steganalysis based on recurrent neural networks. IEEE Signal Processing Letters. 2019, 26(12): 1743-1747

[50] Yang Z L, Wei N, Sheng J Y, et al. TS-CNN: Text steganalysis from semantic space based on convolutional neural network. arXiv preprint arXiv: 2106. 02011, 2021

附 录.

附录列出了实验中本文提出方法在不同参数下所生成隐写摘要文本的样例，以及参考摘要文本和正常生成摘要文本样例. 参考摘要文本是数据集中给出的人工书写的摘要；正常生成摘要文本则是基于贪心采样策略生成的未嵌入秘密信息的摘要文本，即每个时刻选择预测概率最大的单词输出. 隐写摘要文本中加粗单词(包括标点符号)为嵌入时刻，每嵌入时刻随机嵌入 1 比特秘密信息.

例 1.

原始文本	The flight crew of the Delta Air Lines plane that skidded into a fence at LaGuardia Airport last week cited brake issues during the landing, according to an update on Monday from the NTSB. The crew said they did not sense any deceleration from the wheel brake upon landing, despite the auto brakes being set to "max," according to an ongoing investigation by the National Transportation Safety Board. The runway appeared all white in the moments before landing, according to the report. They based their decision to land after receiving a brake action report of "good" from air traffic control, the NTSB said. "The automatic spoilers did not deploy," the crew told the NTSB, "But that the first officer quickly deployed them manually." The captain said he was unable to stop the aircraft from drifting left, according to the report. The Boeing MD-88 sustained significant damage to the left wing, flight spoilers, the nose of the plane and the left-wing fuel tank, according to the NTSB. Delta Flight 1086 departed from Atlanta shortly after 9 a. m. Thursday. LaGuardia was dealing with snow and freezing fog as the flight approached its destination about two hours later. The aircraft briefly circled New York because of issues with snow and ice before touching down shortly after 11 a. m. The plane slid off the runway with its nose busting through a fence before skidding to a halt mere foot from frigid water. Twenty-three passengers received minor injuries, and others were transported to the hospital for evaluation. An NTSB meteorologist is examining the weather conditions at the time of the accident, said the report. The cause of the accident has not been determined.
参考摘要文本	Delta Air Lines flight 1086 skidded into a fence last week at a LaGuardia Airport beset by winter weather. The NTSB says the crew reported they did not sense any deceleration from the wheel brake upon landing.
正常生成摘要文本	The flight crew of the Delta Air Lines plane skidded into a fence. The captain skidded into a fence at LaGuardia Airport last week cited brake issues during the landing. The captain said he was unable to stop the aircraft from drifting left, according to the report.
	本文隐写摘要文本
$\alpha=0.1$ $\beta=0.2$	嵌入秘密信息：11 The flight crew of the Delta Air Lines plane that skidded into a fence. The **captain skidded** into a fence at LaGuardia Airport last week cited brake issues during the landing. The captain said he was unable to stop the aircraft from drifting left, according to the report.
$\alpha=0.1$ $\beta=0.4$	嵌入秘密信息：001 The flight crew of the Delta Air Lines plane skidded into a fence. The captain **skidded** into a fence at LaGuardia Airport last week cited brake issues **during** the landing. They **based** their decision to land after receiving a brake action report of "good".
$\alpha=0.1$ $\beta=0.6$	嵌入秘密信息：010110 The flight crew of the Delta Air Lines plane skidded into a fence. **Crew says they** did not sense any deceleration from the wheel brake **upon** landing, **despite** the auto brakes being set to "max," **crew** says.

（续　表）

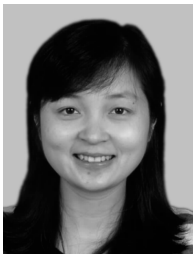
$\alpha=0.1$ $\beta=0.8$	嵌入秘密信息：111001011100 The **flight** crew **of** the Delta Air Lines plane that skidded into a fence. **The captain skidded** into a fence at LaGuardia Airport last week **cited** issues. **Crew told** the **NTSB said they** did not sense any deceleration from the wheel brake upon landing.
$\alpha=0.2$ $\beta=0.2$	嵌入秘密信息：100 The flight crew of the Delta Air Lines plane that skidded into a fence. The **crew said** they did not sense any deceleration from the wheel brake upon landing. The **captain** said he was unable to stop the aircraft from drifting left, according to the report.
$\alpha=0.2$ $\beta=0.4$	嵌入秘密信息：11100 The flight crew of the Delta Air Lines plane that skidded into a fence. The **captain said** he was unable to stop the aircraft from drifting left. **The crew told** the NTSB said he was unable to stop the aircraft from drifting left.
$\alpha=0.2$ $\beta=0.6$	嵌入秘密信息：00110 The flight crew of the Delta Air Lines plane skidded into a fence. **The crew said they** did not sense any deceleration from the wheel brake upon landing, **despite** the auto brakes being set to "max," crew says.
$\alpha=0.2$ $\beta=0.8$	嵌入秘密信息：101001 The flight crew of the Delta Air Lines plane that skidded into a fence. The **captain skidded** into a fence at LaGuardia Airport last week cited brake issues **during** the landing. The **captain said** he was unable to stop the aircraft from drifting left, according to the report.

例 2.

原始文本	Actress Liza Minnelli has entered a rehab facility for her addictions, a spokesman said. "Liza Minnelli has valiantly battled substance abuse over the years and whenever she has needed to seek treatment she has done so," said spokesman Scott Gorenstein. "She is currently making excellent progress at an undisclosed facility." The 68-year-old has struggled with addictions to alcohol and painkillers in the past. Minnelli won an Oscar in 1973 for her performance in "Cabaret." CNN's Topher Gauk-Roger contributed to this report.
参考 摘要文本	Minnelli won an academy award for her performance in "Cabaret." She has struggled with addictions to alcohol and painkillers in the past.
正常生成 摘要文本	Minnelli won an Oscar in 1973 for her performance in "Cabaret." CNN's Topher Gauk-Roger contributed to this report. "She is currently making excellent progress at an undisclosed facility" The 68-year-old has struggled with addictions to alcohol and painkillers in the past.
	本文隐写摘要文本
$\alpha=0.1$ $\beta=0.2$	嵌入秘密信息：1 Minnelli won an Oscar in 1973 for her performance in "Cabaret." **CNN**'s Topher Gauk-Roger contributed to this report. "She is currently making excellent progress at an undisclosed facility" The 68-year-old has struggled with addictions to alcohol and painkillers in the past.
$\alpha=0.1$ $\beta=0.4$	嵌入秘密信息：01 Minnelli won an Oscar in 1973 for her performance in "Cabaret." **CNN**'s Topher Gauk-Roger contributed to this report. "**Liza** Minnelli has valiantly battled substance abuse over the years and whenever she has needed to seek treatment she has done so," said spokesman Scott Gorenstein.
$\alpha=0.1$ $\beta=0.6$	嵌入秘密信息：1111 Minnelli **won** an Oscar in **1973** for her performance in "Cabaret." **CNN**'s Topher Gauk-Roger contributed to this report. "**Liza** Minnelli has valiantly battled substance abuse over the years and whenever she has needed to seek treatment she has done so," said spokesman Scott.
$\alpha=0.1$ $\beta=0.8$	嵌入秘密信息：011000000010011 Actress **Liza** Minnelli **has been entered rehab** facility for **alcohol**. She **has struggled** with addictions to alcohol and painkillers in the past. **CNN**'s **Topher says she is currently** making excellent progress at an undisclosed facility. **Minnelli** won an Oscar in "Cabaret.".
$\alpha=0.2$ $\beta=0.2$	嵌入秘密信息：0 Minnelli won an Oscar in 1973 for her performance in "Cabaret." **CNN**'s Topher Gauk-Roger contributed to this report. "She is currently making excellent progress at an undisclosed facility". The 68-year-old has struggled with addictions to alcohol and painkillers in the past.
$\alpha=0.2$ $\beta=0.4$	嵌入秘密信息：11 Minnelli won an Oscar in 1973 for her performance in "Cabaret." **CNN**'s Topher Gauk-Roger contributed to this report. "**Liza** Minnelli has valiantly battled substance abuse over the years and whenever she has needed to seek treatment she has done so," said spokesman Scott.
$\alpha=0.2$ $\beta=0.6$	嵌入秘密信息：0010 Actress **Liza** Minnelli has entered a rehab facility for her addictions. The 68-year-old has struggled with addictions to alcohol and painkillers in the past. Minnelli won an Oscar in 1973 for her performance in "Cabaret.". "**The** 68-year-old has struggled with addictions to alcohol and painkillers in the past.
$\alpha=0.2$ $\beta=0.8$	嵌入秘密信息：100100 Minnelli **won** the Oscar in **1973** for her performance in "Cabaret". "**The** 68-year-old has struggled with addictions to alcohol and painkillers in **the** past. **CNN**'s Topher Gauk-Roger contributed to this report.

例 3.

原始文本	The father of baby Lily, found by rescuers after her mother's car flipped into a river, says she's doing great and that he feels blessed. Rescuers found the toddler Saturday hanging upside down in the car, which had crashed into a frigid Utah river a day before. Lily's mother, Lynn Jennifer Groesbeck, died in the crash that had landed their car on its roof in the Spanish Fork River. She was 25. Deven Trafny, 34, was out of town on a job at the time of the accident, CNN affiliate KUTV reported. He rushed to his daughter's side as soon as he heard. "Came in, I put my finger in her hand, and I told her her Dad was here, and I love her," he told reporters Wednesday. "I haven't left her bedside since, and I've just been here just sitting next to her waiting for her to get better so she can come home." Trafny said that Lily is awake and has been singing nursery rhymes. Video of the two of them at a hospital shows her waving at a camera. "She knows everything she knew before anything happened. It's amazing. Doctors say it's amazing," he said. How did toddler survive a car crash in Utah river? Lily might have died unseen with her mother had a man not gone fishing in that particular spot Saturday. The angler waded into the river around noon, then noticed the car wheels-up in the water. The fisherman called emergency dispatches. The water was so cold that, when the rescue was over, seven of the men involved had to be treated for hypothermia. They heaved the car onto its side and saw Groesbeck in the driver's seat. It was clear to them that she was dead. Lily was still strapped into her seat, where she may have been for 14 hours, if the wreck occurred at about 10: 30 Friday night, as police believe. Trafny described Groesbeck as the love of his life, according to KUTV: "I'm going to miss her a lot. I still have to deal with that." But he also considers himself lucky to still have his daughter-- healthy and alive. "I'm just blessed. I'm counting all my blessings right now," said Trafny. CNN's Ben Brumfield contributed to this report.
参考摘要文本	Baby Lily was trapped in the car in the cold river for as long as 14 hours. "She knows everything she knew before anything happened. It's amazing," father says. Lily's mother, Lynn Jennifer Groesbeck, died in the crash.
正常生成摘要文本	Lily's mother, Lynn Jennifer Groesbeck, died in the crash that had landed their car on its roof in the Spanish Fork River. She was out of town on a job at the time of the accident, CNN affiliate KUTV reported.
	本文隐写摘要文本
$\alpha=0.1$ $\beta=0.2$	嵌入秘密信息:1 Lily's mother, Lynn Jennifer Groesbeck, died in the crash that had landed their car on its roof in the Spanish Fork River. She was out of town on **a** job at the time of the accident, CNN affiliate KUTV reported.
$\alpha=0.1$ $\beta=0.4$	嵌入秘密信息:101 Lily's mother, Lynn Jennifer Groesbeck, died in the car **flipped** into a river. She had **crashed** into a frigid Utah river a day before. Deven Trafny, 34, was out of town on **a** job at the time of the accident.
$\alpha=0.1$ $\beta=0.6$	嵌入秘密信息:10010001 Lily's mother, Lynn Jennifer Groesbeck, died in the crash, had landed their car on its roof in the Spanish Fork River. **She was 25. Trafny**, 34, was out of town on **a** job at the time of the accident, CNN affiliate KUTV reported.
$\alpha=0.1$ $\beta=0.8$	嵌入秘密信息:0110011000011110 Lynn Jennifer Groesbeck **had** landed their car on its roof **in the** Spanish Fork River. **He was 25. He rushed** to **his** daughter's side **as soon** as **he** heard. He **told her dad** was here, and "I love her", he says.
$\alpha=0.2$ $\beta=0.2$	嵌入秘密信息:1 Lily's mother, Lynn Jennifer Groesbeck, died in the crash that had landed their car on its roof in the Spanish Fork River. She was out of town on **a** job at the time of the accident, CNN affiliate Kutv reported.
$\alpha=0.2$ $\beta=0.4$	嵌入秘密信息:100 Lily's mother, Lynn Jennifer Groesbeck, died in the car **flipped** into a river. She had **crashed** into a frigid Utah river a day before. Deven Trafny, 34, was out of town at **the** time of the accident.
$\alpha=0.2$ $\beta=0.6$	嵌入秘密信息:1100111 Lily's mother, Lynn Jennifer Groesbeck, died in the car, landed in the crash **that** had landed their car on its roof in the Spanish Fork River. He **told reporters** Wednesday **is** awake and has **crashed** into a frigid Utah river a day.
$\alpha=0.2$ $\beta=0.8$	嵌入秘密信息:1110000011 Lily's mother, Lynn Jennifer Groesbeck, died in the car **flipped** into a river. **She had landed** their car on its roof in the **Spanish** Fork River. She **was 25**, **Trafny**, 34, was out of town on a job at the time of the accident.



XIANG Ling-Yun, Ph. D., associate professor. Her research interests include information security, information hiding, digital watermarking, steganalysis, and natural language processing.

YANG Shuang-Hui, M. S. candidate. His research interests include natural language processing and information hiding.

WANG Rong, M. S. candidate. Her research interests include natural language processing and information hiding.

LIU Yu-Hang, M. S. candidate. His research interests include natural language processing and steganalysis.

ZHANG Deng-Yong, Ph. D., associate professor. His research interests include multimedia information security, image processing, pattern recognition, image forensics and information hiding.

Background

As a research interest that lies at the intersection of natural language processing and information security, linguistic steganography, also known as natural language information hiding, plays a crucial role in many important applications such as military, commercial and personal communications for privacy and secret message protection. With the development of neural language model, generative linguistic steganography, which automatically generates steganographic texts to conceal the secret message by combining text generation techniques and steganographic coding methods, has made significant progress in the improvement of hidden capacity. The existing methods mainly focus on controlling and planning the generation process of steganographic texts without any other constraint information. However, they always struggle to generate long steganographic texts, whose quality deteriorate sharply as the text length increases. This can easily lead to poor imperceptibility and security. Therefore, this paper proposes a general sequence to steganographic sequence framework to improve the quality of the generated long steganographic texts by incorporating special constraint information. The framework is composed of a Linguistic-Encoder and a Steganographic-Encoder. On the basis of the proposed framework, a new constraint linguistic steganographic method is proposed by designing and optimizing the Linguistic-Encoder and Steganographic-Encoder. The steganographic texts generated by our method have better qualities than those of normal texts and steganographic texts generated by some other methods. They achieve higher imperceptibility and anti-steganalysis capability.

This paper is supported in part by the National Natural Science Foundation of China project "Research on Multi-Granularity Natural Language Information Hiding with High Capacity and Security" under Grant No. 61972057, which aims at designing advanced natural language information hiding methods with both large capacity and high security, and studying the construction of word-level, phrase-level and sentence-level multi-granularity embedding domains, the design of steganographic distortion function, multi-domain combined information embedding and text selection with high anti-steganalysis capability. And it is also supported in part by the Hunan Provincial Natural Science Foundation of China project "Research on Generative Steganographic Text Detection Based on Domain Adaptation" under Grant No. 2022JJ30623, which focuses on utilizing domain adaptation methods to carry out efficient detection of mismatched and unknown multi-source generative steganographic text, aiming to improve the practicability and generality of linguistic steganalysis. Our research group mainly paid attention to the field of linguistic steganography and steganalysis in past years. More details can be found on the author's publications.