

DOI: 10.20189/j.cnki.CN/61-1527/E.202501013

基于深度学习的生成式文本隐写研究综述

施 昊, 郭文普*, 陈 昊
(火箭军工程大学, 陕西 西安 710025)

摘 要: 针对公开信道下文本信息传输易被敌方发现和窃取的问题, 可通过文本隐写术实现隐蔽通信。相较于传统的文本隐写算法, 基于深度学习的生成式文本隐写算法在隐藏容量、文本质量和安全性等方面更具优势。基于此, 通过结合自然语言处理技术和隐写编码算法, 对基于深度学习的生成式文本隐写算法的研究进展进行归纳总结。首先, 对生成式文本隐写算法的特点和一般流程进行介绍, 分析了生成式文本隐写算法信息隐藏和信息提取 2 个过程, 并推广到基于深度学习的生成式文本隐写模型中, 同时对相关的评价指标进行了分类介绍; 然后, 区分文本生成模型、隐写编码方式和应用场景等不同方向, 将现有算法从神经网络模型、语义和安全性 3 个方面进行综述, 阐述了各算法的实现原理和相互联系, 并对现有算法的性能进行对比和分析; 最后, 从隐藏容量、文本质量、语义和安全性 4 个方面对现有工作进行总结, 基于当前存在的问题, 提出相应的解决思路并对生成式文本隐写算法的未来发展趋势进行了展望。

关键词: 网络安全; 信息隐藏; 深度学习; 生成式文本隐写

中图分类号: TP309

文献标志码: A

开放科学标识码(OSID):

文章编号: 2097-2741(2025)02-124-13



听语音
与作者互动
聊科研

Review on Deep Learning-based Generative Text Steganography

SHI Hao, GUO Wenpu*, CHEN Hao

(Rocket Force University of Engineering, Xi'an 710025, China)

ABSTRACT: To address the issue of text information transmission being easily detected and stolen by the enemy on public channels, covert communication can be achieved through text steganography. Compared with traditional text steganography algorithms, generative text steganography based on deep learning has advantages in terms of hidden capacity, text quality, and security. Therefore, the research progress of generative text steganography algorithms based on deep learning was summarized in this paper by combining natural language processing technologies and steganographic encoding algorithms. First, the characteristics and general process of the generative text steganography algorithms were summarized, and the processes of information hiding and extraction in generative text steganography algorithms were analyzed, which were all extended to

收稿日期: 2024-11-07

第一作者: 施昊 (2001—), 男, 硕士研究生, 主要从事文本隐写研究。E-mail: shzxx109829@163.com

***通信作者:** 郭文普 (1976—), 男, 副教授, 主要从事网络与信息安全研究。E-mail: gwp_403@163.com

引用格式: 施昊, 郭文普, 陈昊. 基于深度学习的生成式文本隐写研究综述[J]. 火箭军工程大学学报, 2025, 39 (2): 124-136. SHI Hao, GUO Wenpu, CHEN Hao. Review on deep learning-based generative text steganography[J]. Journal of Rocket Force University of Engineering, 2025, 39 (2): 124-136.

generative text steganography models based on deep learning. Relevant evaluation indicators were categorized and comprehensively introduced. Then, from different directions such as text generation models, steganographic encoding methods, and application scenarios, the existing algorithms were summarized from three aspects: neural network models, semantics, and security. The implementation principles and interrelationships of each algorithm were elaborated, and performances of existing algorithms were compared and analyzed. Finally, the existing work was summarized in terms of four aspects: hidden capacity, text quality, semantics, and security. Based on existing problems, corresponding solutions were proposed, and future development trends were discussed.

KEYWORDS: network security; information hiding; deep learning; generative text steganography

随着网络技术和信息数字化的不断发展, 数字通信在日常生活中占据着重要地位。文本作为信息传输最广泛的载体之一, 由于网络的开放性, 导致文本信息被恶意窃取的情况层出不穷, 因此对文本隐藏技术的研究具有十分重要的现实意义。加密作为文本隐藏技术的一种, 虽能够较好地保护信息内容, 但因其不具备隐蔽性, 易被敌方察觉, 因而难以实现文本信息在公开信道下进行安全传输的目的。为满足文本信息安全传输的需要, 文本隐写术成为文本隐藏技术领域的研究热点。

文本隐写术是一种将秘密信息隐藏在文本之中, 在传输时不易被敌方察觉其存在的隐藏技术。文本隐写术分为3种: 基于文本修改的隐写术、基于文本选择的隐写术和基于文本生成的隐写术。基于文本修改的隐写术主要通过修改文本特征或内容来隐藏秘密信息, 如改变文本间距^[1]、改变语法结构^[2]、同义词替换^[3]、利用语言特征^[4]等。基于文本修改的隐写术存在隐藏容量小和安全性的问题, 在面对特定的隐写分析算法攻击下, 几乎无法隐藏其秘密信息^[5]。基于文本选择的隐写术通过构建特定的文本数据库, 根据秘密信息与所构建文本数据库之间的对应关系, 挑选尽可能合适的文本作为载体^[6]。基于文本选择的隐写术所选文本与所使用的文本数据库直接相关, 无法根据语境选择载体信息, 并且当秘密信息为长文本时, 导致该秘密信息需要大量的自然文本进行组合来将其进行隐藏, 会出现文本质量差和语义不一致等情况。基于文本生成的隐写术不需要事先构建数据库, 而是利用自然语言处理(Natural Language Processing, NLP)技术, 根据秘密信息生成隐写文本^[7]。然而, 早期基于规则的生成式文本隐写算法和基于马尔可夫的生成式

文本隐写算法均根据算法流程逐个生成文本来隐藏秘密信息, 导致文本在质量、语义和安全性等方面表现较差。

近年来, 深度学习(Deep Learning, DL)在NLP领域中取得重大进展, 其优势在于通过构建具有多层结构的神经网络, 自动提取文本数据中的复杂特征, 捕捉长距离依赖关系, 从而生成连贯且符合语境的文本。基于上述相关背景, 本文首先对生成式文本隐写术的相关概念进行阐述; 然后, 从神经网络模型、语义和安全性3个方面对现有的基于深度学习的生成式文本隐写方法进行了归纳总结; 最后, 对该领域的未来发展趋势进行了展望。

1 生成式文本隐写术的相关概念

1.1 生成式文本隐写术模型

1984年, Simmons提出了“囚牢模型”^[8], 揭开了隐写术发展的序幕。在囚牢模型中, 2名囚犯使用隐写术隐藏秘密信息, 在监狱长的监视下进行传递, 从而策划一个越狱计划。以文本为载体, 对该模型进行建模, 如式(1)所示^[9]。

$$\begin{cases} \text{Emb: CM} \times \text{K} \times \text{SM} \rightarrow \text{ST} \\ \text{Ext: ST} \times \text{K} \rightarrow \text{SM} \end{cases} \quad (1)$$

式中: Emb为信息隐藏; Ext为信息提取; CM为载体文本, 用于隐藏秘密信息; K为密钥, 用于对信息进行加密或包含双方共享的信息; SM为待隐藏的秘密信息; ST为隐写文本。

由式(1)构建文本隐写术模型, 如图1所示。其中, 发送方需要将秘密信息通过公开信道传输给接收方。首先, 发送方需要选择合适的载体文本, 随后在嵌入算法和密钥的映射关系下, 根据

秘密信息生成隐写文本。然后，在公开信道传输过程中，敌方利用隐写分析算法等攻击手段对传递信息进行检测，判断其是否含有秘密信息。最后，接收方得到隐写文本后，通过已知的提取算法和密钥进行信息提取，得到隐写文本中的秘密信息。

在上述模型的基础上，将其推广到基于深度学习的生成式文本隐写模型中，生成过程如图2所示。在隐写文本生成阶段，通过训练好的神经网络模型进行词预测，构建候选池。随后在秘密信息、编码策略和密钥的约束下，从候选池中选择符合要求的词元，生成最终的隐写文本，提取过程则为生成过程的逆过程。

1.2 评价指标

对于生成式文本隐写算法，主要通过隐藏容量、不可感知性和不可检测性3个方面来评价其性

能。在算法的设计过程中，难点在于权衡三者之间的关系，在提高任一指标的同时，可能会导致其他指标的下降。

1.2.1 隐藏容量

隐藏容量是衡量文本中隐藏秘密信息的比特数量，用嵌入率（Embedding Rate，ER）来表示。通常来说，在较少的文本中嵌入更多的信息是反映算法隐藏容量高的重要体现。其具体量化方式有以下2种。

1) 单个词元或字符的嵌入比特数（Bits Per Word，BPW）^[10]。BPW是由隐藏秘密比特数与隐写文本中除无关不可见字符之外的词元数或字符数之比来确定，表达式为

$$ER = \frac{1}{N} \sum_{i=1}^N \frac{B_i}{S_i}$$

(2)

式中： N 为隐写文本中词元或字符的总数； B_i 为

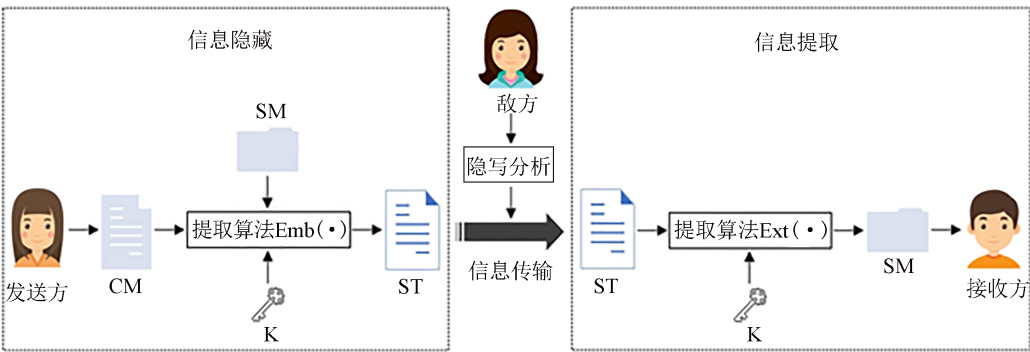


图 1 文本隐写术模型
Fig. 1 Text steganography model

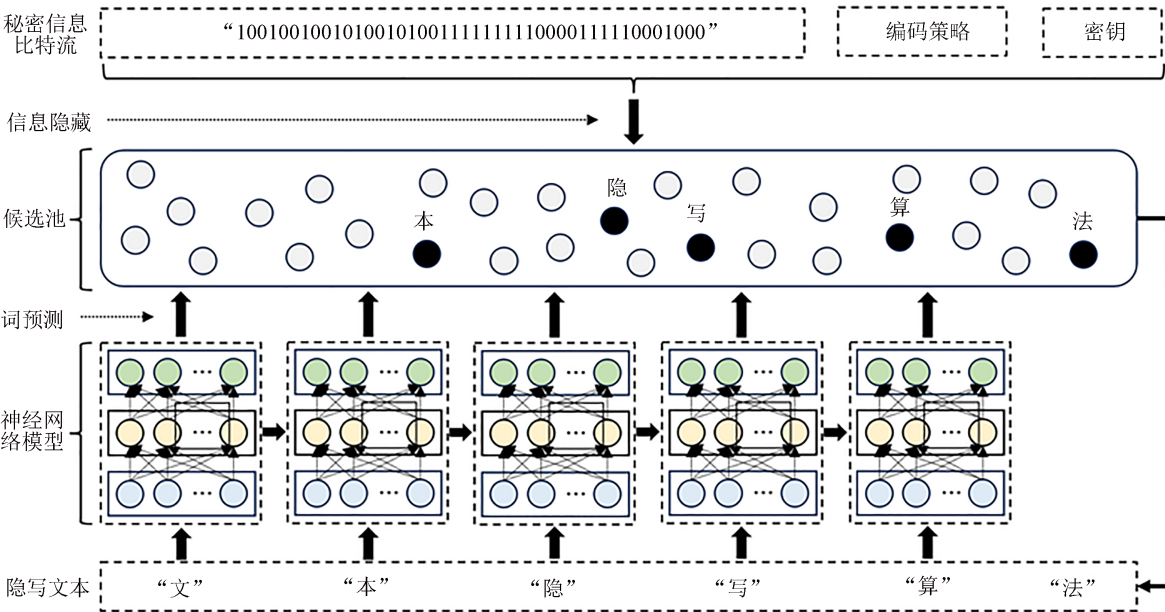


图 2 基于深度学习的生成式文本隐写算法生成文本过程
Fig. 2 Process of text generation by deep learning-based generative text steganography

第 i 个词元或字符中嵌入的秘密信息比特数; S_i 为第 i 个词元或字符的比特长度。

2) 实际嵌入比特数与隐写文本所占比特数的比值, 表达式^[9]为

$$ER = \frac{1}{M} \sum_{i=1}^M \frac{(L_i - 1) \cdot k}{B(s_i)} \quad (3)$$

式中: M 为隐写文本句子总数; L_i 为第 i 个句子的长度; k 为每个单词嵌入的比特数; $B(s_i)$ 为第 i 个句子所占用的比特数。

1.2.2 不可感知性

不可感知性是衡量隐写文本是否在语义上完整、在语序上顺畅, 从而判断能否被肉眼所察觉的指标。在评价过程中, 通常采用主观和客观 2 种方式进行。主观评价是指通过人工进行观察, 评价隐写文本是否与自然文本近似^[11]; 客观评价则使用困惑度 (Perplexity, PPL) 来衡量生成文本模型的隐蔽性能优劣, 表达式^[12]为

$$PPL = 2^{-\frac{1}{n} \log(p(s_i))} = 2^{-\frac{1}{n} \log(p(w_1, w_2, \dots, w_n))} \quad (4)$$

式中: n 为单词总数; $s_i = w_1, w_2, \dots, w_n$ 为生成的句子; $p(s_i)$ 为句子中单词的概率分布, 通过生成文本的语言模型来计算。PPL 实际上是计算隐写文本和训练文本之间语言模型的统计分布差异, 其值越小, 表明隐写文本与训练文本越接近。

随着文本隐写技术对语境要求变高以及隐写分析等检测技术的不断发展, 要求隐写文本的语义必须满足当前语境, 从而确保语义一致性。针对语义一致性采用的指标还有 BLEU (Bilingual Evaluation Understudy)^[13]、METEOR (Metric for Evaluation of Translation with Explicit Ordering)^[14]和 ROUGE (Recall Oriented Understudy for Gisting Evaluation)^[15], 以上 3 种指标值越大, 表明算法生成的隐写文本更加符合当前语境要求。

1.2.3 不可检测性

不可检测性是衡量隐写算法抵抗隐写分析的能力, 要求隐写文本与自然文本之间在统计上的差异较小。KL 散度 (Kullback-Leibler Divergence, KLD)^[16]通常用于计算隐写文本与自然文本句子向量之间的统计分布差异, 因此采用 KLD 来评价两者在统计上的关系, 表达式为

$$D_{KL}(P \parallel Q) = \sum_{i=1}^m [p(x_i) \log(p(x_i)) - p(x_i) \log(q(x_i))] \quad (5)$$

式中: P 和 Q 分别为编码前后的概率分布; $p(x_i)$

和 $q(x_i)$ 分别为分布 P 和 Q 中的第 i 个概率值; m 为隐写文本总数。KLD 值越小, 表明隐写文本与自然文本之间在统计上的差异越小。除 KLD 之外, 常用的评价指标还有 JS 散度 (Jensen-Shannon Divergence, JSD)^[12]。

反隐写分析也是衡量隐写方法不可检测性的重要度量, 该方法是将隐写文本和自然文本同时输入到隐写分析算法中, 通过衡量隐写分析算法的准确率、精度、召回率和 F1-score 等来判断隐写方法的不可检测性。针对基于深度学习的生成式文本隐写算法, 主要包括基于语义检测的 Se-Sy^[17]和 V-A3tS^[18]隐写分析方法和基于长短期记忆 (Long Short-Term Memory, LSTM)^[19]网络和 BERT (Bidirectional Encoder Representations from Transformers)^[20]的统计检测隐写分析方法。

2 基于神经网络的生成式文本隐写算法

近年来, 对基于深度学习的生成式文本隐写算法的研究主要采用神经网络模型, 通过改进网络结构和优化编码策略等方法来实现文本隐写的同时, 进一步提高了生成文本的质量、语义和安全性。根据神经网络模型的不同, 目前基于神经网络模型的生成式文本隐写算法主要分为以下 4 类: 基于 LSTM 的生成式文本隐写算法、基于生成对抗网络 (Generative Adversarial Network, GAN) 的生成式文本隐写算法、基于 Transformer 的生成式文本隐写算法和基于序列到序列 (Sequence to Sequence, Seq2Seq) 的生成式文本隐写算法。

2.1 基于 LSTM 网络的生成式文本隐写算法

LSTM 的核心在于其强大的记忆能力。与传统的循环神经网络 (Recurrent Neural Network, RNN) 相比, LSTM 能够学习到长期依赖关系, 擅长处理序列数据, 因此被广泛应用于文本生成领域中。Fang 等^[7]将 LSTM 引入文本隐写算法中, 同时采用分组编码的方式生成隐写文本, 与传统隐写算法相比, 该方法生成文本的质量和隐藏容量均有所提高。但因编码策略简单, 使得隐写文本在质量方面依然表现较差。在此基础上, Yang 等^[9]分别采用固定长度编码 (Fixed Length Code, FLC) 和霍夫曼编码 (Huffman Coding, HC) 生成隐写文本。与文献[7]相比, 编码策略的提升, 使得隐写文本的质量得到明显改善。为提高隐藏

容量, Xiang 等^[21]提出了基于 LSTM 的字符级隐写方法, 将预测下一个单词改为预测下一个字符, 随后根据秘密信息生成多个隐写文本, 最后选取 PPL 最低的隐写文本作为最终生成文本。与文献[9]相比, 该方法运行速度更快, ER 更高。但改用字符级隐写方法使得隐写文本在安全性上表现较差, 在基于统计的检测方法下难以隐蔽信息。为躲避隐写分析, Varol Arisoy 等^[22]在基于 LSTM 字符级隐写术的基础上, 引入了注意力机制, 通过使用注意力机制考虑关键词, 提高了算法的不可感知性和抗隐写分析能力。

2.2 基于 GAN 的生成式文本隐写算法

GAN 可以分为生成网络和判别网络 2 个部分, 生成网络采用文本生成模型, 生成与自然文本近似的隐写文本来欺骗判别网络。判别网络则采用隐写分析模型, 判别传来的隐写文本和自然文本之间的差异。在上述对抗训练下, 生成网络生成的隐写文本会更加接近于自然文本, 从而提高了生成文本的质量。Yang 等^[23]将 GAN 引入文本隐写算法中, 通过将传统的 GAN 与策略更新算法相结合, 对不同类型的文本数据进行对抗训练, 来生成高质量文本。但该算法在编码策略上仅采用 FLC, 当 ER 提高时, 会导致生成隐写文本的不可感知性急剧下降。为解决上述问题, Zhou 等^[24]将自适应概率分布与 GAN 相结合, 基于候选词的概率进行编码, 确保神经网络尽可能选取分布概率较高的候选词, 从而提高隐写文本的不可感知性。与文献[23]相比, 虽然生成隐写文本的 ER 略有下降, 但 PPL、KLD 均有明显提升。为解决误差积累, 进一步提高生成文本的质量, 黄瑶等^[25]将多头注意力机制与 GAN 相结合, 在生成器训练过程中, 通过多头注意力机制提取到多个注意力得分参与奖惩模块的计算, 与文献[23]相比, 该方法的文本生成质量有明显提升。

2.3 基于 Transformer 的生成式文本隐写算法

Transformer 由编码器和解码器组成, 通过注意力机制, 允许模型同时处理整个序列的元素, 从而有效捕捉全局依赖。李晖等^[26]通过使用基于 Transformer 的编码器和解码器, 并采用 FLC、HC 和两者结合的编码策略构建映射关系, 在动态选词的作用下生成隐写文本。与文献[9]相比, 算法模型和编码策略的改进, 使得生成文本的质量得到提高。

BERT 作为一种多层 Transformer 编码器的预

训练语言模型, 能够捕捉文本的双向上下文关系, 在文本隐写领域中应用广泛^[27]。Ueoka 等^[11]将 BERT 引入文本隐写算法中, 通过替换掩码位置实现隐写文本生成。与文献[24]的思想类似, 为解决 ER 提高时隐写文本出现不可感知性和文本质量急剧下降的问题, Zheng 等^[28]将 BERT 与一致性编码相结合, 确保所选候选词的概率与 BERT 语言模型预测概率成正比, 与文献[11]相比, 提高了隐写文本的不可感知性。单一的 BERT 语言模型在选择嵌入位置时, 仅取决于事先确定的嵌入间隔, 未能考虑所选位置进行词替换后带来的结果, 导致隐写文本质量降低。为解决上述问题, Ding 等^[10]将 BERT 加以规则约束, 引导秘密信息嵌入到合适的位置。同时, 利用多头注意力和图注意力网络提取文本序列的时间和空间特征, 确保隐写前后文本特征的一致性, 提高了生成文本的质量。为确保基于 BERT 模型生成文本的安全性, 提高其反隐写分析能力, Xiang 等^[29]通过构建一个基于卷积神经网络 (Convolutional Neural Network, CNN) 作为判别器的因果感知网络来评估模板中原始单词的不可感知性和可替代的候选词, 从而控制秘密信息的嵌入, 使得隐写文本具有更强的抗隐写分析能力和更低的语义失真。

由于 BERT 引入了位置编码, 使该模型对单个文字位置具有较高的敏感性, 因此在特殊体裁的文本中具有广泛应用。杨婉霞等^[30]提出了一种 BERT 词嵌入结合注意力机制的模型, 在格律模板的约束下, 通过生成宋词来隐藏秘密信息。在此基础上, 秦川等^[31]采用了基于 BERT 改进的自回归语言模型 SongNet 作为宋词生成模型, 通过构建平仄韵词牌、词牌模版、韵律信息和关键词信息等约束条件, 生成宋词体裁的隐写文本。与文献[30]相比, 在隐藏容量方面提升巨大。

随着 Transformer 模型的不断发展, GPT (Generative Pre-trained Transformer) 因其强大的自回归生成和单向上下文建模能力, 在语言生成和文本创作任务中表现出色。Ziegler 等^[32]将 GPT-2 引入文本隐写领域, 通过与算术编码 (Arithmetic Coding, AC) 相结合, 开发了文本隐写程序。与文献[9]相比, 语言模型和编码策略的改进, 使得隐写文本的文本质量得到显著提高。但该方法在设计时并没有考虑长文本的生成任务, 导致生成文本上下文缺乏联系。为解决上述问题, 在文献[32]的基础上, Cao 等^[33]将 GPT-2 推广到

长文本生成任务中，通过设计共享密钥确定主题，在秘密信息、阈值和密钥的约束下生成隐写文本，提高了上下文之间的联系。近几年，大型语言模型（Large Language Model, LLM）成为文本生成的研究热点，因其参数规模具有千亿以上，使得生成文本质量更高。Wang 等^[34]首次将 LLM 引入文本隐写领域，在生成过程中采用区间编码，通过区间调整因子和秘密信息决定时间间隔和候选词，从而生成隐写文本。与其他自回归模型相比，LLM 在文本质量、不可感知性和反隐写分析等方面具有优异的性能。

2.4 基于 Seq2Seq 的生成式文本隐写算法

Seq2Seq 由编码器和解码器组成，编码器用于对输入的序列进行表征，解码器用于将编码器提取出的表征建立起到输出序列的映射。Yang 等^[35]使用双向 RNN 作为编码器，普通 RNN 作为解码器，并在训练过程中与强化学习相结合，利用奖惩机制提高隐写文本的质量。该方法与基于 RNN 的算法相比，在文本质量方面表现优异，但并没有考虑上下文之间的联系，导致生成的隐写文本在语义方面表现较差。为解决上述问题，在 Seq2Seq 的基础上，Yang 等^[36]引入门控循环单元（Gated Recurrent Unit, GRU）解决了长期依赖问题。同时，在解码器中加入注意力机制，确保生成符合上下文语义的隐写文本。为充分利用 Seq2Seq 的优势，研究者提出通过语言翻译来隐藏秘密信息。Yang 等^[37]基于 2 个 Seq2Seq 分别实现英语到德语和德语到英语的翻译。在德语到英语的翻译过程中，采用语义感知的方式实现信息编码，在改变表达但不改变语义的情况下嵌入秘密信息，生成隐写文本。在翻译模型的基础上，为进一步提高生成文本上下文之间的联系，Ding 等^[38]在生成模块的编码器中引入全局上下文融合，将全局特征和局部详细特征进行组合，从而获得更细致的上下文信息。为提高生成文本的质量，Lu 等^[39]结合 GAN 的思想，使用判别网络对基于 Seq2Seq 的翻译模型进行对抗性训练。通过语言模型与判别网络的不断对抗，使得隐写文本质量逐步提高。因 Seq2Seq 能够根据输入序列的特点建立输出序列，使其能够运用到特殊体裁文本中。在此基础上，李欣然等^[40]基于对联载体采用多流注意力结构的 Seq2Seq 框架，由主题词和候选信息构成上联部分，逐字生成下联，同时在生成过程中嵌入秘密信息。

基于 Seq2Seq 的思想，变分自动编码器（Variational Auto-Encoder, VAE）也可以看作编码器和解码器的结构。Yang 等^[12]将 VAE 引入文本隐写领域中，分别采用 LSTM 和 BERT 两种模型作为编码器，LSTM 作为解码器。在信息隐藏过程中，分别采用 HC 和 AC 对候选池进行编码，生成隐写文本。与 RNN^[9]和 GPT^[32]相比，生成文本在安全性方面有显著提升。为解决隐写文本在不同主题下出现分布多样化的问题，Tu 等^[41]采用 2 个独立的 VAE 分别进行主题建模和序列建模，通过引入鉴别器和初等反射变换来控制隐写文本生成方向，提高了生成文本的安全性。

表 1 展示了在数据集 IMDB^[42]下，部分语言模型与编码策略结合生成隐写文本的 ER、PPL 和 KLD。

表 1 基于不同神经网络模型和编码策略的隐写文本 ER、PPL 和 KLD 比较

Table 1 Comparison of ER, PPL and KLD based on different neural network model and coding strategy

理论	指标	IMDB ^[42]		
		ER	PPL	KLD
LSTM+分组编码	ER	1.00	2.00	3.00
	PPL	29.73	50.49	87.02
	KLD	11.84	18.56	29.69
RNN+FLC	ER	1.00	2.00	3.00
	PPL	20.17	29.88	50.09
	KLD	27.59	33.20	29.77
RNN+HC	ER	1.00	1.83	2.50
	PPL	20.23	23.68	28.04
	KLD	19.50	18.62	17.55
GPT+AC	ER	0.44	1.35	2.17
	PPL	18.06	21.56	26.52
	KLD	17.96	16.20	15.56
VAE(LSTM-LSTM)+AC	ER	0.42	1.35	2.19
	PPL	39.16	37.81	51.09
	KLD	11.29	10.31	8.48
VAE(BERT-LSTM)+AC	ER	0.40	1.35	2.20
	PPL	28.87	31.57	47.61
	KLD	11.91	9.33	8.20

从表 1 可以看出，GPT 与 AC 结合生成隐写文本的 PPL 最小，但与其他语言模型和编码算法相比，该方法是以牺牲 ER 为代价得到的。整体来看，随着 ER 的增加，PPL 也在逐渐增加。这是因

为随着 ER 的增加,在对候选池进行编码时,更多概率较低的词被编码,在秘密信息的约束下,会出现选择更多概率较低词的情况,从而造成隐写文本 PPL 增加的现象。同时,VAE 模型生成文本的 KLD 值明显低于其他模型,是因为 VAE 能够充分考虑输入文本的表征,从而确保生成文本更符合自然文本。整体来看,采取充分考虑自然文本整体特征的语言模型和编码策略,能够在 KLD 上取得更好的效果。

3 基于约束语义一致性的生成式文本隐写算法

基于深度学习的自然语言处理技术提高了生成式文本隐写算法生成句子的质量,但缺乏控制文本中语义和上下文特征的能力,导致隐写文本的不可感知性变差。为提高上下文之间的联系,约束句子生成,确保语义一致性已逐渐成为该领域研究的热点。目前主要采用以下 3 种方法:基于语义空间约束、基于语言模型和编码策略改进约束以及基于模板约束。

3.1 基于语义空间约束方法

基于语义空间约束是通过构造一个具备上下文语义关系的语义空间,在该空间和秘密信息的约束下进行候选词的选择,从而提高生成隐写文本语义的改进策略。Zhang 等^[43]将语义空间引入文本隐写算法中,约束每个候选词必须从语义空间中确定。与无约束算法相比,虽然与自然文本概率分布上的差异变大,但在语义方面的表现明显提升。为进一步提高语义空间约束方法的普适性和现实应用,在此方法的基础上,刘玉玲等^[44]基于新媒体平台,将秘密信息映射到情感空间以实现语义空间下的秘密信息隐藏,提高了隐写文本的情感表达,具有较好的语义一致性。为进一步拓展语义空间的范围,Xiang 等^[45]通过构造符合语法的句子释义器来指导句子生成和转换,然后利用相同语义文本具有不同语法表达这一特性,生成多个相同语义的不同句子,在秘密信息的约束下选择隐写文本。与传统的基于词预测替换的生成式算法相比,该方法在保证语义一致性的同时,提高了隐藏容量和安全性。

3.2 基于语言模型和编码策略改进约束方法

通过改进语言模型和编码策略,约束候选词生成和候选词对应的编码值,以确保在秘密信息

约束下,每次候选词的选择都尽可能符合语义要求,最终生成语义一致的隐写文本。Li 等^[46]提出了一种基于重写的文本隐写算法,通过对符合语义要求的载体文本引入噪声,在语言模型的作用下,对输入文本进行降噪处理,同时在降噪的词替换过程中嵌入秘密信息。在编码上采取生成多组信息代表同一秘密信息的方式,从而选择最佳隐写文本。上述降噪模型在确保语义一致性上表现较好,但编码策略较为简单,使得在语义上还有进步空间。因此,在上述降噪模型的基础上,Wang 等^[47]在编码上采取动态语义约束的 HC,通过主题感知语义分数和双向语义约束分布得到阈值,随后根据阈值选取频率最高的候选词构成霍夫曼树,依据秘密信息从树的根部到叶子节点逐步选择候选词。在上述编码策略下,虽然 ER 有所下降,但隐写文本的语义得到了显著提高。一般的生成式文本隐写算法在嵌入位置的选择上通常没有约束,导致出现生成影响文本语义的词汇。针对上述问题,Lu 等^[48]通过使用 PPL 和 KL 散度构建组合公式,计算在 HC 下的失真值,并与设定的阈值进行比较,对大于阈值的位置不嵌入秘密信息。在牺牲部分嵌入位置和隐藏容量的前提下,提高了隐写文本的语义。为增强约束型方法的普适性,向凌云等^[49]构造了一个序列到隐写序列的模型,语言编码器用于编码输入原始约束的文本特征,然后隐写器对获得的编码结果进行解码,最后在秘密信息的控制下生成隐写文本。同时,在编码过程中引入注意力优化单元,优化编码过程中注意力分布情况。在解码过程中,采用基于多候选优化的自适应隐写编码策略,根据候选词的概率分布判断是否嵌入秘密信息,随后根据秘密信息选择最优的位置嵌入,从而生成隐写文本。

3.3 基于模版约束方法

基于模版约束的方法通过构建模板,在不改变模版语义的情况下,对其进行变换、提取或回答,在秘密信息的约束下进行信息隐藏。Wang 等^[50]利用对话双方在对话内容上具有相互关联的特点,通过选取尽可能多的包含秘密信息的数据信息作为模板,然后对数据信息中的秘密信息建立索引,对未出现的秘密信息构成关键词集形成控制信息,从而得到隐写文本。虽然生成的隐写文本牺牲了大量 ER,但在语义表现上远远超过无约束的生成式文本隐写算法。与上述方法不同,Zhang 等^[51]采取基于模板文本生成摘要的方法隐藏

秘密信息。通过输入文本作为模板，在动态规划编码算法和秘密信息的约束下，生成符合模板的摘要文本，即隐写文本。

表2展示了在自采网购平台数据集和大众点评数据集下，通过对编码策略进行改进来约束文本生成，从而实现了较好的语义一致性。从表2可以看出，高语义一致性的实现，均伴随着ER的降低。因为在候选词选择过程中，为尽可能地选取最佳选项，在给定编码值时会将概率分布最高的词设定最短的编码值，以确保在秘密信息的约束下提高选择最佳候选词的概率，进而提高隐写文本的语义，这也导致其ER的下降。

表2 基于编码策略改进方法下的语义表现
Table 2 Semantic representation under improved method based on encoding strategy

理论	ER	数据集	KLD	BLEU-3	BLEU-4	ROUGE
NLS ^[32]	3.02	网购平台数据集 ^[47]	60.29	3.27	2.18	10.61
SPLS ^[47]	2.86		52.84	7.14	5.00	21.05
NLS ^[32]	2.89	大众点评数据集 ^[47]	81.24	3.05	1.67	11.25
SPLS ^[47]	2.69		69.59	7.33	5.11	21.20

4 基于安全性提高的生成式文本隐写算法

在生成式文本隐写算法中，隐写文本质量的提高往往会对模型进行修改或对候选池进行截断、删除等操作，导致隐写文本和自然文本的条件概率分布出现统计差异，带来安全风险。为提高生成式文本隐写算法的安全性，主要采取基于候选池改进和基于语言模型改进的方法。

4.1 基于候选池改进的方法

当前秘密信息所要选择的候选词决定了隐写文本的生成。基于上述情况，Zhang等^[52]首次提出基于改进候选池来提高文本安全性的方法。通过在每一步确定候选词时，根据每个词的概率分布，自适应地对其进行分组，最后依据分组动态嵌入秘密信息。与分组编码和HC相比，该方法生成的隐写文本不仅在反隐写分析方面表现优异，在隐藏容量上也有了明显提高。但上述改进候选池的方法并没有考虑截断操作导致算法出现忽略概率较高词的情况。为解决上述问题，梅佳蒙等^[53]在确定候选词时，采用PPL和KL散度对其进行约束，动态进行截断，从而保留满足约束的候选池。

对于分割歧义出现解码错误的问题，通常采用删除候选池中与所选词前缀具有相同词元的词，使得隐写文本的不可感知性变差。为解决上述问题，Nozaki等^[54]在构建候选池的过程中，对于出现具有前缀关系候选词时，总是选择最长的子词，从而避免分割歧义。上述方法虽能避免分割歧义，但是会造成文本的ER下降。为解决上述问题，Ding等^[55]基于生成模型预测的概率分布构建多个副本，通过选择不同副本嵌入秘密信息，提高了生成文本的安全性，同时在ER方面也表现较好。为进一步解决分割歧义带来的安全性问题，在上述算法的思想下，Yan等^[56]通过构建候选图，找出最大权重无关集，从而确定了所选候选词。该方法极大地提高了生成文本的质量，在不可感知性上表现出色，但同时会导致候选池中部分候选词的概率发生变化，使得隐写文本与自然文本在概率分布上出现差异，带来了安全问题。为解决上述问题，Yan等^[57]采用全情况提取方法，通过设计密钥，使得接收方能够从全部可能的秘密信息中提取出唯一通过密钥校验和验证的结果。Qi等^[58]根据前缀关系对候选词进行分组，随后通过设计密钥，确保发送方和接收方从同一分组实现同步采样。针对字符级隐写方法存在的安全性问题，Yan等^[59]利用熵的概率阈值截断隐写过程的候选池，避免因信息含量过高的字符导致出现未知单词和异常表达，提高了隐写文本的安全性。

4.2 基于语言模型改进的方法

语言模型依据候选词的概率、编码策略和秘密信息生成隐写文本，因此语言模型对隐写文本的安全性起着至关重要的作用。Chen等^[60]提出调整语言模型的概率分布，通过设计一种分类器反向传播更新参数，使隐写文本更接近于自然文本。为设计一种能够提高文本安全的普适性方法，Pang等^[61]对语言模型的Softmax函数进行重建，提高了低频词的条件概率，同时抑制了高频词的条件概率，从而避免了文本发生退化。

基于深度学习的生成式文本隐写算法几乎都采用的是自回归模型，因此不可避免地会出现后续生成文本与上文在语义和内容上的不一致，从而带来安全问题。因此，为解决自回归模型逐词预测带来的安全问题，Wang等^[62]将非自回归模型引入文本隐写算法中。首先，利用约束词序列提取自然文本的语义信息，然后将提取的信息作为渐进式非自回归模型的框架，随后通过预测可嵌

入位置来选择候选词。基于上述文本生成过程,不断进行递归操作,直至将所有的秘密信息隐藏到文本中。

表3展示了在IMDB数据集^[42]微调下,并对候选池进行改进后,隐写文本在安全性方面的表现。理论上,安全的隐写算法应保证KLD值为0,以确保敌方不会获得任何在概率分布上的信息。从表3可以看出,在确保安全性的同时,不可避免地出现隐写文本PPL升高的问题。这是因为在确保隐写文本安全性的前提下,无法保证每个候选词的选择均是最佳选择,这也导致了在文本质量方面出现了下降。

表3 基于候选池改进方法下的安全性表现
Table 3 Security performance under improved method based on candidate pool

理论	截断值	PPL	KLD
Discop ^[55]	16	3.27	0.00
	32	3.50	0.00
	64	3.88	0.00
Discop ^[55] +NWIS ^[56]	16	2.64	2.72
	32	2.83	2.94
	64	3.04	3.41
Discop ^[55] +Sincpool ^[58]	16	3.19	0.00
	32	3.74	0.00
	64	3.88	0.00

5 结论

文本隐写技术在信息和网络安全领域中具有十分广阔的应用前景,但同时也面临着诸多挑战。随着DL的迅速发展,基于深度学习的生成式文本隐写算法通过语言模型和隐写编码策略生成隐写文本,提高了生成文本的隐藏容量、质量和安全性,在文本隐写领域中占据主流地位,但仍存在巨大的改进和发展空间。归纳起来,基于深度学习的生成式文本隐写算法存在以下4方面的问题及改进方向。

1) 隐藏容量

基于深度学习的生成式文本隐写算法与基于文本修改的算法、基于文本选择的算法和传统的生成式算法相比,在生成文本的隐藏容量上有所提高,但仍显不足。隐藏容量与文本的冗余空间

相关,在秘密信息嵌入的过程中,要考虑嵌入信息过多是否会造成生成文本语义和安全性降低,从而导致文本在隐藏容量上受到制约。为增加隐藏容量,可以采用以下方法:一是设计语言模型。通过与压缩算法、采样算法和规则机制等相结合,以提高单个嵌入位的隐藏数量,同时也可以通过选用字符级语言模型,将嵌入单元由单词改为字符来增加嵌入位置;二是改进隐写编码策略。隐写编码算法的改进,使得每个词可以携带更多的秘密信息,从而提高隐藏容量。

2) 文本质量

在现有的研究中,生成隐写文本质量不高是绝大多数算法的共性问题。该问题具体表现在过于追求隐藏容量将会导致隐写文本流畅性不够、过于追求安全性将会导致生成文本内部毫无联系等方面,对生成文本的质量具有极大影响。尽管现有研究在提高文本质量方面做了许多工作,但在ER增加的情况下,仍然会出现生成文本质量下降的情况,甚至还会出现带有语法错误的句子。一般来说,语言模型参数的数量与隐写文本质量成正比。因此为提高生成文本质量,可以采用大语言模型作为文本生成模型。通过研究更高效的信息编码算法和模型结构来提高生成文本的质量,是未来提高生成文本质量的重要手段。

3) 文本语义

生成式文本隐写算法在提高文本隐藏容量和自由度的同时,也会带来文本语义难以控制的问题^[47]。因为语言模型是基于概率的统计分布生成文本,使得生成文本的语义、情感和主题是不确定的,导致难以生成语义一致的文本。未来,生成符合语义的文本可以从以下4个方面考虑:一是通过构建语义空间,限制候选词的选择;二是通过改进编码策略,使得最佳候选词与秘密信息比特流对应;三是通过改进注意力机制或提示词方法,限制文本生成方向;四是通过引入模版,约束文本生成内容。

4) 文本安全性

随着隐写分析技术的发展,隐写文本的安全性也同样面临极大考验,主要表现为:基于神经网络的生成式文本隐写算法生成的隐写文本在隐写分析算法的攻击下将会暴露无疑。为确保接收方能够正确提取秘密信息,对模型、候选池进行修改,但同时也会造成隐写文本与自然文本在统计上出现差异,导致文本安全性下降。因此,为

提高文本的安全性, 可以采用以下4种方法: 一是对候选池进行约束, 减少截断和删除等操作; 二是改进语言模型, 约束嵌入位置选择或候选池的生成, 从而得到更加接近于自然文本的隐写文本; 三是引进对抗机制, 不断优化生成模块; 四是改进编码算法, 约束候选词的编码值, 以减少隐写失真。

综上所述, 生成式文本隐写技术相比于其他文本隐写方法, 极大地提高了生成隐写文本的隐藏容量、不可感知性和不可检测性。随着神经网络和NLP技术的不断发展, 基于深度学习的生成式文本隐写算法将会被广泛应用于信息安全领域中。本文对基于深度学习的生成式文本隐写算法进行了综述, 全面介绍了算法的框架、评价指标、分类方式以及相关典型算法的原理和性能分析。通过汇编现有方法, 为未来基于深度学习的生成式文本隐写方法指明了研究方向。

参考文献

- [1] LOW S H, MAXEMCHUK N F, BRASSIL J T, et al. Document marking and identification using both line and word shifting[C]//Proceedings of the International Conference on Computer Communications. Piscataway, NJ, USA: IEEE, 1995: 853-860.
- [2] 赵敏之, 孙星明, 向华政. 基于虚词变换的自然语言信息隐藏算法研究[J]. 计算机工程与应用, 2006(3): 158-160.
ZHAO M Z, SUN X M, XIANG H Z. Research on the Chinese text steganography based on the modification of the empty word[J]. Computer Engineering and Applications, 2006(3): 158-160.
- [3] WU W, XIANG L Y, WANG W, et al. A synonym substitution-based steganography using LZW compression algorithm[J]. Revista de la Facultad de Ingeniería, 2017, 32(1): 763-770.
- [4] ZHU P, XIANG G L, SONG W N, et al. A text zero-watermarking algorithm based on Chinese phonetic alphabets[J]. Wuhan University Journal of Natural Sciences, 2016, 21(4): 277-282.
- [5] XIANG L Y, GUO G Q, YU J M, et al. A convolutional neural network-based linguistic steganalysis for synonym substitution steganography[J]. Mathematical Biosciences and Engineering, 2020, 17(2): 1041-1058.
- [6] CHEN X Y, SUN H Y, TOBE Y, et al. Coverless information hiding method based on the Chinese mathematical expression[M]//HUANG Z Q, SUN X M, LUO J Z, et al. Cloud Computing and Security. Cham: Springer International Publishing, 2015: 133-143.
- [7] FANG T N, JAGGI M, ARGYRAKI K. Generating steganographic text with LSTMs[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics-Student Research Workshop. Stroudsburg, PA, USA: ACL, 2017: 100-106.
- [8] SIMMONS G J. The Prisoners' problem and the subliminal channel[M]//CHAUM D. Advances in Cryptology. Boston, MA: Springer, 1984: 51-67.
- [9] YANG Z L, GUO X Q, CHEN Z M, et al. RNN-stega: linguistic steganography based on recurrent neural networks[J]. IEEE Transactions on Information Forensics and Security, 2019, 14(5): 1280-1295.
- [10] DING C H, FU Z J, YU Q, et al. Joint linguistic steganography with BERT masked language model and graph attention network[J]. IEEE Transactions on Cognitive and Developmental Systems, 2024, 16(2): 772-781.
- [11] UEOKA H, MURAWAKI Y, KUROHASHI S. Frustratingly easy edit-based linguistic steganography with a masked language model[C]//Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Stroudsburg, PA, USA: ACL, 2021: 5486-5492.
- [12] YANG Z L, ZHANG S Y, HU Y T, et al. VAE-stega: linguistic steganography based on variational auto-encoder[J]. IEEE Transactions on Information Forensics and Security, 2021, 16: 880-895.
- [13] PAPINENI K, ROUKOS S, WARD T, et al. BLEU: a method for automatic evaluation of machine translation[C]//Proceedings of the 40th Annual Meeting on Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2001: 311-318.
- [14] DENKOWSKI M, LAVIE A. Meteor universal: language specific translation evaluation for any target language[C]//Proceedings of the 9th Workshop on Statistical Machine Translation. Stroudsburg, PA, USA: ACL, 2014: 376-380.
- [15] LIN C Y. ROUGE: a package for automatic evaluation of summaries[C]//Proceedings of 2024 Workshop on Text Summarization Branches Out. Stroud-

- sburg, PA, USA: ACL, 2004: 74-81.
- [16] KULLBACK S, LEIBLER R A. On information and sufficiency[J]. The Annals of Mathematical Statistics, 1951, 22(1): 79-86.
- [17] YANG J S, YANG Z L, ZHANG S Y, et al. SeSy: linguistic steganalysis framework integrating semantic and syntactic features[J]. IEEE Signal Processing Letters, 2022, 29: 31-35.
- [18] WANG Y H, ZHANG R, LIU J Y. V-A3tS: a rapid text steganalysis method based on position information and variable parameter multi-head self-attention controlled by length[J]. Journal of Information Security and Applications, 2023, 75: 103512.
- [19] YANG H, BAO Y J, YANG Z L, et al. Linguistic steganalysis via densely connected LSTM with feature pyramid[C]//Proceedings of the 2020 ACM Workshop on Information Hiding and Multimedia Security. New York, USA: ACM, 2020: 5-10.
- [20] PENG W L, LI S, QIAN Z X, et al. Text steganalysis based on hierarchical supervised learning and dual attention mechanism[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023, 31: 3513-3526.
- [21] XIANG L Y, YANG S H, LIU Y H, et al. Novel linguistic steganography based on character-level text generation[J]. Mathematics, 2020, 8(9): 1558.
- [22] VAROL ARISOY M. LZW-CIE: a high-capacity linguistic steganography based on LZW char index encoding[J]. Neural Computing and Applications, 2022, 34 (21): 19117-19145.
- [23] YANG Z L, WEI N, LIU Q H, et al. GAN-TStega: text steganography based on generative adversarial networks[C]//Digital Forensics and Watermarking. Cham: Springer International Publishing, 2020: 18-31.
- [24] ZHOU X J, PENG W L, YANG B Y, et al. Linguistic steganography based on adaptive probability distribution[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 19 (5): 2982-2997.
- [25] 黄瑶, 潘丽丽, 熊思宇, 等. 基于生成对抗网络与多头注意力的文本隐写术[J]. 计算机工程与科学, 2023, 45(10): 1789-1796.
- HUANG Y, PAN L L, XIONG S Y, et al. Text steganography based on generative adversarial networks and multi-head attention[J]. Computer Engineering & Science, 2023, 45(10): 1789-1796.
- [26] 李晖, 金家立, 金纤羽, 等. 基于自动选择编码及动态选词策略的文本隐写方法[J]. 通信学报, 2022, 43(9): 240-253.
- LI H, JIN J L, JIN S Y, et al. Text steganography method based on automatic selection coding and dynamic word selection strategy[J]. Journal on Communications, 2022, 43(9): 240-253.
- [27] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[C]//2019 Conference of the North American Chapter of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2019: 4171-4186.
- [28] ZHENG X Y, WU H Z. Autoregressive linguistic steganography based on BERT and consistency coding[J]. Security and Communication Networks, 2022, 2022: 9092785.
- [29] XIANG L Y, XIA J L, LIU Y F, et al. CPG-LS: causal perception guided linguistic steganography[J]. IEEE Signal Processing Letters, 2023, 30: 1762-1766.
- [30] 杨婉霞, 陈帅, 管磊, 等. 宋词自动生成的信息隐藏算法[J]. 信息网络安全, 2022, 22 (7): 46-54.
- YANG W X, CHEN S, GUAN L, et al. Information steganography algorithm for automatic generation of Song Ci[J]. Netinfo Security, 2022, 22 (7): 46-54.
- [31] 秦川, 李蓉受, 钱振兴, 等. 基于宋词生成的大容量构造式信息隐藏算法[J]. 计算机学报, 2023, 46 (1): 17-30.
- QIN C, LI R S, QIAN Z X, et al. Large-Capacity constructive information hiding based on Song Ci generation[J]. Chinese Journal of Computers, 2023, 46(1): 17-30.
- [32] ZIEGLER Z M, DENG Y T, RUSH A M. Neural linguistic steganography[EB/OL]. (2019-09-03)[2024-11-05]. <https://arxiv.org/abs/1909.01496>.
- [33] CAO Y, ZHOU Z L, CHAKRABORTY C, et al. Generative steganography based on long readable text generation[J]. IEEE Transactions on Computational Social Systems, 2024, 11(4): 4584-4594.
- [34] WANG Y H, SONG R Q, ZHANG R, et al. LLsM: generative linguistic steganography with large language model[EB/OL]. (2024-08-08) [2024-11-05]. <https://arxiv.org/pdf/2401.15656v3>.
- [35] YANG Z L, ZHANG P Y, JIANG M Y, et al. RITS: real-time interactive text steganography based on automatic dialogue model[C]//International Con-

- ference on Cloud Computing and Security. Cham: Springer International Publishing, 2018: 253-264.
- [36] YANG Z, XU Z H, ZHANG R, et al. T-GRU: conTextual gated recurrent unit model for high quality linguistic steganography[C]//2022 IEEE International Workshop on Information Forensics and Security (WIFS). Piscataway, NJ, USA: IEEE, 2022: 1-6.
- [37] YANG T Y, WU H Z, YI B, et al. Semantic-preserving linguistic steganography by pivot translation and semantic-aware bins coding[J]. IEEE Transactions on Dependable and Secure Computing, 2024, 21(1): 139-152.
- [38] DING C H, FU Z J, YANG Z L, et al. Context-aware linguistic steganography model based on neural machine translation[J]. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023, 32: 868-878.
- [39] LU T H, LIU G S, ZHANG R, et al. Robust secret data hiding for transformer-based neural machine translation[C]//2023 International Joint Conference on Neural Networks (IJCNN). Piscataway, NJ, USA: IEEE, 2023: 1-8.
- [40] 李欣然, 李蓉受, 秦川, 等. 基于对联载体的生成式信息隐藏方法[J/OL]. 计算机研究与发展, 2024: 1-11. (2024-08-05) [2024-11-05]. <http://kns.cnki.net/kcms/detail/11.1777.TP.20240805.1334.002.html>.
- LI X R, LI R S, QIN C, et al. Generative information hiding method based on couplet carrier[J/OL]. Journal of Computer Research and Development, 2024: 1-11. (2024-08-05) [2024-11-05]. <http://kns.cnki.net/kcms/detail/11.1777.TP.20240805.1334.002.html>.
- [41] TU H Q, YANG Z L, YANG J S, et al. FET-LM: flow-enhanced variational autoencoder for topic-guided language modeling[J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35(8): 11180-11193.
- [42] MAAS A L, DALY R E, PHAM P T, et al. Learning word vectors for sentiment analysis[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Stroudsburg, PA, USA: ACL, 2011: 142-150.
- [43] ZHANG S Y, YANG Z L, YANG J S, et al. Linguistic steganography: from symbolic space to semantic space[J]. IEEE Signal Processing Letters, 2021, 28: 11-15.
- [44] 刘玉玲, 王翠林, 付章杰. 语义空间下基于情感表达的生成式文本隐写方法[J]. 通信学报, 2023, 44(4): 176-186.
- LIU Y L, WANG C L, FU Z J. Generative text steganography method based on emotional expression in semantic space[J]. Journal on Communications, 2023, 44(4): 176-186.
- [45] XIANG L Y, OU C F, ZENG D J. Linguistic steganography: hiding information in syntax space[J]. IEEE Signal Processing Letters, 2023, 31: 261-265.
- [46] LI F X, WU S X, YU J, et al. Rewriting-stego: generating natural and controllable steganographic text with pre-trained language model[C]//International Conference on Database Systems for Advanced Applications. Cham: Springer, 2023: 617-626.
- [47] WANG S X, LI F X, YU J, et al. Enhancing semantic consistency in linguistic steganography via denosing auto-encoder and semantic-constrained Huffman coding[C]//Natural Language Processing and Chinese Computing. Cham: Springer, 2023: 799-812.
- [48] LU T H, LIU G S, ZHANG R, et al. Neural linguistic steganography with controllable security[C]//2023 International Joint Conference on Neural Networks (IJCNN). Piscataway, NJ, USA: IEEE, 2023: 1-8.
- [49] 向凌云, 杨双辉, 王蓉, 等. 基于序列到隐写序列的约束型自然语言信息隐藏方法[J]. 计算机学报, 2023, 46(8): 1650-1669.
- XIANG L Y, YANG S H, WANG R, et al. Constrained linguistic steganography based on sequence to steganographic sequence model[J]. Chinese Journal of Computers, 2023, 46(8): 1650-1669.
- [50] WANG H L, YANG Z L, YANG J S, et al. Hi-Stega: a hierarchical linguistic steganography framework combining retrieval and generation[C]//Neural Information Processing. Singapore: Springer, 2024: 41-54.
- [51] ZHANG R F, LIU J Y, ZHANG R. Controllable semantic linguistic steganography via summarization generation[C]//2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ, USA: IEEE, 2024: 4560-4564.
- [52] ZHANG S Y, YANG Z L, YANG J S, et al. Provably secure generative linguistic steganography[EB/OL]. (2021-06-03) [2024-11-05]. <http://arxiv>.

- org/abs/2106.02011.
- [53] 梅佳蒙, 任延珍, 王丽娜. 安全性可控的生成式文本隐写算法[J]. 网络与信息安全学报, 2022, 8(3): 53-65.
- MEI J M, REN Y Z, WANG L N. Generation-based linguistic steganography with controllable security[J]. Chinese Journal of Network and Information Security, 2022, 8(3): 53-65.
- [54] NOZAKI J, MURAWAKI Y. Addressing segmentation ambiguity in neural linguistic steganography [EB/OL]. (2022-11-12) [2024-11-05]. <http://arxiv.org/abs/2211.06662>.
- [55] DING J Y, CHEN K J, WANG Y F, et al. Discop: provably secure steganography in practice based on distribution copies[C]//2023 IEEE Symposium on Security and Privacy (SP). Piscataway, NJ, USA: IEEE, 2023: 2238-2255.
- [56] YAN R Y, YANG Y T, SONG T. A secure and disambiguating approach for generative linguistic steganography[J]. IEEE Signal Processing Letters, 2023, 30: 1047-1051.
- [57] YAN R Y, SONG T J, YANG Y T. SegFree: segmentation-free generative linguistic steganographic approach for unsegmented languages[EB/OL]. (2023-09-26) [2024-11-05]. <https://doi.org/10.36227/techrxiv.24199068.v1>.
- [58] QI Y, CHEN K J, ZENG K, et al. Provably secure disambiguating neural linguistic steganography [EB/OL]. (2020-03-26) [2024-11-05]. <http://arxiv.org/abs/2403.17524>.
- [59] YAN R Y, SONG T J, YANG Y T. Token-Free: a tokenization-free generative linguistic steganographic approach with enhanced imperceptibility [C]//2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Piscataway, NJ, USA: IEEE, 2023: 449-455.
- [60] CHEN C, YANG J S, GAO Y, et al. Minimizing distortion in steganography via adaptive language model tuning[C]//International Conference on Neural Information Processing. Singapore: Springer, 2024: 571-584.
- [61] PANG K Y, BAI M H, YANG J S, et al. FREmax: a simple method towards truly secure generative linguistic steganography[C]//2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ, USA: IEEE, 2024: 4755-4759.
- [62] WANG R, XIANG L Y, LIU Y F, et al. PNG-Stega: progressive non-autoregressive generative linguistic steganography[J]. IEEE Signal Processing Letters, 2023, 30: 528-532.
- (编辑: 刘建民)