计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.210100015

基于神经网络机器翻译的自然语言信息隐藏

周小诗　张梓葳　文　娟
中国农业大学信息与电气工程学院　北京 100083
(zhouxiaoshi0713@163.com)

摘　要　生成式自然语言信息隐藏在自然语言生成过程中嵌入秘密信息。目前主流的生成式自然语言隐藏方法采用一个简单的循环神经网络(Recurrent Neural Networks,RNN)或长短时记忆网络(Long Short-Term Memory,LSTM)进行载密文本的生成。这种方法生成的载密文本长度有限,且句子和句子之间没有语义关联。为了解决这个问题,提出了能够生成长句且句与句之间能保持语义关系的机器翻译隐写算法 Seq2Seq-Stega。采用序列到序列(Sequence to Sequence,Seq2Seq)模型作为文本隐写的编码器和解码器,源语句的信息可以保证目标载密句的语义关联性。此外,根据每一时刻模型计算的单词概率分布,设计了候选池的选词策略,并引入了平衡源语句与目标句的贡献度的注意力超参数。通过实验比较了不同选词阈值和注意力参数下模型的隐藏容量和生成文本的质量。与其他3种生成式模型的对比实验表明,该算法能够保持长距离语义关联,并具有较好的抗隐写分析能力。

关键词:自然语言信息隐藏;文本生成;语义距离;自注意力机制;机器翻译
中图法分类号　G316

Natural Language Steganography Based on Neural Machine Translation

ZHOU Xiao-shi,ZHANG Zi-wei and WEN Juan
College of Information and Electrical Engineering,China Agricultural University,Beijing 100083,China

Abstract　Generation-based natural language steganography embeds secret information during text generation under the guidance of secret bitstream. The current generation-based steganographic methods are based on recurrent neural networks (RNN) or long short-term memory (LSTM),which can only generate short stego text because the semantic quality becomes worse as the length of the sentence increases. Moreover,there is hardly any semantic connection between sentences. To address this issue,this paper proposes a neural machine translation steganography algorithm,namely Seq2Seq-Stega,that can generate long text in which semantic relationship maintains well between words and sentences. An encoder-decoder model based on sequence-to-sequence (Seq2Seq) structure is used as our translation model. The source sentence can offer extra information and ensure the semantic relevance between the target stego sentences. In addition,according to the word probability distribution obtained by the model,we design a word selection strategy to form the candidate pool. An attention hyperparameter is introduced to balance the contribution of the source sentence and the target sentence. Experimental results show the hidden capacity and the text quality under different word selection thresholds and attention parameters. Comparative experiments with other three generation-based models show that Seq2Seq-Stega can maintain long-distance semantic connections and better resist steganalysis.

Keywords　Natural language steganography,Text generation,Semantic relevance,Self-attention mechanism,Machine translation

1　引言

信息隐藏或隐写,指的是将秘密信息隐藏于视频[1-3]、图片[4-7]、音频[8-10]、文本[11-14]等可公开的媒体信息中,使人们凭直观的视觉和听觉难以察觉其存在的技术。在秘密通信过程中,将需要隐写的信息进行加密,然后通过载体(covertext)嵌入秘密信息,最终得到一个可以通过信道传递的“隐秘文本”(stegotext)。

最初的文本信息隐藏大多基于修改式的方法,如改变字间距[15]、字大小、字体等方法,这类方式具有形式多样、编码简单、存储方便、传输速度快等优点,成为当时文本隐藏中应用最广泛的一种方式。虽然基于修改式的文本隐写具有较大的隐藏容量,但其载密文本容易受到隐写分析的攻击。随着自然语言信息隐藏成为文本信息隐藏领域的研究热点,主流的文本信息隐藏融入了自然语言处理技术,从格式修改变为更加侧重句子语法或语义的方法,从嵌入方式可以分为修改式的自然语言信息隐藏、生成式的自然语言信息隐藏和基于载体获取式的信息隐藏三大类。早期的修改式的自然语言信息隐藏主要以同义词替换[16-17]、拼写错误引入[18]、句法变换[19],以及语义运算[20]等方式嵌入秘密信息。然而这些方

基金项目:国家自然科学基金(61802410);中国高校科学基金(2019TC047)
This work was supported by the National Natural Science Foundation of China (61802410) and Chinese Universities Scientific Fund (2019TC047).
通信作者:文娟(wenjuan@cau.edu.cn)

法依赖于复杂的句法或语义分析，很难达到较高的准确率，攻击者可以通过对比发现修改的位置从而进行攻击，因此安全性较低。此外，由于文本冗余少，这些方法的嵌入容量较小，因为即使是微小的文本变化也可能产生语义异常或语法错误。

针对这些问题，研究者们提出了不修改文本的隐写方案，在秘密信息的引导下获取或生成载密文本。基于载体获取式的方法旨在找到与秘密信息匹配的一系列文本，如用映射函数[21]从一个大语料库中选择若干个文本作为载密文本。生成式自然语言隐藏根据特定的统计规律或语言模型[22-23]自动生成隐写文本。早期的生成式文本隐写大都基于语法规则，比如基于上下文无关的语法或句子模板。TEXTO 是最早的生成式文本隐写方法，它设计了许多由词组构成的句子模板。在此基础上，根据语句的词性特征，对这些语句进行填充，生成载密文本。其他早期的研究，如 NICETEXT[24] 和 Mimicry[25]，通过语法规则生成载密文本。这类方法与修改式文本隐写的不同之处在于，它们无需原始文本，因此攻击者很难进行对比检测。这种方法虽然嵌入率很高，但由于没有考虑到语义信息，载密文本的上下文语义毫无关联，安全性较差。

为解决语义不相关的问题，一些采用统计语言模型的生成式隐写模型出现了，如使用 n-gram 模型或马尔可夫链[26]来建模语义特征。由于语义建模比较困难，一些统计模型被用于如短笑话[27]、电子邮件[28]、诗歌[29]等特定的体裁。为提高文本的流畅性，Safaka 等使用 n-gram 模型生成了一些候选载密文本，然后对这些载密文本进行人工编辑和润色。这种方式计算每个时刻单词的条件概率 $p(x_i|x_1,x_2,\cdots,x_{i-1})$，其中 x_i 是第 i 个单词。针对数据稀疏以及参数过大的问题，n-gram 模型和马尔可夫模型均引入马尔可夫假设，即认为每个单词只与前面的几个单词相关，即 $p(x_i|x_1,x_2,\cdots,x_{i-1})\approx p(x_i|x_{i-n+1},x_{i-n+2},\cdots,x_{i-1})$。然而，由马尔可夫假设可知，随着词语距离的增加，词之间的语义关联度会减小，导致整篇文章的句子之间的语义关联较弱。

近年来，神经网络为建模文本语义长距依赖性提供了新的解决方案。Luo 等[29]将 RNN 编码-解码器(Encoder-Decoder)与语法模板相结合，生成载密汉语诗歌，提高了生成的载密文本间的长距离语义相关性。Fang 等[21]采用长短期记忆网络(Long Short-Term Memory，LSTM)，从特定的词库中选择编码与秘密信息匹配的载密词。Yang 等[30]提出了一种基于循环神经网络(Recurrent Neural Network，RNN)的信息隐藏算法 RNN-Stega，在嵌入容量和文本质量方面都达到了最先进的性能。该算法首先根据条件概率分布对候选词进行编码，然后选择与当前秘密信息比特流相同的词来嵌入。由于传统模型的长期记忆力远不及 RNN 和 LSTM，基于深度学习的模型逐渐取代了传统的文本隐写模型。然而，RNN 记忆单元也存在一定限制，随着产生的词汇越来越多，之前的语义信息最终会被忽略，从而导致整个文本的语义一致性变弱。因此，在生成长句或多句的载密篇章时如何保持语义相关性，仍然是自然语言信息隐藏的一个挑战。

本文将神经网络机器翻译(Neural Machine Translation，NMT)模型与隐写模型相结合，提出了 Seq2Seq-Stega 模型。Seq2Seq-Stega 根据源文本 x 和之前生成的目标词来生成目标语言的载密文本 y。在生成目标词时，源文本为句子之间的语义相关性提供了有用的信息。例如，生成单词 y_{n+1} 时，Seq2Seq-Stega 不仅考虑了之前生成的单词 $y_1,y_2,\cdots,y_n$，还考虑了对应的源句 $x_1,x_2,\cdots,x_m$。此外，该模型还使用了一个注意力参数来平衡源语句和目标句双方信息对目标词的影响。为了进一步提高隐写文本的质量，本文还引入了一种选词策略来构建候选池。实验表明，Seq2Seq-Stega 具有生成多个语义相关长句的能力。此外，Seq2Seq-Stega 在抗隐写分析实验中也表现出色。

本文的创新主要有 3 个方面。

(1)Seq2Seq-Stega 解决了语义关联性随着生成句子长度的增加而降低的问题。采用基于注意力机制的 Encoder-Decoder 架构，在文本生成过程中动态嵌入秘密信息。由于生成的文本需要与较远的词保持语义联系，每个目标词的选择都会参考源文本和已经生成的目标文本两部分。

(2)为了缓解嵌入秘密信息而造成的文本质量下降问题，设计了一种基于概率方差的动态选词策略，通过方差阈值来自适应地构建候选池，动态调节每一时刻的嵌入容量。

(3)引入了注意力超参数，研究了不同注意力参数和方差阈值对文本的嵌入容量和文本质量的影响。

2 相关工作

2.1 基于机器翻译的文本信息隐藏

同一个源语句采用不同的翻译器翻译，将会得到不同的翻译句，Grothoff 等利用这个特点首先提出了一种机器翻译信息隐藏(Translation-based Steganography，TBS)模型 LiT(Lost in Translation)[31]。这种隐藏方式采用多个翻译器翻译同一个源句，得到多个不同的翻译句。先对每个翻译器进行哈夫曼编码，再选择编码与秘密信息比特相对应的翻译器生成的翻译句作为最终的载密句。之后，Grothoff 等对 LiT 的编码方式进行了改进，提出了 LiJtT(Lost in just the translation)[31-32]。与 LiT 的编码方式不同的是，LiJtT 是直接对生成的翻译句编码，每个翻译句都根据一个密钥转换成一个哈希值，然后选择最低有效位与当前秘密信息比特匹配的句子作为载密句。

无论是 LiT 还是 LiJtT，都需要多个翻译器参与。模型的嵌入容量取决于得到的翻译句的多样性。如果某一时刻没有任何翻译句的 LSB 与秘密信息比特匹配，则 LiJtT 无法实现嵌密，从而导致嵌入失败。为了解决这个问题，Meng 等提出了 LinL(Lost in n-best List)[33]。它不再使用多个翻译器，而是使用一个统计机器翻译模型来获得 n-best 个翻译句。例如，当 $n=1$ 时，给定一个源语句，翻译句 $\hat{t}$ 是使得条件概率 $p(t|s)$ 最大的一句，即：

$$\hat{t}=\arg\max_t p(t|s) \tag{1}$$

由贝叶斯公式 $p(t|s)=\frac{p(t)p(s|t)}{p(s)}$ 可得：

$$\hat{t}=\arg\max_t p(t)p(s|t) \tag{2}$$

其中，$p(t)$ 是目标文本的语言模型，$p(s|t)$ 是用双语料库训练的翻译模型。

LinL 使用 n-best 搜索算法选出 n-best 翻译句，并对这些句子进行编码，以便进行信息嵌入。与 LiT 和 LiJtT 相比，

LinL 具有更好的抗失真性能和更高的嵌入容量。

现有的 TBS 模型都是基于统计机器翻译(StatiStical Machine Translation，SMT)模型的。SMT 的一个问题是采用马尔可夫假设来计算语言模型，即假设下一个单词的生成只取决于其前面的几个单词，因此得到的翻译文本质量比较差。

2.2 基于神经网络的机器翻译

最近，NMT 在英-德、英-法等多个机器翻译任务中取得了优异的性能[34-35]。与 SMT 不同的是，NMT 是由 Encoder-Decoder 组成。Encoder 将源句编码成固定长度的向量，然后将其送到 Decoder 端以生成目标语言的翻译[23]，如图 1 所示。

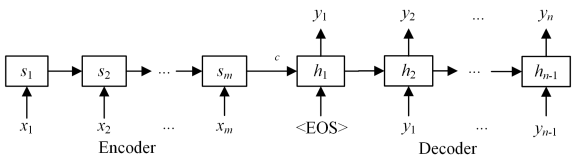


图 1 神经网络机器翻译模型

Fig. 1 Neural machine translation model

假设由 $x=(x_1,\cdots,x_m)$ 表示 Encoder 当前输入的源语句，m 是 x 中的单词数；$y=(y_1,\cdots,y_n)$ 是 Decoder 最终输出的目标句，n 是 y 中包含的单词数。在 t 时刻，NMT 模型通过计算条件概率来预测该时刻的目标词 y_t，即：

$$p(y_t\mid y_1,y_2,\cdots,y_{t-1},x)=\mathrm{softmax}(g(h_t)) \tag{3}$$

其中，h_t 是根据式(4)计算出的递归隐层状态，g 是将 h 转换为词向量的转换函数，词向量的维度等于词典大小。

$$\begin{aligned} h_t&=f(h_{t-1},y_{t-1}),t\geqslant 2\\ h_1&=f(c,y_0),t=1 \end{aligned} \tag{4}$$

其中，c 是从 Encoder 得到的源语句表示，f 是非线性函数，可以是 RNN，LSTM，GRU 或 Transformer。每个句子都有一个句子开始标志〈SOS〉和结束标志〈EOS〉，令 $y_0=$〈SOS〉。

由于 NMT 在各项翻译任务上的表现都超过了 SMT，因此引起了人们的广泛关注。而注意力机制可以使模型在生成目标词时特别注意源语句中的某些特定词，显著提高了生成的翻译文本质量[36]。

3 模型原理与架构

为了得到具有句间语义相关性的高质量隐写文本，本文提出了一种基于 NMT 的隐写模型 Seq2Seq-Stega。为了能够充分利用源文本中的有用信息，在生成过程中采用了注意力机制。传统的基于 SMT 的 TBS 是根据源语句生成多个翻译句，然后对每个句子进行编码，最后选择与秘密信息对应的句子作为最终的目标句。与之不同的是，Seq2Seq-Stega 是在目标句生成过程中根据秘密信息动态选择生成词，从而达到嵌入秘密信息的目的。

由于所选单词并不总是概率最大的单词，在嵌入过程中或多或少会导致生成的文本质量下降。本文的主要工作就是在信息嵌入过程中保持句间语义，同时尽可能减小因选词导致的文本质量下降问题。

3.1 模型框架

为了提高翻译准确率，本文采用基于注意力机制的 NMT 模型作为翻译模型。将注意力机制用在 Encoder-Decoder 结构中，可以使得模型动态地关注输入的某个特定部分，从而更有效地完成自然语言处理任务。例如在机器翻译中，注意力机制可以在预测输出值 y_t 时，找到与其关联程度高的源语言单词。

基于注意力机制的神经网络机器翻译模型的框架如图 2、图 3 所示。下面将分别介绍模型的解码器和编码器的工作原理及其作用。

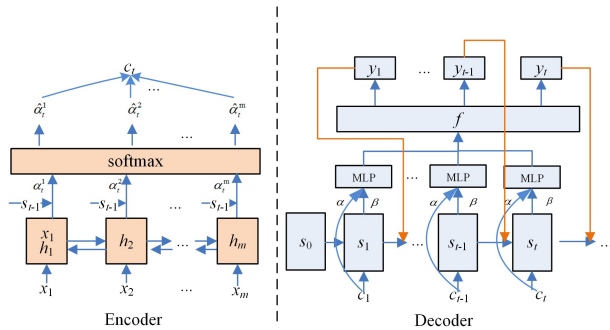


图 2 Seq2Seq-Stega 模型架构图

Fig. 2 Framework of Seq2Seq-Stega model

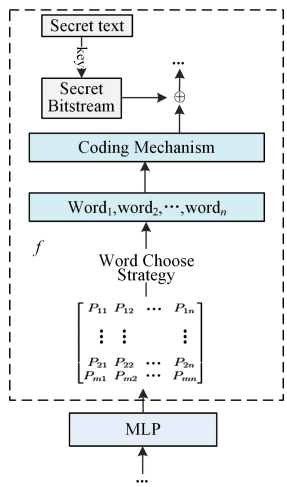


图 3 f 展开图

Fig. 3 Expanded diagram of f

假设编码器的当前输入句表示为 $\vec{x}=(x_1,x_2,\cdots,x_m)$，其中 m 为 $\vec{x}$ 中单词的个数。隐藏层 h 可以采用 RNN、LSTM 或门控循环单元(Gate Recurrent Unit，GRU)中的任意一种形式。在本文的实验中，使用双向 LSTM(Bi-directional LSTM，Bi-LSTM)作为隐藏单元。Bi-LSTM 能更好地捕获句子中上下文的信息，每个单词的隐藏状态不仅压缩了当前单词之前的信息，还压缩了当前词之后的信息。即每个单词 x_t 都可以表示为 h_t，这是前向隐藏状态 $\overrightarrow{h}_t$ 和后向隐藏状态 $\overleftarrow{h}_t$ 的融合。

$$\overrightarrow{h}_t=f_{\mathrm{LSTM}}(x_t,\overrightarrow{h}_{t-1}) \tag{5}$$

$$\overleftarrow{h}_t=f_{\mathrm{LSTM}}(x_t,\overleftarrow{h}_{t+1}) \tag{6}$$

$$h_t=[\overrightarrow{h}_t^{\mathrm{T}};\overleftarrow{h}_t^{\mathrm{T}}]^{\mathrm{T}} \tag{7}$$

解码器的最终输出是 $y=(y_1,y_2,\cdots,y_n)$，其中 n 是目标句 y 包含的单词数。预测当前词 y_t 的概率分布可以表示为：

$$p(y_t\mid y_1,y_2,\cdots,y_{t-1},\vec{x})=MLP(y_{t-1},s_t,c_t) \tag{8}$$

其中，MLP 是一个多层感知器组件，s_t 是隐藏层的输出状态，即：

$$s_t=f_{\mathrm{LSTM}}(s_{t-1},y_{t-1},c_t) \tag{9}$$

c_t 是除了 h_t 外的又一隐藏状态，其计算公式如下：

$$c_t=\sum_{i=1}^{m}\hat{\alpha}_t^i h_i \tag{10}$$

其中，$\hat{\alpha}_t^i$ 是注意力权重，公式如下：

$$\hat{\alpha}_t^i=\mathrm{softmax}(\alpha_t^i) \tag{11}$$

α_t^i 是当前词 y_t 对源语句中单词 x_i 的参照权重，即：

$$\alpha_t^i=a(s_{t-1},h_t) \tag{12}$$

组件 a 是前馈神经网络，它与 NMT 模型的另一部分共同训练。

3.2 动态选词策略

将选词策略作用于上一步得到的概率分布就可以生成候选池，选词策略的优劣直接决定了候选池中元素的质量好坏。本次实验通过对概率分布的方差设限来决定一个词能否进入候选池。

首先需要从得到的概率分布中选取概率值 top8 的词，排名第一的概率值称做 p_1，对应了在不嵌密的情况下 t 时刻生成的目标词 $\mathrm{target\_word}_1$。分别统计 top8 的词中，前 2，4，8 个的词的概率方差值 $\mathrm{var}(n)(n=2,4,8)$，即：

$$\mathrm{var}(n)=\frac{\left[p_1-\left(\frac{p_1+p_2+\cdots+p_n}{n}\right)\right]^2+\cdots+\left[p_n-\left(\frac{p_1+p_2+\cdots+p_n}{n}\right)\right]^2}{n}<\varepsilon \tag{13}$$

其中，p_i 表示 t 时刻依概率排序的第 i 个单词 w_i 的概率值。ε 为方差阈值。如果概率最大的前 8 个词的概率方差值满足阈值条件，即 var(8)小于 ε 时，就把这 8 词都添加到候选池中。否则，则减小 n 为 4，考查 var(4)是否小于 ε。满足条件则把前 4 个词加入候选词，不满足条件则减小 n 为 2，继续判断 var(2)是否小于 ε。满足则当前候选池的词为 2，不满足则当前位置不进行嵌入，直接输出最大概率的词。ε 取不同的值将会产生不同的候选池。ε 的取值对模型的影响将在实验部分详细介绍。

本文的选词策略会使得每一时刻的候选池中的词数可能不相同，即每一时刻的嵌入容量会根据当前的条件概率分布自动调节，是一种自适应的秘密信息嵌入策略。

3.3 注意力参数的调整

在本文模型中，t 时刻的目标词的概率分布取决于源文本和已经生成的目标文本两部分，其计算公式如下：

$$p(y_t\mid y_1,y_2,\cdots,y_{t-1};x_1,x_2,\cdots,x_n)=MLP(\alpha\cdot C_t,S_t) \tag{14}$$

其中，参数 α 为注意力超参数，满足 $\alpha\in[0,1]$。MLP 是一个多层感知器。在训练模型时，α 设置为 1；在生成时，通过调整 α 的值来控制目标句对源语句的依赖程度。当 $\alpha=0$ 时，t 时刻词的生成只依赖于已经生成的目标句，这就是传统的基于生成式的文本隐写。由于缺少对源语句依赖，生成的目标句间缺乏语义相关性。随着 α 的增大，t 时刻的载密词会渐渐依赖于源语句。当 $\alpha=1$ 时，载密词对源语句和目标句的依赖程度达到一致。

本文手动选择了不同的注意力参数权重来测试模型生成的文本质量和隐写能力。不同的权值将会得到不同的概率分布，进而改变候选池的选词。

3.4 嵌入与提取算法

嵌入算法的主要思想是根据语言模型生成的概率分布构建候选池，然后选择与当前时刻秘密信息对应的词作为最终的载密词。具体嵌密过程如算法 1 所示。例如，假设秘密信息比特流 $B=\{01011011\cdots\}$，源文本为“他希望双方进一步加强交流和合作。”候选池大小为 2，当前秘密信息比特位是 0，则选择概率值最大的词作为载密词，下一比特为 1，则选择概率值次大的词作为载密词。根据秘密信息比特流得到的最终载密句为“He said he hopes that the two sides will further strengthen their exchanges and cooperation”。

算法 1 秘密信息嵌入算法

输入：秘密信息比特流 B；源文本 C；beamsize bs；方差阈值 ε；权值 α，β
输出：目标载密文本

1. 处理数据并训练模型；
2. While B 不为空 do
3. 从源文本 C 中读取一个句子；
4. If 没有到句子结尾 then
5. 根据式(14)通过模型计算下一个词的概率分布；
6. End if
7. For bs=8；bs>0，do
8. If var(bs) <ε then
9. 将 bs 个词都加入候选池；
10. Else
11. bs=bs−2
12. End if
13. End for
14. 根据候选池中候选词的概率分布构建二叉树，对候选词进行二进制编码；
15. 选择编码与当前要嵌入的秘密信息比特相对应的词作为目标词生成；
16. End while
17. Return 生成的载密文本

秘密信息提取的过程与嵌入相似。接收方共享源文本和相同参数的 NMT 模型。然后采用相同的方法构建候选池。对候选词进行编码。最后将从发送方接收到的载密文本与候选词对比，提取出对应的秘密信息比特。具体提取算法如算法 2 所示。

算法 2 秘密信息提取算法

输入：目标载密文本；源文本 C；beamsize bs；方差阈值 ε；权值 α，β
输出：秘密信息比特流 B

1. 处理数据并训练模型；
2. 从源文本 C 中读取一个句子；
3. If 没有到句子结尾 then
4. 根据式(14)通过模型计算下一个词的概率分布；
5. End if
6. For bs=8；bs>0，do
7. If var(bs) <ε then
8. 将这 bs 个词都加入候选池；
9. Else
10. bs=bs−2
11. End if
12. End for
13. 根据候选池中候选词的概率分布构建二叉树，对候选词进行二进制编码；
14. 将接收到的载密句与候选词对比，提取出对应的秘密信息比特；
15. 将提取出的比特信息加入到 B 中；
16. Return 秘密信息比特流 B

4 实验与结果分析

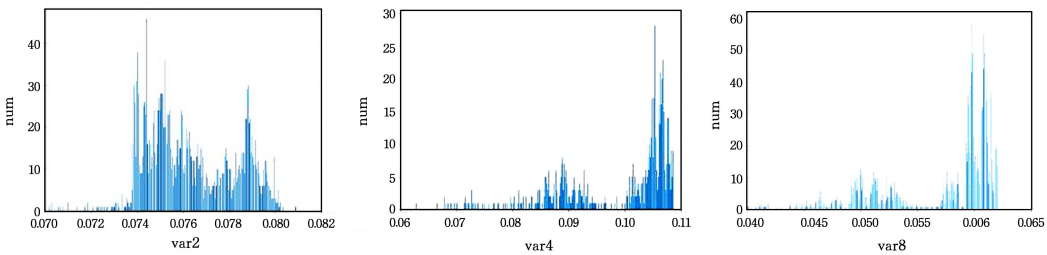
本文为了测试生成载密文本的嵌入率、文本质量以及模型的安全性，进行了一系列实验，并通过与其他生成式隐写模型对比来测试本模型在保持文本语义方面的性能。

4.1 数据处理与模型训练

本文采用从公开新闻网站获得的中-英平行语料库作为模型的数据集。该语料库共包含 1 252 977 句新闻语料。这些句子的最大长度和平均长度分别为 98 和 34。其中训练、验证和测试数据集按照 8∶1∶1 的比例进行划分。在模型训练之前，首先要对数据进行预处理，包括对特殊符号、网站链接、数字号码等的清除工作。

本文中所有的测试实验均基于 Ubuntu16.4 操作系统，Dell GX2080Ti GPU，CUDA9.0 进行。模型使用 Tensorflow[37] 搭建，Python3.6 编写。该模型的超参数设置如下：编码、解码分别设置了 6 层叠加，每一层采用 8 头注意力机制；每个词映射为 512 维的向量；模型预训练时采用 dropout[38] 正则化来避免过拟合（$dropout=0.2$）；优化算法采用 Adam，学习率初始值为 0.0003。此外，batch size 设置为 64，迭代次数为 75。

为了保证生成的载密文本的质量，设置合理的方差阈值是非常重要的。本文首先分别统计了不嵌密的情况下，top2，top4，top8 词的方差分布，其对应的直方图分别如图 4 所示。



(a)top2 词条件概率方差直方图　(b)top4 词条件概率方差直方图　(c)top8 词条件概率方差直方图

图 4　top2，top4，top8 词的条件概率方差直方图

Fig. 4　Probability-variance histogram of top2，top4，top8 words

根据得到的直方图，手动选取了 0.07，0.072，0.074，0.076 为 top2 的 4 个阈值；0.07，0.08，0.09，0.1 为 top4 的 4 个阈值；0.045，0.05，0.055，0.06 为 top8 的 4 个阈值。方差阈值对个评估指标的影响将在以下的实验中详细展示。

4.2 评估指标

在机器翻译系统中，通常采用 BLEU（Bilingual Evaluation Understudy）[39] 和 PPL（perplexity）评估文本质量。BLEU 是衡量机器翻译与专业翻译相似性的一种评估方法，*BLEU* 值越大，表示翻译的文本质量越高，其计算方法如下：

$$BLEU_N = b(C,S)\exp\left(\sum_{n=1}^{N}\omega_n \log CP_n(C,S)\right) \tag{15}$$

其中，$b(C,S)$是惩罚因子。

$$b(C,S)=\begin{cases}1, & l_c \geqslant l_s \\ e^{1-\frac{l_s}{l_c}}, & l_c \leqslant l_s\end{cases} \tag{16}$$

其中，l_c为待评价句的长度，l_s为参考句的有效长度（多个参考句时选择与l_c最接近的长度）。$CP_n(C,S)$是生成的翻译文本与语料库中的参考句的重合精度：

$$CP_n(C,S)=\frac{\sum_i \sum_k \min(h_k(c_i), \max h_k(s_{ij}))}{\sum_i \sum_k h_k(c_i)} \tag{17}$$

PPL 是衡量语言模型质量的一种评估方式。语言模型可以看作一个句子或段落上的概率分布，表示着文本中生成一个句子的概率。当一个语言模型训练好后，可以采用 *PPL* 来评估模型训练的好坏。*PPL* 值越小表示模型训练得越好，其计算方法如下：

$$perplexity = 2^{-\frac{1}{N}\sum_{i=1}^{N}\log P(s_i)} \tag{18}$$

其中，s_i 表示 t_i 时刻生成的第 i 个句子；N 是生成的句子总数；$P(s_i)$是通过语言模型计算出来的概率值。本文使用 BLEU 验证文本质量，并使用 PPL 测试生成文本的统计性能。

4.3 α 的影响

表 1 列出了 α 取不同值时生成的目标文本例句。

表 1　同一源语句 α 取不同值下生成的目标句对比

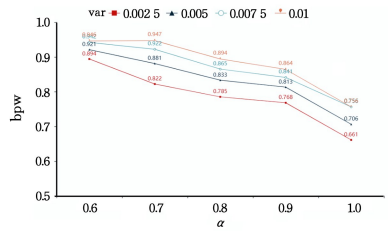
Table 1　Some examples of translations obtained under different α

α	翻译句
1	It is not the level and number of people that we have seen.
0.9	There has been a long time but not the level and number of people we have seen now.
0.8	There is always a problem but not the level and numbers that we have seen now.
0.7	However it is not the case that we have seen in the level and number of people.
0.6	It is not enough to see that the current level of work.
1	During the passage of typhoon disaster on board were found in the sky and some trees on the ground were found damaged by road safety.
0.9	During the passage of typhoon certain trees on the slope were posed a safety hazard in august last year.
0.8	During the passage of typhoon some trees on the slope were once burning and posed a risk of fallen trees.
0.7	During the passage of typhoon course there were many trees on the slope of sheungyiu last august and posed a challenge to road safety.
0.6	During the course of typhoon it was forbidden to grow down on the ground and posed a chance to prevent the spread of land.
1	The best way is to encourage enterprises of the two countries to explore areas and content of cooperation and the government has given positive support.
0.9	The best way is to encourage the companies of both sides to explore areas and content of cooperation and the government should give positive support.
0.8	The best way is to encourage the enterprises of the two countries to explore areas and content of cooperation and to offer positive support.
0.7	The best way to promote cooperation is to encourage enterprises of the two sides to explore new ways to expand cooperation and to give them positive support.
0.6	The best way to promote cooperation is to encourage enterprises to discuss ways to expand cooperation.

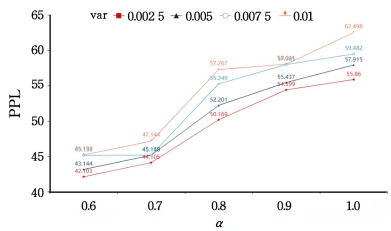
从表 1 中可以看到，在不同注意参数下生成的句子是不同的，但它们具有相似的语义属性，因此须进一步研究注意力参数与模型性能之间的关系。本节将讨论不同的 α 值对模型嵌入率和生成文本质量的影响。为了方便讨论，以 $bs=2$ 为例。

(1)α 对嵌入率的影响

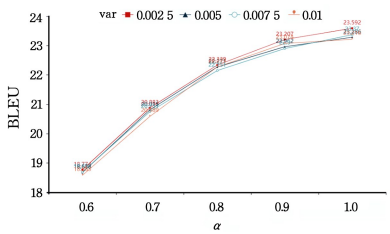
从图 5(a)可以看出，随着 α 的增加，不同模型的嵌入率 bpw 均在减小。这是因为 α 的大小反映了当前的生成词 y_t 对源语句的依赖程度。当 α 的值从 1 降到 0.6 时，当前词 y_t 的限制减少了，从而使得候选池元素的可选择范围增加了，最终使候选池的数量增加。因此，降低 α 值可以增加模型的嵌入容量。



(a)不同 α，ε 取值对模型 bpw 的影响



(b)不同 α，ε 取值对 PPL 的影响



(c)不同 α，ε 取值对 $BLEU$ 的影响

图 5 不同 α，ε 取值对模型 bpw，PPL，$BLEU$ 的影响

Fig. 5 bpw, PPL, $BLEU$ performance of different α, ε

(2)α 对 PPL 的影响

从图 5(b)可以看出，随着 α 取值的增加，PPL 值也在增加，这与 bpw 正好相反。这说明适当降低 α 的取值不仅可以增加模型的嵌入容量，还可以减小语言模型的复杂度。这是因为，当减少 y_t 对源语句的依赖程度时，y_t 的生成会更多地依赖已经生成的部分。正是这一部分提供了更多有用的信息，使得最终生成的句子分布更加接近目标文本的真实分布。

(3)α 对 $BLEU$ 的影响

从图 5(c)可以看出，$BLEU$ 的值会随着 α 的降低而降低，这进一步证明了改变 α 的取值会影响 y_t 对源语句的依赖程度。

表 2 列出了在不同方差阈值 ε 和注意力参数 α 以及不同 bs 取值下，PPL，$BLEU$ 和 bpw 的实验结果。根据表 2 可以得出，随着 α 的减小，PPL 和 $BLEU$ 也在减小，而 bpw 在增加。此外，随着 bs 和方差阈值 ε 的增加，模型的嵌入容量和复杂度也在增加，且 $BLEU$ 在减小。

表 2 不同 ε，α，bs 对 PPL，$BLEU$ 和 bpw 的影响

Table 2 Effect of different ε, α, bs on bpw, $BLEU$, PPL

ε	bs	α	bpw	PPL	$BLEU$
0.07	2	1	0.663	55.860	23.592
0.072			0.715	57.915	23.286
0.074			0.758	59.482	23.370
0.076			0.758	62.498	23.216
0.07	4		0.778	63.507	23.516
0.08			1.059	67.658	23.202
0.09			1.120	72.800	23.316
0.1			1.174	75.913	23.050
0.045	8		0.841	69.731	23.368
0.05			0.981	73.123	23.131
0.055			1.104	79.709	22.898
0.06			1.320	81.373	22.468
0.07	2	0.9	0.770	54.399	23.207
0.072			0.815	55.437	22.962
0.074			0.843	57.985	22.900
0.076			0.866	58.031	23.074
0.07	4		1.180	63.046	23.213
0.08			1.265	63.268	22.921
0.09			1.297	68.503	23.034
0.1			1.390	73.664	22.833
0.045	8		1.243	64.391	22.970
0.05			1.395	72.113	22.669
0.055			1.457	74.640	22.699
0.06			1.650	79.044	22.236
0.07	2	0.8	0.787	50.169	22.339
0.072			0.835	52.201	22.271
0.074			0.867	55.249	22.151
0.076			0.896	57.267	22.273
0.07	4		1.199	59.267	22.245
0.08			1.301	60.397	22.144
0.09			1.332	62.409	22.143
0.1			1.446	67.547	21.813
0.045	8		1.265	62.408	22.269
0.05			1.452	71.735	21.803
0.055			1.518	72.801	21.674
0.06			1.762	73.345	21.196
0.07	2	0.7	0.824	44.105	20.883
0.072			0.883	45.148	20.815
0.074			0.924	45.167	20.751
0.076			0.949	47.144	20.589
0.07	4		1.265	52.200	20.932
0.08			1.393	52.270	20.804
0.09			1.488	55.270	20.669
0.1			1.572	62.448	20.272
0.045	8		1.352	51.329	20.810
0.05			1.604	59.586	20.516
0.055			1.803	63.716	20.095
0.06			1.988	65.112	19.777
0.07	2	0.6	0.896	42.103	18.774
0.072			0.923	43.144	18.678
0.074			0.944	45.138	18.697
0.076			0.948	45.193	18.583
0.07	4		1.389	45.177	18.700
0.08			1.553	50.258	18.515
0.09			1.678	50.327	18.108
0.1			1.775	52.433	18.063
0.045	8		1.548	44.348	18.510
0.05			1.878	46.549	18.080
0.055			2.102	50.751	18.054
0.06			2.211	55.065	17.519

当嵌入率过大时，由于候选池中的候选词质量也会受到影响，因此会导致生成的载密文本质量下降，例如，源语句为“约三百三十公顷的前滨及海床将受工程影响其范围在今日四月十四日宪报内载明。”当嵌密率过大时，得到的载密句为

“The scope of the service is well affected by the three stages of the project which is scheduled for today april 15 in 2004.”。当将嵌密率调小后得到的比较好的载密句为“An area of approximately 330 hectares of foreshore and sea bed are affected by the works as required in the gazette today april 14.”。

4.4 模型对比

本文对比了3种生成式自然语言信息隐藏模型，包括两种不同的马尔可夫文本隐写模型[40-41]和一种基于循环神经网络的生成式文本隐写模型(RNN-stega[30])，并采用NFZ-WDA[42]和LS-CNN[43]两种隐写分析工具来检测载密文本的安全性。NFZ-WDA[42]是一种专门用于检测基于机器翻译的神经网络隐写分析模型。它采用n-gram算法检测载密文本与参照文本之间的统计特征变化。LS-CNN则是应用卷积神经网络CNN提取特征并检测载密文本与参照文本之间的分布差异。

本文并没有采用常规的训练方法，即针对不同的bs分别训练隐写分析工具。由于在现实世界中，攻击者得到载密文本往往是多种不同数据混合而成，因此，为了更加接近真实情况，本文将不同bs的载密文本混合，训练出一个隐写分析模型。例如，当训练LS-CNN检测载密文本时，用于训练的载密文本包含不同有效载荷bs、注意力参数α和方差阈值ε下生成的共15000句载密句。

此外，为了验证生成的载密句之间的长距离语义相关性，计算了生成的载密句的语义距离以及句子的平均长度。语义距离使用开源的OpenAIGPT[44]模型[1)]来计算。不同模型的隐写分析准确率和语义距离结果如表3所列。

表3 相同数据集下的不同模型性能对比

Table 3 Comparisions of methods performance for each evaluation indication on the same News data set

模型	ε	bpw	NFZ-WDA	LS-CNN	语义距离	平均句长	PPL	$BLEU$
Markov[40]	—	1	82.5	95.2	354.317	17	493.992	0.994
		2	80.5	94.8	355.293	17	577.628	0.731
		3	82.5	95	368.158	23	585.311	0.863
Markov[41]		1	88.5	94.9	329.113	19	294.578	1.681
		2	84.5	96.6	360.997	20	486.043	0.973
		3	82	96.8	373.532	25	531.080	0.607
RNN-Stega[30]	—	1	76.5	80.5	280.612	26	44.080	10.362
		2	73	88	324.685	28	67.915	8.041
		3	64.5	89.8	331.454	31	136.542	5.679
Seq2Seq-Stega	0.045	0.841	**53**	**56.5**	**87.916**	46	69.731	**23.368**
	0.05	0.981	53.5	58.2	89.941	46	73.123	23.131
	0.055	1.104	53.5	59	93.218	46	79.709	22.898
	0.06	1.320	56	61.5	94.301	46	81.373	22.468
Seq2Seq-Stega	0.045	1.548	55.5	**53.4**	**92.019**	**47**	**44.348**	18.510
	0.05	1.878	52.5	55	96.428	**47**	46.549	18.08
	0.055	2.102	57.5	59.7	106.440	**47**	50.751	18.054
	0.06	**2.211**	55.5	65.5	113.629	**47**	55.065	17.519

从表3可以看出，本文提出的模型在抵抗隐写分析攻击和保持句间语义上的性能均高于其他3种模型。与RNN-stega[30]相比，本文模型生成的载密文本在相同嵌入率下，NFZ-WDA[42]和LS-CNN[43]的检测正确率分别降低了21.5%和28.3%。这说明该模型具有较高的安全性，更能抵抗恶意攻击。此外，尽管该模型生成的载密文本的平均句长比较长，但句间的语义距离却比较小，说明即使生成长句子，模型也能够很好地保持长句间的语义关系。

结束语 本文提出了保持语义关系的文本隐写模型。为了保证文本质量，采用了一种动态选词策略，根据单词的条件概率分布自适应地确定候选池。本文还引入了注意力参数，以平衡源文本和目标文本对预测载密词的贡献。同时本文对模型的嵌入率、文本质量、抗隐写分析能力和语义距离进行了验证实验。结果表明，本文提出的模型在抗隐写分析能力和保持语义相关性方面优于其他3种隐写模型。

参 考 文 献

[1] ZHANG H,CAO Y,ZHAO X F. Motion vector-based video steganography with preserved local optimality[J]. Multimedia Tools and Applications,2016,75(2):13503-13519.

[2] LIN T J,CHUNG K L,CHANG P C,et al. An improved DCT-based perturbation scheme for high capacity data hiding in H.264/AVC intra frames[J]. Journal of Systems & Software,2013,86(3):604-614.

[3] JANGID S,SHARMA S. High PSNR based video steganography by MLC(multi-level clustering) algorithm[C]// 2017 International Conference on Intelligent Computing and Control Systems (ICICCS). 2017.

[4] BAI Y Q,JIANG G Y,ZHU Z J,et al. Reversible data hiding scheme for high dynamic range images based on multiple prediction error expansion[J]. Signal Processing:Image Communication,2020,91.

[5] EVSUTIN O,MELMAN A,MESHCHERYAKOV R. Algorithm of error-free information embedding into the DCT domain of digital images based on the QIM method using adaptive masking of distortions[J]. Signal Processing,2020,179(10):107811-107829

[6] PU W. Deep SAR Imaging and Motion Compensation[J]. IEEE Transactions on Image Processing,2021,PP(99):1-1.

[7] NANDAL V,SINGH P. Hybrid Optimized Image Steganography with Cryptography[C]// Computational Methods and Data Engineering. 2021.

1) https://github.com/AlexanderYogurt/pytorch-pretrained-BERT

[8] ATOUM M S. Evolutionary Detection Accuracy of Secret Data in Audio Steganography for Securing 5G-Enabled Internet of Things[J]. Symmetry, 2020, 12(12): 2071-2088.

[9] SOLIMAN N F, KHALIL M I, ALGARNI A D, et al. Efficient HEVC Steganography Approach Based on Audio Compression and Encryption in QFFT Domain for Secure Multimedia Communication[J]. Multimedia Tools and Applications, 2020(2): 4789-4823.

[10] JIANG S, YE D, HUANG J, et al. SmartSteganogaphy: Lightweight generative audio steganography model for smart embedding application[J]. Journal of Network and Computer Applications, 2020, 165: 102689.

[11] WANG X, YANG L T, SONG L, et al. A Tensor-based Multi-Attributes Visual Feature Recognition Method for Industrial Intelligence[J]. IEEE Transactions on Industrial Informatics, 2020(99): 1-1.

[12] JIA J, ZHANG G, HU C, et al. Information hiding method for long distance transmission in multi-channel IOT based on symmetric encryption algorithm[J]. Journal of Ambient Intelligence and Humanized Computing, 2021, 10(2): 1007-1017.

[13] YANG Z, ZHANG S, HU Y, et al. VAE-Stega: Linguistic Steganography Based on Variational Auto-Encoder[J]. IEEE Transactions on Information Forensics and Security, 2020, 16(10): 1109-1124.

[14] CHAUDHARY S, DAVE M, SANGHI A. Aggrandize text security and hiding data through text steganography[C]// 2016 IEEE 7th Power India International Conference (PIICON). IEEE, 2016.

[15] FU Z J, SUN X M, ZHOU L, et al. New forensic methods for ooxml format documents[C]// 2013 In International Workshop on Digital Watermarking, 2013.

[16] BARMAW I, AR I. Linguistic Based Steganography Using Lexical Substitution and Syntactical Transformation[C]// International Conference on It Convergence & Security. IEEE, 2016: 1-6.

[17] ZHANG J, WANG W, YANG X, et al. A word-frequency-preserving steganographic method based on synonym substitution [J]. International Journal of Computational Science and Engineering, 2016, 1(1): 1.

[18] TOPKARA M, TOPKARA U, ATALLAH M J. Information hiding through errors: a confusing approach[J]. Proc Spie, 2007, 6505.

[19] AGRAWAL R, SHARMA M, SINGH B K. Hiding Patient Information in Medical Images: A Robust Watermarking Algorithm for Healthcare System[M]// Advances in Biomedical Engineering and Technology. 2021.

[20] JIN C, ZHANG D, PAN M. Chinese Text Information Hiding Based on Paraphrasing Technology[C]// Information Science & Management Engineering. IEEE, 2010.

[21] KANG H, WU H, ZHANG X. Generative Text Steganography Based on LSTM Network and Attention Mechanism with Keywords[J]. Electronic Imaging, 2020, 291(8): 1-8.

[22] GROTHOFF C, GROTHOFF K, ALKHUTOVA L, et al. Translation-Based Steganography[C]// 7th International Workshop on Information Hiding. Springer, Berlin, Heidelberg, 2005.

[23] MINH-THANG LUONG† *. Addressing the Rare Word Problem in Neural Machine Translation[J]. Bulletin of University of Agricultural Sciences & Veterinary Medicine Cluj Napoca Veterinary Medicine, 2015, 27(2): 82-86.

[24] KARTIKA A S. Steganografi linguistik metode NICETEXT menggunakan kata dan variasi pola kalimat dasar bahasa indonesia[J]. UT-Computer Science, 2014, 1740(3): 865-880.

[25] POLIDORI C, NIEVES-ALDREY J L, GILBERT F, et al. Hidden in taxonomy: Batesian mimicry by a syrphid fly towards a Patagonian bumblebee[J]. Insect Conservation & Diversity, 2014, 7(1): 32-40.

[26] SHNIPEROV A N, NIKITINA K A. A text steganography method based on Markov chains[J]. Automatic Control & Computer Sciences, 2016, 50(8): 802-808.

[27] DESOK Y, ABDELRAHMA N. Jokestega: Automatic joke generation-based steganography methodology [J]. International Journal of Security & Networks, 2012, 7(3): 148-160.

[28] MANSOOR F, MOHSEN R. An email-based high capacity text steganography using repeating characters[J]. International Journal of Computers & Applications, 2018: 1-7.

[29] LUO Y, HUANG Y, LI F, et al. Text Steganography Based on Ci-poetry Generation Using Markov Chain Model [J]. Ksii Transactions on Internet & Information Systems, 2016, 10(9): 4568-4584.

[30] 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.

[31] STUTSMAN R, GROTHOFF C, ATALLAH M, et al. Lost in just the translation[C]// Proceedings of the 2006 ACM Symposium on Applied Computing. ACM, 2006: 338-345.

[32] GROTHOFF C, GROTHOFF K, ALKHUTOVA L, et al. Translation-based steganography[C]// International Workshop on Information Hiding. 2009.

[33] MENG P, SHI Y Q, HUANG L, et al. LinL: Lost in n-best List [C]// International Workshop on Information Hiding. Springer Berlin Heidelberg, 2011.

[34] AHMADNIA B, DORR B J. Impact of a New Word Embedding Cost Function on Farsi-Spanish Low-Resource Neural Machine Translation[C]// The Thirty-Third International Flairs Conference. 2020.

[35] O'BRIEN S, ROSSETTI A. Neural machine translation and the evolution of the localisation sector: Implications for training[J]. The Journal of Internationalization and Localization, 2020, 7(1): 95-121.

[36] BAHDANAU D, CHO K, BENGIO Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. Computer ence, 2014, 1405(2): 36-50.

[37] ABAD I, MARTI N. TensorFlow: Learning Functions at Scale [J]. Acm Sigplan Notices A Monthly Publication of the Special Interest Group on Programming Languages, 2016, 51(9): 1-1.

[38] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

(下转第 584 页)

[8] SAHAI A, WATERS B. Fuzzy identity-based encryption [C]// Advances in Cryptology (EUROCRYPT 2005). Berlin: Springer, 2005: 457-473.
[9] DAN B, FRANKLIN M. Identity-based encryption from the Weil pairing [C]// Proceedings of the 21st Annual International Cryptology Conference on Advances in Cryptology. Berlin: Springer, 2001: 213-229.
[10] BETHENCOURT J, SAHAI A, WATERS B. Ciphertext-policy attribute-based encryption [C]// Proceedings of the 2007 IEEE Symposium on Security and Privacy. Piscataway, NJ: IEEE, 2007: 321-334.
[11] GOYAL V, PANDEY O, SAHAI A, et al. Attribute-based encryption for fine-grained access control of encrypted data [C]// Proceedings of the 13th ACM Conference on Computer and Communications Security. New York, NY: ACM, 2006: 89-98.
[12] LI S, XU M Z. Attribute-based public encryption with keyword search[J]. Chinese Journal of Computers, 2014, 37(5): 1017-1024.
[13] ZHENG Q J, XU S H, ATENIESE G. VABKS: Verifiable attribute-based keyword search over outsourced encrypted data [C]// Proceedings of the IEEE Conference on Computer Communications. Piscataway, NJ: IEEE, 2014: 522-530.
[14] LIANG K, SUSILO W. Searchable attribute-based mechanism with efficient data sharing for secure cloud storage[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(9): 1981-1992.
[15] SONG Y, HAN Z, CHEN D, et al. Attribute-based encryption supporting arbitrary conjunctive key word search[J]. Journal on Communications, 2016, 37(8): 77-85.
[16] SUN W, YU S, LOU W, et al. Protecting your right: verifiable attribute-based keyword search with fine-grained owner-enforced search authorization in the cloud[J]. IEEE Transactions on Parallel and Distributed Systems, 2016, 27(4): 1187-1198.
[17] WU Q Y, MA J F, LI H, et al. Multi-keyword search over encrypted data with user revocation[J]. Journal on Communications, 2017, 38(8): 183-193.
[18] YAN X X, MENG H. Ciphertext policy attribute-based encryption scheme supporting direct revocation[J]. Journal on Communications, 2016, 37(5): 44-50.
[19] IBRAIMI L, TANG Q, HARTEL P, et al. Efficient and provable secure ciphertext-policy attribute-based encryption schemes [C]// Proceedings ofthe 5th International Conference on Information Security Practice and Experience, Berlin: Springer, 2009: 1-12.
[20] LI J, WANG Q, WANG C, et al. Fuzzy keyword search over encrypted data in cloud computing [C]// Proceedings of the IEEE Conference on Information Communications, Piscataway, NJ: IEEE, 2010: 441-445.
[21] WANG Y, FAN K. Effective CP-ABE with Hidden Access Policy[J]. Journal of Computer Research and Development, 2019, 56(10): 2151-2159.
[22] HE H, ZHANG J, GU J G, et al. A fine-grained and lightweight data access control scheme for WSN-integrated cloud computing [J]. Cluster Computing, 2017, 20(2): 1457-1472.
[23] RUIXUAN L, CHENGLIN S, HENG H, et al. A Lightweight Secure Data Sharing Scheme for Mobile Cloud Computing[J]. IEEE Transactions on Cloud Computing, 2018, 6(2): 344-357.

HE Heng, born in 1981, Ph.D, associate professor, is a member of China Computer Federation. His main research interests include network security, cloud computing and ciphertext retrieval.

(上接第 564 页)

[39] PAPINENI K, ROUKOS S, WARD T, et al. Bleu: a method for automatic evaluation of machine translation [C] // Association for Computa-tional Linguistics, 2002.
[40] YANG Z, JIN S, HUANG Y, et al. Automatically Generate Steganographic Text Based on Markov Model and Huffman Coding[J]. arXiv: 1811.04720.
[41] SHNIPEROV A N, NIKITINA K A. A text steganography method based on Markov chains [J]. Automatic Control & Computer Sciences, 2016, 50(8): 802-808.
[42] CHEN Z, HUANG L, MENG P, et al. Blind Linguistic Steganalysis against Translation Based Steganography[C]// International Workshop on Digital Watermarking. Springer, Berlin, Heidelberg, 2010.
[43] WEN J, ZHOU X, ZHONG P, et al. Convolutional Neural Network Based Text Steganalysis[J]. IEEE Signal Processing Letters, 2019, PP(3): 1-1.
[44] LEE J S, HSIANG J. Patent claim generation by fine-tuning OpenAI GPT-2 [J]. World Patent Information, 2020, 62: 101983.

ZHOU Xiao-shi, born in 1994, postgraduate. Her main research interests include deep transfer learning for image classification, deep learning for text information steganography, natural language processing, machine learning.

WEN Juan, born in 1982, Ph.D, associate professor. Her main research interests include artificial intelligence, information hiding, and natural language processing.