ICStega: Image Captioning-based Semantically Controllable Linguistic Steganography

Xilong Wang, Yaofei Wang*, Kejiang Chen, Jinyang Ding, Weiming Zhang*, and Nenghai Yu

Abstract

Nowadays, social media has become the preferred communication platform but brought security threats, while generation-based Linguistic steganography can hide secret data in text. However, most existing approaches cannot control the semantic of text. In this paper, we created a novel image captioning-based stegosystem, where semantics of stego text can be controlled. To balance the conflict between payload capacity and semantic preservation, we proposed a new sampling method to cutoff low-probability words. Experiments have shown that our method can control semantic accuracy & diversity, payload capacity, and security at the same time.

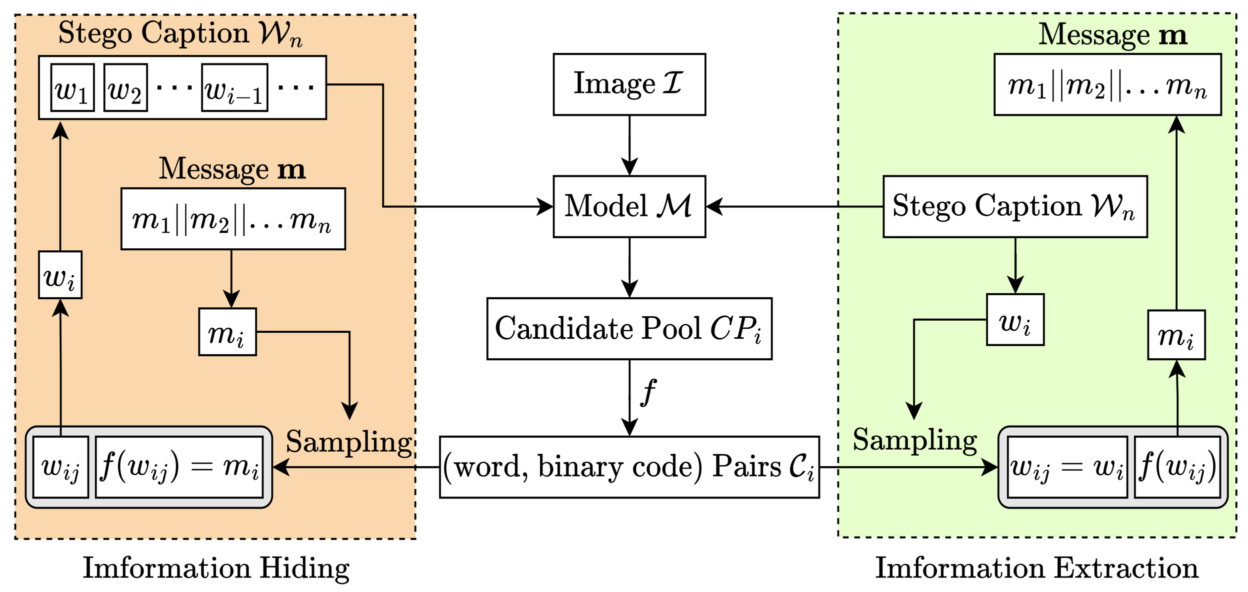
Keywords: Linguistic Steganography, Image Captioning, Behavioral Security

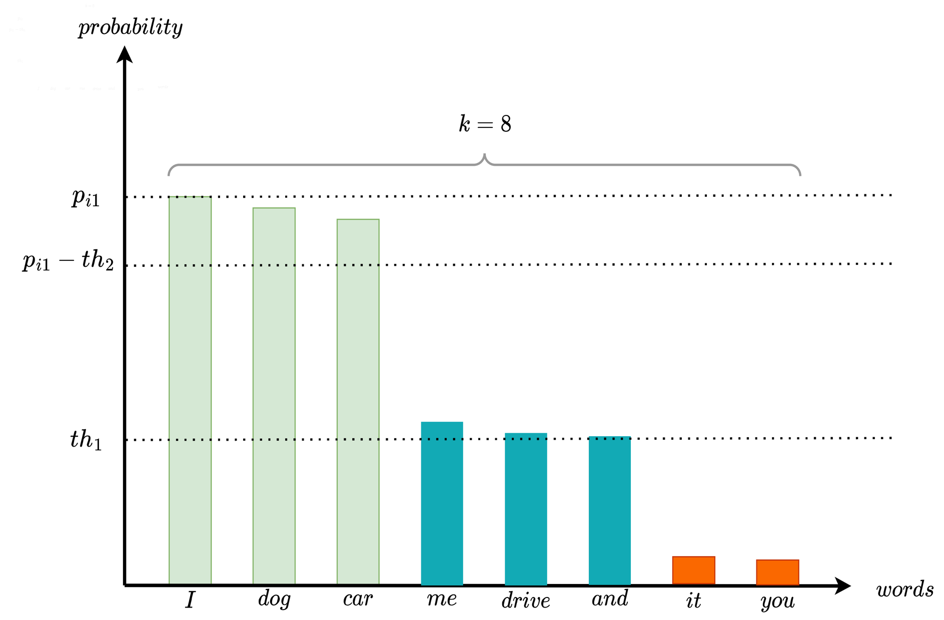
Introduction

Steganography is the science of hiding secret message in inconspicuous objects. In this paper, we propose a novel image captioning-based linguistic steganography (ICStega) where the secret messages are embedded into the generated captions. The main contributions are as follows: (a) We created a new application scenario for linguistic steganography where the secret messages are transmitted by sending the image-stego text pairs to social media. (b) the stego text of our method is semantically controllable, which can describe images in words. (c) we proposed an optimized strategy to build the candidate pool, so as to improve the semantic accuracy and further control the semantic expression of stego text.

Method

As shown in the figure on the left, in the information hiding process, based on an uploaded image $\mathcal{I}$, the image captioning model $\mathcal{M}$ generates the conditional probability of words in the vocabulary. Next, a candidate pool (CP) containing words with the largest probabilities is built according to the conditional probability distribution. In this process, we propose a new method called Two-Parameter Semantic Control Sampling to cutoff words with low probabilities as shown in figure on the right. Then, words in CP were assigned binary codes by a certain encoding strategy and sampled to form the stego caption carrying the secret message $\boldsymbol{m}$.





Related Work

Generally, linguistic steganography (LS) can be divided into two categories: edit-based and generation-based. Compared to edit-based linguistic steganography, generation-based approaches largely improve the payload capacity. To improve the performance of generation-based LS, Yang et al. proposed Huffman coding. However, most generation-based steganography cannot control the semantics of the stego text and thus cannot satisfy semantically controllable application scenarios. Image captioning (IC) is the task of describing an image in words. The words in the caption are predicted one by one according to the conditional probability distribution. This feature allows us to perform generation-based linguistic steganography based on IC.

Results

In this paper, we propose an image captioning-based steganography approach, so we need to consider the accuracy of captions, i.e., how relevant captions and images are. To this end, we use two evaluation metrics $BLEU_4$ and METEOR. In addition, we utilize LSA and Self-CIDEr to measure the diversity of captions, i.e., describing a single image with different captions. ICStega does not degrade the accuracy while reaching a certain degree of diversity, thus finding a balance between these two factors with the benefit of Two-Parameter Semantic Control Sampling.

	top k	$BLEU_4\uparrow$	METEOR↑	LSA↑	Self-CIDEr↑
Captions with Top-K	2	0.320	0.278	0.506	0.721
	4	0.251	0.257	0.596	0.833
	8	0.215	0.243	0.627	0.871
	16	0.204	0.233	0.661	0.891
	32	0.187	0.228	0.657	0.897
	64	0.177	0.224	0.672	0.905
ICStega		0.312	0.271	0.517	0.717

Conclusion

In this paper, we put forward ICStega, where the semantics of the stego text can be controlled. Unlike previous generation-based methods, as an image captioning task, we need to pay special attention to the semantic accuracy of stego captions. Therefore, we also put forward a new method called Two-Parameter Semantic Control Sampling to cutoff words with low probabilities. We show through metrics and examples that our approach not only has better payload capacity, imperceptibility, and security performance but also can properly describe the content of the image, which satisfies the need of sending the semantically related image-text pairs in social media. In future work, we will further enhance the security, semantic accuracy, and certain robustness without losing payload capacity.

Contact Information

Name *Xilong Wang*
Email wxl20010508@mail.ustc.edu.cn
Homepage http://home.ustc.edu.cn/~wxl20010508

Thanks for your attention