به‌نام خدا

فرم پیشنهاد پروژه‌ی درس بازیابی پیشرفته اطلاعات

1401-01

اعضای گروه:

آرین تشکر ۴۰۰۲۳۴۹۴
شهاب مهره‌کش ۴۰۰۳۲۱۳۴

عنوان پروژه:

نهان نگاری متنی با استفاده از یک Language model از پیش آموزش داده شده

الف – توضیح مساله

در این پروژه هدف طراحی یک نهان نگار متنی برای پنهان کردن یک پیام مخفی در بستر یک متن است به طوری که متن تولید شده از متون واقعی زبان طبیعی غیرقابل تشخیص باشد و بتوان از آن به عنوان یک چارچوب برای برقراری ارتباط امن استفاده کرد.

ب- کارهای مهم قبلی

در [1] با استفاده از یک شبکه GAN مبتنی بر LSTM یک Language model با قابلیت تخمین احتمال کلمه بعدی به شرط دنباله کلمات پیشین ارائه شده است. می‌توان با استفاده از این فرض که کلمات بعدی با احتمال بالا هم معنی هستند و یا ادامه جمله به نحوی با همه آن کلمات با معنی است، یک پیام مخفی را درون متن تولید شده پنهان کرد. در [2] همین ایده با استفاده از Attention LSTM بر روی زبان چینی دنبال شده است. [1] Yang, Z., Wei, N., Liu, Q., Huang, Y., & Zhang, Y. (2019, November). GAN-TStega: Text steganography based on generative adversarial networks. In *International Workshop on Digital Watermarking* (pp. 18-31). Springer, Cham. [2] Kang, H., Wu, H., & Zhang, X. (2020). Generative text steganography based on LSTM network and attention mechanism with keywords. *Electronic Imaging*, *2020*(4), 291-1.

ج- مشخصات دیتاست و ویژگی‌های موجود در آن

در این پروژه از دو مجموعه داده Image COCO و EMNLP WMT17 برای Fine-tune کردن مدل استفاده شده است که در ادامه آن‌ها را معرفی می‌کنیم.

۱- Image COCO : مجموعه داده ای برای شرح تصاویر (Image Captioning) که شامل بیش از یک و نیم میلیون عنوان است که بیش از ۳۳۰۰۰۰ تصویر را توصیف می کند[3].

۲- EMNLP WMT17 : یک مجموعه داده برای ترجمه ماشینی است که داده های اخبار انگلیسی را در آن انتخاب کرده و برای پروژه مورد اسفاده قرار گرفته است[4].

[3] Chen, X., Fang, H., Lin, T. Y., Vedantam, R., Gupta, S., Dollár, P., & Zitnick, C. L. (2015). Microsoft coco captions: Data collection and evaluation server. *arXiv preprint arXiv:1504.00325*.

[4] https://www.statmt.org/wmt17/translation-task.html

د- روش پیشنهادی شما

در روش پیشنهادی ما به جای آموزش یک Language Model به صورت تخاصمی (GAN) از یک Language Model از پیش آموزش داده شده استفاده خواهد شد که بر روی متون هدف مورد نظر Fine-tune شده است.

هدف آن است که با استفاده از یک Language Model از پیش آموزش داده شده بتوان:

۱- متون با کیفیت تر و با قدرت پنهان پذیری بالاتر تولید کرد.

۲- پیوستگی معنایی بین جملات برقرار باشد.

نحوه ی پنهان سازی اطلاعات بر مبنای این امر است که در صورت استفاده از یک Language Model به اندازه کافی پیچیده، کلمه های با احتمال بالا به شرط context قبلی (کلمه های پیشین) همه کلمه های کم و بیش مناسبی برای ادامه جمله و تولید جمله های با معنی می باشند.

بدین ترتیب، پیام مخفی ابتدا به صورت یک دنباله از بیت ها در خواهد آمد. سپس قطعات k - بیتی در هر مرحله از این دنباله بیت ها جدا می شوند. این عدد k – بیتی به مبنای ۱۰ (مانند n) تبدیل می شود. با شروع از یک کلمه ی دلخواه و یا تصادفی و با استفاده از Language Model، کلمه های بعدی به شرط context قبلی انتخاب خواهند شد. با این تفاوت که به جای انتخاب محتمل ترین کلمه در هر مرحله، کلمه n ام (مبنای ۱۰ عدد k-بیتی برداشته شده) به عنوان کلمه بعدی انتخاب می شود و بدین ترتیب رشته k – بیتی درون این کلمه ذخیره می شود. در حین رمز گشایی این متن، دقیقاً عکس همین عملیات صورت می گیرد. با شروع از همان کلمه تصادفی اولیه و با دانستن کلمه بعدی، به دنبال رتبه کلمه بعدی در پیش بینی های Language Model می گردیم و سپس این رتبه را به مبنای ۲ تبدیل می کنیم تا به همان k – بیت ابتدایی برسیم.

به عنوان مثال، متون جدول ۱ از روش ارائه شده در مقاله GAN-TStega تولید شده اند. همچنین در این جدول تعداد بیت های مخفی شده در هر کلمه (عدد k توضیح داده شده در بالا) را نیز می توان مشاهده کرد.

Embedding rate (bpw)	Generated steganographic sentences
0	A person riding a motorbike with people A person wearing red shirt standing next to a shop Man is riding a motorcycle down a beach
1	A cat sitting on top of a wooden bathroom tub A motorcycle parked in front of a temple A pine apple in a corner
2	A safety conscious rock outside of formation in the night direction A blurry mirror cuts a sink and medicine pen A narrow bathroom with a mirror looking toilet behind some ruins

جدول ۱: نمونه متون نهان نگاری شده با استفاده از چارچوب GAN-TStega

ه- معیار(های) ارزیابی نتایج

برای ارزیابی کیفیت متن تولید شده از مقایسه توزیع احتمالی آن با توزیع احتمالی یک Corpus از متون واقعی استفاده می‌شود. یکی از راه‌های انجام این مقایسه استفاده از معیار Perplexity است که به صورت زیر تعریف می‌شود:

$$Perplexity(T_{gen}) = 2^{-\frac{1}{N}\sum_{i=1}^{m} \log_2 p(w_i)}$$

که w_i، i-امین سمبل متن تولید شده است.

برای مقایسه روش پیشنهادی با روش پیشین از مقایسه معنایی متون تولید شده استفاده می‌شود.