ИССЛЕДОВАНИЕ ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ МЕТРИКИ RIBES ДЛЯ ОЦЕНКИ ЛИНГВИСТИЧЕСКИХ СТЕГОСИСТЕМ

К.А. Ахрамеева, *доцент, к.т.н., Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича, oklaba@mail.ru;*
Д.Ю. Мицковский, *Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, denism111198@yandex.ru.*

УДК 004.7

Аннотация. В данной статье представлены результаты исследования использования метрики *RIBES* для оценки стеготекста лингвистических стегосистем, а также приводится оценка метрики *RIBES* в качестве оценки текста.

Ключевые слова: лингвистическая стеганография; метрика *RIBES*; метрики оценки перевода.

RESEARCH ON EVALUATING THE SAFETY OF LINGUISTIC STEGOSYSTEMS BASED ON THE RIBES METRIC

Kseniya Ahrameeva, *associate professor, St. Petersburg state university of telecommunications n/a prof. M.A. Bonch-Bruevich;*
Denis Mickovskij, *St. Petersburg National Research University of Information Technologies, Mechanics and Optics.*

Annotation. This article presents the results of a study of the use of the *RIBES* metric for evaluating the stegotext of linguistic stegosystems, and also provides an assessment of the *RIBES* metric as a text evaluation.

Keywords: linguistic steganography; *RIBES* metric; translation evaluation metrics.

Лингвистическая стеганография является перспективным направлением в защите информации ввиду сложности стегоанализа данных стегосистем, однако вместе с тем оценка безопасности данных систем является проблематичной, ввиду необходимости каким-либо образом точно определить качество стеготекста. Для осуществления такой проверки в масштабах сети необходима программа, а для программы нужна определённая метрика – числовой показатель, который может отразить оценку качества стеготекста, примерно совпадающую с оценкой стеготекста человеком, поскольку, программа не может «понять» смысл слов в стеготексте. Необходимость нахождения программного решения подтверждается наличием множества исследований в данной области, например в работах [1-4]. В качестве одного из решений для программной реализации оценивания предлагается использовать метрику оценки перевода *RIBES.*

Лингвистическая стеганография скрывает стеганограмму, изменяя информацию, которая кодируется на основе лингвистического порядка текста. То есть, для скрытия сообщений изменяется сам текст оригинального сообщения – контейнера. Одним из основных преобразований, используемых в лингвистической стеганографии, является, например, замена синонимов, предполагающая использование пар синонимов для передачи стеганограмм. Например, в тексте «Дэвид завел машину и, въехав во двор, поставил ее под каштаном у фасада, затем вытащил чемодан, портфель и висевший на вешалке джинсовый костюм и понес в дом». Можно использовать пары абсолютных

синонимов (выражения, которые можно заменить другим выражением в любом контексте без изменения его смысла) «машина-автомобиль», «затем-потом», «портфель-сумка» для передачи бит, где первое слово из пар синонимов принимается за 0, а второе – за 1. Например: Дэвид завел автомобиль и, въехав во двор, поставил ее под каштаном у фасада, затем вытащил чемодан, сумку и висевший на вешалке джинсовый костюм и понес в дом. В данном случае передаётся 101, то есть можно передать, как минимум, три бита.

Ещё одним методом лингвистической стеганографии является стеганография с изменением порядка слов. Для передачи стегосообщения используется порядок слов в предложении. Например:

В четверг, в центральной части города проходят фестивали и ярмарки.
L *T* *V* *S* *M*

Данное предложение можно преобразовать в «В центральной части города, в четверг, проходят ярмарки и фестивали». Что даёт последовательность *TLVMS*. И так далее. Всего, при данном разделении, возможно 5! = 120 перестановок, но вариаций, при которых сохраняется естественность текста, гораздо меньше, а именно 8: *LTVSM, LTVMS, TLVMS, TLVSM, TVMSL, TVSML, LVMST, LVSMT*. То есть в предложение можно вложить три бита [6]. Описанные выше методы являются одними из основных типов изменений стеготекста, поэтому данные преобразования были рассмотрены при изучении работы метрики.

Используемые до сих пор методы оценки безопасности лингвистических стегосистем можно разделить на две категории: автоматическая оценка и оценка человеком. Для автоматической метрики используются метрики оценки машинного перевода (такие как *BLEU* [4]).

Метрика *RIBES* фокусируется на порядке слов. Для расчёта используются коэффициенты ранговой корреляции, ρ – коэффициент Спирмена, показывающий насколько порядковый номер слова в стеготексте отличается от порядкового номера слова в эталоне, и τ – коэффициент Кендалла, показывающий в каком направлении порядок слов стеготекста отличается от эталона – в отрицательном или в положительном. Коэффициенты вычисляются по формулам (1-4)

$$\rho = 1 - 6\frac{\sum d^2}{n^3 - n}, \quad (1)$$

$$\tau = \frac{\sum S - \sum Q}{\frac{1}{2}n(n-1)}, \quad (2)$$

где: *d* – различия в значениях рангов, *n* – общее количество рангов, *S* – количество рангов после *n*-го ранга, превышающих значение *n*-го ранга, *Q* – количество рангов после *n*-го ранга, чьё значение меньше значения *n*-го ранга

Эти Ранговые меры могут быть нормализованы для обеспечения положительных значений:

- нормализованный коэффициент Спирмена ρ (*NSR*):

$$NSR = \frac{\rho + 1}{2} \quad (3)$$

- нормализованный коэффициент Кендалла τ (*NKT*):

$$NKT = \frac{\tau + 1}{2} \quad (4)$$

Для учёта изменений слов в метрике также учитывается точность. С учётом нормализованных ранговых коэффициентов и точности метрика *RIBES* вычисляется по формуле (5) [5]. Чем выше значение метрики, тем лучше оцениваемый текст.

$$RIBES = \frac{NSR_1 * P^{\alpha} + NST_1 * P^{\alpha}}{2}, \quad (5)$$

где: α – параметр в диапазоне $0 < \alpha < 1$.

В табл. 1 представлен расчёт метрик *RIBES* для стеготекста с изменением порядка слов.

Таблица 1.

Слово предложения	Первое предложение	Второе предложение
Нормализованный коэффициент Спирмана	0,97	0,99
Нормализованный коэффициент Кендалла	0,94	0,96
Точность	1	1
RIBES	96%	98%
Общее значение *RIBES*	97%	

$$RIBES_1 = \frac{NSR_1 * P^{\alpha} + NST_1 * P^{\alpha}}{2} = \frac{0.97*1 + 0.94*1}{2} = 0{,}96 \quad (6)$$

$$RIBES_2 = \frac{NSR_2 * P^{\alpha} + NST_2 * P^{\alpha}}{2} = \frac{0.99*1 + 0.96*1}{2} = 0{,}98 \quad (7)$$

$$RIBES_{\text{общ}} = \frac{RIBES_1 + RIBES_2}{2} = \frac{0.96+0.98}{2} = 0{,}97 \quad (8)$$

Исследование производится на основе текста, который преобразовывается с помощью двух разных лингистических стегосистем (синонимическая и на основе изменения порядка слов текста), а также комбинации изменений.

К тексту подбираются четыре ссылочных текста, которые используются метриками оценки текста для совершения расчёта. Расчёты *RIBES* и ранговых коэффициентов представлены в табл. 1-4.

Так, как и у стеготекста, и у эталона полностью совпадают слова, то очевидно, что точность будет равна единице для обоих предложений. При этом значения *RIBES* при всех значениях коэффициента *α* будут равны.

RIBES рассчитывается согласно ранжированию слов, исходя из их порядка в предложениях эталона и стеготекста, при этом в расчёте также учитываются совпадения слов с помощью использования точности в расчётах, чтобы текст, с полным совпадением порядка нескольких слов, находящихся и в стеготексте, и в эталоне, но с большим числом случайных слов, не могли быть оценены положительно. Чем больше совпадений в порядке слов и больше совпадений слов в текстах, тем выше значение *RIBES*.

Для стеготекста на основе замены синонимов расчеты гораздо более тривиальны. Так как порядок всех не заменённых слов совпадает, то очевидно, что сумма коэффициента $\sum d^2$ будет равна, соответственно коэффициент Спирмана $\rho = 1 - 0 = 1$, как и нормализованный коэффициент $NSR = \frac{1+1}{2}$ = 1 для обоих предложений.

Расчёт метрики *RIBES* для стеготекста с изменением порядка слов представлен на формулах (6-8). Для стеготекста с заменой синонимов для первых и вторых предложений были получены значения точности 0,73 и 0,91 соответственно. Расчёт метрики *RIBES* представлены на формулах (9-17).

Для стеготекста с комбинированием изменений для первых и вторых предложений также были получены значения точности 0,73 и 0,91, так как точность

при изменении порядка слов не изменилась относительно исходного предложения. Расчёт метрики *RIBES* представлен проведен с помощью формул (18-26).

В табл. 2 представлен расчёт метрик *RIBES* для стеготекста с заменой синонимов.

Таблица 2.

Слово предложения	Первое предложение	Второе предложение
Нормализованный коэффициент Спирмана	1	1
Нормализованный коэффициент Кендалла	1	1
Точность	$\frac{11}{15}$ = 0,73	$\frac{21}{23}$ = 0,91
RIBES (при α = 0)	100%	100%
RIBES (при α = 0,5)	85%	95%
RIBES (при α = 1)	73%	91%
Общее значение *RIBES* (при α = 0)	100%	
Общее значение *RIBES* (при α = 0,5)	90%	
Общее значение *RIBES* (при α = 1)	82%	

$$RIBES_1(\alpha = 0) = \frac{NSR_1 * P^{\alpha} + NST_1 * P^{\alpha}}{2} = \frac{1*0{,}73^0 + 1*0{,}73^0}{2} = 1 \quad (9)$$

$$RIBES_2(\alpha = 0) = \frac{NSR_2 * P^{\alpha} + NST_2 * P^{\alpha}}{2} = \frac{1*0{,}91^0 + 1*0{,}91^0}{2} = 1 \quad (10)$$

$$RIBES_1(\alpha = 0.5) = \frac{1*0{,}73^{0,5} + 1*0{,}73^{0,5}}{2} = 0{,}85 \quad (11)$$

$$RIBES_2(\alpha = 0.5) = \frac{1*0{,}91^{0,5} + 1*0{,}91^{0,5}}{2} = 0{,}95 \quad (12)$$

$$RIBES_1(\alpha = 1) = \frac{1*0{,}73^1 + 1*0{,}73^1}{2} = 0{,}73 \quad (13)$$

$$RIBES_2(\alpha = 1) = \frac{1*0{,}91^1 + 1*0{,}91^1}{2} = 0{,}91 \quad (14)$$

$$RIBES_{\text{общ}}(\alpha = 0) = \frac{RIBES_1 + RIBES_2}{2} = \frac{1+1}{2} = 1 \quad (15)$$

$$RIBES_{\text{общ}}(\alpha = 0.5) = \frac{0{,}85+0{,}95}{2} = 0{,}9 \quad (16)$$

$$RIBES_{\text{общ}}(\alpha = 1) = \frac{0{,}73+0{,}91}{2} = 0{,}82 \quad (17)$$

В талб. 3. представлен расчёт метрик *RIBES* для стеготекста с заменой синонимов и изменением порядка слов.

Таблица 3.

Слово предложения	Первое предложение	Второе предложение
Нормализованный коэффициент Спирмана	1	0,98
Нормализованный коэффициент Кендалла	0,98	0,95

Слово предложения	Первое предложение	Второе предложение
Точность	$\frac{11}{15}$ = 0,73	$\frac{21}{23}$ = 0,91
RIBES (при α = 0)	99%	97%
RIBES (при α = 0,5)	85%	92%
RIBES (при α = 1)	73%	88%
Общее значение *RIBES* (при α = 0)	98%	
Общее значение *RIBES* (при α = 0,5)	89%	
Общее значение *RIBES* (при α = 1)	81%	

$$RIBES_1(\alpha = 0) = \frac{NSR_1 * P^{\alpha} + NST_1 * P^{\alpha}}{2} = \frac{1*0{,}73^0 + 0{,}98*0{,}73^0}{2} = 0{,}99 \quad (18)$$

$$RIBES_2(\alpha = 0) = \frac{NSR_2 * P^{\alpha} + NST_2 * P^{\alpha}}{2} = \frac{0{,}98*0{,}91^0 + 0{,}95*0{,}91^0}{2} = 0{,}97 \quad (19)$$

$$RIBES_1(\alpha = 0{,}5) = \frac{1*0{,}73^{0{,}5} + 0{,}98*0{,}73^{0{,}5}}{2} = 0{,}85 \quad (20)$$

$$RIBES_2(\alpha = 0{,}5) = \frac{0{,}98*0{,}91^{0{,}5} + 0{,}95*0{,}91^{0{,}5}}{2} = 0{,}92 \quad (21)$$

$$RIBES_1(\alpha = 1) = \frac{1*0{,}73^1 + 0{,}98*0{,}73^1}{2} = 0{,}73 \quad (22)$$

$$RIBES_2(\alpha = 1) = \frac{0{,}98*0{,}91^1 + 0{,}95*0{,}91^1}{2} = 0{,}88 \quad (23)$$

$$RIBES_{\text{общ}}(\alpha = 0) = \frac{RIBES_1 + RIBES_2}{2} = \frac{0{,}99+0{,}97}{2} = 0{,}98 \quad (24)$$

$$RIBES_{\text{общ}}(\alpha = 0{,}5) = \frac{0{,}85+0{,}92}{2} = 0{,}89 \quad (25)$$

$$RIBES_{\text{общ}}(\alpha = 1) = \frac{0{,}73+0{,}88}{2} = 0{,}81 \quad (26)$$

При расчёте значений для текста со случайными изменениями была получена схожая точность (поскольку часть слов оригинального текста не заменялась), лишь с небольшим уменьшением по значению. Но при этом в связи со случайными изменениями в порядке слов сильно уменьшились значения нормализованных коэффициентов, в связи с чем снизилось и само значение метрики. Это показывает, что основную функцию определения несоответствий выполняют именно коэффициенты ранговой корреляции. Расчёты метрики *RIBES* основаны на формулах (27-35).

В табл. 4 представлен расчёт метрик *RIBES* для стеготекста со случайными заменами слов и их порядка.

Таблица 4.

Слово предложения	Первое предложение	Второе предложение
Нормализованный коэффициент Спирмана	0,74	0,75
Нормализованный коэффициент Кендалла	0,69	0,75
Точность	$\frac{11}{15}$=0,73	$\frac{17}{23}$=0,74
RIBES (при α = 0)	72%	75%

Слово предложения	Первое предложение	Второе предложение
RIBES (при α = 0,5)	61%	65%
RIBES (при α = 1)	52%	56%
Общее значение *RIBES* (при α = 0)	74%	
Общее значение *RIBES* (при α = 0,5)	63%	
Общее значение *RIBES* (при α = 1)	54%	

$$RIBES_1(\alpha = 0) = \frac{NSR_1 * P^{\alpha} + NST_1 * P^{\alpha}}{2} = \frac{0{,}74 * 0{,}73^0 + 0{,}69 * 0{,}73^0}{2} = 0{,}72 \quad (27)$$

$$RIBES_2(\alpha = 0) = \frac{NSR_2 * P^{\alpha} + NST_2 * P^{\alpha}}{2} = \frac{0{,}75 * 0{,}74^0 + 0{,}75 * 0{,}74^0}{2} = 0{,}75 \quad (28)$$

$$RIBES_1(\alpha = 0{,}5) = \frac{0{,}74 * 0{,}73^{0{,}5} + 0{,}69 * 0{,}73^{0{,}5}}{2} = 0{,}61 \quad (29)$$

$$RIBES_2(\alpha = 0{,}5) = \frac{0{,}75 * 0{,}74^{0{,}5} + 0{,}75 * 0{,}74^{0{,}5}}{2} = 0{,}65 \quad (30)$$

$$RIBES_1(\alpha = 1) = \frac{0{,}74 * 0{,}73^1 + 0{,}69 * 0{,}73^1}{2} = 0{,}52 \quad (31)$$

$$RIBES_2(\alpha = 1) = \frac{0{,}75 * 0{,}74^1 + 0{,}75 * 0{,}74^1}{2} = 0{,}56 \quad (32)$$

$$RIBES_{\text{общ}}(\alpha = 0) = \frac{RIBES_1 + RIBES_2}{2} = \frac{0{,}72 + 0{,}75}{2} = 0{,}74 \quad (33)$$

$$RIBES_{\text{общ}}(\alpha = 0{,}5) = \frac{0{,}61 + 0{,}65}{2} = 0{,}63 \quad (34)$$

$$RIBES_{\text{общ}}(\alpha = 1) = \frac{0{,}52 + 0{,}56}{2} = 0{,}54 \quad (35)$$

В результате исследования было выяснено, что метрика *RIBES* является эффективной для использования в оценке лингвистических стегосистем при определённом значении коэффицента учёта точности. Метрика *RIBES* при коэффициенте учёта точности α равного 0 не учитывает изменения слов, поэтому показывает более высокие оценки, в том числе для текста со случайными преобразованиями. Учёт точности также недостаточен при коэффициенте 0,5. Однако для коэффициента равного 1 метрика *RIBES* показывает наиболее точные оценки. Оценки для стеготекстов меняются незначительно для разных стегопреобразований, при этом текст со случайными преобразованиями получает достаточно низкую оценку.

Литература

1. Zachary M. Ziegler, Yuntian Deng, Alexander M. Rush., 2019. Neural Linguistic Steganography.
2. Ching − Yun Chang, Stephen Clark, 2012, The Secret's in the Word Order: Text-to-Text Generation for Linguistic Steganography.
3. Ching − Yun Chang, Stephen Clark, 2014, Practical Linguistic Steganography using Contextual Synonym Substitution and a Novel Vertex Coding Method.
4. Ахрамеева К.А., Герлинг Е.Ю., Мицковский Д.Ю., Прудников С.В. Использование метрики BLEU для оценки естественности текста лингвистических стегосистем // Вестник Российского нового университета, серия «Сложные системы: модели, анализ и управление», 2020. – 5 с.
5. Krzysztof Wołk. Machine Learning in Translation Corpora Processing, 2019. – 264 p.

6. Коржик В.И., Небаева К.А., Герлинг Е.Ю., Догиль П.С., Федянин И.А. Цифровая стеганография и цифровые водяные знаки. Часть 1. Цифровая стеганография. [монография]: – СПбГУТ. – СПб., 2016. – 226 с.
7. Красов А.В., Штеренберг С.И., Фахрутдинов Р.М., Рыжаков Д.В., Пестов И.Е. Анализ информационной безопасности предприятия на основе сбора данных пользователей с открытых ресурсов и мониторинга информационных ресурсов с использованием машинного обучения // T-Comm: Телекоммуникации и транспорт, 2018. – Т. 12. – № 10. – С. 36-40.
8. Коржик В.И., Нгуен З.К., Годлевский А.К. Оценка стегоключей для стегосистем, использующих стойкое шифрование вложенных сообщений // Проблемы информационной безопасности. Компьютерные системы, 2018. – № 3. – С. 26-36.
9. Сахаров Д.В., Левин М.В., Фостач Е.С., Виткова Л.А. Исследование механизмов обеспечения защищенного доступа к данным, размещенным в облачной инфраструктуре // Наукоемкие технологии в космических исследованиях Земли, 2017. – Т. 9. – № 2. – С. 40-46.
10. Герлинг Е.Ю., Ахрамеева К.А. Метод Лингвистической стеганографии, основанный на опорном слове // I-methods, 2019. – Т. 11. – № 4. – С. 1-9.
11. Герлинг Е.Ю. Исследование эффективности методов обнаружения стегосистем, использующих широкополосное вложение // Телекоммуникации, 2014. – № 1. – С. 6-12.