УДК 003.27

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ ИСПОЛЬЗОВАНИЯ МЕТРИК ОЦЕНКИ МАШИННОГО ПЕРЕВОДА ДЛЯ ОЦЕНКИ ЛИНГВИСТИЧЕСКИХ СТЕГОСИСТЕМ С ТОЧКИ ЗРЕНИЯ БЕЗОПАСНОСТИ

Мицковский Д.Ю. Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики

Научный руководитель – PhD Аксёнов В.Е.

Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики

В докладе представлены основные результаты исследования возможности использования метрик оценки машинного перевода для оценки лингвистических стегосистем Исследуемые метрики предполагается использовать для обучения стегосистемы.

Введение. Лингвистическая стеганография основана на маскировке конфиденциальных данных с помощью лингвистического порядка элементов текста в покрывающих данных (например, на основе замен пар синонимов, на основе изменения порядка слов и т.д.). Такие стегопреобразования сложнее реализовать в программе, поскольку из-за того, что алгоритмы не могут понимать смысл слов, генерируемые стеготексты зачастую представляют из себя нечитаемые сообщения, которые злоумышленник-посредник может довольно легко опознать как обёртки для стеганограмм. Для создания автоматической стегосистемы нужно решить проблему автоматической генерации текста. Перспективным решением может оказаться использование RNN-сетей, которые уже используются для генерации текстов. Однако, для обучения RNN сети нужна определённая метрика ошибки, которая могла бы достоверно оценить качество генерируемого стеготекста. Перспективными для использования являются метрики оценки машинного перевода, которые сравнивают машинный перевод с существующими текстовыми данными.

Основная часть.

Метрика BLEU рассчитывается согласно количеству совпадений n-грамм между стеготекстом и эталонным текстом, составленным человеком. Чем выше значение метрики, тем лучше текст. По метрике BLEU текст оценивается по шкале от 0 до 1. Чем ближе к единице, тем больше совпадений наблюдается с их человеческими референтными образцами текста и, таким образом, тем лучше система.

Метрика WER рассчитывается как число ошибок, деленное на общее количество слов. Получение WER начинается со сложения замен, вставок и удалений в стеготексте относительно эталона, которые происходят в последовательности распознанных слов. Затем это число делится на общее количество первоначально использованных слов. Чем выше значение метрики, тем хуже текст. Метрика дает более высокие оценки последовательным совпадающим словам.

Метрика TER рассчитывается как количество правок, которые необходимо внести в стеготекст, чтобы тот точно соответствовал эталонному тексту. Правки могут включать удаление, вставку, замену, а также, в отличии от WER, изменение порядка слов. Чем больше правок необходимо, тем больше значение метрики, а соответственно хуже оцениваемый текст.

Метрика METEOR рассчитывается согласно совпадениям n-грамм стеготекста и эталона, при этом в расчёте метрики учитывает размер совпадающих n-грамм. Для этого стеготекст делится на группы n-грамм, совпадающих с группами n-грамм в эталоне. Чем больше длина совпадающих n-грамм, тем больше значение метрики METEOR, т.е. лучше текст.

Метрика RIBES фокусируется на порядке слов. В этом случае используются коэффициенты ранговой корреляции, основанные на порядке слов, для сравнения машинного и эталонного текстов. Чем выше значение метрики, тем лучше текст. Для вычисления метрики используются коэффициенты ранговой корреляции, а каждому слову в тексте присваивается

числовой ранг – соответствующий порядку расположения слова в тексте. Основными коэффициентами ранговой корреляции являются ρ – коэффициент Спирмена, который измеряет расстояние насколько порядковый номер слова в стеготексте отличается от порядкового номера слова в эталоне, и τ – коэффициент Кендалла, который указывает на то, в каком направлении порядок слов стеготекста отличается от эталона – в отрицательном или в положительном. Ранговые меры могут быть нормализованы для обеспечения положительных значений. Для учёта изменений слов в метрике также учитывается точность. С учётом нормализованных ранговых коэффициентов и точности метрика вычисляется по формуле, где α-коэффициент учёта точности в диапазоне $0 < \alpha < 1$.

Для исследования изменений значений метрик оценки использовался эталонный текст, тексты с заменой синонимов и порядка слов без сильного изменения по смыслу. Также для проверки работы метрик был использован эталон с внесением случайных замен слов и порядка, при сохранении некоторых частей текста эталона.

По полученным значениям метрики BLEU можно увидеть, что метрика показывает корректные значения для текста с заменой синонимов, однако её значение сильно уменьшается при изменении порядка слов в тексте. Это происходит из-за того, что при изменении порядка слов число совпадающих биграмм и триграмм значительно снижается, а вместе с тем снижается и значение метрики. Аналогичная ситуация проявляется и при совмещении преобразований. При этом текст со случайными преобразованиями оценён корректно.

Согласно результатам, метрика WER имеет аналогичную проблему, поскольку изменение порядка слов вызывает значительные изменения в структуре текста относительно эталонного текста, при этом метрика WER учитывает изменение порядка слова относительно эталона как удаление слова со старой позиции и добавления на новую (то есть изменение порядка слова считается как два изменения относительно эталона).

В отличии от WER значение метрики TER практически не влияет изменение порядка слов, поскольку метрикой учитывается изменение порядка слов и на каждое изменение порядка слов необходимо одно исправление.

Метрика METEOR имеет наименьшие различия в оценках для разных стегосистем, поскольку метрика учитывает не только замены слов с помощью учёта точности и полноты, но также и порядок слов с помощью учёта совпавших сочетаний слов. Однако метрика в большей степени учитывает совпадения словосочетаний, чем отдельных слов, поэтому при этом метрика более высоко оценивает текст со случайными изменениями, поскольку из-за нахождения совпадений словосочетаний оценка метрики увеличилась.

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. Оценки для стеготекстов меняются незначительно для разных стегопреобразований, при этом текст со случайными преобразованиями получает достаточно низкую оценку.

Выводы. В результате исследования наилучшие результаты показали метриками TER и RIBES при коэффициенте учёта точности 1. При изменении стегопреобразований оценка стеготекстов менялась менее значительно, чем для других метрик, при этом текст со случайными преобразованиями был оценён низко. Метрики BLEU и WER также показали хорошие результаты при оценке стеготекстов с заменой синонимов, однако для использования при оценивании стегосистем с изменением порядка не рекомендуются. Метрика METEOR показала худший результат среди прочих метрик из-за более высокой оценки текста со случайными преобразованиями, что является недопустимым.

Мицковский Д.Ю. (автор) Подпись

Аксенов В.Е. (научный руководитель) Подпись