

Нейронный машинный перевод в аспекте качества перевода

Научный руководитель – Коканова Елена Сергеевна

Бурянина Мария Васильевна

Студент (бакалавр)

Северный (Арктический) федеральный университет имени М.В. Ломоносова, Высшая школа социально-гуманитарных наук и международной коммуникации, Архангельск, Россия

E-mail: buryaninamaria@yandex.ru

Настоящее исследование направлено на анализ оценки качества машинного перевода с использованием искусственных нейронных сетей (нейронного машинного перевода, НМП) на основе реальных примеров сопоставления переводов человека и НМП.

Сегодня существуют три разновидности машинного перевода (МП): на основе правил, статистический, нейронный. Введение НМП позволило улучшить качество МП и приблизиться к эталонному переводу, выполненному человеком [3].

Для анализа был использован отрывок статьи *Arctic* из англоязычной энциклопедии «Британника», а также его переводы на русский язык, выполненные автором работы и НМП Amazon translate. На основе сопоставления делается вывод о возможности замены человека машиной. Выбор типа текста обоснован наличием в нем только когнитивной информации, то есть НМП может успешно справиться с переводом.

Единой методики оценки качества перевода нет. К автоматическим оценкам относятся метрики BLEU, F-measure, TER. Они основаны на соответствии переведенных слов и длины предложений [2]. В силу различия строя русского и английского языков, показатели автоматических метрик могут быть занижены.

Для проведения анализа используется усовершенствованный вариант шкалы, изначально предложенный в Chartered Institute of Linguists (GB) [1]. Оценка качества письменного перевода определяется по следующим аспектам:

1. Понимание, точность и стиль.
2. Грамматика, связность текста, логичность изложения.
3. Технические аспекты: пунктуация, правописание, передача имен и наименований, дат, чисел.

За каждый аспект можно получить от 1 до 10 баллов. Перевод считается приемлемым, если набрано не менее 12 баллов (по 4 за каждый аспект). Собственный перевод, согласно шкале, оценивается в 24 балла (по 8 за каждый аспект).

Обращаясь к первому аспекту, мы можем отметить, что в переводе НМП информация передана верно; используемая лексика соответствует тематике ИТ и в некоторых случаях совпадает с переводом человека: *hydrocarbons* - углеводороды, *petroliferous province* - нефтеносная провинция, *continental shelf* - континентальный шельф. Общий процент совпадений - 40%. Слов в тексте собственного перевода - 358, в тексте НМП - 311.

Однако в нескольких случаях НМП не учитывает контекст и допускает некоторые неточности. В частности:

ИТ: *There is also gold, tin, nickel, copper <...>. For the first four, the north provides...*

Собственный перевод: *Также имеются запасы золота, олова, никеля, меди <...>. Север страны представляет собой <...> самый крупный источник первых четырех пород.*

НМП: *Есть также золото, олово, никель, медь <...>. Для первых четырех стран север <...> является крупнейшим источником в стране.*

В ИТ существительное после числительного *four* не названо, но из предыдущих предложений мы понимаем, что речь идет о перечисленных горных породах. Вместо словосочетания *горные породы* используется существительное *страны*. Ошибка связана с эллипсисом, основной проблемой для МП.

Особое внимание хочется обратить на наличие буквализмов в НМП:

ИТ: *there could be enormous expansion if it is required and thought desirable.*

Собственный перевод: ... *однако при возникновении необходимости и желания, объемы их добычи можно было бы увеличить.*

НМП: ... *однако, если это необходимо и считается желательным, может быть огромное расширение.*

Существительное *expansion* в тексте НМП передано прямым соответствием *расширение*. Это делает ПТ буквальным и не совсем понятным для реципиента. В ПТ, выполненном человеком, использован прием экспликации, что способствует более точной передаче информации.

Указанные выше ошибки не мешают читателю воспринимать информацию, указанную в тексте. Поэтому НМП по первому аспекту оценивается в 4 балла.

Применительно ко второму аспекту структура предложений текста НМП вполне приемлема, но в некоторых случаях наблюдаются сохранение структуры английских предложений (буквальный перевод) и необоснованное опущение сказуемого. Рассмотрим пример перевода предложения со сказуемым *contain*:

ИТ: *Northwestern Siberia contains a petroliferous province. . .*

Собственный перевод: *На северо-западе Сибири была обнаружена нефтеносная провинция. . .*

НМП: *Северо-Западная Сибирь содержит нефтеносную провинцию. . .*

В переводе НМП наблюдается нарушение лексической сочетаемости. Добавим, что в ПТ следовало изменить структуру предложения.

В переводах двух предложений, выполненных НМП, опущено сказуемое. Пример одного из них:

ИТ: *Further large discoveries are likely.*

Собственный перевод: *Возможны дальнейшие разработки новых месторождений.*

НМП: *Скорее всего, дальнейшие крупные открытия.*

Из-за этого смысл предложений становится не совсем понятным. В остальных случаях структура изложения логична и ясна. Принимая во внимание некоторые нарушения языковой нормы и необоснованные опущения в тексте НМП, мы оцениваем перевод в 4 балла по второму аспекту.

Проводя оценку технических параметров текста, мы можем отметить, что ошибок в пунктуации, передаче дат и числовой информации в ПТ нет. Однако имеются две грубые ошибки при переводе топонимов. Первая связана с тем, что при передаче названия субъекта РФ из него было опущено нарицательное имя в тексте НМП: *Саха* вместо *Республика Саха*. Вторая ошибка связана с передачей на русский язык месторождения, находящегося на Аляске (*Alaskan field*). В тексте НМП было выбрано несуществующее слово *Аляскасский*. Машина использовала неверную модель образования прилагательных. Верным был бы вариант *Аляскинский*. Собственный вариант перевода – *месторождение Аляски*.

По третьему аспекту мы оцениваем МП также на 4 балла. Это связано с наличием ошибок в передаче топонимов. Отметим также, что при оценке качества МП возможно не такое жесткое следование критериям, как при оценке перевода, выполненного человеком. Причем допущенные ошибки не введут читателя в заблуждение. Как видно, НМП оценивается в 12 баллов из 30 возможных.

Проведенное нами исследование позволяет сделать вывод о том, что качество перевода энциклопедического текста НМП можно назвать удовлетворительным. Однако для использования в профессиональной сфере он должен пройти процесс постредактирования человеком. Интересным представляется оценка качества перевода разножанровых текстов НМП.

Источники и литература

- 1) Убин И.И., Пушнов И.А. Перевод в Европейском Сообществе: объем перевода и оценка качества перевода. Обзор/ Под ред. И.И. Убина – М., 2011.
- 2) Shterionov D., Way A. Human versus quality evaluation of NMT and PBSMT// Machine Translation, 2018, Volume 32, Issue 3, P. 217-235.
- 3) Нейронный перевод: новый тренд и его перспективы: <https://help.smartcat.ai/hc/ru/articles/360017711291-16-03-2017>.