PBML

Пражский вестник математической лингвистики

НОМЕР 108 ИЮНЬ 2017 109–120

Нейронный машинный перевод - это новое слово в искусстве?

Шейла Кастильо,[а] Джосс Муркенс,[а] Федерико Гаспари,[а] Ясер Каликсто,[а] Джон Тинсли,[б] Энди Уэй[а]

[а] Центр ADAPT, Дублинский городской университет
[б] Культовые переводческие машины

Абстрактный

В этой статье обсуждается нейронный машинный перевод (NMT), новая парадигма в области машинного перевода, сравнивается качество систем NMT со статистическим машинным переводом, описываются три исследования с использованием автоматических методов и методов оценки, проводимых человеком. Результаты автоматической оценки, представленные для NMT, очень многообещающие, однако человеческие оценки показывают неоднозначные результаты. Мы сообщаем о повышении беглости речи, но о непоследовательных результатах с точки зрения адекватности и усилий после редактирования. NMT, несомненно, представляет собой шаг вперед в области машинного перевода, но сообществу следует быть осторожным, чтобы не перепродать его.

1. Введение

С момента его создания различные теории и практики машинного перевода (МП) приходили и уходили, и каждая новая волна вызывала огромное волнение и ожидание в этой области. Тем не менее, от первых коммерческих систем, основанных на правилах, до более поздних статистических моделей, как правило, было большое несоответствие между высокими ожиданиями того, что должна выполнять МП, и тем, что она в действительности может дать. Совсем недавно нейронный подход (NMT) появился как новая парадигма в системах машинного перевода, вызвав интерес в академических кругах и промышленности, превзойдя статистические системы на основе фраз (PBSMT), основанные в основном на впечатляющих результатах автоматической оценки (Bahdanau et al., 2015; Sennrich et al., 2016; Bojar et al., 2016). Но превосходят ли результаты NMT результаты SMT при использовании человеческой оценки? Можем ли мы на данном этапе утверждать, что NMT - это новая современная парадигма для производства? В этой статье обсуждается качество систем NMT по сравнению с современным SMT.

 Автор, ответственный за переписку: sheila.castilho@adaptcentre.ie
Цитируйте как: Шейла Кастильо, Джосс Муркенс, Федерико Гаспари, Ясер Каликсто, Джон Тинсли, Энди Уэй. Нейронный машинный перевод - это новое слово в искусстве ?. Пражский вестник математической лингвистики № 108, 2017, стр. 109-120. Doi: 10.1515 / pralin-2017-0013.

систем, сообщив о трех случаях использования, в которых оценщики сравнивали результаты NMT и SMT для ряда языковых пар. Основываясь на полученных данных, мы утверждаем, что даже несмотря на то, что NMT демонстрирует значительные улучшения для некоторых языковых пар и конкретных областей, все еще есть много возможностей для исследований и улучшений, прежде чем можно будет сделать широкие обобщения.

Остальная часть статьи организована следующим образом: в разделе 2 мы проводим обзор существующей литературы, касающейся систем NMT. В разделе 3 мы описываем три случая использования, в которых системы NMT сравнивались с системами SMT и проводилась оценка человеком: в разделе 3.1 представлено исследование с использованием изображений для машинного перевода списков товаров электронной коммерции, созданных пользователями, с двумя системами NMT и одной системой SMT для Англо-немецкая языковая пара; В разделе 3.2 описывается небольшая человеческая оценка с упором на патентную область для китайского языка, а в разделе 3.3 описывается крупномасштабная человеческая оценка для области MOOC с учетом переводов с английского на четыре целевых языка (немецкий, греческий, португальский и русский).). Наконец, в разделе 4 мы обсуждаем основные выводы из вариантов использования, подробно останавливаясь на том, как оценивался NMT,

2. Расцвет моделей нейронного машинного перевода.

Нейронные модели включают построение сквозной нейронной сети, которая отображает выровненные двуязычные тексты, которые при заданном вводном предложении Икс для перевода, обычно обучается для максимизации вероятности целевой последовательности Y без дополнительной внешней лингвистической информации. В последнее время всплеск интереса к NMT вызван применением глубоких нейронных сетей (DNN) для создания сквозных*кодировщик-декодер* модели (Kalchbrenner, Blunsom, 2013; Sutskever et al., 2014; Cho et al., 2014). Bahdanau et al. (2015) впервые представили механизм внимания в структуре кодировщика-декодера NMT, который обучен обращать внимание на соответствующие слова исходного языка при генерации каждого слова целевого предложения. Некоторые важные недавние разработки в NMT включают улучшение механизма внимания, включая лингвистическую информацию или включение большего количества языков в модель (Luong et al., 2015; Sennrich and Haddow, 2016)

Сообщалось об улучшениях NMT по сравнению с системами PBSMT в общих задачах, где NMT занимал первое место над системами SMT в шести из 12 языковых пар для задач перевода (Bojar et al., 2016). Кроме того, для задачи автоматического постредактирования было обнаружено, что нейронные сквозные системы представляют собой «значительный шаг вперед» по сравнению с базовым статистическим подходом. В других недавних исследованиях сообщалось о повышении качества при сравнении NMT с SMT с использованием автоматических показателей (Bahdanau et al., 2015; Jean et al., 2015) или небольших человеческих оценок (Bentivogli et al., 2016; Wu et al. , 2016). Wu et al. (2016) сообщают, что их система NMT превосходит подходы SMT (для английского на испанский, французский, упрощенный китайский и обратно), особенно для морфологически богатых языков, с впечатляющими оценками людей. Bentivogli et al.

по сравнению с наиболее производительной системой SMT, с меньшим количеством ошибок порядка слов, лексических и морфологических ошибок, что позволяет сделать вывод о том, что NMT «значительно продвинул вперед современное состояние», особенно для морфологически богатых языков.

Торал и Санчес-Картахена (2017) сравнивают NMT и PBSMT для девяти языковых пар (с английского на чешский, немецкий, румынский, русский и с английского на финский) с двигателями, обученными для данных новостных тестов WMT. Для вывода NMT получены лучшие результаты автоматической оценки, чем для вывода PBSMT для всех языковых пар, кроме русско-английского и румынско-английского. Повышенное изменение порядка в системах NMT приводит к тому, что системы NMT работают лучше, чем SMT, в отношении ошибок перестановки и изменения порядка во всех языковых парах. Однако они также сообщают, что SMT, по-видимому, работает лучше, чем NMT для сегментов длиной более 40 слов, при применении метрики автоматической оценки chrF1 (Popović, 2015).

Этот обзор недавней работы предполагает, что NMT значительно улучшил эту область, особенно если принять во внимание современные автоматические метрики оценки. Однако прогресс не всегда очевиден. В разделе 3 представлены три случая использования, в которых NMT сравнивался с SMT и оценивался с помощью человеческих оценок. Выясняется, что в зависимости от различных доменов и различных исследуемых языковых пар NMT не всегда дает наилучшие результаты.

3. Сценарии использования

Каждый вариант использования ориентирован на отдельный домен и охватывает разный набор языковых пар. Во-первых, в разделе 3.1 рассматривается протокол NMT для электронной коммерции, описываются важные части более обширного исследования, о котором подробно сообщается в Calixto et al. (2017b). Второй вариант использования (раздел 3.2) - это оценка, выполненная Iconic Translation Machines Ltd.[1], целью которого было выяснить, может ли NMT обеспечить лучший перевод патентной области, чем SMT. Наконец, третий и последний вариант использования (обсуждаемый в разделе 3.3) - это сравнение, проведенное в рамках проекта TraMOOC, финансируемого ЕС, на данных, взятых из массовых открытых онлайн-курсов (MOOC) на английском языке.

3.1. NMT для листинга товаров электронной коммерции

Обычный вариант использования в электронной коммерции заключается в использовании MT для обеспечения максимально широкого доступа к описаниям продуктов, отзывам пользователей и комментариям (например, на специализированных форумах), независимо от родного языка или страны происхождения клиентов. В предыдущей работе Calixto et al. (2017a) сравнили качество переводов списков продуктов, полученных с помощью мультимодальной модели NMT, с двумя подходами, основанными только на тексте: традиционной NMT, основанной на внимании, и моделью PBSMT. Переводы оценивались с использованием автоматических показателей, а также посредством качественной оценки, конечной целью которой было проверить, улучшает ли обучение системы NMT с доступом к изображениям продуктов качество вывода для переводов с английского на немецкий.

1 http://iconictranslation.com/

MT Systems - В этом эксперименте сравнивались три разные системы (1) а*ПБСМТ* базовая модель, построенная с помощью Moses SMT Toolkit (Koehn et al., 2007), (2) текстовая модель NMT (*Не более*$_T$) и (3) мультимодальная модель NMT (*Не более*$_M$), более подробно описанный в Calixto et al. (2017b), который расширяет модель, ориентированную только на текст, и вводит*визуальный компонент* включить *местный* визуальные особенности.

Набор данных состоит из списков продуктов и изображений с 23, 697 обучающие кортежи, каждый из которых содержит *(я)* список товаров на английском языке, *(II)* список продуктов на немецком языке, и *(iii)* изображение продукта. Наборы для валидации и тестирования имеют480 а также 444 кортежи соответственно. Следует учитывать, что перевод списков товаров, созданных пользователями, сопряжен с особыми проблемами, например, потому, что они часто не грамматичны и их трудно интерпретировать даже для носителя языка. В частности, в списках на обоих языках есть много разрозненных ключевых слов и / или фраз, склеенных вместе, а также несколько опечаток. Все это сложности, которые делают мультимодальное машинное программирование списков продуктов сложной задачей, поскольку существует множество трудностей, связанных с обработкой списков и изображений.

Оценка - Для качественной оценки людей двуязычным носителям немецкого языка было предложено (1) *оценить мультимодальную адекватность* переводов (количество участников N = 18); и (2)*классифицировать* переводы разных моделей от лучших к худшим (количество участников N = 18). Для *мультимодальная оценка адекватности* участникам был представлен список продуктов на английском языке, изображение продукта и перевод, созданный одной из моделей, не зная, какая модель. Затем их спросили, какая часть значения источника также выражена в переводе с учетом изображения продукта с использованием 4-балльной шкалы Лайкерта (где*4 = Ничего подобного* а также *1 = Все это*). Для*рейтинг* При оценке участникам были представлены изображение продукта и три перевода, полученные с разных моделей для определенного списка продуктов на английском языке (без указания моделей), и их попросили расположить переводы от лучшего к худшему.

Автоматическая оценка проводилась с использованием четырех широко распространенных автоматических показателей MT: BLEU4, METEOR (Denkowski and Lavie, 2014), TER (Snover et al., 2006) и chrF3.

Полученные результаты - В таблице 1 некоторые автоматические метрики сравниваются с оценками адекватности переводов, сделанными людьми.

Модель	BLEU4↑	METEOP↑	TER↓	chrF3↑	Достаточность↓
Не более$_T$	22,5	40,0	58,0	56,7	2,71 ± .48
Не более$_M$	25,1$_†$	42,6$_†$	**55,5**$_†$	58,6	**2.36** ± .47
ПБСМТ	**27,4**$_{†‡}$	**45,8**$_{†‡}$	**55,4**$_†$	**61,6**	**2.36** ± .47

Таблица 1. Адекватность переводов и четыре автоматических метрики для списков продуктов и изображений. По первым трем показателям результаты значительно лучше, чем у NMT$_T$ (†) или NMT$_M$ (‡) с участием p <0,01.

с двумя базовыми линиями только для текста, PBSMT и NMT$_T$, и одна мультимодальная модель NMT$_M$.

Модель PBSMT превосходит обе модели NMT в соответствии с BLEU, METEOR и chrF3. Однако разницы нет - разногласия между NMT$_M$ модель и PBSMT в соответствии с оценками TER.

Кроме того, оценки адекватности для обеих этих моделей, NMT_T и PBSMT, в среднем совпадают в соответствии с оценками, полученными при оценке людей.

Тем не менее, несмотря на то, что обе модели дают одинаково адекватный результат, переводы, полученные с помощью PBSMT, оцениваются людьми лучше, чем 56,3% времени, а переводы, полученные с помощью мультимодальной модели NMT_M считаются лучшими 24,8% времени. Эти результаты предполагают, что, хотя модели NMT иногда могут достигать автоматических показателей MT PBSMT, в соответствии с этим вариантом использования оценщики не предпочитают их.

3.2. NMT для патентной области

Оценка, представленная в этом разделе, была основана на совместном проекте, выполненном группой машинного перевода в Центре ADAPT, Городским университетом Дублина (Ирландия) и Iconic Translation Machines Ltd. (Iconic), коммерческим поставщиком машинного перевода, базирующимся в Дублине (Ирландия). . Iconic разрабатывает для своих пользователей специализированные движки машинного перевода, часто обращаясь к языковым парам и типам контента, которые создают большие проблемы для машинного перевода. Одной из таких комбинаций, пользующихся особым спросом, является китайская патентная информация для перевода на английский язык, при этом в 2016 году было переведено более 100 миллионов слов машинным переводом.

Цель этой оценки состояла в том, чтобы сравнить характеристики зрелых китайских и английских патентных двигателей MT, используемых в производстве Iconic, с новыми двигателями NMT, разработанными в Центре ADAPT по принципу «от яблок к яблокам», обученным на тех же доступных данных.

Областью оценки были названия и рефераты химических патентов (см. Таблицу 2). Этот тип контента имеет особые характеристики, которые создают проблемы для машинного перевода, в том числе очень технический контент со специальной терминологией, названиями химических компонентов, а также буквенно-цифровыми и аминокислотными последовательностями. Заголовки и аннотации раздела сами патенты весьма различны: заголовки короткие, в среднем 8,2 жетона, и написаны в стандартном телеграфном стиле; рефераты обычно содержат от 2 до 6 довольно длинных предложений со средней длиной 42,5 лексемы.

Описание	**Пары предложений**	**Слова (источник)**
Химические рефераты	1 076 894	50 198 888
Химические названия	350 840	2 868 121
Общий патент	11 931 127	324 222 969
Глоссарии	1,575	1,575
Общий	**13 358 861**	**377 291 553**

Таблица 2. Использование тренировочных данных для Iconic и NMT двигателестроение

MT Systems - Механизмы Iconic MT основаны на запатентованной архитектуре Ensemble. TM который сочетает в себе элементы машинного перевода на основе фраз, синтаксиса и правил машинного перевода, а также автоматическое постредактирование. Механизмы были тщательно настроены в течение ряда лет для патентной области, с использованием нескольких различных моделей перевода и языков, а также включения терминологии, специфичной для контента.

Движки ADAPT / Iconic NMT были реализованы с использованием моделей, основанных на внимании, построенных с помощью Nematus.[2] с использованием различных комбинаций данных (учитывая, что домены немного отличаются, используются все данные, т. е. только данные внутри домена, а также данные внутри домена плюс различные части более общих данных, выбранных с помощью выбора данных). Мы также настроили различные наборы для разработки заголовков и аннотаций. Четыре лучших движка разработки w для оценки. Обе системы были обучены с использованием одних и тех же данных, которые включали смесь очень специфичных для контента данных в предметной области, более общих патентных данных (включая поддомен по химии) и технических глоссариев.

Оценка - Двигатели оценивались отдельно по их характеристикам в заголовках и аннотациях, с двумя разными наборами тестов, каждый из которых содержал 1123 сегмента. Была проведена стандартная автоматическая оценка, и баллы BLEU представлены в таблице 3. Оценка на людях также проводилась для сравнения характеристик двух двигателей. Два рецензента оценили 100 случайно выбранных сегментов из вышеупомянутых наборов тестов двумя способами: слепым ранжированием лучшего перевода (с учетом ссылки) и анализом ошибок для определения основной ошибки перевода в данном сегменте. Таксономия ошибок состояла из знаков препинания, части речи, пропусков, дополнений, неправильной терминологии, дословного перевода и словоформы. Сегменты были выбраны случайным образом из тестового набора, так что 25% сегментов были короткими предложениями (т. Е. Они содержали <10 слов), 25% были длинными предложениями (т. Е.

Результаты автоматической оценки показывают, что NMT немного превосходит SMT по заголовкам, тогда как система SMT превосходит NMT по рефератам. Что касается человеческой оценки, в целом система SMT была признана «лучшей» в 54% случаев по сравнению с 39% для NMT. При рассмотрении длины предложения система SMT была оценена как «лучшая» в 84% случаев для коротких предложений по сравнению с только 8% для системы NMT; и был лучшим в 58% случаев для длинных предложений (> 40 токенов) по сравнению с 33% для NMT. Система NMT была оценена как «лучшая» чаще, чем система SMT, только для предложений средней длины (> 10 <40 слов), с 57% предпочтений против 36% для SMT.

Полученные результаты - Типы ошибок, обнаруженные в выходных данных NMT, были высокими для пропусков (37% ошибок, обнаруженных в сегментах, против 8% для Система SMT), тогда как для SMT ошибки заключались в структуре предложений (35% сегментов против 10% для системы NMT).

Для сегментов без ошибок было обнаружено, что 25% сегментов из системы SMT не содержат ошибок, по сравнению только с 2% сегментов из системы NMT. Эти результаты снова показывают, что система NMT превосходит систему SMT в отношении автоматических показателей (для заголовков), но человеческая оценка по-прежнему предпочитает систему SMT.

Система	Титулы (BLEU)	Аннотации (BLEU)
Iconic MT	31,99	***28,32***
Нейронный MT	***37,52***	13,39

Таблица 3. Результаты автоматической оценки MT для названий химических патентов и рефераты.

2 https://github.com/rsennrich/nematus

3.3. NMT для домена MOOC

Оценка, представленная в этом разделе, была проведена в рамках проекта TraMOOC (Перевод для массовых открытых онлайн-курсов), финансируемого ЕС.[3], который является совместным проектом Horizon 2020, направленным на обеспечение надежного машинного перевода для MOOK. PB-SMT и система NMT сравнивались по четырем направлениям перевода (т.е. с английского (EN) на немецкий (DE), греческий (EL), португальский (PT) и русский (RU)) в серии обширных оценочных заданий. Целью этого сравнения было решить, какая система обеспечит более качественные переводы для предметной области проекта.

MT Systems - В качестве SMT на основе фраз использовался Moses, а системы NMT - это сети кодировщиков-декодеров внимания, которые были обучены с помощью Nematus. Механизмы MT были обучены на больших объемах обучающих данных из различных источников: обучающих данных WMT.[4] и OPUS[5], TED от WIT3[6], Корпус образовательных областей QCRI (QED)[7], корпус MOOK Coursera и собственный сборник образовательных данных. Количество используемых обучающих данных показано в таблице 4.

Поскольку эта оценка была предназначена для определения наиболее эффективной системы машинного перевода для перевода MOOC, наборы тестов были извлечены из реальных данных MOOC (одна тысяча английских сегментов - для задачи ранжирования использовалась всего сто сегментов). Эти данные включали пояснительные тексты, субтитры из видеолекций или пользовательский контент (UGC) на студенческих форумах или в разделах комментариев ресурсов электронного обучения.

Данные пользовательского контента часто были плохо сформулированы и содержали частые грамматические ошибки. В других текстах использовалась более стандартная грамматика и синтаксис, но содержалась специализированная терминология и неконтекстные переменные и формулы.

Язык перевода	DE	EL	PT	RU
Вне домена	23,78	30,73	31,97	21,30
В домене	0,14	0,58	2,31	0,27

Таблица 4. Размер обучающих данных для обучения машин MT для EN→* направление перевода (количество пар предложений, в миллионы).

Оценка - Для оценки использовались автоматические метрики (BLEU, METEOR и HTER (Snover et al., 2006)), а также выполнялась оценка на людях. Оценка человеком проводилась профессиональными переводчиками (три для EL, PT и RU, и два для DE) и состояла из: i) постредактирования (PE) вывода MT для достижения пригодного для публикации качества в окончательном отредактированном тексте, ii) рейтинг беглости и адекватности (т. е.

3 http://tramooc.eu/

4 http://www.statmt.org/wmt16/

5 http://opus.lingfil.uu.se/

6 http://www.clg.ox.ac.uk/tedcorpus

7 http://alt.qcri.org/resources/qedcorpus/

степень, в которой целевой сегмент отражает значение исходного сегмента) по 4-балльной шкале Лайкерта для каждого сегмента, и iii) выполнение аннотации ошибок с использованием простой таксономии (которая включала: флективную морфологию, порядок слов, пропуск, добавление, и неправильное ранжирование).

Полученные результаты - Автоматическая оценка (см. Таблицу 3) показала, что NMT превзошел SMT по показателям BLEU и METEOR для немецкого, греческого и русского языков (статистически значимо при парном сравнении одностороннего ANOVA ($p < 0,05$)).

По португальскому языку можно наблюдать только умеренные улучшения. Оценки HTER показывают, что при использовании выходных данных системы SMT для всех целевых языков требовалось больше PE (статистически не значимо). Эти результаты показывают, что при вмешательстве человека учтено (постредактирование), прирост с NMT был менее стабильным.

Lang.	Система	BLEU	METEOR	HTER	Fluency	Адекватность
DE	SMT	41,5	33,6	49,0	2,60	2,85
	Не более	61,2 †	42,7 †	32,2	2,95	2,79
EL	SMT	47,0	35,8	45,1	2,86	3,44
	Не более	56,6 †	40,1 †	38,0	3,08	3,46
PT	SMT	57,0	41,6	33,4	3,15	3,73
	Не более	59,9	43,4	31,6	3,22	3,79
RU	SMT	41,9	33,7	44,6	2,70	2,98
	Не более	57,3 †	40,65 †	33,9	3,08	3,12

Таблица 5. Результаты автоматической оценки (статистически значимые результаты, отмеченные знаком †), беглость и адекватность

Оценка человека - Что касается человеческой оценки *Беглость*Хотя не было обнаружено статистически значимых различий, NMT был оценен как более беглый, чем SMT для всех языковых пар (Таблица 5). Результаты для*достаточность* были менее последовательными, с более высокими средними баллами для немецкого SMT. Эти результаты показывают, что по мере того, как NMT становится понятнее, при оценке того, какая часть смысла, выраженного в источнике, появляется в переводе, SMT немного лучше или равен NMT.

Взяв во внимание *аннотация ошибки* задача, общее количество проблем, выявленных в выходных данных, было больше для SMT, чем для NMT для всех языковых пар.

Более того, количество сегментов без ошибок было больше для NMT во всех языковых парах. Также было обнаружено, что выходные данные NMT содержат меньше ошибок порядка слов и ошибок флективной морфологии на всех целевых языках. Однако вывод SMT содержал меньше ошибок пропуска, добавления или неправильного перевода для EN-EL, чем вывод NMT; он также показал меньше пропусков, чем система NMT для EN-PT, в то время как EN-RU SMT показал меньше ошибок перевода, чем система NMT. Интересно, что для немецкого языка ошибки флективной морфологии составляют 49% от всех

Lang.	Системные	технические усилие	временные усилие	WPS
DE	SMT	5,8	74,8	0,21
	Не более	3.9	72,8	0,22
EL	SMT	13,9	77,7	0,22
	Не более	12,5	70,4	0,24
PT	SMT	3.8	57,7	0,29
	Не более	3,6	55,19	0,30
RU	SMT	7,5	104,6	0,14
	Не более	7.2	105,6	0,14

Таблица 6. Технические (нажатия клавиш / сегмент) и временные усилия после редактирования (секунды / сегмент) и слов в секунду (WPS)

ошибок, обнаруженных в выводе NMT, больше, чем в SMT (где флективная морфология составляет 43% ошибок). С уважением к*постредактирование*результаты показывают, что меньшее количество сегментов NMT, по мнению участников, требует редактирования (но со статистической значимостью только для немецкого языка ($p < 0,05$, где $M = 0,06$, $SE = 0,04$)). Средняя пропускная способность или временные усилия (Таблица 6) были лишь незначительно улучшены для постредактирования на немецком, греческом и португальском языках с помощью NMT, в то время как временные затраты для англо-русского языка были ниже для SMT на уровне сегмента. Эти результаты также воспроизводятся в словах в секунду (WPS).

Технические усилия по постредактированию были сокращены для NMT во всех языковых парах за счет измерения фактических нажатий клавиш (Таблица 6) или минимального количества правок, необходимых для перехода от предварительно отредактированного текста к постредактированному (HTER в Таблице 5). Отзывы участников показали, что им труднее идентифицировать ошибки NMT, тогда как ошибки порядка слов и несоответствия, требующие исправления, обнаруживались быстрее в выводе SMT.

Наконец, что касается *рейтинг* Во всех языковых парах участники оценивали результаты NMT, особенно английский-немецкий. 53% предпочитали NMT для коротких сегментов (20 токенов или меньше) и 61% предпочли NMT для длинных сегментов (более 20 токенов). В заключение, для рассматриваемых языковых пар (EN-DE, EN-EL, EN-PT и EN-RU) и для конкретного домена MOOC была улучшена беглость речи и уменьшены ошибки порядка слов при использовании NMT. При использовании NMT требуется меньшее количество сегментов, требующих постредактирования, особенно из-за меньшего количества морфологических ошибок. Однако не было явного улучшения в отношении ошибок пропуска и неправильного перевода при сравнении SMT и NMT. Также не было значительного снижения усилий после редактирования,

4. Обсуждение и заключение

NMT вызвал большой ажиотаж, особенно потому, что индустрия переводов стремится улучшить качество машинного перевода, чтобы минимизировать затраты (Moorkens, 2017). Хотя при сравнении NMT с другими парадигмами машинного перевода с использованием автоматических показателей сообщаются многообещающие результаты, когда к сравнению добавляется человеческая оценка, результаты еще не столь однозначны. Мы попытались проиллюстрировать это утверждение на трех примерах использования, сравнивающих NMT с системами SMT, где оценка также выполнялась людьми.

Результаты, представленные в разделе 3.1 для переводов списков продуктов, показывают, что модели NMT действительно очень многообещающие, особенно с учетом того, что современная система PBMST была развернута уже довольно давно, в то время как модели NMT - особенно мультимодальная система NMT - были разработаны в течение более короткого периода времени. Однако система PBSMT по-прежнему обеспечивает лучший перевод при оценке как с помощью автоматических показателей, так и показателей оценки, проводимых человеком. Тот же результат можно наблюдать в разделе 3.2, где модели NMT быстро приближаются к автоматическим оценкам SMT.

в течение нескольких месяцев после развертывания патентной области. Важно отметить, что для обоих вариантов использования 3.1 и 3.2 обучающие данные - это те же обучающие данные, которые используются в их повседневной работе, что делает их реальными результатами.

Наконец, обширная человеческая оценка, описанная в разделе 3.3 для домена MOOC, показывает, что NMT хорошо работает с точки зрения автоматических показателей (кроме португальского, где улучшение незначительно), но несовместимо с адекватностью и усилиями после редактирования. Несмотря на то, что нейронная модель демонстрирует повышение беглости речи, она также показывает большее количество ошибок упущения, добавления и неправильного перевода. Решение перейти на модель NMT в качестве предпочтительной системы машинного перевода для проекта TraMOOC подтверждает, что нейронные модели очень многообещающие, хотя на их разработку тратится мало времени по сравнению с давно существующими системами PBSMT.

Хотя результаты автоматической оценки, опубликованные для NMT, несомненно, впечатляют, до сих пор может показаться, что NMT не полностью достиг качества SMT, основываясь на оценке человека. Мы считаем, что к шумихе, созданной в области машинного перевода с появлением нейронных моделей, следует относиться осторожно. Перепродажа технологии, которая все еще нуждается в дополнительных исследованиях, может негативно повлиять на MT, как уже было замечено ранее с системами SMT (особенно с выпуском свободно доступного инструментария Moses в 2006 году, который упростил каждому обучать свой собственный MT. system), когда утверждалось, что МП производит переводы «почти человеческого качества» и что МП «крадет рабочие места переводчиков», делая переводчиков «просто постредакторами МП».*против* их тип противостояния.

NMT, без сомнения, представляет собой шаг вперед в области машинного перевода. Однако есть также ограничения для нейронных моделей, которые нельзя игнорировать и которые все еще необходимо решить. На наш взгляд, на данном этапе исследователи и отрасль должны проявлять осторожность, чтобы не обещать слишком многого, и позволять проводить дополнительные исследования для устранения ограничений NMT и проводить более обширные человеческие оценки, обращаясь к как можно большему количеству типов текста, областей и языка. пары по возможности.

Благодарности

Проект TraMOOC получил финансирование в рамках исследовательской и инновационной программы Европейского Союза Horizon 2020 в рамках грантового соглашения № 644333. Центр ADAPT по технологиям цифрового контента при Городском университете Дублина финансируется в рамках Программы исследовательских центров Ирландии Научного фонда (грант 13 / RC / 2106).) и софинансируется Европейским фондом регионального развития.

Библиография

Богданау, Дмитрий, Кёнхён Чо и Йошуа Бенжио. Нейронный машинный перевод Совместное обучение выравниванию и переводу. В*Международная конференция по обучающимся представительствам, ICLR 2015*, Сан-Диего, Калифорния, 2015.

Бентивольи, Луиза, Арианна Бизацца, Мауро Четтоло и Марчелло Федерико. Нейронная система против Качество фразового машинного перевода: пример из практики. *CoRR*, абс / 1608.04631, 2016. URL http://arxiv.org/abs/1608.04631.

Бояр, Ондржей, Райен Чаттерджи, Кристиан Федерманн, Иветт Грэм, Барри Хаддоу, Маттиас Гек, Антонио Химено Йепес, Филипп Коэн, Варвара Логачева, Кристоф Монц, Маттео Негри, Орели Невеол, Мариана Невес, Мартин Попел, Мэтт Пост, Рафаэль Рубино, Каролина Скартон, Люсия Специа, Марко Турчи, Карин Верспур и Маркос Зампиери. Результаты конференции по машинному переводу 2016 года. В*Труды Первой конференции по машинному переводу*, страницы 131–198, Берлин, Германия, август 2016 г. Ассоциация компьютерной лингвистики.

Каликсто, Ясер, Даниэль Штайн, Евгений Матусов, Пинту Лохар, Шейла Кастильо и Энди Уэй. Нас- Использование изображений для улучшения машинного перевода списков товаров для электронной коммерции. В*Труды 15-й конференции Европейского отделения Ассоциации компьютерной лингвистики: Том 2, Краткие статьи*, страницы 637–643, Валенсия, Испания, 2017a. URLhttp: //www.aclweb. org / anthology / E17-2101.

Каликсто, Ясер, Цюнь Лю и Ник Кэмпбелл. Дважды внимательный декодер для мультимодальных Нейронный машинный перевод. В*Труды 55-го Ежегодного собрания Ассоциации компьютерной лингвистики - Том 1*, Ванкувер, Канада (статья принята), 2017b. URL https://arxiv.org/abs/1702.01287.

Чо, Кёнхён, Барт ван Мерриенбоер, Чаглар Гульчере, Дмитрий Богданау, Фетхи Бугарес, Хольгер Швенк и Йошуа Бенжио. Изучение представлений фраз с использованием RNN Encoder – Decoder для статистического машинного перевода. В*Материалы конференции 2014 г. по эмпирическим методам обработки естественного языка (EMNLP)*, страницы 1724–1734, Доха, Катар, 2014. URL http://www.aclweb.org/anthology/D14-1179.

Денковски, Майкл и Алон Лави. Meteor Universal: оценка языкового перевода Использование для любого целевого языка. В*Материалы девятого семинара по статистическому машинному переводу*, страницы 376–380, Балтимор, Мэриленд, США, 2014. URL http://www.aclweb.org/ антология / W14-3348.

Жан, Себастьен, Кёнхён Чо, Роланд Мемишевич и Йошуа Бенжио. При использовании очень больших Целевой словарь для нейронного машинного перевода. В*Труды 53-го ежегодного собрания Ассоциации компьютерной лингвистики и 7-й совместной международной конференции по обработке естественного языка (Том 1: Длинные статьи)*, страницы 1–10, Пекин, Китай, 2015. URL http://www.aclweb.org/anthology/P15-1001.

Кальчбреннер, Нал и Фил Блансом. Рекуррентные модели непрерывного перевода. В*Труды конференции 2013 г. по эмпирическим методам обработки естественного языка, EMNLP 2013*, страницы 1700–1709, Сиэтл, октябрь 2013 г.

Коэн, Филипп, Хиеу Хоанг, Александра Берч, Крис Каллисон-Берч, Марчелло Федерико, Никола Бертольди, Брук Коуэн, Уэйд Шен, Кристин Моран, Ричард Зенс, Крис Дайер, Ондржей Бояр, Александра Константин и Эван Хербст. Моисей: набор инструментов с открытым исходным кодом для статистического машинного перевода. В*Материалы демонстрационной и стендовой сессий ACL-2007*, страницы 177–180, Прага, Чешская Республика, 2007. Ассоциация компьютерной лингвистики.

Луонг, Тханг, Хиеу Фам и Кристофер Д. Мэннинг. Эффективные подходы к вниманию- на основе нейронного машинного перевода. В*Материалы конференции по эмпирическим методам 2015 г.*

в обработке естественного языка (EMNLP), страницы 1412-1421, Лиссабон, Португалия, 2015. ISBN 978-1-941643-32-7.

Муркенс, Джосс. Под давлением: перевод во времена жесткой экономии.*Перспективы*, 25 (3): 1-14, 2017. DOI: 10.1080 / 0907676X.2017.1285331. URLhttp://dx.doi.org/10.1080/0907676X.2017.1285331.

Попович, Майя. chrF: n-грамм символа F-оценка для автоматической оценки MT. В*Труды Десятый семинар по статистическому машинному переводу*, страницы 392-395, Лиссабон, Португалия, сентябрь 2015 г.

Сеннрих, Рико и Барри Хэддоу. Функции лингвистического ввода улучшают передачу нейронных машин. lation. В*Труды Первой конференции по машинному переводу*, страницы 83-91, Берлин, Германия, август 2016 г.

Сеннрих, Рико, Барри Хэддоу и Александра Берч. Эдинбургский нейронный машинный перевод Системы для WMT 16. В *Труды Первой конференции по машинному переводу (WMT16)*, Берлин, Германия, 2016.

Сновер, Мэтью, Бонни Дорр, Ричард Шварц, Линнеа Миччулла и Джон Махул. А изучение скорости редактирования переводов с целевой человеческой аннотацией. В*Труды ассоциации машинного перевода в Северной и Южной Америке*, том 200 (6), 2006.

Суцкевер, Илья, Ориол Виньялс, Куок В Ле. Последовательность для последовательного обучения с помощью нейронных сетей Сети. В*Достижения в системах обработки нейронной информации*, страницы 3104-3112, Монреаль, Канада, 2014 г.

Тораль, Антонио и Виктор М. Санчес-Картахена. Многогранная оценка нейронной вер- Sus Phrase-Based Machine Translation для 9 языковых направлений. В*Труды 15-й конференции Европейского отделения Ассоциации компьютерной лингвистики: Том 1, Длинные статьи*, страницы 1063-1073, Валенсия, Испания, апрель 2017 г. Association for Computational Лингвистика. URLhttp://www.aclweb.org/anthology/E17-1100.

Ву, Юнхуэй, Майк Шустер, Чжифэн Чен, Куок В. Ле, Мохаммад Норузи, Вольфганг Машери, Максим Крикун, Юань Цао, Цинь Гао, Клаус Машери, Джефф Клингнер, Апурва Шах, Мелвин Джонсон, Сяобин Лю, Лукаш Кайзер, Стефан Гоус, Йошикиё Като, Таку Кудо, Хидето Казава, Кейт Стивенс, Джорджент Куриан, Ниша Вэй Ван, Клифф Янг, Джейсон Смит, Джейсон Риза, Алекс Рудник, Ориол Виньялс, Грег Коррадо, Макдуф Хьюз и Джеффри Дин. Система нейронного машинного перевода Google: устранение разрыва между человеческим и машинным переводом.*CoRR*, абс / 1609.08144, 2016.

Адрес для корреспонденции:
Шейла Кастильо
sheila.castilho@adaptcentre.ie
Центр ADAPT, Дублинский городской университет