

УДК 81'322.4

DOI 10.22213/2618-9763-2024-4-106-116

Е. С. Ошанова, кандидат филологических наук, доцент

Ижевский государственный технический университет имени М. Т. Калашникова, Ижевск, Россия

Д. А. Поносова

МБОУ «Средняя общеобразовательная школа № 8», Ижевск, Россия

АНАЛИЗ СИСТЕМ НЕЙРОННОГО МАШИННОГО ПЕРЕВОДА С ПОМОЩЬЮ МЕТРИК АВТОМАТИЗИРОВАННОЙ ОЦЕНКИ

В статье представлены современные метрики автоматизированной оценки качества нейронного машинного перевода и рассмотрены два способа оценки результатов, которые включают в себя оценку при помощи специальных метрик автоматизированной оценки, а также оценку переводчиком-экспертом. Для проведения анализа и оценки протестированы системы нейронного машинного перевода. В качестве кандидатов в работе представлены системы Google Translate и DeepL Translate, которые применяют в работе нейросетевой подход. В качестве метрик для оценки рассмотрены: METEOR как традиционная референсная метрика, СОМЕТ как нейросетевая референсная метрика и СОМЕТ-Kiwi как нейросетевая безреференсная метрика. Необходимо указать, что на современном этапе нейросетевые метрики лучше всего соотносятся с человеческими оценками качества машинного перевода. Отмечается, что даже модели с нейросетевым подходом обучаются на данных, предоставленных человеком, т. е. на сегодняшний день невозможно избавиться от эталонов или оценок качества, выполненных экспертами. Данные метрики используют для упрощения оценки качества, они позволяют лучше исследовать машинный перевод, его особенности и предельные возможности.

В рамках анализа был выбран фрагмент исходного текста, выполнен его перевод на целевые языки при помощи выбранных систем нейронного машинного перевода для получения переводов-кандидатов, а затем указан для каждого из них текст-эталон.

Результаты оценки при помощи метрик дали возможность оценить, насколько близок машинный перевод к переводу, выполненному человеком, а также выяснить, на каком этапе развития находятся современные системы машинного перевода. Экспертная оценка помогла понять, насколько эффективно системы справляются с переводом.

Ключевые слова: нейронный машинный перевод; метрики автоматизированной оценки качества; системы нейронного машинного перевода; эталонный перевод.

Введение

Из-за скорости развития технологий ученые и исследователи со всего мира имеют больше возможностей для проведения исследований в области машинного перевода. Это приводит к тому, что количество новой информации для перевода растет, а следовательно, переводчикам необходимы современные инструменты, которые позволят ускорить темп их работы, а также повысить качество переводимых текстов. Одной из таких технологий является нейронный машинный перевод (НМП). Объектом для анализа в данной работе является нейронный машинный перевод как один из подходов к машинному переводу.

Целью исследования является определение эффективности применения систем ней-

ронного машинного перевода на основе метрик автоматизированной оценки качества.

Для достижения поставленной цели необходимо решить ряд задач:

1. Рассмотреть нейросетевой подход к машинному переводу.
2. Рассмотреть существующие метрики автоматизированной оценки.
3. Проанализировать результаты автоматизированной оценки НМП.

Нейронный машинный перевод

В первую очередь необходимо представить определение для ключевого понятия «нейронный машинный перевод».

Машинный перевод претерпевал огромное количество изменений в ходе своего существования. С развитием технологий

у ученых появлялось больше возможностей, и появление новых идей не заставляло себя ждать. Одной из них стала идея об искусственном интеллекте (ИИ), представленная миру еще в 1956 г. на семинаре в Дартмутском колледже в США. В 2016 г. компания *Google* представила систему, которая использовала нейросетевой подход к машинному переводу – *Google Neural Machine Translation system (GNMT)* [1]. Улучшенная система с повышенной точностью перевода теперь позволяла переводить предложение целиком, без разбивания его на сегменты. В 2017 г. немецкой компанией *DeepL GmbH* миру был представлен новейший онлайн-переводчик на основе нейронных сетей *DeepL Translate*.

Следует отметить труды отечественных и зарубежных исследователей в области общей и компьютерной лингвистики, в частности по теории машинного перевода, таких как И. А. Улиткин [2], М. Сновер (*M. Snover*) [3], Д. Богданау (*D. Bahdanau*) [4].

Стоит указать, что ученые при исследовании систем машинного перевода указывали на их недостатки, отмечая медлительность, отсутствие точности и дороговизну [5], а также невозможность реализации знаний об окружающем нас мире через данные системы [6, с. 173]. Несмотря на это ряд исследователей подчеркивает и перспективность подобных разработок. Так, в своей работе 1992 г. А. Гросс продемонстрировал, что с общими переводами, требующими знания реального мира, лучше справляется человек, а тексты с математическими и абстрактными понятиями наиболее качественно переводятся системами машинного перевода [7, с. 103].

Система машинного перевода (МП) с нейросетевым подходом значительно отличается от системы, основанной на правилах, которой для выполнения перевода необходимо разбить входное высказывание на фрагменты и сопоставить их с фрагментами на целевом языке. Задача системы с нейросетевым подходом – обработать все входное высказывание в процессе перевода целиком. В системах НМП на вход принимается некоторая совокупность элементов (слова, предложения), которые обрабатываются ею

последовательно и на выходе образуют ту же совокупность элементов уже на целевом языке. Таким образом, одним из преимуществ применения нейронных сетей можно назвать способность к тщательному изучению связей между языками и их структурой, а также способность к сохранению контекста [8]. Отметим, что в базе систем НМП лежат словари с редкими словами и открытые словари, которые содержат числа, имена и т. д.

Машинный перевод, а в особенности НМП, на данном этапе развития не всегда справляется с задачей правильного подбора необходимого варианта, а потому для оценки результатов такого вида перевода были созданы специальные метрики автоматизированной оценки. Отметим, что автоматическая оценка – прежде всего сравнение переведенного текста с неким эталонным переводом. Эталоном может послужить перевод, выполненный человеком, а также результат машинного перевода, отредактированный экспертом. По определению О. В. Митрениной, метрика – это способ числовой оценки различий между результатом МП и эталонном [9].

Таким образом, системы машинного перевода прошли долгий путь развития: от прямого перевода к интерлингвальному подходу, который так и остался только на уровне идеи [10], затем к трансферному подходу, основанному на правилах, и к переводу с использованием предварительного редактирования текста [11], далее – к статистическим системам машинного перевода и, наконец, к нейронному машинному переводу [12].

Метрики автоматизированной оценки

В качестве примера рассмотрим три разные метрики.

METEOR (Metric for Evaluation of Translation with Explicit Ordering) – является одной из традиционных референсных метрик качества машинного перевода. Помимо применения *n*-граммного метода, оценивание посредством данной метрики основывается на подсчете совпадений слов, которые называют «униграммами». Совпадениями считаются не только идентичные слова, но

и наличие однокоренных слов или синонимов. Итоговая оценка включает в себя штрафы за краткость и за отсутствие совпадений в n -граммах [13].

Улучшенная *референсная нейросетевая метрика COMET*. Интересна эта метрика тем, что она работает на базе мультязычной модели *XLM-RoBERTa*, которая, в отличие от других моделей, принимает на вход не только эталон и перевод-кандидат, но и исходный текст. Модель метрики обучается на входных данных, а также на оценках экспертов-переводчиков [14].

В настоящее время в области изучения искусственного интеллекта было выделено направление, которое называется «обработка естественного языка» (*Natural Language Processing, NLP*). В *NLP* существует метод, который позволяет оценивать качество перевода в отсутствие эталонного перевода, выполненного человеком. Называется данный метод *Quality Estimation (QE)*. Смысл его заключается в том, что специальные языковые модели «дообучают» для того, чтобы они могли сравнивать и оценивать перевод-кандидат непосредственно с первоисточником на исходном языке. Обучаются модели на специальных наборах данных (*dataset*), содержащих оценки перевода, выполненные как профессиональными переводчиками, так и рядовыми пользователями [15].

Один из подходов к оценке переводов – это *MQM (Multidimensional Quality Metrics)*, где оценка выполняется переводчиками-экспертами по специально заданным критериям. Эти данные применяют для обучения *безреференсных нейросетевых метрик*. К такой метрике относится *COMET-Kiwi*, которая объединяет в себе две модели *QE*: одна из них обучена на оценках перевода человеком, а также на наборе данных *MLQE-PE (Multilingual Quality Estimation and Post-Editing Dataset)* и выполняет одновременное преобразование МП и первоисточника; вторая модель обучена на данных *MQM* и не использует для оценки референс. Итоговая объединенная оценка получается путем сравнения извлеченных данных при помощи специальных настраиваемых параметров [16].

Анализ и результаты

В работе представим анализ и оценку результатов нейронного машинного перевода фрагментов оригинального текста произведения *Physics of the Future* Митио Каку на русский и немецкий языки при помощи специальных метрик автоматизированной оценки, а также самостоятельной оценки результатов.

В рамках анализа был выбран небольшой фрагмент исходного текста, состоящий из 10 предложений. Данный фрагмент был разбит на сегменты по предложениям, чтобы упростить процесс его обработки метриками, произведена проверка на соответствие структур предложений исходного текста и эталонов для более точных результатов. Так, например, выявлено, что переводчицей на немецкий язык был применен прием членения предложения, из-за чего при сегментации эти два члененных предложения были выделены в один сегмент.

Исходный текст

“(1) Another great prophet of technology was Leonardo da Vinci, painter, thinker, and visionary. (2) In the late 1400s, he drew beautiful, accurate diagrams of machines that would one day fill the skies: sketches of parachutes, helicopters, hang gliders, and even airplanes. (3) Remarkably, many of his inventions would have flown. (4) His flying machines, however, needed one more ingredient: at least a 1-horsepower motor, something that would not be available for another 400 years.

(5) What is equally astonishing is that Leonardo sketched the blueprint for a mechanical adding machine, which was perhaps 150 years ahead of its time. (6) In 1967, a misplaced manuscript was reanalyzed, revealing his idea for an adding machine with thirteen digital wheels. (7) If one turned a crank, the gears inside turned in sequence performing the arithmetic calculations. (8) The machine was built in 1968 and it worked.

(9) In addition, in the 1950s another manuscript was uncovered which contained a sketch for a warrior automaton, wearing German-Italian armor, that could sit up and move its arms, neck, and jaw. (10) It, too, was subsequently built and found to work” [17, с. 21].

Перевод-эталон (переводчик Наталья Лисова, рус. яз.)

«(1) Еще одним великим пророком техники был Леонардо да Винчи, художник, мыслитель и мечтатель.

(2) Он оставил нам красивые и точные чертежи машин, которые должны были когда-нибудь подняться в небо: наброски парашютов, вертолетов, дельтапланов и даже самолетов. (3) Замечательно, что многие из его изобретений действительно полетели бы, хотя придуманы в конце XV в. (4) Его летающим машинам не хватало одного единственного ингредиента: мотора мощностью по крайней мере в одну лошадиную силу, которому предстояло появиться на свет лишь через 400 лет.

(5) Не менее поразителен тот факт, что Леонардо оставил нам набросок механического суммирующего устройства, обогнавшего время примерно на 150 лет. (6) В затерявшейся рукописи, вновь открытой в 1967 г., был обнаружен проект суммирующей машины с цифровыми колесиками на тринадцать разрядов. (7) При повороте ручки колесики внутри должны были прокручиваться и производить арифметические вычисления. (8) Машина, построенная по этому чертежу в 1968 г., на самом деле работала.

(9) Кроме того, в 1950-х гг. была обнаружена еще одна рукопись Леонардо с наброском робота-воина в итало-германских доспехах, способного садиться, двигать руками, шеей и челюстью. (10) Робот по этому проекту также был построен и оказался вполне функциональным» [18, с. 6–7].

Перевод-кандидат *Google Translate* (рус. яз.)

«(1) Другим великим пророком технологий был Леонардо да Винчи, художник, мыслитель и визионер. (2) В конце 1400-х годов он нарисовал прекрасные, точные схемы машин, которые однажды заполнят небо: эскизы парашютов, вертолетов, дельтапланов и даже самолетов. (3) Примечательно, что многие из его изобретений летали. (4) Однако его летательным аппаратам требовался еще один ингредиент: по край-

ней мере двигатель мощностью в 1 лошадиную силу, чего не будет еще 400 лет.

(5) Не менее удивительно то, что Леонардо набросал чертеж механической арифмометрической машины, которая, возможно, опередила свое время на 150 лет. (6) В 1967 году была повторно проанализирована утерянная рукопись, раскрывшая его идею арифмометра с тринадцатью цифровыми колесами. (7) Если повернуть рукоятку, шестеренки внутри последовательно вращались, выполняя арифметические вычисления. (8) Машина была построена в 1968 году, и она работала.

(9) Кроме того, в 1950-х годах была обнаружена еще одна рукопись, которая содержала эскиз воина-автомата, одетого в немецко-итальянские доспехи, который мог сидеть и двигать руками, шеей и челюстью. (10) Она также была впоследствии построена и оказалась работоспособной».

Перевод-кандидат *DeepL Translate* (рус. яз.)

«(1) Другим великим пророком технологий был Леонардо да Винчи, художник, мыслитель и провидец. (2) В конце 1400-х годов он нарисовал красивые, точные схемы машин, которые однажды заполнят небо: эскизы парашютов, вертолетов, дельтапланов и даже самолетов. (3) Примечательно, что многие из его изобретений могли бы летать. (4) Однако его летающим машинам требовался еще один ингредиент: мотор мощностью не менее 1 л. с., который не будет доступен еще 400 лет.

(5) Не менее удивительно и то, что Леонардо набросал чертеж механической суммирующей машины, которая, возможно, опередила свое время на 150 лет. (6) В 1967 году была заново проанализирована потерянная рукопись, в которой обнаружилась его идея о суммирующей машине с тринадцатью цифровыми колесами. (7) Если повернуть кривошип, шестеренки внутри последовательно вращались, выполняя арифметические вычисления. (8) Машина была построена в 1968 году и работала.

(9) Кроме того, в 1950-х годах была обнаружена еще одна рукопись, содержащая

¹ Источник табл. 1 – URL: <https://huggingface.co/spaces/evaluate-metric/meteor> (дата обращения: 20.10.2024).

эскиз автомата-воина в немецко-итальянских доспехах, который мог сидеть и двигать руками, шеей и челюстью. (10) Впоследствии он тоже был построен и оказался работоспособным».

Перевод-эталон (переводчик Моника Нихаус, нем. яз.)

“(1) Ein anderer großer Prophet der technischen Entwicklung war Leonardo da Vinci, Maler, Denker und Visionär.

(2) Gegen Ende des 15. Jahrhunderts fertigte er wunderbare, präzise Zeichnungen von Maschinen an, die eines Tages den Himmel füllen würden: Skizzen von Fallschirmen, Helikoptern, Hängegleitern und sogar Flugzeugen.

(3) Bemerkenswerterweise wären viele seiner Erfindungen tatsächlich geflogen.

(4) Seinen Flugapparaten fehlte jedoch eine weitere Komponente: zumindest ein 1-PS-Motor, aber ein solcher Antrieb sollte noch weitere 400 Jahre Zukunftsmusik bleiben.

(5) Ebenso erstaunlich ist, dass Leonardo eine mechanische Addiermaschine entwarf, die ihrer Zeit um rund 150 Jahre voraus war.

(6) Im Jahr 1967 findet man eine Sammlung von Skizzenblättern und Texten von Leonardo da Vinci, den sogenannten Codex Madrid. Der Codex enthält eine Skizze, die als Basis für eine Addiermaschine mit 13 Rädern interpretiert wurde.

(7) Wenn man eine Kurbel drehte, drehten sich die Zahnräder in Folge und führten arithmetische Berechnungen durch.

(8) Leonardos Maschine wurde 1968 von IBM nachgebaut und ausgestellt.

(9) Zudem wurde in den 1950er Jahren ein anderes Manuskript mit der Skizze eines automatischen Kriegers in deutsch-italienischer Uniform entdeckt, der sich aufsetzen und Arme, Hals und Kiefer bewegen konnte. (10) Auch dieser Automat wurde später gebaut und erwies sich als funktionsfähig” [19, c. 16–17].

Перевод-кандидат *Google Translate* (нем. яз.)

“(1) Ein weiterer großer Prophet der Technologie war Leonardo da Vinci, Maler, Denker und Visionär. (2) Ende des 15. Jahrhunderts zeichnete er wunderschöne, genaue Diagramme von Maschinen, die eines Tages den Himmel

bevölkern würden: Skizzen von Fallschirmen, Hubschraubern, Drachenfliegern und sogar Flugzeugen. (3) Bemerkenswerterweise wären viele seiner Erfindungen geflogen. (4) Seine Flugmaschinen benötigten jedoch noch eine weitere Zutat: mindestens einen 1-PS-Motor, etwas, das erst 400 Jahre später verfügbar sein würde.

(5) Ebenso erstaunlich ist, dass Leonardo den Bauplan für eine mechanische Rechenmaschine skizzierte, die ihrer Zeit vielleicht 150 Jahre voraus war. (6) 1967 wurde ein verlegtes Manuskript erneut analysiert und enthüllte seine Idee einer Rechenmaschine mit dreizehn Ziffernrädern. (7) Wenn man eine Kurbel drehte, drehten sich die Zahnräder im Inneren nacheinander und führten die arithmetischen Berechnungen aus. (8) Die Maschine wurde 1968 gebaut und funktionierte.

(9) Darüber hinaus wurde in den 1950er Jahren ein weiteres Manuskript entdeckt, das eine Skizze für einen Kriegerautomaten in deutsch-italienischer Rüstung enthielt, der aufrecht sitzen und seine Arme, seinen Hals und seinen Kiefer bewegen konnte. (10) Auch dieser wurde später gebaut und funktionierte”.

Перевод-кандидат *DeepL Translate* (нем. яз.)

“(1) Ein weiterer großer Prophet der Technik war Leonardo da Vinci, Maler, Denker und Visionär. (2) Ende des 14. Jahrhunderts zeichnete er wunderschöne, präzise Diagramme von Maschinen, die eines Tages den Himmel bevölkern würden: Skizzen von Fallschirmen, Hubschraubern, Hängegleitern und sogar Flugzeugen. (3) Bemerkenswerterweise wären viele seiner Erfindungen auch geflogen. (4) Seine Flugmaschinen benötigten jedoch noch eine weitere Zutat: einen Motor mit mindestens 1 PS, der erst in 400 Jahren verfügbar sein würde.

(5) Erstaunlich ist auch, dass Leonardo den Bauplan für eine mechanische Rechenmaschine entwarf, die ihrer Zeit vielleicht 150 Jahre voraus war. (6) Im Jahr 1967 wurde ein verlegtes Manuskript wieder ausgewertet, das seine Idee für eine Rechenmaschine mit dreizehn digitalen Rädern enthüllte. (7) Wenn man eine Kurbel drehte, drehten sich die Zahnräder im Inneren nacheinander und führten die arithmetischen Berechnungen durch. (8) Die Maschine wurde 1968 gebaut, und sie funktionierte.

(9) Außerdem wurde in den 1950er Jahren ein weiteres Manuskript entdeckt, das eine Skizze für einen Kriegerautomaten in einer deutsch-italienischen Rüstung enthielt, der sich aufsetzen und seine Arme, seinen Hals und seinen Kiefer bewegen konnte. (10) Auch er wurde später gebaut und erwies sich als funktionstüchtig”.

Далее, необходимо выполнить оценку результатов НМП при помощи метрик:

Таблица 1. Показатели референсной метрики *METEOR*¹

Table 1. *METEOR* Reference Metrics Indicators

Сегмент	<i>Google (en-ru)</i>	<i>DeepL (en-ru)</i>	<i>Google (en-de)</i>	<i>DeepL (en-de)</i>
(1)	0,6938	0,6938	0,7991	0,7991
(2)	0,4351	0,4352	0,6355	0,6579
(3)	0,3507	0,3458	0,8676	0,8413
(4)	0,3596	0,2424	0,2643	0,1684
(5)	0,4099	0,4078	0,7603	0,6471
(6)	0,1687	0,1571	0,0453	0,1780
(7)	0,2165	0,2166	0,7962	0,8589
(8)	0,1603	0,1616	0,4261	0,4213
(9)	0,6101	0,6516	0,5905	0,6414
(10)	0,0934	0,3709	0,5647	0,7282
Среднее значение	0,3498	0,3683	0,5749	0,5941

Как можно заметить, среднее значение оценки у языковой пары «английский – русский» довольно низкое, что может быть связано с тем, что структура предложений данных языков отличается. Кроме того, профессиональные переводчики часто применяют различные переводческие приемы и трансформации, тогда как системы МП зачастую выполняют дословный перевод, и к тому же могут столкнуться с такой проблемой, как отсутствие эквивалента слова или выражения в параллельных данных языковой пары. Это может привести к тому, что система начнет самостоятельно придумывать новые слова для сохранения целостности предложения, а это, в свою очередь, станет очередной ошибкой для метрики в процессе оценки. Можно отметить, что перевод с английского языка на немецкий получил общую оценку выше среднего, что связано с тем, что немецкое предложение имеет фиксированную структуру, а значит, кан-

METEOR, *COMET* и *COMET-Kiwi*, принцип работы и особенности.

1. *METEOR* – значение оценки данной метрики зависит от результатов сравнения перевода-кандидата с переводом-эталоном. Результат оценки базируется на количестве совпадений униграмм и *n*-грамм, а также зависит от того, насколько отличается длина предложения перевода-кандидата от эталона. Диапазон оценивания по умолчанию от 0 до 1 (табл. 1).

дидат и эталон будут иметь больше совпадений.

Одни из самых низких результатов получили переводы системы *Google* сегментов (10) в переводе на русский – 0,0934, и (6) в переводе на немецкий – 0,0453.

Обращая внимание на результаты в целом, можно сделать вывод о том, что данная метрика «очень строга» в своей оценке, но только из-за того, что довольно примитивна, ведь она является одной из первых созданных метрик. Однако для общего понимания принципа работы таких типов метрик этого может быть вполне достаточно.

2. *COMET* – нейросетевая метрика, которая для оценки использует не только перевод-эталон, но и сам исходный текст. Метрики такого типа способны обрабатывать все входные данные, а также обучаться на них для выполнения последующих оценок. Метрика способна «предсказывать» перевод исходного текста и сравнивать с этими дан-

¹ Источник табл. 1 – URL: <https://huggingface.co/spaces/evaluate-metric/meteor> (дата обращения: 20.10.2024).

ными перевод-кандидат. Кроме того, назначением метрик такого типа является сравнение совпадений смысла высказываний на

исходном и целевом языках. Максимальное значение, которое можно получить по данной метрике – 1, минимальное – 0 (табл. 2).

Таблица 2. Показатели референсной нейросетевой метрики COMET¹

Table 2. Indicators of the COMET reference neural network metric

Сегмент	<i>Google(en-ru)</i>	<i>DeepL(en-ru)</i>	<i>Google(en-de)</i>	<i>DeepL(en-de)</i>
(1)	0,9541	0,9481	0,9395	0,9436
(2)	0,8792	0,8864	0,8941	0,9094
(3)	0,8414	0,8485	0,9295	0,9439
(4)	0,8341	0,8343	0,8184	0,7768
(5)	0,8818	0,8874	0,869	0,8716
(6)	0,796	0,8041	0,6643	0,6436
(7)	0,8075	0,7241	0,8338	0,8444
(8)	0,9125	0,898	0,7695	0,7730
(9)	0,8532	0,8124	0,8324	0,8407
(10)	0,8798	0,865	0,8778	0,9169
Среднее значение	0,864	0,8508	0,8428	0,8464

Результаты проведенной с помощью метрики *COMET* оценки значительно выше по сравнению с результатами метрики *METEOR*. Минимальное значение оценки перевода сегмента (6) на немецкий язык системой *DeepL* имеет значение выше среднего – 0,6436, а самое высокое достигло значения 0,9541 (сегмент (1)).

Рассмотрим перевод сегмента (6) на немецкий язык Моникой Нихаус: *Im Jahr 1967 findet man eine Sammlung von Skizzenblättern und Texten von Leonardo da Vinci, den sogenannten Codex Madrid. Der Codex enthält eine Skizze, die als Basis für eine Addiermaschine mit 13 Rädern interpretiert wurde*; и перевод этого же сегмента, выполненный системой DeepL: *Im Jahr 1967 wurde ein verlegtes Manuskript wieder ausgewertet, das seine Idee für eine Rechenmaschine mit dreizehn digitalen Rädern enthüllte*.

Можно отметить, что переводчица внесла действительно много значимых изменений в структуру предложения и применила некоторые переводческие трансформации (членение предложения, дополнение), тогда как в результате НМП наблюдается лишь

дословность и конкретика, а также отсутствие логических ошибок.

3. *COMET-Kiwi*: данная метрика является улучшенной версией метрики *COMET*. Особенность данного типа метрик в том, что они не нуждаются в эталоне для сравнения, т. е. они способны обрабатывать и сравнивать исходный текст с переводом-кандидатом. Модели нейросетевых метрик нового поколения обучаются на специальных наборах данных, которые содержат оценки переводов, выполненные профессиональными переводчиками по специально заданным критериям. Диапазон оценки этой метрики такой же: от 0 до 1 (табл. 3).

Сразу же отметим, что полученные результаты очень близки к результатам оценки метрикой *COMET*. А минимальное (0,7667) и максимальное значения (0,8962) имеют разницу между собой всего лишь чуть больше 0,1. Как видно из результатов, способность нейросетей к прогнозированию и проведению анализа значительно повышает шанс получить высокую оценку, т. к. отличия в порядке слов или длина высказывания не считаются за ошибку.

¹ Источник табл. 2 – URL: <https://github.com/Unbabel/COMET> (дата обращения: 20.10.2024).

Таблица 3. Показатели оценок безреференсной нейросетевой метрики COMET-kiwi¹

Table 3. Evaluation indicators of the reference-free neural network metric COMET-kiwi

Сегмент	Google(en-ru)	DeepL(en-ru)	Google(en-de)	DeepL(en-de)
(1)	0,896	0,8962	0,8749	0,8744
(2)	0,8654	0,8654	0,821	0,8602
(3)	0,8266	0,8476	0,8365	0,8356
(4)	0,8625	0,8682	0,8361	0,8377
(5)	0,8425	0,8524	0,8257	0,8245
(6)	0,8392	0,8472	0,8088	0,7947
(7)	0,8377	0,7667	0,8268	0,8294
(8)	0,888	0,8898	0,882	0,8818
(9)	0,8694	0,8623	0,8496	0,852
(10)	0,8503	0,8557	0,8462	0,8562
Среднее значение	0,8578	0,8552	0,8408	0,8446

Для лучшего понимания рассмотрим сегмент с минимальной оценкой ((7), *DeepL en-ru*): *If one turned a crank, the gears inside turned in sequence performing the arithmetic calculations* (исходный текст); *если повернуть кривошип, шестеренки внутри последовательно вращались, выполняя арифметические вычисления* (перевод-кандидат). Видим, что результат основывается на несовпадении временных форм глаголов в оригинале и переводе, а также на неточном переводе слова *crank* (рукоятка, рычаг).

Выводы

Обобщая результаты проведенного исследования, можно отметить, что современные метрики действительно хорошо проявляют себя в оценке того, был ли сохранен смысл предложения в переводе. Следовательно, данный способ оценки является скорее инструментом для определения уровня развития самого направления машинного перевода, чем инструментом для выделения ошибок или определения правильности/неправильности перевода (профессионального или машинного) как такового. Однако, если к переводу предъявляются такие требования, как сохранение исходных структур и близость к оригинальному тексту, то в таком случае данный способ может быть эффективным.

На основе проведенного анализа выявлено, как системы НМП *Google Translate* и *DeepL Translate* справляются с переводом

текстов в жанре научно-популярной литературы. При помощи метрик выяснено, насколько сильно нейронный машинный перевод близок к эталонному человеческому. Таким образом, нейросетевые метрики нового поколения показали высокую оценку соответствия, однако традиционные метрики, которые ценят дословные совпадения, существенно от них отстают.

Библиографические ссылки

1. Дьяченко И. Н., Матыченко Ю. В. Нейронный машинный перевод: преимущества, сложности, перспективы // Язык и литература в поликультурном пространстве. 2020. № 6. С. 28–33. EDN: KOINQX
2. Улиткин И. А. Автоматическая оценка качества машинного перевода научно-технического текста // Вестник Московского государственного областного университета. Серия: Лингвистика. 2016. № 4. С. 174–182. DOI: 10.18384/2310-712X-2016-4-174-182. EDN: WPZGQP
3. A Study of Translation Edit Rate with Targeted Human Annotation / M. Snover, B. Dorr, R. Schwartz, L. Micciulla, J. Makhoul // AMTA. 2006. URL: <https://aclanthology.org/2006.amta-papers.25> (дата обращения: 13.04.2024).
4. Bahdanau D., Cho K., Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate // arXiv, 2016. URL: arxiv.org/pdf/1409.0473 (дата обращения: 20.04.2024).
5. Language and Machines: Computers in Translation and Linguistics: a report by the Automatic

¹ Источник табл. 3 – URL: <https://huggingface.co/Unbabel/wmt22-cometkiwi-da> (дата обращения: 20.10.2024).

Language Processing Advisory Committee, National Academy of Science / J. R. Pierce, J. B. Carroll, E. P. Hamp, D. G. Hays [et al.]. Washington, DC : The National Academic Press, 1966. 138 p.

6. *Austermühl F.* Electronic Tools for Translators. Manchester : St. Jerome, 2001. 202 p.

7. *Gross A.* Limitations of Computers as Translation Tools // *Computers and Translation*. London : Routledge, 1992. Pp. 96–130.

8. *Раренко М. Б.* Машинный перевод: от перевода «по правилам» к нейронному переводу (обзор) // *Социальные и гуманитарные науки. Отечественная и зарубежная литература. Сер. 6. Языкознание: Реферативный журнал*. 2021. № 3. С. 70–79. DOI: 10.31249/ling/2021.03.05. EDN: ZSKAVN

9. Прикладная и компьютерная лингвистика : монография / ред. О. В. Митренина, И. С. Николаев, Т. М. Ландо. Москва : URSS, 2016. 320 с. С. 156–189. EDN: WEYYGT

10. *Quah C. K.* Translation and Technology. Basingstoke : Palgrave, 2006. 221 p.

11. *Hutchins J.* Current commercial machine translation systems and computer-based translation tools: System types and their uses // *International Journal of Translation*. 2005. Vol. 17 (1-2). Pp. 5–38.

12. *Улиткин И. А.* Автоматическая оценка качества машинного перевода научного текста: 5 лет спустя // *Вестник Московского государственного областного университета. Серия: Лингвистика*. 2022. № 1. С. 47–59. DOI: 10.18384/2310-712X-2022-1-47-59. EDN: GTAUUM

13. Эволюция метрик качества машинного перевода. Часть 1 // Хабр : сайт. URL: <https://habr.com/ru/articles/745642/> (дата обращения: 13.04.2024).

14. Эволюция метрик качества машинного перевода. Часть 2 // Хабр : сайт. URL: <https://habr.com/ru/articles/748496/> (дата обращения: 15.04.2024).

15. Эволюция метрик качества машинного перевода. Часть 3 // Хабр : сайт. URL: <https://habr.com/ru/articles/758522/> (дата обращения: 16.04.2024).

16. Эволюция метрик качества машинного перевода. Часть 3 // Хабр : сайт. URL: <https://habr.com/ru/articles/758522/> (дата обращения: 16.04.2024).

17. *Kaku M.* Physics of the Future: How Science Will Shape Human Destiny and Our Daily Lives by the Year 2100. New York : Doubleday, 2011. 416 p.

18. *Каку М.* Физика будущего / пер. Наталья Лисова ; ред. Мария Миловидова. Москва : Альпина нон-фикшн, 2012. 584 с. ISBN 978-5-91671-164-6

19. *Kaku M.* Die Physik der Zukunft. Unser Leben in 100 Jahren. Reinbek bei Hamburg: Rowohlt Verlag, 2012. 602 s.

References

1. D'yachenko I.N., Matychenko YU.V. [Neural Machine Translation: Advantages, Difficulties, Prospects]. *Yazyk i literatura v polikul'turnom prostranstve*, 2020, no. 6, pp. 28-33. (in Russ.).

2. Ulitkin I. A. [Automatic assessment of the quality of machine translation of scientific and technical text]. *Vestnik Moskovskogo gosudarstvennogo oblastnogo universiteta. Seriya: Lingvistika*, 2016, no. 4, pp. 174-182. (in Russ.). DOI: 10.18384/2310-712X-2016-4-174-182

3. Snover M., Dorr B., Schwartz R., Micciulla L., Makhoul J. A Study of Translation Edit Rate with Targeted Human Annotation. AMTA. 2006. (in Engl.). Available at: <https://aclanthology.org/2006.amta-papers.25> (accessed 13.04.2024).

4. Bahdanau D., Cho K., Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate. arXiv, 2016. (in Engl.). Available at: arxiv.org/pdf/1409.0473 (accessed 20.04.2024).

5. Pierce J. R., Carroll J. B., Hamp E. P., Hays D. G. [et al.]. Language and Machines: Computers in Translation and Linguistics: a report by the Automatic Language Processing Advisory Committee, National Academy of Science. Washington, DC, The National Academic Press, 1966, 138 p. (in Engl.).

6. *Austermühl F.* Electronic Tools for Translators. Manchester, St. Jerome, 2001. 202 p. (in Engl.).

7. *Gross A.* Limitations of Computers as Translation Tools. Computers and Translation. London, Routledge, 1992, pp. 96-130. (in Engl.).

8. *Rarenko M.B.* [Machine Translation: From Translation “by the Rules” to Neural Translation (Review)]. *Social'nye i gumanitarnye nauki. Otechestvennaya i zarubezhnaya literatura. Ser. 6, Yazykoznanie: Referativnyj zhurnal*, 2021, no. 3, pp. 70-79. (in Russ.). DOI: 10.31249/ling/2021.03.05. EDN: ZSKAVN

9. Red. O.V., Mitrenina I.S., Nikolaev T.M. *Lando Prikladnaja i komp'juternaja lingvistika : monografija* [Applied and computational linguistics, monograph]. Moscow, URSS, 2016, 320 p. (in Russ.). EDN: WEYYGT

10. *Quah C. K.* Translation and Technology. Basingstoke. Palgrave, 2006, 221 p. (in Engl.).

11. *Hutchins J.* Current commercial machine translation systems and computer-based translation tools: System types and their uses. *International Journal of Translation*, 2005, vol. 17 (1-2), pp. 5-38. (in Engl.).

12. Ulitkin I. A. [Automatic evaluation of machine translation quality of a scientific text: five years later]. *Bulletin of the Moscow Region State University. Series: Linguistics*, 2022, no. 1, pp. 47-59. (in Russ.). DOI: 10.18384/2310-712X-2022-1-47-59. EDN: GTAUUM

13. *Evolyuciya metric kachestva mashinnogo perevoda. CHast' 1* [Evolution of machine translation quality metrics. Part 1]. *Habr.* (in Russ.). Available at: <https://habr.com/ru/articles/745642/> (accessed 13.04.2024).

14. *Evolyuciya metric kachestva mashinnogo perevoda. CHast' 2*. [Evolution of machine translation quality metrics. Part 2]. *Habr.* (in Russ.). Available at: <https://habr.com/ru/articles/748496/> (accessed 15.04.2024).

15. *Evolyuciya metric kachestva mashinnogo perevoda. CHast' 3* [Evolution of machine transla-

tion quality metrics. Part 3]. *Habr.* (in Russ.). Available at: <https://habr.com/ru/articles/758522/> (accessed 16.04.2024).

16. *Evolyuciya metric kachestva mashinnogo perevoda. CHast' 3* [Evolution of machine translation quality metrics. Part 3]. *Habr.* (in Russ.). Available at: <https://habr.com/ru/articles/758522/> (accessed 16.04.2024).

17. Kaku M. *Physics of the Future: How Science Will Shape Human Destiny and Our Daily Lives by the Year 2100*. New York, Doubleday, 2011, 416 p. (in Engl.).

18. Kaku M. *Fizika budushhego / per. Natal'ja Lisova ; red. Marija Milovidova* [Physics of the future]. Moscow, Alpina non-fiction, 2011, 416 p. (in Russ.).

19. Kaku M. *Die Physik der Zukunft. Unser Leben in 100 Jahren*. Reinbek bei Hamburg. Rowohlt Verlag, 2012, 602 p. (in Germ.).

E. S. Oshanova, Candidate of Philological Sciences, Associate Professor
Kalashnikov Izhevsk State Technical University, Izhevsk, Russia

D. A. Ponosova

Municipal budgetary educational institution "Secondary school no. 8", Izhevsk, Russia

THE ANALYSIS OF NEURAL MACHINE TRANSLATION SYSTEMS USING AUTOMATED MACHINE TRANSLATION EVALUATION METRICS

This paper introduces current-day automated metrics for neural machine translation evaluation and discusses two ways of result evaluation, which include evaluation using specific automated metrics as well as evaluation by an expert translator. Neural machine translation systems were tested for analysis and evaluation purposes. Google Translate and DeepL Translate systems, which apply a neural network approach, are presented as candidates. The following metrics were considered for evaluation: METEOR as a traditional reference-based metric, COMET as a neural reference-based metric, and COMET-kiwi as a neural reference-free metric. It can be noted that even if these metrics are automated, there is a huge human impact in this automation. Even models with a neural network approach are trained on data provided by humans, as today it is impossible to get rid of reference translations or quality estimations made by experts. Metrics allow to understand and better investigate machine translation, its features and limitations, as well as to determine the direction of its development.

As a part of the analysis, a piece of source text was selected for translation, translated into the target languages using selected neural machine translation systems to get candidate translations, and then a reference translation was specified for each of them.

The results of the metrics evaluation were helpful for understanding how close machine translation is to human translation, and also helped to learn more about the current stage of machine translation systems development. Expert evaluation helped to understand how well such systems are doing in terms of translation performance.

Keywords: neural machine translation; automated evaluation metrics; neural machine translation systems; reference.

Получена: 02.10.2024

ГРНТИ 16.31.21

Образец цитирования

Ошанова Е. С., Поносова Д. А. Анализ факторов внешней среды дополнительного профессионального образования // Социально-экономическое управление: теория и практика. 2024. Т. 20, № 4. С. 106–116. DOI: 10.22213/2618-9763-2024-4-106-116

For Citation

Oshanova E.S., Ponosova D.A. [The analysis of neural machine translation systems using automated machine translation evaluation metrics]. *Social'no-ekonomiceskoe upravlenie: teoria i praktika*, 2024, vol. 20, no. 4, pp. 106-116 (in Russ.). DOI: 10.22213/2618-9763-2024-4-106-116