

Технологии машинного перевода. Нейроперевод

Р. Г. Мифтахова*, Е. М. Черепанова

Башкирский государственный университет

Россия, Республика Башкортостан, 450076 г. Уфа, улица Заки Валиди, 32.

**Email: miftahovar@yandex.ru*

Технологии машинного перевода постоянно совершенствуются, и каждое новое решение значительно улучшает качество перевода. На смену, казалось бы, хорошо зарекомендовавшим себя методам статистического машинного перевода приходят методы нейронного перевода. Статья ставит целью отследить динамику развития методов обработки естественного языка в машинном переводе, изучить принцип работы нейроперевода и его перспективы. В статье проанализированы различные подходы к моделированию машинного перевода с обоснованием преимуществ и недостатков каждой модели.

Ключевые слова: нейроперевод, машинный перевод, обработка естественного языка.

В 2007 году британский информатик Джеффри Хинтон предложил метод обратного распространения ошибки для тренировки многослойной нейронной сети. Такой метод, обучающий глубокую нейронную сеть с большим количеством слоев, был назван методом «бэкпропа» и использовался для распознавания изображений и речи, переводах, компьютерных играх и т.д. На сегодняшний день глубокие нейронные сети могут с легкостью обходить даже лучшие системы распознавания изображений, а все достижения в области исследований искусственном интеллекте сделаны именно благодаря прорыву Джеффри Хинтона. [9] Таким образом, взяв все самое лучшее, что имеется в мозге человека, а именно самоорганизацию, способность учиться и обобщать, а также устойчивость к сбоям (ошибкам), человечество совершило новый виток в истории, изменив привычное представление об искусственном интеллекте, и, в частности, о нейронных сетях. Главное, что необходимо помнить – это то, что не существует одной, единой нейросети, а существует несколько парадигм нейросетей по их обучению и использованию [4, 6].

Машинный перевод играет значительную роль в коммерции. Ежегодно на переводческую сферу тратится до 40 миллиардов долларов, речь идет только о сумме, потраченной на перевод специалистами-переводчиками. Машинный перевод – необходимое условие для работы в таких государственно-политических образованиях, как например Евросоюз, для поддержания переговоров и сотрудничества на различных уровнях. Сегодня существенное развитие МП получает так же в странах Азии, так как существует множество отраслей его применения, от торговых и социальных до военных и по-

литических задач. Сам МП день превратился в одну большую индустрию. Ежедневно Google переводит до 100 миллиарда слов. В прошлом году компания Facebook запустила собственный машинный переводчик.

В современном мире большое внимание уделяется применению искусственных нейросетей для обработки естественного языка.

«Машинный перевод активно развивается последние двадцать лет, однако, до идеала еще далеко – ведь даже люди, чтобы стать хорошими переводчиками, учатся годами, – считает разработчик Яндекс переводчика, – Думаю, что самая главная проблема на нашем пути – это контекст, понимание которого необходимо для успешной коммуникации. Профессиональный переводчик, к примеру, не просто владеет двумя языками, чаще всего он также является специалистом в какой-то области. Владение конкретной областью знаний, понимание того, что именно ты переводишь, осознание различий между языками – все это пока еще очень сложные задачи для компьютера».

Люди свободно могут пользоваться несколькими языками. Например, при переводе с английского на русский, человек понимает, когда нужно менять порядок слов, когда слова должны быть согласованы, например, по роду. А для компьютера это очень сложная задача.

Нейросистема моделирует весь процесс машинного перевода на основе уникальной искусственной нейросети. Искусственная нейросеть состоит из уровней искусственных нейронов, которые связаны, так называемым, весом, т.е. параметрами сети. Ключевой элемент нейросети заключается в его способности автоматически исправлять собственные параметры во время обучающей фазы (в среднем 4–5 недель). С технической точки зрения, сгенерированный вывод сравнивается с ожидаемым эталоном, затем скорректированный вариант отправляется вновь в систему для настройки параметров и сетевых связей.

Такая технология использует сложные алгоритмы на стадии Глубокого обучения, давая возможность системе «обучаться» и, на основе переведенного текста генерировать правила языка и обеспечивать вполне адекватный перевод.

В цифровом мире, языковые барьеры до сих пор стоят на пути к быстрому развертыванию бизнес стратегий на глобальном рынке. «Технология нейроперевода представляет собой новый шаг для общения людей в целом и для туризма, в частности. Это великолепное поле возможностей, новые горизонты и захватывающее лингвистическое путешествие!», – так отзывается о системах нейроперевода Доминик Аузиас, основатель компании Petit Fute. А это мнение генерального директора компании Лекселера: «Система понимала, что я хочу сказать и переводила очень быстро. По большей части, терминология определялась правильно, а предложения звучали как из уст человека».

Компания SYSTRAN успешно разрабатывает технологию нейроперевода. По мнению Дж. Сенелларта, технология нейросетей прекрасно подходит для адаптации перевода в отдельном домене, таком как юридический, технический, фармацевтический, что так необходимо глобальным компаниям и организациям. Генеррирование нейросистем предлагает новые возможности для адаптации доменов.

Google также объявил, что технология нейронного машинного перевода теперь используется для перевода с русского языка на английский и обратно. Нововведение должно существенно повысить качество перевода. Оно запущено как в веб-версии Google Translate, так и в приложениях сервиса для iOS и Android.

Как в статистическом машинном переводе, так и в нейросетях используются огромные базы данных, на которых происходит обучение модели. Статистический метод работает на основе параллельных двуязычных корпусов, то есть огромного количества переведенных людьми предложений. Благодаря большому объему памяти статистический переводчик запоминает конкретные фразы, даже, если они встречались только один или два раза. Такой метод не всегда может обработать целое предложение, поэтому часто фразы не согласованы между собой.

Тогда как машинный переводчик на основе нейросетей способен «разпознать» предложение целиком, он «понимает» структуру отношений между словами, а значит, переведенный на его основе материал более естественен.

Обе вышеназванные системы перевода имеют свои преимущества и недостатки. Поэтому Яндекс решил запустить гибридную систему. В гибридной системе исходный текст переводится с помощью двух технологий, получившийся результат оценивается системой оценки качества машинного перевода BLEU, после чего пользователь получает вариант перевода, получившего больше очков.

Какова вероятность того, что машинный перевод достигнет совершенства и специалисты переводчики останутся без работы? Скорее всего переводчики станут чаще прибегать к помощи машины. До недавнего времени качество машинного перевода было неутешительным. Например, правка текста, переведенного машинным способом, зачастую занимала больше времени, чем тот же перевод, выполненный человеком самостоятельно. Людям легче понять и воспроизвести ту информацию о тексте, которая недоступна компьютеру – например, стилистические особенности оригинала. Компьютер этого пока не умеет.

Но машинный перевод предоставляет взаимовыгодные возможности: не только человеку видны ошибки компьютера, но и компьютер может подсказать человеку, что где-то кроется ошибка. Таким образом, они могут «обучать» друг друга. Есть, действительно, опасение, что искусственный интеллект оставит многих специалистов без работы.

Но компьютер не заменит человека, а расширит его возможности. С его помощью люди будут переводить с быстрее и лучше, но определенно не прекратят этого делать.

Более того, машинный перевод способствует сохранению редких языков.

С появлением машинных переводчиков, в интернете появляется больше информации на редких языках, больше корпусных данных, которые, к тому же, полезны желающим изучать эти языки.

Литература

1. Анненков. А. В России создан смысловой процессор. <http://d-russia.ru/v-rossii-sozdan-smyslovoj-processor.html>
2. Морозкина. Е. А. Мифтахова Р. Г.. Влияние информационных технологий на развитие лингвистических норм. Вестник БашГУ. 2012. №1.– С. 162–164
3. Birch A., Blunsom Ph., Osborne M. A quantitative analysis
4. of reordering phenomena. In Proceedings of the Fourth Workshop on Statistical Machine Translation, pages 197–205. Association for Computational Linguistics, 2009
5. Bodomo, A. B. 2009. Computer-Mediated Communication for Linguistics and Literacy: Technology and Natural Language Education. IGI Global, Hershey, PA, USA. 374 pages.
6. David Kriesel, A Brief Introduction to Neural Networks, 2007, 286p., available at <http://www.dkriesel.com>
7. Lopez A. Statistical machine translation. ACM Computing Surveys
8. (CSUR), 40(3):8, 2008
9. Menezes A., C. Cherry. 2005. Dependency treelet translation: Syntactically informed phrasal smt. Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics, pages 271–279.
10. Yamada K., K. Knight. A syntax-based statistical translation model. In Proceedings of the 39th Annual Meeting on Association for Computational Linguistics, pages 523–530. Association for Computational Linguistics, 2001
11. <https://ain.ua/2017/03/03/kak-rabotayut-nejroseti/>

Machine translation technologies. Neurotranslation

R. G. Miftakhova*, E. M. Cherepanova

Bashkir State University

32 Zaki Validi Street, 450074 Ufa, Republic of Bashkortostan, Russia.

**Email: miftahovar@yandex.ru*

The technologies of machine translation are constantly changing, and every new solution improves the quality of machine translation. The seemingly well-established methods of statistical machine translation are being replaced by the methods of neural translation. The article aims to trace the development of methods for natural language processing in machine translation, to determine the degree of elaboration of neural models and its prospects. Various approaches to the modeling of machine translation with the justification of the advantages and disadvantages of each model were investigated.

Keywords: neurotranslation, machine translation, natural language processing.