

А. М.СОЛОВЧУК

Брест, БрГУ имени А. С.Пушкина

РАЗВИТИЕ НЕЙРОННОГО МАШИННОГО ПЕРЕВОДА

Нейронный машинный перевод (англ. NeuralMachineTranslation, NMT) — это подход к машинному переводу, который использует большую искусственную нейронную сеть. Он отличается от методов машинного перевода, основанных на статистике фраз, которые используют отдельно разработанные подкомпоненты.

Системы машинного перевода прошли различные стадии развития, при прохождении каждой из которых качество перевода значительно повышалось. До начала 90-х годов прошлого века почти все системы основывались на наборах правил.

Позже ученые IBM предложили статистический метод, который использует уже переведенные примеры. Для этого вам нужно достаточное количество пар предложений в исходном и желаемом языке и, используя алгоритмы, автоматически обучить модель перевода отдельных слов. Эти алгоритмы позволили выяснить для каждой пары предложений, какие исходные слова связаны с определенными словами перевода, то есть получить соответствие слов. Вместе с моделью целевого языка, основанной на существующих текстах, это позволило нам переводить слова в контексте, строить перевод, переходя от одного слова в исходном предложении к другому, пытаясь достичь синтаксически правильного порядка слов.

В начале 2000-х эти системы были значительно улучшены благодаря использованию перевода фраз. Фразы для приверженцев статистического машинного перевода — это группы последовательных слов, от семи до одного. Это позволило системе запомнить характерные, часто встречающиеся фразы, запомнить контекст. Такие системы хорошо переводят предложения, состоящие из фраз, которые система уже получила во время обучения. Но качество перевода было значительно снижено, если система столкнулась с чем-то новым. Эти системы использовались различными онлайн-переводчиками с момента их создания.

Наконец, в 2014–2016 гг. В машинном переводе произошел большой прорыв. Системы на основе нейронных сетей очень быстро завоевали умы ученых и также стали использоваться в коммерческих приложениях. Качество перевода с использованием нейронных сетей настолько улучшилось, что зачастую становится невозможным отличить перевод, сделанный человеком, от машинного перевода, сделанного компьютером, особенно при рассмотрении перевода отдельных предложений. [1].

Сервисы перевода компаний Google, Яндекс, Microsoft и PROMT уже используют нейронный перевод. Google использует нейронный машинный перевод Google (GNMT) вместо ранее используемых статистических методов. Майкрософт использует похожую технологию для перевода речи (в том числе в Майкрософт Переводчике и Skype Переводчике). Гарвардской группой по обработке естественного языка была выпущена OpenNMT, система нейронного машинного перевода с открытым исходным кодом. Яндекс.Переводчик имеет гибридную модель: свой вариант перевода предлагает и статистическая модель, и нейросеть. После этого технология CatBoost, в основе которой лежит машинное обучение, будет выбирать лучший из полученных результатов [2].

Основные проблемы машинного перевода на текущий момент это: распространение систем нейронного перевода на большее количество языковых пар и зависимость нейронных сетей от типа данных, на которых они тренируются

Список использованной литературы

1. Системный Блок{ъ} - Хроника перехода культуры в цифру [Электронный ресурс]. – Режим доступа :<https://sysblok.ru/nlp/nejronnye-seti-v-mashinnom-perevode-status-kvo/> – Дата доступа : 20.03.20

2. «Википедия — свободная энциклопедия»[Электронный ресурс]. – Режим доступа : https://ru.wikipedia.org/wiki/Нейронный_машинный_перевод – Дата доступа : 20.03.20