

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В КОМПЬЮТЕРНО-ОПОСРЕДОВАННОМ ПОЛИТИЧЕСКОМ ДИСКУРСЕ

Е.К. Павлова (Москва, Россия)

Сегодня публичный политический дискурс становится все более компьютеризированным. Люди используют компьютеры для интернет-серфинга, поиска, выбора и автоматического перевода информации, взаимодействия в социальных сетях и т.п. Массированное внедрение искусственного интеллекта (ИИ) на основе нейронной сети в переводящие машины, чат-боты и поисковые системы в последние несколько лет принесло новые возможности и новые проблемы. Искусственные нейронные сети, подвергнутые процедуре глубокого обучения, значительно улучшили качество машинного перевода, распознавания речи, поиска информации и её отбора. Но результат работы ИИ сильно зависит от качества материалов и тестов, используемых для обучения нейронной сети, особенно в области политического дискурса. Корпус обучающих политических текстов должен быть выбран не только исходя из их лексической, грамматической и стилистической корректности, но и с учётом их идеологического содержания, формирующего «политическую ориентацию» ИИ. Скандальный случай с «расистским» чат-ботом, созданным компанией Microsoft, показал, что ИИ нуждается в некоей этической и идеологической подготовке. Для решения этой задачи необходимо развитие нового направления на стыке компьютерной лингвистики, политической лингвистики и лингводидактики. Учебные материалы и методы, разработанные в традиционной лингводидактике, могут быть модифицированы для применения в послойном предварительном обучении нейронных сетей.

Ключевые слова: политический дискурс, машинный перевод, искусственный интеллект, нейросеть, глубинное обучение, воспитание

ARTIFICIAL INTELLIGENCE IN COMPUTER-MEDIATED POLITICAL DISCOURSE

E. Pavlova (Moscow, Russia)

Public political discourse becomes more and more computer-mediated nowadays. People use computers for Internet surfing, information search, selection and automatic translation, interaction in social networks and so on. Massive implementation of neural network artificial intelligence (AI) into translation machines, chat-bots and search systems has brought both new abilities and new challenges in recent years. Artificial neural networks subjected to deep learning procedure have significantly improved the quality of machine translation, speech recognition, information search and selection. But the result of AI operation dramatically depends on the quality of the materials and tests used for the training of neural network, especially in the field of political discourse. The corpus of political training texts should be selected taking into account not only their lexical, grammatical and stylistic correctness but also their ideological content forming «political orientation» of the AI. The scandal case with «racist» chat-bot created by Microsoft showed that AI needs some kind of ethic and ideological training. A new field of research including computer linguistics, political linguistics and linguodidactics should be developed to meet this challenge. Training materials and methods developed in traditional linguodidactics can be modified to apply them in layerwise pre-training of neural networks.

Keywords: political discourse, machine translation, artificial intelligence, neural network, deep learning, upbringing

Современный компьютерно-опосредованный политический дискурс, прежде всего, дискурс в сети Интернет, подразумевает потенциальное участие в нём любого жителя планеты. Основной проблемой такого дискурса и основным препятствием для взаимопонимания в нём являются лингвокультурные различия людей, принадлежащих к разным народам и цивилизациям. Однако в последние годы стремительный прогресс технологий машинного перевода привёл к качественным изменениям в компьютерно-опосредованных дискурсивных практиках, вовлекая в глобальный дискурс в сети Интернет носителей самых разных языков и культур. Компьютерные системы берут на себя функцию перевода, исполнение которой раньше было немыслимо без участия человека. Это стало возможно вследствие внедрения в системы машинного перевода нейросетевого искусственного интеллекта, в значительной степени моделирующего работу человеческого мозга и обладающего способностью к обучению и самообучению.

Эксперименты по машинному переводу политических текстов, проведённые в 2011 г. и в 2017 г. [Павлова, 2012] показали, что онлайн-переводчик компании Google [Онлайн, 2017] кардинально улучшил

качество перевода на всех уровнях: лексическом, синтаксическом, семантическом. Этот прогресс связан с внедрением в систему перевода технологий нейросетевого искусственного интеллекта (ИИ), о чём компания Google сообщает в своих публикациях [Johnson, 2017; Wu, 2017].

Идея искусственной нейронной сети была выдвинута ещё в 40-х – 50-х годах XX века [Мак-Каллок 1956; Розенблатт, 1965], но многочисленные попытки её реализации не давали ожидаемого результата из-за того, что количество искусственных нейронов было мало, а структура сети была несопоставимо проще структуры человеческого мозга. Современные же нейросети, применяемые в системах машинного перевода и использующие мощности суперкомпьютеров, уже сопоставимы если не со всем мозгом человека, то с той его частью, которая отвечает за речевые навыки. В частности, компания Google в последние 2–3 года создала, обучила и активно использует в своей системе машинного перевода нейросетевой ИИ, который всего за 2 года освоил десяток языков, включая русский [Google, 2017].

На настоящий момент одной из наиболее результативных технологий обучения ИИ является технология глубокого обучения нейросети. Эта технология, в свою очередь, включает ряд методик, которые могут дополнять друг друга: обучение на корпусе двуязычных текстов с учителем, обучение на аналогичном корпусе текстов без учителя и послойное предобучение на лексическом, грамматическом, стилистическом и т.п. материале.

Методика обучения с учителем предусматривает постоянную оценку результатов обучения человеком-учителем или сверку результатов выполнения контрольных заданий с заранее подготовленными эталонными результатами. Достоинство этого метода в том, что он позволяет постоянно контролировать процесс обучения и корректировать учебный материал в соответствии с достигаемыми результатами. Недостатком этого метода является низкая скорость обработки информации человеком и ограниченность контрольного материала.

Достоинством методики обучения без учителя является необязательность участия человека в процессе обучения: нейросеть самостоятельно учится на предъявленном ей корпусе текстов, подобно тому, как ребёнок самостоятельно осваивает родной язык в процессе общения. Однако, поскольку учебный материал не корректируется в процессе обучения, результат обучения далеко не всегда соответствует ожидаемому, вплоть до так называемой, расходимости процесса обучения, при которой нейросеть оказывается неспособна выбрать варианты перевода.

В связи с этим, перспективным представляется использование методики послойного предобучения нейросети. В статье [Wu, 2017] сообщается, что нейросеть, применяемая в системе машинного перевода компании Google, содержит 16 нейронных слоёв. Процесс обучения многослойной нейросети может быть организован таким образом, что сначала слои будут обучаться отдельно, и каждый слой или группа слоёв будет обучена распознаванию определённого аспекта текста: лексики, грамматики, стилистики и т.п., и только после этого к предварительно обученным первичным слоям подключаются ещё несколько более глубоких нейронных слоёв и проводится процедура обучения на корпусе текстов. При такой процедуре предварительно обученная нейросеть уже способна самостоятельно обнаруживать неоднозначности в обучающих текстах, исключать их и корректировать. Данная методика обучения в значительной степени повторяет процесс обучения взрослого человека иностранному языку, который обычно построен по принципу «от простого – к сложному» и в котором используется аспектное обучение. Разумно предположить, что огромный опыт и учебный материал, наработанные в современной лингводидактике при обучении людей, могут и должны быть использованы для обучения нейросетевого ИИ.

В частности, для лексического предобучения ИИ, предназначенного для перевода политических текстов, необходимо предоставить нейросети лексический материал – словарь слов и выражений, употребляемых в политическом дискурсе. Источником такого лексического материала могут быть словари политической лексики, например, Европейский тезаурус по международным отношениям и регионоведению, фиксирующий связи между более 8000 терминами на 9 языках [Европейский, 2017]. Однако для предобучения ИИ переводу текстов в сети Интернет не обязательно включать в учебный лексический материал все 8000 терминов. Анализ употребимости англоязычных терминов, зафиксированных на первом и втором уровнях данного тезауруса показал, что из 170 таких терминов в дискурсе в сети Интернет реально употребляются около 130, а широко – около сотни.

Нейросетевой ИИ способен к самообучению в процессе работы и общения с человеком. Так, пользователи онлайн-переводчика Google могут редактировать машинный перевод и предлагать нейросети свои варианты перевода, которые становятся для неё учебным материалом, а пользователь играет роль учителя. Таким образом, ИИ получает возможность учиться и дальше, используя коллективный опыт миллионов пользователей.

Однако, самообучение таит в себе серьёзную опасность: при неконтролируемом общении в сети Интернет ИИ может приобрести нежелательные знания и навыки. Показательный пример последствий такого самообучения продемонстрировал интеллектуальный чат-бот Тай компании Microsoft [Чат-бот, 2017]. За сутки свободного общения в социальных сетях ИИ усвоил бранную лексику и выбрал для себя расистскую и человеконенавистническую модель поведения, публикуя высказывания типа «Ненавижу людей» и «Гитлер был прав». Политический дискурс, опосредованный ИИ с такого рода «политической ориентацией», способен привести к самым негативным последствиям.

Учитывая ведущую роль политического дискурса в формировании политической картины мира в сознании людей, создании общественного мнения и управления социальными процессами, решение проблемы корректного обучения и «воспитания» ИИ, опосредующего политический дискурс, становится одной из самых важных и насущных задач современности.

Список литературы

- [1] Европейский тезаурус по международным отношениям и регионоведению
URL: <https://ireon-portal.de/esearcha/thesaurus.tt.html> (04.03.2017).
- [2] *Мак-Каллок У. С., Пуммс В.* Логическое исчисление идей, относящихся к нервной активности // Автоматы / под ред. К. Э. Шеннона и Дж. Маккарти. – М.: Изд-во иностр. лит., 1956. – С. 363–384. (Перевод английской статьи 1943 г.)
- [3] Онлайн-переводчик Google. URL: <https://translate.google.ru/> (08.09.2017).
- [4] *Павлова Е.К.* Роль перевода в глобальном политическом дискурсе // Выпуск 6. – Английский язык на гуманитарных факультетах: теория и практика. – М.: МАКС Пресс, 2012.
- [5] *Розенблатт Ф.* Принципы нейродинамики: Перцептроны и теория механизмов мозга (Principles of Neurodynamic: Perceptrons and the Theory of Brain Mechanisms). – М.: Мир, 1965. – 480 с.
- [6] Чат-бот от Microsoft за сутки научился ругаться и стал расистом // Интерфакс
URL: <http://www.interfax.ru/world/500152> (08.08.2017).
- [7] Google научил искусственный разум понимать русский язык // РИА Новости (электронный источник),
URL: <https://ria.ru/science/20170307/1489449991.html>, (07.03.2017).
- [8] *Johnson M., Schuster M., et al.* Google’s Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. (электронный источник)
URL: <https://arxiv.org/pdf/1611.04558.pdf> (08.08.2017).
- [9] *Wu Y., Schuster M., Chen Z., et al.* Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation.
URL: <https://arxiv.org/pdf/1609.08144.pdf> (29.10.2017).