

АНАЛИЗ ДООБУЧЕННЫХ ЯЗЫКОВЫХ МОДЕЛЕЙ ДЛЯ ЗАДАЧИ ГЕНЕРАЦИИ ДИСТРАКТОРОВ

Меньщиков М.А. (Петрозаводский государственный университет)

Научный руководитель – доцент, кандидат физико-математических наук Корзун Д.Ж.
(Петрозаводский государственный университет)

Введение. Качество тестовых заданий с множественным выбором (Multiple Choice Questions) сильно зависит от созданных к ним неправильных ответов: дистракторов. Если эти отвлекающие варианты не могут запутать тестируемых участников, то правильный ответ может быть найден из логических соображений, никак не основанных на знании предметной области, понимание которой проверяется в данном вопросе. В результате качество всего задания сильно ухудшится, а сам тест теряет возможность объективной оценки знаний аттестуемых. Тем не менее, подбор подходящих дистракторов является сложной задачей. Вместо того, чтобы быть очевидным неправильным вариантом, дистрактор должен обладать правдоподобностью, которая сбивает с толку аттестуемых, не усвоивших проверяемые предмет. Хороший дистрактор должен быть грамматически правильным с точки зрения вопроса и семантически соответствовать его контексту. Составители заданий должны добиться правдоподобия дистрактора, при этом не задев присущую ему некорректность. Таким образом, ручная подготовка дистракторов требует много времени и ресурсов. На сегодняшний день актуальной задачей является разработка методов автоматической генерация “хороших” отвлекающих вариантов ответа, которые смогут облегчить работу составителей вопросов и уменьшить количество необходимых знаний/опыта для их составления по конкретной области [1].

За последние годы языковые модели на основе Transformer-архитектуры прошли путь от экспериментальных исследований до внедрения в большинство задач по обработке естественного языка (NLP). Их популярность началось с ULM-FiT- и BERT-моделей, которые показали потенциал предварительного обучения (Pre-training) большой нейронной сети с использованием подхода обучения без учителя (Unsupervised Learning). После предварительного обучения эти нейронные сети можно дообучить (Fine-tune) и применить для решения разных конкретных задач. Это привело к существенным улучшениям в решении задач классификации эмоционального состояния по тексту, генерации ответов на вопросы, извлечения признаков из неструктурированных данных и многих других. С появлением GPT2-архитектуры, был сделан большой скачок по качеству генерации текстовых данных. Было показано, что производительность языковой модели продолжает улучшаться с увеличением её размера: начиная от 117 миллионов параметров для самой маленькой модели GPT-2 до 175 миллиардов параметров для самой большой из моделей GPT-3.

В рамках данного исследования внимание будет сконцентрировано на применении предобученных языковых моделей для генерации грамматически верных и подходящих под контекст вопроса неправильных вариантов ответа (дистракторов).

Основная часть. Для некоторой предметной области есть четыре классов вопросов, для которых мы хотим научиться автоматически получать дистракторы: определения, слова-индикаторы, “если..., то...”, “при условии...”. Выдвигаются следующие гипотезы, которые будут проверены в рамках данной работы:

- 1) Языковая модель, дообученная на всех классах вопросов, даёт лучшее качество генерируемых дистракторов по сравнению с подходом по созданию группы моделей, генерирующих неправильные ответы для заданных классов.
- 2) Мультиязыковые модели после дообучения позволяют получить качество генерируемых дистракторов не хуже по сравнению с моделями, генерирующих текст только на одном языке.

Исследование может быть разбито на следующие этапы:

- 1) Обучение моделей генерации дистракторов для каждого класса вопросов.

- 2) Обучение общей модели, генерирующей дистракторы для всех классов вопросов на одном языке [2, 3].
- 3) Обучение мультязыковой модели, генерирующей дистракторы для всех классов вопросов [4].

В качестве основного языка, для которого планируется проводить анализ качества генерируемых дистракторов, выбран русский. Для всех экспериментов планируется взять готовые языковые модели из открытого репозитория HuggingFace [5] и дообучить их под данную задачу: будут использованы T5- и GPT2-архитектуры. Будет опробовано два формата входной последовательности токенов, которая будет подаваться моделям в процессе дообучения: текстовый фрагмент с контекстом вопроса, правильный ответ, вопрос и, либо один, либо сразу несколько эталонных дистракторов, которые мы хотим научиться автоматически генерировать. В дообучении мультязыковых моделей будет участвовать RACE-датасет на английском языке. Для оценки качества полученных дистракторов будут использоваться следующие метрики: BLEU-2, ROUGE-L и BertScore. Для каждой группы моделей будут проведены эксперименты по подбору оптимальной стратегии получения следующего токена в генерируемой последовательности: GreedySearch, BeamSearch и Sampling стратегии.

В результате работы для четырёх классов вопросов будут получены 24 языковых модели генерации дистракторов: 10 моделей на основе ruGPT2, 10 моделей на основе ruT5, 2 модели на основе mGPT2 и 2 модели на основе mT5. Помимо метрик оценки дистракторов, описанных выше, будет проведён их ручной анализ на предмет наличия грамматических ошибок и степени их связи с контекстом вопроса. На основании этих данных будет определён оптимальный способ автоматического получения неправильных вариантов ответа для заданного вопроса при помощи предобученных языковых моделей.

Выводы. В результате проведённых экспериментов будет выполнена оценка предложенных методик генерации дистракторов и сделано заключение о выполнимости выдвинутых гипотез. Ожидается, что модели, обученные на всех классах вопросов, будут давать лучшее качество. Полученное решение может быть интегрировано в системы автоматического создания тестовых заданий для образовательных учреждениях.

Информация о финансовой поддержке. Проект ООО «Универтус» по теме «Разработка программного ассистента для аттестации профессиональных знаний на основе методов искусственного интеллекта» при финансовой поддержке Фонда содействия инновациям.

Список использованных источников.

1. Смирнов Н., Семёнов Н., Абашкина Е., Храмов К. Автоматическая генерация вопросов по текстовому документу // XVI Всероссийская научно-практическая конференция, Петрозаводск. – 2022. – С. 88-91.
2. Offerijns, J., Verberne, S., Verhoef, T. Better Distractors: Transformer-based Distractor Generation and Multiple Choice Question Filtering [Электронный ресурс] // arxiv.org. – 2020. – URL: <https://arxiv.org/abs/2010.09598> (дата обращения: 17.02.2023)
3. Qiu Z., Wu X., Fan W. Automatic Distractor Generation for Multiple Choice Questions in Standard Tests // In Proceedings of the 28th International Conference on Computational Linguistics, Barcelona, Spain (Online). – 2020. – С. 2096–2106.
4. Armengol-Estapé J., Bonet G., Melero M. On the Multilingual Capabilities of Very Large-Scale English Language Models // In Proceedings of the Thirteenth Language Resources and Evaluation Conference, Marseille, France. – 2022. – С. 3056–3068.
5. HuggingFace [Электронный ресурс]. – 2023. – URL: <https://huggingface.co/> (дата обращения 16.01.2023)

Меньшиков М.А. (автор)

Подпись

Корзун Д.Ж. (научный руководитель)

Подпись

