

ГЕНЕРАЦИЯ SQL-ЗАПРОСОВ ИЗ ФРАЗ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ

Бушков Николай Алексеевич

Аспирант

Факультет биологической и медицинской физики МФТИ, Москва, Россия

E-mail: bushkov@phystech.edu

Большие объёмы информации, накопленной человечеством в ходе своего развития, содержатся в структурированных хранилищах. Реляционные базы данных на сегодняшний день являются одними из наиболее распространённых типов таких хранилищ, а вместе с ними и язык запросов - Structured Query Language (SQL). SQL - это язык программирования, и для его использования требуется специальная подготовка, создающая барьер для применения его рядовыми специалистами-аналитикам. По этой причине задача автоматического перевода запросов на естественном языке, таком как английский или русский, в SQL-запросы имеют высокую практическую значимость.

Распознавание смысла запроса пользования, а точнее более общая задача семантического парсинга, решается исследователями в области искусственного интеллекта уже многие годы, однако качество текущих лучших решений не позволяет широко внедрять их в индустрии. Эти решения основаны на глубоких искусственных нейронных сетях, методе машинного обучения, в последние годы набравшем популярность благодаря выросшим вычислительным ресурсам. Если быть точнее, то для задач обработки последовательностей (например, последовательностей слов) в подавляющем большинстве случаев используются рекуррентные сети [1]. Целью настоящей работы является упростить и ускорить решение задачи генерации SQL-запросов из фраз на естественном языке с помощью адаптации нерекуррентной архитектуры Transformer [2], не ухудшив при этом качества решения, которое для этой задачи составляет 68% (точность совпадения результатов запроса на тестсете WikiSQL [1]).

Литература

1. Zhong V., Xiong C, Socher R.: Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning
<https://arxiv.org/abs/1709.00103>
2. Vaswani A. et al.: Attention Is All You Need
<https://arxiv.org/abs/1706.03762>