

Обучение двунаправленного энкодера с ускоренным вниманием через скрытые состояния



Специалисты СПбПУ ускорили «мозг» языковых моделей. Новая программа помогает быстрее обучать языковые модели, которые экономно расходуют вычислительные ресурсы и лучше справляются с анализом длинных текстов.

В Санкт Петербургском политехническом университете Петра Великого разработали программу, которая ускоряет обучение современных языковых моделей — тех самых алгоритмов, что лежат в основе умных поисковиков, чат ботов и систем анализа текстов. Программа позволяет заранее «натренировать» модель на больших массивах текстов, а затем быстро дообучать ее под конкретные задачи: от работы с документами до анализа различных последовательностей.

Главный эффект разработки в том, что она помогает обойти одно из главных ограничений таких моделей — дорогую работу с длинными текстами. Внутри используется особый механизм, который иначе распределяет «внимание» модели к словам в предложении и опирается на дополнительные скрытые состояния. Благодаря этому снижается нагрузка на

вычислительные ресурсы, и работа с объемными документами — технической документацией, юридическими текстами, отчетами или логами — становится заметно менее затратной.

Модель обучается в два этапа: программа сама рассчитывает, сколько шагов обучения нужно сделать, а после каждой эпохи сохраняет параметры и веса. Это дает возможность строить собственные предобученные языковые модели на данных компании или отрасли, возвращаться к нужным этапам обучения и надежно переносить модель на новые задачи — от классификации и поиска по текстам до извлечения ключевой информации.

Разработка снижает порог входа в создание своих языковых моделей. Университеты, исследовательские группы и компании с ограниченным вычислительным бюджетом получают инструмент, который позволяет работать не только с крупными универсальными моделями, но и обучать собственные решения на специализированных корпусах. Это особенно важно для сфер, где критична отраслевой контекст: права и финансы, медицина, промышленность, энергетика, высокотехнологичный сектор.

Подробнее — в опубликованном патенте № [2026664030](#) на [сайте ЦИСИТТ](#).