

АКТИВНОЕ ОБУЧЕНИЕ ДЛЯ КЛАССИФИКАЦИИ НАМЕРЕНИЙ В КОНТЕКСТНОМ ОБУЧЕНИИ БОЛЬШИХ ЯЗЫКОВЫХ МОДЕЛЕЙ

ACTIVE LEARNING FOR INTENT CLASSIFICATION IN IN-CONTEXT LEARNING OF LARGE LANGUAGE MODELS

D. Dranga

Summary. In modern natural language processing, large language models have demonstrated high effectiveness across various tasks, including intent classification in dialogue systems. However, selecting the optimal set of examples for contextual learning remains challenging, especially with a limited annotation budget. This paper proposes an adapted active learning algorithm for the efficient selection of examples in contextual few-shot learning of large language models for intent classification tasks. The proposed method takes model uncertainty into account and ensures diversity in the selected examples, thereby enhancing classification accuracy. Experimental studies on three different datasets have shown that the developed approach outperforms alternative methods, providing higher classification quality for both closed models (e.g., GPT-4) and open models that perform well in Russian (e.g., Gemma 2 27b).

Keywords: active learning, contextual learning, large language models, intent classification, few-shot learning, uncertainty estimation.

Дранга Даниил Игоревич

*Национальный исследовательский технологический университет МИСиС, Москва, Российская Федерация
dranga.daniil@gmail.com*

Аннотация. В современной обработке естественного языка большие языковые модели продемонстрировали высокую эффективность в различных задачах, включая классификацию намерений пользователя в диалоговых системах. Однако выбор оптимального набора примеров для контекстного обучения остается вызовом, особенно при ограниченном бюджете на разметку данных. В данной работе предлагается адаптированный алгоритм активного обучения для эффективного подбора примеров в контекстном обучении больших языковых моделей для задачи классификации намерений. Предложенный метод учитывает неопределенность модели и обеспечивает разнообразие выбранных примеров, способствуя улучшению точности классификации. Экспериментальные исследования показали, что разработанный подход превосходит альтернативные методы, обеспечивая более высокое качество классификации как для закрытых моделей (например, GPT-4), так и для открытых моделей, демонстрирующих хорошие результаты на русском языке (например, Gemma 2 27b).

Ключевые слова: активное обучение, контекстное обучение, большие языковые модели, классификация намерений, few-shot обучение, оценка неопределенности.

Введение

Диалоговые системы и виртуальные помощники, предоставляющие взаимодействие на естественном языке, стали неотъемлемой частью жизни. Ключевая задача в их разработке — классификация намерений пользователя по его сообщениям, что напрямую влияет на качество систем и удовлетворенность пользователей.

С появлением больших языковых моделей (BERT [1], GPT-3 [2], PaLM [3] и др.) произошел значительный прогресс в обработке естественного языка. Эти модели решают задачи генерации текста, перевода и понимания контекста, часто превосходя предыдущие подходы. Перспективной методикой их использования является контекстное обучение (in-context learning) [2], где модель генерирует ответы, опираясь на несколько примеров без явного обучения.

Контекстное обучение актуально при ограниченном количестве размеченных данных или необходимости быстрой адаптации модели без переобучения. Однако его эффективность зависит от выбора контекстных при-

меров [4]. В задачах с множеством классов, как классификация намерений, критичен выбор наиболее информативных примеров [5].

Активное обучение [6] оптимизирует сбор размеченных данных, отбирая наиболее информативные для модели примеры, и успешно применяется в областях: компьютерное зрение [7], обработка естественного языка [8], системы рекомендаций [9]. Однако его интеграция в контекстное обучение с большими языковыми моделями малоизучена [10].

Мы предлагаем гибридный алгоритм, сочетающий активное и безнадзорное обучение для оптимизации выбора примеров в контекстном обучении больших языковых моделей при классификации намерений. Алгоритм учитывает особенности контекстного обучения и интегрирует оценку неопределенности модели, повышая эффективность использования ограниченного бюджета разметки данных.

Некоторые исследования улучшают контекстное обучение оптимизацией выбора примеров. В [11] рассматриваются принципы активного обучения в этом кон-

тексте, оцениваются стратегии выборки данных. В [12] анализируются zero-shot и few-shot техники для классификации намерений. Однако эти подходы не учитывают специфику выбора примеров при ограниченном бюджете разметки и не интегрируют оценку неопределенности модели.

Наш подход комбинирует кластеризацию данных для учета их структуры и активное обучение с оценкой неопределенности модели для отбора информативных примеров. Подобные идеи были в [13] с методом ActiveLLM для активного обучения больших языковых моделей, но без фокуса на классификации намерений и специфики выбора примеров в контексте.

Цель работы — продемонстрировать эффективность предложенного алгоритма на реальных данных и сравнить с современными методами, включая большие языковые модели в zero-shot и few-shot режимах. Мы рассматриваем GPT-4 [14] и Gemma 2 27b [15], показывающие хорошие результаты на русском языке, и анализируем их производительность в сравнении с нашим подходом.

Основная часть

Контекстное обучение и классификация намерений

Контекстное обучение (in-context learning) [2] представляет собой подход, при котором большая языковая модель способна выполнять новую задачу, опираясь на несколько примеров (контекст), предоставленных в виде входных данных, без явного обновления весов модели. Это особенно полезно в тех ситуациях, когда обучение или дообучение модели на новых данных затруднительно или нежелательно.

В задаче классификации намерений цель состоит в том, чтобы на основании текстового сообщения пользователя определить его намерение из заданного на-

бора классов. Пусть имеется множество классов $Y = \{y_1, y_2, \dots, y_c\}$ и набор неразмеченных пользовательских сообщений $U = \{x_1, x_2, \dots, x_n\}$. Требуется выбрать небольшое подмножество размеченных примеров $L = \{(x_i, y_i)\}$, которое будет использоваться в контексте для контекстного обучения модели для максимизации качества классификации.

Активное обучение для выбора примеров

Активное обучение [6] направлено на эффективный отбор наиболее информативных для модели примеров при ограниченном бюджете на разметку. Традиционные стратегии активного обучения включают отбор примеров с максимальной неопределенностью модели [16], использование кластеризации для представления структуры данных [17], а также методы, учитывающие разнообразие и репрезентативность выбранных примеров [18]. Рисунок 1 иллюстрирует процесс активного обучения с отбором объектов из выборки.

В задаче контекстного обучения с большими языковыми моделями выбор примеров приобретает особую значимость, поскольку количество примеров в контексте ограничено (обычно от 1 до 50), и от их качества зависит способность модели правильно обобщать на новые данные [4]. Кроме того, необходимо учитывать, что не все примеры оказывают одинаковое влияние на модель, и выбор нерепрезентативных или избыточных примеров может снизить эффективность контекстного обучения [19]. Активный отбор примеров для контекстного обучения можно рассматривать как задачу активного обучения объектов из выборки с одной итерацией, представлено на Рисунке 2.

Для классификации тестового примера алгоритм отбирает из размеченных данных наиболее релевантные примеры, которые используются в качестве контекста в затравке (prompt) для обучения модели. Эти примеры, включенные в контекст, позволяют языковой модели

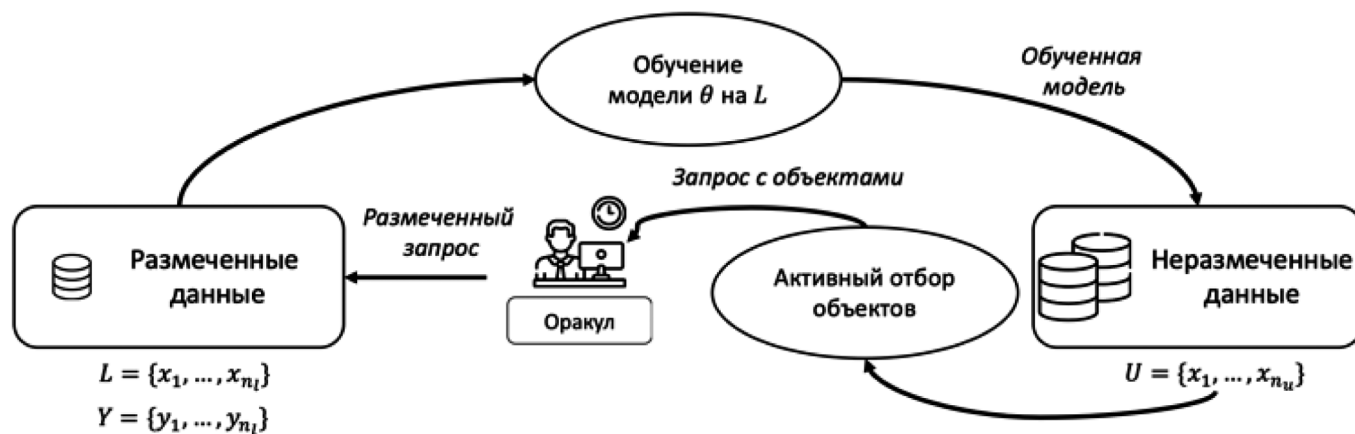


Рис. 1. Схема работы активного обучения



Контекстное обучение с активно отбираемыми примерами

Рис. 2. Активное контекстное обучение с одной итерацией

предсказывать класс тестового примера на основе предоставленного окружения. Качество отобранных примеров существенно влияет на точность классификации и общее качество работы модели.

Для эффективного активного обучения важно уметь оценивать неопределенность модели в отношении неразмеченных примеров. Оценка неопределенности позволяет выбирать примеры, по которым модель «сомневается» больше всего, тем самым повышая прирост знаний при разметке этих примеров [16]. Оценка неопределенности в контекстном обучении измеряется через перплексию, которая отражает уверенность модели в предсказании текста без меток. Перплексия рассчитывается как экспонента средней логарифмической вероятности токенов последовательности, включая ввод, но без метки, предсказанных моделью [20].

Постановка задачи

В задаче классификации намерений с использованием больших языковых моделей важно не только точное определение намерений, но и оптимальный выбор контекста для контекстного обучения. Этот подход предполагает, что модель получает несколько примеров, релевантных тестовому примеру, что помогает ей обобщать на новые данные без изменения своих параметров. Цель — выбрать небольшое подмножество наиболее значимых и репрезентативных примеров из размеченной выборки L для улучшения классификации намерений в тестовом примере.

Пусть имеется:

- Множество классов $Y = \{y_1, y_2, \dots, y_C\}$, где каждый класс соответствует определённому намерению пользователя.
- Набор неразмеченных сообщений пользователей $U = \{x_1, x_2, \dots, x_N\}$, которые модель должна уметь классифицировать.
- Размеченное множество $L = \{(x_i, y_i) \mid x_i \in R^{EMB_DIM}, y_i \in Y\}$, содержащее пары «сообщение–намерение», где x_i — векторное представление сообщения.

В задаче контекстного обучения необходимо выбрать небольшое подмножество $S \subset L$, состоящее из K примеров, которые будут представлены модели в качестве контекста для классификации нового тестового примера x_{test} . Это подмножество должно быть таким, чтобы оно максимально способствовало правильной классификации намерений для x_{test} , учитывая, что размер контекста ограничен.

Особенности задачи:

- **Ограничение на размер контекста:** Основное ограничение — количество примеров K в контексте; бюджет разметки не ограничен, так как выбор идёт из уже размеченной выборки.
- **Релевантность к тестовому примеру:** Примеры в контексте должны быть релевантны x_{test} . Учитывается семантическая близость между примерами в L и x_{test} .
- **Оценка неопределённости модели:** Для повышения эффективности отбора оценивается степень неопределённости модели по каждому примеру.

Таким образом, задача активного выбора примеров для контекстного обучения направлена на формирование оптимального подмножества S из L , позволяющего модели эффективно обобщать на x_{test} и минимизировать ошибки в классификации намерений.

Предлагаемый подход

В статье предлагается гибридный итеративный метод активного обучения для задачи классификации намерений с использованием больших языковых моделей. Полный алгоритм был предложен в работе 24. В текущей постановке метод адаптирован для задачи выбора контекстных примеров (in-context examples), где целью является подбор K примеров из размеченной выборки, которые обеспечат максимальное качество классификации для модели в контексте обучения на тестовом примере.

Алгоритм состоит из нескольких этапов:

1. **Инициализация:** задаются параметры алгоритма — область поиска для кластеризации (ϵ), минимальное число объектов для кластеров (minPts), порог сходства (Sim), количество отбираемых примеров K и тестовый пример (test_example). Эти параметры устанавливают начальные условия и настраивают процесс кластеризации.
2. **Кластеризация данных:** применяется алгоритм DBSCAN к размеченной выборке для выявления семантически связанных групп объектов, исключая шумовые данные. Если значимые кластеры не обнаружены, параметры ϵ и minPts корректируются для учета менее плотных областей данных.
3. **Оценка значимости кластеров:** для каждого кластера вычисляются три метрики:
 - Степень неуверенности модели для объектов кластера (энтропия классификации).
 - Расстояние до тестового примера (косинусное сходство векторов).
 - Размер кластера (число объектов).
 Метрики нормализуются и преобразуются в ранги; итоговый рейтинг кластера рассчитывается как среднее этих рангов.
4. **Отбор ключевых точек:** из кластеров с высоким рейтингом отбираются ключевые точки, наиболее релевантные тестовому примеру. Они ранжируются по значимости, и для каждой вычисляется сходство с тестовым примером. Выбираются K объектов с наибольшей семантической близостью, формируя динамический контекст, оптимизированный под конкретный пример.
5. **Обучение модели:** выбранный контекст используется для дообучения модели на размеченной выборке, что позволяет модели адаптироваться к текущему тестовому примеру и точнее его интерпретировать.

В отличие от исходной задачи с ограничением бюджета аннотаций, здесь цель — динамически подобрать K наиболее значимых примеров для формирования контекста, улучшая классификацию намерений для каждого тестового примера.

Результаты экспериментов

Эффективность предлагаемого алгоритма оценивалась на датасете MASSIVE Intent, содержащем более 1 миллиона запросов с аннотациями намерений на 51 языке, включая русский. Этот набор данных предоставляет разнообразные задачи по пониманию естественного языка, позволяя моделям оценивать производительность в условиях многоязычия и различных сценариев.

В исследовании использовались две современные модели: **GPT-4** и **Gemma 2 27B**.

Для подготовки данных были получены эмбединги всех неразмеченных примеров с помощью модели MUSE (Lample et al., 2017), предоставляющей многоязычные векторные представления текстовых данных.

Задача исследования — классификация намерений в пользовательских запросах. Для оценки качества использовались метрики:

- **Точность классификации намерений.**
- **Макро F1-мера**, учитывающая несбалансированность классов.

Эксперименты проводились с различным количеством контекстных примеров ($k = 10, 30, 100$) для оценки влияния объёма контекста на производительность классификации.

Использовались следующие методы выбора контекстных примеров:

- **Zero-shot:** модель без предоставления примеров.
- **Случайный выбор:** случайные k примеров из размеченного пула.
- **Отбор по неуверенности:** выбор наиболее неопределённых примеров по энтропии модели.
- **Предлагаемый алгоритм:** гибридный подход, учитывающий неопределённость и структуру данных.

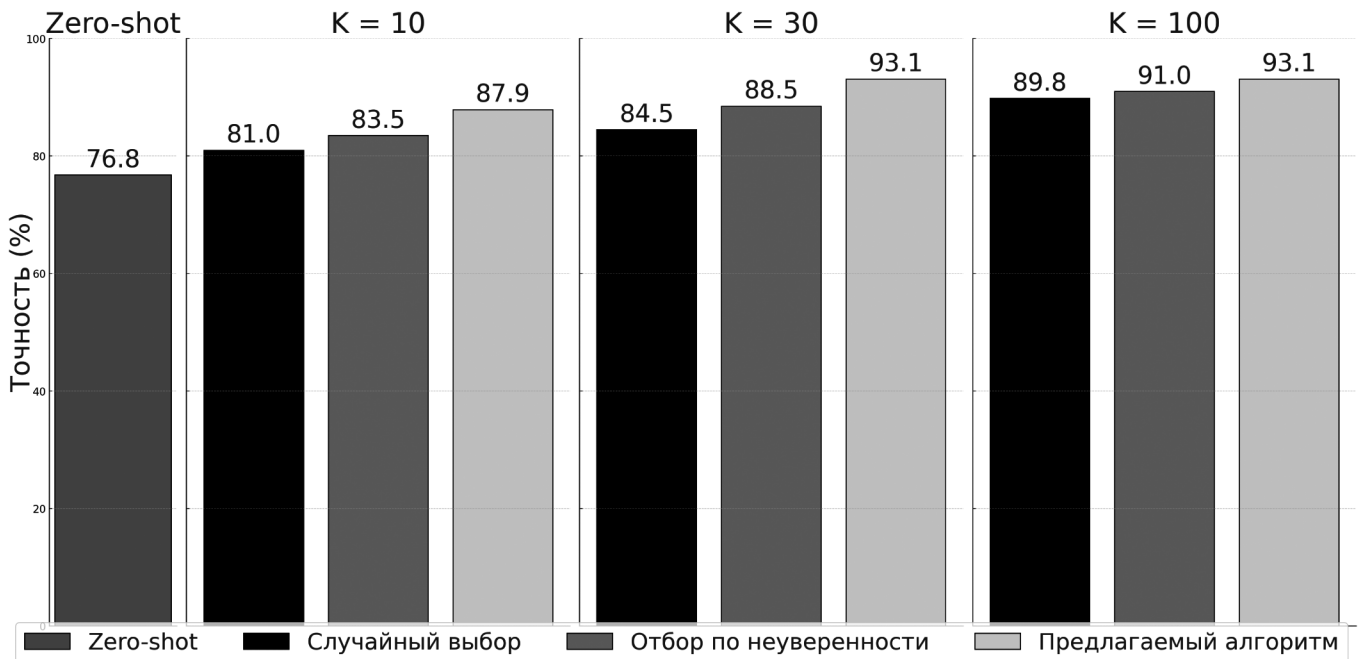
Параметры экспериментов включали значения k : 10, 30, 100. Всего проведено 72 эксперимента, учитывая комбинации 2 моделей, 3 датасетов, 4 методов выбора примеров и 3 значений k .

Такой набор экспериментов оценивает различные подходы активного отбора примеров для контекстного обучения в задаче классификации намерений, как для закрытых, так и открытых многоязычных моделей. Также сравнивается обучение без примеров с обучением на разном количестве примеров и эффективность классических методов активного обучения при отборе из неразмеченных данных.

Эксперименты показали значительное улучшение производительности классификации намерений при использовании предлагаемого алгоритма. На моделях GPT-4 и Gemma 2 27B при $k = 10, 30, 100$ наш метод стабильно превосходил случайный выбор и отбор по неуверенности. Наибольший прирост точности достигался при увеличении k до 100, что подтверждает эффективность алгоритма в выборе информативных примеров, особенно важно при ограниченном числе доступных данных.

Сравнение методов отбора показало, что наш гибридный подход, сочетающий кластеризацию и оценку неопределённости, превосходит классические методы активного обучения. Алгоритм обеспечивает репрезен-

Точность для GPT-4 с Zero-shot и Few-shot с разным количеством примеров



Точность для Gemma 2 27b с Zero-shot и Few-shot с разным количеством примеров

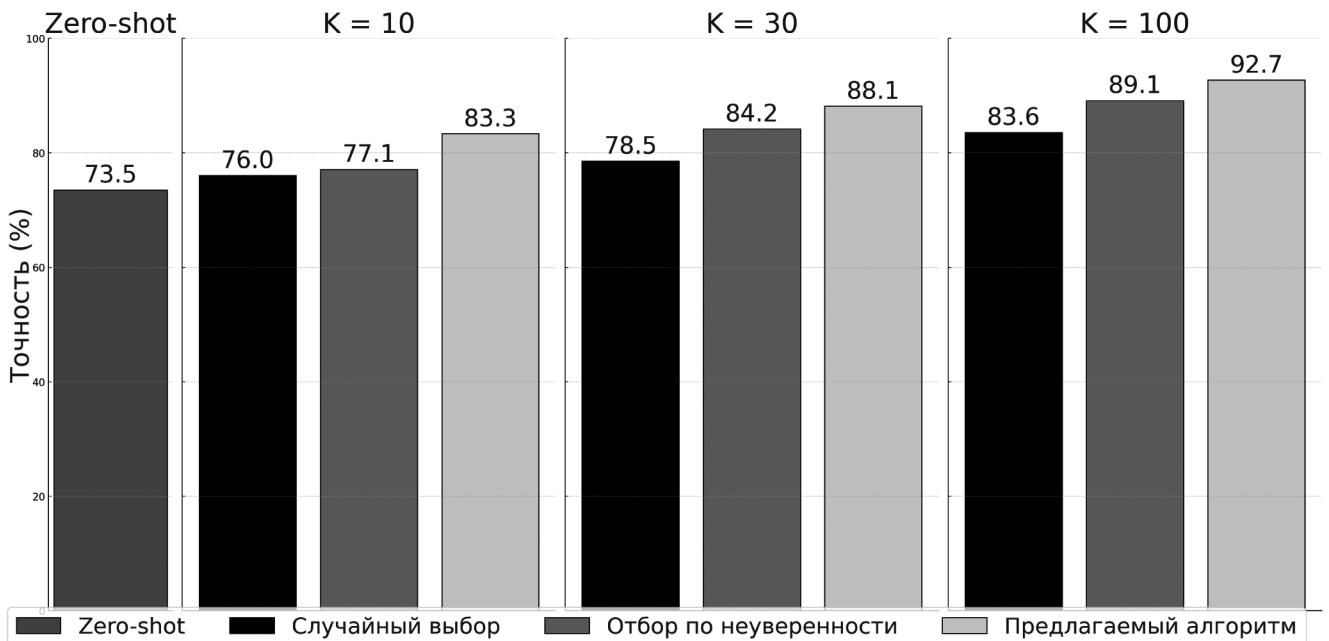


Рис. 3. Сравнение точности различных методов выбора контекстных примеров в задачах классификации намерений для моделей GPT-4 и Gemma 2 27b

тативность и разнообразие примеров, учитывая неопределённость. Эксперименты на многоязычном датасете MASSIVE Intent подтвердили, что наш подход стабильно улучшает производительность классификации намерений в условиях многоязычия, делая его подходящим для реальных приложений, таких как виртуальные ассистенты и диалоговые системы.

Выводы

В данной работе предложен новый гибридный алгоритм выбора примеров для контекстного обучения больших языковых моделей, который сочетает оценку неопределенности и структурный анализ данных. Этот подход позволяет существенно повысить точность клас-

сификации намерений в условиях ограниченного числа примеров, особенно по сравнению с методами случайного выбора и отбора по неуверенности. Эксперименты показали, что предлагаемый метод обеспечивает лучшие результаты на задачах классификации намерений как

для GPT-4, так и для Gemma 2 27B, что подчеркивает его эффективность для широкого спектра языковых задач.

Благодарность. Работа выполнена под руководством кандидата технических наук, А.С. Кожаринова.

ЛИТЕРАТУРА

1. Devlin J. et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805. 2018.
2. Brown T.B. et al. Language Models are Few-Shot Learners. arXiv preprint arXiv:2005.14165. 2020.
3. Chowdhery A. et al. PaLM: Scaling Language Modeling with Pathways. arXiv preprint arXiv:2204.02311. 2022.
4. Liu H. et al. ILLUMINER: Instruction-tuned Large Language Models as Few-shot Intent Classifier and Slot Filler. arXiv preprint arXiv:2301.12825. 2024.
5. Sahu G. et al. Data Augmentation for Intent Classification with Off-the-shelf Large Language Models. Proceedings of the 4th Workshop on NLP for Conversational AI. 2022.
6. Settles B. Active Learning Literature Survey. University of Wisconsin-Madison. 2009.
7. Gal Y., Ghahramani Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. Proceedings of the 33rd International Conference on Machine Learning. 2016.
8. Tong S., Koller D. Support vector machine active learning with applications to text classification. Journal of machine learning research. 2001. — Т. 2. — №. Nov. — С. 45–66.
9. Bu Y., Small K. Active Learning in Recommendation Systems with Multi-level User Preferences. Proceedings of the 2018 IEEE International Conference on Big Data. 2018.
10. Margatina K. et al. Active Learning Principles for In-Context Learning with Large Language Models. arXiv preprint arXiv:2302.12246. 2023.
11. Parikh S. et al. Exploring Zero and Few-shot Techniques for Intent Classification. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics. 2023.
12. Muldrew W. et al. Active Preference Learning for Large Language Models. arXiv preprint arXiv:2301.11270. 2024.
13. Wang T., Chen S., Jia R. One-round Active Learning. arXiv preprint arXiv:2104.11843. 2021.
14. OpenAI. GPT-4 Technical Report. arXiv preprint arXiv:2303.08774. 2023.
15. Gemma Team. Gemma: Open Models Based on Gemini Research and Technology. arXiv preprint arXiv:2403.08295. 2024.
16. Lewis D.D., Gale W.A. A sequential algorithm for training text classifiers. Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval. New York, NY, USA: ACM. 1994.
17. Zhdanov F. Diverse mini-batch active learning. arXiv preprint arXiv:1901.05954. 2019.
18. Huang S.-J. Active Learning by Querying Informative and Representative Examples. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2014.
19. Liu Y. et al. What Makes Good In-Context Examples for GPT-3? arXiv preprint arXiv:2101.06804. 2021.
20. Gonen H., Iyer S., Blevins T., Smith N.A., Zettlemoyer L. Demystifying Prompts in Language Models via Perplexity Estimation. arXiv preprint arXiv:2212.04037. 2022.
21. Ester M. et al. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. 1996.
22. Coucke A. et al. Snips Voice Platform: an embedded Spoken Language Understanding system for private-by-design voice interfaces. arXiv preprint arXiv:1805.10190. 2018.
23. Larson S. et al. An Evaluation Dataset for Intent Classification and Out-of-Scope Prediction. Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. 2019.
24. Дранга Д.И. Гибридный алгоритм активного обучения для определения намерений в диалоговых системах с неразмеченными и шумными данными // Вестник Череповецкого государственного университета. 2023. № 5 (116). С. 39–51.

© Дранга Даниил Игоревич (dranga.daniil@gmail.com)

Журнал «Современная наука: актуальные проблемы теории и практики»