

адаптивных алгоритмов машинного обучения является перспективным направлением для развития мобильных приложений в сфере ментального здоровья.

Литература

1. Иванов, П. К. Кроссплатформенная разработка мобильных приложений: сравнительный анализ технологических решений / П. К. Иванов, М. В. Сидорова // Вестник компьютерных технологий. – 2023. – № 4 (12). – С. 45–52.
2. Петров, А. С. Применение методов машинного обучения в мобильных приложениях для здоровья / А. С. Петров // Цифровая медицина. – 2022. – Т. 15, № 3. – С. 78–85.
3. Козлова, Е. В. Методы анализа пользовательской активности в health-приложениях / Е. В. Козлова, Д. С. Новиков // Информационные технологии в здравоохранении. – 2024. – № 1 (8). – С. 112–125.

ПРОЕКТИРОВАНИЕ И РЕАЛИЗАЦИЯ ВЫСОКОПРОИЗВОДИТЕЛЬНОЙ СИСТЕМЫ ХРАНЕНИЯ И ИНДЕКСИРОВАНИЯ НЕЙРОДАНЫХ С СЕМАНТИЧЕСКИМ ПОИСКОМ

Е. А. Клецко, Н. С. Иванченко, П. В. Травничева

*Витебский государственный университет имени П. М. Машерова,
Республика Беларусь*

Рассмотрен процесс проектирования и реализации высокопроизводительной системы хранения и индексации нейроданных, обеспечивающей масштабируемость, низкую задержку при операциях чтения/записи и эффективный семантический поиск по векторным представлениям. Основное внимание уделено вопросам выбора технологий векторного поиска, проектирования распределенной архитектуры и реализации интер-фейса управления.

Ключевые слова: нейроданные, векторная база данных, семантический поиск, HNSW, ChromaDB, FastAPI, Kubernetes.

DESIGN AND IMPLEMENTATION OF A HIGH-PERFORMANCE STORAGE AND INDEXING SYSTEM FOR NEURODATA WITH SEMANTIC SEARCH

E. A. Kletsko, N. S. Ivanchenko, P. V. Travnicheva

Vitebsk State University named after P. M. Masherov, Republic of Belarus

The paper discusses the design and implementation of a high-performance storage and indexing system for neurodata, providing scalability, low read/write latency, and efficient semantic search over vector representations. The main attention is paid to the selection of vector search technologies, the design of a distributed architecture, and the implementation of a management interface.

Keywords: neurodata, vector database, semantic search, HNSW, ChromaDB, FastAPI, Kubernetes.

В современную эпоху больших данных и искусственного интеллекта нейросетевые модели становятся неотъемлемой частью множества сфер деятельности. С увеличением сложности и масштабов моделей экспоненциально растут объемы нейроданных – весов, эмбедингов и метаданных. Это создает потребность в эффективных методах их хранения, индексации и семантического поиска, что особенно актуально для задач дообучения моделей, управления версиями и MLOps. Современные системы хранения данных, такие как реляционные и NoSQL базы данных [1], не всегда способны эффективно рабо-

тать с высокоразмерными векторными представлениями и выполнять семантический поиск по сходству. В связи с этим возникает необходимость в разработке специализированного программного обеспечения, способного обеспечить высокопроизводительное хранение и интеллектуальный поиск по нейроданным.

Цель работы заключалась в проектировании и реализации высокопроизводительной системы хранения и индексации нейроданных, обеспечивающей масштабируемость, низкую задержку при операциях чтения/записи, эффективный семантический поиск по векторным представлениям и универсальность применения в различных конвейерах машинного обучения.

Методика проектирования включала анализ существующих технологий и практическую реализацию системы. В результате изучения были выбраны следующие технологии: ChromaDB / Weaviate; FastAPI; Docker & Kubernetes.

ChromaDB / Weaviate: эти векторные базы данных были выбраны за их открытость, производительность и богатый функционал для семантического поиска. Они предоставляют эффективные методы индексации, такие как HNSW, что критически важно для работы с высокоразмерными данными [2].

FastAPI: для создания высокопроизводительного API был выбран фреймворк FastAPI. Этот фреймворк обеспечивает простоту разработки, автоматическую генерацию документации и асинхронную обработку запросов, что снижает задержки при обслуживании клиентов.

Docker & Kubernetes: технологии контейнеризации и оркестрации были выбраны за свою способность обеспечить масштабируемость, отказоустойчивость и простоту развертывания системы в различных средах [3].

Интерфейс приложения для управления нейроданными разработан с акцентом на удобство и интуитивность использования. Реализованы функции загрузки моделей (весов и эмбеддингов), семантического поиска по векторным представлениям и интерактивная панель мониторинга для отображения метрик производительности системы, статистики использования и результатов поисковых запросов.

Полученные результаты показывают, что использование современных векторных баз данных и микросервисной архитектуры позволяет создать масштабируемое и производительное решение для хранения и интеллектуального поиска по нейроданным, что является важным шагом в развитии инфраструктуры для машинного обучения и MLOps.

Разработана высокопроизводительная и масштабируемая система для хранения и семантического поиска по нейроданным. Реализованная архитектура демонстрирует возможность создания эффективных решений для управления большими объемами векторных данных в конвейерах машинного обучения. Перспективным направлением дальнейших исследований является оптимизация алгоритмов индексации и повышение эффективности распределенного поиска.

Литература

1. Костюков, И. В. Современные системы управления базами данных / И. В. Костюков. – М. : Наука, 2020.
2. Шафеев, А. П. Векторная семантика и поиск в больших данных / А. П. Шафеев, Д. В. Соловьев. – СПб. : БХВ-Петербург, 2022.
3. Лебедев, С. Н. Контейнеризация и оркестрация приложений с использованием Docker и Kubernetes / С. Н. Лебедев. – М. : ДМК Пресс, 2021.