

Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет
информатики и радиоэлектроники

УДК 004.6:004.032.26

Шараев
Евгений Владимирович

Оптимизация планировщика запросов к БД посредством алгоритмов
машинного обучения

АВТОРЕФЕРАТ

на соискание степени магистра информатики и вычислительной техники

по специальности 1-40 81 02 Технологии виртуализации и облачных
вычислений

Научный руководитель
Луцик Юрий Александрович
к.т.н., доцент

Минск 2019

КРАТКОЕ ВВЕДЕНИЕ

Использование реляционных систем управления базами данных (РСУБД) является крайне распространенным способом хранения информации.

Современные РСУБД предоставляют множество различных конфигураций и функциональных преимуществ, включая определенную универсальность для описания различных моделей данных, достаточно высокую скорость запросов доступа и управления данными, высокую надёжность, отказоустойчивость, различные средства для восстановления после сбоя и так далее.

Для доступа, добавления и изменения данных используется язык структурированных запросов (SQL). Поскольку SQL является декларативным языком, база данных сама должна выбрать способ, который будет использован при осуществлении доступа к данным. Существует множество подходов для улучшения производительности запросов, включающих сбор дополнительной статистики или даже применение подсказок планировщику запросов. Вместе с этим существует возможность применения альтернативного подхода с применением алгоритмов машинного обучения.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Целью данной работы является реализация модуля оптимизации запросов к реляционной системе управления базами данных PostgreSQL. В работе решаются задачи подбора наиболее оптимальных алгоритмов машинного обучения с точки зрения оптимизации запросов к базе данных, задачи подбора оптимальной архитектуры модуля оптимизации, а также проведение тестирования готовой реализации.

Объектом данного исследования является реляционная система управления базами данных PostgreSQL. Предметом исследования является оптимизация планировщика запросов к реляционной системе управления базами данных PostgreSQL.

Практическая значимость работы заключается в возможности повышения предельной скорости выполнения запросов на извлечение данных. Теоретическая значимость заключается в рассмотрении более широкого набора способов получения признаков, чем в существующих решениях, приведении замечаний о возможностях улучшения существующих решений, а также реализации части модуля, отвечающего за работу с алгоритмами машинного обучения, отдельно от сервера базы данных, что может упростить дальнейшие исследования возможных подходов к оптимизации и к работе с признаками.

Личный вклад заключается в разработке архитектуры модуля оптимизации, его реализации и проведении тестирования производительности полученной системы посредством использования стандартизированных бенчмарков и тестов, специально составленных для реализации.

Результаты работы опубликованы в сборниках материалов конференций:

1. Информационные технологии и системы, Минск, 2018.
2. BIG DATA and Advanced Analytics, Минск, 2019.
3. CSITIC, Харьков, 2019.
4. Научному прогрессу - творчество молодых, Йошкар-Ола, 2019.

Общий объем работы составляет 66 страниц, из которых основного текста – 57 страниц, список использованных источников из 38 наименований и 2 приложения.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

В первой главе данной работы обсуждаются существующие решения и подходы к оптимизации планировщика запросов к реляционной системе управления базами данных. В ней приведены как подходы к оптимизации для PostgreSQL, так и к некоторым другим реляционным системам управления базами данных.

Во второй главе приводится обоснование выбора реляционной системы управления базами данных PostgreSQL, а также описание возможных подходов к реализации модуля оптимизации планировщика.

В третьей главе приводится описание возможных подходов к получению признаков, которые будут использованы для работы модуля оптимизации. Были проанализированы преимущества и недостатки подходов в зависимости от типов данных и операторов в запросах. Также были проанализированы возможные решения для их преобразования для повышения качества работы алгоритмов машинного обучения.

В четвёртой главе приводится описание возможных алгоритмов машинного обучения, их основные характеристики, достоинства, недостатки и возможности адаптации к изменениям данных. Кроме того, были проанализированы некоторые способы увеличения скорости работы алгоритмов посредством применения аппроксимаций.

Пятая глава содержит основную информацию о тестировании готового модуля в рамках реляционной системы управления базами данных PostgreSQL. В ней приведены основные принципы проведения тестирования на производительность, перечислены утилиты, позволяющие автоматизировать процесс тестирования, указаны возможные метрики для оценки производительности модуля, а также приведены результаты работы модуля оптимизации на некоторых микробенчмарках и стандартизированных тестах.

ЗАКЛЮЧЕНИЕ

В рамках работы были проанализированы возможные подходы для реализации модуля оптимизации планировщика запросов к базам данных.

Была предложена архитектура, которая позволяет разделить модуль, осуществляющий оптимизацию, и сервер базы данных, что должно обеспечить большую масштабируемость и более высокую производительность.

Был проведен сравнительный анализ алгоритмов машинного обучения, которые могут быть использованы для оптимизации запросов посредством предсказания значений мощности. Были проанализированы возможные подходы к получению признаков описаний объектов, их обработке и масштабированию.

Также была продемонстрирована эффективность использования алгоритмических композиций с применением метамоделей-регрессоров и смесей экспертов.

Были проведены испытания с использованием микробенчмарков и стандартизированных тестов на производительность, в результате которых были получены экспериментальные подтверждения эффективности реализованного модуля. Повышение скорости составило около 20-40% в среднем, однако для некоторых видов запросов увеличение скорости выполнения происходит в несколько раз.

СПИСОК ПУБЛИКАЦИЙ АВТОРА

[1-А.] Шараев, Е. В. Оптимизация запросов к базе данных с помощью алгоритмов искусственного интеллекта / Е. В. Шараев // Информационные технологии и системы 2018 – Минск, 2018. – С. 136 – 137.

[2-А.] Sharayeu, Y. Machine learning and lambda architecture methods application to optimize postgresql database performance / Y. Sharayeu // BIG DATA and Advanced Analytics – Минск, 2019. – С. 80 – 81.

[3-А.] Sharayeu, Y. Approaches on benchmarking postgresql optimized with machine learning Approaches on benchmarking postgresql optimized with machine learning/ Е. В. Шараев // Y. Sharayeu – Харьков, 2019. – С. 42 – 43.

[4-А.] Шараев, Е. В. Использование алгоритмических композиций при оптимизации PostgreSQL методами машинного обучения / Е. В. Шараев // Научному прогрессу - творчество молодых 2019 – Йошкар-Ола, 2019.