

На правах рукописи

Азимов Рустам Шухратуллович

Алгоритмы выполнения навигационных запросов к
графам использующие операции линейной алгебры

Специальность 05.13.11 —
Математическое и программное обеспечение вычислительных
машин, комплексов и компьютерных сетей

Автореферат
диссертации на соискание учёной степени
кандидата физико-математических наук

Санкт-Петербург — 2022

Работа выполнена на кафедре системного программирования Санкт-Петербургского государственного университета

Научный руководитель: кандидат физико-математических наук, доцент
Кознов Дмитрий Владимирович

Официальные оппоненты: Оппонент1,
доктор физико-математических наук, профессор,
федеральное государственное бюджетное учреждение науки Институт систем информатики им.
А.П. Ершова Сибирского отделения Российской академии наук, директор

Оппонент2,
кандидат технических наук, доцент,
федеральное государственное автономное образовательное учреждение высшего образования
“Санкт-Петербургский политехнический университет Петра Великого”, исполняющий обязанности заведующего кафедрой

Ведущая организация: Федеральное государственное бюджетное учреждение науки Институт системного программирования Российской академии наук (ИСП РАН)

Защита состоится _____ г. в _____ часов на заседании диссертационного совета Д 212.232.51 на базе Санкт-Петербургского государственного университета по адресу: 198504, Санкт-Петербург, Петродворец, Университетский пр., 28, математико-механический факультет, ауд. 405.

С диссертацией можно ознакомиться в Научной библиотеке Санкт-Петербургского государственного университета по адресу: 199034, Санкт-Петербург, Университетская наб., д. 7/9, а также на сайте <http://spbu.ru/science/disser/>.

Автореферат разослан _____ 20____ года

Ученый секретарь

диссертационного совета

Д 212.232.51, д.ф.-м.н., профессор

Демьянович Юрий Казимирович

Общая характеристика работы

Актуальность темы исследования

Графы используются в качестве структуры данных для представления больших объемов информации в компактной и удобной для анализа форме во многих областях, например, в биоинформатике, в графовых базах данных, при статическом анализе программ. При этом оказывается необходимым выявлять сложные зависимости между вершинами графа, которые выражаются путями между ними. Для этого необходимо решать задачу достижимости с определенными ограничениями на пути в графе, которая отвечает на вопрос о существовании путей, удовлетворяющих данным ограничениям между каждой парой вершин. Кроме того, в некоторых областях, в качестве доказательства существования таких путей необходимо предъявить все или хотя бы один из них.

Для описания ограничений естественно помечать ребра графа символами из некоторого конечного алфавита и выделять пути с помощью формальных грамматик над тем же алфавитом (регулярные выражения, контекстно-свободные грамматики). В настоящее время активно исследуются ограничения в виде контекстно-свободных (КС) грамматик, так как они позволяют описывать более широкий класс запросов, чем регулярные выражения.

Однако большинство существующих алгоритмов в данной области имеют низкую производительность на больших графах, что затрудняет их анализ. Хорошую производительность могут показать алгоритмы, использующие матричный подход к поиску путей в графе с КС ограничениями. При активном использовании матричных операций линейной алгебры в процессе анализа графов, может быть применен широкий класс матричных оптимизаций, например, разреженное представление матриц и параллельное вычисление. На текущий момент существующий матричный алгоритм в данной области позволяет лишь решить задачу достижимости, т.е. установить факт наличия между двумя вершинами пути определенного вида, при этом сам путь не предоставляется.

Кроме того, в данной области не исследована возможность использования более сложных матричных операций, таких как, например, произведение Кронекера. Данное направление интересно тем, что позволяет оперировать матрицами большего размера и уменьшить общее количество операций над ними. Это может позволить получить еще больший прирост производительности после применения матричных оптимизаций.

Таким образом, для большинства типов ограничений необходима разработка алгоритмов поиска путей, использующих различные операции линейной алгебры и имеющих высокую производительность на больших графах.

Степень разработанности темы исследования

TODO

Объект исследования

Объектом исследования являются алгоритмы выполнения навигационных запросов к графам, использующие операции линейной алгебры.

Цель и задачи диссертационной работы

Целью данной работы является разработка алгоритмов выполнения навигационных запросов к графам, использующих операции линейной алгебры и имеющих высокую производительность на больших графах.

Достижение поставленной цели обеспечивается решением следующих задач.

1. Разработать подход к вычислению КС-запросов к графам, использующий операции линейной алгебры.
2. Разработать алгоритм вычисления КС-запросов к графам, использующий предложенный подход и позволяющий предоставлять по одному искомому пути для каждой пары вершин, если они существуют.
3. Разработать алгоритм вычисления КС-запросов к графам, использующий предложенный подход и позволяющий предоставлять все искомые пути для каждой пары вершин.
4. Разработать алгоритм вычисления КС-запросов к графам, использующий предложенный подход и произведение Кронекера в качестве основной операции линейной алгебры.

Методология и методы исследования

TODO

Положения, выносимые на защиту

1. Разработан подход к вычислению КС-запросов к графам, использующий операции линейной алгебры.
2. Разработан алгоритм вычисления КС-запросов к графам, использующий предложенный подход и позволяющий предоставлять по одному искомому пути для каждой пары вершин, если они существуют. Доказана завершаемость и корректность предложенного алгоритма. Проведено экспериментальное исследование.

3. Разработан алгоритм вычисления КС-запросов к графам, использующий предложенный подход и позволяющий предоставлять все искомые пути для каждой пары вершин. Доказана завершаемость и корректность предложенного алгоритма. Проведено экспериментальное исследование.
4. Разработан алгоритм вычисления КС-запросов к графам, использующий предложенный подход и произведение Кронекера в качестве основной операции линейной алгебры. Доказана завершаемость и корректность предложенного алгоритма. Проведено экспериментальное исследование.

Научная новизна

Научная новизна полученных в ходе исследования результатов заключается в следующем.

1. Подход, предложенный в диссертации, отличается от аналогов (работы Семёна Григорьева, Джелле Хеллингса) активным использованием матричных операций в процессе вычисления запросов. Это позволяет как применять широкий класс матричных оптимизаций и получать высокую производительность при работе с большими графами.
2. Алгоритм, предложенный в диссертации, отличается от аналогов (работы Семёна Григорьева, Джелле Хеллингса, Сяованга Чжана) активным использованием матричных операций в процессе вычисления запросов и (формулировка улучшения собственного результата?) отличается от матричного алгоритма Азимова Рустама возможностью построения по одному искомому пути для каждой пары вершин, если они существуют. Это позволяет как применять широкий класс матричных оптимизаций, так и предъявлять пути в качестве доказательства отношения определенного вида между парами вершин, что является важным результатом анализа во многих областях.
3. Алгоритм, предложенный в диссертации, отличается от аналогов (работы Семёна Григорьева, Джелле Хеллингса, Сяованга Чжана) активным использованием матричных операций в процессе вычисления запросов и (формулировка улучшения собственного результата?) отличается от матричного алгоритма Азимова Рустама возможностью построения всех искомых путей для каждой пары вершин.
4. Алгоритм, предложенный в диссертации, отличается от аналогов (работы Семёна Григорьева, Джелле Хеллингса, Сяованга Чжана) активным использованием матричных операций в процессе вычисления запросов и (формулировка улучшения собственного результата?) отличается

от матричного алгоритма Азимова Рустама использованием в процессе вычисления запросов произведения Кронекера и представлением КС-грамматики запроса в виде рекурсивного автомата. Это позволяет улучшить производительность на больших графах за счёт оптимизаций над матрицами большего размера и уменьшения количества операций над ними.

Теоретическая и практическая значимость работы

TODO

Степень достоверности и апробация результатов

Достоверность и обоснованность результатов исследования опирается на использование формальных методов исследуемой области, выполнение формальных доказательств и инженерные эксперименты.

Основные результаты работы были доложены на ряде международных научных конференций: GRADES'20, ADBIS'20, SIGMOD'21 (еще не приняли). Дополнительной апробацией является то, что разработка предложенных алгоритмов была поддержана грантом РФФИ №18-11-00100 и грантом РФФИ №19-37-90101.

Публикации по теме диссертации

Все результаты диссертации изложены в 4 [1–4] научных работах, которые содержат основные результаты работы и индексируются Scopus. Работы [1–3] написаны в соавторстве. В [1] Р. Азимову принадлежит разработка алгоритма, доказательство его корректности и завершаемости, работа над текстом. В [2, 3] Р. Азимову принадлежит работа над доказательствами корректности и завершаемости алгоритма, работа над текстом.

Работа [4] еще не написана.

Публикация [5] ВАК и SCOPUS матричный алгоритм для конъюнктивных грамматик. Статья [6] SCOPUS первоначальный матричный алгоритм GRADES'18. Статья [7] ВАК конъюнктивные труды ИСП РАН. Статья [8] РИНЦ про конъюнктивные и выступление на конференции PLC'17.

Список литературы

1. Азимов Р. Ш. Context-Free Path Querying with Single-Path Semantics by Matrix Multiplication / Терехов А., Хорошев А., Азимов Р., Григорьев С. // In Proceedings of the 3rd Joint International Workshop on Graph Data

2. Азимов Р. Ш. Context-Free Path Querying by Kronecker Product / Орачев Е., Эпельбаум И., Азимов Р., Григорьев С. // In Proceedings of the 24th European Conference on Advances in Databases and Information Systems (ADBIS'20)
3. Азимов Р. Ш. Context-Free Path Querying by Kronecker Product большая версия / Орачев Е., Эпельбаум И., Азимов Р., Григорьев С. // In Proceedings of the (SIGMOD'21)
4. Азимов Р. Ш. Ненаписанная работа матричный алгоритм по всем путям
5. Азимов Р. Ш. Path Querying with Conjunctive Grammars by Matrix Multiplication / Азимов Р., Григорьев С. // Programming and Computer Software. – 2019. – Т. 45. – №. 7. – С. 357-364.
6. Азимов Р. Ш. Context-Free Path Querying by Matrix Multiplication / Азимов Р., Григорьев С. // In Proceedings of the 1st Joint International Workshop on Graph Data Management Experiences & Systems (GRADES) and Network Data Analytics (NDA) (GRADES-NDA'18)
7. Азимов Р. Ш. Синтаксический анализ графов с использованием конъюнктивных грамматик / Азимов Р., Григорьев С. // Труды ИСП РАН, 2018, том 1 вып. 2, с. 3-4.
8. Азимов Р. Ш. Синтаксический анализ графов и задача генерации строк с ограничениями / Азимов Р., Григорьев С. // Сборник трудов конференции PLC 2017, с. 24-27.