

???

Екатерина Шеметова

16 июня 2020 г.

1 Основные данные проекта

1.1 Название проекта

Варианты:

- 1) Улучшенные оценки сложности проблемы пустоты пересечения автоматов
- 2) Алгебраические, автомато-теоретические свойства формальных языков и оценки сложности алгоритмов для поиска путей с ограничениями в терминах формальных языков.

1.2 Основной код (по классификатору РФФИ)

07-365 Специализированные методы и алгоритмы обработки и анализа больших данных

1.3 Ключевые слова (указываются отдельные слова и словосочетания, наиболее полно отражающие содержание проекта: не более 15, строчными буквами, через запятые)

поиск путей в графах, пересечение автоматов, теория формальных языков, контекстно-свободные языки, теория сложности, теория графов

1.4 Аннотация проекта (кратко, в том числе – актуальность, уровень значимости и научная новизна исследования; ожидаемые результаты и их значимость; аннотация будет опубликована на сайте РФФИ, если проект получит поддержку)

Задача поиска путей в графе является фундаментальной алгоритмической задачей. Часто оказывается необходимым находить не просто пути, связывающие вершины, а пути с некоторыми ограничениями. Одним из популярных видов ограничений для графов с помеченными

рёбрами являются ограничения в терминах формальных языков. Например, ограничения в терминах регулярных языков используются для поиска паттернов в графах и запросах к графовым базам данных, а ограничения в терминах контекстно-свободных языков активно применяются в задачах статического анализа программного кода (проблема CFL-reachability). Также существуют более широкие классы языков, такие как языки, задаваемые конъюнктивными и булевыми грамматиками, позволяющие формулировать более сложные ограничения. Задача поиска путей с ограничениями в терминах формальных языков может быть также сформулирована как задача проверки пустоты конечного автомата и автомата, допускающего язык, задающий ограничения на пути.

Известно, что для некоторых фиксированных подклассов языков и графов задача поиска путей с ограничениями в терминах формальных языков может быть решена быстрее, чем в общем случае для произвольного графа и класса языка. Например, задача поиска путей с контекстно-свободными ограничениями P -полна для произвольного графа, но в случае линейных языков, являющихся подклассом контекстно-свободных, данная задача лежит в классе NC при фиксированном языке. При этом являются практически не изученными свойства графов или языков, благодаря которым задача может быть решена эффективнее, чем в общем случае.

Проект посвящен исследованию алгебраических и автомато-теоретических характеристик входных данных задачи поиска путей с ограничениями в терминах формальных языков: языка, задающего ограничения и помеченного графа (конечного автомата). Планируется изучить и обобщить свойства подклассов языков (автоматов) и графов, влияющие на эффективность алгоритмов поиска путей, что поможет найти большой класс языков и графов с асимптотически лучшим, по сравнению с общим случаем, временем работы алгоритмов поиска путей. Исследование направлено как на получение более эффективных алгоритмов для частных случаев, полезных на практике, так и на улучшение оценок для существующих алгоритмов.

1.5 Название проекта (на английском языке)

?

1.6 Ключевые слова (на английском языке)(приводится не более 15 слов)

path querying, formal language theory, context-free language reachability, complexity theory, graph theory

1.7 Аннотация проекта на английском языке (кратко, в том числе - актуальность, уровень фундаментальности и научная новизна; ожидаемые результаты и их значимость)

Graph reachability problem is a fundamental algorithmic problem. One of the popular variant of this problem is path quering with constraints in terms of formal languages. For example, the regular

path queries are used in finding patterns in a large graph and graph database query evaluation. The context-free path queries are applicable in static code analysis (the CFL-reachability problem). Also there are conjunctive and boolean grammars, which describe more complex constraints. Also the formal language reachability problem can be considered as the intersection emptiness problem for a finite automaton and an automata accepting a given formal language.

It is known that the problem of path quering for some fixed subclasses of formal languages and a graphs can be solved faster than in the general case of arbitrary graph and the whole language class. For example, context-free reachability problem is known to be P-complete, but in the case of linear languages, which is subclass of context-free languages, this problem lies in NC. Additionally, the properties of the input language or the graph, which imply effectiveness of reachability algorithm are not fully discovered.

The goal of the project is to investigate an algebraic and automata-theoretic properties of the input graph (finite automaton) and the input formal language. Discovering and generalizing of the properties implying the effectiveness of reachability algorithm are planned. This is helpful for finding the big class of languages and graphs, for which the formal language reachability problem has more effective solution. The project focuses on obtaining effectictive algorithms for a specific cases and on estimating better bounds for a known algorithms.

2 Содержание проекта

2.1 Цель и задачи проекта

Целью проекта является получение улучшенных оценок сложности алгоритмов поиска путей с ограничениями в терминах формальных языков. Достижение поставленной цели обеспечивается решением следующих задач.

1. Исследовать влияние на сложность задачи поиска путей алгебраических и автоматно-теоретических характеристик языка, задающего ограничения на пути.
2. Исследовать влияние на сложность задачи поиска путей структурных и сложностных характеристик входного графа.
3. Получить оценки сложности алгоритмов в зависимости от характеристик графа и/или формального языка.
4. Получить классы языков/графов, для которых задача поиска путей может быть решена более эффективно, и предложить алгоритмы для них.

2.2 Направление из Стратегии научно-технологического развития Российской Федерации (при наличии) (выбор из справочника)

1) Переход к передовым цифровым, интеллектуальным производственным технологиям, роботизированным системам, к новым материалам и способам конструирования, создание систем обработки больших объемов данных, машинного обучения и искусственного интеллекта;

2.3 Анализ современного состояния исследований в данной области (приводится обзор исследований в данной области со ссылками на публикации в научной литературе)

Задача поиска путей в графе является фундаментальной алгоритмической задачей. Часто оказывается необходимым находить не просто пути, связывающие вершины, а пути с некоторыми ограничениями. Одним из популярных видов ограничений для графов с помеченными рёбрами являются ограничения в терминах формальных языков. Например, ограничения в терминах регулярных языков используются для поиска паттернов в графах и запросах к графовым базам данных, а ограничения в терминах контекстно-свободных языков активно применяются в задачах статического анализа программного кода (проблема CFL-reachability). Также существуют более широкие классы языков, такие как языки, задаваемые конъюнктивными и булевыми грамматиками, позволяющие формулировать более сложные ограничения.

Задача поиска путей с ограничениями в терминах формальных языков может быть сформулирована следующим образом. Зафиксируем язык L . Для данного помеченного ориентированного графа, метки которого являются символами из алфавита языка L , и двух выделенных вершин s и t нужно ответить на вопрос, существует ли путь от s до t , метки рёбер которого составляют слово из языка L .

Задача поиска путей с ограничениями в терминах формальных языков может быть также сформулирована как задача проверки пустоты конечного автомата и автомата, допускающего язык, задающий ограничения на пути.

Основные результаты в данной области включают в себя оценки сложности по памяти для задач поиска путей с ограничениями в терминах регулярных, контекстно-свободных, контекстно-зависимых языков в произвольных графах (Komarath B., Sarma J., Sunil K S. On the Complexity of L-reachability, 2014; Holzer M, Kutrib M, Leiter U. Nodes connected by path languages, 2011). Также в данных работах рассмотрены задачи поиска путей с ограничениями в терминах упомянутых выше языков и фиксированных подклассов графов — ациклических графов и деревьев.

Некоторые характеристики графов и языков, влияющие на сложность алгоритма изучены в нескольких работах. Показано, что одной из характеристик языка, влияющей на эффективность параллельного алгоритма поиска путей, является так называемый рациональный индекс языка (Ullman JD, Van Gelder A. Parallel complexity of logical query programs, 1986; Rubtsov A, Vyalyi M, Regular realizability problems and context-free languages, 2015; Ramaswamy, Vidhya et al. Space complexity of reachability testing in labelled graphs, 2019). К сожалению, данный показатель носит описательный характер, отсутствуют доказанные алгебраические или автомато-теоретические свойства языка, приводящие к той или иной величине рационального индекса. Иными словами, не изучено, почему одни языки имеют "хороший" рациональный индекс, и, следовательно, эффективный параллельный алгоритм, а другие нет.

Давно известно, что задача поиска путей с ограничениями в терминах контекстно-свободных и регулярных языков может быть решена эффективнее, если входной граф — ациклический (Yannakakis, M. Graph-theoretic methods in database theory, 1990). Также показано, что задача поиска путей с ограничениями в терминах булевых грамматик разрешима на ациклических графах (Shemetova E.N., Grigorev S.V. Path querying on acyclic graphs using Boolean grammars,

2019). Также существуют алгебраические характеристики помеченного графа, влияющие на эффективность параллельного алгоритма поиска путей с контекстно-свободными ограничениями (Ganardi M, Hucke D, König D, Lohrey M. Circuit evaluation for finite semirings, 2016). Авторы данной работы оставляют открытым вопрос о свойствах помеченного графа (конечного автомата), соответствующих найденным характеристикам.

2.4 Предлагаемые методы и подходы к решению поставленных задач (включая детальный план проводимых исследований)

В рамках данного исследования планируется изучить влияние таких автомато-теоретических свойств, как динамика изменения высоты стека автомата с магазинной памятью, характеристик языка стека (storage languages) на величину рационального индекса и сложность по памяти задачи поиска путей с контекстно-свободными ограничениями. Также будут рассмотрены подклассы языков, обладающие особыми алгебраическими свойствами, например, языки с полиномиальной плотностью.

Также планируется исследовать влияние на эффективность поиска путей структурных и сложностных характеристик входного графа, таких как ориентированная древесная и путевая ширина, *DAG-width* и другие.

Далее будут выявлены подклассы языков и графов, обладающие свойствами, позволяющими построить эффективный алгоритм поиска путей для этих подклассов, разработаны соответствующие алгоритмы и доказана их теоретическая сложность.

Затем полученные характеристики будут использованы для улучшения оценок сложности текущих алгоритмов задачи поиска путей с ограничениями в терминах языка, для которого найдены эффективные подклассы, или построения эффективных приближенных алгоритмов для данных задач.

2.5 Новизна исследования, заявленного в проекте (формулируется новая научная идея, обосновывается новизна предлагаемой постановки и решения заявленной проблемы)

В данном исследовании будут получены и обобщены свойства формальных языков и графов, влияющие на сложность алгоритма поиска путей с ограничениями в терминах формальных языков. Также будет получен большой класс языков и графов с асимптотически лучшим, по сравнению с общим случаем, временем работы алгоритмов поиска путей.

Исследование направлено как на получение более эффективных алгоритмов для частных случаев, полезных на практике, так и на улучшение оценок для существующих алгоритмов.

2.6 Ожидаемые по окончании проекта научные результаты

Доказаны оценки сложности алгоритма поиска путей с контекстно-свободными ограничениями в зависимости от динамики изменения высоты стека автомата с магазинной памятью,

а также характеристик языка стека. Доказаны оценки сложности алгоритма поиска путей с ограничениями в терминах формальных языков для входных графов разной сложности. Выявлены классы языков и графов, для которых существуют более эффективные алгоритмы, соответствующие алгоритмы предложены. Теоретические результаты и оценки сложности опубликованы в журнале, индексируемом в Scopus или ВАК. Предложенные алгоритмы представлены на конференции и опубликованы в сборнике материалов конференции, индексируемом в Scopus.

2.7 Научный задел Научного руководителя по тематике проекта

Ф. А. заполняет.

2.8 Педагогический задел Научного руководителя (обязательно указать, количество аспирантов, из них – количество защитивших диссертацию; количество ученых, защитивших диссертации на соискание ученой степени доктора наук)

Ф. А. заполняет.

2.9 Список основных публикаций Научного руководителя в рецензируемых журналах (не менее 5)

Ф. А. заполняет.

2.10 Название диссертационной работы Аспиранта

Субкубическая временная сложность задачи поиска путей с контекстно-свободными ограничениями.

2.11 Основные цели и задачи диссертационного исследования

Целью диссертационного исследования является разработка субкубического алгоритма для решения задачи поиска путей с контекстно-свободными ограничениями.

Достижение поставленной цели обеспечивается решением следующих задач.

1. Выявить подклассы контекстно-свободных языков и графов, для которых временная оценка сложности алгоритма поиска путей может быть улучшена до субкубической.
2. Разработать алгоритмы поиска путей с ограничениями в терминах контекстно-свободных языков, использующие алгебраические свойства графов и контекстно-свободных языков.

3. Изучить возможности сведения задачи поиска путей с контекстно-свободными ограничениями к другим алгоритмическим задачам, например, к задаче динамического транзитивного замыкания, и получить соответствующие верхние и нижние оценки сложности.

2.12 Список основных (не более 5) публикаций Аспиранта в рецензируемых журналах

Аспирант имеет следующую публикацию в рецензируемом журнале.

1. Шеметова Е.Н., Григорьев С.В. Задача поиска путей в ациклических графах с ограничениями в терминах булевых грамматик. Труды Института системного программирования РАН. 2019; 31(4):211-226.

2.13 Научный задел Аспиранта по тематике проекта (необходимо указать сколько выступлений на конференциях; список всех публикаций; прочие достижения (премии, награды, гранты))

Аспирант обладает опытом в применении теории формальных языков к различным задачам, что подтверждается публикацией и участием в качестве исполнителя в гранте РФФИ «Логические и алгебраические методы в теории формальных языков» (18-11-00100).

Аспирант имеет следующие публикации.

1. ВАК, Шеметова Е.Н., Григорьев С.В. Задача поиска путей в ациклических графах с ограничениями в терминах булевых грамматик. Труды Института системного программирования РАН. 2019; 31(4):211-226.

2.14 Дата приказа о переводе на второй курс аспирантуры

01.07.2020