УДК 004.825

Теоретико-множественные операции над грамматиками и их трактовка в различных математических дисциплинах

А. Г. Гурин*[1], А. В. Григорьев*[2]
*[1] аспирант, Донецкий национальный технический университет,
gurin.sasha.1996@gmail.com
*[2] к.т.н., доцент кафедры ПИ, Донецкий национальный технический университет,
grigorievalvl@mail.ru, SPIN-код: 8830-2813

Аннотация
В статье исследуется взаимосвязь теоретико-множественных операций над грамматиками с математическими дисциплинами. Актуальность работы обусловлена необходимостью эффективной обработки сложных данных в современных информационных системах. Методология включает анализ графовых структур, тензорных представлений и полугрупп. Особое внимание уделяется применению теоретико-множественных операций к различным математическим объектам. Результаты демонстрируют перспективность предложенного подхода для решения задач синтаксического анализа, машинного перевода и оптимизации программного кода. Практическая значимость работы заключается в создании новых инструментов для работы с базами знаний.

Введение

Актуальность представленного исследования обусловлена несколькими взаимосвязанными факторами, стимулирующими развитие компьютерных наук, лингвистики и математики.

Современные приложения генерируют огромные объемы сложных данных, часто обладающих замысловатыми реляционными структурами. Традиционные теоретико-множественные операции и графовые методы анализа недостаточны для эффективной обработки и извлечения содержательной информации из таких данных. Современный этап работы с инструментами и базой знаний можно охарактеризовать использованием графов и тензоров, который как инструменты имеют ряд ограничений.

Значительный прогресс в тензорной алгебре и машинном обучении предоставляет мощные инструменты для работы с многомерными данными. Возможность представления сложных графовых структур в виде тензоров открывает перспективы применения этих передовых инструментов к задачам, ранее неразрешимым традиционными методами. Полугруппы же несут с собой ряд дополнительных операций и возможность создавать более сложные базы знаний. Так же это позволяет создавать более эффективные алгоритмы и выявлять скрытые закономерности в данных.

Графы зависимостей и деревья разбора являются фундаментальными структурами в анализе программ. Эффективные алгоритмы для работы с этими структурами критически важны для повышения производительности в таких задачах, как синтаксический анализ, машинный перевод, оптимизация кода и автоматическая верификация программ.

Применение модифицированных теоретико-множественных операций к тензорным представлениям графов и математическим полугруппам расширяет границы традиционных алгебраических методов. А возможность взаимоиспользовать инструменты этих математических сущностей дает больше гибкости в использовании на сложных баз знаний. Это исследование может привести к открытию новых математических соотношений и алгоритмов с более широкими применениями за пределами непосредственных предметных областей [1].

Целью данной работы является показ применения теоретико-множественных операций над формальными грамматиками с использованием разных подходов.

В этой статье рассмотрим: что такое теоретико-множественные операции; как представить формальную грамматику в виде графа; взаимосвязь между графом и тензором; взаимосвязь между тензором и полугруппой; набор инструментов, которые могут расширить работу с базами знаний.

Что такое теоретико-множественные операции

Теоретико-множественные операции (ТМО) — это операции, которые определены над множествами. Они позволяют создавать новые множества из существующих, основываясь на отношениях между элементами этих множеств.

ТМО над графами — это операции, которые определены на множествах вершин и ребер графов, используя базовые теоретико-множественные операции (объединение, пересечение, разность, дополнение). Однако, поскольку граф — это не просто множество, а структура с определенными отношениями (связности) между элементами, применение этих операций требует определенной осторожности и, зачастую, приводит к результатам, не всегда интуитивно понятным.

Рассмотрим основные теоретико-множественные операции и их применение к графам:

- Объединение графов ($G_1 \cup G_2$) — новый граф, содержащий все вершины и ребра из обоих исходных графов G_1 и G_2. Если вершины с одинаковым именем присутствуют в обоих графах, они объединяются в одну вершину в результирующем графе. Ребра также объединяются, но важно учитывать, что могут возникнуть кратные ребра между одними и теми же вершинами.

- Пересечение графов ($G_1 \cap G_2$) — новый граф, содержащий только те вершины и ребра, которые присутствуют одновременно в обоих исходных графах G_1 и G_2. Если вершина присутствует в обоих графах, но ребра, связанные с этой вершиной, различны, то в результирующем графе будет только эта вершина, без ребер.

- Разность графов ($G_1 \setminus G_2$) — новый граф, содержащий все вершины и ребра из графа G_1, которые *не* присутствуют в графе G_2. Вершины из G_2, отсутствующие в G_1, игнорируются. Ребра, присутствующие в G_2, удаляются из G_1.

- Дополнение графа (G^C) — это граф на том же множестве вершин, но содержащий *только те ребра*, которые *отсутствуют* в исходном графе G.

Рассмотрим на примере. Есть два графа:

G_1: Вершины {A, B, C}, Ребра {(A, B), (B, C)}

G_2: Вершины {B, C, D}, Ребра {(B, C), (C, D)}

Тогда:

- $G_1 \cup G_2$ (Объединение): Вершины {A, B, C, D}, Ребра {(A, B), (B, C), (B, C), (C, D)}

- $G_1 \cap G_2$ (Пересечение): Вершины {B, C}, Ребра {(B, C)}

- $G_1 \setminus G_2$ (Разность): Вершины {A, B}, Ребра {(A, B)}

- $G_2 \setminus G_1$ (Разность): Вершины {C, D}, Ребра {(C, D)}

В целом, теоретико-множественные операции над графами представляют собой полезный инструмент для анализа и манипулирования графовыми структурами, но требуют внимательного рассмотрения специфики графовых представлений и свойств. Они часто используются в качестве базовых операций в более сложных алгоритмах обработки графов [2].

Граф, как часть формальной грамматики

Граф может выступать для грамматики в нескольких важных ролях, в зависимости от контекста и целей анализа:

– представление структуры предложений (для естественного языка): граф может отображать синтаксические зависимости между словами в предложении: в котором вершины графа соответствуют словам, а ребра — зависимостям между ними (например, субъект-предикат, модификатор-существительное);

– представление структуры программного кода: граф может представлять зависимости между элементами программного кода (функции, переменные, операторы), который используется в статическом анализе кода, оптимизации компиляторов и автоматическом тестировании;

– представление зависимостей в правилах грамматики: а более абстрактном смысле, граф может показать зависимости между правилами грамматики, например, вершины могут быть правилами, а ребра — указанием на то, что одно правило использует нетерминал, определенный другим правилом.

– дерево разбора — это частный случай дерева (ориентированного ациклического графа), оно показывает, как строка выводится по правилам грамматики, в котором корень дерева — стартовый символ грамматики, листья — терминалы, а внутренние узлы — нетерминалы, а анализ дерева разбора позволяет понять структуру выведенной строки и проводить различные виды анализа.

Граф может отображать все возможные выводы грамматики для данной строки. Вершины могут соответствовать промежуточным выводам, а ребра — применению правил грамматики. В случае с графом для работы с грамматикой мы будем использовать И-ИЛИ-дерево. Базовые графы создают рекурсию и цикличность, что усложняет понимание и работу, но использование ТМО с ними сильно затруднено. И-ИЛИ-дерево достаточно структурировано и масштабируемый инструмент. Так как И-ИЛИ-дерево остается все равно графом, это дает нам возможность использовать все операции с графами. Все вышеперечисленные ТМО могут использоваться с И-ИЛИ-деревом. Важно учесть, что через логические операции и будем исследовать и осуществлять переход на другие математические сущности: тензора и полугруппы.

И-ИЛИ-дерево - это дерево, узлы которого могут быть двух типов: И-узел или ИЛИ-узел (см. рис. 1). Вариантом дерева И/ИЛИ называется дерево, полученное из данного путём отсечения всех дуг, кроме одной, у ИЛИ-узлов. Корнем варианта будет корень исходного дерева. Мощностью дерева И/ИЛИ называется количество вариантов, которое оно содержит, представленного на рис. 2 [3,4].

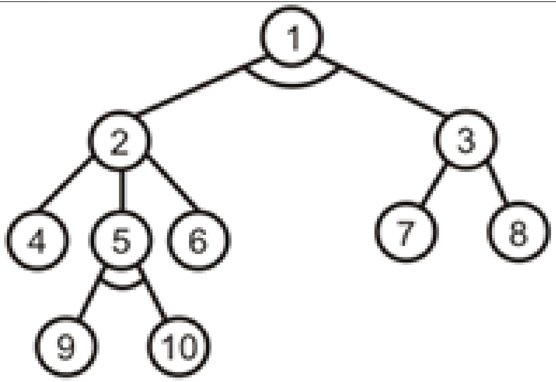

Рисунок 1 - И-ИЛИ-дерево [4]

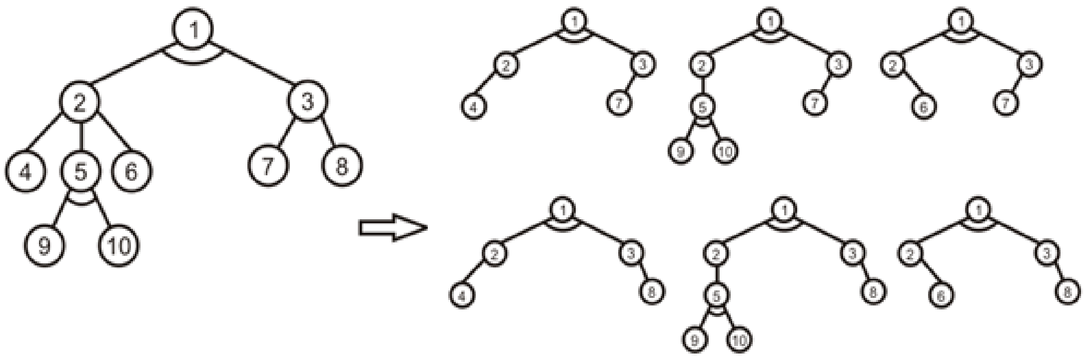

Рисунок 2 - Мощность И-ИЛИ-дерева [4]

И-ИЛИ-дерево — это структура данных для представления пространства состояний в задачах поиска решений [5,6], сложных логических условий в продукционных системах [7], а также они дают преимущества при работе с формальной грамматикой. А формальная грамматика как раз и позволить создать более универсальный инструмент, которые позволит создавать базы знаний и в области электроники, в области программировании и на уровне естественных языков.

Формальная грамматика – это математическая система правил, определяющая структуру и синтаксис формального языка. Она служит эффективным инструментом формализации, который позволяет четко определить правила построения языковых конструкций. Данная система выступает в качестве регулятора, определяющего допустимые комбинации символов.

В рамках теории формальных языков грамматика выступает как математический инструмент, устанавливающий правила построения корректных языковых выражений. Она определяет, какие последовательности символов являются допустимыми в данном языке, а какие следует считать некорректными согласно установленным правилам.

Следует подчеркнуть, что формальная грамматика не только устанавливает правила построения выражений, но и обеспечивает основу для анализа и создания языковых конструкций, что делает ее ключевым элементом в компьютерных науках, лингвистике и теории автоматов.

ТМО над тензорами

Так как напрямую граф никак не связан с полугруппами, мы должны сначала придать им более подходящую форму для взаимодействия с полугруппами. Этой формой сущности являются тензоры.

Представить И-ИЛИ-дерево напрямую в виде обычного тензора довольно сложно, потому что структура И-ИЛИ-дерева не является естественным образом многомерным массивом. Тензоры обычно представляют собой многомерные массивы с упорядоченными индексами, а структура И-ИЛИ-дерева иерархична и нерегулярна. Однако, можно использовать тензорные представления, которые кодируют информацию об И-ИЛИ-дереве, но не напрямую отображают его структуру. Рассмотрим несколько вариантов.

Важно отметить, что работы по введению тензоров в образовательную систему уже проводились [8]. Есть работы, где исследует использование нейросетей в образовательной системе [9].

Можно использовать тензорное представление с использованием кодирования узлов. Для этого нужно каждому узлу И-ИЛИ-дерева присваивается уникальный числовой идентификатор. Тип узла (И или ИЛИ) можно кодировать, например, с помощью битового флага в идентификаторе или в отдельном тензоре. Затем создается тензор второго ранга (матрица), представляющий смежность между узлами. Элемент A_{ij} равен 1, если существует ребро от узла _i_ к узлу _j_, и 0 в противном случае. Это представление хранит информацию о структуре дерева, но не явно отображает тип узлов (И или ИЛИ). Для кодирования типа узлов потребуется дополнительный тензор. Так же можно использовать дополнительные тензоры для хранения информации о типах узлов, значениях листьев (если есть) и других атрибутах узлов. Дополнительные тензоры могут расширить возможности применения алгоритмов в различных областях.

Еще одним вариантов является тензорное представление, основанное на пути к листьям. Для каждого листа И-ИЛИ-дерева можно определить путь от корня до этого листа. Этот путь можно представить как последовательность идентификаторов узлов. Каждый путь можно представить как вектор, где каждый элемент вектора — это идентификатор узла на пути. Все пути можно объединить в тензор, где каждое измерение соответствует узлу пути. Пустые элементы вектора можно заполнить специальным значением (например, -1). Это представление не отображает непосредственно дерево, но кодирует информацию о нём через множество путей.

Стоит отметить еще тензорное представление на основе свертки. Этот подход более сложный и требует применения методов, подобных свёрточным нейронным сетям. И-ИЛИ-дерево может быть представлено как графическая структура, а затем применяется свёртка, которая генерирует тензорное представление. Такой подход эффективен для задач машинного обучения, где И-ИЛИ-дерево используется как входные данные.

После представления графом в виде тензора, мы можем начать взаимодействовать на уровне ТМО. Теоретико-множественные операции над тензорами не являются стандартными операциями в линейной алгебре или тензорном анализе, как, например, тензорное произведение или свертка. Тензоры обычно рассматриваются как элементы векторных пространств, над которыми определены линейные операции (сложение, умножение на скаляр и т.д.). Однако, можно определить аналогии теоретико-множественных операций, если рассматривать тензоры как многомерные массивы или сопоставлять им множества индексов.

Рассмотрим возможные подходы к определению теоретико-множественных операций для тензоров:

– операции над элементами тензора - если рассматривать тензор как многомерный массив чисел, можно определять теоретико-множественные операции над элементами этого массива, однако это будет операция над множеством чисел, а не над тензорами как математическими объектами;

– операции над множествами индексов - можно сопоставить тензору множество индексов, которые определяют его элементы, тогда теоретико-множественные операции можно применить к этим множествам индексов;

– операции над подпространствами, порожденными тензорами - если тензоры рассматривать как элементы векторного пространства, им можно сопоставить подпространства, порожденные этими тензорами, тогда теоретико-множественные операции можно трактовать как операции над этими подпространствами, например, пересечение двух подпространств порождает новое подпространство.

Не существует единого общепринятого определения теоретико-множественных операций над тензорами. Подходы, описанные выше, представляют собой возможные способы интерпретации таких операций, но их использование требует четкого определения и зависит от контекста задачи [10].

Операции над полугруппами

Полугруппа — это алгебраическая структура, состоящая из множества S и ассоциативной бинарной операции *, которая отображает две любые пары элементов из множества S в третий элемент, также принадлежащий S. Ассоциативность означает, что для любых трёх элементов a, b, и c из S выполняется следующее равенство по формуле 1:

$$(a\ \ b)\ \ c = a\ \ (b\ \ c). \qquad (1)$$

Это означает, что порядок вычисления операции не влияет на результат.

Примеры полугрупп:

– множество натуральных чисел с операцией сложения: $(\mathbb{N}, +)$ — это полугруппа, так как сложение ассоциативно.

– множество всех строк над некоторым алфавитом с операцией конкатенации (сцепления): если у нас есть алфавит Σ, то множество Σ* всех возможных строк над этим алфавитом образует полугруппу с операцией конкатенации.

Примеры ассоциативных бинарных операций:

– сложение чисел для любых действительных, комплексных, целых и т.д. чисел:

$$(a + b) + c = a + (b + c); \qquad (2)$$

– умножение чисел для любых действительных, комплексных, целых и т.д. чисел:

$$(a * b) * c = a * (b * c); \qquad (3)$$

– конкатенация строк: объединение строк. Если a = "Hello", b = " ", c = "World!", то (a ++ b) ++ c = a ++ (b ++ c) = "Hello World!";

– объединение множеств:

$$(A \cup B) \cup C = A \cup (B \cup C); \tag{4}$$

– пересечение множеств:

$$(A \cap B) \cap C = A \cap (B \cap C); \tag{5}$$

– композиция функций: если f, g, и h — функции, то $(f \circ g) \circ h = f \circ (g \circ h)$, где $\circ$ обозначает композицию функций.

Вы видим, что у полугрупп нет так называемых "теоретико-множественных операций", однако существует набор логических операций, которые не только в точности повторяют их функциональность, но и предоставляют значительно более широкие возможности для математического анализа. В частности, помимо стандартных теоретико-множественных операций, полугруппы обладают расширенным набором специфических операций, которые значительно обогащают их математический аппарат и открывают новые перспективы для исследования.

Это фундаментальное различие между теоретико-множественным и полугрупповым подходами открывает перед нами широкие возможности для глубокого и всестороннего анализа, а также установления взаимозаменяемости между различными математическими сущностями, которые на первый взгляд не связанны между собой. При более детальном рассмотрении становится очевидным, что именно эти дополнительные операции создают богатый инструментарий для выявления скрытых связей и закономерностей.

Данный методологический подход позволяет нам не только детально исследовать, но и математически строго доказывать логические взаимосвязи между различными сущностями, что в свою очередь существенно расширяет и обогащает инструментарий для работы с базой знаний.

Вывод

Важно отметить, что предложенный подход не лишен ограничений. Необходимо дальнейшее исследование для оптимизации вычислительной сложности алгоритмов, особенно при работе с большими графами знаний и сложными тензорными представлениями. Кроме того, нужно разработать методики эффективного сопоставления различных типов графовых структур с тензорными представлениями, учитывая специфику структуры данных и целевые задачи. Более глубокое изучение взаимосвязи между свойствами полугрупп и модифицированными теоретико-множественными операциями позволит разработать более эффективные и гибкие алгоритмы.

В заключение, проведенное исследование демонстрирует перспективность использования тензорных представлений для анализа графовых структур, в частности, деревьев И-ИЛИ, применяемых в задачах обработки естественного языка и анализа программного кода. Предложенный подход, основанный на модификации традиционных теоретико-множественных операций и применении их к тензорным представлениям графов и полугруппам, позволяет эффективно обрабатывать сложные реляционные данные. Полученные результаты указывают на возможность создания более производительных алгоритмов для решения задач синтаксического анализа, машинного перевода и оптимизации программного кода.

Дальнейшие исследования могут быть направлены на разработку конкретных алгоритмов на основе предложенного подхода, а также на изучение применимости данной методологии к другим типам графовых структур и областям применения. Перспективным направлением является исследование связи между модифицированными теоретико-множественными операциями над тензорами и свойствами полугрупп, что может привести к открытию новых математических закономерностей и алгоритмов. В целом, предложенный подход открывает новые возможности для эффективной обработки и анализа больших объемов сложных данных с реляционными связями.

В заключение, представленное исследование закладывает фундаментальную основу для разработки и внедрения инновационных инструментов, предназначенных для комплексной работы с базами знаний. Эти инструменты позволят не только эффективно представлять и модернизировать существующие базы знаний, но и синтезировать новые на основе прошлых знаний. Особенно важно отметить, что данное исследование открывает широкие перспективы в области логики входа-выходов, базирующейся на уникальном сочетании трех ключевых математических концепций: графовых структур, обеспечивающих топологическое представление данных, тензорной алгебры, позволяющей работать с многомерными массивами информации, и теории полугрупп, предоставляющей мощный аппарат для анализа алгебраических структур.

Следующим этапом исследования станет формализация логических связей, фундаментальных требований и системных ограничений. Такая подход позволит на строгом математическом уровне провести анализ предложенного метода, что в конечном итоге приведет либо к его научному обоснованию и подтверждению, либо к выявлению существенных ограничений и опровержений.

Литература

1. Гурин, А. Г. Обзор методов представления онтологий с физической семантикой / А. Г. Гурин, А. В. Григорьев // Сборник материалов VIII Всероссийской научно-технической конференции «Современные информационные технологии в образовании и научных исследованиях» (СИТОНИ-2023). - Донецк, 2023. - С. 87-92.

2. Григорьев, А.В. Алгоритм выполнения теоретико-множественных операций над грамматиками в среде специализированной оболочки для создания интеллектуальных САПР, Научные труды Донецкого государственного технического университета Серия: Проблемы моделирования и автоматизации проектирования динамических систем (МАП-2002). — № 52. — Донецьк: ДонНТУ, 2002. — С. 83–93.

3. Григорьев, А.В. Специфика выполнения теоретико-множественных операций над контекстно-свободными грамматиками в условиях различных форм дополнительных семантических правил в семиотической модели интеллектуальных САПР / А.В. Григорьев // Научные труды Донецкого национального технического университета. Серия «Проблемы моделирования и автоматизации проектирования динамических систем» (МАП – 2006). – Выпуск 5 (116). – Донецк: ДонНТУ, 2006. – С. 91-104.

4. Метод генерации тестовых заданий на основе деревьев И/ИЛИ и его программная реализация - Хабр [Электронный ресурс] / Интернет-ресурс. Режим доступа: https://habr.com/ru/articles/198068/. – Загл. с экрана.

5. Russell, S. Artificial Intelligence: A Modern Approach / S. Russell, P. Norvig. – 4th ed. – Pearson, 2020. – 1136 p. – ISBN 978-0134610993.

6. Nilsson, N. J. Principles of Artificial Intelligence / N. J. Nilsson. – Springer, 1982. – 476 p. – ISBN 978-3-540-11340-1.

7. Люгер, Дж. Ф. Искусственный интеллект: стратегии и методы решения сложных проблем / Дж. Ф. Люгер; пер. с англ. – М.: Вильямс, 2003. – 864 с.

8. Андриевская, Н. К. Обобщенная модифицированная модель представления текстовых информационных ресурсов / Н. К. Андриевская // Информатика и кибернетика. – 2020. – № 4 (22). – С. 21–30.

9. Николаев, А. А. Международный опыт и перспективы использования искусственного интеллекта в образовании / А. А. Николаев, М. Ю. Кузнецов, В. А. Николаев // Управление образованием: теория и практика. – 2024. – Т. 14, № 5-1. – DOI 10.25726/e8567-8724-8003-k. – EDN ZNISEA.

10. Полугруппа - Википедия [Электронный ресурс] / Интернет-ресурс. Режим доступа: https://ru.wikipedia.org/wiki/Полугруппа. – Загл. с экрана.

Гурин А.Г., Григорьев А.В. Теоретико-множественные операции над грамматиками и их трактовка в различных математических дисциплинах. *В статье исследуется взаимосвязь теоретико-множественных операций над грамматиками с математическими дисциплинами. Актуальность работы обусловлена необходимостью эффективной обработки сложных данных в современных информационных системах. Методология включает анализ графовых структур, тензорных представлений и полугрупп. Особое внимание уделяется применению теоретико-множественных операций к различным математическим объектам. Результаты демонстрируют перспективность предложенного подхода для решения задач синтаксического анализа, машинного перевода и оптимизации программного кода. Практическая значимость работы заключается в создании новых инструментов для работы с базами знаний.*

Ключевые слова: *формальная грамматика, онтология, тензор, полугруппа, граф, и-или-дерево.*

Gurin A.G., Grigoriev A.V. Set-theoretic operations on grammars and their interpretation in various mathematical disciplines. *The article examines the relationship between set-theoretic operations on grammars and mathematical disciplines. The relevance of the work is due to the need for efficient processing of complex data in modern information systems. The methodology includes the analysis of graph structures, tensor representations, and semigroups. Special attention is paid to the application of set-theoretic operations to various mathematical objects. The results demonstrate the promise of the proposed approach for solving problems of syntactic analysis, machine translation, and software code optimization. The practical significance of the work lies in the creation of new tools for working with knowledge bases.*

Keywords: *formal grammars, ontology, tensor, semigroup, graf, and-or-tree.*

Статья поступила в редакцию 21.09.2025
Рекомендована к публикации профессором Зори С. А.