ОТЗЫВ

члена диссертационного совета Кузнецова Николая Владимировича на диссертацию Юрия Олеговича Костюкова «Автоматический вывод индуктивных инвариантов программ с алгебраическими типами данных», представленную на соискание учёной степени кандидата физико-математических наук по специальности 2.3.5. Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей

В современном мире программное обеспечение играет все большую роль, и, соответственно, его надежность становится критически важной. В связи с этим формальные методы (formal methods) призваны сыграть важную роль, создавая применимые на практике методы верификации сложных программных систем. В последнее время появились и оказали решающее влияние на методы верификации такие подходы как бинарные диаграммы решений и символьная проверка моделей. В совокупности с методами статического анализа эти подходы привели к появлению SAT- и SMT-решателей, а также разнообразных целевых инструментов, основывающихся на этих решателях.

На заре использования формальных методов основное внимание уделялось поддержке "классических" типов данных, таких как целые числа и массивы. С ростом популярности алгебраических типов данных (АТД), которые обладают рекурсивной структурой и широко используются в функциональных языках программирования для представления сложных структур данных, возникла необходимость в разработке соответствующих методов верификации. Однако до настоящего времени АТД оставались мало исследованными, что создавало препятствия для автоматизированного доказательства корректности программ, использующих эти типы данных. Сложность верификации программ с алгебраическими типами данных заключается в необходимости вывода индуктивных инвариантов, необходимых для доказательства корректности программ: существующие подходы часто требуют от пользователя ручного вывода таких инвариантов, что является очень трудоемким делом. Преодоление этого барьера и разработка методов автоматического вывода индуктивных инвариантов, способных адекватно работать с АТД, стали ключевыми задачами данной работы.

В ходе работы были получены следующие результаты.

1. Метод автоматического вывода индуктивных инвариантов на основе автоматов над деревьями и поиска конечных моделей, включая доказательство корректности метода.
2. Метод автоматического вывода индуктивных инвариантов на основе трансформации программы и поиске конечных моделей в классе синхронных автоматов над деревьями, включая доказательство корректности метода.
3. Новый класс индуктивных инвариантов, основанный на булевой комбинации классических инвариантов и автоматов над деревьями; также предложен

эффективный метод совместного вывода индуктивных инвариантов в этом классе по средством вывода инвариантов в подклассах.

4. Теоретическое исследование существующих и предложенных в рамках диссертации классов индуктивных инвариантов, включая формулировку и доказательство лемм о «накачке» для языка ограничений и для языка ограничений расширенного функцией размера терма.

Таким образом, **теоретической ценность** данной диссертации является новые методы автоматического вывода индуктивных инвариантов, новый класс индуктивных инвариантов, а также леммы о «накачке» для языков первого порядка. Все эти результаты обладают несомненной **научной новизной** и открывают интересные возможности для дальнейших теоретических исследований в этой области, например, могут быть перенесены на теорию массивов и ряд других областей теоретической верификации.

Практическая ценность работы заключается в предложенной пилотной реализации предложенных методов, а также в том, что он могут быть использованы при создании статических анализаторов для языков с алгебраическими типами данных – а эти языки в настоящее время активно используются в индустрии (например, Rust, Scala, Solidity, Haskell и OCaml).

Сделаем краткий обзор диссертационной работы по главам.

Во введении обосновывается актуальность представленного диссертационного исследования, приведена постановка задачи и сформулированы результаты, которые выносятся на защиту.

В первой главе дается обзор существующих в области результатов, даются основные определения и приводятся наиболее существенные результаты, используемые в данной работе. Формально определяется язык ограничений, логика первого порядка и алгебраические типы данных, системы дизъюнктом Хорна, базовые понятия формальных языков над деревьями.

Во второй главе излагается новый метод автоматического вывода индуктивных инвариантов систем для АТД на основе автоматов над деревьями при помощи инструментов автоматического доказательства теорем. В этой главе, помимо изложения самого метода, представлено доказательство его корректности. Также описывается, как предложенный метод может быть применён для вывода регулярных инвариантов при помощи инструментов поиска конечных моделей.

В третьей главе описан новый автоматический метод вывода индуктивных инвариантов на основе трансформации программы и поиске конечных моделей в классе синхронных автоматов над деревьями. Последние являются естественным расширением расширения автоматов над деревьями, способными выражать синхронные отношения. Предложенный метод сопровождён необходимыми теоретическими построениями, а именно, доказано, что метод корректен и полон для вывода синхронных регулярных инвариантов.

В четвертой главе представлен новый класс индуктивных инвариантов, основанный на булевой комбинации классических инвариантов и автоматов над деревьями. Этот класс оказывается существенно шире существующих классов и позволяет выводить значительное количество новых инвариантов. Для этого класса представлен эффективный метод совместного (collaborative) вывода индуктивных инвариантов в этом классе по средством

вывода инвариантов в подклассах. В этой главе доказывается, что предложенный метод корректен при условии корректности методов вывода инвариантов в подклассах. Также доказано, что алгоритм построения абстрактного контрпримера по контрпримеру к остаточной системе имеет линейную сложность.

В пятой главе представлено теоретическое сопоставление существующих и предложенных в работе классов индуктивных инвариантов для программ с алгебраическими типами данных. Рассмотрены только те классы, известные из литературы, для которых существуют полностью автоматические методы вывода инвариантов. В главе выполнено значительное количество теоретических построений для доказательства формальных отношений рассматриваемых классов. В частности, важным теоретическим результатом является формулировка и доказательство лемм о «накачке» для языка ограничений и для языка ограничений расширенного функцией размера терма. Эти леммы позволяют доказывать невыразимость множества в данном языке, что даёт возможность отделять классы индуктивных инвариантов друг от друга.

В шестой главе представлена программная реализация предложенных методов, а также результаты экспериментального исследования. Описана архитектура предложенного программного инструмента с использованием диаграмм компонент UML.

В заключении диссертации кратко изложены основные результаты диссертации, приведены направления дальнейших исследовательских работ.

Следует отметить, что каждая глава снабжена выводами и небольшой обзорной частью, что позволяет читателю легко удерживать нить повествования. Также, текст диссертации содержит большое количество примеров и рисунков, делающих излагаемый материал доступным для широкого круга специалистов.

К представленной диссертации имеется ряд **замечаний**.

1. Язык диссертации в целом часто страдает от излишней лаконичности и чрезмерно больших и сложных предложений. Рекомендую автору использовать более «легковесный» стиль изложения.
2. Весьма запутанно изложена архитектура программного инструмента в той ее части, где описывается реализация предложенных в диссертации теоретических методов (стр. 76-77).
3. Эксперименты (глава 6) описаны весьма кратко, исследовательские вопросы весьма сложны для понимания, нет описания используемых метрик, графики для представления результатов (стр. 87–88) неочевидны (возможно, графики недостаточно подробно описаны?).
4. Доказательства ключевых теорем о синхронных языках над деревьями крайне сложны для понимания (глава 3, теоремы 10 и 11): нотация для всех возможных синтаксических комбинаций термов (см. стр. 39–40) могла быть лучше проработана.
5. Описание коллаборативного вывода комбинированных инвариантов (глава 4) можно существенно упростить: вместо того, чтобы начинать с систем переходов (раздел 4.1, стр. 46–55), а затем описывать дизъюнкты Хорна как их частный случай (раздел 4.2, стр. 56–61), можно было бы сразу описывать подход на дизъюнктах Хорна.

Тем не менее, указанные замечания не являются существенными, не влияют на корректность представленных результатов и их научную новизну.

Таким образом, можно сделать вывод о том, что диссертация Ю.О. Костюкова на тему «Автоматический вывод индуктивных инвариантов программ с алгебраическими типами данных» соответствует требованиям, установленным Приказом от 19.11.2021 № 11181/1 «О порядке присуждения научных степеней в Санкт-Петербургском государственном университете», а сам соискатель Юрий Олегович Костюков заслуживает присуждения ему ученой степени кандидата физико-математических наук по специальности 2.3.5. Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей. Нарушения пунктов 9 и 11 указанного Порядка в диссертации не обнаружены.

Член диссертационного совета,
доктор физико-математических наук,
член-корреспондент РАН, профессор,
заведующий кафедрой прикладной кибернетики Санкт-Петербургского государственного университета

Николай Владимирович Кузнецов

15.02.2024