

Отзыв

официального оппонента Кознова Дмитрия Владимировича на диссертацию Качанова Владимира Владимировича «Методы и программные средства автоматического рецензирования исходного кода», представленную на соискание ученой степени кандидата технических наук по специальности 2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей

Актуальность работы

С ростом масштабов и сложности программных систем проблема поддержания больших кодовых баз в консистентном состоянии становится все более актуальной и превращается из технологической задачи в бизнес-задачу. Отсутствие адекватных решений здесь может привести к существенному удорожанию поддержке существующих и выпуску новых продуктов.

Процесс рецензирования программного кода представляет собой важную активность, способную существенно улучшить состояние активно меняющейся кодовой базы. Задача рецензирования вносимых изменений заключается не только в поиске грубых ошибок, но и проверку соответствия кода установленным требованиям проекта, оценку его читаемости, качества архитектурных решений и потенциала для оптимизации.

В диссертационной работе предлагается метод и программный комплекс для автоматического рецензирования программного кода, основанный на применении методов машинного обучения и специализированной языковой модели Qwen2.5-Coder-32B. Важным аспектом практической применимости подобных систем является высокий уровень доверия к ним со стороны разработчиков. Значительное количество ложноположительных срабатываний или частая генерация неприменимых рекомендаций может привести к отказу от использования такой системы. Соответственно, в работе уделено особое внимание оценке качества генерации рецензий и практической ценности предлагаемых исправлений кода. Стоит также отметить разработку методов детектирования признаков технического долга — категории антипаттернов, которые не имеют строгих формальных определений. Подобные дефекты могут долгое время оставаться в коде, существенно увеличивая стоимость разработки в критические моменты (например, при выпуске новой версии системы, нового продукта линейки про-

дуктов), потребовав срочного исправления критических ошибок и недоработок (влияющих, например, на производительность), реализации «костылей» для внедрения новой функциональности и т.д. Таким образом, автоматизация рецензирования исходного кода является актуальной задачей.

Краткое описание работы

Диссертация состоит из введения, четырех глав и заключения.

Во введении сформулированы цель и задачи работы, обоснована их актуальность, сформулированы результаты, выносимые на защиту, описана их научная новизна и практическая значимость, а также выполнен обзор публикаций автора.

В первой главе представлен анализ существующих в данной области исследований, в частности, по автоматизации рецензирования кода и детектированию антипаттернов. Сравнительный обзор существующих инструментов (PMD, SonarQube, Designite) выявил низкую согласованность их результатов, что обосновывает необходимость новых методов.

Во второй главе представлен разработанный алгоритм сбора размеченных антипаттернов из истории проектов с открытым исходным кодом. Модели, обученные на этих данных, показали значительное улучшение эффективности поиска антипаттернов в коде (прирост F1-меры составил до 78%).

В третьей главе детально рассмотрена задача рецензирования программного кода. Проанализированы существующие подходы, предложены модификации для улучшения качества данных, а также новые методы для их обработки, исследованы метрики оценки сходства исходного кода, наиболее подходящие для целевой задачи.

В четвертой главе представлены эксперименты для задачи генерации рецензий. Подход с большими языковыми моделями продемонстрировал наилучший результат. Представленные в третьей главе результаты позволили увеличить долю применимых сгенерированных комментариев до 37%, а долю неплохих — до 77%.

В заключении приводятся результаты диссертационной работы.

Результаты работы

В ходе диссертационного исследования были получены следующие результаты.

- Разработан и реализован метод сбора данных об исправлениях антипаттернов кода из истории разработки проектов с открытым исходным кодом. Показано, что собранные таким образом данные позволяют обучать новые более точные классификаторы.

- Разработаны и реализованы методы детектирования антипаттернов исходного кода, сочетающие метрики кода с методами машинного обучения.
- Полученный подход позволяет повысить точность и полноту результатов классификации.
- Разработаны методы автоматического рецензирования исходного кода, основанные на применении языковых моделей, которые предлагают разработчику применимые советы по улучшению исходного кода. Показана эффективность внедренных новых модификаций в существующие методы.
- Разработан и реализован программный комплекс, сочетающий работу детекторов антипаттернов исходного кода в систему автоматического рецензирования.

Эти результаты соответствуют паспорту специальности 2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей; они также обладают теоретической новизной и практической значимостью, позволяя усовершенствовать существующие инструменты анализа программного кода и повысить качество разработки ПО.

Достоверность и обоснованность результатов

Достоверность результатов подтверждается экспериментальными исследованиями, проведенными на реальных проектах, а также сравнением с существующими подходами.

Основные результаты докладывались и обсуждались на 3 научных конференциях. По теме диссертационной работы автором опубликовано 6 статей, включая 5 статей в рецензируемых журналах, рекомендованных ВАК РФ. Кроме того, было получено свидетельство о регистрации программы для ЭВМ.

Теоретическая и практическая значимость

Теоретическая значимость представленного исследования заключается в развитии методологической базы автоматического анализа кода посредством сочетания статического анализа, машинного обучения и больших языковых моделей.

Практическая значимость состоит в создании готового к программному инструменту, который позволяет надеяться на сокращение издержек при сопровождении крупных кодовых баз и повышение качества процесса рецензирования изменений исходного кода. Также данный инструмент эффективно выявляет дефекты и улучшает читаемость кода. Следует отметить, что представлен-

ный инструмент прошел практическую апробацию, о чём имеется акт внедрения.

Замечания

1. Описание результатов работы в «Заключении», а также в автореферате, существенно выиграло бы от указания использованных технологий (алгоритмов машинного обучения, использованных больших моделей). В текущем варианте читателю приходится основательно знакомиться с текстом работы, чтобы выяснить несложную вещь – а что же было сделано диссертантом? Выяснение этого факта затрудняется наличием большого количества экспериментов (и, соответственно, опробованных методов), описанных по ходу изложения результатов в тексте работы.
2. В главе 3 при описании экспериментов с большими языковыми моделями отсутствуют указания гиперпараметров запусков. Эта информация является важным аспектом воспроизводимости результатов.
3. Таблицы 4.7 и 4.8 (точнее, указанные в них большие модели) неочевидным образом соответствуют друг другу, хотя первая обсуждает большие модели применительно к пониманию/генерации программного кода, а вторая – применительно к генерации рецензий программного кода.
4. Работа несомненно выиграла бы, если бы содержала акцент на больших кодовых базах, включая описание сценариев использования предложенных методов и инструмента.

Однако указанные замечания не снижают значимости полученных результатов и не влияют на общую положительную оценку диссертационного исследования.

Заключение

Диссертация Качанова Владимира Владимировича является завершенной научно-исследовательской работой, содержащей теоретические и практические результаты, имеющие важное значение для области анализа исходного кода программного обеспечения. Работа соответствует требованиям ВАК, предъявляемым к кандидатским диссертациям по специальности 2.3.5 – математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей, а автор заслуживает присуждения ученой степени кандидата технических наук.

Профессор кафедры системного
программирования Санкт-Петербургского
государственного университета, доктор
технических наук

Кознов Дмитрий Владимирович

24.11.2025