

УДК 1082

**ОБЪЕДИНЕНИЕ КЛАСТЕРИЗАЦИИ И  
КЛАССИФИКАЦИИ ДЛЯ ОЦЕНКИ  
КАЧЕСТВА ПРОГРАММНОГО  
ОБЕСПЕЧЕНИЯ***Шабанов Р. М., Левченков А. Н.*

Донской государственной технической  
университет Ростов-на-Дону, Российская  
федерация

[reslin@gmail.com](mailto:reslin@gmail.com)

Исходный код и интеллектуальный анализ метрик были использованы в успешном оценивании качества программного обеспечения. Представлен подход к интеллектуальному анализу данных, который включает в себя кластеризацию классов Java, а также классификацию извлеченных кластеров для оценки качества внутреннего программного обеспечения. Используются классы Java как сущности и статические метрики и атрибуты интеллектуального анализа данных.

**Ключевые слова:** кластеризация метрик, оценка качества, анализ программных средств, объединение и классификация.

**Введение.** Оценка качества является важным вопросом при разработке программного обеспечения, решаемым различными методами, во многих случаях, включая метрики. Увеличивающийся объем кода порождает повышение спроса на автоматизацию процесса. Опыт показывает, что сбор и анализ мелкозернистых тестовых дефектов из большого, сложного программного обеспечения сложны. Было показано, что интеллектуальный анализ данных облегчает оценку качества исходного кода, а также метрики, извлеченные из кода. Статический анализ с использованием метрик качества программного обеспечения означает, что код анализируется без необходимости выполнения программы. Инструменты статического анализа для поиска низкоуровневых ошибок в программировании особенно полезны для оценки ремонтпригодности, понятности, компиляции и расширяемости программного обеспечения. Статический анализ также можно использовать в процессе разработки, обеспечивая раннее обнаружение неисправностей на точке, где код не полностью функционален или исполняем. Несколько инструментов такого рода уже существуют. Однако проблемы масштабирования размера препятствуют оценке больших наборов данных. Статический анализ вряд ли будет принят для улучшения качества программного обеспечения на практике, если оно не выходит за рамки небольших показателей.

**Основная часть.** Для решения этой проблемы предлагается статический анализ с использованием объектно-ориентированной метрики в сочетании с методами интеллектуального анализа данных, используемого для проверки больших и действующих программных средств. Использование этих показателей отражает атрибуты исходного кода программы, такие как объем, размер, сложность, зацепление. Наш подход применим для систем Java, который может быть легко

UDC 1082

**COMBINING CLUSTERING AND  
CLASSIFICATION FOR SOFTWARE  
QUALITY EVALUATION***Shabanov R.M., Levchenkov A. N.*

Don State Technical University, Rostov-on-Don,  
Russian Federation

[reslin@gmail.com](mailto:reslin@gmail.com)

Source code and metric mining have been used to successfully assist with software quality evaluation. This paper presents a data mining approach, which incorporates clustering Java classes, as well as classifying extracted clusters, in order to assess internal software quality. We use Java classes as entities and static metrics as attributes for data mining.

**Keywords:** metrics clustering, quality assessment, software analysis, aggregation and classification.

расширен для обслуживания других объектно-ориентированных языков. Интеллектуальный анализ данных в статическом анализе позволяет управлять большими объемами данных и способен давать неожиданные результаты. Эта работа не только фокусируется на оценке качества программного обеспечения, но также рассматривает пригодность метрик для анализа, который является полезной и новой частью процесса изучения. В дальнейшем мы используем кластеры как метки классов и используем C4.5 алгоритм классификации дерева решений при оценивании выбранной метрики, чтобы показать их важность для определения кластеров и оценки качества программного обеспечения. Использование C4.5 позволяет установить, какие метрики играют важную роль в процессе оценки и основные показатели, которые не влияют на построение дерева. Мы использовали исходный код из различных больших систем Java с открытым исходным кодом для проверки этого подхода. Эксперименты показали, что объединение кластеризации с классификацией дает лучшие результаты, чем самостоятельные кластеры.

Несмотря на положительные результаты такого подхода, существуют определенные ограничения. Самым важным ограничением является отсутствие полной автоматизации, так как решения должны приниматься во время процесса интеллектуального анализа данных. Это ограничение может быть преодолено в ходе дальнейшей работы над доступностью больших проектов с открытым исходным кодом и возможностью отслеживания больших проектов компании.

**Выводы.** Предлагаемый подход дает хорошие результаты в области обнаружения ошибок с предварительной обработкой данных, группирования/организации классов с кластеризацией методов и категоризации кластеров с использованием модели описания с классификацией. Руководители проектов имеют возможность лучше понять преимущества и недостатки кода. Этот подход облегчает инженерам по техническому обслуживанию определить классы, которые подвержены ошибкам, сложны в понимании и поддержке, а также оценить возможности расширения системы.