

Секция «Нейронные сети в научной работе студента (совместно с Сибирским государственным университетом науки и технологий имени М.Ф. Решетнева)»

Применение нейронных сетей для автоматического анализа и документирования кода

Научный руководитель – Дубровина Оксана Васильевна

Белоусова Екатерина Евгеньевна

Студент (бакалавр)

Тамбовский государственный технический университет, Тамбовская область, Россия

E-mail: Katya78947@bk.ru

Программисты, независимо от опыта и уровня квалификации, постоянно сталкиваются с необходимостью анализа чужого кода. Это происходит в различных ситуациях, таких как работа в команде, поддержка унаследованного кода, исследование open-source проектов или при обучении. Однако, понимание чужого кода часто является сложной и трудоемкой задачей, особенно для студентов, которые только начинают свой путь в программировании.

Во-первых, сложность анализа чужого кода может заключаться в незнакомом синтаксисе или стиле в ситуации, когда код был написан на незнакомом языке программирования или с использованием другого стиля кодирования.

Во-вторых, существует проблема некачественной документации или ее отсутствия в целом. Поддержание актуальной и качественной документации – критически важная, но часто игнорируемая задача в разработке программного обеспечения. Ручное документирование трудоемко, подвержено ошибкам и часто устаревает. При таком подходе создания документации разработчики могут экономить время и силы, что приводит к ряду последствий:

- если в коде присутствуют только базовые комментарии, то по ним сложно или невозможно будет объяснить сложные алгоритмы и логику работы;
- если документация не обновляется при внесении изменений в код, то это приводит к расхождению между кодом и его описанием;
- если описана только часть функциональности кода, то это затрудняет понимание общей картины;
- если документация написана нечетким и сложным языком, то это затрудняет ее понимание.

Наконец, при анализе чужого кода программисту может не хватать информации о целях и задачах проекта, о принятых решениях и об используемых технологиях.

Существуют инструменты, которые могут генерировать документацию на основе комментариев в коде, однако они часто не способны понять семантику кода и предоставить высокоуровневое описание его функциональности. Они только извлекают информацию из комментариев и структуры кода, не анализируя логику работы и взаимосвязи между различными элементами кода.

Например, Doxygen – один из самых популярных инструментов для автоматической генерации документации из исходного кода. Он поддерживает множество языков программирования (C++, C, Java, Python и другие) и генерирует документацию в различных форматах (HTML, LaTeX, PDF и другие). Doxygen анализирует код и генерирует документацию на основе специальных комментариев (docstrings) в коде. Его недостатками являются зависимость от комментариев и отсутствие понимания семантики кода [1].

Javadoc, Sphinx и другие так же, как и Doxygen являются полезными инструментами и позволяют упростить процесс создания документации. Однако, они имеют существенное

ограничение: не могут автоматически генерировать описания алгоритмов, логики работы или архитектуры системы, если эта информация явно не представлена в коде или комментариях.

В последние годы наблюдается значительный рост интереса к применению методов машинного обучения, и, в частности, нейронных сетей, для решения различных задач, связанных с анализом кода.

Возможный функционал нейросетей для анализа кода:

- извлечение информации о его функциональности, даже если документация отсутствует или неполная;
- понимание семантики кода, то есть его смысл и предназначение;
- создание визуализации структуры кода, что позволяет программистам лучше понимать взаимосвязи между различными элементами кода.

Для анализа синтаксического дерева (AST) и извлечения информации о структуре кода используются сверточные нейронные сети (CNN) или графовые нейронные сети (GNN) – Graph Convolutional Networks (GCN), Graph Attention Networks (GAT). Для анализа последовательностей токенов и понимания семантики кода применяются рекуррентные нейронные сети (RNN) – LSTM, GRU или трансформеры – BERT, GPT, CodeBERT, CodeT5. Объединение методов анализа AST и семантического анализа позволяет получить более полную информацию о коде [2].

Возможный функционал нейросетей для документирования кода:

- генерация комментариев для кода, описывающие его функциональность и логику работы;
- создание документации на основе анализа кода и комментариев.

Documatic – это инновационная нейросеть, созданная для автоматизации процесса создания документации к программному коду [3]. Он предлагает множество уникальных возможностей:

- создание документации по текстовым запросам, с помощью анализа структуры кода и комментариев разработчиков, что позволяет быстро получать полные и точные описания функций, классов и методов;
- объяснение работы отдельных частей кода, что особенно полезно для новых членов команды или при работе с устаревшими проектами;
- построение карты зависимостей между файлами и компонентами проекта, что помогает лучше понять архитектуру приложения и упрощает его поддержку;
- генерация не только новой документации, но и улучшение уже имеющейся, что обеспечивает более точные и последовательные описания.

Подводя итог, следует отметить, что использование нейронных сетей для автоматического анализа и документирования кода не только облегчает работу с существующими проектами, но и предоставляет студентам-программистам ценный инструмент для обучения и исследований. Это позволяет им быстрее осваивать новые технологии, эффективно работать в команде и создавать более качественное программное обеспечение.

Источники и литература

- 1) Абросимов, Р. А. Применение методов машинного обучения в статическом анализе кода // ТОГУ-Старт: Фундаментальные и прикладные исследования молодых. Хабаровск, 2022. С. 213-217.
- 2) Гольдберг, Й. Нейросетевые методы в обработке естественного языка: руководство / Й. Гольдберг; перевод с английского А. А. Слинкина. — Москва: ДМК Пресс, 2019. — 282 с.

- 3) Соробин, А. Б. Сверточные нейронные сети: примеры реализаций: учебно-методическое пособие / А. Б. Соробин. — Москва: РТУ МИРЭА, 2020. — 159 с.