

Непрерывная интеграция методов обработки ДЗЗ для исследования Байкальской природной территории

Ю.В. Авраменко¹, А.К. Попова¹, Г.М. Ружников¹, Р.К. Фёдоров¹

¹Институт динамики систем и теории управления имени В.М. Матросова Сибирского отделения РАН, Лермонтова 134, Иркутск, Россия, 664033

Аннотация

В работе рассмотрена реализация непрерывной интеграции методов обработки ДЗЗ для включения новых методов в цифровую платформу Байкальской природной территории. Предложенный подход позволяет упростить сравнение, применение новых и известных методов, технологий проведения научных исследований Байкальской природной территории.

Ключевые слова

Пространственные данные, ДЗЗ, мониторинг, сервисы

1. Введение

Дистанционное зондирование Земли (ДЗЗ) является единственным регулярным источником данных, покрывающих всю территорию Байкальской природной территории (БПТ). Результаты обработки данных ДЗЗ используются для решения множества практических задач. Существуют различные цифровые платформы, реализующие регулярную обработку данных ДЗЗ и публикацию результатов для общего доступа. Обычно набор методов обработки в таких системах ограничен и расширяется небольшим коллективом разработчиков. В наши дни активно растёт число и эффективность методов и технологий обработки ДЗЗ (особенно на основе нейронных сетей), которые необходимо включать в цифровые платформы. Поэтому одной из основных задач является выбор наиболее эффективного метода обработки данных ДЗЗ на основе объективного сравнения результатов применения методов на реальных данных и интеграция метода в цифровую платформу. Применение включает установку и настройку программного обеспечения, обучение метода и программирование (если требуется), то есть, по сути, его можно сравнить с разработкой некоторой функции программного обеспечения. Для процедуры выбора часто привлекают профильных специалистов. Таким образом, актуально проведение открытых конкурсов для выбора эффективного метода, по аналогии с Kaggle [2] и другими системами [3], на реальных задачах, сравнение результатов и включение в цифровую платформу.

2. Непрерывная интеграция новых методов обработки и анализа

Цифровая платформа должна реализовывать жизненный цикл методов обработки данных ДЗЗ, поддерживая непрерывную интеграцию новых методов на основе сравнения их и применения наиболее эффективных. Жизненный цикл представлен на Рисунке 1.



Рисунок 1: Жизненный цикл методов обработки данных ДЗЗ

Непрерывная интеграция (Continuous Integration) – это практика разработки программного обеспечения, включающая постоянное объединение рабочих версий в базовую. Часть этапов

жизненного цикла методов обработки данных ДЗЗ уже реализована в ИДСТУ СО РАН. Рассмотрим этапы подробнее.

Разработка и обучение методов в рамках цифровой платформы выполняется с помощью системы Jupyter Notebook [4], эффективного инструмента интерактивного анализа данных. Она объединяет математические уравнения, программный код и визуализации, а поддержка языка Python способствует использованию большого числа библиотек обработки данных NumPy, Scipy, Pandas, Matplotlib, Seaborn, SciKit-Learn, TensorFlow, Theano, Keras и т.д. [5]. Проведена интеграция Jupyter Notebook и прикладной цифровой платформы. Обучающие выборки формируются на основе пространственно-временных данных Байкальской природной территории.

Добавление новых методов осуществляется регистрацией в каталоге WPS сервисов. Разработанные в Jupyter Notebook методы могут быть опубликованы в виде сервисов с помощью PyWPS. Обработка данных предполагает автоматическое (без участия разработчиков) выполнение всех методов на верификационной выборке.

Выбор метода производится на основе результатов работы на верификационной выборке. Учитывая размеры территории БПТ и объем данных, критериями выбора метода являются качество и время работы. Применение подразумевает запуск выбранного метода для всей территории БПТ, сохранение и регистрацию результатов в каталогах данных.

3. Заключение

Непрерывная интеграция методов ДЗЗ позволит снизить трудоемкость применения и сравнения новых методов для проведения исследований Байкальской природной территории. Открытая цифровая платформа обеспечивает подключение независимых разработчиков.

4. Благодарности

Работа выполнена при частичной поддержке гранта Минобрнауки России № 075-15-2020-787, ЦКП ИИВС ИРНОК.

5. Литература

- [1] Осипов, Ю.М. Цифровая платформа как институт эпохи технологического прорыва / Ю.М. Осипов, Т.Н. Юдина, И.З. Гелисханов // Экономические стратегии. – 2018. Т. 5, № 155. – С. 22-29.
- [2] Welcome Kaggle to Google Cloud // Google Cloud Platform Blog [Electronic resource]. – Access mode: <https://cloud.google.com/blog/products/gcp/welcome-kaggle-to-google-cloud> (дата обращения 25.10.2020).
- [3] Borowski, D. The 10 Best Coding Challenge, 2018 [Electronic resource]. – Access mode: <https://medium.com/coderbyte/the-10-best-coding-challenge-websites-for-2018-12b57645b654> (дата обращения 25.10.2020).
- [4] Project Jupyter exists to develop open-source software, open-standards, and services for interactive computing across dozens of programming languages [Electronic resource]. – Access mode: <https://jupyter.org> (25.10.2020).
- [5] Stancin, I. An overview and comparison of free Python libraries for data mining and big data analysis / I. Stancin, A. Jovic // 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). – Opatija, CROATIA, 2019. – P. 977-982.