

УДК 004.89

DOI: 10.31673/2412-9070.2021.024952

О. М. НІКОЛАЄНКО, студент;

О. В. ПОЛОНЕВИЧ, канд. техн. наук, доцент,
Державний університет телекомунікацій, Київ

ВИКОРИСТАННЯ ФРЕЙМВОРКУ MLOPS ДЛЯ ЗБІЛЬШЕННЯ РІВНЯ ЖИТТЄВОГО ЦИКЛУ МАШИННОГО НАВЧАННЯ В ІОТ

Останніми роками особливої популярності набули пристрої IoT, що призвело до великих обсягів даних та попиту на зв'язок з низькою затримкою, зумовлюючи попит на мережі 5G. Цей зсув в інфраструктурі дає змогу приймати рішення в режимі реального часу з використанням штучного інтелекту для додатків IoT. Штучний інтелект речей (AIoT) — це поєднання технологій штучного інтелекту (AI) з інфраструктурою Інтернету речей (IoT) для досягнення більш ефективних операцій IoT та прийняття рішень. З'являються крайові обчислення, щоб активувати системи AIoT. Крайові обчислення дають можливість генерувати статистику та приймати рішення у джерелі даних, зменшуючи обсяг даних, що надсилаються до хмари та центрального сховища. Екосистема для полегшення крайових обчислень для додатків AIoT стала важливою щодо прийняття рішень у режимі реального часу.

Ключові слова: хмарна платформа; штучний інтелект; Інтернет речей; крайові обчислення; великі дані; машинне навчання.

Вступ

З погляду MLOps (операцій машинного навчання) IoT Edge — це ще одна платформа розгортання. Однак, якщо ми розгортаємо моделі на IoT Edge, нам потрібно зважати на деякі додаткові міркування. Моделі MLOps, націлені на IoT Edge, мають працювати в автономному режимі. Моделі IoT більш сприйнятливі до дрейфу даних через високу швидкість передавання даних. Моделі машинного навчання IoT мають бути розгорнуті на різних цільових платформах, і нам потрібно використовувати можливості цих платформ.

Аналіз дослідження. Дослідження в галузі програмної інженерії часто мають міждисциплінарний характер, і ця стаття не є винятком. Сфера досліджень програмної інженерії часто визначається як перетин інформаційних технологій, бізнесу та оброблення даних. У цьому дослідженні діловий аспект зосереджено на приватних та державних компаніях, які прагнуть оптимізувати ресурси, підвищити ефективність та модернізувати свої наявні служби, щоб вони були розумнішими та забезпечувались завдяки силі штучного інтелекту та за допомогою даних у реальному часі, які генеруються з використанням Інтернету речей. У соціальному контексті дослідження ці підприємства мають на меті оптимізувати ресурси з погляду енергії, часу, грошей та людських ресурсів. Оптимізація цих ресурсів сприятиме підвищенню ефективності, зростанню та впровадженню сучасних технологій, зокрема крайових обчислень, штучного інтелекту речей (AIoT) та мереж із низькою затримкою.

Для розв'язання нашого дослідницького питання будемо слідувати методу дизайну, запропонованому Wieringa. Наука про дизайн — це проектування та дослідження артефактів у контексті. Артефакти, які ми вивчаємо, покликані взаємодіяти з проблемним контекстом, аби щось покращити в цьому контексті (рис. 1).

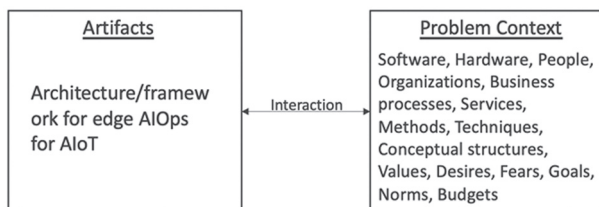


Рис. 1. Предмет науки про дизайн: артефакт, що взаємодіє з контекстом

У нашому разі артефактом у контексті буде основний фреймворк MLOps для систем AIoT, який ітеративно взаємодіє з контекстом проблеми, що зводиться до нашого питання. Також будемо оцінювати та перевіряти масштабованість та надійність артефактів у контексті.

Метою дослідження є принципи ефективного застосування фреймворку MLOps для збільшення рівня циклу машинного навчання в системах IoT.

Основна частина

Для керування пристроями IoT, великими даними та швидшого прийняття рішень крайові обчислення пропонують можливості здійснювати обчислення, близькі до джерела даних. Це ідеальний вибір, коли йдеться про випадки з пристроями IoT, низькі затримки та операції в реальному часі, оскільки для масштабного обслуговування IoT-додатків важливо мати правильну синергію для хмари та краю. Незабаром рішення IoT будуть охоплювати набагато ширший спектр вимог, беручи масштабованість та надійність за основу.

© О. М. Ніколаєнко, О. В. Полоневич, 2021

Оскільки ми бачимо, що інтегрований інтелект стрімко впроваджується в галузі, життєво важливо синергізувати крайові обчислення та штучний інтелект, щоб надалі приймати рішення в режимі реального часу поблизу джерел даних для надійних та масштабованих систем.

Edge Computing (крайові (периферійні) обчислення) — це процес виконання обчислювальних завдань, фізично близьких до цільових пристроїв, а не в хмарі або на самому пристрої. Це дає можливість отримувати знання, уявлення та приймати рішення поблизу джерела даних.

Призначення крайових обчислень — наблизити обчислення до джерела даних та розвантажити централізовані обчислення до децентралізованих. Крайові обчислення дають змогу застосовувати різні алгоритми машинного навчання, уможливаючи створення нових вражень та нових можливостей у багатьох галузях промисловості, починаючи від підімкненого будинку і закінчуючи безпекою, спостереженням та автомобілебудуванням.

Це також забезпечує безпечну та надійну роботу для оброблення даних та координації роботи кількох пристроїв. Оглядову схему побудови безпечної та надійної інтелектуальної архітектури на краю зображено на рис. 2.

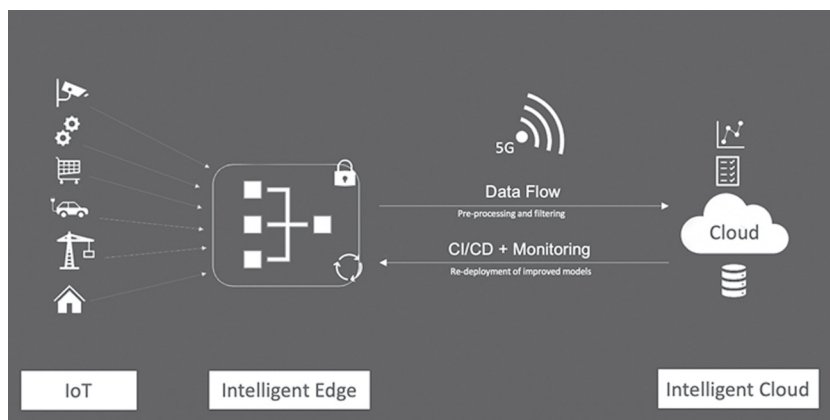


Рис. 2. Схема побудови інтелектуальної архітектури на краю

MLOps. Операції машинного навчання (MLOps) засновано на принципах та практиці DevOps, що підвищують ефективність робочих процесів, наприклад неперервна інтеграція, доставляння та розгортання. MLOps застосовує ці принципи до процесу машинного навчання з метою:

- швидкого експериментування та розроблення моделей;
- швидкого впровадження моделей у виробництво;
- гарантії якості.

Машинне навчання **Azure** надає такі можливості MLOps:

- створювати відтворювані конвеєри машинного навчання. Конвеєри машинного навчання дають змогу визначати повторювані та повторно використовувані кроки для процесів підготовки даних, навчання та оцінювання;
- створювати повторно використовувані програмні середовища для навчання і розгортання моделей;
- реєструвати, упаковувати та розгортати моделі з будь-якого місця, а також відстежувати зв'язані метадані, потрібні для використання моделі;
- здійснювати збір даних керування для неперервного життєвого циклу машинного навчання. Реєстрована інформація може мати в собі публікації моделі, історію внесених змін та коли моделі було розгорнуто або використовувались у процесі;
- повідомляти та попереджати про події в життєвому циклі машинного навчання. Наприклад, завершення експерименту, реєстрація моделі, розгортання моделі та виявлення відхилення даних;
- здійснювати моніторинг додатків машинного навчання щодо операційних проблем і проблем, пов'язаних із машинним навчанням; порівнювати вхідні дані моделі між навчанням і висновком, досліджувати метрики, специфічні для моделі; забезпечувати моніторинг і оповіщення у вашій інфраструктурі машинного навчання;
- автоматизувати неперервний життєвий цикл машинного навчання за допомогою машинного навчання Azure та Azure Pipelines. Використання конвеєрів дає можливість часто оновлювати моделі, тестувати нові моделі та постійно впроваджувати нові моделі машинного навчання разом з іншими додатками та службами.

Принципи використання фреймворку MLOps для збільшення рівня циклу машинного навчання у системах IoT за допомогою платформи Azure IoT Edge передбачають розгляд трьох кроків.

Перший крок — побудова та навчання (рис. 3). На цьому кроці створюються відтворювані моделі та навчальні конвеєри багаторазового використання. Коли конвеєр CI (*Continuous Integration*) запускається, він щоразу під час реєстрації коду публікує оновлений конвеєр машинного навчання Azure після побудови коду та запуску набору тестів. Конвеєр побудови має низку модульних тестів і тестів якості коду.

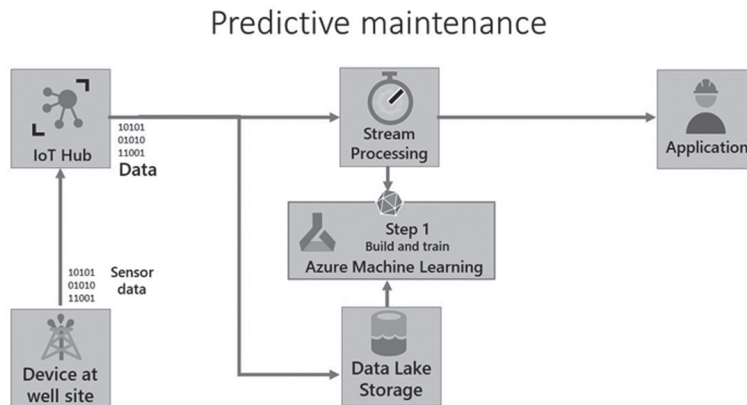


Рис. 3. Перший крок: побудова та навчання

Другий крок — пакування та розгортання (рис. 4). На цьому кроці здійснюється пакування, перевірка та розгортання моделі. Конвеєр спрацьовує щоразу, коли доступний новий артефакт. Зареєстрована модель упаковується разом зі сценарієм оцінювання та залежностями Python (файл Conda YAML) в образ Docker для подальшого запуску моделі на сервері.

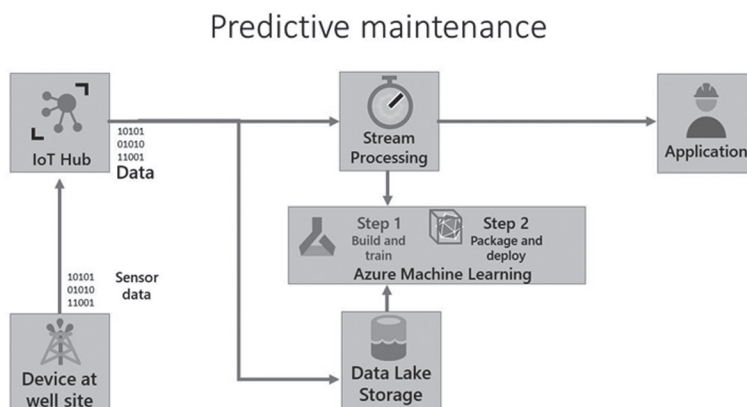


Рис. 4. Другий крок: пакування та розгортання

Третій крок — моніторинг та повторне навчання (рис. 5). Конвеєр машинного навчання організовує процес перекваліфікації моделі асинхронно. Будь-яке повторне навчання може бути запущено за розкладом або коли нові дані стануть доступними, викликавши опубліковану кінцеву точку REST з попереднього кроку. На цьому кроці буде здійснено перекваліфікацію, оцінювання та реєстрацію навченої моделі.

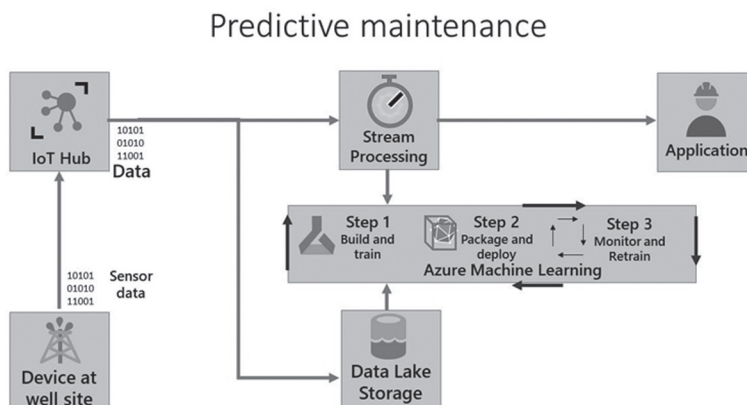


Рис. 5. Третій крок: моніторинг та повторне навчання

Висновки

Таким чином, фреймворк MLOps у поєднанні з хмарною платформою Microsoft Azure дасть можливість полегшити впровадження машинного навчання на краю для додатків AIoT та надасть відповідні для цього послуги, інструменти та методи для оптимізації витрат, операцій та ресурсів для спрощення ефективних масштабних хмарних операцій, що забезпечить збільшення життєвого рівня циклу машинного навчання, зокрема доставляння, розгортання та моніторинг моделей на кінцевих пристроях IoT.

Список використаної літератури

1. *Общие сведения об MLOps для IoT Edge* [Електронний ресурс]. URL: <https://docs.microsoft.com/ru-ru/learn/modules/intro-mlops-iot-edge/1-introduction>
2. *MLOps: Model management, deployment, and monitoring with Azure Machine Learning* [Електронний ресурс]. URL: <https://docs.microsoft.com/en-us/azure/machine-learning/concept-model-management-and-deployment>
3. *How MLOps works for IoT Edge* [Електронний ресурс]. URL: <https://docs.microsoft.com/uk-ua/learn/modules/intro-mlops-iot-edge/3-how-mlops-works>
4. *How DevOps For Machine Learning (MLOps) Works For IoT Edge With Azure IoT Edge* [Електронний ресурс]. URL: <https://nayeen.info/devops-for-machine-learning-with-azure-iot-edge/>
5. *Edge MLOps framework for AIoT applications* [Електронний ресурс]. URL: https://www.theseus.fi/bitstream/handle/10024/342167/Raj_Emmanuel.pdf?sequence=2&isAllowed=y
6. *Wieringa Roel J. Design Science Methodology for Information Systems and Software Engineering* [Електронний ресурс]. 2014. URL: https://ris.utwente.nl/ws/files/5123327/Wieringa_2014_-_Design_science_methodology_-_front_matter.pdf.

А. Н. Николаенко, О. В. Полоневич

ИСПОЛЬЗОВАНИЕ ФРЕЙМВОРКА MLOPS ДЛЯ УВЕЛИЧЕНИЯ УРОВНЯ ЖИЗНЕННОГО ЦИКЛА МАШИННОГО ОБУЧЕНИЯ В IoT

В последние годы особую популярность приобрели устройства IoT, что привело к большим объемам данных и спроса на связь с низкой задержкой, вызывая спрос на сети 5G. Этот сдвиг в инфраструктуре позволяет принимать решения в режиме реального времени с использованием искусственного интеллекта для приложений IoT. Искусственный интеллект вещей (AIoT) — это сочетание технологий искусственного интеллекта (AI) с инфраструктурой Интернета вещей (IoT) для достижения более эффективных операций IoT и принятия решений. Появляются краевые вычисления, чтобы активировать системы AIoT. Краевые вычисления дают возможность генерировать статистику и принимать решения в источнике данных, уменьшая объем данных, которые направляются в облака и центральное хранилище. Экосистема для облегчения краевых вычислений для приложений AIoT стала важной для принятия решений в режиме реального времени.

Ключевые слова: облачная платформа; искусственный интеллект; Интернет вещей; краевые вычисления; большие данные; машинное обучение.

O. M. Nikolaenko, O. V. Polonevich

USING MLOPS FRAMEWORK TO INCREASE MACHINE LEARNING LIFE CYCLE IN IoT

Recent years have witnessed a boom in IoT devices, leading to high data volumes and low latency demand, which has led to demand for the 5G networks. This infrastructure shift allows for real-time decision-making using artificial intelligence for IoT applications. Artificial intelligence (AIoT) is a combination of artificial intelligence (AI) technology with the Internet of Things infrastructure (IoT) to achieve more efficient IoT operations and decision making. Edge computing is emerging to enable AIoT applications. Edge computing enables generating insights and making decisions at the data source, reducing the amount of data sent to the cloud and central repository. An ecosystem to facilitate edge computing for AIoT applications has become essential to make real-time decisions.

Edge computing is the process of performing computational tasks that are physically close to the target devices, rather than in the cloud or on the device itself. This allows you to gain knowledge, ideas, and decisions at the data source. The purpose of edge computing should bring the calculations closer to the data source and unload the centralized calculations to the decentralized ones. Edge computing allows the application of various machine learning algorithms to create new experiences and new opportunities in many industries, from a connected house to security, surveillance, and automotive.

As for MLOps, IoT Edge is another deployment platform. However, if we use models on the IoT Edge, we need to consider some additional considerations. IoT Edge-targeted MLOps models must run offline mode. IoT models are more susceptible to data drift due to high data rates. IoT machine learning models need to be deployed on different target platforms, and we need to use the capabilities of these platforms.

Keywords: cloud platform; artificial intelligence; Internet of Things; edge computing; big data; machine learning.