IoT 온디바이스 AI 실현을 위한 AI 모델 레포지토리

이석준* · 최충재 · 성낙명

한국전자기술연구원

AI Model Repository for Realizing IoT On-device AI

Seokjun Lee* · Chungjae Choe · Nakmyung Sung

Korea Electronic Technology Institute

E-mail : sjlee88@keti.re.kr / cjchoe1@keti.re.kr / nmsung@keti.re.kr

요 약

IoT 디바이스에서 on-device AI를 수행할 때, 타겟 서비스나 디바이스의 환경에 따라 필요한 AI 모델이 달라질 수 있다. 또한, 기존 AI 모델도 federated learning과 같이 추가적인 데이터를 이용해 트레이닝을 하거나 보다 향상된 새로운 기법을 사용하는 등 업데이트가 일어날 수 있다. 이에 따라 IoT 디바이스에서 양질의 AI 서비스를 수행하기 위해서는 상황에 따라 필요한 AI 모델을 선택적으로 사용하거나 최적화된 최신 버전의 AI 모델로 업데이트 할 수 있어야 한다. 본 논문에서는 이를 지원하기 위한 AI 모델 레포지토리를 제안한다. 레포지토리는 AI 모델의 등록, 검색, 관리 및 배포를 지원하며 실사용을 위한 웹 포털을 포함한다. 제안하는 시스템의 실효성 확인을 위해 Node.js와 Vue.js로 구현하여 동작을 확인하였다.

ABSTRACT

When IoT device performs on-device AI, the device is required to use various AI models selectively according to target service and surrounding environment. Also, AI model can be updated by additional training such as federated learning or adapting the improved technique. Hence, for successful on-device AI, IoT device should acquire various AI models selectively or update previous AI model to new one. In this paper, we propose AI model repository to tackle this issue. The repository supports AI model registration, searching, management, and deployment along with dashboard for practical usage. We implemented it using Node.js and Vue.js to verify it works well.

키워드

IoT device, on-device AI, AI model repository

Ⅰ. 서 론

AI 기술이 발전하고, 디바이스의 컴퓨팅 능력이 향상됨에 따라 IoT 디바이스에서도 AI 기반의 서비스 제공이 가능하게 되었다. 이러한 on-device AI를 수행할 때, 타겟 서비스나 디바이스의 환경에 따라 필요한 AI 모델이 달라질 수 있다. 예를 들어, 지능형 CCTV의 경우 방범 서비스를 위해서는 사람을 탐지해야 하지만 불법주정차 단속을 위해서는 차량과 번호판의 탐지가 필요하다. 이 경우 서로 다른 AI 모델을 사용해야 한다. 또한, 기존 AI 모델도 federated learning [1]과 같이 추가적인 데이터를 이용해 트레이닝을 하거나 보다 향상된 새로운 기법을 사용하는 등 업데이트가 일어날 수 있다. 이에 따라 IoT 디바이스에서 양질의 AI 서비스를 수행하기 위해서는 상황에 따라 필요한 AI 모델을 선택적으로 사용하거나 최적화된 최신 버전의 AI 모델로 업데이트 할 수 있어야 한다. 본 논문에서는 이를 지원하기 위한 AI 모델 레포지토리를 제안한다. 레포지토리는 서버 레벨에서 동작하면서 IoT 디바이스가 편리하

* corresponding author

게 원하는 AI 모델을 얻을 수 있도록 한다. 구체적으로 AI 모델의 등록, 검색, 관리 및 배포를 지원하며 실사용을 위한 웹 포털을 포함한다. 우리는 제안하는 시스템의 실효성 확인을 위해 Node.js와 Vue.js로 구현하여 동작을 확인하였다.

II. IoT 디바이스용 AI 모델 레포지토리

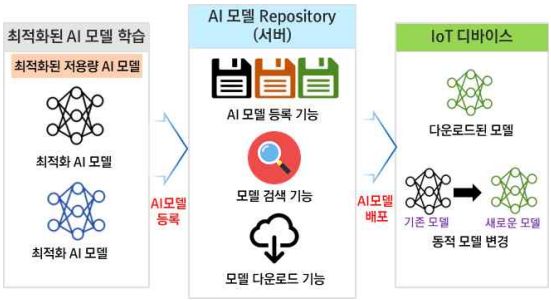

그림 1. AI 모델 레포지토리 개요도

AI 모델 레포지토리의 구조는 그림 1과 같다. 개발자 혹은 사용자가 최적화된 AI 모델을 학습하여 다양한 모델을 생성한 다음, AI 모델 레포지토리에 등록한다. 이 때 원활한 검색을 위해 모델의 특징을 등록하는데, 제안하는 시스템에서는 이를 컨텍스트라고 정의한다. 등록된 모델은 컨텍스트를 기반으로 관리되며, 검색 조건도 컨텍스트를 활용한다. IoT 디바이스는 필요 시 AI 모델 레포지토리에서 AI 모델을 검색한 다음, 배포 기능을 이용해 모델을 다운로드받는다. 이후 선택적으로 여러 AI 모델을 활용하여 on-device AI를 수행한다.

표 1. AI 모델 레포지토리 요구사항

구분	기능	설명
고속 지능분석 모델 관리	모델 조회	사용자 UI/UX 를 통해 등록된 고속 지능분석 모델 검색 기능을 제공
		조회된 모델에 대한 키워드 검색 기능을 제공
		모델에 대한 상세 정보 확인
	모델 등록	학습 기반 고속 지능분석 모델 등록을 위한 사용자 UI/UX 제공
		등록 모델 관리를 위한 DB 저장 및 조회
	모델 배포	웹 포털 사이트를 통해 제공하는 RESTFul API 를 통해 등록된 모든 모델을 다운로드 가능하도록 지원한다.
웹 서비스	UI/UX	사용자 편의성을 고려한 UI/UX 제공
		효율적인 정보 전달과 카테고리 이동을 위한 서브 UI 개발
	Context 연관 검색	Context 키워드와 연관된 최적의 모델 검색 및 표시

표 1은 AI 모델 레포지토리의 구체적인 요구사항을 나타낸다. 우리는 추상적인 AI 모델 레포지토리의 개념을 구체화하기 위하여 모델 관리와 웹 서비스의 2가지 카테고리로 요구사항을 정의하였다. 모델 관리 측면에서는 모델 조회, 모델 등록, 모델 배포의 3가지 요구사항이 존재하며, 모델 조회에는 검색 기능이 포함된다. 웹 서비스는 모델의 등록 및 관리를 위해 실질적으로 필요한 웹 포털 관련 요구사항을 정의했으며, 크게 UI/UX 기능과 검색 기능이 속한다.

III. 시스템 구현

그림 2. AI 모델 레포지토리 웹 포털 화면

제안하는 AI 모델 레포지토리의 동작을 확인하기 위해 우리는 Node.js [2]와 Vue.js [3]를 이용하여 표 1의 요구사항을 만족시키는 시스템을 구현하였다. 그림 2는 웹 포털의 화면이다. 관리 포털 웹 사이트에 접속을 하면 서비스별로 등록된 지능분석 관리 모델의 목록이 표시되고 상단 메뉴에는 모델 및 컨텐스트 검색과 사용자 관리 메뉴가 표시되며 왼쪽 사이드 메뉴를 통해 모델의 등록 및 서비스 타입별 선택 메뉴가 표시된다. 관리 포털 웹 사이트는 UI 프레임워크를 Vue.JS 기반으로 구현하여 화면 표시 기능을 강화하고 직관적인 사용자 UI를 제공한다.

그림 3은 모델 등록 화면을 나타낸다. 학습기반 고속 지능분석 모델 생성을 통해 만들어진 모델을 웹 사이트에 등록하기 위한 기능을 제공하며 서비스 타입, 컨텍스트 정보 및 모델 파일과 상세 설명을 위한 Readme 파일의 업로드가 가능하다. 업로드된 정보들은 웹 사이트내의 모델 관리 로직을 통해 분류, 저장되며 추후 모델 조회, 배포시 활용된다

그림 4는 모델 검색 화면이다. Conext 연관 검색을 통해 입력된 키워드와 연관된 최적의 모델을 검색하여 사용자에게 표시하는 기능을 제공한다.

그림 3. AI 모델 등록 화면

그림 4. AI 모델 검색 화면

그림 5는 모델 배포 화면이다. 사용자에 의해 등록된 고속지능 모델은 2가지 방법을 통해 배포가 가능하다. 하나는 다운로드 버튼을 통해 직접 파일을 저장하는 방법과 RESTFul API를 통해 모델의 파일을 내려받는 방법을 제공한다.

그림 5. AI 모델 배포 화면

Ⅴ. 결 론

본 논문에서는 IoT 디바이스가 on-device AI를 성공적으로 수행하기 위해 동적 AI 모델 사용을 지원할 수 있는 AI 모델 레포지토리를 제안하였다. AI 모델 레포지토리의 구성과 요구사항을 정의하였고, 이를 실제 서버상에 구현하여 그 동작을 확인하였다. AI 모델 레포지토리는 다양한 AI 모델의 등록과 관리를 용이하게 하고, IoT 디바이스가 필요로 하는 AI 모델을 쉽게 검색하고 얻을 수 있도록 하여 on-device AI의 실현에 실질적인 도움을 제공할 것으로 기대된다.

Acknowledgement

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (NO. 2022-0-00959, 드론 및 로봇 분야에 적용 가능한 5G 환경 온디바이스 IoT 고속 지능 HW 및 SW 엔진 기술 개발).

References

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, B.A.Y. Arcas, “Communication-efficient learning of deep networks from decentralized data,” in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, pp. 1273-1282, 2017.

[2] Node.js [Internet]. Available : https://nodejs.org/.

[3] Vue.js - The Progressive JavaScript Framework [Internet]. Available : https://vuejs.org/.