

chapter 3-1

컨테이너 기반 AI 서비스 구성 기술



김승우 || 전자부품연구원 책임연구원

I. 결과물 개요

개발목표시기	2020. 12.	기술성숙도(TRL)	개발 전	개발 후
			TRL 3	TRL 5
결과물 형태	Server S/W	검증방법	자체검증	
Keywords	클라우드 플랫폼, 엣지 플랫폼, AI 서비스			
외부기술요소	Tensorflow, Docker, Kubernetes	관리성	특허(출원 중), SW	

II. 기술의 개념 및 내용

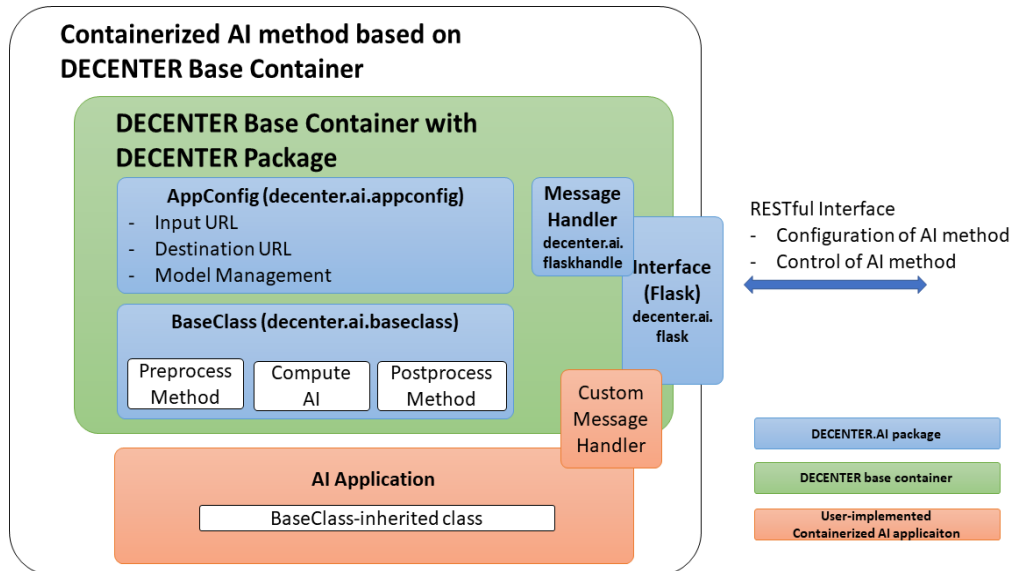
1. 기술의 개념

- AI 연산은 GPU 등 고성능 컴퓨팅 자원을 요구하고 있어, 이러한 AI 알고리즘 기반의 서비스를 구성하기 위해서는 생성된 알고리즘을 서비스 플랫폼화 하는 작업이 필수적

* 본 내용은 김승우 책임연구원(☎ 02-6388-6653)에게 문의하시기 바랍니다.

** 본 내용은 필자의 주관적인 의견이며 IITP의 공식적인 입장이 아님을 밝힙니다.

***정보통신기획평가원은 현재 개발 진행 및 완료 예정인 ICT R&D 성과 결과물을 과제 종료 이전에 공개하는 "ICT R&D 사업화를 위한 기술예고"를 2014년부터 실시하고 있는 바, 본 칼럼에서는 이를 통해 공개한 결과물의 기술이전, 사업화 등 기술 활용도 제고를 위해 매주 1~2건의 관련 기술을 소개함



[그림 1] 기술개념도

으로 요구됨

- 본 기술은 학습을 통해 생성된 AI 알고리즘에 대해 Docker 기술을 적용하여 효율적으로 컨테이너를 구성할 수 있게 함
- 또한, 해당 컨테이너를 외부에서 제어할 수 있는 정형적 인터페이스를 제공함과 함께 인터페이스에 대한 손쉬운 확장을 지원함
- 상기의 기술을 통해 AI 서비스 개발자는 AI 기술에 대한 심도 있는 지식 없이도 손쉽게 AI 서비스를 컨테이너 형태로 활용 가능하며, 다양한 다른 웹 서비스에 연계하여 자신이 개발하고자 하는 서비스로 구성이 가능함

2. 기술의 상세내용 및 사업화 제약사항

➤ 기술의 상세내용

- AI 알고리즘의 컨테이너화를 지원하기 위한 파이썬 패키지 모듈
- Tensorflow 기반의 컨테이너화 구성을 지원하는 Docker Base Image
- 외부에서 컨테이너 내부의 AI 알고리즘을 호출하기 위한 정형화된 인터페이스

- CPU 및 GPU(NVidia 계열)에서 운용 가능
- VGG16, Yolo v3에 대한 Reference Service 구성 제공
- 기술이전 범위
 - 소스 코드
 - 기술 교육
 - 서비스 개발 가이드 라인
- 사업화 제약사항
 - 외부 라이브러리 의존성(Tensorflow, Docker, Flask 등)

III. 국내외 기술 동향 및 경쟁력

1. 국내 기술 동향

- 국내에서는 알고리즘의 고도화를 통한 성능 개선에 연구가 집중되어 있으며, 생성된 알고리즘의 배포에 대한 연구는 아직까지 초기 단계임
 - 네이버 등 주요 AI 연구 기업은 음성 처리 및 영상 처리에 대한 알고리즘 성능 개선 및 고도화에 대한 연구를 지속적으로 수행 중
 - KT, SKT, LG U Plus 등 주요 통신 사업자는 음성 기반 명령을 처리할 수 있고, 해당 명령을 자사의 서비스에 연계할 수 있는 AI 스피커를 출시하였으나 AI 서비스에 대한 패키징 및 배포 기술에 대한 내용은 아직 확인되지 않음

2. 해외 기술 동향

- 해외서는 주요 ICT 기업을 중심으로 생성된 AI 알고리즘의 배포에 대한 기술 개발이 진행 중임
 - Google은 자사의 Tensorflow 기반 모델을 웹 인터페이스로 제공할 수 있는 Tensorflow Serving 기술을 개발 및 서비스 중에 있음

- Microsoft는 자사의 Azure Edge 및 IoT 플랫폼을 통해 학습된 모델을 엣지 혹은 클라우드에 배포할 수 있는 서비스를 운영 중

3. 표준화 동향

- AI 모델의 배포 최적화 기술에 대한 표준이 국제 표준으로 진행 중
 - ISO/IEC SC29 WG11 산하 NNR(Neural Network Representation) 그룹은 생성된 모델의 효율적 배포를 위한 모델의 표현, 압축 및 메타 데이터 기술에 대한 표준화를 진행 중에 있음

4. 기술적 경쟁력

경쟁기술	본 기술의 우수성 및 차별성
Tensorflow Serving	분석 결과의 단순 출력 외에 영상 기반 overlay 처리 기능 지원
Azure	AI 모델간 연동 기술 지원

IV. 국내외 시장 동향 및 전망

1. 국내 시장 동향 및 전망

- 국내 인공지능 산업은 시장 형성 단계로 일부 대기업 및 IT기업에서 인공지능 연구에 투자하고 있음

2. 해외 시장 동향 및 전망

- 시장조사기관인 IDC는 인공지능 시장 규모를 2015년 약 1,270억 달러에서 2017년 약 1,650억 달러로 연평균 14.0% 성장할 것으로 전망
- Tractica는 기업용 인공지능시스템 시장이 2015년 2억 달러에서 2024년 111억 달러로 연평균 56.1% 성장할 것으로 전망하였으며, Marketsandmarkets은 2014년 41억

9,700만 달러에 이르는 인공지능 시장이 2015년부터 2020년까지 53.7%의 연평균성장률을 보일 것으로 전망함

3. 제품화 및 활용 분야

활용 분야(제품/서비스)	제품 및 활용 분야 세부내용
엣지 카메라	AI 모델의 원격 관리 및 운영을 지원하는 엣지 카메라
홈 엣지 셋탑 박스	홈 내 AI 서비스의 운영을 지원하는 엣지 셋탑 박스

V. 기대효과

1. 기술도입으로 인한 경제적 효과

- 클라우드 기반 AI 서비스 구성을 위한 개발 및 관리 비용 절감
- AI 알고리즘 자체에 대한 컨테이너화를 통해 기존 클라우드 기반 서비스와의 효율적 연계를 지원

2. 기술사업화로 인한 파급효과

- AI 기반 서비스 활성화가 기대
- AI 알고리즘을 활용하는 서비스 구성에 있어, AI 알고리즘/플랫폼 운영 등 사전 지식에 대한 요구사항을 축소하고 클라우드 서비스 형태로 구성함으로써 보다 손쉽게 AI 알고리즘 활용이 가능