Azure 클라우드 플랫폼의 가상서버 호스팅을 이용한 데이터 수집환경 및 분석에 관한 연구

이재규[O], 조인표[*], 이상엽[*]
[O]전자부품연구원 지능융합SW연구센터,
[*]전자부품연구원 지능융합SW연구센터
e-mail: {jaekyu, inpyo.cho, syublee}@keti.re.kr[O*]

A study on data collection environment and analysis using virtual server hosting of Azure cloud platform

Jaekyu Lee[O], Inpyo Cho[*], Sangyub Lee[*]
[O]Intelligent Convergence SW Research Center, Korea Electronics Technology Institute,
[*]Intelligent Convergence SW Research Center, Korea Electronics Technology Institute

● 요 약 ●

본 논문에서는 Azure 클라우드 플랫폼의 가상서버 호스팅을 이용해 데이터 수집 환경을 구축하고, Azure에서 제공하는 자동화된 기계학습(Automated Machine Learning, AutoML)을 기반으로 데이터 분석 방법에 관한 연구를 수행했다. 가상 서버 호스팅 환경에 LAMP(Linux, Apache, MySQL, PHP)를 설치하여 데이터 수집환경을 구축했으며, 수집된 데이터를 Azure AutoML에 적용하여 자동화된 기계학습을 수행했다. Azure AutoML은 소모적이고 반복적인 기계학습 모델 개발을 자동화하는 프로세스로써 기계학습 솔루션 구현하는데 시간과 자원(Resource)를 절약할 수 있다. 특히, AutoML은 수집된 데이터를 분류와 회귀 및 예측하는데 있어서 학습점수(Training Score)를 기반으로 보유한 데이터에 가장 적합한 기계학습 모델의 순위를 제공한다. 이는 데이터 분석에 필요한 기계학습 모델을 개발하는데 있어서 개발 초기 단계부터 코드를 설계하지 않아도 되며, 전체 기계학습 시스템을 개발 및 구현하기 전에 모델의 구성과 시스템을 설계해볼 수 있기 때문에 매우 효율적으로 활용될 수 있다. 본 논문에서는 NPU(Neural Processing Unit) 학습에 필요한 데이터 수집 환경에 관한 연구를 수행했으며, Azure AutoML을 기반으로 데이터 분류와 회귀 등 가장 효율적인 알고리즘 선정에 관한 연구를 수행했다.

키워드: 데이터 수집(Data Collection), MS 애저(MS Azure), LAMP(Linux, Apache, MySQL, PHP), 자동화된 기계학습(Automated Machine Learning), 데이터 분석(Data Analysis)

I. Introduction

현재, 인공지능과 빅데이터는 매우 활발히 연구되고 있는 분야중 하나이다. 빅데이터와 인공지능 기술은 기존의 다양한 문제를 해결하는데 사용되고 있으며, 제조기업 등 산업체의 생산성을 향상시키는데 적극적으로 활용되어지고 있다.[1] 또한, 우리는 임베디드 인공지능 시스템 개발에 관한 연구를 수행하고 있다. 임베디드 인공지능 시스템은 저전력/저사양 플랫폼에서 운영 가능한 인공지능 시스템을 개발하는 것이며, 신경망 처리장치(Neural Processing Unit, NPU)가 사용된다. 임베디드 플랫폼 기반 인공지능 시스템을 구현하기 위해서는 고수준의 최적화 작업이 필요하며, 수많은 데이터 기반의 시뮬레이션 및 테스트가 요구된다. 본 논문에서는 NPU 학습을 위한 데이터 수집 환경과 Azure AutoML기반 범용 기계학습 최적화 모델 분석에 관한 연구를 수행했다.

II. Background

우리는 임베디드 시스템을 기반으로 NPU를 이용해 소형의 저전력 PCB 검사기를 개발에 관한 연구를 수행하고 있다. NPU는 Neural Processing Unit으로 인공 신경망 처리장치를 의미한다. 최근 임베디드 시스템 기반 NPU나 뉴로모픽 칩을 이용해 인공지능 시스템 개발하는 연구가 활발하게 진행되고 있다.[2] 임베디드 시스템 및 NPU를

이용한 인공지능 시스템을 설계하기 위해서는 양질의 데이터 확보가 매우 중요하다. 본 논문에서는 NPU 학습에 필요한 데이터 수집 환경 구축에 대한 연구를 수행했으며, Azure AutoML을 이용해 최적 알고리즘을 도출하는 연구를 수행했다. AutoML을 이용하면 실제 임베디드 시스템을 위한 기계학습 알고리즘 설계 및 포팅(Porting) 이전에 데이터에 적합한 알고리즘을 도출 할 수 있기 때문에 효과적으로 시스템을 구성할 수 있다.

III. Data collection environment and analysis

데이터 수집환경은 Azure에서 제공하는 가상서버를 기반으로 LAMP를 이용해 서버 구축 및 데이터 수집환경을 구성했다. 즉, Linux(Ubuntu 18.04) 운영체제에 Apache2 서버를 설치했으며, 데이터베이스는 MySQL(5.7)을 사용했다. 웹페이지 구성은 PHP를 기반으로 설계 및 구현했다. 아래 그림은 각각 Apache2 서버와 MySQL 그리고 PHP 설치가 완료된 화면을 캡처한 것이다. Azure는 마이크로소프트의 클라우드 컴퓨팅 플랫폼으로 가상서버와 DB, 인공지능 API 등 다양한 서비스를 제공한다.[3]

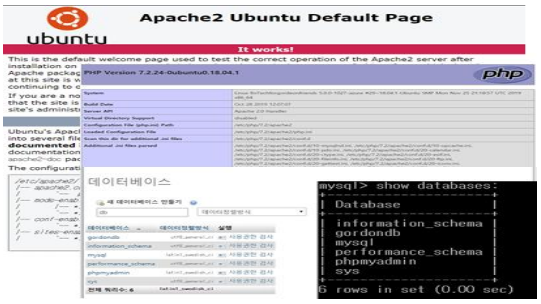

Fig. 1. Ubuntu Linux기반 AMP 환경 구축

데이터 분석은 Azure의 자동화된 기계학습(AutoML)을 기반으로 수집된 데이터에 가장 적합한 분류 모델을 선정했다. 본 논문에서 수집한 데이터에 적합한 모델로는 그림 2에서와 같이 “Voting Ensemble”이 학습 및 분류에서 최적의 모델로 선정되었다. 이처럼 Azure AutoML을 이용해 데이터에 가장 적합한 기계학습 모델을 선정한 후 임베디드 인공지능 시스템에 모델을 적용하는 순서로 개발을 수행하면 임베디드 시스템 기반 인공지능 응용을 개발하는데 있어서 보다 효율적으로 시스템을 개발 및 구현 할 수 있다.

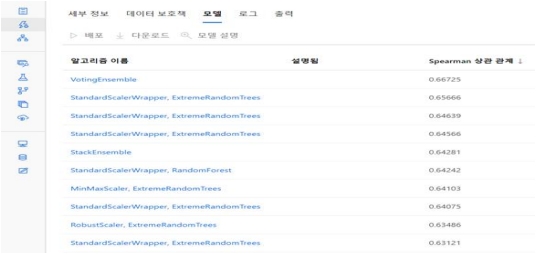

Fig. 2. Azure AutoML기반 최적 모델 선정

IV. Conclusions

본 논문에서는 Microsoft Azure 클라우드 플랫폼의 가상서버를 기반으로 NPU 학습을 위한 데이터 수집환경 구축과 AutoML을 이용한 효율적인 분석방법에 관한 연구를 수행했다. 특히, LAMP를 이용한 데이터 수집 환경 구축에 관한 연구와 Azure AutoML기반 데이터에 최적화된 기계학습 모델 도출 방법에 관한 연구를 수행하고 절차를 기술했다. 본 논문과 같이 Azure AutoML을 이용해 최적 알고리즘을 도출한 후 기계학습 모델을 설계하고 임베디드 시스템 기반 인공지능 시스템을 구현하면, 보다 빠르고 효율적인 기계학습 솔루션을 설계 및 구현 할 수 있을 것이다.

ACKNOWLEDGEMENT

본 연구는 산업통상자원부 및 한국 산업 기술 평가 관리원의 산업 핵심 기술개발 사업의 일환으로 수행하였습니다. (20006566: 뉴로모픽 기반 임베디드 인공지능 모듈 및 시스템 기술 개발)

REFERENCES

[1] Li, Liangzhi, Kaoru Ota, and Mianxiong Dong. "Deep learning for smart industry: Efficient manufacture inspection system with fog computing," IEEE Transactions on Industrial Informatics, Vol. 14, No. 10, pp. 4665-4673, 2018.

[2] Yang, Fan, and Michel Paindavoine. "Implementation of an RBF neural network on embedded systems: real-time face tracking and identity verification," IEEE Transactions on Neural Networks Vol. 14, No. 5, pp. 1162-1175, 2003.

[3] Chappell, David. "Introducing windows azure," Microsoft, Dec 2009.