DBpia

논문 21-46-09-24 The Journal of Korean Institute of Communications and Information Sciences '21-09 Vol.46 No.09
https://doi.org/10.7840/kics.2021.46.9.1539

소형가전제품 HS Code와 기기부호에 대한 인공지능 지도학습 데이터셋 연구

오 영 주◆, 박 대 우°

Dataset for Artificial Intelligence Supervised Learning on HS Code and Device Code of Small Home Appliance

Young-joo Oh◆, Dea-woo Park°

요 약

세계 무역 장벽의 일종으로 '법적의무인증'이 요구되고 있다. '법적의무인증'은 제품을 인식하고 특징을 구분하는 것이 필요하다. 그러나 이러한 작업은 수작업으로 HS Code를 구분하고, 기기부호에 대한 인증 작업을 수행하고 있다. 4차 산업혁명 기술을 활용하여 무역장벽을 개선해야 할 필요성이 있다. 본 논문에서는 KC 시험소의 성적서 이미지와 품질표시항목을 정제하여, 라벨링과 어노테이션으로 인공지능 지도학습용 데이터셋을 만든다. 이 데이터셋을 학습한 인공지능은 제품을 인식하고, 특징도 구분하는 것이 가능하다. 본 논문은 세계 무역 장벽의 해결을 위한 기초연구로써 이바지할 것이다.

키워드 : 인공지능, HS 코드, 빅 데이터, 욜로v4, 지도학습, 레이블
Key Words : Artificial Intelligence, HS Code, Big Data, Yolov4, Supervised Learning, Label

ABSTRACT

As a kind of barrier to world trade, 'Legal Compulsory Certification' is required. 'Legal Compulsory Certification' requires recognizing the product and distinguishing its characteristics. However, in this work, the HS Code is manually classified and the device code is authenticated. There is a need to improve trade barriers by utilizing the 4th industrial revolution technology. In this paper, we make a dataset for artificial intelligence supervised learning through labeling and annotation by refining the image and quality indication items of the KC laboratory. AI that has learned this dataset can recognize products and distinguish features. This paper will contribute as a basic research for resolving world trade barriers.

Ⅰ. 서 론

세계는 COVID-19 PANDEMIC 사태와 미국과 중국의 무역 갈등으로 불안한 세계 환경에 처해 있다. 대한민국은 미국과 중국과의 무역거래뿐만 아니라, 각국의 보호무역주의에 대한 문제로 무역에 대한 많은 갈등 요인을 갖고 있다[1].

세계 주요 국가들은 자유 무역 협정과 보호주의를 사용하여, 비관세 무역장벽과 자국 산업에 혜택을 주는 관세를 부과하고 있다. 관세의 부과뿐만 아니라, 여러 가지의 인증과 서류상의 규제를 통해 비관세 장벽도 실시하고 있다[2]. 대한민국은 2021년부터 세계

◆ First Author : Hoseo Graduate School of Venture, young10041@chol.com, 정회원
° Corresponding Author : Hoseo Graduate School of Venture, prof_pdw@naver.com, 정회원
논문번호 : KICS202108-186-0-SE, Received July 31, 2021; Revised August 11, 2021; Accepted August 11, 2021

www.dbpia.co.kr

10대 경제 교역국으로, 발달한 ICT(Information & Communications Technology)기술을 바탕으로 가전제품을 세계에 수출하고, 수입하는 국가이다. 특히 소형가전제품은 생활에 밀착되어 국민의 생활에 많이 사용된다. 대한민국에는 총 79개의 법적 의무 인증 시스템이 있어 비관세 보호 장치로 국내의 산업을 보고하고 있다[3]. 세계 무역과 관련하여 제품의 분류는 국제통일 상품분류체계에 의해, 대외 무역거래 제품을 총괄적으로 분류한 HS Code (Harmonized Commodity Description and Coding System Code)를 부여한다. 소형가전제품들은 "방송 및 통신 장비 적합성 평가"시스템의 판매전 인증을 받아야만 합법적인 판매가 가능하다. 이러한 문제를 해결하기 위해서는 제품의 품목을 정확히 분류하는 것은 매우 중요하다.

하지만 세계 무역에서 비관세 장벽에서 사용되는 소형가전제품의 무역 통관과 판매 관련 인증 문서는 종목마다 달라서 전문가들이 일일이 수작업에 의존하여 시간과 비용의 과다 소요되어, 정부의 비관세 무역장벽의 방법으로도 사용된다.

본 논문에서는 소형가전제품에 대한 수출과 수입 및 관세 비관세 정책에 부여되는 무역 인증서류 수작업을 개선하도록 연구한다. 인공지능(Artificial Intelligence)이 지도 학습(Supervised Learning)할 수 있는 소형가전제품 HS Code를 데이터셋(Dataset)으로 분류한다. 소형가전제품 품목분류의 HS Code 찾는 인공지능 지도학습 알고리즘(Algorithm)을 적용하여 구현하는 라벨링(Labeling)과 어노테이션(Annotation)에 대하여 연구한다. 소형가전제품의 넓은 제품의 범위를 빅 데이터(Big Data)로 분석한다. 국내에서 판매되는 KC 전자파의 적합성 사전 평가가 인공지능 기계학습(Machine Learning)을 통해 유효한지 판단할 수 있는 인공지능 모델링을 설계한다[4].

Ⅱ. 관련연구

2.1 소형 가전제품의 KC 적합등록

본 논문에서는 소형가전제품에 대한 인공지능 지도학습 데이터셋의 원천 데이터로 KC 적합등록 시험소의 시험 성적서 자료에 포함된 이미지와 품질표시를 사용한다. KC 적합등록 대상의 소형가전제품의 약 180개 제품군이 있으며, 기기부호(Common Device Code)를 적용하여 분류하고 있다. 여기서 기기부호는 "전파법" 제58조의2부터 제58조의4 및 같은 법 시행령 제77조의2부터 제77조의8에 따른 "방송통신기자재 등의 적합성평가에 관한 고시" 고시의 "별표 1의 적합성평가 대상기자재의 11항 전자파장애를 주거나 전자파로부터 영향을 받는 기기"에 따르고 있다. 표 1. KC 적합등록 소형가전제품의 기기부호(Device Code) 목록은 이 논문에 적용할 소형가전제품에 대한 기기부호만을 정리하였다.

표 1. KC 적합등록의 소형 가전제품의 기기부호 목록
Table 1. Common Device Code List on KC EMC

Common Device Code Label List				
CLN11	IRN11	DSM11	MWO11	WSM11
HDR11	WRM11	KTC11	CUK11	MAT11
STW11	WRM21	POT11	COL11	RAM11
SEM11	CHG11	HRY11	STV11	MSG11
SCD11	SUN11	BTH11	AIR11	FAN11
BTH21	MST11	DIS11	PET11	SKT11
SKT11	CHR11	CHR11	BOL11	BRH11
VRM11	DUS11	MIL11	CTN11	GAM11
CTN21	VRM21	CLE11	COW11	ETL11
WDM11	AQU11	BBL11	BLD11	MST11
ROL11	KCN11	POL11	ELD11	WAT21
WTP11	PJT11	BIN11	BIN11	DOR11

소형가전제품 HS Code에 대한 인공지능 지도학습 데이터셋의 원천데이터로는 1,000종 이상의 소형가전제품에 대한 20만장 이상의 이미지가 필요하다. KC 적합등록 성적서의 제품 사진과 품질표시, 품목분류의 이미지와 제품정보, 그리고 직접 촬영한 소형가전제품의 사진으로 원천 데이터 확보한다. KC 적합등록 대상의 소형가전제품의 이미지 기반으로 물체 검출기의 성능 개선하기 위해 인공지능 학습용 데이터셋 구축하고 한다.

2.2 빅 데이터 분석

인공지능 데이터셋의 관심도와 시장의 증가에 따라 기업들은 인수·합병 전략으로 인공지능 산업에 대비하고 있으며, 전 세계 인공지능 학습용 데이터셋 시장도 성장함에 따라 인공지능의 성능 향상을 위해 고품질 인공지용 학습용 데이터셋에 대한 관심도가 증가하고 있다. 인공지능 학습용 데이터셋 시장의 규모는 '18년 5억 달러에서 '23년 12억 달러로 전망 성장할 것으로 전망하고 있다. 산업계에서는 시장에서의 입지와 데이터셋 품질 향상의 수단으로 인수·합병, 협업을 통해 시장을 확대하고 있고, 호주 글로벌 인공지능 기업 에이펜은 미국의 인공지능 데이터 기업인 피규어 에잇트를 인수하기 위해 한화 약 3,500억 원에 투자유치를 계획하고 있다[5]. 산업별 인공지능 학습용 데이터셋 시장의 규모는 IT, 자동차, 헬스케어(Health Care), 전자상거래, 금융 분야 순이다. 고품질 데이터

www.dbpia.co.kr

셋은 IT 산업의 컴퓨터 비전, 데이터 분석 등의 서비스 품질을 향상시키고, 이 요인으로 IT분야가 학습용 데이터 시장에서 높은 점유율을 차지하고 있다[6].

소형가전제품의 객체 인식 성능 향상을 위해 소형가전제품객체의 이미지 학습 데이터 구축하기 위하여 데이터 수집, 정제, 가공, 인공지능 모델링을 통한 프로세스로 진행하여 인공지능 데이터셋 구축에 적합한 빅 데이터를 분석은 필수이다[7].

2.3 인공지능 지도학습

컴퓨터 비전 기술은 개발과 연구 및 응용과 적용함에 있어 특정한 산업분야에서 한정적으로 연구되고 있다. 컴퓨터 비전 기술이란, 인공지능의 분야 중에서 사람의 시각 기능을 카메라가 영상이미지를 촬영하고, 알고리즘으로 분석하고 판별함으로 인공지능이 대신 수행하는 것이다. 컴퓨터 비전 기술의 연구에서 객체의 영상이미지에 대한 인식 기술 개발에 주로 사용되는 YOLO(You Only Look Once) 알고리즘이 있다[8].

이미지 객체 감지 알고리즘인 YOLOv4는 영상이미지에 대한 인식을 활용하기 위한 기술 개발에서는 주요 기술로 다뤄지고 있고, YOLOv4 알고리즘은 최근에 개선하거나 연구 상황에 맞춰 제안되었다[9].

영상이미지에서의 객체 인식은 사람의 시각 정보의 분석돠 해석을 기계가 대신하는 연구 분야이며 영상감시, 얼굴 인식, IoT, 로봇 제어, 자율주행, 보안 등에 활용됨으로써 여러 분야에서 핵심기술로 사용되고 있으며, 사물의 객체 인식에 대한 분야를 연구하는 사람들은 Pascal, ImageNet, 그리고 MS COCO 등의 발전을 도모하고 있다.

이미지 인식 기술은 이미지에 나타난 객체가 무엇인지를 인식하는 기술로써, 이미지에 하나의 대상만 포함하고 있다고 보고 대상을 분류하는 이미지 분류, 대상의 위치에 사각형을 그려주어 위치를 표시하는 이미지 찾기, 이미지에 포함된 여러 대상을 찾고 위치를 사각형으로 표시하는 이미지 인식, 이미지에 포함된 대상을 정확히 선으로 분할하는 대상 분할 등이 있다[10].

인공지능의 객체 인식과 검출은 컴퓨터 비전 기술 기반의 객체 인식 및 검출 기술과 관련한 모델링과 알고리즘에 기반을 둔다. 인공지능 기술의 근간이 되는 FCNN(Fully-Connected Neural Network)에서부터 1989년 발표된 합성곱 신경망(Convolutional neural network), 2012년 깊어지는 합성망 신경망(Deep Convolutional neural network) 모델인 AlexNet의 등장으로 많은 연구팀들이 인공지능 알고리즘을 활용하게 되었으며, 2014년에는 VGG Net과 GooLeNet가 소개되었으며 GooLeNet은 학습과정에의 과적합 문제를 해결하기 위해 네트워크 구조를 개선함으로써 향상된 인식률을 보였으며, 이후 소개된 ResNet는 인공지능 네트워크 구조가 깊어질수록 학습과정에서 발생하는 오류 문제를 해결하기 위해 제안된 Residual Learning 머신러닝 학습 방법이다[11]. 또한, 이미지넷(ImageNet)은 약 10억장에 이르는 이미지를 다운로드 했고, 아마존(Amazon)은 클라우드소싱(Crowdsourcing) 기술을 사용해 이미지를 정리하고 분류된 이미지를 라벨링 하였다. 이러한 인공지능의 기반 객체 인식기술과 검출 기술은 구글 포토 서비스, 불량품을 선별, 스마트팩토리(Smart Factory)의 업무 효율화와 공정관리, 제품의 안전 관리와 생산성 관리, 물류 상품의 관리나 손상 탐지를 위해 활용되고 있다.

Ⅲ. 소형가전제품의 HS Code 인공지능 데이터셋 분석과 설계

3.1 소형가전제품 HS Code 분석

소형가전제품 중 '전기기기 및 전동기기류'에 속하는 기기부호(Common Device Code)에 따라 분류된 제품군이 있다.

표 1. 의 소형가전제품 기기부호에 따라 분류된 소형 가전제품군에는 여러 개의 유사한 기능의 제품으로 나누어진다. 제품명으로 포함되는 모든 제품의 이미지들을 원천 데이터로 사용한다. 이러한 이미지를 라벨링 하기 위한 기준이 필요하다. 표 2 는 소형가전제품의 원천데이터 분류체계이다. 이 표는 기기부호, 제품명, HS Code로 구성되어 있다. 각 기기부호에는 여러 개의 제품과 각 제품에 대한 해당되는 HS Code를 정의하였다.

원천 데이터의 이미지를 품목분류에 따라 분석하여 HS Code를 결정할 때 사용된다. 이 논문에서 수집된 원천데이터는 여러 가지 방법으로 수집된 소형가전제품 이미지 데이터와 직접 촬영한 소형가전제품 이미지 데이터를 관세법에 따른 품목분류와 KC 적합성 평가에 따른 품질표시를 기준으로 HS Code를 산출 하고, 라벨링의 기준이 되는 데이터에 대한 기준을 만들었다.

표 2. 는 소형가전제품에 대한 전체를 보여주는 것은 아니고, 라벨링 기준으로 사용될 소형가전제품의 원천데이터 분류체계의 일부이다. 이 표에서 기준이 되는 기기부호는 전파법에 따른 분류 코드로 KC 적합성을 평가시 제품의 분류 체계이다. HS Code는 관세

www.dbpia.co.kr

표 2. 소형가전제품의 원천데이터 분류체계
Table 2. Classification of Raw Data for Small Home Appliances

	Device Code	Product Name	HS Code
1	CLN11	Vacuum Cleaner	8508.19-9000
		Water suction Cleaner	8508.11-9000
		Electric Floor Cleaner	8508.11-9000
		Electric Surface Washer	8508.19-9000
		Steam Sleaner	8508.11-9000
2	IRN11	Electric Drying Iron	8451.21-0000
		Steam Iron	8451.29-0000
		Wrinkle Smoothing	8451.30-1000
		Trouser Press	8451.21-0000
		Ironing Press	8451.30-1000
3	DSM11	Electric Dishwasher	8422.11-0000
		Electric Dish Dryer	8422.11-0000
4	MWO11	Electric Stove	8516.50-0000
		Electric Oven	8516.60-1000
		Electric Grill	8516.60-1000
		Electric Hob	8516.79-9000
		Electric Stove	8516.60-9000
		Electric Heater	8516.60-9000
		Electric Toaster	8516.72-0000
		Electric Frying Pan	8516.60-9000
		Electric Meat	8516.60-9000
		Waffle Machine	8516.60-9000
		Hot Plate	8516.60-9000

법에 따른 품목분류 체계이다.

그리고 제품명은 일반적인 소형가전제품의 대표적 제품명이다. 이러한 소형가전제품 원시데이터를 정제하고 표 2. 의 소형가전제품의 원천데이터 분류체계에 따라 모든 이미지를 라벨링 작업을 하였다.

3.2 소형가전제품 데이터 수집

소형가전제품의 원천데이터(Raw Data)는 3가지 방법 수집된다. 첫 번째 전파법에 의거한 KC 적합성 시험과정에서 생성되고 법적으로 보관하게 되어 있는 이미지 데이터와 품질표시 데이터, 두 번째 관세법에 따른 품목분류 과정에서 검토가 완료된 이미지와 판정기준 데이터, 마지막으로 확보 방안은 직접 구매한 제품을 촬영하여 이미지 데이터를 품목분류와 품질표시에 따라 분류한 데이터를 원천데이터이다. 이렇게 수집한 원천데이터를 정재하고 가공하여 라벨링과 어노데이션 작업을 하여 가공하여 소형가전제품의 데이터셋을 구현하기 위하여 수집한다.

그림 1. 은 소형가전제품 원시데이터 분류 절차이다. 여기서는 직접 구매하여 촬영한 데이터 수집 방법은 포함시키지 않았다. KC 시험소와 수출입 관세법인의 기업 활동에서 관리하는 엔터프라이즈 데이터 웨어하우스(Enterprise Data Warehouse, EDW)에서 소형가전제품의 원천데이터를 수집한다.

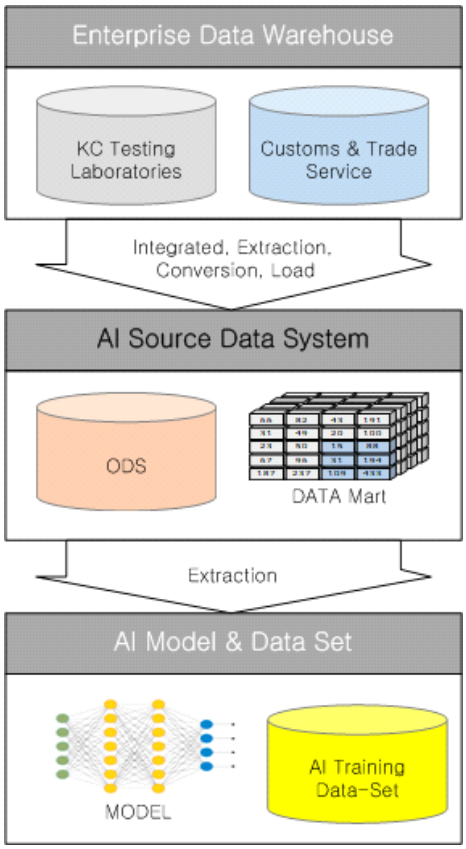

그림 1. 소형가전제품 원시데이터 분류 절차
Fig. 1. Small Home Appliance Raw Data Classification Procedure

KC 인증기관에 의뢰한 제품이 시장에서 판매에 적합한지에 관련된 시험을 하며, 각 제품을 촬영한 제품의 이미지와 품질표시의 이미지를 보관하고, 그 모든 정보를 성적서 등과 함께 관리한다. 관세법인의 품목분류 업무 중 발생한 각 기업의 정보시스템의 원천데이터를 수집하고 보관한다. 이 논문에 필요한 원천데이터는 시험소나 관세법인 등에서 운영하는 엔터프라이즈 데이터 웨어하우스에서 수집한다.

그림 1. 은 KC 인증 시험소나 관세법인의 엔터프라이즈 데이터 웨어하우스에서 수집된 원천데이터를 통합(Integrated), 추출(Extraction), 변환(Conversion)과 적재(Load) 과정을 거쳐서 정제된 데이터를 운영데이터 스토어(Operational Data Stor, ODS)나 데이

www.dbpia.co.kr

터 마트(Data Mart)에 분산 저장한다.

다음 단계에서는 인공지능 원천 데이터 시스템(AI Source Data System)에서 인공지능 모델링의 설계나 응용에 따라 데이터를 추출하고 가공한다. 인공지능 학습용 데이터셋(AI Trning Dataset)은 이러한 과정을 통하여 구축되게 된다.

3.3 소형가전제품 이미지 내 오브젝트 분석

그림 1. 과 같이 원천데이터로부터 데이터를 구축은 소형가전제품의 이미지 내의 오브젝트(Object)를 분석에서부터 시작된다. 즉, 인공지능이 학습할 후 있는 오브젝트를 분석하고 정의하여 원천데이터에서 오브젝트를 추출, 변환하여야 한다. 이 오브젝트는 모델링이나 응용플랫폼의 영향을 받기도 한다.

표 3. 은 소형가전제품의 원천데이터의 이미지 데이터에서 "품질표시나 주의문구"에 대한 OCR 데이터(Optical Character Recognition Data), 품질마크나 주의마트와 같은 정의가 명확한 마크에 대한 이미지 분류 데이터 이미지 데이터에서는 도형으로 표시되는 도형이미지 데이터 등에 대해 정리하고 있다. 이러한 소형가전제품의 이미지 데이터들의 범위는 제품 자체의 표시되어 있는 이미지 데이터뿐만 아니라 포장재에 표시되어 있는 이미지 데이터들을 학습하도록 한다.

실제 상황에서는 제품 자체의 사진으로 인식되는 경우도 있지만 포장된 상태의 정보가 필요한 경우도 발생한다. 예를 들어 현장에서 판매되는 제품은 포장박스 등으로 포장된 상태로 판매가 되고 있다. 이렇게 포장된 제품에 대해서 적용 할 수 있는 데이터가 필요하다. 이러한 이미지는 제품의 형상에 대한 정보만 포함 하는 것이 아니라, 제품의 품질표시나 주의사항 또는 품질 마크나 주의 마크 등이 이 제품을 규정하고 기능을 판단하는데, 더 중요한 단서가 되어 인식률을 높일 것으로 생각할 수 있다. 물론 법률적으로도 제품 자체에도 품질표시나 마크가 표기되어있다. 그래서 제품이나 포장재의 이미지에 포함이 되어 있는 문자를 OCR 데이터에 포함하여 제품의 인식률을 높이고자 한다. 물론 품질표시나 주의사항 이외에 광고 효과를 높이기 위한 문자도 포함되어 있을 것이고, 이 문자 역시 제품을 분류하는데 도움이 된다.

또한, 제품이나 포장재에 표기가 되어 있는 품질마크와 주의마크도 제품을 분류하는데 좋은 키워드가 될 것이기 때문에 이 부분도 인식 할 수 있도록 라벨링과 어노테이션을 적용한다.

표 3. 소형가전제품의 라벨 또는 어노테테션 데이터 분류표
Table 3. The classification of Raw Data for Small Home Appliances

Data	AI Model	Description
Quality indication OCR data	The Quality – LaTeX transform AI model	- Convert the text of the quality mark in the image of the product or packaging to LaTeX. - Created through a typesetting library. - Expected to be able to increase the Small Home Appliance recognition rate in images.
Caution statement OCR data	The Caution – LaTeX transform AI model	- Convert the text of the Caution in the image of the Small Home Appliance or Packaging. - Created through a typesetting library. - Expected to be able to increase the product recognition rate in images.
Shape Classification Data	Shape Image Classification AI Model	- Classifies the shapes included in the images of the products or the packing. - Expected to increase product recognition rate by the shape of the product image.
Quality Mark Classification Data	Quality Mark – AI Image Model	- Classifies the type of Quality Mark included in the Product or Packaging Image. - Expected to increase the Recognition rate by the Quality Mark.
Caution Mark Classification Data	Caution Mark – AI Image Model	- Classifies the type of Caution Mark included in the Product or Packaging Image. - Increase the Small Home Appliance Recognition rate by the Caution Mark.

3.4 소형가전제품 빅 데이터 분석

지금까지 데이터 가공 절차를 통하여, 소형전자제품 이미지의 훈련 이미지, 검증 이미지 및 평가(Test) 이미지 등을 정의하고 데이터화 하였다. 이러한 데이터는 빅 데이터 분석에 따라 KC인증, 품목분류, 등 인공지능 응용서비스(KC인증 품질표시 지능화 서비스 등)에 활용될 학습 데이터셋을 구축하게 된다.

그림 2. 는 빅 데이터 분석에 따라 정리된 원천데이터로 일반적인 인공지능 학습용 원천데이터의 가공과 검증까지의 절차를 설명하고 있다.

여기서, 소형가전제품의 빅 데이터 분석에 따라 인공지능 학습용 원천데이터의 가공 방법 및 절차를 설

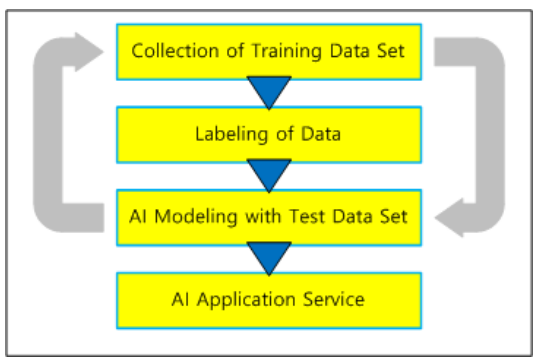

그림 2. 인공지능 학습용 원천데이터 가공 방법 및 절차
Fig. 2. The Processing Method and The Procedure for AI Learning Raw Data

www.dbpia.co.kr

명하고 있다. 즉, 소형가전제품의 훈련 데이터를 수집한다. 각 데이터의 이미지에서 정의 된 오브젝트를 품목분류에 따른 HS Code와 적합성평가에 분류한 기기 부호에 따른 분류와 품질표시에 대한 라벨링을 하고, 이미지 내의 문자나 마크 및 도형에 대한 어노테이션 과정을 통하여 데이터셋을 구축한다.

이러한 데이터셋은 인공지능 모델링을 하고 이 데이터셋으로 학습한 인공지능 응용서비스에 적용함으로써 데이터의 유용성을 평가 할 수 있다. 이 데이터셋에 대한 검증은 다음 테마의 논문에서 다룰 예정이다.

Ⅳ. 소형가전제품 HS Code 데이터셋의 인공지능 지도학습 분석 설계

4.1 소형가전제품 HS Code 데이터셋의 인공지능 지도학습 분석 설계

인공지능의 지도학습용 데이터셋을 구현하기 위하여 소형가전제품에 대한 원천데이터를 어떻게 설계하는지 그림 3. 에서 설명하고 있다. KC 시험소의 이미지 데이터를 가공하여 이미지 데이터와 OCR 데이터로 분류하고, 이 데이터로 구분 된 제품은 품목분류 방식에 따라 HS Code를 결정한다.

인공지능이 원천데이터를 바로 학습할 수 없기 때문에 3장에서 제시한 아이디어로 분석하고 설계하여 데이터셋으로 구현했다. 그림 3. 에서 원천데이터는 KC 적합등록 시험소와 관세법인의 엔터프라이즈 데이터 웨어하우스(EDW)에 있다. 이 데이터 웨어하우스에 저장이 되어 있는 소형가전제품의 이미지 원시데이터(Small Appliances Image Souce Data)에서 제품의 이미지 관련 데이터 항목(Items)을 도출한다. 이 소형가전 제품의 이미지로 데이터베이스(Data Base) 스키마 설계(Schema Design)을 한다. 소형가전제품 이미지 원시데이터 시스템(Small Appliances Image Souce Data System)을 개발한다.

이 소형가전제품 이미지 원시데이터 시스템으로 소형가전제품 인공지능 학습용 데이터셋(Small Appliances AI Learning Dataset)을 설계한다. 이 시스템은 소형가전제품 관련 학습데이터의 항목들(Items) 도출하고 정의한다. 즉, OCR, 도형분류(Classifications)나 도형기호(Symbols)로부터 인공지능 학습용 오브젝트를 도출한다.

또한 관세/인증 전문가들이 정의한 학습데이터용 라벨을 작성하고 교차로 검증했다. 이러한 과정은 소형가전제품의 학습용 데이터 시스템(Small Appliances Learning Data System)을 개발의 필요성

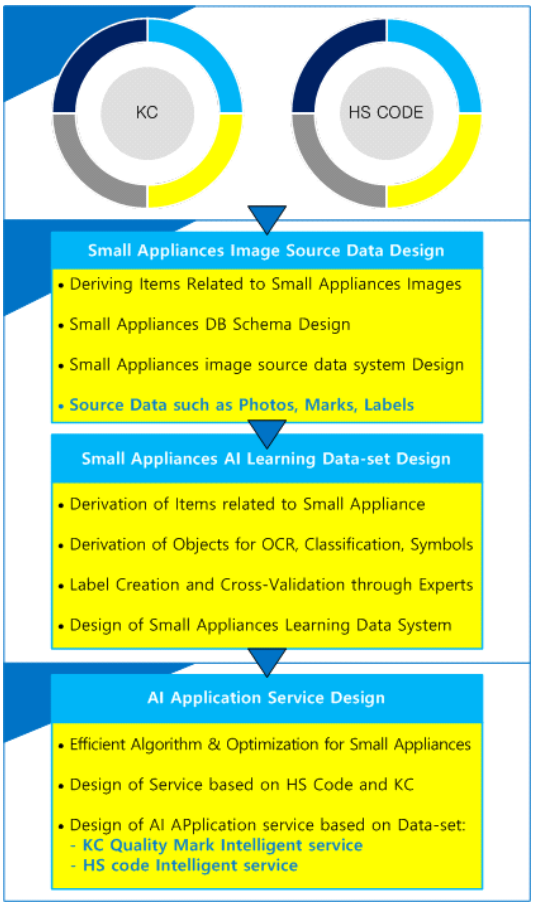

그림 3. 인공지능 원천 & 학습용 데이터 구축, 인공지능 모델과 응용 서비스 설계
Fig. 3. AI Data Construction, AI model, AI Application Servive Design

과 가능성을 제시하고 있다.

이 시스템에서 설계되고 구현된 데이터셋이 HS Code를 산출과 KC 인증 이미지 데이터 분석에 효율적(Efficient)이고 최적화(Optimization)된 알고리즘이다. 그리고 품목분류(HS Code)와 KC 인증에 기반을 둔 서비스의 디자인의 적용이 가능하다. 향후, 이 데이터셋은 "KC 품질표시 지능화 서비스(HS Code Intelligent Service)"의 구현을 가능하게 할 것이다.

4.2 인공지능 지도학습 데이터셋 샘플링 설계

아래의 그림 4. 는 인공지능 학습용 데이터셋 구축 절차이다. 소형가전제품의 데이터셋 구축 절차는 "수집(Data Collection) - 정제(Data Refinement) - 가공(Data Processing) - 모델링(Modelling)" 순서로 진행한다. 주의 할 점은 여러 주기로 나누어 진행하는 것을 설명하고 있다. 그림 4.의 데이터 구축 프로세스 중

www.dbpia.co.kr

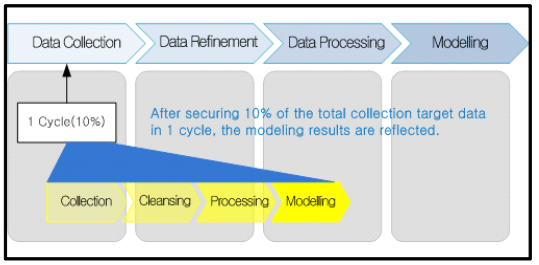

그림 4. 인공지능 학습용 데이터셋 구축 절차
Fig. 4. Dataset Building Procedures for AI Training

수집 단계에서 목표 데이터의 10% 수집 후, 10%의 데이터를 대상으로 "수집-정제-가공-모델링"을 선제적으로 수행하여 사전 검증 및 개선사항 도출하면서 반복적으로 설계한다.

수집하고자 하는 소형가전제품의 각각 데이터 총량, 이미지 수 기준의 10% 획득 완료 시 1 사이클(Cycle)을 중심으로 검사하고, 이때 얻은 데이터셋을 1 사이클 데이터셋(Cycle Dataset)이라 명명한다. 획득된 소형가전제품의 데이터는 정제와 라벨링을 거쳐 인공지능 학습 모델에 바로 적용할 수 있는 형태로 가공한다.

획득된 소형가전제품의 데이터는 8 : 1 : 1 비율로 나누어 학습하고 검증하며 시험에 활용한다. 각 데이터 종류의 약 10%로 학습이 진행된 것이므로, 전체 데이터가 주어졌을 때보다 성능이 낮을 것으로 예상된다. 물론, 전체 소형가전제품의 데이터가 주어졌을 때 편향성이 감소하고, 분산이 증가하여 모델이 더 높은 일반화 능력을 지녀야 할 수 있기에 1 사이클 데이터셋에서 성능이 더 높을 수도 있다. 이렇게 구축하는 소형가전제품의 데이터에 대해서 각각의 인공지능 기계학습 모델을 이용하여 학습하도록 연구한다. 각 데이터의 특성을 한 가지로 묶어 진행하기에 서로 차이점이 많아 통합하여 진행하지 않는다.

4.3 인공지능 지도학습 라벨링/어노테이션 설계

소형 가전제품 이미지 데이터에 대한 어노테이션 데이터와 기타 공통 데이터(ID, HS CODE, 품질표시)를 하나의 JSON 파일로 저장하고 관리한다. 그 이유는 특정한 소프트웨어에 종속되지 않는 포맷과 동시에 구조화된 어노테이션 정보를 저장하기에 적합하기 때문이다.

표 4.는 소형가전제품의 어노테이션 정보의 저장 구조이다. 소형가전제품의 이미지 데이터에 대한 어노테이션 데이터와 기타 공통 데이터를 JSON 파일(File)에 저장하고 관리하는 구조를 보여주고 있다. 여

표 4. 어노테이션 정보 저장 구조
Table 4. Annotation Information Storage Structures

Attribute Name	Attribute Description	Type	Example
project[].annotate[].id	ID	string	ANTT0_1
project[].annotate[].count	Number of Points	int	324
project[].annotate[].delivery_date	Creation Date	date	2021-04-01 9:48
project[].annotate[].label	Device Classification	string	table_01
project[].annotate[].category	Classification	int	0: EMC 1: Safty
project[].annotate[].memo	Other Information	string	main table
project[].annotate[].media_type	Media type	int	0: image, 1: video, 2: link
project[].annotate[].media_data	Media data	binary	-
project[].annotate[].minx	Min. X-coordinate	double	10203.48
project[].annotate[].maxx	Max X-coordinate	double	30243.21
project[].annotate[].miny	Min. Y-coordinate	double	2132.39
project[].annotate[].maxy	Max. Y-coordinate	double	4593.49
project[].annotate[].minz	Min. Z-coordinate	double	23
project[].annotate[].maxz	Max. Z-coordinate	double	243.23

기서, ID(Identification)는 어노테이션의 분류와 분류별 개수를 이용하여 고유하게 부여한다. 소형가전제품의 이미지 등의 미디어는 배열 형태로 복수 등록한다. 소형가전제품의 이미지 데이터 항목의 경우 JSON 내에 BINARY 형태로 값을 저장된다.

표 5. 는 소형가전제품의 라벨링 및 어노테이션 구조 설명 및 구축 샘플이다. 샘플 작업은 통일된 기준의 상세한 가이드를 수립하여 진행한다.

라벨링의 형식은 객체를 직사각형의 박스 안에 포함되게 하고, 객체 전체가 커버되고 박스 안에 객체 외의 여백을 최소가 되도록 지정한다.

많은 소형가전제품의 객체를 라벨링 하기 때문에 라벨링 대상에 대한 정확한 인지, 작업 대상을 사전에 작업 내용에 대한 정보를 제공하여 작업자가 충분히 인지고 작업해야 한다. 즉, 소형 가전제품 객체 식별의 정확한 예측 모델 구축을 위해 객체에 대한 정확한 라벨링(Bounding Box) 진행. 라벨링을 하는 크기에 맞추어 라벨링을 형성해야 한다. 포괄적으로 넓게 잡아 객체와 박스사이 여백을 생성하거나 너무 좁게 잡아 이미지가 박스 바깥으로 나가게 형성되지 않도록 작업하는 것이 중요한다.

이러한 소형가전제품에 대한 데이터셋을 검증하기 위해서는 인공지능 지도학습 모델링이 필요하다. 이 모델링에 대한 검출기의 성능을 검증하기위한 모델의 성능 지표로는 mAP를 적용하는 것이 일반적이며, 여기서 mAP(mean Average Precision)는 클래스당 AP(Average Precision)를 구하여 이 값들을 평균 낸 값이다.

적용 모델로는 YOLOv4를 사용하는 것을 계획하고 있다. YOLOv4는 최신 딥러닝 기법을 적용하고 있

www.dbpia.co.kr

표 5. 라벨링 및 어노테이션 구조 설명 및 구축 예시
Table 5. Labeling and Annotation Structure Description and Construction Examples

Name	Description	Example(json)
id	ID	{ "id":"ANTT0_1", "count":342, "delivery_date": "2021-03-31 14:39:31", "label":{ "category":1, "name":"table01", "memo":"main table", "media":[{ "media_type":"1", "media_data":, }], [{ "media_type": "0", "media_data": , }] }, "boundary":{ "minx":10203.48, "maxx":30243.21, "miny":2132.39, "maxy":4593.49, "minz":23, "maxz":243.23, }, "pointcloud":, }
count	Number of Points	
delivery_date	Creation Date	
	(yyyy-MM-dd hh:mm:ss)	
label		
name	Device Code	
category	HS Code	
memo	Classification	
media_type	Media type	
	(ex: image)	
media_data	Media data	
boundary		
minx	Min. X-coordinate	
maxx	Max X-coordinate	
miny	Min. Y-coordinate	
maxy	Max. Y-coordinate	
minz	Min. Z-coordinate	
maxz	Max. Z-coordinate	
3D data		
pointcloud	Image Data	

으며 성능 향상을 되었고, 객체 인식의 취약한 점을 개선하기 위해 입력 해상도(Input Resolution)을 큰 데이터를 사용하는 것을 고려하고 있다.

V. 결 론

미국의 바이든 행정부의 중국에 대한 무역 정책은 강성 대응으로 진행하고 있다. 또한 COVID-19 PANDEMIC 사태도 4차 유행으로 들어가면서, 대한민국의 수입과 수출은 더욱 어려워지고, 실제적인 무역 갈등이 일어나면서, 발달된 4차 혁명 기술인 인공지능을 활용한 세계 무역의 디지털 개선이 요구되고 있다.

본 논문에서 국내에서 판매되는 소형가전제품에 대한 HS Code를 결정하고, KC 적합성 대상을 분류할 수 있는 인공지능 지도학습용 데이터셋을 분석하였다. HS 코드에 대한 품목분류 알고리즘이 적용된 인공지능 기계학습 모델을 설계하였다. 본 논문에서는 KC 적합등록 대상의 소형가전제품의 이미지로 HS Code를 구분하는 인공지능용 학습용 데이터셋을 구현하기 위한 품목분류의 법령에 따른 인공지능 데이터셋 라벨링과 어노테이션에 대한 연구하였다.

또한, YOLOv4로 품목분류 응용프로그램에 적합한 모델링을 설계하여 이 데이터셋을 유용성을 확인하는 은 다음의 논문에서 다룰 예정이다.

본 논문의 연구 결과는 소형가전제품에 대한 품목분류를 수행하는 기기부호로 분류함으로써 품질의 정확성의 개선을 가져와 인식률을 높일 수 있다. 또한, KC 적합등록에서 정확한 기기부호는 비관세 무역대응의 인공지능 응용프로그램의 유용성도 확보하게 하여, 세계 무역에서의 비관세 장벽으로 불리는 수작업들을 인공지능을 활용하여 반자동화 하여 세계 무역의 발전에 기여한 기초자료를 제공 할 것이다.

향후 연구로는 소형가전제품의 데이터셋을 기반으로 다른 제품군의 품목분류와 무역 처리를 확장하도록 연구되어야 할 것이다.

References

[1] J. H. Min, "COVID-19, global leadership, and US-China relations," *IFANS FOCUS*, May 2020.

[2] J. D. Kim, B. Y. Choi, J. H. Um, and M. C. Jung, "Analysis of the status of non-tariff measures in korea: Focusing on the establishment of a db to identify non-tariff measures," *Korea Inst. for Int. Econ. Policy*, Dec. 2018.

[3] D. G. Jeong, "*Safety certification system and KC certification*," Retrieved Aug. 08, 2019, from https://brunch.co.kr/@jdglaw1/42

[4] N. K. Kim, "*Smarter than humans" speech recognition 'AI assistant' development competition* (2016), Retrieved Jun. 05, 2021, from https://www.yna.co.kr/view/AKR20160614130800009?input=1195m

[5] K. S. Yang, "*Key Competitiveness for Digital Transformation 'AI/Data'* (2021)," Retrieved Jan. 05, 2021, from https://www.apple-economy.com/news/articleView.html?idxno=64240

[6] J. H. Jang, "*Policy direction for effectiveness improvement of data business for ai learning*," NIA, IT & Future Strategy Report, Nov. 2020.

[7] D. S. Park, "*Artificial Intelligence (AI)*

www.dbpia.co.kr

Powerhouse Korea, Data Is Correct," Retrieved Dec. 04, 2019, from https://m.etnews.com/20191204000002?obj=Tzo4OiJzdGRDbGFzcyI6Mjp7czo3OiJyZWZlcmVyIjtOO3M6NzoiZm9yd2FyZCI7czoxMzoid2ViIHRvIG1vYmlsZSI7fQ%3D%3D

[8] D. Y. Kang, "Popular object detection algorithms in deep learning," in *Proc. KIPS Conf.*, vol. 26, no. 1, pp. 427-430, May 2019.

[9] A. Bochkovskiy, C. Y. Wang and H. Y. Mark Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv:2004.10934, Apr. 23. 2020.

[10] Y. C. Yoon, J. H. Sang, and S. M. Park, "Trands of plant image processing technology," *Electron. and Telecommun. Trends.*, vol. 33, no. 4, pp. 54-60, Aug. 2018.

[11] J. M. Song, S. B. Lee, and A. R. Park, "A study on the industrial application of image recognition technology," *J. Korea Contents Assoc.*, vol. 20, no. 7, pp. 86-96, 2020.

오 영 주 (Young-joo Kim)

1988년 2월 : 국민대학교 전자공학과 졸업

1990년 2월 : 국민대학교 전자공학과 석사

2019년 3월~현재 : 호서대학교 벤처대학원 융합공학과 박사과정

<관심분야> 전자공학, 통신공학, 광통신 공학, 융합공학

박 대 우 (Dea-woo Park)

2004년 2월 : 숭실대학교 대학원 컴퓨터공학과 공학박사

20007년 3월~현재 : 호서대학교 벤처대학원 융합공학과 교수

<관심분야> 사이버보안, 정보보호, 해킹포렌식, ICT, 네트워크, 인공지능

[ORCID:0000-0003-4073-2065]

www.dbpia.co.kr