Federated Learning

Feature Connectivity between Server and Clients in a Federated Learning

Oct. 2022

Intelligent Information Processing Lab

JunHo Yoon

Contents

01. Related Work

Reference

클래스 불균형 문제에 연합학습 적용을 위한 최적화 기법 연구

이현수*, 홍성은**[1], 방준일**[2], 김화종***

Study of Optimization Techniques to Apply Federated Learning on Class Imbalance Problems

Hyeonsu Lee*, Seongeun Hong**[1], Junil Bang**[2], and Hwajong Kim***

이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No. 2018-0-00261) 또한 2019년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2019007059)

요 약

고도로 발달된 개인 정보 식별 기술에 의해 개인에 대한 식별이 용이해지면서, 정보 사회에서 정보 주체의 권리를 보장할 수 있는 다양한 방안이 요구되고 있다. 연합학습은 이러한 요구에 의해 제안된 기계학습 방식으로 데이터를 비공개로 유지하면서 기계학습 알고리즘을 훈련하기 위한 특정 접근 방식이다. 본 논문에서는 개인정보보호 이슈에 민감한 의료 산업에 연합학습을 적용할 때 발생할 가능성이 있는 문제점을 파악하기 위해 망막 환자 데이터셋을 실제 의료기관이 데이터를 보유하고 있는 환경처럼 데이터 분포를 불균형하게 분할했다. 여기서 발생하는 클래스 불균형 문제에 다양한 학습 최적화 기법을 적용한 실험을 진행한 결과, 언더 샘플링 및 TopkAvg 기법을 적용한 실험에서 F1 score 0.96을 달성했으며, 학습 시간도 단축시켰다.

Abstract

Recently, as highly advanced personal identification technology has made it easier to identify individuals, various measures are required to guarantee the rights of information subjects in the information society. Federated learning is a machine learning approach proposed by these needs, a specific approach to educating machine learning algorithms while keeping the data private. In this paper, in order to identify problems that may arise when applying federated learning to the medical industry, which is sensitive to privacy issues, a retinal patient data set, was disproportionately distributed like the environment in which the actual medical institution holds the data. As a result of experiments applying various learning optimization techniques to class imbalance problems that occur here, F1 score 0.96 was achieved in experiments with under sampling and TopkAvg techniques, and the learning time was also shortened.

Keywords

imbalance data, federated learning, sampling, optimization

■ Title
- 클래스 불균형 문제에 연합학습 적용을 위한 최적화 기법 연구

■ Journal
- 한국정보기술과학학회논문지

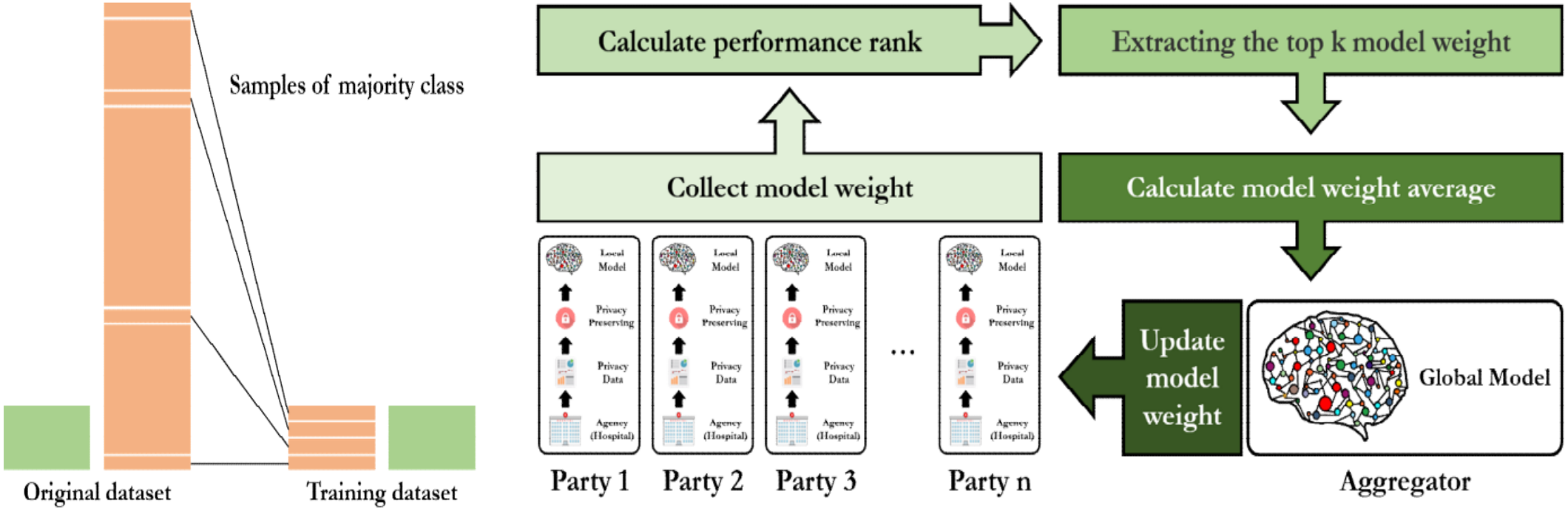
■ Published
- Jan. 2021

■ Citiation(Google Scholar)
- 0

Reference

■ 클래스 불균형 문제에 연합학습 적용을 위한 최적화 기법 연구

- 불균형 데이터 문제가 가져오는 성능 저하를 개선하기 위해 언더 샘플링 기법과 TopkAvg 기법을 적용함



Reference

■ 클래스 불균형 문제에 연합학습 적용을 위한 최적화 기법 연구

- 실험 결과 불균형 데이터셋에 대하여 샘플링 기법의 효과를 입증할 수 있었으며, TopkAvg 기법을 사용하여 학습 시간 또한 4~16초 감소하였다.

Distribution	Party name	Best performance
balance	party1	**0.95**
	party2	**0.95**
	party3	0.94
	party4	**0.95**
imbalance	party1	0.50
	party2	0.25
	party3	0.74
	party4	0.87

Distribution	Party name	Best Performance
class weight	party1	0.50
	party2	0.25
	party3	0.90
	party4	0.80
over sampling	party1	0.94
	party2	0.94
	party3	0.94
	party4	0.94
under sampling	party1	0.94
	party2	0.94
	party3	0.94
	party4	**0.95**

Reference

Data Poisoning Attacks on Federated Machine Learning

Gan Sun, *Member, IEEE*, Yang Cong, *Senior Member, IEEE*, Jiahua Dong, Qiang Wang, Lingjuan Lyu, and Ji Liu

***Abstract*—Federated machine learning which enables resource-constrained node devices (e.g., Internet of Things (IoT) devices and smartphones) to establish a knowledge-shared model while keeping the raw data local, could provide privacy preservation, and economic benefit by designing an effective communication protocol. However, this communication protocol can be adopted by attackers to launch data poisoning attacks for different nodes, which has been shown as a big threat to most machine learning models. Therefore, we in this article intend to study the model vulnerability of federated machine learning, and even on IoT systems. To be specific, we here attempt to attacking a popular federated multitask learning framework, which uses a general multitask learning framework to handle statistical challenges in the federated learning setting. The problem of calculating optimal poisoning attacks on federated multitask learning is formulated as a bilevel program, which is adaptive to the arbitrary selection of *target* nodes and *source attacking* nodes. We then propose a novel systems-aware optimization method, called as attack on federated learning (AT^{2}FL), to efficiently derive the implicit gradients for poisoned data, and further attain optimal attack strategies in the federated machine learning. This is an earlier work, to our knowledge, that explores attacking federated machine learning via data poisoning. Finally, experiments on several real-world data sets demonstrate that when the attackers directly poison the *target* nodes or indirectly poison the related nodes via using the communication protocol, the federated multitask learning model is sensitive to both poisoning attacks.**

***Index Terms*—Bilevel optimization, data poisoning, federated machine learning, multitask learning.**

Manuscript received June 9, 2021; revised September 21, 2021 and October 19, 2021; accepted November 4, 2021. Date of publication November 17, 2021; date of current version June 23, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62003336 and Grant 62073205; in part by the National Postdoctoral Innovative Talents Support Program under Grant BX20200353; in part by the State Key Laboratory of Robotics under Grant 2022-Z06; and in part by the Nature Foundation of Liaoning Province of China under Grant 2020-KF-11-01. *(Gan Sun and Jiahua Dong contributed equally to this work.) (Corresponding author: Yang Cong.)*

Gan Sun and Yang Cong are with the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China, and also with the Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China (e-mail: sungan1412@gmail.com; congyang81@gmail.com).

Jiahua Dong is with the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China, also with the Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: dongjiahua@sia.cn).

Qiang Wang was with the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China, also with Shenyang University, Shenyang 110044, China (e-mail: wangqiang@sia.cn).

Lingjuan Lyu is with Sony AI, Tokyo 108-0075, Japan (e-mail: lingjuanlvsmile@gmail.com).

Ji Liu is with the Beijing Kuaishou Technology Company, Ltd., Beijing 100005, China (e-mail: jiliu@kwai.com).

Digital Object Identifier 10.1109/JIOT.2021.3128646

I. Introduction

MACHINE learning has been widely applied into a broad array of applications, e.g., spam filtering [35], lesions segmentation [11], natural gas price prediction [1], and Internet of Things (IoT) devices [10], [16], [33]. Among these applications, the reliability or security of the machine learning system has been a great concern, including adversaries [15], [34]. For example, for a product recommendation system [28], researchers can either rely on public E-commerce platforms, e.g., Taobao or Amazon Mechanical Turk, or collect training data by private teams. Unfortunately, both of these above systems have the opportunity of being injected corrupted or poisoned data by attackers, which could be a security risk to the physical objects or the IoT systems. To improve the robustness and reliability of existing machine learning and IoT systems, it is critical to study how well machine learning performs under the poisoning attacks.

For the attack strategy on existing machine learning methods, it can be partitioned into two categories: 1) causative attacks and 2) exploratory attacks [3]. Causative attacks methods affect machine learning models via controlling over training data, whereas exploratory attacks methods could take use of misclassifications without affecting the training phase. However, more previous researches on poisoning attacks focus on the scenarios that training samples are collected in a centralized location, or the training samples are sent to a centralized location via a distributed data collection network, e.g., autoregressive models [1], support vector machines (SVMs) [5], and collaborative filtering [20]. There exist scarce works studying poisoning attacks on federated machine learning [18], [30], [32], where the training data are distributed across multiple IoT devices (e.g., users' mobile devices: phones/tablets), and may be privacy sensitivity. To further improve its robustness, in this article, our work explores how to attack the federated learning system via data poisoning.

For federated machine learning [26], [27], [31], [32], its main idea is to build a knowledge-shared machine learning models while guaranteeing data privacy, where the raw data are distributed on multiple local devices. Even though most recent progressions have been achieved on tackling

■ Title
- Data Poisoning Attacks on Federated Machine Learning

■ Journal
- IEEE Internet of Things

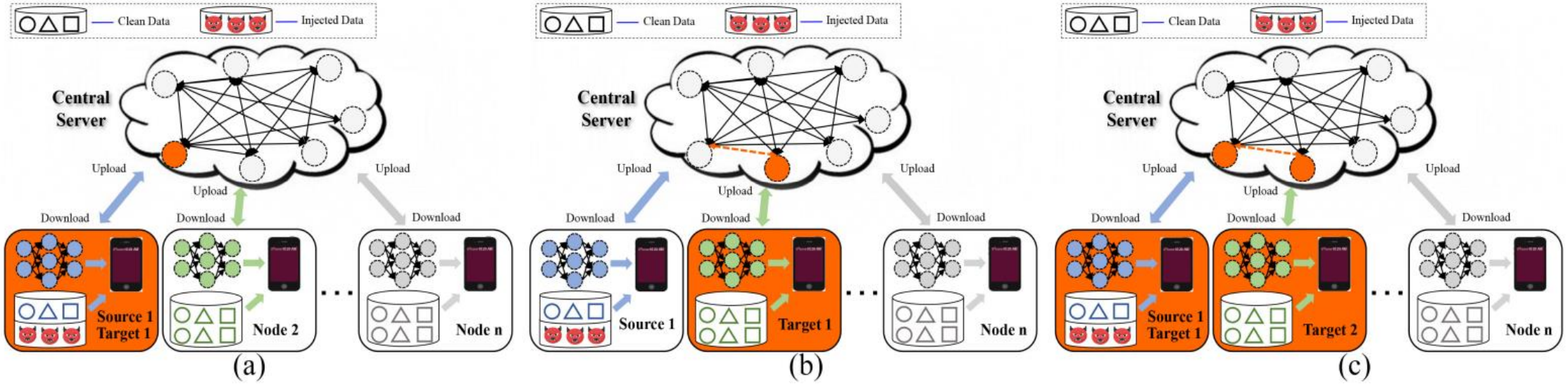
■ Published
- Nov. 2021

■ Citiation(Google Scholar)
- 57

Reference

■ Data Poisoning Attacks on Federated Machine Learning

- 샘플링 방식의 경우 주요 정보 손실 또는 과적합 및 노이즈나 이상치에 민감한 문제가 존재함
- 따라서 GAN을 사용한 Client 내에서 데이터를 증강하여 사용한 방법이 있으나, 기존 데이터에서의 특징 정보가 사라지거나 데이터가 공격 받으면 학습 능력이 저하됨
- 본 논문에서는 Client를 3가지 방식(직접 공격, 간접 공격, 하이브리드 공격)을 사용하여 공격시 에러율을 분석함

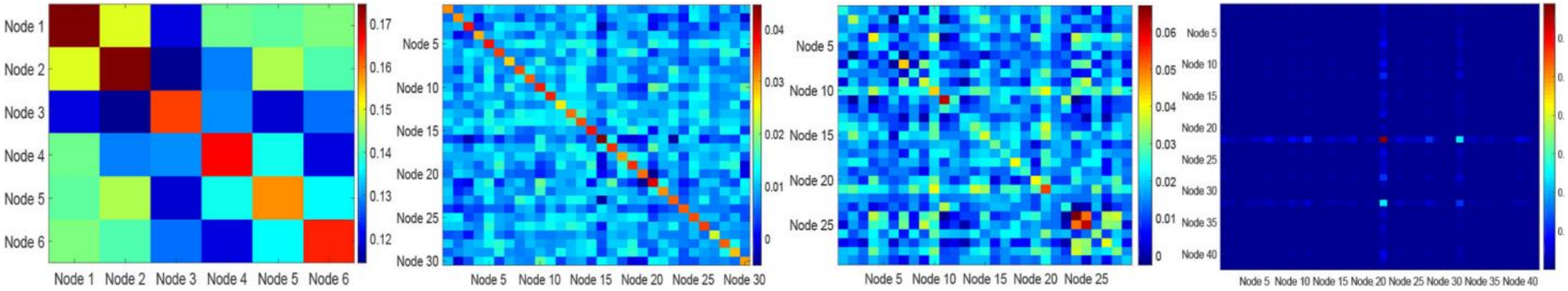


Reference

■ Data Poisoning Attacks on Federated Machine Learning

- 실험 결과 데이터에 대하여 직접적인 공격을 수행하는 경우 에러율이 가장 높았으며, 이외에도 에러율이 높은 것을 보아 연합학습 방식에서는 감염 공격에 모두 민감한 것을 알 수 있었음

	Metrics	Non attacks	Random direct attacks	Random indirect attacks	Random hybird attacks	Direct attacks	Indirect attacks	Hybird attacks
EndAD	Error(%)	6.881±0.52	7.659±1.14	6.888±0.45	7.154±0.16	**28.588±3.74**	7.324±0.62	16.190±2.26
Human Activity	Error(%)	2.586±0.84	3.275±0.71	2.894±0.83	3.172±0.69	**29.422±2.96**	3.438±0.34	17.829±2.75
Landmine	Error(%)	5.682±0.28	5.975±0.36	5.735±0.36	5.819±0.22	**13.648±0.54**	7.428±0.39	9.579±0.27
	Avg. Error (%)	5.049±0.55	5.636±0.74	5.172±0.55	5.382±0.36	**23.886±2.41**	6.069±0.45	14.533±1.76
Parkinson-Total	RMSE	6.302±0.45	13.651±2.10	6.633±0.75	11.145±1.83	**44.939±3.21**	7.763±0.82	21.990±3.17
Parkinson-Moter	RMSE	4.125±0.50	11.472±2.51	5.046±1.14	9.422±1.81	**32.992±3.78**	6.866±1.21	16.956±3.78
	Avg. RMSE(%)	5.213±0.48	12.562±2.31	5.839±0.95	10.284±1.82	**38.966±3.49**	7.314±1.02	19.473±3.48



Reference

Learning Transferable Visual Models From Natural Language Supervision

Alec Radford [*1] Jong Wook Kim [*1] Chris Hallacy [1] Aditya Ramesh [1] Gabriel Goh [1] Sandhini Agarwal [1]
Girish Sastry [1] Amanda Askell [1] Pamela Mishkin [1] Jack Clark [1] Gretchen Krueger [1] Ilya Sutskever [1]

Abstract

State-of-the-art computer vision systems are trained to predict a fixed set of predetermined object categories. This restricted form of supervision limits their generality and usability since additional labeled data is needed to specify any other visual concept. Learning directly from raw text about images is a promising alternative which leverages a much broader source of supervision. We demonstrate that the simple pre-training task of predicting which caption goes with which image is an efficient and scalable way to learn SOTA image representations from scratch on a dataset of 400 million (image, text) pairs collected from the internet. After pre-training, natural language is used to reference learned visual concepts (or describe new ones) enabling zero-shot transfer of the model to downstream tasks. We study the performance of this approach by benchmarking on over 30 different existing computer vision datasets, spanning tasks such as OCR, action recognition in videos, geo-localization, and many types of fine-grained object classification. The model transfers non-trivially to most tasks and is often competitive with a fully supervised baseline without the need for any dataset specific training. For instance, we match the accuracy of the original ResNet-50 on ImageNet zero-shot without needing to use any of the 1.28 million training examples it was trained on. We release our code and pre-trained model weights at https://github.com/OpenAI/CLIP.

1. Introduction and Motivating Work

Pre-training methods which learn directly from raw text have revolutionized NLP over the last few years (Dai & Le, 2015; Peters et al., 2018; Howard & Ruder, 2018; Radford et al., 2018; Devlin et al., 2018; Raffel et al., 2019).

[*]Equal contribution [1]OpenAI, San Francisco, CA 94110, USA. Correspondence to: <{alec, jongwook}@openai.com>.

Task-agnostic objectives such as autoregressive and masked language modeling have scaled across many orders of magnitude in compute, model capacity, and data, steadily improving capabilities. The development of "text-to-text" as a standardized input-output interface (McCann et al., 2018; Radford et al., 2019; Raffel et al., 2019) has enabled task-agnostic architectures to zero-shot transfer to downstream datasets removing the need for specialized output heads or dataset specific customization. Flagship systems like GPT-3 (Brown et al., 2020) are now competitive across many tasks with bespoke models while requiring little to no dataset specific training data.

These results suggest that the aggregate supervision accessible to modern pre-training methods within web-scale collections of text surpasses that of high-quality crowd-labeled NLP datasets. However, in other fields such as computer vision it is still standard practice to pre-train models on crowd-labeled datasets such as ImageNet (Deng et al., 2009). Could scalable pre-training methods which learn directly from web text result in a similar breakthrough in computer vision? Prior work is encouraging.

Over 20 years ago Mori et al. (1999) explored improving content based image retrieval by training a model to predict the nouns and adjectives in text documents paired with images. Quattoni et al. (2007) demonstrated it was possible to learn more data efficient image representations via manifold learning in the weight space of classifiers trained to predict words in captions associated with images. Srivastava & Salakhutdinov (2012) explored deep representation learning by training multimodal Deep Boltzmann Machines on top of low-level image and text tag features. Joulin et al. (2016) modernized this line of work and demonstrated that CNNs trained to predict words in image captions learn useful image representations. They converted the title, description, and hashtag metadata of images in the YFCC100M dataset (Thomee et al., 2016) into a bag-of-words multi-label classification task and showed that pre-training AlexNet (Krizhevsky et al., 2012) to predict these labels learned representations which preformed similarly to ImageNet-based pre-training on transfer tasks. Li et al. (2017) then extended this approach to predicting phrase n-grams in addition to individual words and demonstrated the ability of their system to zero-shot transfer to other image

■ Title

- Learning Transferable Visual Models From Natural Language Supervision

■ Journal

- International Conference on Machine Learning. PMLR 2021

■ Published

- Jul. 2021

■ Citiation(Google Scholar)

- 2247

Connecting Text and Images

■ Approach

- Natural Language Supervision
  - CLIP는 자연어를 supoervision으로 학습하기 때문에 Labeling 작업이 필요 없으며, 각 Modality에 대한 Representaion Learning을 수행 할 수 있고 다른 종류의 Task로도 유연하게 Zero-shot Transfer이 가능함

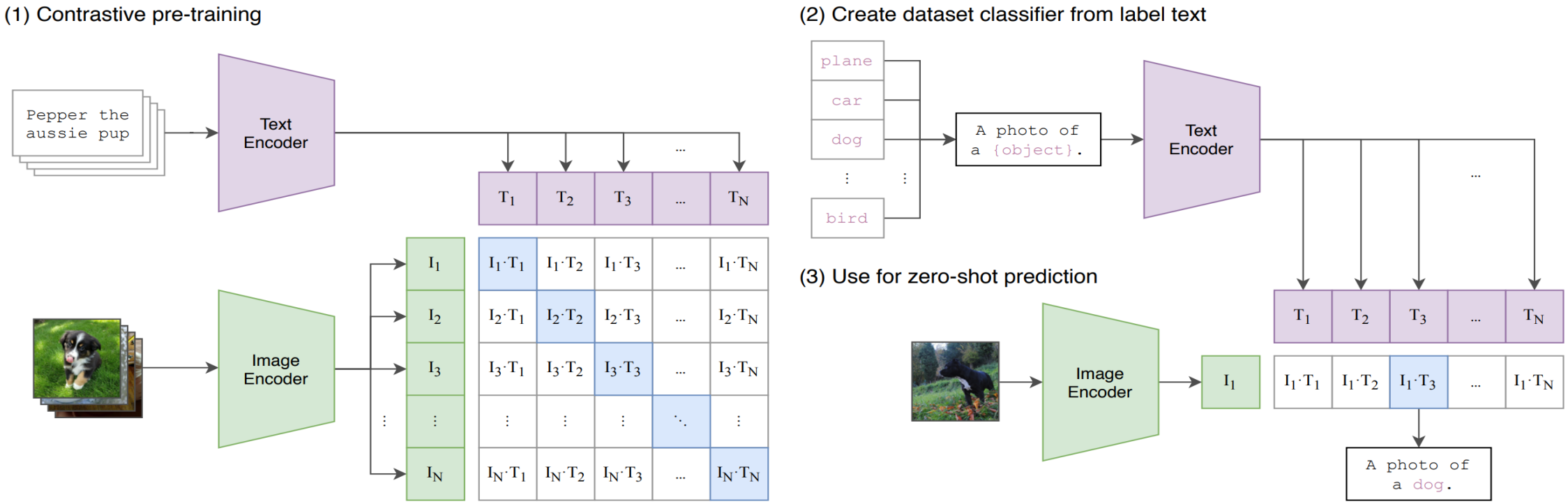
- Creating a Sufficiently Large Dataset
  - MS-COCO, Visual Genome 데이터는 품질은 좋으나 양이 적음
  - YFCC100M은 데이터의 양은 많으나 데이터셋 품질이 다름
  - 따라서 CLIP에서는 WIT(WebImageText)라고 하는 새로운 데이터셋을 만듦

- Selecting an Efficient Pre-Training Method
  - image와 text를 하나의 공통된 공간으로 보냄
  - positive pair에서의 유사도는 최대화 하고 negative pair에서의 유사도는 최소화하도록 학습함
  - 따라서 Multi-modal Embedding Space를 학습함

Connecting Text and Images

■ Using CLIP for zero-shot Transfer

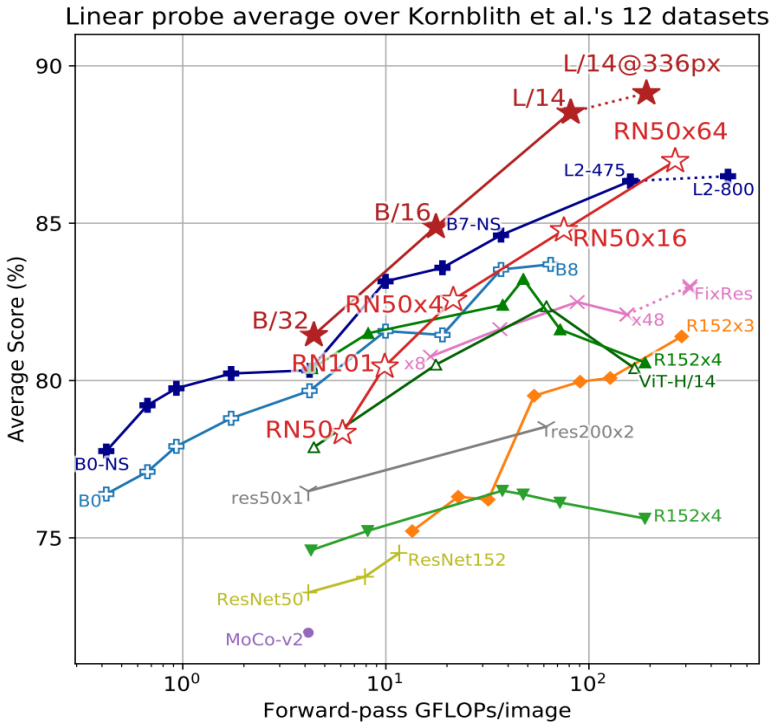
- 학습된 Encoder를 사용하여 각 Modality에 대한 Featur를 추출하여 Class를 예측함

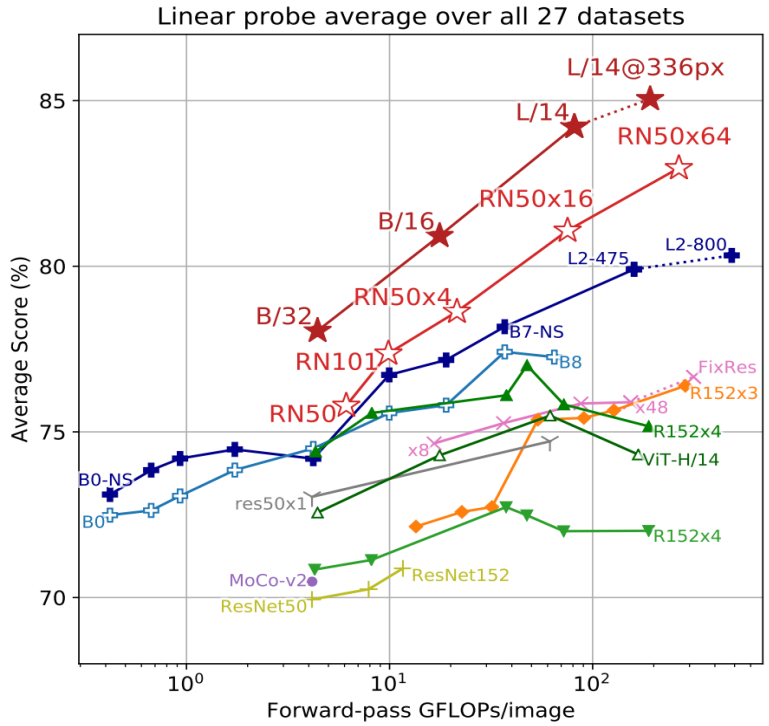


Connecting Text and Images

■ Result

- **CLIP은 Zero-shot Learning과 이미지에서 추출한 Feature를 Downstream Task에서 적용한 것 모두 높은 성능을 달성함**





	Accuracy	Majority Vote on Full Dataset	Accuracy on Guesses	Majority Vote Accuracy on Guesses
Zero-shot human	53.7	57.0	69.7	63.9
Zero-shot CLIP	**93.5**	**93.5**	**93.5**	**93.5**
One-shot human	75.7	80.3	78.5	81.2
Two-shot human	75.7	85.0	79.2	86.1

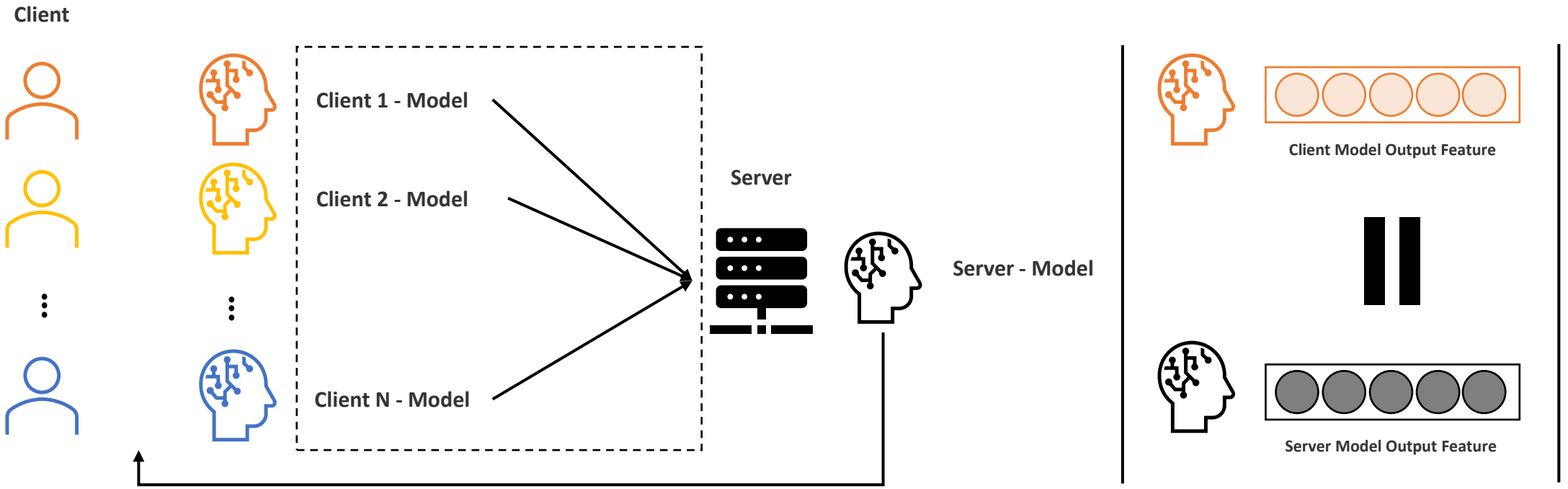
	aYahoo	ImageNet	SUN
Visual N-Grams	72.4	11.5	23.0
CLIP	**98.4**	**76.2**	**58.5**

02. Research Proposal

Proposed Method

■ Title

- (국문) 연합학습 환경에서의 서버와 클라이언트 간의 특징 연결
- (영문) Feature Connectivity between Server and Clients in a Federated Learning



Introduction

■ Non-IID Data

- 연합학습 시 아래와 같이 Client 간의 데이터 불균형 문제가 심각하며, 이를 해결하기 위해 각 Client가 데이터의 일부분을 중앙 서버에 제출해서 공용 데이터셋을 만들어 해결할 수 있으나, 데이터의 보안성을 지킨다는 연합학습의 취지를 훼손함

  1) 데이터 클래스 불균형
  2) 데이터 분포 불균형
  3) 데이터 크기 불균형

■ 기술동향

- 따라서 다음과 같이 다양한 방식을 사용하여 Non-IID 문제를 해결하기 위한 연구가 수행되고 있으나, 이와 같이 데이터를 수정 및 보완하는 방식의 경우 결과적으로 데이터의 보안성을 훼손함
  1) Data Augmentation
  2) Transfer Learning
  3) Client Clustering
  4) Knowledge Distililation

Approach

■ Feature Connectivity

- Multi-modal Learning에서 SOTA 성능을 달성한 기법인 CLIP과 같이 Clients와 Server간의 Feature를 통합함
- 따라서 Non-IID 상황임에도 모든 Client가 동일한 수준의 Feature를 추출하는 Global Encoder 학습이 가능함

■ Clients ←→ Server

- Clients에서 Server로 보내는 모델 가중치의 경우 기존의 Federated Learning 학습 방식과 동일하게 진행하되, 학습 시간 감소 및 불필요한 정보 감소를 위해 TopkAvg를 사용하여 일부 가중치만 통합함

■ Training Flow

- (1) Clients는 Server에 전송한 모델 가중치를 사용하여 Feature추출하고 (2) Server에서 받은 모델 가중치를 사용하여 Feature를 추출하여 (3) 각 Feature가 유사해지도록 학습함
- 최종적으로 추출한 Feature를 사용하여 Classification Task를 수행함
- 따라서 각 Clients 들은 Non-IID 상황에도 데이터를 공유하지 않고 Global한 Encoder 학습이 가능함

Thank you ☺

JunHo Yoon 윤준호
Department of Computer Engineering, Gachon University | Researcher

Tel. +82-31-750-8822 Mobile. +82-10-9110-6257
E-mail. junho6257@gachon.ac.kr