Machine Learning Generalization:
Domain Adaptation & Federated Learning

고려대학교 산업경영공학과
데이터 어낼리틱스 및 헬스케어 시스템 연구실
권정을(Internship)

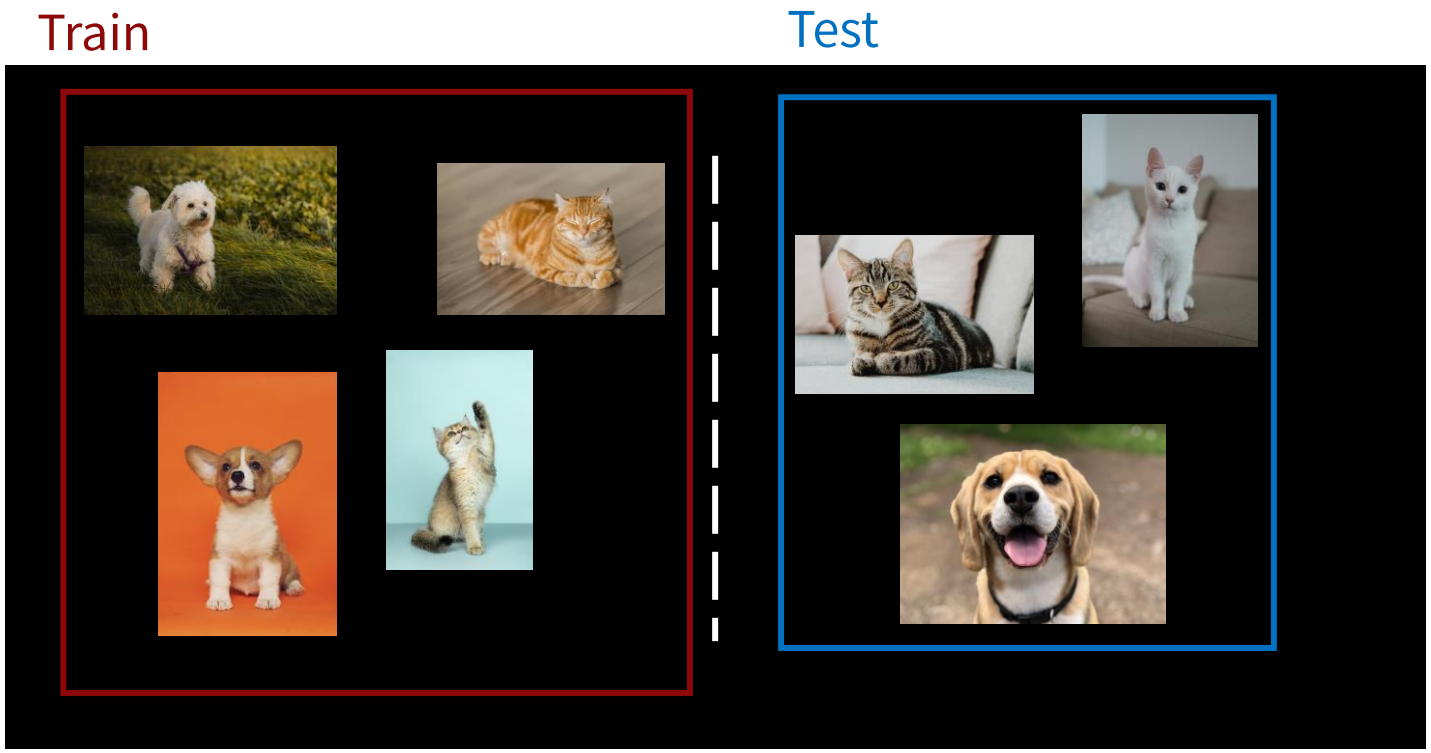
Keywords: domain adaptation, federated learning, representation learning

DAHS
Data Analytics and Healthcare Systems LAB

INDEX

01 Introduction: Usually we try to..

- Let’s consider the binary classification problem
- Train data for Model Training/Test data for Model Evaluation



01 Introduction: Usually we try to..

Empirical error

- $R_s = \{(x_i^s, y_i^s)\}_{i=1}^{m_s} \sim (\mathcal{P}_S)^{m_s}$, a labeld sample drawn i.i.d. from $\mathcal{P}_S$
- Associated **empirical error** of an hypothesis h:

$$R_s(h) = \frac{1}{m_s}\sum_{i=1}^{m_s} I[h(x_i^s) \neq y_i^s]$$

틀린 개수 / 전체 샘플 수 **즉, 경험적 오차**

Classical PAC result: From the same distribution

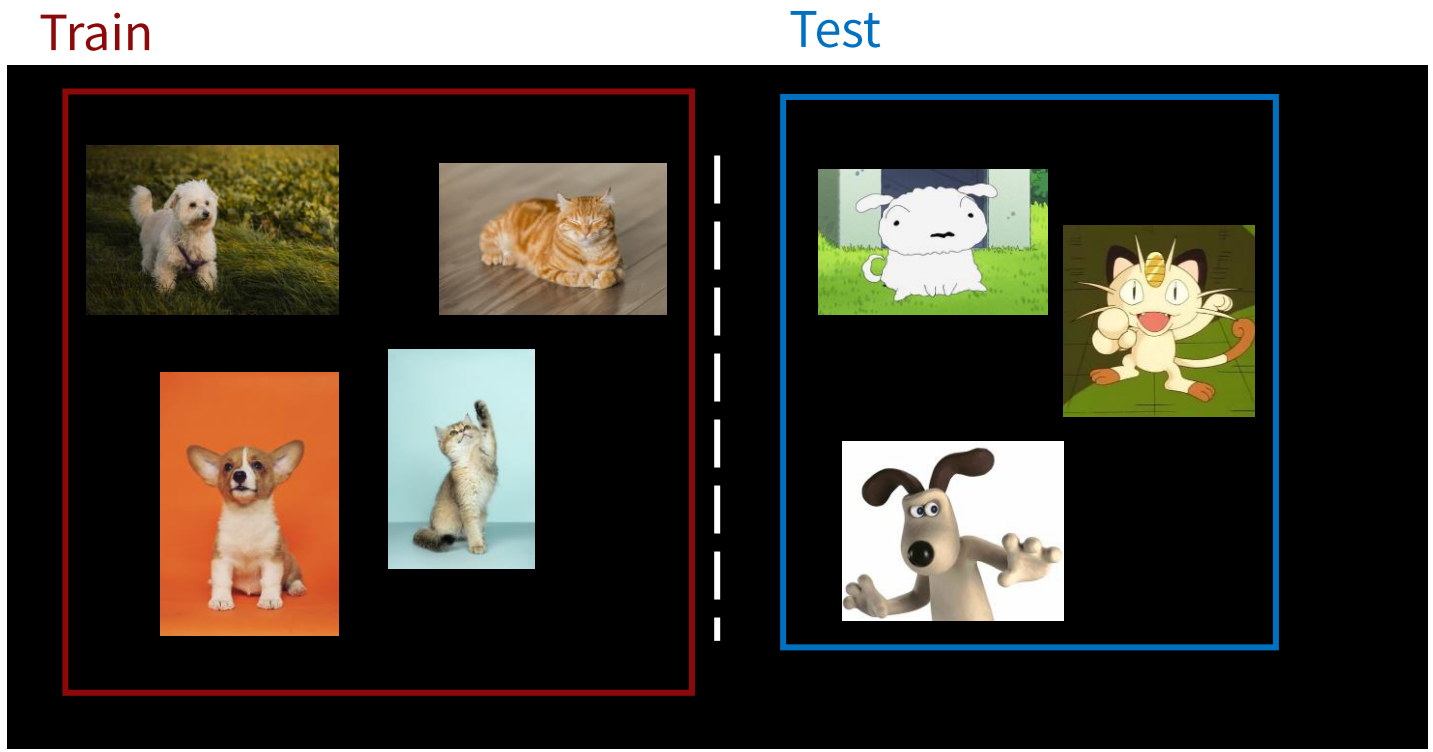
$$R_{P_s}(h) \leq R_s(h) + O\left(\frac{complexity(h)}{\sqrt{m_s}}\right)$$

Error of unseen set

- 데이터의 독립성, 동일 분포 가정(i.i.d)은 데이터를 Train, Validation, Test set으로 분할하여 학습 능력을 평가하는 행위에 대한 당위성을 부여하며 머신러닝에서 **강력한 가정**
- (i.i.d를 가정했을 때) Train error $R_s(h)$와 Model complexity $O(\frac{complexity(h)}{\sqrt{m_s}})$는 Trade-off 관계이며, Overfitting 개념을 설명할 수 있음

01 Introduction: Usually we try to..

- If our data does not satisfy the i.i.d. assumption..
  - *Train data와 Test data가 i.i.d 가정을 위배한다면?*



Domain shift

*Real Image to Animation

ϵ_{test}

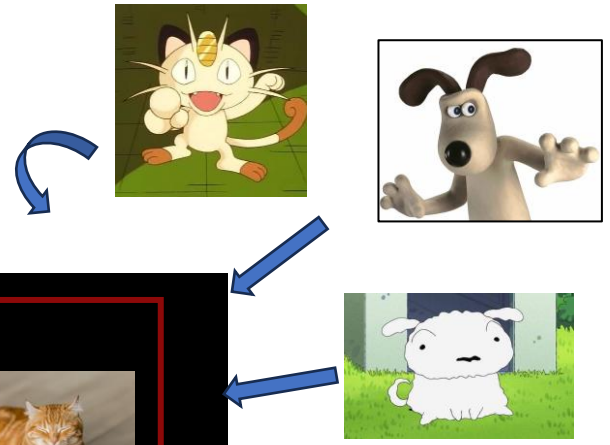
*error of test dataset

DAHS
Data Analytics and Healthcare Systems LAB

01 Introduction: Usually we try to..

- Domain shift 해결방안
  - 라벨링이 되어 있는 해당 도메인의 고품질 데이터 수집 진행
  - 수집된 데이터를 바탕으로 모델 재학습

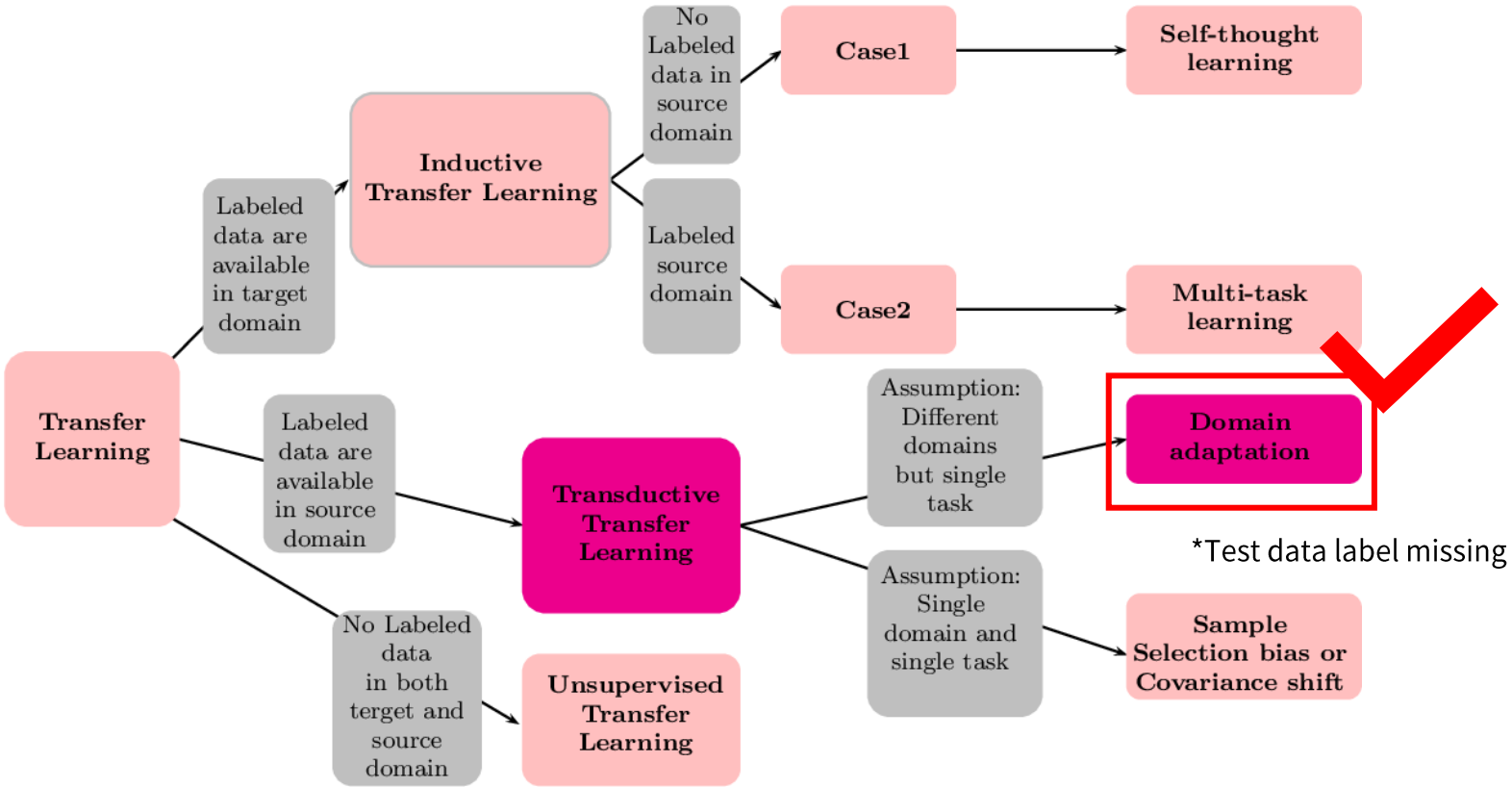
- 한계
  - ➢ Labeled 고품질 데이터 수집은 시간과 비용이 수반
  - ➢ 현실적인 어려움 존재



Transfer Learning

: 한 문제를 해결하기 위해서 얻은 지식과 정보를
다른 문제를 푸는데 사용하는 방식

02 Taxonomy of Transfer Learning



*Test data label missing + different domain

03 Example: Domain Adaption for sentiment analysis

	Electronics	Video games
✓	(1) Compact; easy to operate; very good picture quality; looks sharp!	(2) A very good game! It is action packed and full of excitement. I am very much hooked on this game.
✓	(3) I purchased this unit from Circuit City and I was very excited about the quality of the picture. It is really nice and sharp.	(4) Very realistic shooting action and good plots. We played this and were hooked.
✗	(5) It is also quite blurry in very dark settings. I will never_buy HP again.	(6) It is so boring. I am extremely unhappy and will probably never_buy UbiSoft again.

- Source specific(Electronics): Compact, sharp, blurry
- Target specific(Video games): realistic, boring, hooked
- Domain independent: good, never buy, unhappy

Domain Adaption(DA):

Domain specific 한 학습이 아닌 Domain independent 한 학습
즉, general 한 모델을 얻기 위해 general 한 feature 를 학습하자!

- 전자기기 고객 평가(X), 긍정/부정 label(Y)
- 비디오 게임 고객평가(X)
- Target label은 모르지만 Source, Target 두 도메인 모두에서 label을 잘 맞추는 Classifier h를 찾고 싶음

Notations

- $X \subseteq \mathbb{R}^d$ input space, $Y = \{0, 1\}$ output space
- P_S **source** domain: distribution over $X \times Y$
  D_S marginal distribution over X
- P_T **target** domain: different distribution over $X \times Y$
  D_T marginal distribution over X
- $\mathcal{H} \subseteq Y^X$: hypothesis class

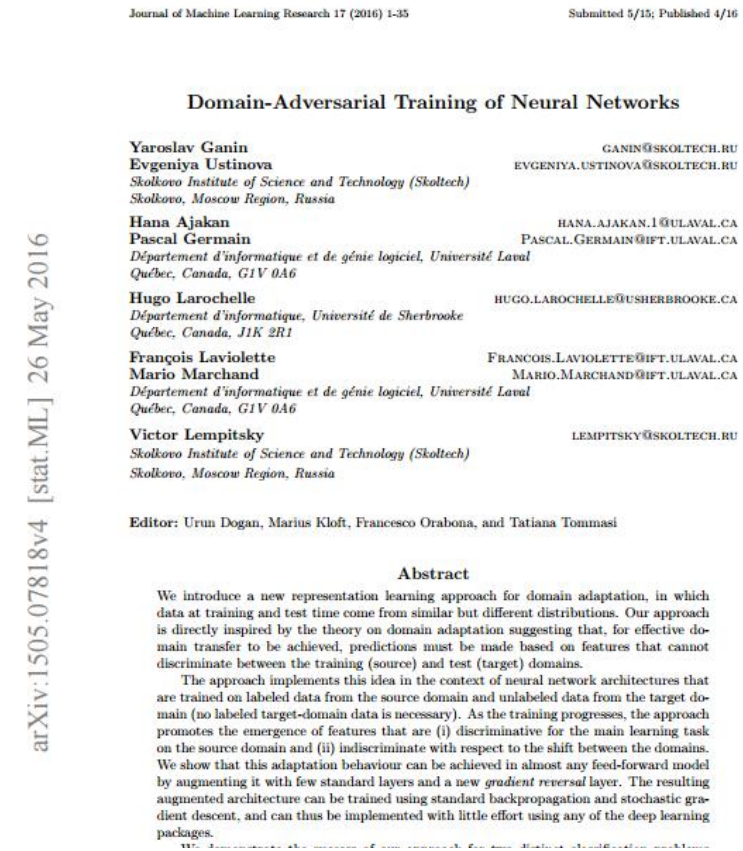
Expected error of hypothesis h: $X \to Y$

- $R_{P_S}(h) = E_{(x^s, y^s) \sim P_S} I[h(x^s) \neq y^s]$ source domain error
- $R_{P_r}(h) = E_{(x^t, y^t) \sim P_r} I[h(x^t) \neq y^t]$ target domain error

Domain Adaptation: find $h \in \mathcal{H}$ with R_{P_r} small from data $\sim D_T$ and P_S

DAHS
Data Analytics and Healthcare Systems LAB

05 Paper: DANN



Journal of Machine Learning Research 17 (2016) 1-35 Submitted 5/15; Published 4/16

Domain-Adversarial Training of Neural Networks

Yaroslav Ganin GANIN@SKOLTECH.RU
Evgeniya Ustinova EVGENIYA.USTINOVA@SKOLTECH.RU
Skolkovo Institute of Science and Technology (Skoltech)
Skolkovo, Moscow Region, Russia

Hana Ajakan HANA.AJAKAN.1@ULAVAL.CA
Pascal Germain PASCAL.GERMAIN@IFT.ULAVAL.CA
Département d'informatique et de génie logiciel, Université Laval
Québec, Canada, G1V 0A6

Hugo Larochelle HUGO.LAROCHELLE@USHERBROOKE.CA
Département d'informatique, Université de Sherbrooke
Québec, Canada, J1K 2R1

François Laviolette FRANCOIS.LAVIOLETTE@IFT.ULAVAL.CA
Mario Marchand MARIO.MARCHAND@IFT.ULAVAL.CA
Département d'informatique et de génie logiciel, Université Laval
Québec, Canada, G1V 0A6

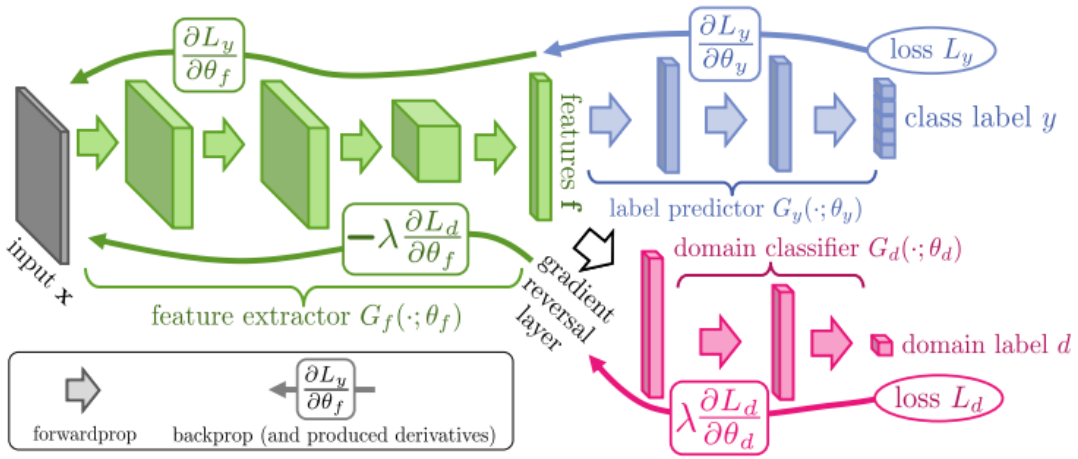
Victor Lempitsky LEMPITSKY@SKOLTECH.RU
Skolkovo Institute of Science and Technology (Skoltech)
Skolkovo, Moscow Region, Russia

Editor: Urun Dogan, Marius Kloft, Francesco Orabona, and Tatiana Tommasi

Abstract

We introduce a new representation learning approach for domain adaptation, in which data at training and test time come from similar but different distributions. Our approach is directly inspired by the theory on domain adaptation suggesting that, for effective domain transfer to be achieved, predictions must be made based on features that cannot discriminate between the training (source) and test (target) domains.

The approach implements this idea in the context of neural network architectures that are trained on labeled data from the source domain and unlabeled data from the target domain (no labeled target-domain data is necessary). As the training progresses, the approach promotes the emergence of features that are (i) discriminative for the main learning task on the source domain and (ii) indiscriminate with respect to the shift between the domains. We show that this adaptation behaviour can be achieved in almost any feed-forward model by augmenting it with few standard layers and a new *gradient reversal* layer. The resulting augmented architecture can be trained using standard backpropagation and stochastic gradient descent, and can thus be implemented with little effort using any of the deep learning packages.



[PDF] **Domain-adversarial** training of **neural networks**

Y Ganin, E Ustinova, H Ajakan, P Germain... - The journal of machine ..., 2016 - jmlr.org

... Finally, we evaluate **domain-adversarial** descriptor learning in the ... We apply **domainadversarial** learning, as we consider a ... we demonstrate that **domain-adversarial** learning can ...

☆ 저장 99 인용 7099회 인용 관련 학술자료 전체 23개의 버전 Web of Science: 2507 »

*2023/07/12

DAHS
Data Analytics and Healthcare Systems LAB

- 기존 전략은 최대한 적은 parameter로 Train error가 최소가 되는 model을 탐색하는 것

$$\underline{\epsilon_{test}} \leq \underline{\epsilon_{train}} + complexity$$

Error of unseen set　　Error of Train set

Train, Test set are same distribution !

- Train set과 Test set가 **이질적인 분포를 갖고 있는 상황**에서는 새로운 전략이 필요함
- Train error는 쉽게 줄일 수 있는 요소이므로 그대로 활용해 다음과 같은 접근이 가능

$$\underline{\epsilon_{test}} \leq \underline{\epsilon_{train}} + \ ???$$

Error of unseen set　　Error of Train set

❖ $\mathcal{H} - divergence$

- 두 확률 분포의 측정값이 벌어질 수 있는 가장 큰 값을 뜻함
- $\mathcal{H} - divergence$ 값이 크면 두 도메인 간의 차이가 클 것이라는 의미로 직관적인 이해가 가능

$$\epsilon_{test} \leq \epsilon_{train} + \mathcal{H} - divergence$$

Error of unseen set / Error of Train set

$$= \epsilon_{train} + d_{\mathcal{H}}(D_S^X, D_T^X)$$

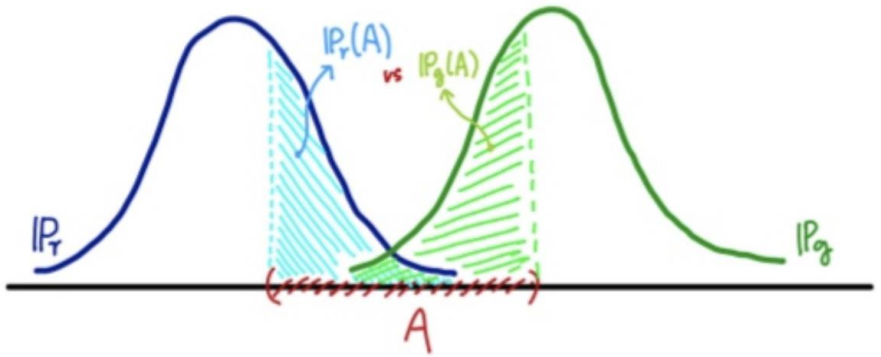
$$= \epsilon_{train} + \int |P_{r_{x\sim D_S^X}}(x) - P_{r_{x\sim D_T^X}}(x)|dx$$

$$= \epsilon_{train} + 2\sup_{\eta\in\mathcal{H}} |P_{r_{x\sim D_S^X}}(x) - P_{r_{x\sim D_T^X}}(x)|$$

데이터가 Source domain 에서 나왔을 확률

데이터가 Target domain 에서 나왔을 확률

$$\delta(\mathbb{P}_r, \mathbb{P}_g) = \sup_{A\in\Sigma} |\mathbb{P}_r(A) - \mathbb{P}_g(A)|$$



❖ $\mathcal{H} - divergence$

$$\epsilon_{test} \leq \epsilon_{train} + \mathcal{H} - divergence$$
$$= \epsilon_{train} + \ d_{\mathcal{H}}(D_S^X, D_T^X)$$
$$= \epsilon_{train} + \int |P_{r_{x\sim D_S^X}}(x) - P_{r_{x\sim D_T^X}}(x)|dx$$
$$= \epsilon_{train} + \ 2\sup_{\eta\in\mathcal{H}} |P_{r_{x\sim D_S^X}}(x) - P_{r_{x\sim D_T^X}}(x)|$$

현실적으로 Classifier 집합인 가설공간 $\mathcal{H}$의 **모든 경우의 수를 고려할 수 없기** 때문에 위 식의 해를 구할 수 없음

$\mathcal{H}$의 capacity에 따라 $\mathcal{H}$-divergence의 값이 계속 바뀜(linear, polynomial …)

That is, the $\mathcal{H}$-divergence relies on the capacity of the hypothesis class $\mathcal{H}$ to distinguish between examples generated by $\mathcal{D}_{\mathrm{S}}^X$ from examples generated by $\mathcal{D}_{\mathrm{T}}^X$. Ben-David et al. (2006, 2010) proved that, for a symmetric hypothesis class $\mathcal{H}$, one can compute the *empirical* $\mathcal{H}$*-divergence* between two samples $S \sim (\mathcal{D}_{\mathrm{S}}^X)^n$ and $T \sim (\mathcal{D}_{\mathrm{T}}^X)^{n'}$ by computing

$\mathcal{H} - divergence$

$$\epsilon_{test} \leq \epsilon_{train} + \mathcal{H} - divergence$$
$$= \epsilon_{train} + \ d_{\mathcal{H}}(D_S^X, D_T^X)$$
$$= \epsilon_{train} + \int |P_{r_{x\sim D_S^X}}(x) - P_{r_{x\sim D_T^X}}(x)|dx$$
$$= \epsilon_{train} + \ 2\sup_{\eta\in\mathcal{H}} |P_{r_{x\sim D_S^X}}(x) - P_{r_{x\sim D_T^X}}(x)|$$

$\mathcal{A} - distance$

$$\epsilon_{test} \leq \epsilon_{train} + \mathcal{H} - divergence$$
$$= \epsilon_{train} + \ d_{\mathcal{H}}(D_S^X, D_T^X)$$
$$= \epsilon_{train} + \int |P_{r_{x\sim D_S^X}}(x) - P_{r_{x\sim D_T^X}}(x)|dx$$
$$\leq \epsilon_{train} + 2\sup_{\mathcal{A}\in\mathcal{H}} |P_{r_{x\sim D_S^X}}(A) - P_{r_{x\sim D_T^X}}(A)| + \lambda + o(\sqrt{\frac{dlog\left(\frac{m'}{d}\right) + log\left(\frac{1}{\delta}\right)}{m'}})$$

따라서 가설공간 $\mathcal{H}$를 Neural Network로 표현할 수 있는 함수 공간 $\mathcal{A}$로 제한하면
$\mathcal{H} - divergence$에 최대한 approximate 할 수 있음

함수 공간 제한으로 인해 생기는 차이는 λ 와 model complexity

$$\lambda = \epsilon_{D_s}(h^*) + \epsilon_{D_t}(h^*)$$
$$h^* = argmin_{h\in\mathcal{H}}\ \epsilon_{D_s}(h) + \epsilon_{D_t}(h)$$ (Ground Truth Classifier)

DANN

$$\epsilon_{test} \leq \epsilon_{train} + \mathcal{H} - divergence$$

$$= \epsilon_{train} + d_{\mathcal{H}}(D_S^X, D_T^X)$$

$$= \epsilon_{train} + \int |P_{r_{x\sim D_S^X}}(x) - P_{r_{x\sim D_T^X}}(x)|dx$$

$$\leq \epsilon_{train} + 2\sup_{A\in\mathcal{H}} |P_{r_{x\sim D_S^X}}(A) - P_{r_{x\sim D_T^X}}(A)| + \lambda + o\left(\sqrt{\frac{dlog\left(\frac{m'}{d}\right) + log\left(\frac{1}{\delta}\right)}{m'}}\right)$$

구할 수 없는 값

Neural Net 학습을 통해 조절 가능한 Component

h 가 정해지면 항상 fix인 값

따라서 Source 분포와 Target 분포의 차이를 줄이면서
(= 모델이 도메인 간의 차이를 구분하지 못하게 하면서)
(= 모델이 domain에 independent한 feature를 학습하게 하면서)

Train error(Source classification error)를 줄인다면

Domain adaptation이 가능하다는 것이 본 논문의 논리

DANN

$$\epsilon_{test} \le \epsilon_{train} + \mathcal{H} - divergence$$

$$= \epsilon_{train} + d_{\mathcal{H}}(D_S^X, D_T^X)$$

$$= \epsilon_{train} + \int |P_{r_{x\sim D_S^X}}(x) - P_{r_{x\sim D_T^X}}(x)|dx$$

$$\le \epsilon_{train} + \underbrace{2\sup_{\mathcal{A}\in\mathcal{H}} |P_{r_{x\sim D_S^X}}(A) - P_{r_{x\sim D_T^X}}(A)|}_{\hat{d}_{\mathcal{A}}} + \lambda + o\left(\sqrt{\frac{dlog\left(\frac{m'}{d}\right) + log\left(\frac{1}{\delta}\right)}{m'}}\right)$$

A-distance는 함수 공간을 제한하여 구한 값이기 때문에 우리가 정한 h를 통해 얻은 entropy ϵ 를 사용하여 다음과 같이 표현 가능

$$\hat{d}_{\mathcal{A}} = 2(1 - 2\epsilon)$$

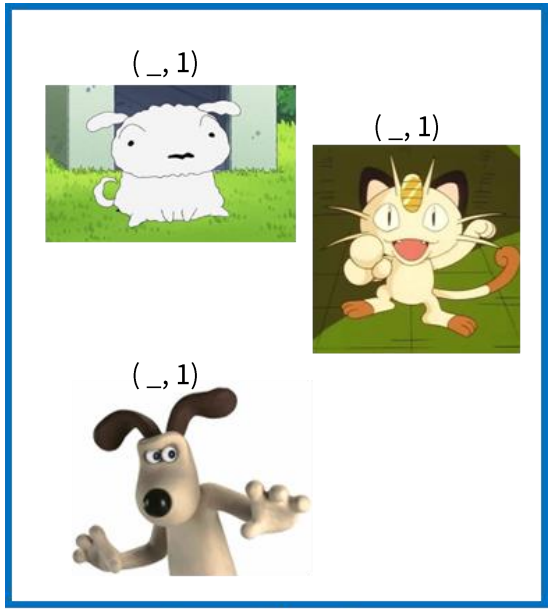
두 분포간 차이를 모델이 구분하지 못한다
$\epsilon = \frac{1}{2}$, distance는 0

두 분포간 차이를 모델이 잘 구분한다
$\epsilon = 0\ or\ 1$, distance는 2

❖ Domain – Adversarial Training of Neural Networks

- Label이 존재하는 source domain / Label이 존재하지 않는 Target domain data가 있을 때
- 추가적으로 Source / Target domain 을 구분하는 작업을 진행 (Source = 0, Target = 1) ➡ Distance 계산에 사용

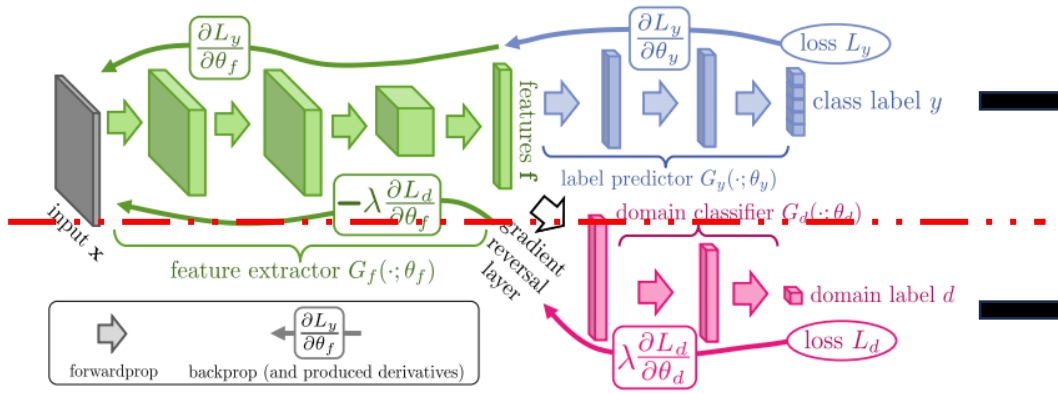
Source domain



Target domain

❖ Domain – Adversarial Training of Neural Networks

$$E(\mathbf{W}, \mathbf{V}, \mathbf{b}, \mathbf{c}, \mathbf{u}, z) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_y^i(\mathbf{W}, \mathbf{b}, \mathbf{V}, \mathbf{c}) - \lambda\left(\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_d^i(\mathbf{W}, \mathbf{b}, \mathbf{u}, z) + \frac{1}{n'}\sum_{i=n+1}^{N}\mathcal{L}_d^i(\mathbf{W}, \mathbf{b}, \mathbf{u}, z)\right)$$



Source domain 의 데이터를 잘 분류하도록 학습(Cross-entropy minimize)

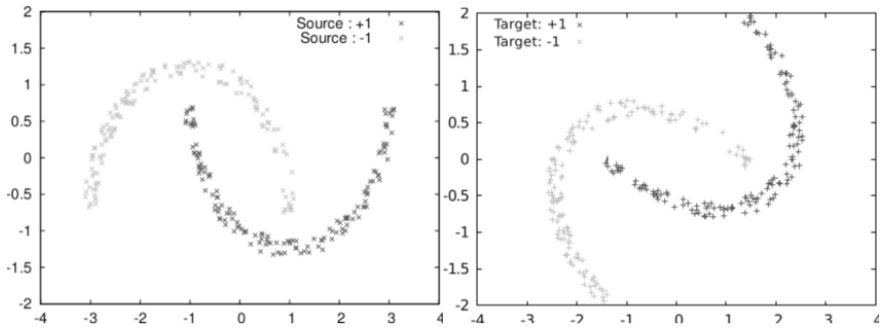
Source 와 Target을 잘 분류하도록 학습하다가(distance maximize)
Gradient에 – 를 붙여 두 도메인을 구분하지 못하도록 업데이트(distance minimize)

즉, 제안 모델은 도메인 간의 차이를 줄이는 Inter representation을 통해 general한 feature를 생성하고,

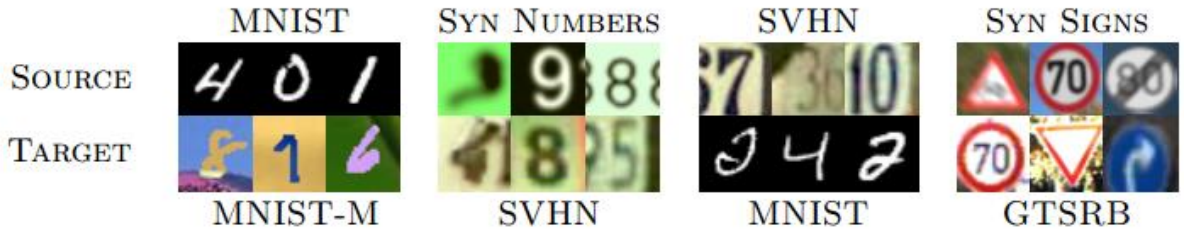
이를 바탕으로 Source domain에서의 Classification error를 줄여 Domain adaptation을 달성할 수 있음

❖ Domain – Adversarial Training of Neural Networks

- Tabular data) **The inter-twining moons**: 2차원에 plot 했을 때 두 개의 달이 꼬리를 물고 있는 형태를 나타내는 데이터셋
  - Source data는 본 데이터 셋을 그대로 활용하고, Target data는 이를 35°회전하여 생성

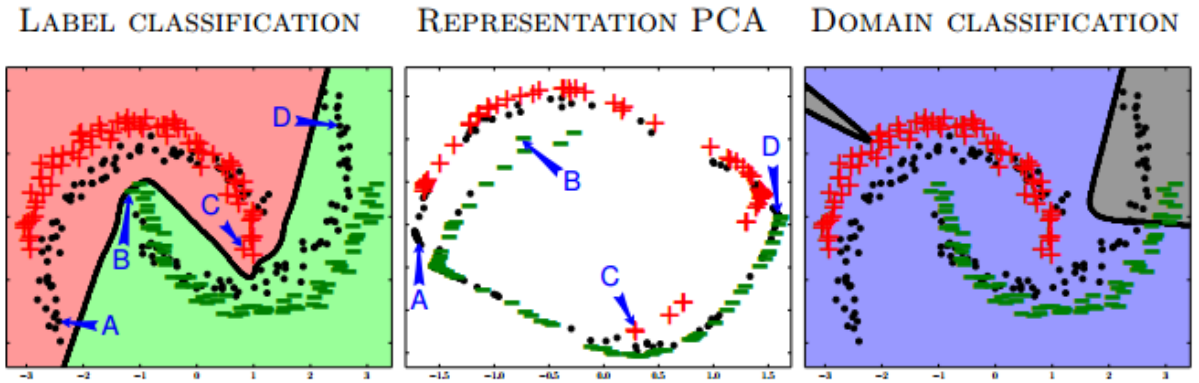


- Image data) **(MNIST→MNIST-M), (Synthetic numbers→SVHN), (Synthetic Signs→GTSRB)**
  - Source domain, Target domain Image pair



❖ Domain – Adversarial Training of Neural Networks

- + (source, label = 1), - (source, label = 0), black dot (target, label = unknown)



➢ **Label Classification**: Source data 를 잘 분류 해냄을 알 수 있음

➢ **Representation PCA**: Latent space 상에서 Target data가 Source data 가 거의 동일하게 겹쳐 있음(도메인 간 구분 불가)
3 차원 공간으로 생각했을 때, 같은 Class로 예상 되는 A와 C(B와 D)는 같은 Space에 존재

❖ Domain – Adversarial Training of Neural Networks

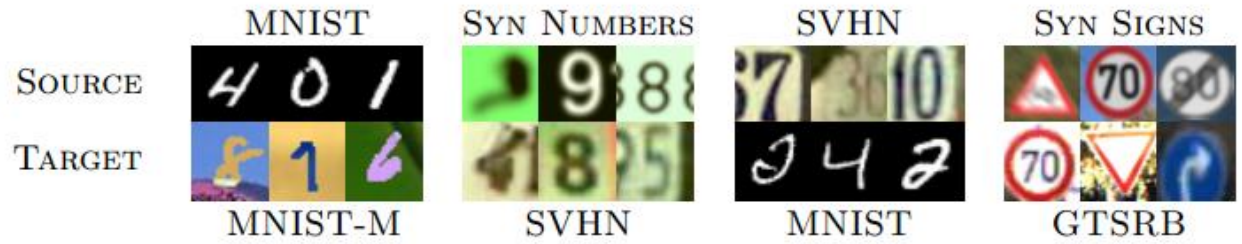


Figure 6: Examples of domain pairs used in the experiments. See Section 5.2.4 for details.

METHOD	SOURCE	MNIST	SYN NUMBERS	SVHN	SYN SIGNS
	TARGET	MNIST-M	SVHN	MNIST	GTSRB
SOURCE ONLY		.5225	.8674	.5490	.7900
SA (Fernando et al., 2013)		.5690 (4.1%)	.8644 (−5.5%)	.5932 (9.9%)	.8165 (12.7%)
DANN		**.7666** (52.9%)	**.9109** (79.7%)	**.7385** (42.6%)	**.8865** (46.4%)
TRAIN ON TARGET		.9596	.9220	.9942	.9980

❖ 본 논문을 한마디로 정의하면?

- ✓ Target domain dataset 이 unlabeled 인 상황에서 Feed forward Neural network의 Domain adaptation을 위한 새로운 접근 방식 제안

❖ 본 논문의 문제점 혹은 부족한 부분은?

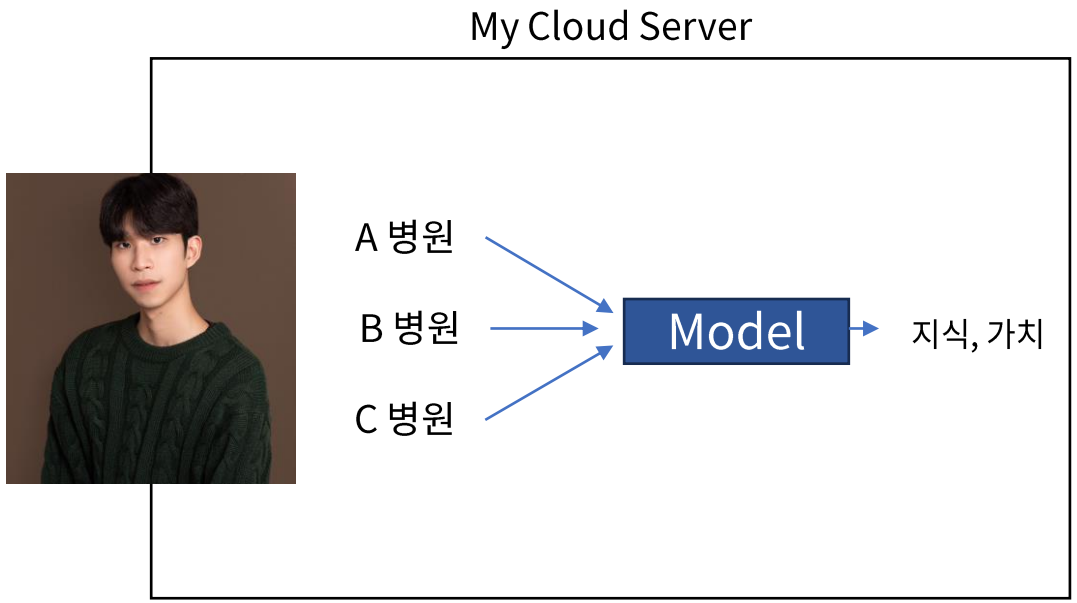
- ✓ 제안한 아키텍처로 학습된 모델이 궁극적으로 Target domain data 도 잘 분류하는지 제시하지 않음
- ✓ $\mathcal{H} - divergence$ **사용 이유**: KL-divergence, Earth mover distance와 같은 분포 사이의 유사성을 측정하는 기법은 다양하기 때문에 $\mathcal{H} - divergence$ 를 사용한 저자의 이유가 궁금함
- ✓ **Data leakage와 어떤 차이**가 존재하는가: Target data를 모델 학습에 사용했다는 관점에서 data leakage로 볼 수 있음. Generalization을 위해서 외부 데이터 정보의 사용은 필수적인가?

❖ 본 논문과 관련된 본인의 아이디어는?

- ✓ 개선하기
  - 적대적 구조의 두 Classifier 사용으로 인한 비효율성 해결 방안 고민(1 가지의 classifier 만 사용해서 연산량을 줄이고, 비슷한 성능을 낼 수는 없을까?)
- ✓ 적용하기
  - Early prediction disease in the intensive care unit using Machine Learning
  - 서로 다른 Time resolution 을 가지고 수집된 데이터도 도메인이 다르다고 볼 수 있다면
    - ✓ Circulatory failure: MIMIC, eICU 를 활용한 연구에서 domain adaptation 실험을 진행하고 성능을 제시할 수 있을 것 같음
  - Model이 domain 에 bias 가 없다는 것을 어필할 수 있을 것

❖ 일반적인 Work Flow

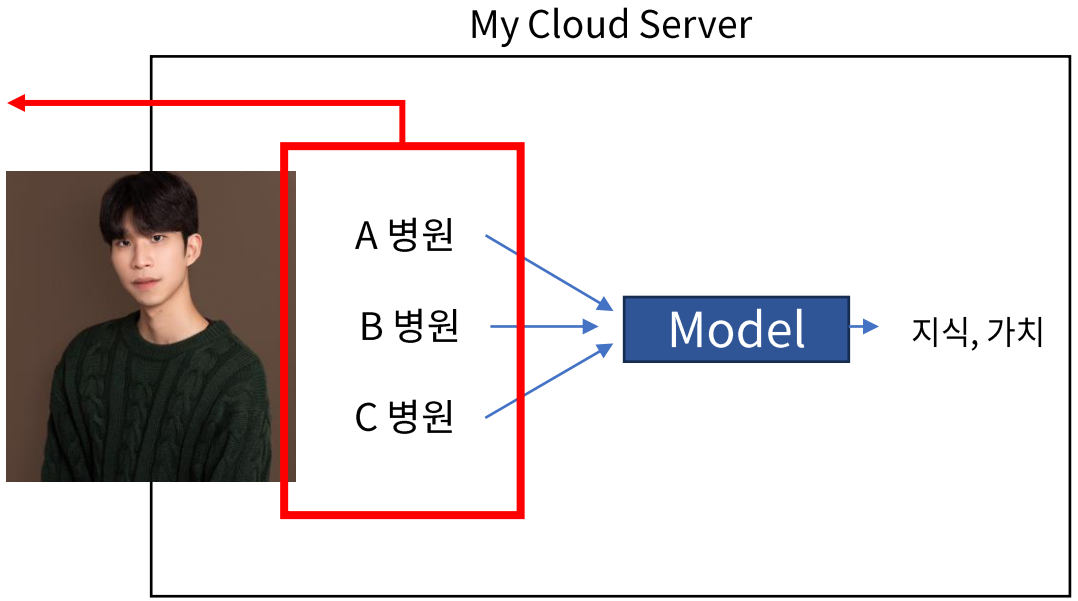
- 데이터과학자는 개인 또는 회사(기관) 클라우드에 데이터를 수집하고 이를 분석하여 지식과 가치를 창출함
- 이는 효율성, 분석 시간, 비용, 개인정보 측면에서 많은 제약이 따름
- 특히, **의료데이터**는 데이터 격리(기관 별)와 개인정보 문제로 데이터 접근이 쉽지 않음
- 충분한 데이터에 접근하지 못한다면 ML은 전체적인 잠재력을 발휘하지 못해 임상 실무로의 전환이 어려움



❖ 일반적인 Work Flow

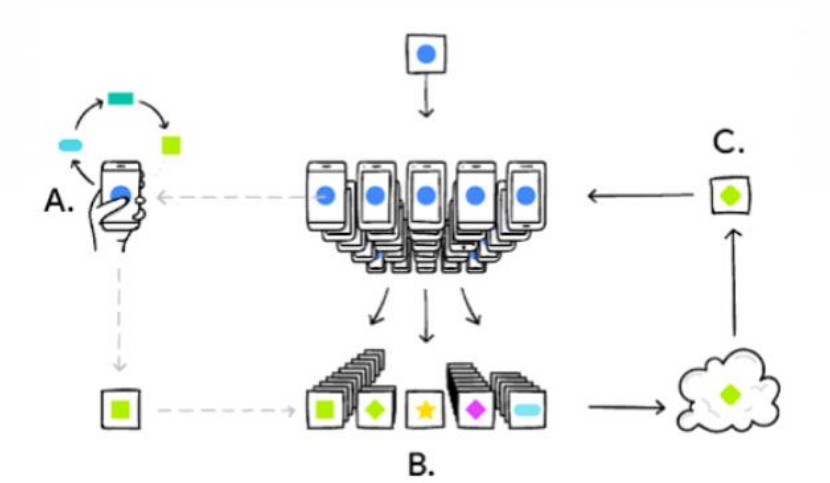
- 데이터과학자는 개인 또는 회사(기관) 클라우드에 데이터를 수집하고 이를 분석하여 지식과 가치를 창출함
- 이는 효율성, 분석 시간, 비용, 개인정보 측면에서 많은 제약이 따름
- 특히, **의료데이터**는 데이터 격리(기관 별)와 개인정보 문제로 데이터 접근이 쉽지 않음
- 충분한 데이터에 접근하지 못한다면 ML은 전체적인 잠재력을 발휘하지 못해 임상 실무로의 전환이 어려움

데이터 수집에 제한적



❖ 연합학습(FL)

- 기기나 기관 등이 여러 위치에서 분산 저장된 데이터를 직접 공유하지 않고 서로 협력하며 인공지능 모델을 학습하는 분산형 학습 기법

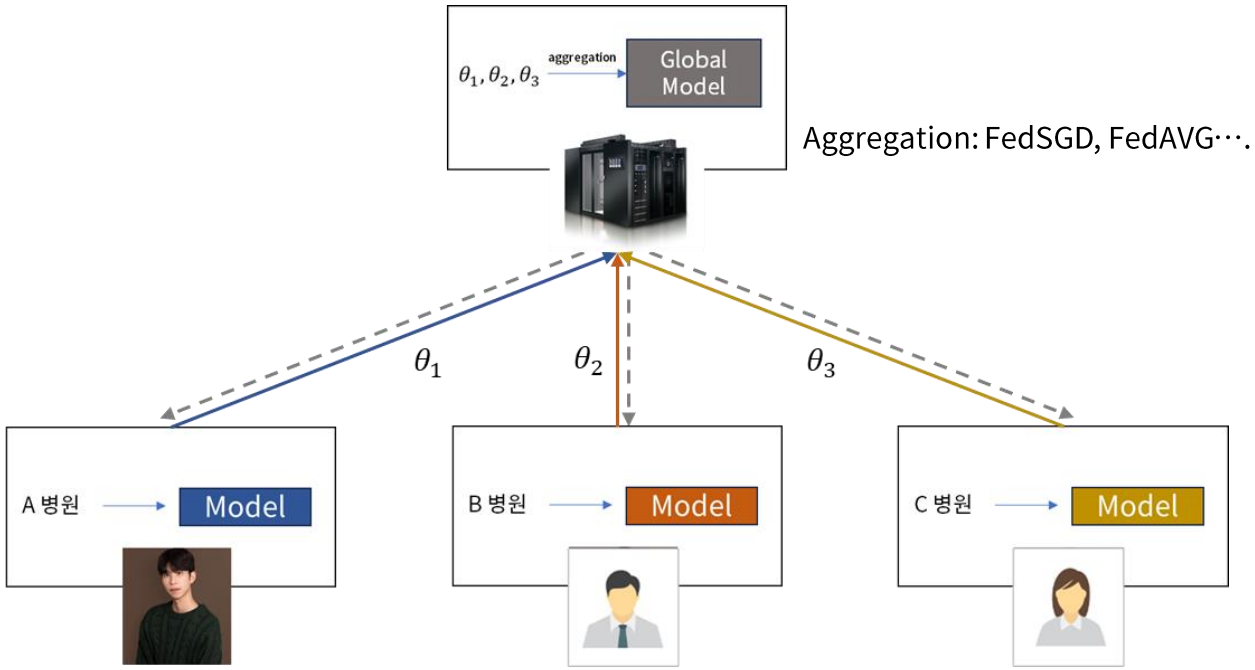


<중앙 서버가 존재하는 연합 학습 구조>

(A)개인 클라우드에서 인공지능 모델이 저장된 데이터에 맞게 학습
(B)에서 다양한 사용자의 학습 파라미터가 글로벌 모델이 저장되어 있는 중앙 서버로 전송
(C)에서 글로벌 모델의 학습 파라미터를 개인의 모바일로 전송하며 전체 과정을 반복

❖ 연합학습(FL)

- 연합학습 접근 방식은 개별 클라이언트에서 데이터를 집계하는 대신 모델 학습을 수행하고 업데이트된 **모델의 파라미터들만** 중앙 서버에 전달되도록 설계됨
- 중앙 서버에서는 대규모 클라이언트들로부터 학습된 로컬 모델 파라미터들을 집계하고 이를 평준화 하여 글로벌 모델을 학습



*예시는 서버-클라이언트 연합학습 구조

❖ 연합학습(FL)

- 장점:
  1. 개인정보가 보호되어야 하는 상황에서 데이터 유출 없이 학습이 가능
  2. 수만 개의 로컬 디바이스의 데이터를 모두 중앙 서버로 전송할 필요 없이 로컬 모델의 업데이트 정보만을 주고 받아 비용을 절감 가능
  3. ML model의 일반화 성능을 높일 수 있음

- 단점:
  1. 연합학습은 수만 개의 장치가 참여하기 때문에 통신에 과부하가 발생할 수 있음
  2. 연합학습 모델은 통합 클라우드에 고용량으로 저장되어 있어 단일 장치에 적용할 수 없음
  3. 연합학습 모델의 추론은 네트워크의 지연으로 실시간 요구에 응하기 어려움

❖ 연합학습(FL)

- 장점:
  1. 개인정보가 보호되어야 하는 상황에서 데이터 유출 없이 학습이 가능
  2. 수만 개의 로컬 디바이스의 데이터를 모두 중앙 서버로 전송할 필요 없이 로컬 모델의 업데이트 정보만을 주고 받아 비용을 절감 가능
  3. ML model의 일반화 성능을 높일 수 있음

- 단점:
  1. 연합학습은 수만 개의 장치가 참여하기 때문에 통신에 과부하가 발생할 수 있음
  2. 연합학습 모델은 통합 클라우드에 고용량으로 저장되어 있어 단일 장치에 적용할 수 없음
  3. 연합학습 모델의 추론은 네트워크의 지연으로 실시간 요구에 응하기 어려움

연합학습을 제대로 구현하기 위해서는 모델을 공유를 위한 **기관 별 협조**와
네트워크 운용관리가 가능한 **Field의 기술적 지원**,
서버를 관리하고, 모델을 개발할 수 있는 **데이터 엔지니어, 분석가**가 필요

08 Discussion 2

- Domain Adaptation 과 Federated Learning 은 모두 ML의 generalization을 달성할 수 있는 학습 기법으로 이해 가능
- 데이터 소싱 및 기관의 협조가 불가능한 상황에서는 DA, FL 모두 실현할 수 없음 → 최악의 상황, 일자리를 잃을 수 있음
- Federated Learning은 데이터 소싱의 제약이 없더라도, 높은 수준을 가진 인력과 자본 등의 부가적 요소가 필수적임(개인이 Cover할 수 있는 범위를 지나침)
- 그에 반해 Domain Adaptation 은 데이터 소싱 제약만 없어도, FL에 비해 비교적 쉬운 구현이 가능함

➢ 어쩌면 이 두가지 방식은 **목적**이 다를 수 있음

- ✓ Federated Learning: ML의 일반화 보다는 데이터의 **직접적인 유출 없이 파라미터 전송만으로 모델을 구축**하는 것
  - ✓ 기관 또는 회사의 소중한 자산인 데이터를 지키면서 거대한 모델을 만드는 것이 주 목적
- ✓ Domain Adaptation: **Labeling이 되어 있지 않은** 세상에 존재하는 다양한 **데이터만을 가지고 모델을 일반화** 하는 것
  - ✓ 사실 labeling이 되어 있는 Target domain dataset이 있고 local에 가져올 수 있다면 Fine tune 만으로 Domain adaptation을 이룰 수 있을 것으로 예상

09 Reference

[1] Ganin, Yaroslav, et al. "Domain-adversarial training of neural networks." *The journal of machine learning research* 17.1 (2016): 2096-2030.

[2] Rieke, Nicola, et al. "The future of digital health with federated learning." *NPJ digital medicine* 3.1 (2020): 119.

[3] 한국지능정보사회진흥원(NIA), “산업분야에서 연합학습 동향 및 시사점 "

Thank you

Question?