딥러닝 세미나 Season #7

Zero-Shot Learning by Convex Combination of Semantic Embeddings

2014, *ICLR*, M. Norouzi et al.

https://arxiv.org/pdf/1312.5650.pdf

한양대학교
컴퓨터 소프트웨어학과
인공지능 연구실
조건희

Introduction

〈 Classic Machine Learning Approach 〉

10가지 클래스를 분류하는 image classifier 에 대해 생각해보자

〈 Classic Machine Learning Approach 〉

10가지 클래스를 분류하는 image classifier 에 대해 생각해보자

이미지를 처리해야 하므로

ResNet, VGG, Inception 등을 이용하면 되겠지?

〈 Classic Machine Learning Approach 〉

10가지 클래스를 분류하는 image classifier 에 대해 생각해보자

이미지를 처리해야 하므로

ResNet, VGG, Inception 등을 이용하면 되겠지?

학습에 이용할 데이터셋은

10가지 클래스에 대한 이미지를 각각 준비하면 되겠다

〈 Classic Machine Learning Approach 〉

10가지 클래스를 분류하는 image classifier 에 대해 생각해보자

이미지를 처리해야 하므로

ResNet, VGG, Inception 등을 이용하면 되겠지?

학습에 이용할 데이터셋은

10가지 클래스에 대한 이미지를 각각 준비하면 되겠다

이제 학습을 돌려보자!

〈 Classic Machine Learning Approach 〉

이제 클래스를 좀더 세분화해서

10,000가지 클래스를 분류하는 image classifier 에 대해 생각해보자

Introduction

〈 Classic Machine Learning Approach 〉

이제 클래스를 좀더 세분화해서

10,000가지 클래스를 분류하는 image classifier 에 대해 생각해보자

ResNet, VGG, Inception 등은 그대로 이용하면 될 것 같은데,

〈 Classic Machine Learning Approach 〉

이제 클래스를 좀더 세분화해서

10,000가지 클래스를 분류하는 image classifier 에 대해 생각해보자

ResNet, VGG, Inception 등은 그대로 이용하면 될 것 같은데,

10,000가지 클래스의 이미지 데이터는 어떻게 구하지?

〈 Classic Machine Learning Approach 〉

이제 클래스를 좀더 세분화해서

10,000가지 클래스를 분류하는 image classifier 에 대해 생각해보자

ResNet, VGG, Inception 등은 그대로 이용하면 될 것 같은데,

10,000가지 클래스의 이미지 데이터는 어떻게 구하지?

게다가 열심히 학습했던 10가지 클래스 분류 모델은

재활용이 안되네 ㅠㅠ

Introduction

정리하자면,

n-way classification 문제에서 n이 아주 클 경우에는

기존 기계학습 방법을 그대로 적용하기 어려움

Introduction

정리하자면,

n-way classification 문제에서 n이 아주 클 경우에는

기존 기계학습 방법을 그대로 적용하기 어려움

그래서

이미지의 semantic embedding을 이용하는 방법이 나옴

바로 Zero-Shot Learning

Zero-Shot Learning

$$\mathcal{D}_0 \equiv \{(\mathbb{x}_i, y_i)\}_{i=1}^{m}$$
데이터셋 크기 : m

$$\mathbb{x}_i \in \mathbb{R}^p$$
p 차원 벡터

$$y_i \in \mathcal{Y}_0 \equiv \underbrace{\{1, \ldots, n_0\}}_{\text{클래스: } n_0\text{개}}$$

Zero-Shot Learning

$\mathbb{x}_i$
labeled
training
image

$$\mathcal{D}_0 \equiv \{(\mathbb{x}_i, y_i)\}_{i=1}^{m}$$ 데이터셋 크기 : $\boldsymbol{m}$

$$\mathbb{x}_i \in \mathbb{R}^p$$ $\boldsymbol{p}$ 차원 벡터

$$y_i \in \mathcal{Y}_0 \equiv \underbrace{\{1, \dots, n_0\}}_{\text{클래스: } n_0 \text{개}}$$

$\mathbb{x}'_j$
test
image

$$\mathcal{D}_1 \equiv \{(\mathbb{x}'_j, y'_j)\}_{j=1}^{m'}$$ 데이터셋 크기 : $\boldsymbol{m'}$

$$\mathbb{x}'_j \in \mathbb{R}^p$$ $\boldsymbol{p}$ 차원 벡터

$$y'_j \in \mathcal{Y}_1 \equiv \underbrace{\{n_0 + 1, \dots, n_0 + n_1\}}_{\text{클래스: } n_1 \text{개}}$$

Zero-Shot Learning

〈 제로샷 러닝의 목적 〉

$\mathcal{D}_0$으로 학습한 *classifier*가

$\mathcal{D}_1$에도 잘 적용되도록 하는 것

Zero-Shot Learning

〈 제로샷 러닝의 목적 〉

$\mathcal{D}_0$으로 학습한 *classifier*가

$\mathcal{D}_1$에도 잘 적용되도록 하는 것

$\mathcal{Y}_0 \cap \mathcal{Y}_1 = \emptyset$ 이므로

다른 정보가 주어지지 않으면 **불가능**

Zero-Shot Learning

〈 제로샷 러닝의 목적 〉

$\mathcal{D}_0$으로 학습한 *classifier*가

$\mathcal{D}_1$에도 잘 적용되도록 하는 것

$\mathcal{Y}_0 \cap \mathcal{Y}_1 = \emptyset$ 이므로

다른 정보가 주어지지 않으면 **불가능**

모든 label y $(1 \le y \le n_0 + n_1)$ 에 대해서

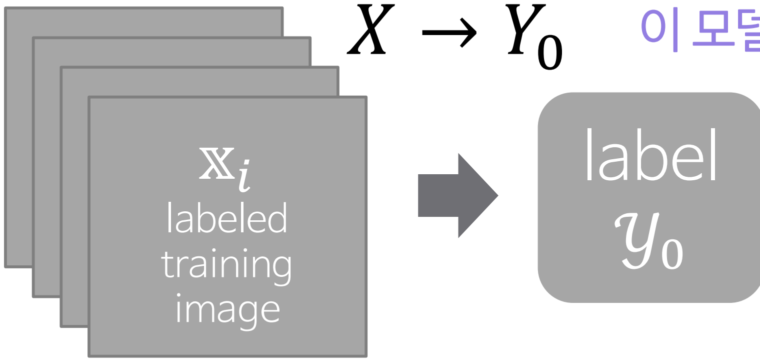
semantic embedding

$s(y) \in \mathcal{S} \equiv \mathbb{R}^q$ ***q* 차원 벡터**

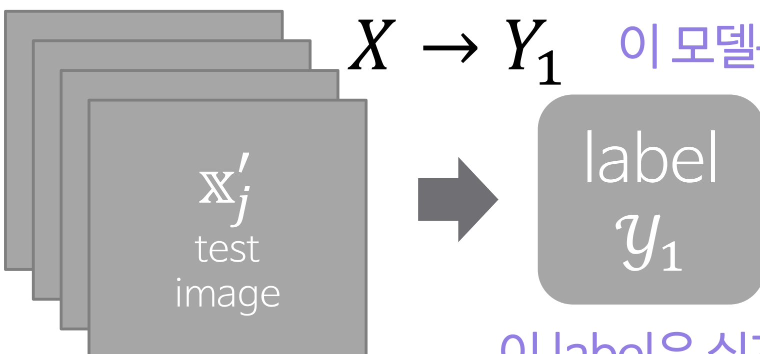
가 존재하면 **가능**

Zero-Shot Learning

$X \rightarrow Y_0$ 이 모델은 학습 가능



$X \rightarrow Y_1$ 이 모델은 학습 불가능



이 label은 실제로는 존재하지 않음

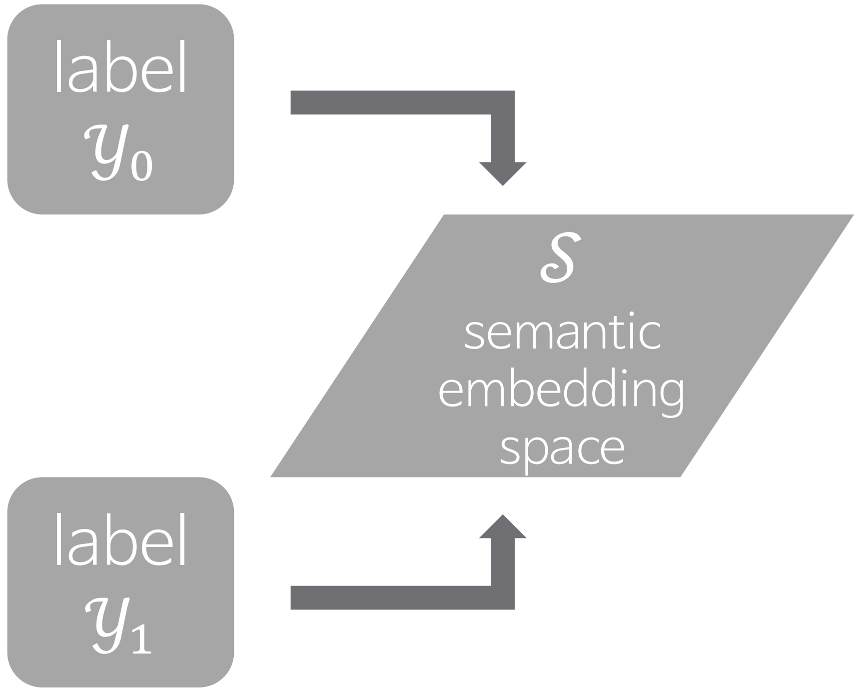
Zero-Shot Learning

Word2Vec

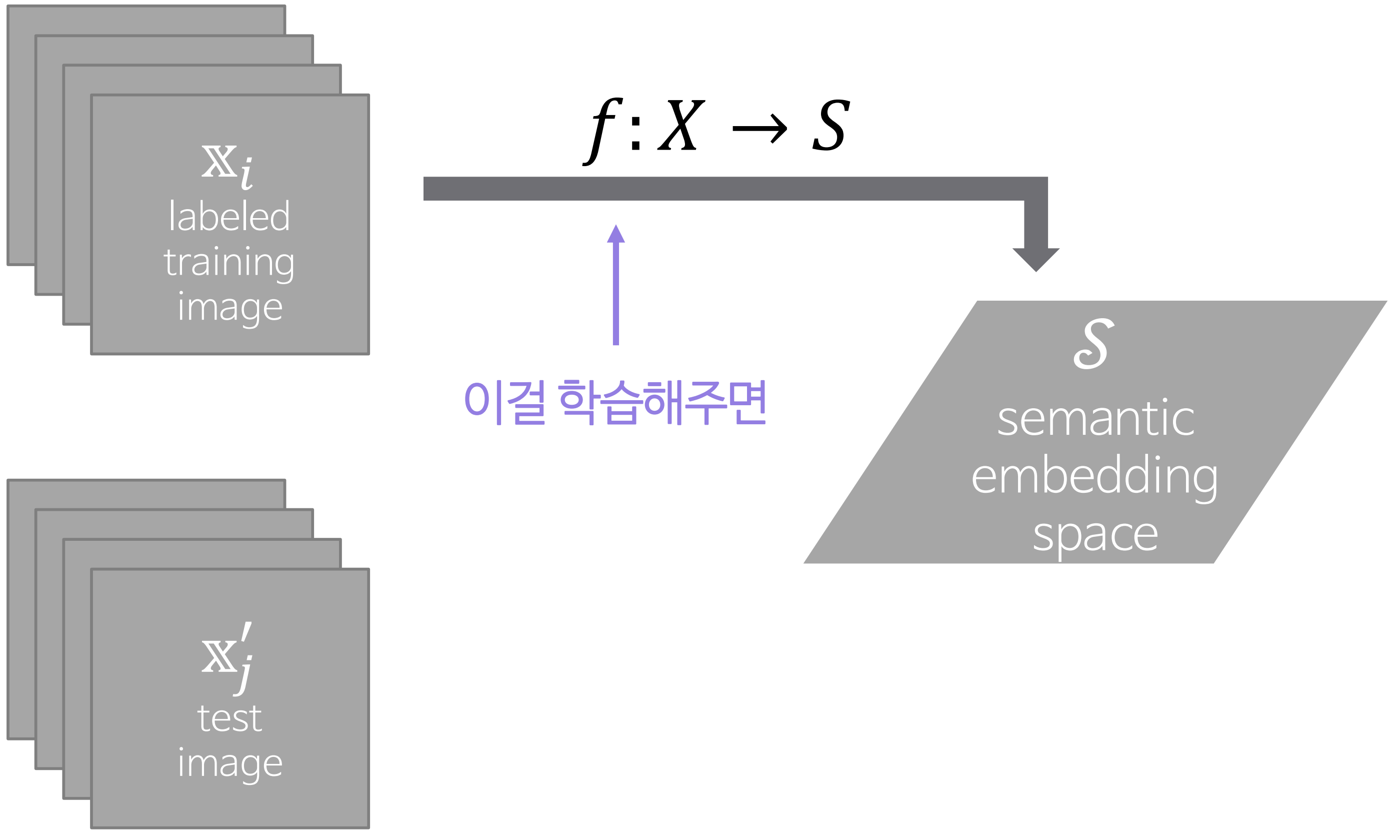
단어의 의미가 비슷하면

semantic embedding space 상의

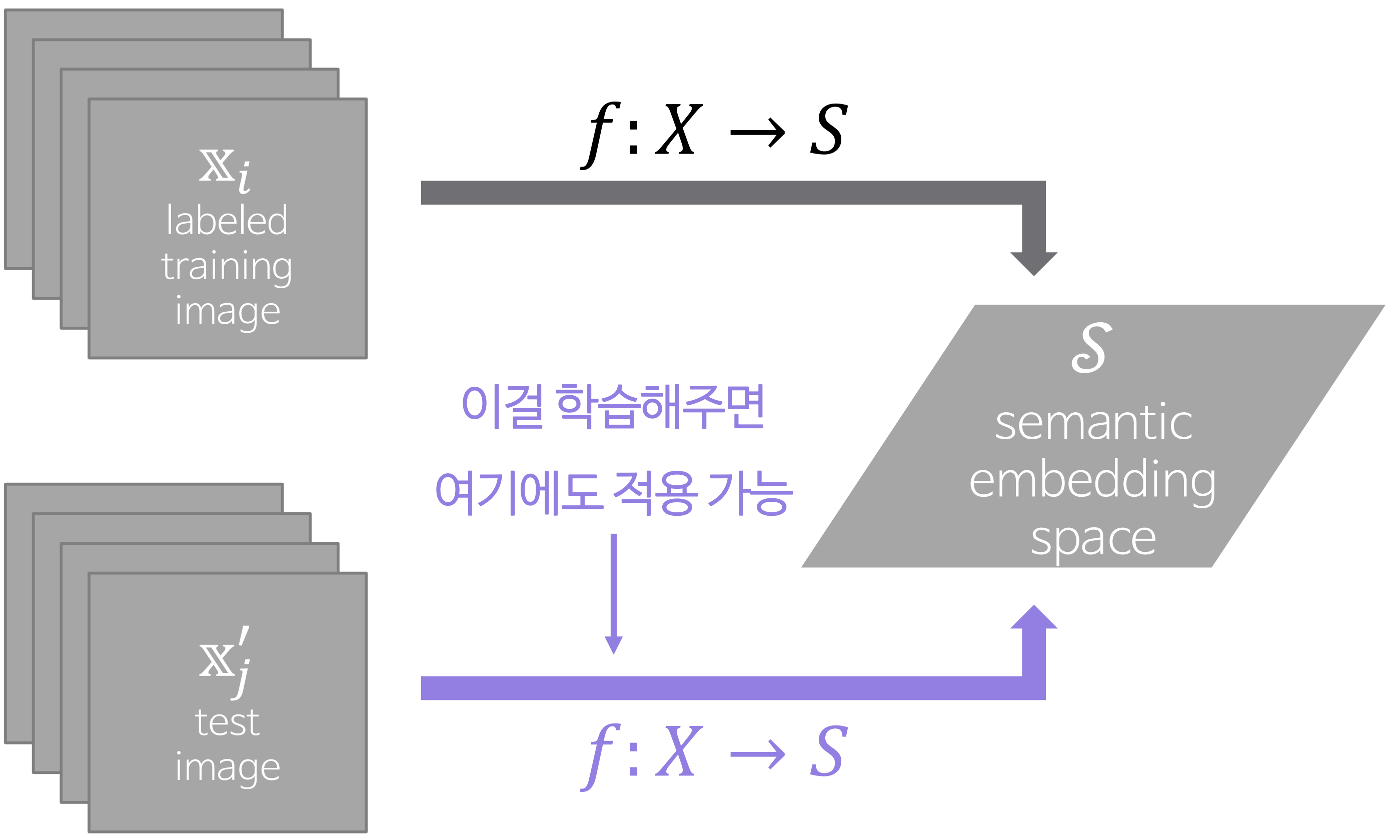
벡터 좌표도 비슷하도록 학습함



Zero-Shot Learning

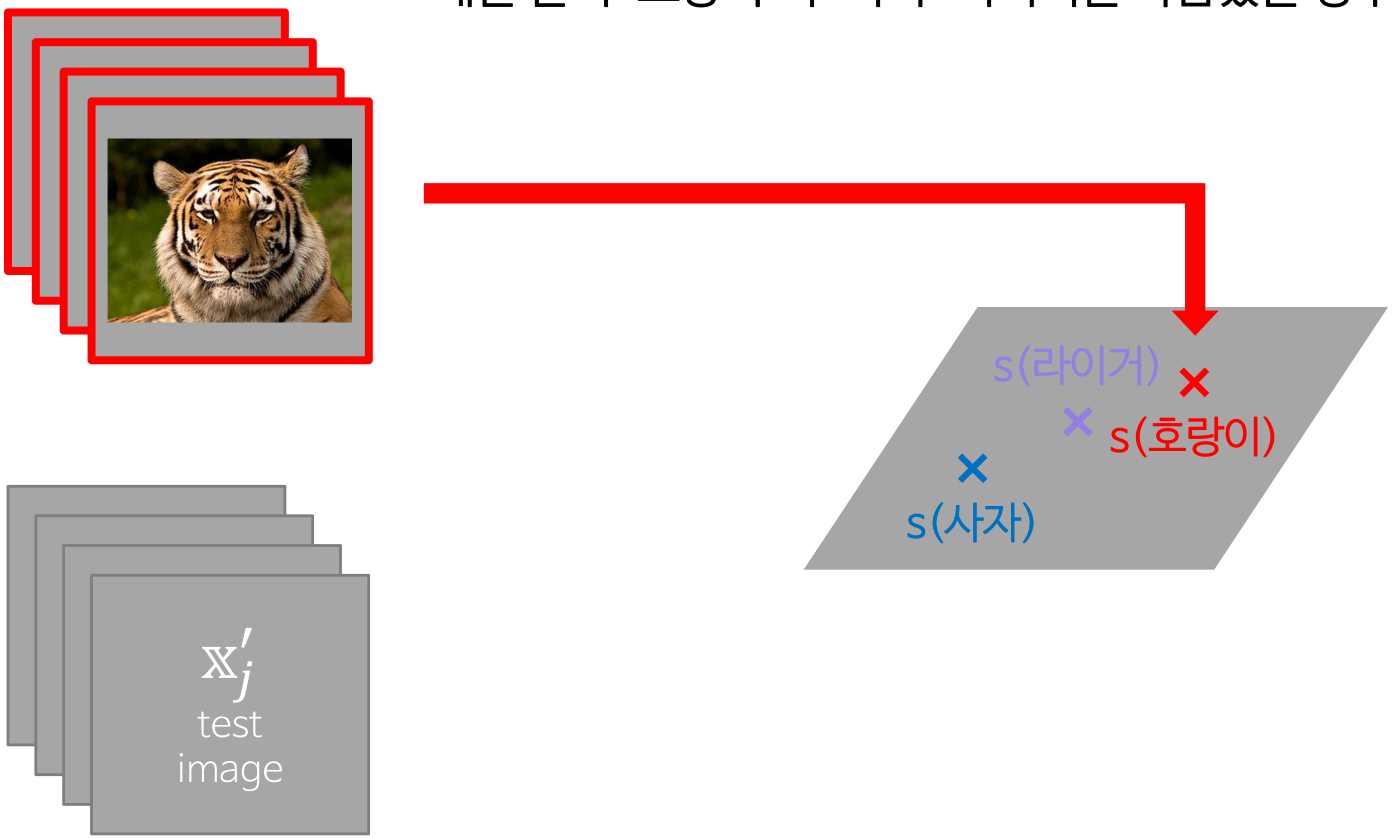


Zero-Shot Learning



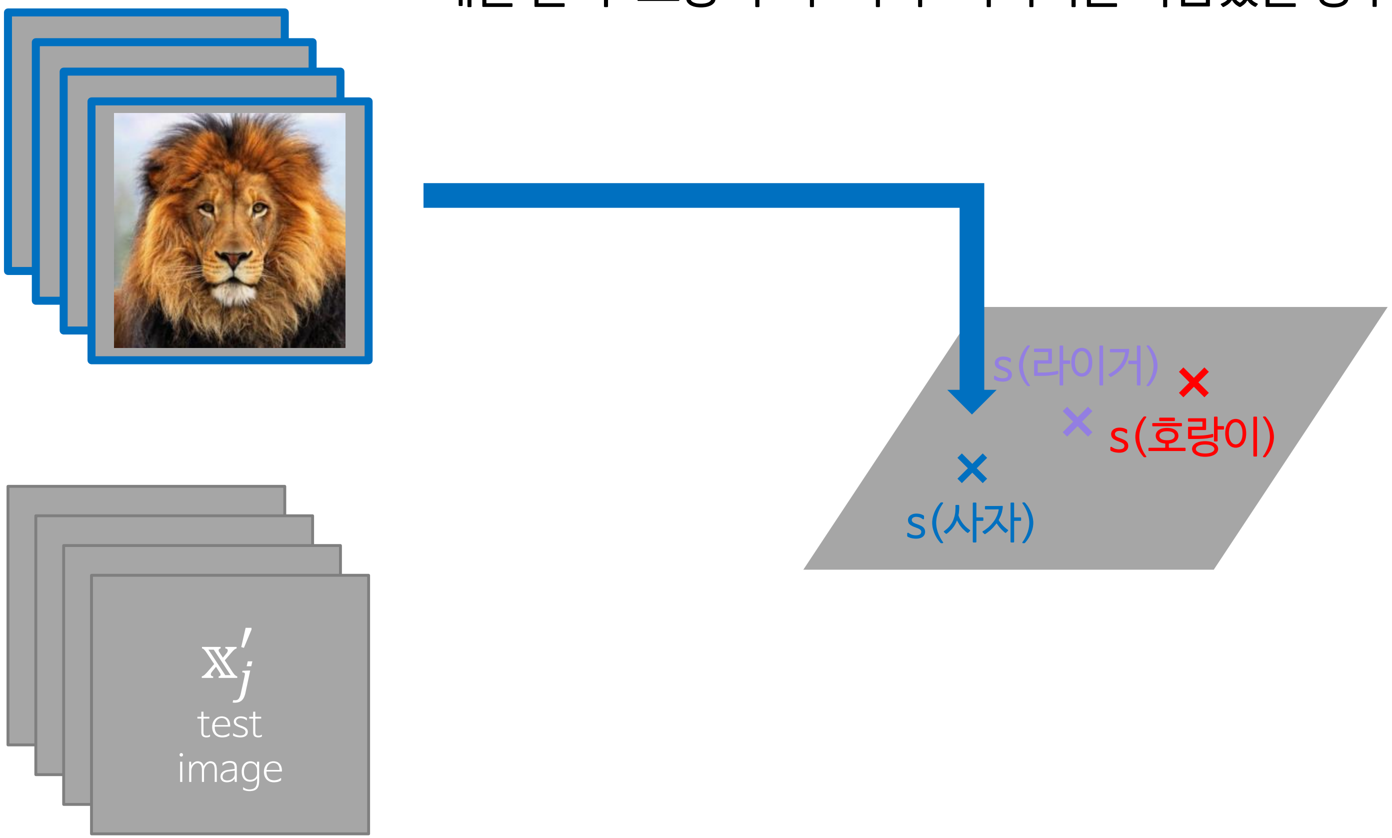
Zero-Shot Learning

예를 들어 ‘호랑이’와 ‘사자’ 이미지를 학습했을 경우



Zero-Shot Learning

예를 들어 '호랑이'와 '사자' 이미지를 학습했을 경우



Zero-Shot Learning

예를 들어 ‘호랑이’와 ‘사자’ 이미지를 학습했을 경우

‘라이거’라는 label은 학습되지 않았지만

분류할 수 있게 됨

ConSE: Convex combination of semantic embeddings

기존 제로샷은

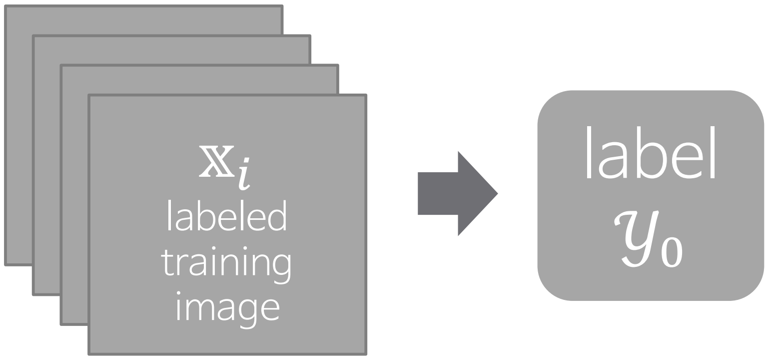
$f: X \rightarrow S$ 를 **직접** 학습하지만

이 논문에서 제안하는 모델(ConSE)은
$f: X \rightarrow S$ 를 간접적으로 학습함

오히려 classic machine learning 처럼

기존 classifier를 그대로 가져다 이용

ConSE: Convex combination of semantic embeddings



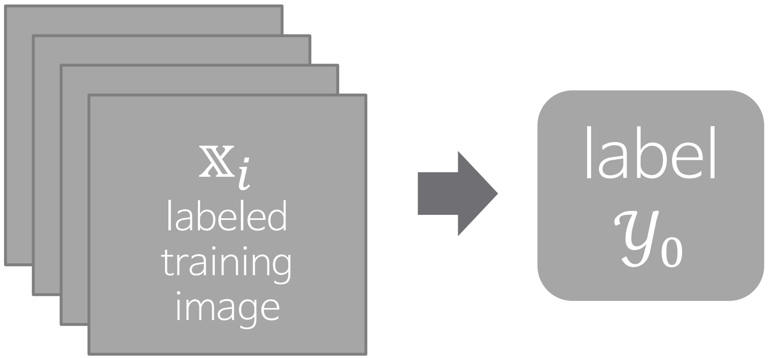
classifier 학습

$$p_0(y|\mathbb{x})$$

이미지 x 의 label이 y 일 확률

$$\sum_{y=1}^{n_0} p_0(y|\mathbb{x}) = 1$$

ConSE: Convex combination of semantic embeddings



$$p_0(y|\mathbb{x})$$

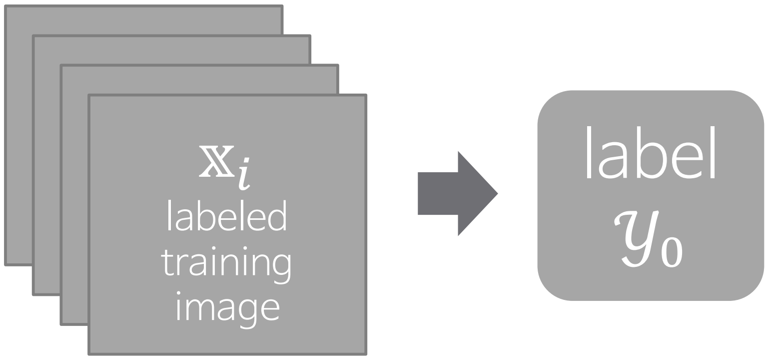
이미지 x 의 label이 y 일 확률

$$\widehat{y_0}(\mathbb{x}, 1) \equiv \underset{y \in \mathcal{Y}_0}{\operatorname{argmax}}\, p_0(y|\mathbb{x})$$

most likely training label

(즉, 이 classifier는 x의 label을 이 label로 판단할 것이다)

ConSE: Convex combination of semantic embeddings



$$p_0(y|\mathbb{x})$$

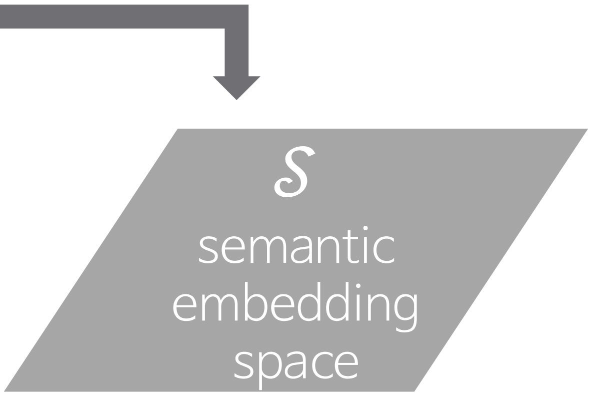
이미지 x 의 라벨이 y 일 확률

$$\widehat{y_0}(\mathbb{x}, t)$$

t^{th} most likely training label

ConSE: Convex combination of semantic embeddings

$f: X \to S$



$p_0(y|\mathbb{x})$: 이미지 x 의 라벨이 y 일 확률

$\widehat{y_0}(\mathbb{x}, t)$: t^{th} most likely training label

$$f(\mathbb{x}) = \frac{1}{Z}\sum_{t=1}^{T} p(\widehat{y_0}(\mathbb{x}, t)|\mathbb{x}) \cdot s(\widehat{y_0}(\mathbb{x}, t))$$

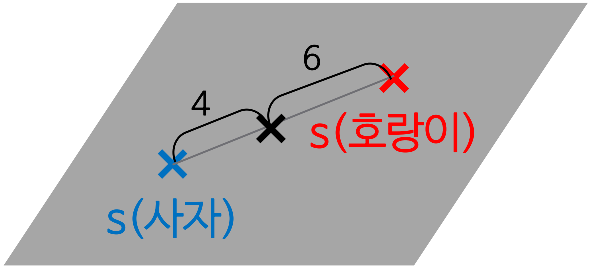
ConSE: Convex combination of semantic embeddings

‘라이거’ 이미지를 classifier에 넣었을 때 각 label일 확률

y		$p_0(y\|\mathbb{x})$
$\widehat{y_0}(\mathbb{x},1)$	사자	0.6
$\widehat{y_0}(\mathbb{x},2)$	호랑이	0.4

$$f(\mathbb{x}) = 0.6 \cdot s(\text{'사자'}) + 0.4 \cdot s(\text{'호랑이'})$$

$$\approx s(\text{'라이거'})$$



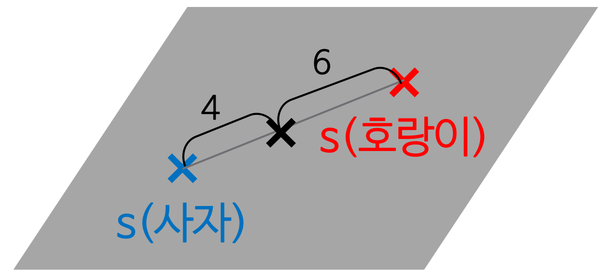
$$f(\mathbb{x}) = \frac{1}{Z}\sum_{t=1}^{T} p(\widehat{y_0}(\mathbb{x},t)|\mathbb{x}) \cdot s(\widehat{y_0}(\mathbb{x},t))$$

ConSE: Convex combination of semantic embeddings

‘라이거’ 이미지를 classifier에 넣었을 때 각 label일 확률

y		$p_0(y\|\mathbb{x})$
$\widehat{y_0}(\mathbb{x},1)$	사자	0.6
$\widehat{y_0}(\mathbb{x},2)$	호랑이	0.4

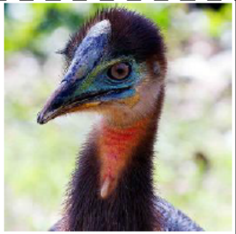
$$f(\mathbb{x}) = 0.6 \cdot s(\text{'사자'}) + 0.4 \cdot s(\text{'호랑이'})$$
$$\approx s(\text{'라이거'})$$



cosine similarity

$$\widehat{y_1}(\mathbb{x},1) \equiv \underset{y' \in \mathcal{Y}_1}{\operatorname{argmax}}\, cos(f(\mathbb{x}), s(y'))$$

Result

Test Image	Softmax Baseline [7]	DeViSE [6]	ConSE (10)
	wig fur coat Saluki, gazelle hound Afghan hound, Afghan stole	water spaniel tea gown bridal gown, wedding gown spaniel tights, leotards	business suit **dress, frock** hairpiece, false hair, postiche swimsuit, swimwear, bathing suit kit, outfit
	ostrich, Struthio camelus black stork, Ciconia nigra vulture crane peacock	heron owl, bird of Minerva, bird of night hawk bird of prey, raptor, raptorial bird finch	**ratite, ratite bird, flightless bird** peafowl, bird of Juno common spoonbill New World vulture, cathartid Greek partridge, rock partridge
	sea lion plane, carpenter's plane cowboy boot loggerhead, loggerhead turtle goose	elephant turtle turtleneck, turtle, polo-neck flip-flop, thong handcart, pushcart, cart, go-cart	California sea lion **Steller sea lion** Australian sea lion South American sea lion eared seal
	hamster broccoli Pomeranian capuchin, ringtail weasel	**golden hamster, Syrian hamster** rhesus, rhesus monkey pipe shaker American mink, Mustela vison	**golden hamster, Syrian hamster** rodent, gnawer Eurasian hamster rhesus, rhesus monkey rabbit, coney, cony

Result

Test Label Set	# Candidate Labels	Model	Flat hit@k (%) 1	2	5	10	20
2-hops	1,589	DeViSE	6.0	10.0	18.1	26.4	36.4
		ConSE(1)	9.3	14.4	23.7	30.8	38.7
		ConSE(10)	**9.4**	**15.1**	**24.7**	**32.7**	**41.8**
		ConSE(1000)	9.2	14.8	24.1	32.1	41.1
2-hops (+1K)	1,589 +1000	DeViSE	**0.8**	2.7	7.9	14.2	22.7
		ConSE(1)	0.2	**7.1**	**17.2**	24.0	31.8
		ConSE(10)	0.3	6.2	17.0	**24.9**	**33.5**
		ConSE(1000)	0.3	6.2	16.7	24.5	32.9
3-hops	7,860	DeViSE	1.7	2.9	5.3	8.2	12.5
		ConSE(1)	2.6	4.2	7.3	10.8	14.8
		ConSE(10)	**2.7**	**4.4**	**7.8**	**11.5**	**16.1**
		ConSE(1000)	2.6	4.3	7.6	11.3	15.7
3-hops (+1K)	7,860 +1000	DeViSE	**0.5**	1.4	3.4	5.9	9.7
		ConSE(1)	0.2	**2.4**	**5.9**	9.3	13.4
		ConSE(10)	0.2	2.2	**5.9**	**9.7**	**14.3**
		ConSE(1000)	0.2	2.2	5.8	9.5	14.0
ImageNet 2011 21K	20,841	DeViSE	0.8	1.4	2.5	3.9	6.0
		ConSE(1)	1.3	2.1	3.6	5.4	7.6
		ConSE(10)	**1.4**	**2.2**	**3.9**	**5.8**	**8.3**
		ConSE(1000)	1.3	2.1	3.8	5.6	8.1
ImageNet 2011 21K (+1K)	20,841 +1000	DeViSE	**0.3**	0.8	1.9	3.2	5.3
		ConSE(1)	0.1	1.2	3.0	4.8	7.0
		ConSE(10)	0.2	1.2	3.0	**5.0**	**7.5**
		ConSE(1000)	0.2	1.2	3.0	4.9	7.3

Result

Test Label Set	Model	Hierarchical precision@k				
		1	2	5	10	20
2-hops	DeViSE	0.06	0.152	0.192	0.217	0.233
	ConSE(10)	**0.094**	**0.214**	**0.247**	**0.269**	**0.284**
2-hops (+1K)	Softmax baseline	0	**0.236**	0.181	0.174	0.179
	DeViSE	**0.008**	0.204	0.196	0.201	0.214
	ConSE(10)	0.003	0.234	**0.254**	**0.260**	**0.271**
3-hops	DeViSE	0.017	0.037	0.191	0.214	0.236
	ConSE(10)	**0.027**	**0.053**	**0.202**	**0.224**	**0.247**
3-hops (+1K)	Softmax baseline	0	0.053	0.157	0.143	0.130
	DeViSE	**0.005**	0.053	0.192	0.201	0.214
	ConSE(10)	0.002	**0.061**	**0.211**	**0.225**	**0.240**
ImageNet 2011 21K	DeViSE	0.008	0.017	0.072	0.085	0.096
	ConSE(10)	**0.014**	**0.025**	**0.078**	**0.092**	**0.104**
ImageNet 2011 21K (+1K)	Softmax baseline	0	0.023	0.071	0.069	0.065
	DeViSE	**0.003**	0.025	0.083	0.092	0.101
	ConSE(10)	0.002	**0.029**	**0.086**	**0.097**	**0.105**

Result

Test Label Set	Model	Hierarchical precision@k				
		1	2	5	10	20
2-hops	DeViSE	0.06	0.152	0.192	0.217	0.233
	ConSE(10)	**0.094**	**0.214**	**0.247**	**0.269**	**0.284**
2-hops (+1K)	Softmax baseline	0	**0.236**	0.181	0.174	0.179
	DeViSE	**0.008**	0.204	0.196	0.201	0.214
	ConSE(10)	0.003	0.234	**0.254**	**0.260**	**0.271**
3-hops	DeViSE	0.017	0.037	0.191	0.214	0.236
	ConSE(10)	**0.027**	**0.053**	**0.202**	**0.224**	**0.247**
3-hops (+1K)	Softmax baseline	0	0.053	0.157	0.143	0.130
	DeViSE	**0.005**	0.053	0.192	0.201	0.214
	ConSE(10)	0.002	**0.061**	**0.211**	**0.225**	**0.240**
ImageNet 2011 21K	DeViSE	0.008	0.017	0.072	0.085	0.096
	ConSE(10)	**0.014**	**0.025**	**0.078**	**0.092**	**0.104**
ImageNet 2011 21K (+1K)	Softmax baseline	0	0.023	0.071	0.069	0.065
	DeViSE	**0.003**	0.025	0.083	0.092	0.101
	ConSE(10)	0.002	**0.029**	**0.086**	**0.097**	**0.105**

Result

Test Label Set	Model	Hierarchical precision@k				
		1	2	5	10	20
ImageNet 2011 1K	Softmax baseline	**0.556**	**0.452**	0.342	0.313	0.319
	DeViSE	0.532	0.447	**0.352**	**0.331**	**0.341**
	ConSE (1)	0.551	0.422	0.32	0.297	0.313
	ConSE (10)	0.543	0.447	0.348	0.322	0.337
	ConSE (1000)	0.539	0.442	0.344	0.319	0.335

Test Label Set	Model	Flat hit@k (%)			
		1	2	5	10
ImageNet 2011 1K	Softmax baseline	**55.6**	**67.4**	**78.5**	**85.0**
	DeViSE	53.2	65.2	76.7	83.3
	ConSE (1)	55.1	57.7	60.9	63.5
	ConSE (10)	54.3	61.9	68.0	71.6
	ConSE (1000)	53.9	61.1	67.0	70.6

감사합니다