Open DMQA Seminar

Contrastive Learning for Sentence Embedding

2023. 04. 28.

Data Mining & Quality Analytics Lab

정재윤

발표자 소개

❖ 정재윤 (Jaeyoon Jeong)

- 고려대학교 산업경영공학과 석사 과정 (2021.09 ~)
- Data Mining & Quality Analytics Lab (김성범 교수님)

❖ Research Interest

- Anomaly Detection and its application
- Application of deep learning and machine learning algorithms

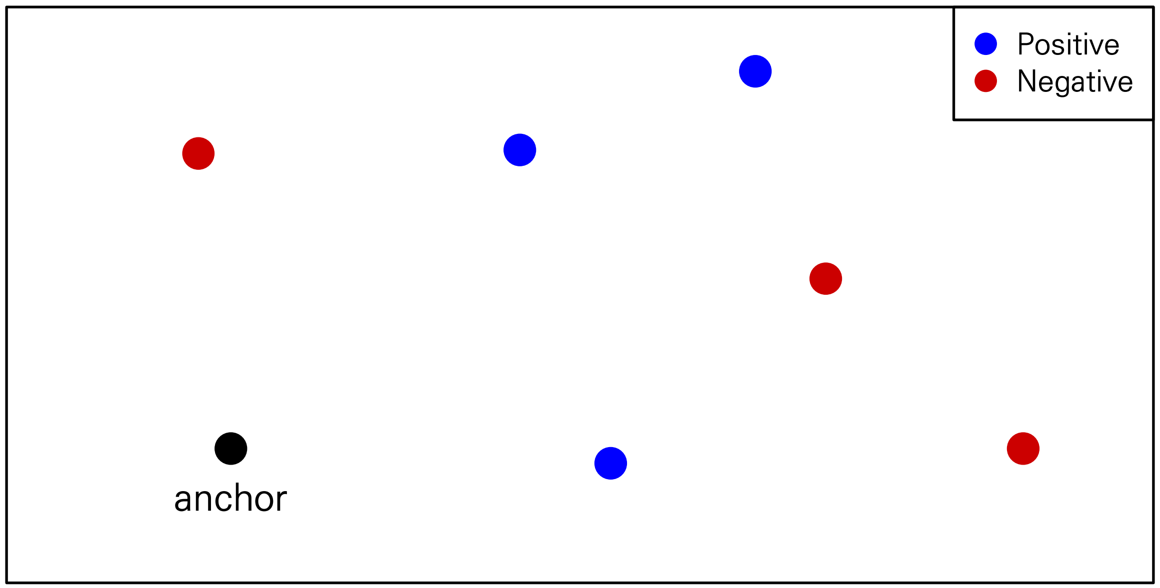
❖ Contact

- E-mail: jj950310@korea.ac.kr

목차

❖ Contrastive learning

- Metric Learning 방법론 중 하나로 특정한 기준을 바탕으로 데이터 간의 거리 함수를 학습하는 방법론
- Anchor를 기준으로 Positive samples은 가깝도록, Negative samples은 멀어지도록 학습

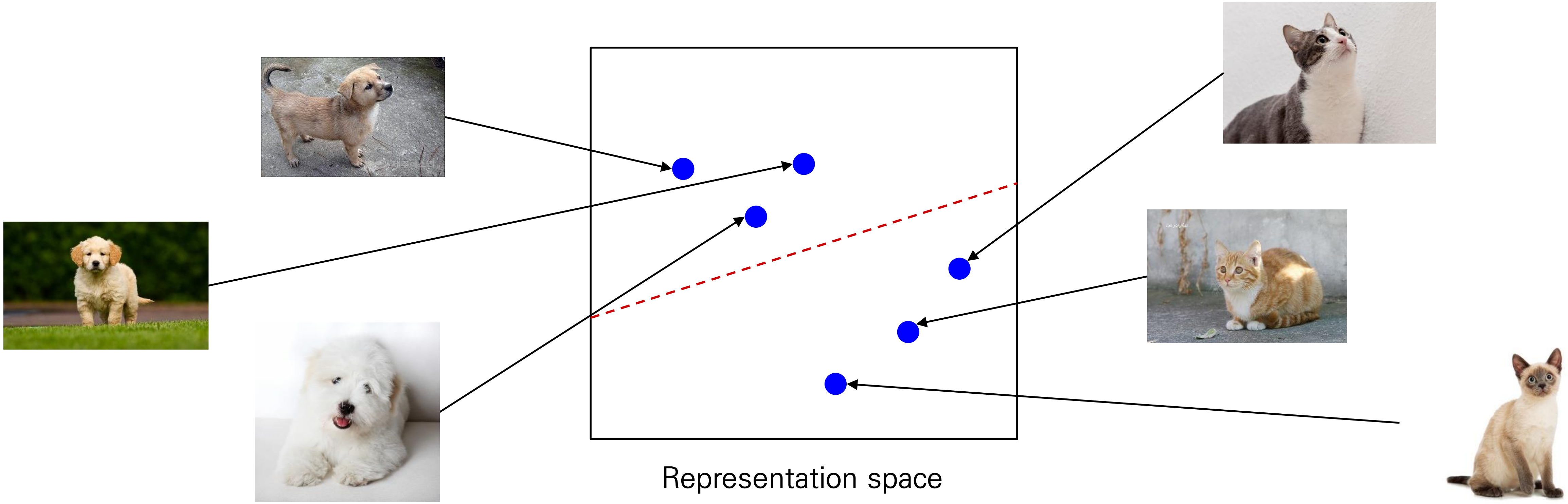


Introduction

Contrastive learning in Vision

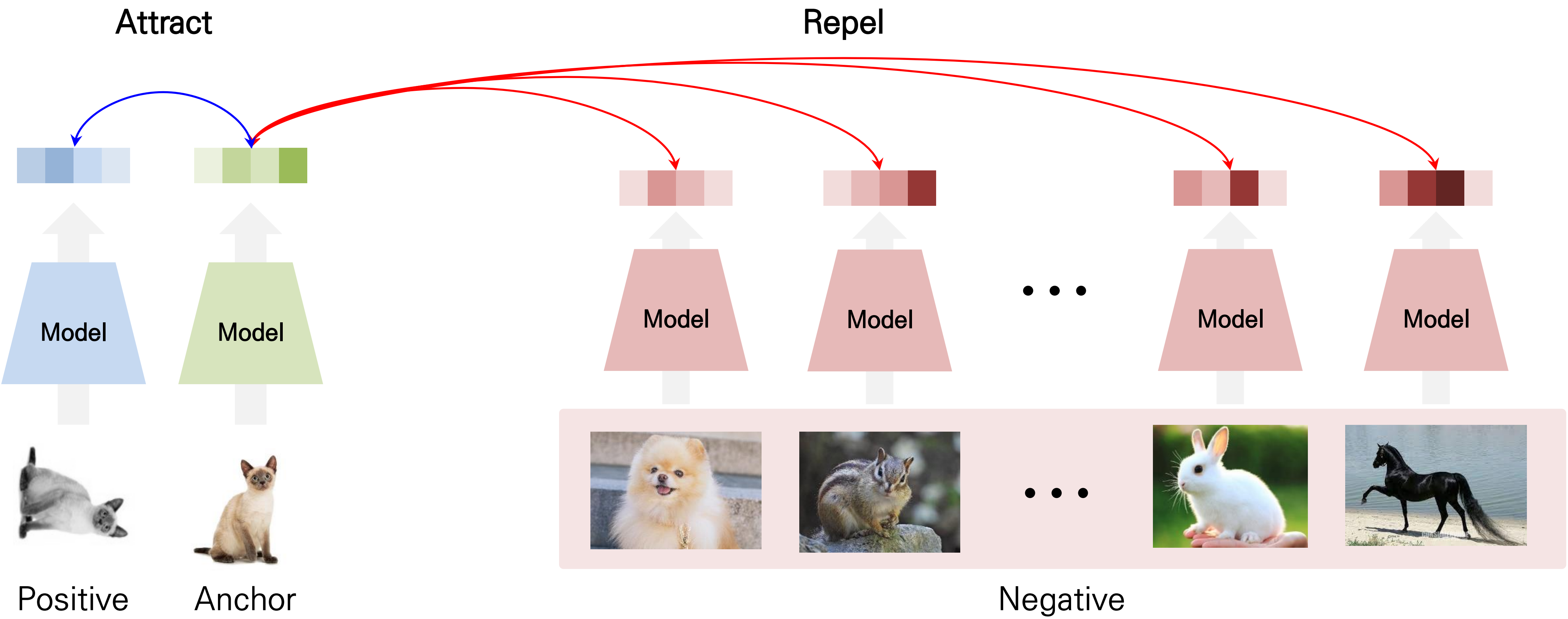
❖ Contrastive learning in Vision

- Self-supervised learning의 방법론으로 사용
- 레이블이 없는 상황에서도, 보다 좋은 representation을 학습하기 위한 방법론으로 발전



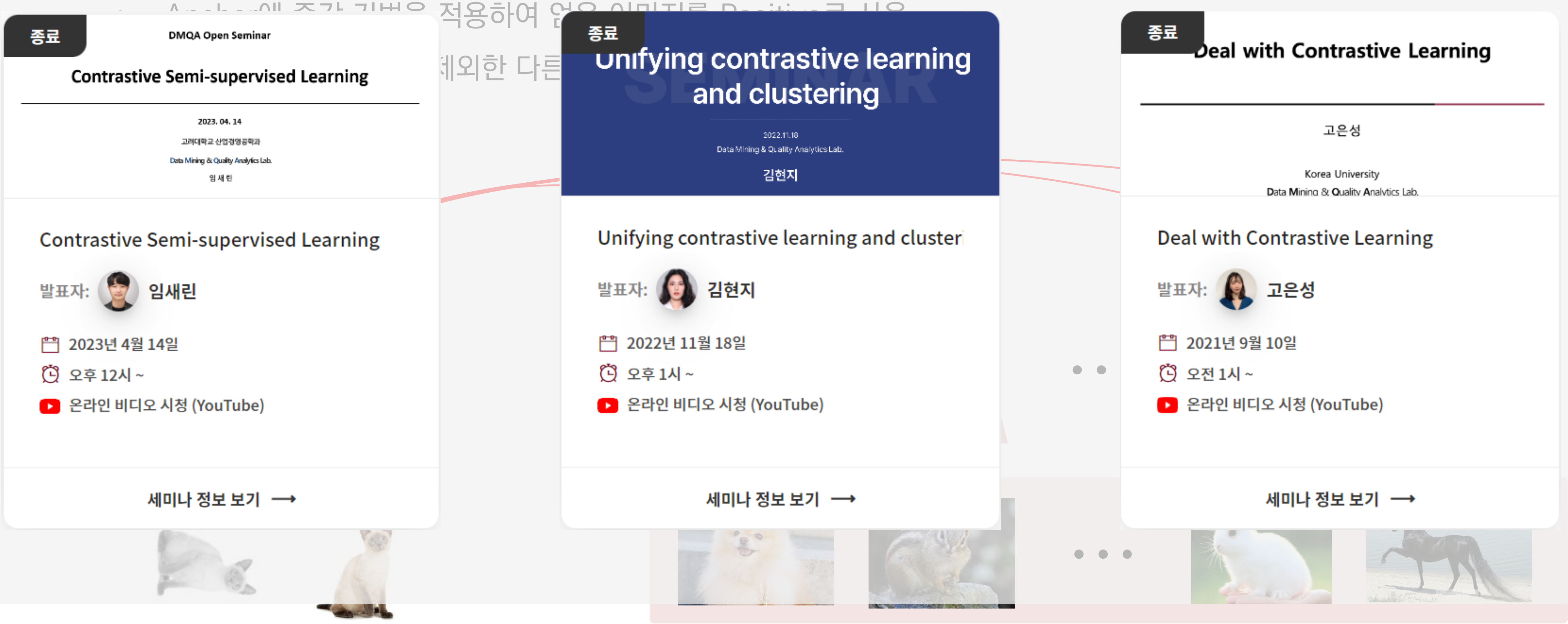
❖ Contrastive learning in Vision

- Anchor에 증강 기법을 적용하여 얻은 이미지를 Positive sample로 사용
- Anchor와 Positive sample을 제외한 다른 이미지들을 모두 Negative samples로 사용



Introduction
Contrastive learning in Vision

❖ Contrastive learning in Vision



Positive Anchor Negative

❖ Contrastive learning in Vision

- Positive sample에 적용되는 data augmentation 기법에 따라 모델의 성능이 크게 변함
- Augmentation 방법이 다양하고, augmentation을 진행해도 Anchor가 가진 레이블 정보가 유지됨

Anchor rotation Color distort Crop and resize

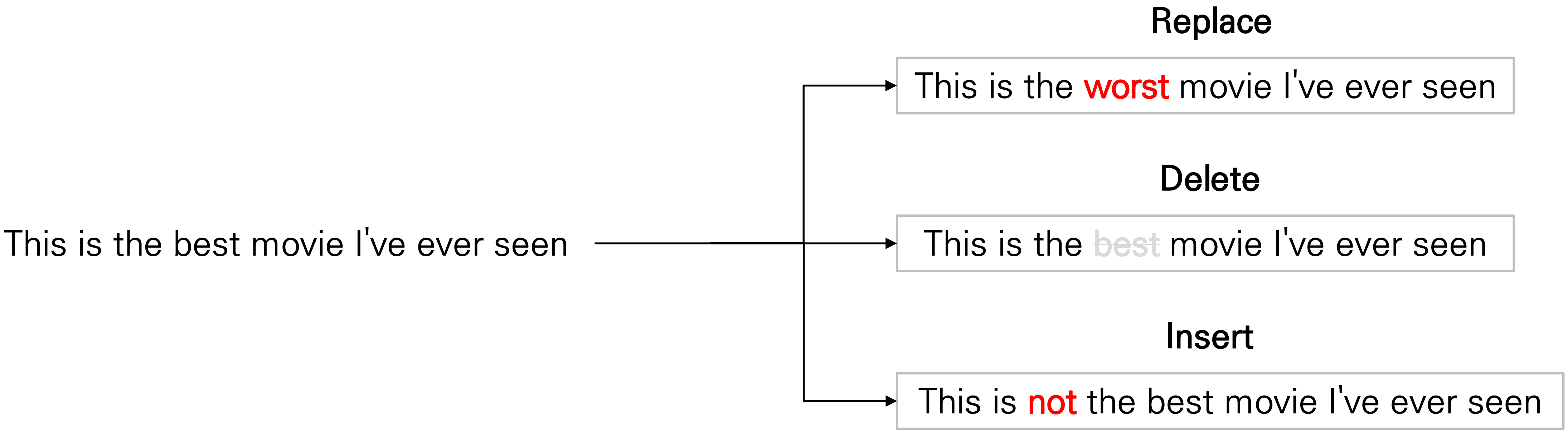
Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In *International conference on machine learning* (pp. 1597-1607). PMLR.

Introduction

Contrastive learning in NLP

❖ Contrastive learning in NLP

- 이미지와는 달리, augmentation 방법론이 비교적 제한적임
- Augmentation을 잘못 적용하면, anchor가 가진 레이블 정보를 크게 해침
- 따라서 Contrastive learning을 적용하기에 조심스러움



Introduction

Contrastive learning in NLP

❖ Contrastive learning in NLP

- 최근 자연어 처리 연구 분야 중 Sentence embedding 쪽에서 contrastive learning을 적용해 성능을 개선시킴

DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations

DeCLUTR

SimCSE: Simple Contrastive Learning of Sentence Embeddings

SimCSE

DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings

DiffCSE

Contrastive learning for sentence embedding

❖ DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations

- 2021년 ACL에 발표된 논문으로, 2023년 4월 25일 기준으로 총 276회 인용
- Document에 대한 anchor-positive를 다양하게 구성해 contrastive learning을 적용

DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations

John Giorgi[1,5,6] Osvald Nitski[2,7] Bo Wang[1,4,6,7,†] Gary Bader[1,3,5,†]
[1]Department of Computer Science, University of Toronto
[2]Faculty of Applied Science and Engineering, University of Toronto
[3]Department of Molecular Genetics, University of Toronto
[4]Department of Laboratory Medicine and Pathobiology, University of Toronto
[5]Terrence Donnelly Centre for Cellular & Biomolecular Research
[6]Vector Institute for Artificial Intelligence
[7]Peter Munk Cardiac Center, University Health Network
[†]Co-senior authors
{john.giorgi, osvald.nitski, gary.bader}@mail.utoronto.ca
bowang@vectorinstitute.ai

Abstract

Sentence embeddings are an important component of many natural language processing (NLP) systems. Like word embeddings, sentence embeddings are typically learned on

2014). Recent work has demonstrated strong transfer task performance using pretrained *sentence* embeddings. These fixed-length vectors, often referred to as "universal" sentence embeddings, are typically learned on large corpora and then trans-

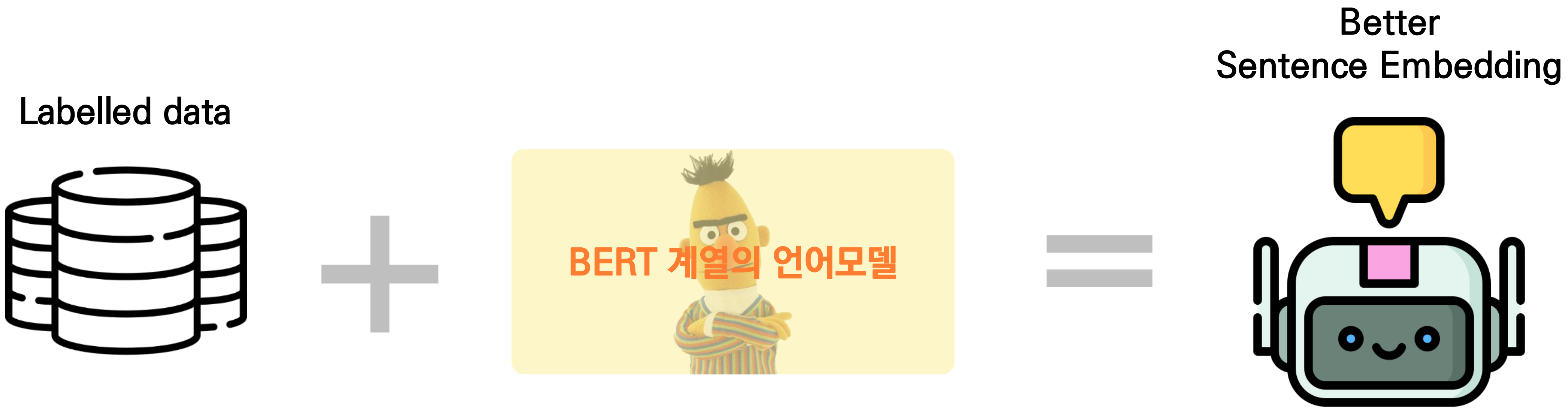
Giorgi, J., Nitski, O., Bader, G.D., & Wang, B. (2020). DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations. ArXiv, abs/2006.03659.

Contrastive learning for sentence embedding

DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations

❖ 연구 배경

- 최근 자연어 처리 연구에서는 Sentence embedding을 활용해 좋은 성능을 얻음
- 그러나 좋은 Sentence embedding을 학습하기 위해서는 대량의 labelled data가 필요함
- 본 논문에서는 레이블이 없는 데이터를 Contrastive learning을 사용해 좋은 sentence embedding을 학습하고자 함

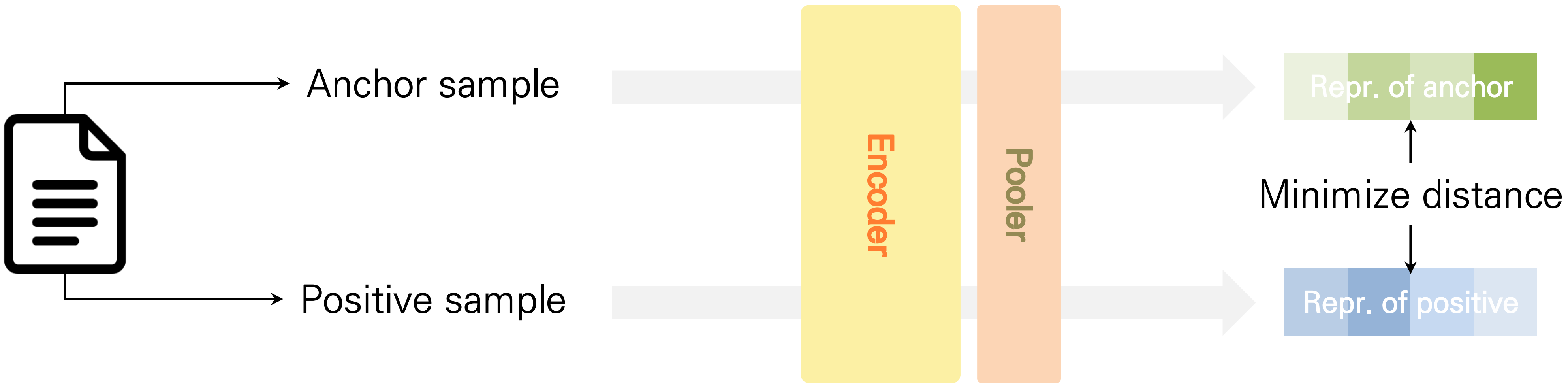


Contrastive learning for sentence embedding

DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations

❖ 전체 구조

- 기본적인 구조는 Vision에서의 contrastive learning 구조와 유사
- Anchor와 positive를 같은 encoder 및 pooler로 통과시킴
- 산출된 Embedding vector들의 거리를 최소화

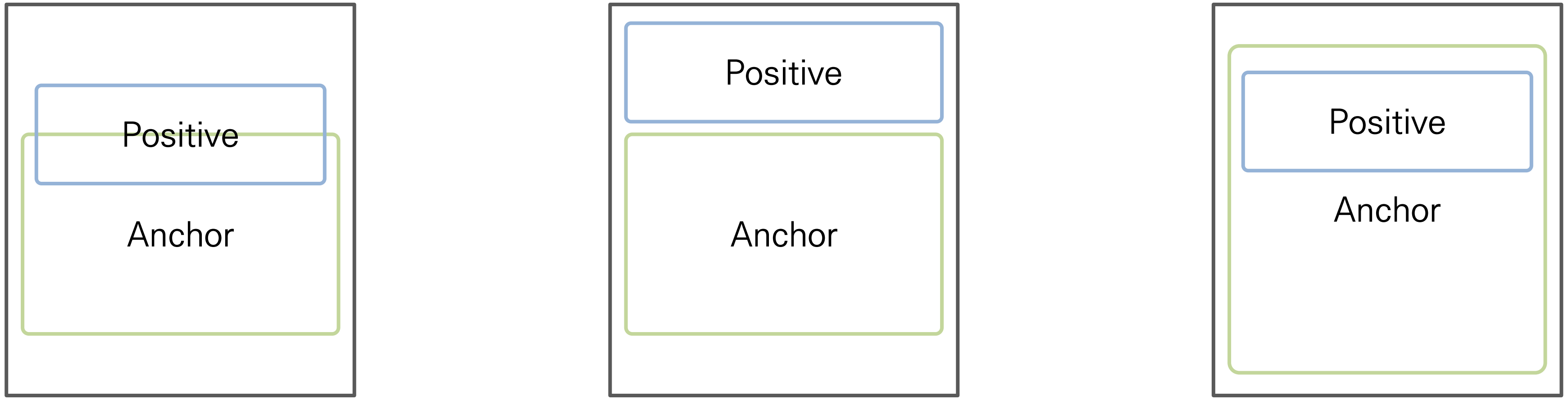


Contrastive learning for sentence embedding

DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations

❖ Anchor – Positive 구성 방법

- 하나의 Batch는 batch size만큼의 documents로 구성
- 선택된 Document에서 anchor와 positive를 아래의 방법으로 추출
  - ✓ Anchor 문장과 일부가 겹치게 positive를 구성
  - ✓ Anchor 문장과 겹치지 않게 전후의 인접한 문장으로 positive를 구성
  - ✓ Anchor 문장 내의 문장으로 positive를 구성



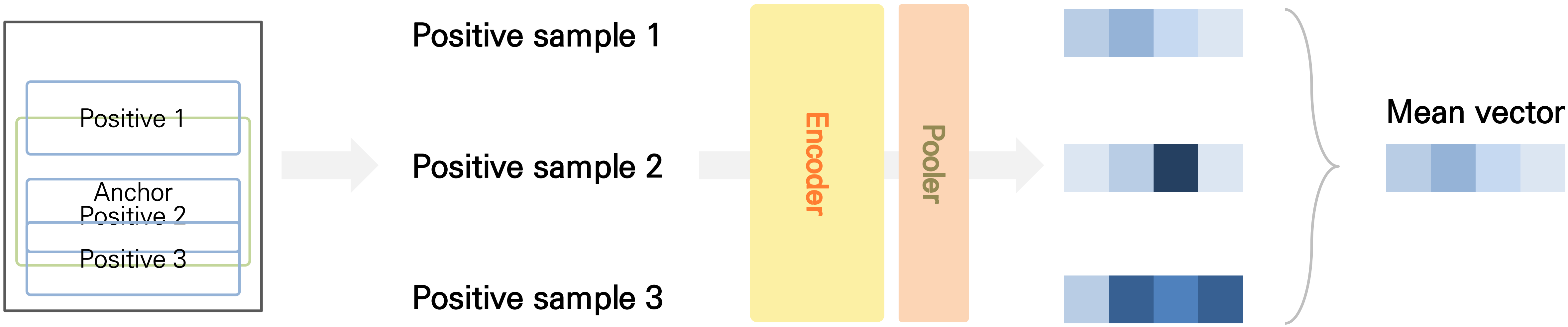
3 type of Anchor – Positive pair

Contrastive learning for sentence embedding

DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations

❖ Anchor – Positive 구성 방법 (One anchor, many positive)

- 한 개의 Anchor에 여러 개의 positive sample을 수집함
- Positive sample에 대한 vector의 평균을 positive sample로 사용
- 한 개의 Positive를 사용하는 것보다 다수의 positive의 평균이 보다 성능을 개선

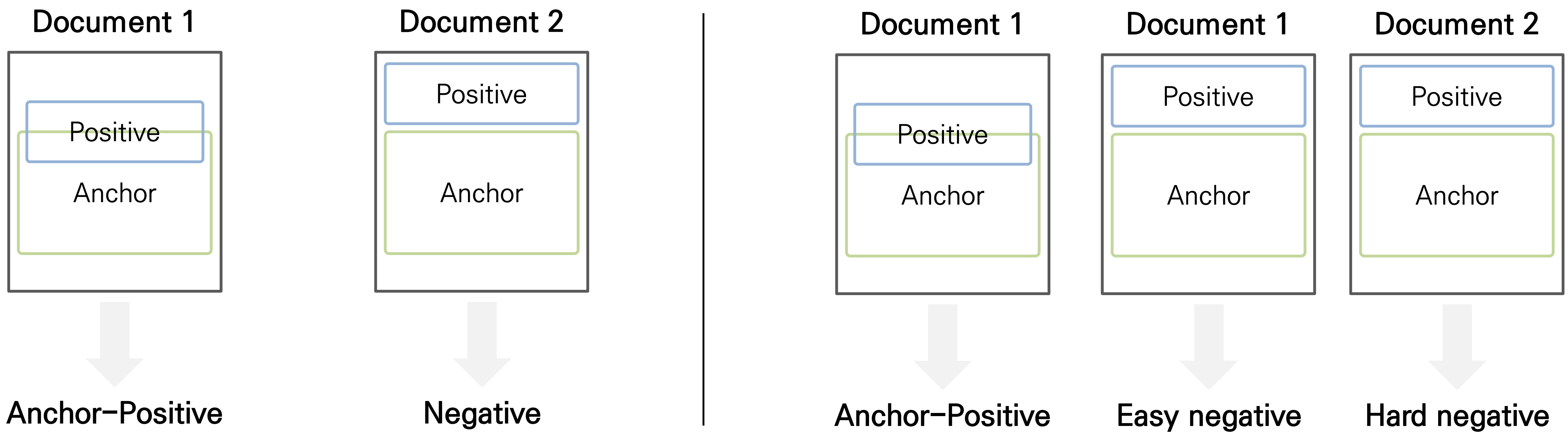


Contrastive learning for sentence embedding

DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations

❖ Anchor – Positive 구성 방법 (Many Anchor)

- 하나의 Document에서 여러 개의 anchor 생성 가능
- Document의 Anchor가 하나일 때, batch 내 다른 document의 anchor – positive sample를 negative로 사용
- Document의 Anchor가 여러 개일 때, 같은 document의 다른 anchor – positive sample를 hard negative로 사용



Contrastive learning for sentence embedding

DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations

❖ 실험 결과

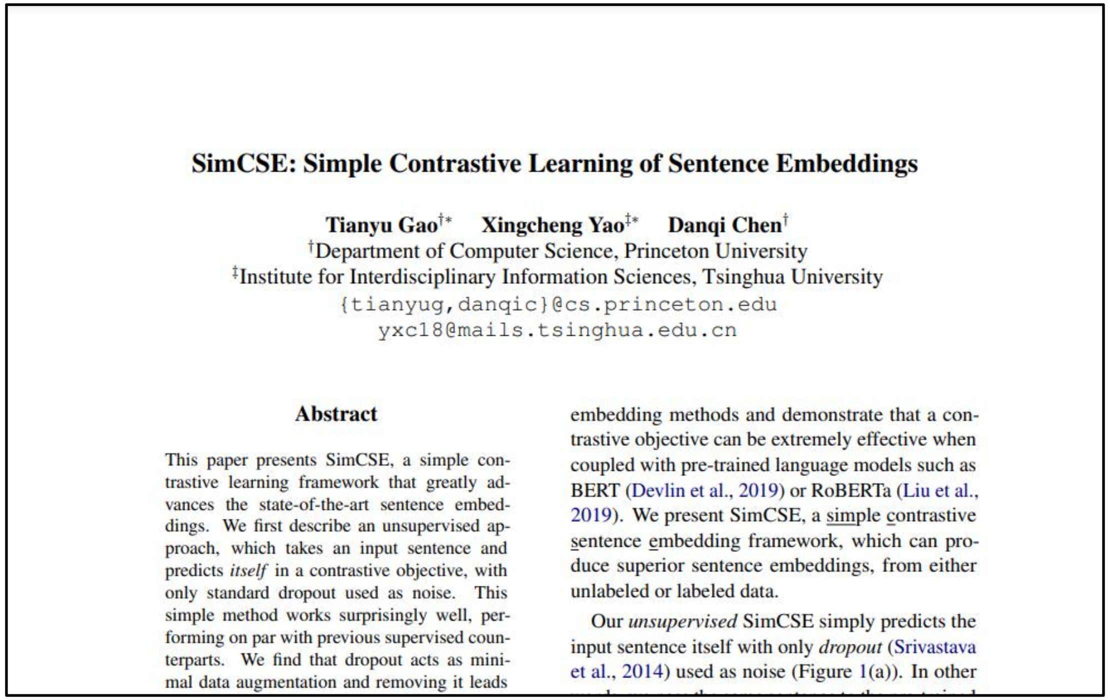
- Transformer-small과 Transformer-base는 각각 DistilRoBERTa와 RoBERTa-base를 의미
- Downstream task에서 보통 좋은 성능을 보인다는 Transformer 계열의 언어모델보다 DeCLUTR가 전반적으로 높은 성능을 확인함

Model	CR	MR	MPQA	SUBJ	SST2	SST5	TREC	MRPC	SNLI	Avg.	Δ
					Bag-of-words (BoW) weak baselines						
GloVe	78.78	77.70	87.76	91.25	80.29	44.48	83.00	73.39/81.45	65.85	65.47	-13.63
fastText	79.18	78.45	87.88	91.53	82.15	45.16	83.60	74.49/82.44	68.79	68.56	-10.54
					Supervised and semi-supervised						
InferSent	84.37	79.42	89.04	93.03	84.24	45.34	90.80	76.35/83.48	84.16	76.00	-3.10
USE	85.70	79.38	88.89	93.11	84.90	46.11	**95.00**	72.41/82.01	83.25	78.89	-0.21
Sent. Transformers	**90.78**	84.98	88.72	92.67	**90.55**	**52.76**	87.40	76.64/82.99	**84.18**	77.19	-1.91
					Unsupervised						
QuickThoughts	86.00	82.40	**90.20**	94.80	87.60	–	92.40	**76.90/84.00**	–	–	–
Transformer-small	86.60	82.12	87.04	94.77	88.03	49.50	91.60	74.55/81.75	71.88	72.58	-6.52
Transformer-base	88.19	84.35	86.49	95.28	89.46	51.27	93.20	74.20/81.44	72.19	72.70	-6.40
DeCLUTR-small	87.52 ↑	82.79 ↑	87.87 ↑	94.96 ↑	87.64 ↓	48.42 ↓	90.80 ↓	75.36/82.70 ↑	73.59 ↑	77.50 ↑	-1.60
DeCLUTR-base	90.68 ↑	**85.16** ↑	88.52 ↑	**95.78** ↑	90.01 ↑	51.18 ↓	93.20 ↑	74.61/82.65 ↑	74.74 ↑	**79.10** ↑	–
Model	**SICK-E**	**SICK-R**	**STS-B**	**COCO**	**STS12***	**STS13***	**STS14***	**STS15***	**STS16***		
GloVe	78.89	72.30	62.86	0.40	53.44	51.24	55.71	59.62	57.93	–	–
fastText	79.01	72.98	68.26	0.40	58.85	58.83	63.42	69.05	68.24	–	–
InferSent	**86.30**	**83.06**	78.48	**65.84**	62.90	56.08	66.36	74.01	72.89	–	–
USE	85.37	81.53	**81.50**	62.42	**68.87**	71.70	**72.76**	**83.88**	**82.78**	–	–
Sent. Transformers	82.97	79.17	74.28	60.96	64.10	65.63	69.80	74.71	72.85	–	–
QuickThoughts	–	–	–	60.55	–	–	–	–	–	–	–
Transformer-small	81.96	77.51	70.31	60.48	53.99	45.53	57.23	65.57	63.51	–	–
Transformer-base	80.29	76.84	69.62	60.14	53.28	46.10	56.17	64.69	62.79	–	–
DeCLUTR-small	83.46 ↑	77.66 ↑	77.51 ↑	60.85 ↑	63.66 ↑	68.93 ↑	70.40 ↑	78.25 ↑	77.74 ↑	–	–
DeCLUTR-base	83.84 ↑	78.62 ↑	79.39 ↑	62.35 ↑	63.56 ↑	**72.58** ↑	71.70 ↑	79.95 ↑	79.59 ↑	–	–

Contrastive learning for sentence embedding

❖ SimCSE: Simple Contrastive Learning of Sentence Embeddings

- 2021년 EMNLP에 발표된 논문으로, 2023년 4월 25일 기준으로 총 1017회 인용
- Dropout noise를 augmentation 기법으로 활용하는 간단한 방법을 통해 당시 SOTA 달성



SimCSE: Simple Contrastive Learning of Sentence Embeddings

Tianyu Gao[†*] Xingcheng Yao[‡*] Danqi Chen[†]
[†]Department of Computer Science, Princeton University
[‡]Institute for Interdisciplinary Information Sciences, Tsinghua University
{tianyug,danqic}@cs.princeton.edu
yxc18@mails.tsinghua.edu.cn

Abstract

This paper presents SimCSE, a simple contrastive learning framework that greatly advances the state-of-the-art sentence embeddings. We first describe an unsupervised approach, which takes an input sentence and predicts *itself* in a contrastive objective, with only standard dropout used as noise. This simple method works surprisingly well, performing on par with previous supervised counterparts. We find that dropout acts as minimal data augmentation and removing it leads

embedding methods and demonstrate that a contrastive objective can be extremely effective when coupled with pre-trained language models such as BERT (Devlin et al., 2019) or RoBERTa (Liu et al., 2019). We present SimCSE, a simple contrastive sentence embedding framework, which can produce superior sentence embeddings, from either unlabeled or labeled data.

Our *unsupervised* SimCSE simply predicts the input sentence itself with only *dropout* (Srivastava et al., 2014) used as noise (Figure 1(a)). In other

Gao, T., Yao, X., & Chen, D. (2021). Simcse: Simple contrastive learning of sentence embeddings. arXiv preprint arXiv:2104.08821.

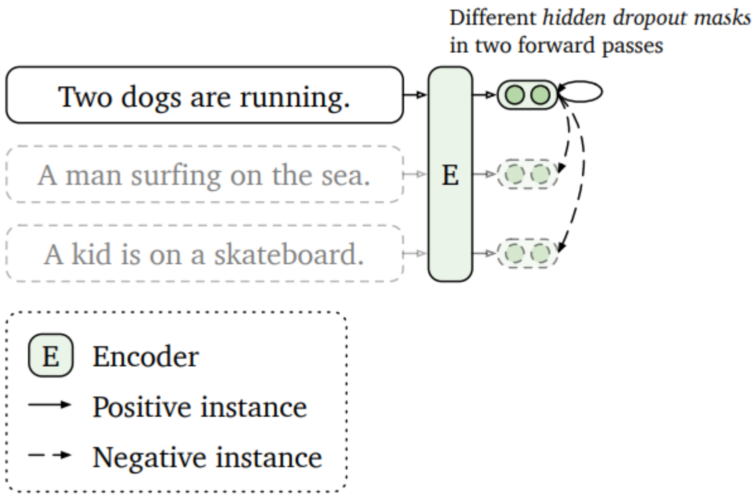
Contrastive learning for sentence embedding

SimCSE: Simple Contrastive Learning of Sentence Embeddings

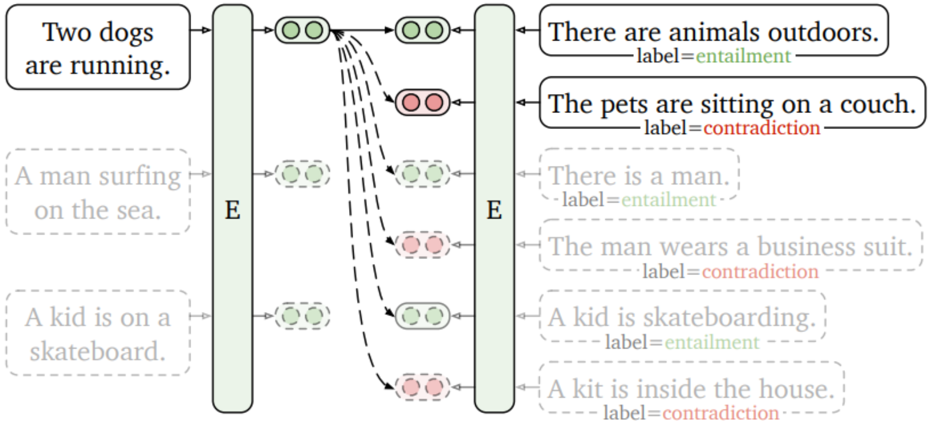
❖ 제안 방법론

- 논문에서는 Supervised SimCSE과 unsupervised SimCSE 2가지 제안 방법론을 제시
- 핵심이 되는 부분은 Unsupervised SimCSE 방법론

Unsupervised SimCSE



Supervised SimCSE

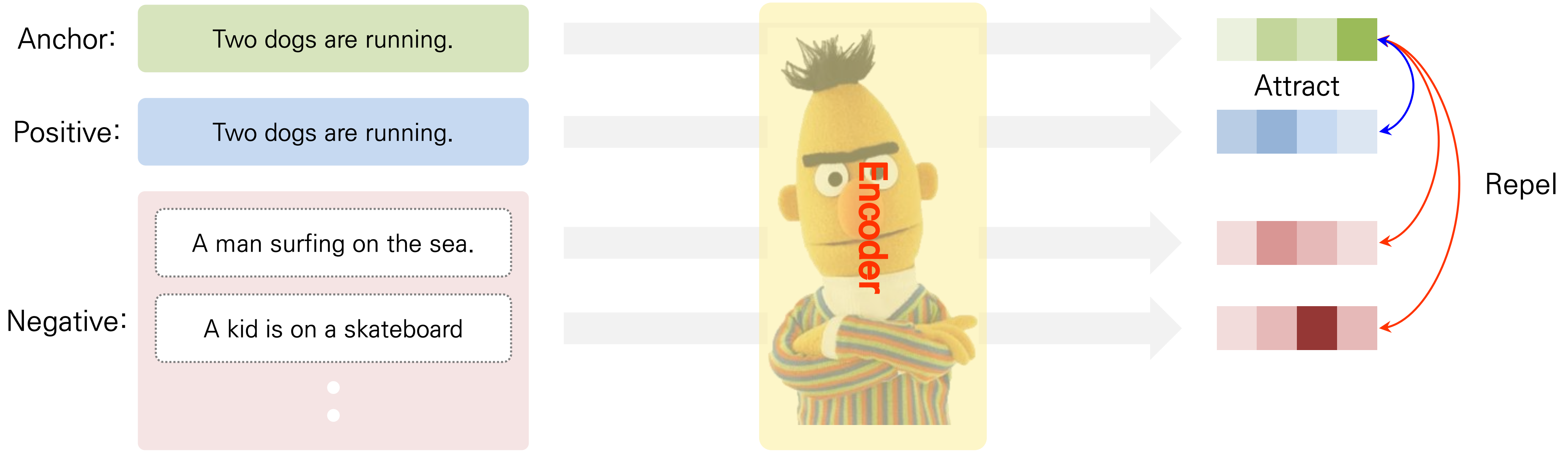


Contrastive learning for sentence embedding

SimCSE: Simple Contrastive Learning of Sentence Embeddings

❖ Unsupervised SimCSE

- 기존의 augmentation과는 달리, Encoder 내의 random dropout을 augmentation으로 활용
- 같은 문장을 Encoder에 2번 통과시켜, 얻은 2개의 embedding을 각각 anchor, positive로 사용함



Contrastive learning for sentence embedding

SimCSE: Simple Contrastive Learning of Sentence Embeddings

❖ Supervised SimCSE

- 기존 자연어 처리 분야에서 활용되는 NLI계열의 데이터셋을 활용해 Contrastive learning을 진행
- NLI 데이터셋은 각 관측치가 두 개의 문장으로 구성되며, 레이블로 두 문장의 관계를 가지고 있는 데이터셋
  - ✓ Entailment: 두 문장이 서로 참인 관계 → Positive
  - ✓ Neutral: 두 문장이 서로 중립인 관계
  - ✓ Contradiction: 두 문장이 서로 거짓인 관계 → Negative

Contrastive learning for sentence embedding

SimCSE: Simple Contrastive Learning of Sentence Embeddings

❖ 실험 결과

- 실험에서는 BERT-base, RoBERTa-base, RoBERTa-large를 encoder로 사용함
- 실험결과, Unsupervised model 중 가장 좋은 sentence embedding 성능을 달성
- 뿐만 아니라 supervised model 중에서도 가장 좋은 sentence embedding을 가짐을 실험적으로 보임

Model	STS12	STS13	STS14	STS15	STS16	STS-B	SICK-R	Avg.
Unsupervised models								
GloVe embeddings (avg.)♣	55.14	70.66	59.73	68.25	63.66	58.02	53.76	61.32
$\text{BERT}_{\text{base}}$ (first-last avg.)	39.70	59.38	49.67	66.03	66.19	53.87	62.06	56.70
$\text{BERT}_{\text{base}}$-flow	58.40	67.10	60.85	75.16	71.22	68.66	64.47	66.55
$\text{BERT}_{\text{base}}$-whitening	57.83	66.90	60.90	75.08	71.31	68.24	63.73	66.28
IS-$\text{BERT}_{\text{base}}$♡	56.77	69.24	61.21	75.23	70.16	69.21	64.25	66.58
CT-$\text{BERT}_{\text{base}}$	61.63	76.80	68.47	77.50	76.48	74.31	69.19	72.05
∗ SimCSE-$\text{BERT}_{\text{base}}$	**68.40**	**82.41**	**74.38**	**80.91**	**78.56**	**76.85**	**72.23**	**76.25**
$\text{RoBERTa}_{\text{base}}$ (first-last avg.)	40.88	58.74	49.07	65.63	61.48	58.55	61.63	56.57
$\text{RoBERTa}_{\text{base}}$-whitening	46.99	63.24	57.23	71.36	68.99	61.36	62.91	61.73
DeCLUTR-$\text{RoBERTa}_{\text{base}}$	52.41	75.19	65.52	77.12	78.63	72.41	**68.62**	69.99
∗ SimCSE-$\text{RoBERTa}_{\text{base}}$	**70.16**	**81.77**	**73.24**	**81.36**	**80.65**	**80.22**	68.56	**76.57**
∗ SimCSE-$\text{RoBERTa}_{\text{large}}$	**72.86**	**83.99**	**75.62**	**84.77**	**81.80**	**81.98**	**71.26**	**78.90**

Contrastive learning for sentence embedding

SimCSE: Simple Contrastive Learning of Sentence Embeddings

❖ 실험 결과

- 실험에서는 BERT-base, RoBERTa-base, RoBERTa-large를 encoder로 사용함
- 실험결과, Unsupervised model 중 가장 좋은 sentence embedding 성능을 달성
- 뿐만 아니라 supervised model 중에서도 가장 좋은 sentence embedding을 가짐을 실험적으로 보임

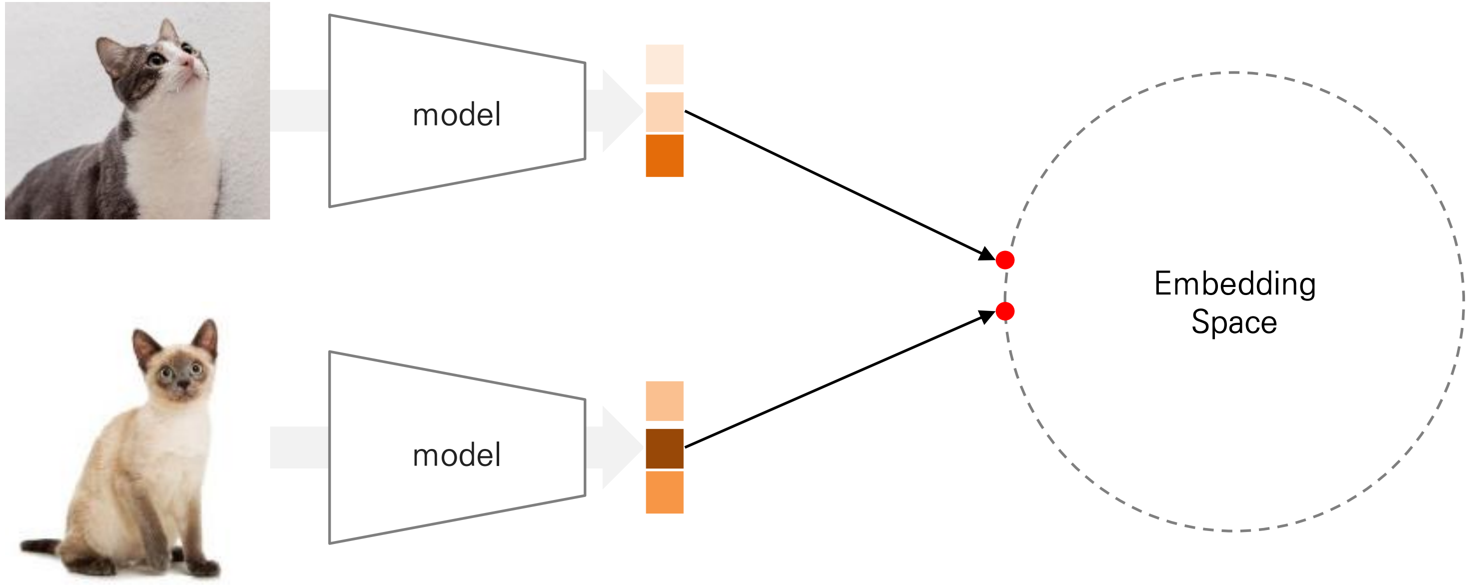
Supervised models								
InferSent-GloVe♣	52.86	66.75	62.15	72.77	66.87	68.03	65.65	65.01
Universal Sentence Encoder♣	64.49	67.80	64.61	76.83	73.18	74.92	76.69	71.22
$\text{SBERT}_{\text{base}}$♣	70.97	76.53	73.19	79.09	74.30	77.03	72.91	74.89
$\text{SBERT}_{\text{base}}$-flow	69.78	77.27	74.35	82.01	77.46	79.12	76.21	76.60
$\text{SBERT}_{\text{base}}$-whitening	69.65	77.57	74.66	82.27	78.39	79.52	76.91	77.00
CT-$\text{SBERT}_{\text{base}}$	74.84	83.20	78.07	83.84	77.93	81.46	76.42	79.39
∗ SimCSE-$\text{BERT}_{\text{base}}$	**75.30**	**84.67**	**80.19**	**85.40**	**80.82**	**84.25**	**80.39**	**81.57**
$\text{SRoBERTa}_{\text{base}}$♣	71.54	72.49	70.80	78.74	73.69	77.77	74.46	74.21
$\text{SRoBERTa}_{\text{base}}$-whitening	70.46	77.07	74.46	81.64	76.43	79.49	76.65	76.60
∗ SimCSE-$\text{RoBERTa}_{\text{base}}$	**76.53**	**85.21**	**80.95**	**86.03**	**82.57**	**85.83**	**80.50**	**82.52**
∗ SimCSE-$\text{RoBERTa}_{\text{large}}$	**77.46**	**87.27**	**82.36**	**86.66**	**83.93**	**86.70**	**81.95**	**83.76**

Contrastive learning for sentence embedding

SimCSE: Simple Contrastive Learning of Sentence Embeddings

❖ Alignment & Uniformity

- 최근 연구에서 주장하는 좋은 Embedding은 사소한 특징에는 변동이 없으며, 최대한 많은 정보를 보존해야 함
- Alignment: 유사한 sample들은 유사한 feature를 가져야 함
- Uniformity: sample들에 대한 정보는 최대한 많이 가질 수 있게 보존해야 한다.



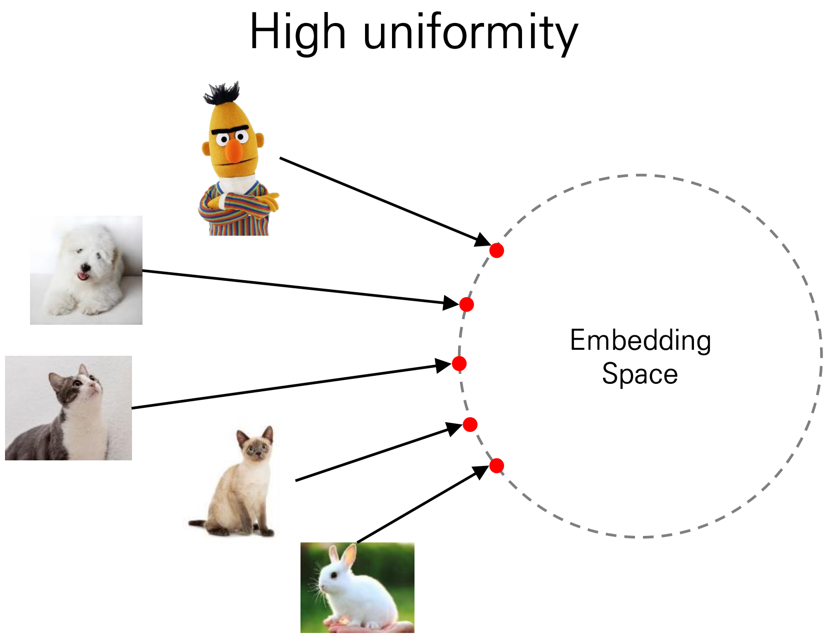
Wang, T., & Isola, P. (2020, November). Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In International Conference on Machine Learning (pp. 9929-9939). PMLR.

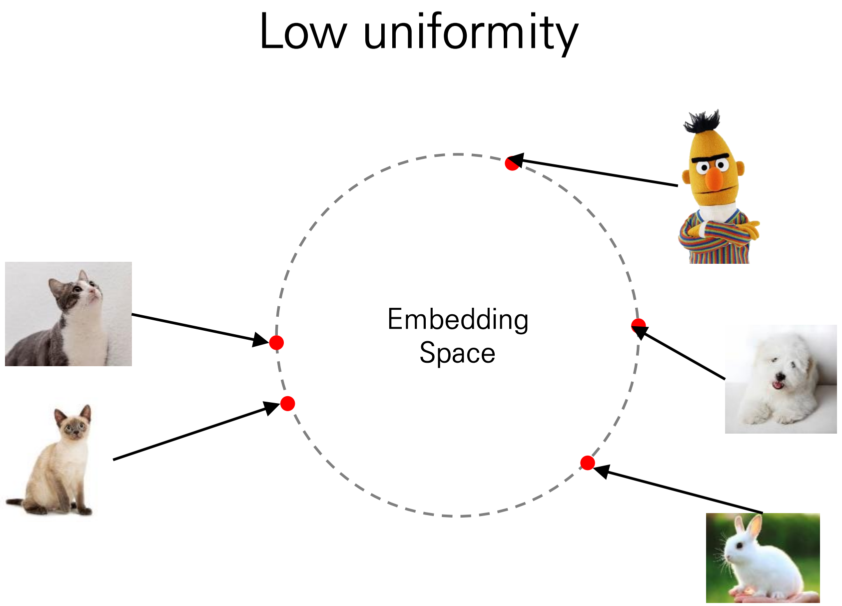
Contrastive learning for sentence embedding

SimCSE: Simple Contrastive Learning of Sentence Embeddings

❖ Alignment & Uniformity

- 최근 연구에서 주장하는 좋은 Embedding은 사소한 특징에는 변동이 없으며, 최대한 많은 정보를 보존해야 함
- Alignment: 유사한 sample들은 유사한 feature를 가져야 함
- Uniformity: sample들에 대한 정보는 최대한 많이 가질 수 있게 보존해야 함





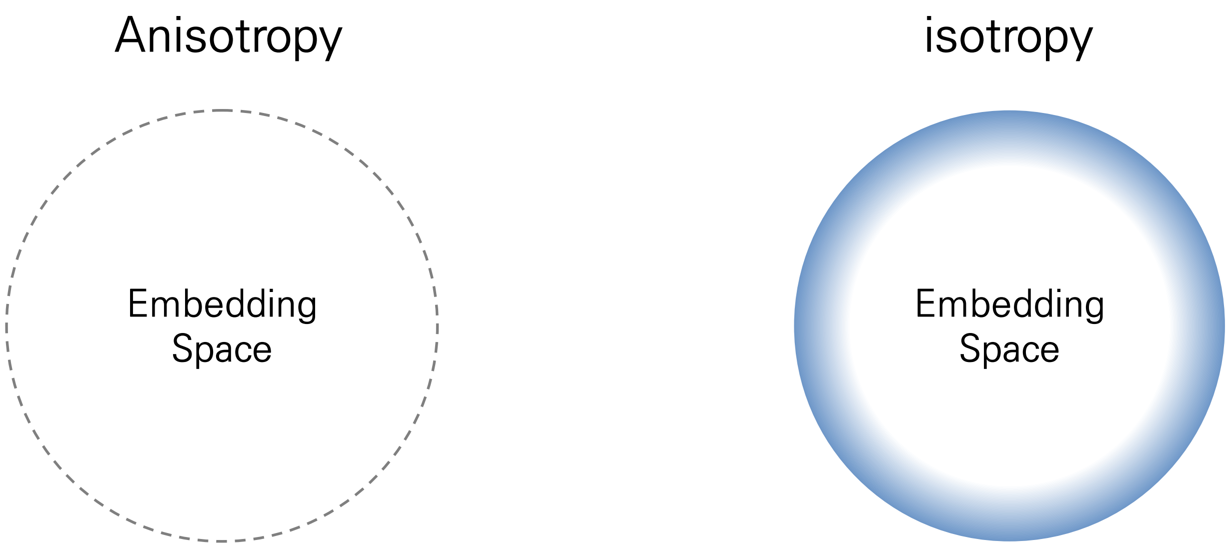
Wang, T., & Isola, P. (2020, November). Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In International Conference on Machine Learning (pp. 9929-9939). PMLR.

Contrastive learning for sentence embedding

SimCSE: Simple Contrastive Learning of Sentence Embeddings

❖ Alignment & Uniformity

- 언어모델의 상위 레이어로 갈수록 해당 레이어의 embedding 역시 Alignment와 Uniformity를 만족하지 못함
- 따라서 모델의 embedding이 한쪽에 몰리는 Anisotropy 문제가 발생
- SimCSE는 해당 문제를 Contrastive learning을 통해 개선함

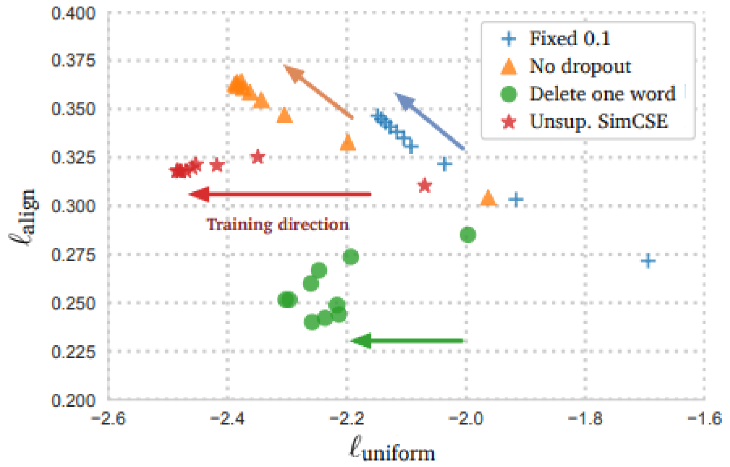


Contrastive learning for sentence embedding

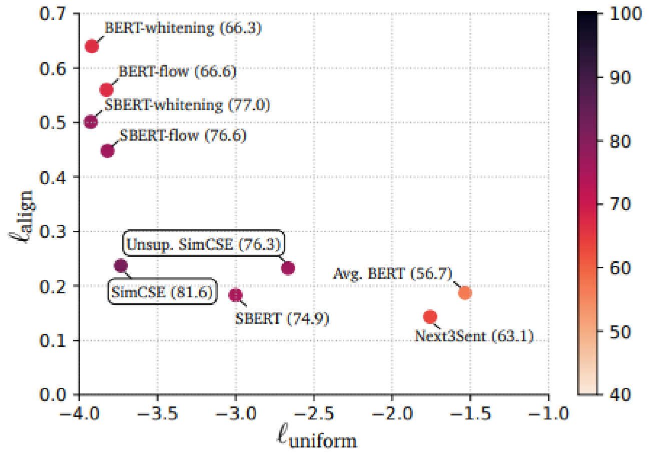
SimCSE: Simple Contrastive Learning of Sentence Embeddings

❖ 실험 결과

- 다른 data augmentation과는 달리, SimCSE는 align을 유지하면서 uniform이 줄어드는 것을 확인
- 또한 다른 모델들에 비해 alignment와 uniformity가 좋은 것을 실험적으로 증명



Augmentation 방법 별 align, uniform



모델 별 align, uniform

Contrastive learning for sentence embedding

❖ DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings

- 2022년 NAACL에 발표된 논문으로, 2023년 4월 25일 기준으로 총 35회 인용
- ELECTRA를 적용한 contrastive learning을 활용해 SimCSE를 이기고 SOTA 달성

DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings

Yung-Sung Chuang[†] Rumen Dangovski[†] Hongyin Luo[†] Yang Zhang[‡] Shiyu Chang[*]
Marin Soljačić[†] Shang-Wen Li[◇] Wen-tau Yih[◇] Yoon Kim[†] James Glass[†]
Massachusetts Institute of Technology[†] Meta AI[◇]
MIT-IBM Watson AI Lab[‡] UC Santa Barbara[*]
yungsung@mit.edu

Abstract

We propose DiffCSE, an unsupervised contrastive learning framework for learning sentence embeddings. DiffCSE learns sentence embeddings that are sensitive to the difference between the original sentence and an edited sentence, where the edited sentence is obtained by stochastically masking out the original sentence and then sampling from a masked language model. We show that DiffSCE is an instance of equivariant contrastive learning (Dangovski et al., 2021), which generalizes contrastive learning and learns representations that are insensitive to certain types of augmentations and sensitive to other "harm-

be more similar to one another than negative pairs. While different data augmentations (random cropping, color jitter, rotations, etc.) have been found to be crucial for pretraining vision models (Chen et al., 2020), such augmentations have generally been unsuccessful when applied to contrastive learning of sentence embeddings. Indeed, Gao et al. (2021) find that constructing positive pairs via a simple dropout-based augmentation works much better than more complex augmentations such as word deletions or replacements based on synonyms or masked language models. This is perhaps unsurprising in hindsight; while the training objective in contrastive learning encourages representations

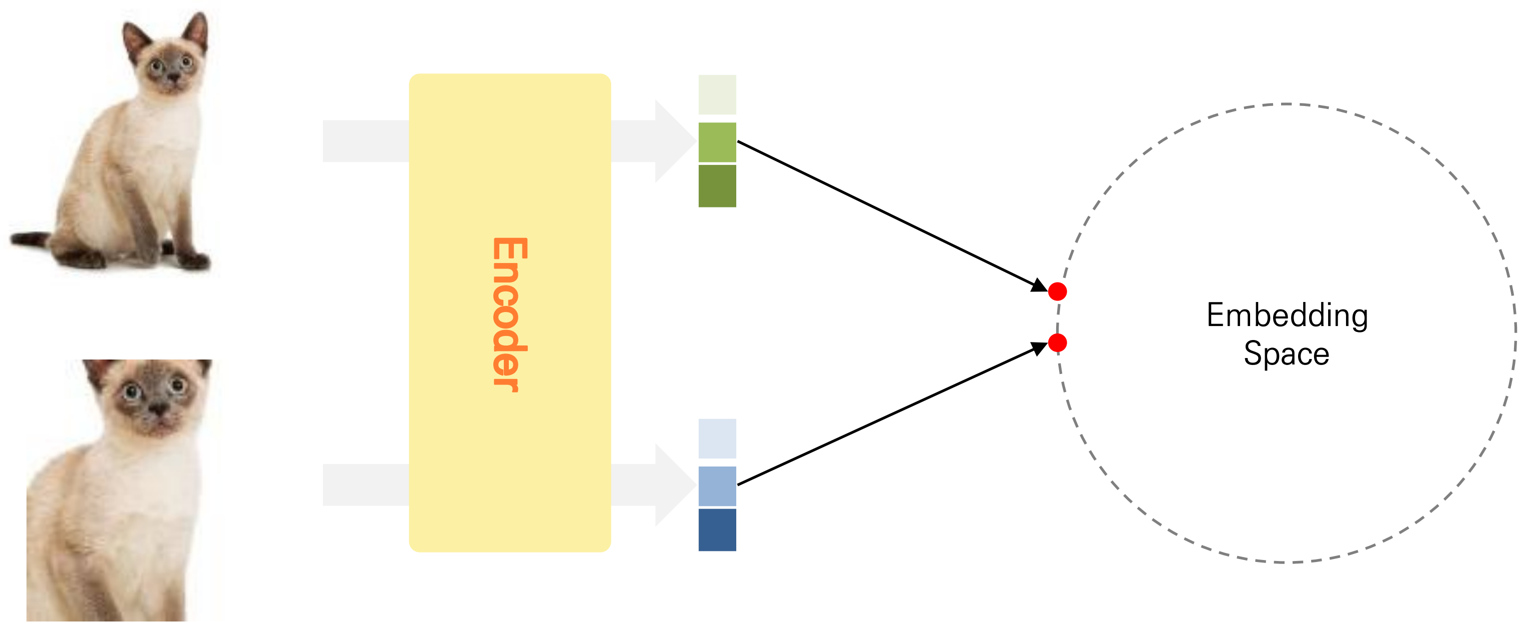
Chuang, Y. S., Dangovski, R., Luo, H., Zhang, Y., Chang, S., Soljačić, M., ... & Glass, J. (2022). DiffCSE: Difference-based contrastive learning for sentence embeddings. arXiv preprint arXiv:2204.10298.

Contrastive learning for sentence embedding

DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings

❖ 연구 배경

- Dropout을 사용한 contrastive learning의 성능은 당연히 좋음
- Contrastive learning의 목적은 embedding 차원에서 positive와 anchor가 최대한 유사해야 함 → Invariant

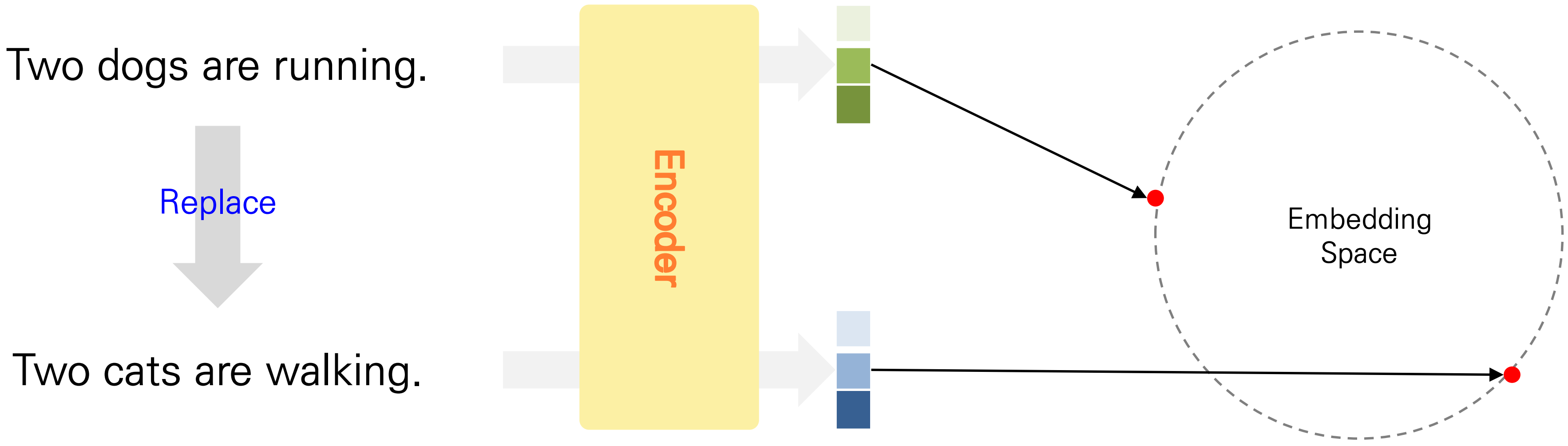


Contrastive learning for sentence embedding

DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings

❖ 연구 배경

- 그러나 data augmentation으로 내포한 의미가 달라진다면, 이 역시 반영할 수 있어야 함
- 즉, 좋은 Sentence embedding이란 의미를 바꾸는 data augmentation에 대해서는 Invariant하지 않아도 됨

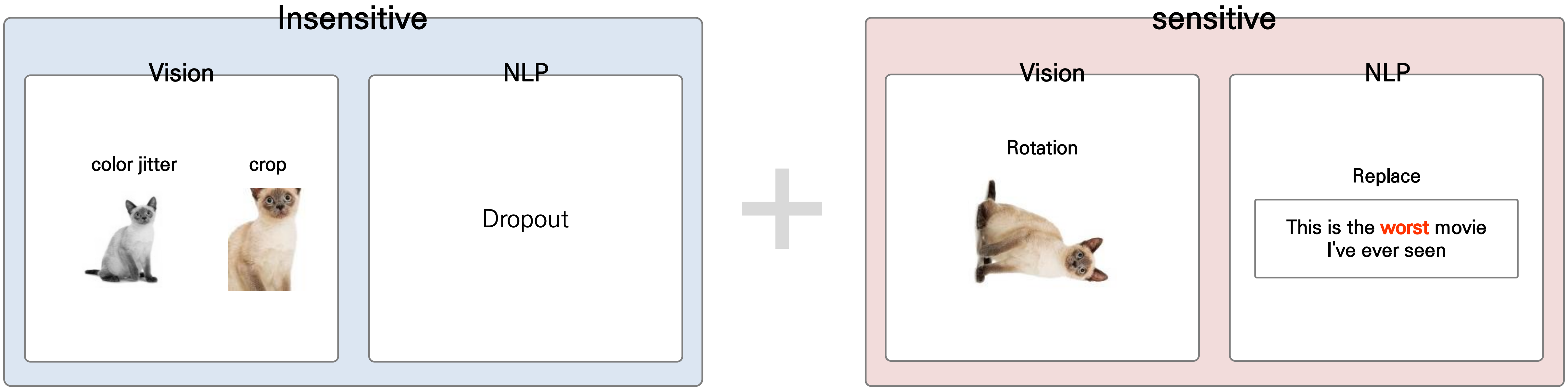


Contrastive learning for sentence embedding

DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings

❖ Equivariant Contrastive Learning

- Data augmentation 기법에는 sensitive와 insensitive 방법으로 구분할 수 있음
- 두 방법을 적절히 섞어서 학습을 진행하면, 좋은 Embedding을 찾을 수 있음



Dangovski, R., Jing, L., Loh, C., Han, S., Srivastava, A., Cheung, B., ... & Soljačić, M. (2021). Equivariant contrastive learning. arXiv preprint arXiv:2111.00899.

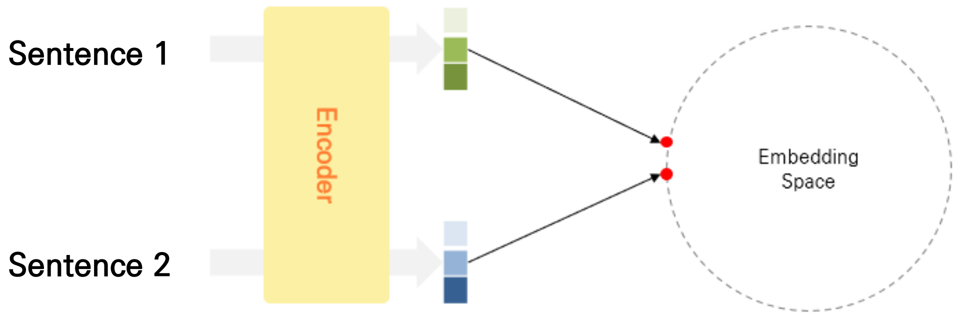
Contrastive learning for sentence embedding

DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings

❖ Equivariant Contrastive Learning

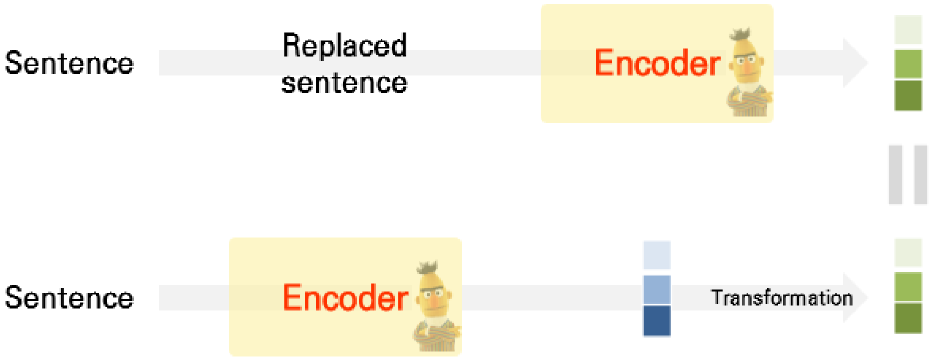
- 이 때, Insensitive한 data augmentation은 invariant하게 학습
- Sensitive한 data augmentation은 equivariant하게 학습

Invariant



Equivariant

$$f(T(x)) = T'(f(x))$$



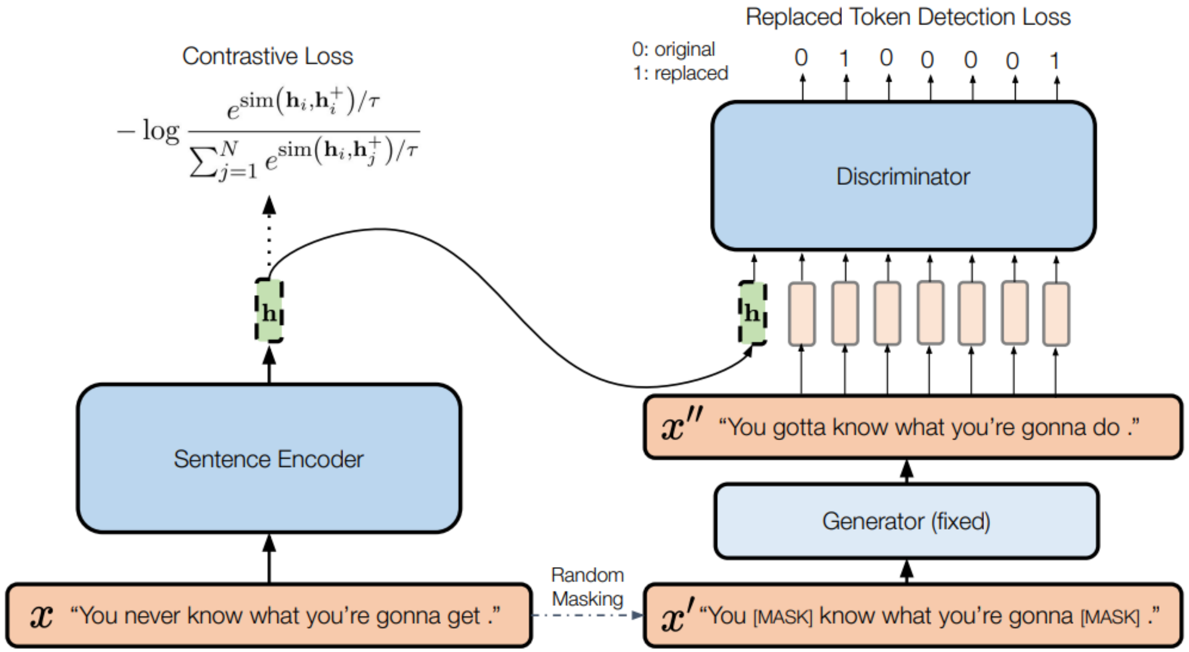
Dangovski, R., Jing, L., Loh, C., Han, S., Srivastava, A., Cheung, B., ... & Soljačić, M. (2021). Equivariant contrastive learning. arXiv preprint arXiv:2111.00899.

Contrastive learning for sentence embedding

DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings

❖ DiffCSE

- DiffCSE의 전체적인 구조는 sentence encoder, generator, discriminator로 구성
- 크게 Contrastive loss와 Replaced Token Detection loss를 활용하는 구조로 구분

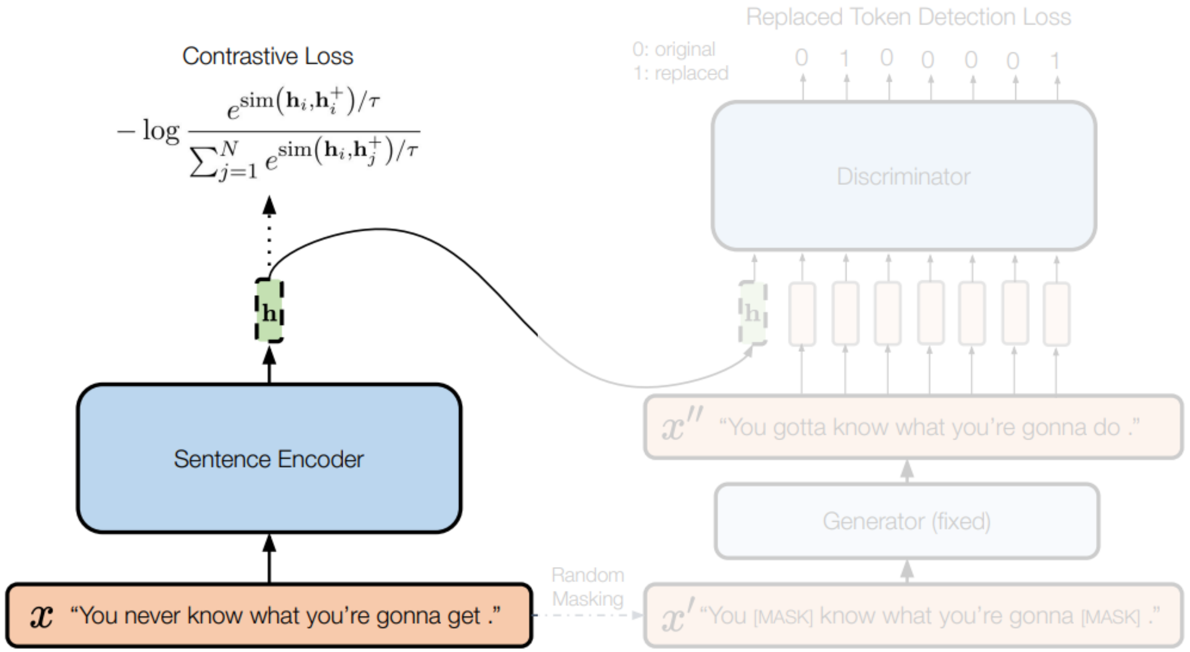


Contrastive learning for sentence embedding

DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings

❖ DiffCSE

- SimCSE와 동일한 Contrastive learning을 진행
- 해당 loss를 통해 Invariant를 만족하고자 함

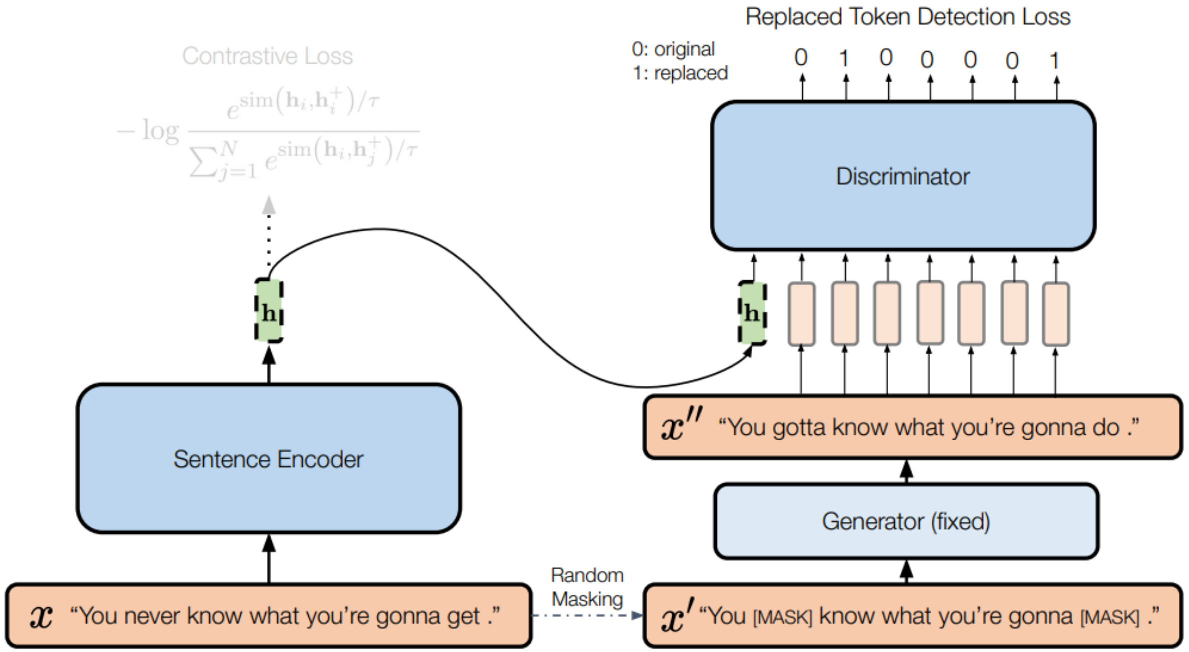


Contrastive learning for sentence embedding

DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings

❖ DiffCSE

- 언어모델인 ELECTRA의 학습 방법을 통해 Equivariant를 얻고자 함
- 원본 문장에 일부 토큰을 마스킹 후, Generator를 통해 다른 토큰으로 대체
- Discriminator는 Anchor의 representation vector를 조건으로 받아, 생성된 문장 중 어떤 토큰이 대체됐는지를 판별

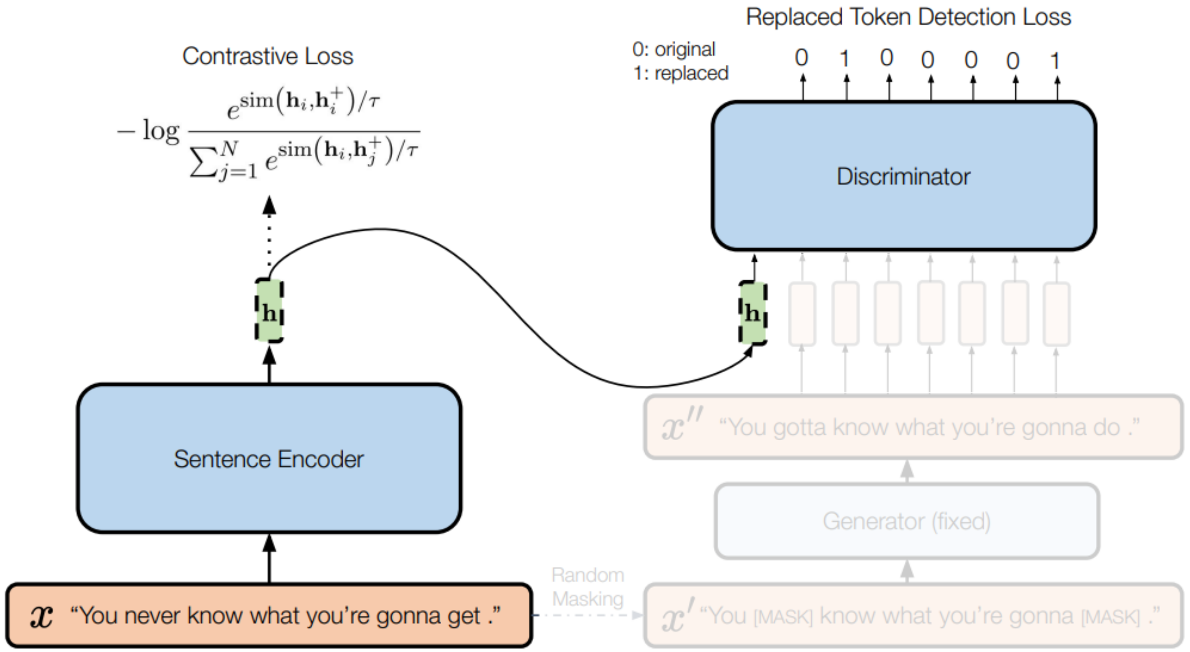


Contrastive learning for sentence embedding

DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings

❖ DiffCSE

- Backpropagation은 discriminator와 sentence encoder만 진행
- Condition으로 들어왔던 h를 통해 sentence encoder의 weight도 업데이트



Contrastive learning for sentence embedding

DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings

❖ 실험 결과

- Sentence encoder와 discriminator는 BERT/RoBERTa를 사용, generator는 DistilBERT/DistilRoBERTa를 사용
- 제안하는 DiffCSE의 성능이 가장 좋음을 실험적으로 증명함

Model	STS12	STS13	STS14	STS15	STS16	STS-B	SICK-R	Avg.
GloVe embeddings (avg.)♣	55.14	70.66	59.73	68.25	63.66	58.02	53.76	61.32
$BERT_{base}$ (first-last avg.)◇	39.70	59.38	49.67	66.03	66.19	53.87	62.06	56.70
$BERT_{base}$-flow◇	58.40	67.10	60.85	75.16	71.22	68.66	64.47	66.55
$BERT_{base}$-whitening◇	57.83	66.90	60.90	75.08	71.31	68.24	63.73	66.28
IS-$BERT_{base}$ ♡	56.77	69.24	61.21	75.23	70.16	69.21	64.25	66.58
CMLM-$BERT_{base}$ ♠ (1TB data)	58.20	61.07	61.67	73.32	74.88	76.60	64.80	67.22
CT-$BERT_{base}$ ◇	61.63	76.80	68.47	77.50	76.48	74.31	69.19	72.05
SG-OPT-$BERT_{base}$ †	66.84	80.13	71.23	81.56	77.17	77.23	68.16	74.62
SimCSE-$BERT_{base}$ ◇	68.40	82.41	74.38	80.91	78.56	76.85	**72.23**	76.25
∗ SimCSE-$BERT_{base}$(reproduce)	70.82	82.24	73.25	81.38	77.06	77.24	71.16	76.16
∗ DiffCSE-$BERT_{base}$	**72.28**	**84.43**	**76.47**	**83.90**	**80.54**	**80.59**	71.23	**78.49**
$RoBERTa_{base}$ (first-last avg.)◇	40.88	58.74	49.07	65.63	61.48	58.55	61.63	56.57
$RoBERTa_{base}$-whitening◇	46.99	63.24	57.23	71.36	68.99	61.36	62.91	61.73
DeCLUTR-$RoBERTa_{base}$ ◇	52.41	75.19	65.52	77.12	78.63	72.41	68.62	69.99
SimCSE-$RoBERTa_{base}$ ◇	**70.16**	81.77	73.24	81.36	80.65	80.22	68.56	76.57
∗ SimCSE-$RoBERTa_{base}$(reproduce)	68.60	81.36	73.16	81.61	80.76	80.58	68.83	76.41
∗ DiffCSE-$RoBERTa_{base}$	70.05	**83.43**	**75.49**	**82.81**	**82.12**	**82.38**	**71.19**	**78.21**

Conclusion

❖ Summary

- 자연어 분야에서는 언어모델의 성능을 높이기 위해 sentence embedding을 활용
- 보다 좋은 sentence embedding을 학습하기 위해, contrastive learning을 적용하고자 함
- 하지만 이미지 데이터와는 달리, 자연어에 data augmentation을 적용하면 ancho의 의미가 크게 손상됨
- 이번 세미나에서는 anchor-positive data를 잘 구축해, 효과적으로 contrastive learning을 자연어 분야에 적용한 방법론들을 소개


DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations

DeCLUTR


SimCSE: Simple Contrastive Learning of Sentence Embeddings

SimCSE


DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings

DiffCSE

Thank you

Reference

- Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PMLR.
- Giorgi, J., Nitski, O., Bader, G.D., & Wang, B. (2020). DeCLUTR: Deep Contrastive Learning for Unsupervised Textual Representations. ArXiv, abs/2006.03659.
- Gao, T., Yao, X., & Chen, D. (2021). Simcse: Simple contrastive learning of sentence embeddings. arXiv preprint arXiv:2104.08821.
- Wang, T., & Isola, P. (2020, November). Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In International Conference on Machine Learning (pp. 9929-9939). PMLR.
- Chuang, Y. S., Dangovski, R., Luo, H., Zhang, Y., Chang, S., Soljačić, M., ... & Glass, J. (2022). DiffCSE: Difference-based contrastive learning for sentence embeddings. arXiv preprint arXiv:2204.10298.
- Dangovski, R., Jing, L., Loh, C., Han, S., Srivastava, A., Cheung, B., ... & Soljačić, M. (2021). Equivariant contrastive learning. arXiv preprint arXiv:2111.00899.
- SimCSE Paper review (https://www.youtube.com/watch?v=c6sW3hO81gg&feature=youtu.be)
- DiffCSE Paper review (https://www.youtube.com/watch?v=40vStqJ2k3w)
- SimCSE review velog (https://velog.io/@lm_minjin/%EB%85%BC%EB%AC%B8-%EB%A6%AC%EB%B7%B0-SimCSE-Simple-Contrastive-Learning-of-Sentence-Embeddings)
- DiffCSE review velog (https://velog.io/@gunny1254/DiffCSE-Difference-based-Contrastive-Learning-for-Sentence-Embeddings)