

[배경]

- (1) Contrastive Learning에 대한 연구는 Data Augmentation, Non-linear Transformation, 모델의 구조적 변화에 대해서 집중적으로 연구가 되고 있음
- (2) 그러나, Negative Sample들에 대한 연구는 매우 저조한 상태

[개요]

- (1) Negative Sample의 Quality는 Feature Representation에 큰 영향을 줄 수 있음.
 - 학습 시 Collapse현상을 방지 가능
 - False Negative가 Negative Sample로 취급된다면, 학습에 역효과를 줄 수 있음.
- (2) 학습 효율성 증가 가능
 - 너무 쉬운 Sample들은 GPU RAM만 차지할 뿐, 학습 개선에는 큰 영향을 주지 못함
 - ** 또한, 너무 쉬운 Sample들은 학습 수렴을 느리게 할 수 있음.

[Negative Sampler가 만족해야 할 조건]

- (1) Efficient: 적은 시간 및 Computation Cost (너무 쉬운 것은 배제, 기준과 동일한 것은 배제)
- (2) Effective: 적당히 어려운 것은 학습에 효과적이고, 수렴 속도 가속화
- (3) Stable: 데이터의 특성에 강건하고, Collapse가 일어나지 않아야 함
- (4) Data-independent: 데이터의 특이 정보에 의지하지 않아야 함. (일반화된 특징 학습)

[Negative Sampler의 4가지 구분]

Category	Subcategory	Representative Approaches
Static NS	Random NS	Random selection [Chopra <i>et al.</i> , 2005]
	Popularity-biased NS	Word2vec [Mikolov <i>et al.</i> , 2013]; PNS [Rendle and Freudenthaler, 2014]
Dynamic NS	Query-dependent DNS	DNS [Zhang <i>et al.</i> , 2013] [Chen <i>et al.</i> , 2018]; ANCE [Xiong <i>et al.</i> , 2020] PinSage [Ying <i>et al.</i> , 2018]; IS [Robinson <i>et al.</i> , 2020]
	Positive-dependent DNS	Max-sampling [Rao <i>et al.</i> , 2016]; GNEG [Zhang and Zweigenbaum, 2018] ϵ -Truncated UNS [Sun <i>et al.</i> , 2018]; TransEdge [Sun <i>et al.</i> , 2019]
	Hybrid DNS	FaceNet [Schroff <i>et al.</i> , 2015]; WAPR [Weston <i>et al.</i> , 2010] MoChi [Kalantidis <i>et al.</i> , 2020]; MixGCF [Huang <i>et al.</i> , 2021]
Adversarial NS	Discrete Sampling	IRGAN [Wang <i>et al.</i> , 2017]; GraphGAN [Wang <i>et al.</i> , 2018a] KBGAN [Cai and Wang, 2017]; KGPolicy [Wang <i>et al.</i> , 2020]
	Continuous Sampling	DAML [Duan <i>et al.</i> , 2018]; AdvIR [Park and Chang, 2019] HeGAN [Hu <i>et al.</i> , 2019]; AdCo [Hu <i>et al.</i> , 2021]
Efficient NS	In-Batch NS	SimCLR [Chen <i>et al.</i> , 2020]; PBG [Lerer <i>et al.</i> , 2019] ABS [Choi <i>et al.</i> , 2021]; GraphVite [Zhu <i>et al.</i> , 2019]
	Caching Mechanism	NSCaching [Zhang <i>et al.</i> , 2019]; MoCo [He <i>et al.</i> , 2020]

(1) Static Negative Sampling

- Random하게, 또는 모집단에서 각 데이터의 비율을 고려하여 Sampling
- 빠른 속도, 데이터가 많을수록 좋음
- 그러나, 정해진 Query에 대해 정해진 답만 뱉어서 학습에 한계가 있음

(2) Dynamic Negative Sampling

■ Positive와 구분하기 어려운 Hard Negative 데이터를 찾아내자.

- Example1: 모델이 오분류하는 데이터 == Hard Negative
- Example2: Score 기반 접근 (Hard Negative일수록 높은 Score를 책정하여 판단)

■ 3개의 유형으로 다시 분류 가능

1) Query-dependent DNS

- ✓ 학습할 때, 각 Iteration별로 Score가 높은 Sample들만 골라 학습
- ✓ Supervision이 없더라도, Query의 Embedding과 유사한 데이터들을 Hard Negative로 판단 가능.

2) Positive-dependent DNS

- ✓ Positive와 유사도를 최대화함으로써, 가장 유사한 Negative 선택
- ✓ Query-dependent와 차이는 Query에 의존하는지, Positive에 의존하는지 차이
- ✓ Labeled 상황을 가정한다면, Query(호랑이), Positive(호랑이), Negative(그 외)

3) Hybrid DNS

- ✓ 위 두 방법론을 결합
- ✓ Query를 기반으로 1차 필터링 후, Positive를 기반으로 2차 필터링
- ✓ Ex) FaceNet: Semi-hard Example들로 학습
- ✓ Ex) Mochi: Feature Level에서 Positive와 Hard Negative를 섞어 More hard Negative 생성

** 그 외에, Sample level에서 mixup을 하거나, Hard Negative끼리 섞는 연구도 존재

■ 해당 방법론들의 특징

- Hard Negative들은 학습에 빠른 수렴 및 Gradient 학습에 큰 도움을 줄 수 있지만, 학습 및 Score 산정에 오랜 시간이 걸릴 수 있다.
- Hard Negative는 False Negative가 될 수 있다. 이들의 경계는 명확하게 할 필요가 있으며, 아직 명확한 방법은 없다. (미리 정한 Range에서만 뽑는 방식 등을 활용)
- False Negative가 늘 문제가 된다. 오히려 이들을 학습할 수 있기에, 성능이 떨어질 수도 있음.

(3) Adversarial Negative Sampling

- GAN을 활용해서 Discriminator를 속일 수 있는 Hard Negative를 Generation하거나 Negative Sample의 분포를 추정
 - Discrete Sampling (NLP / Graph)와 Continuous Sampling(Image)으로 나뉨
 - Continuous Sampling은 Generation을 의미 (즉, Negative Sample을 Generation 하거나, Representation Network에서 동시에 Generation하는 연구도 존재)
- GAN을 통해 정말 고퀄의 Hard Negative를 만들 수 있을 것. 그러나 여전히, False Negative의 위험 존재
- 또한, 학습이 불안정하고, 학습 Cost가 크다는 단점도 존재

(4) Efficient Negative Sampling

- Negative Sampling을 어떻게 효율적으로 할까? 전체 데이터를 모두 활용할 수 있다면 좋을 것 같은데, 현실적으로 쉽지 않음.
 - In-Batch Negative Sampling: 배치 내에서 Negative Sample을 정의 (SimCLR)
 - Caching Mechanism: FIFO의 원리로 Memory Bank나, Queue를 활용 (MoCo)

[차후 연구방향]

- (1) False Negative를 완화하자.
- (2) Easy Negative와 Hard Negative를 동시에 활용하자
 - Easy Negative는 학습 안정성에 도움을 줄 수 있고, 초기 학습에도 큰 영향을 줄 수 있음
 - 쉬운 것부터 학습하는 Curriculum Learning 등의 Approach도 Good
- (3) Hard Negative의 양과 질에 대한 Trade-Off
 - Hard Negative가 많으면 좋지만, 너무 많으면 학습 비효율성 야기 가능
- (4) Negative Sampling을 배제하고 학습하는 것도 가능
 - 모든 데이터를 다 Negative로 보되, 가중치를 부여하는 방식으로 학습
 - Negative를 아예 정의하지 않고, Positive로만 학습하는 방법론도 존재 (BYOL)