

Zero-shot Recognition via Semantic Embeddings and Knowledge Graphs (CVPR 2018)

Jonghyun Lee

Seoul National University

Zero-shot Recognition

- Okapi : zebra-striped four legged animal with a brown torso and a deer-like face
- Deer, zebra가 학습 데이터에 포함되고 okapi는 없는 상황에서 Okapi와 deer, zebra의 relation을 통해 처음 본 okapi class도 잘 맞추도록 만들어보는 것이 목적

Okapi



Deer



Zebra



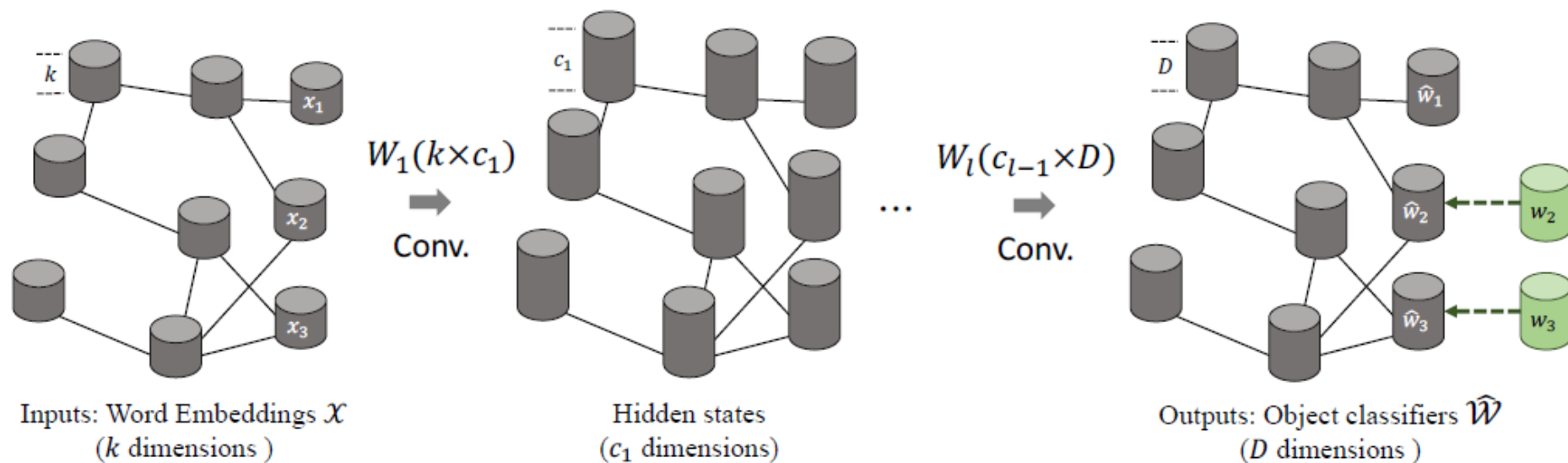
Zero-shot Recognition

- Training 중 한 번도 보지 못한 category를 맞추는 image classifier를 얻는 것이 목표
- 당시의 previous work들은 image의 feature와 category의 word embedding을 matching 시키는 방법을 고안
- Knowledge graph를 사용하는 방법도 존재
- 본 논문에서는 두 방법을 통합

GCN for Zero-shot Learning

- 총 n 개의 category
 - Image와 label이 같이 존재하는 m 개의 category를 포함
- Class의 word embedding과 그 relationship을 이용해 relationship graph 생성 가능
- 일반 GCN
 - Input : image features
 - Output : 각 feature가 속한 class
- GCN for zero-shot learning
 - Input : category들의 embedding vectors
 - Output : 각 category들의 classifier (a logistic regression model)

GCN for Zero-shot Learning



- Output : $\mathcal{W} = \{w_i\}_{i=1}^n$
- 각 w_i 들은 a logistic regression model의 weight
 - Image들의 features가 D-dim $\rightarrow w_i$ 도 D-dim

GCN for Zero-shot Learning

- 각 class를 word embedding한다. (ex. GloVe)
- 우리가 알고있는 m개의 class에 대해서 logistic regression classifier weight를 학습시킨다. (w_i 학습, $i = 1, \dots, m$)
- Class word embedding을 node로 relationship graph를 만들고 이를 GCN을 통과시켜 output을 내보낸다. (\widehat{w}_i 학습, $i = 1, \dots, n$)
- 그 output과 우리가 알고있는 m개의 class의 logistic regression weight간의 MSE를 줄여 둘의 space를 인접하도록 만든다.
 - Loss term : $\frac{1}{m} \sum_{i=1}^m L_{mse}(\widehat{w}_i, w_i)$
- Test할 때 GCN 통해 얻은 n개의 각 class의 logistic regression weight를 이용해 classificatio을 진행한다.

Experiments

- ImageNet에서 한 번도 보지 않은 class에 대한 classification 결과
- Training set : ImageNet2012 1K dataset
- WordNet의 sub-graph를 knowledge graph로 사용
- Base network : Inception-v1, ResNet-50

Experiments




- 2-hops가 가장 training set과 비슷함
- All이 가장 training set과 차이 큼
- Common label 존재하지 않음
- Metric : top-k accuracy
- 2-hops에서는 약 20% point차이로 앞지름

Test Set	Model	ConvNets	Hit@k (%)				
			1	2	5	10	20
2-hops	ConSE [4]	Inception-v1	8.3	12.9	21.8	30.9	41.7
	ConSE(us)	Inception-v1	12.4	18.4	25.3	28.5	31.8
	SYNC [4]	Inception-v1	10.5	17.7	28.6	40.1	52.0
	EXEM [5]	Inception-v1	12.5	19.5	32.3	43.7	55.2
	Ours	Inception-v1	18.5	31.3	50.1	62.4	72.0
	Ours	ResNet-50	19.8	33.3	53.2	65.4	74.6
3-hops	ConSE [4]	Inception-v1	2.6	4.1	7.3	11.1	16.4
	ConSE(us)	Inception-v1	3.2	4.9	7.6	9.7	11.4
	SYNC [4]	Inception-v1	2.9	4.9	9.2	14.2	20.9
	EXEM [5]	Inception-v1	3.6	5.9	10.7	16.1	23.1
	Ours	Inception-v1	3.8	6.9	13.1	18.8	26.0
	Ours	ResNet-50	4.1	7.5	14.2	20.2	27.7
All	ConSE [4]	Inception-v1	1.3	2.1	3.8	5.8	8.7
	ConSE(us)	Inception-v1	1.5	2.2	3.6	4.6	5.7
	SYNC [4]	Inception-v1	1.4	2.4	4.5	7.1	10.9
	EXEM [5]	Inception-v1	1.8	2.9	5.3	8.2	12.2
	Ours	Inception-v1	1.7	3.0	5.8	8.4	11.8
	Ours	ResNet-50	1.8	3.3	6.3	9.1	12.7

Experiments

- Word Embedding마다 성능 차이 적음
 - 다른 baseline은 차이 큼
- Output label도 다른 baseline에 비해 train label에 bias 적음

Model	Word Embedding	Hit@ k (%)				
		1	2	5	10	20
[53]	GloVe	7.8	11.5	17.2	21.2	25.6
Ours	GloVe	18.5	31.3	50.1	62.4	72.0
[53]	FastText	9.8	16.4	27.8	37.6	48.4
Ours	FastText	18.7	30.8	49.6	62.0	71.5
[53]	GoogleNews	13.0	20.6	33.5	44.1	55.2
Ours	GoogleNews	18.3	31.6	51.1	63.4	73.0

Test Image	ConSE (10)	Ours
	panthera tigris(train) tiger cat (train) felis onca (train) leopard (train) tiger shark (train)	tigress (test) bengal tiger (test) panthera tigris (train) tiger cub (test) tiger cat (train)
	rock beauty (train) ringlet (train) flagpole (train) large slipper (test) yellow slipper (train)	butterfly fish (test) rock beauty (train) damsel fish (test) atoll (test) barrier reef (test)
	tractor (train) reaper (train) thresher (train) trailer truck (train) motortruck (test)	tracked vehicle (test) tractor (train) propelled vehicle (test) reaper (train) forklift (train)

Conclusion

- Zero-shot recognition을 word embedding과 GCN with knowledge graph를 연관시켜 novel한 방법론을 제시
- Zero-shot recognition task에서 state-of-the-art 성능을 큰 차이로 얻어냄