



DMAIS Lab Seminar

<ConvE>

2024-07-09

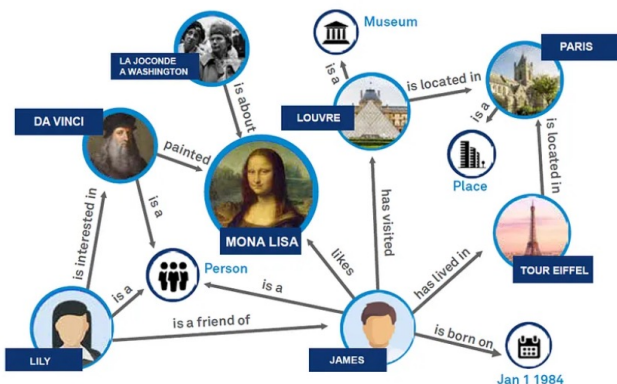
presenter : Moon Soo Ho

Data Mining And Intelligence Systems (DMAIS) Lab
School of Computer Science and Engineering
Chung-Ang University

- ☐ Main interest
- ☐ Problems with previous models
- ☐ Architecture
- ☐ Experiments
- ☐ Conclusion

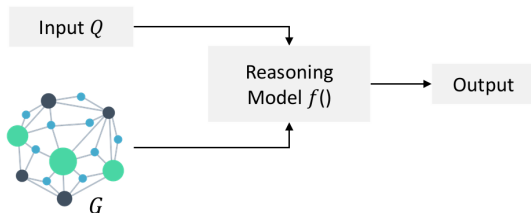
Knowledge Graph(KG;지식 그래프)란

- 지식 그래프는 실세계 엔티티 간의 관계와 상호작용을 특정 edge(relation)로 정보의 연결성과 의미를 담은 구조화된 방향 그래프
- 지식 그래프는 검색 엔진, 추천 시스템, 자연어 처리, 사실 확인등에 사용이 가능
- Triplet으로 entity-relation-entity(h, r, t)를 표현함 e.g. (Da Vinci, painted, Mona Lisa)



KG의 작동 방법

- Q : 쿼리 그래프, 자연어 질문 등 G 를 통해 추론하고자 하는 것
- G : 지식 그래프(일반 그래프와의 차이점은 edge의 종류가 여러가지이다)
- f : 추론 함수



본 논문에서 해결하고자 하는 문제

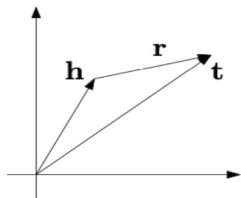
- KG embedding : entity와 relation을 저차원으로 임베딩하여 link prediction, similarity check 등의 task를 수행함.
주로 h, r, t 를 임베딩 먼저 하고 이 임베딩 벡터들의 관계를 비교, 학습함.

Problem with previous models

문제점

- (a) 얕은 추론 깊이 e.g. TransE
- (b) 기존 commonly used dataset 자체의 문제점(inverse relation)
- (c) 파라미터 개수 제약 e.g. RESCAL, DistMult
- (d) multi-layer embedding의 overfitting 문제

(a)



좌측 그림은 TransE의 학습 방법이다. $\mathbf{h} + \mathbf{r} \approx \mathbf{t}$ 의 관계를 만족시키도록 학습하기 때문에 연산이 가볍고 1-to-1 relation 처리가 가능하지만 1-to-N, N-to-1, N-to-N 같은 relation에 적용할 수 없다.

(b) Train set : (Dad, Fatherof, Son)

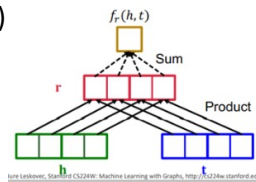


Test set : (Son, Childof, Dad)

Inverse relation은 좌측 예시와 같이 역의 관계를 갖는 것을 뜻한다. Train set에 있는 데이터는 학습과정에서 보이면 안되지만 모델의 학습 과정에서 inverse relation까지 함께 학습할 위험이 있다. 이는 memorization으로 인한 artificially high accuracy로 이어질 수 있다.

Problem with previous models

(c)



$$\mathbf{h}^T \text{diag}(\mathbf{r}) \mathbf{t}$$

DistMult

DistMult는 h, r, t 의 tri-linear dot product의 결과를 학습한다.

하지만 연산 구조상 파라미터의 개수는 임베딩 차원과 반드시 같아야 한다(feature 1개당 parameter 1개). 즉 모델의 성능을 높이기 위해 메모리 과잉을 감수해야한다.

(d) [\(Nickel et al. 2016\)](#)에서 multi-layer를 사용하는 KG embedding model은 overfitting에 취약함을 주장하였다.

Problem with previous models

해결(ConvE)

- (a) 얇은 추론 깊이

벡터 합성이 아닌 deep networks를 이용해 더 많은 representations를 학습할 수 있음

- (b) 파라미터 개수 제약

임베딩 생성에 multi layer를 적용해 임베딩 차원과 파라미터의 개수가 독립적이게 함
결론적으로 DistMult보다 17배 적은 파라미터를 이용한 ConvE가 비슷한 성능냄


- (c) 기존 commonly used dataset 자체의 문제점(inverse relation)


필요에 따라 inverse relation leakage의 위험이 제거된 transformed dataset(FB15k-237, WN18RR)을 사용하기를 권고함

- (d) multi-layer embedding의 overfitting문제

이 문제를 제기한 논문에서 이미 해결책을 내놓았고(regularization, cross-validation, using not too complex model)
본 논문에서는 drop-out을 활용함

*Background

 : concatenation border(where interactions occur)

 : Filter(size = k or k*k)

•2D convolution을 사용하는 이유

$$([a \ a \ a]; [b \ b \ b]) = [a \ \boxed{a \ a} \mid b \ b \ b]$$

두 벡터의 interaction 범위가 1D에서는 제한적이고 오직 k에만 의존함

$$\left(\begin{bmatrix} a & a & a \\ a & a & a \end{bmatrix}; \begin{bmatrix} b & b & b \\ b & b & b \end{bmatrix} \right) = \begin{bmatrix} a & a & a \\ \boxed{a} \mid \boxed{a} & a \\ b & b & b \\ \boxed{b} \mid \boxed{b} & b \end{bmatrix} m \times n$$

반면 1D 벡터를 2D로 reshape하고 concatenate하면 interaction 범위는 k, n에 의존함

$$\begin{bmatrix} a & a & a \\ \boxed{b} & \boxed{b} & b \\ \boxed{a} & \boxed{a} & a \\ b & b & b \end{bmatrix}$$

이 원리를 확장하면 interaction 범위를 k, m, n에 의존하게 할 수 있음

이로써 2D Convolution은 다양한 representative를 학습할 수 있음

*BackGround

•기호 및 scoring function

KG :

$$\mathcal{G} = \{(s, r, o)\} \subseteq \mathcal{E} \times \mathcal{R} \times \mathcal{E}$$

$$r \in \mathcal{R} \quad \leftarrow \text{relation}$$

$$s, o \in \mathcal{E} \quad \leftarrow \text{entity}$$

$$\mathbf{e}_s, \mathbf{e}_o \in \mathbb{R}^k \quad \leftarrow \text{embedding}$$

Scoring function :

$$\psi : \mathcal{E} \times \mathcal{R} \times \mathcal{E} \mapsto \mathbb{R}$$

$$x = (s, r, o) \quad \leftarrow \text{input(triple)}$$

$$\psi(x) \in \mathbb{R}$$

$$\psi(s, r, o) = \psi_r(\mathbf{e}_s, \mathbf{e}_o) \in \mathbb{R}$$

특정 relation에 대한 scoring function이 다름

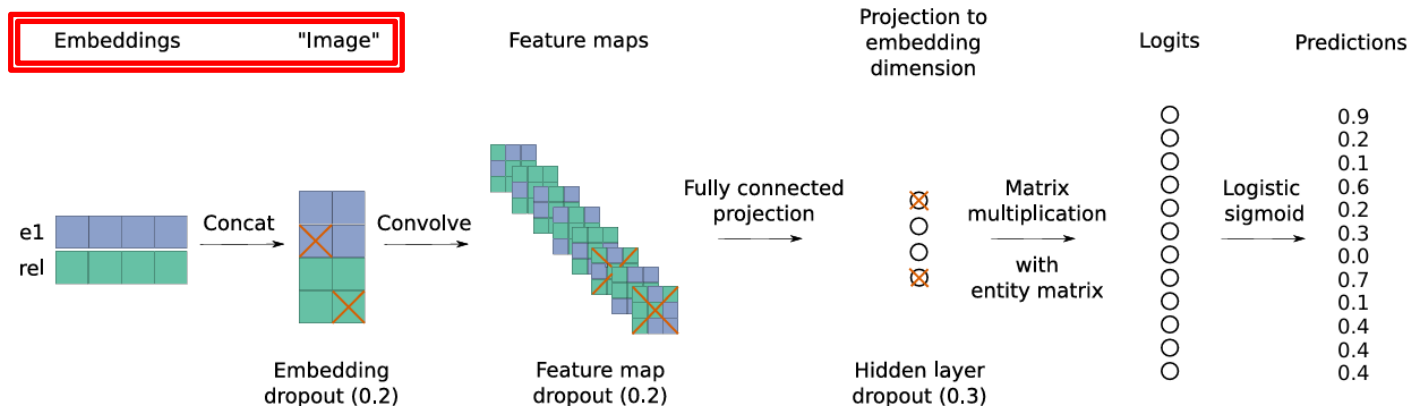
임베딩 된 entity를 입력으로 score(s와 o가 r을 통해 얼마나 연관성이 있는지)를 계산함

Complete form of the scoring function

$$\psi_r(\mathbf{e}_s, \mathbf{e}_o) = f(\text{vec}(f([\overline{\mathbf{e}}_s; \overline{\mathbf{r}}_r] * \omega))\mathbf{W})\mathbf{e}_o$$

Architecture

$$\psi_r(\mathbf{e}_s, \mathbf{e}_o) = f(\text{vec}(f([\underline{\mathbf{e}}_s; \underline{\mathbf{r}}_r] * \omega))\mathbf{W})\mathbf{e}_o$$

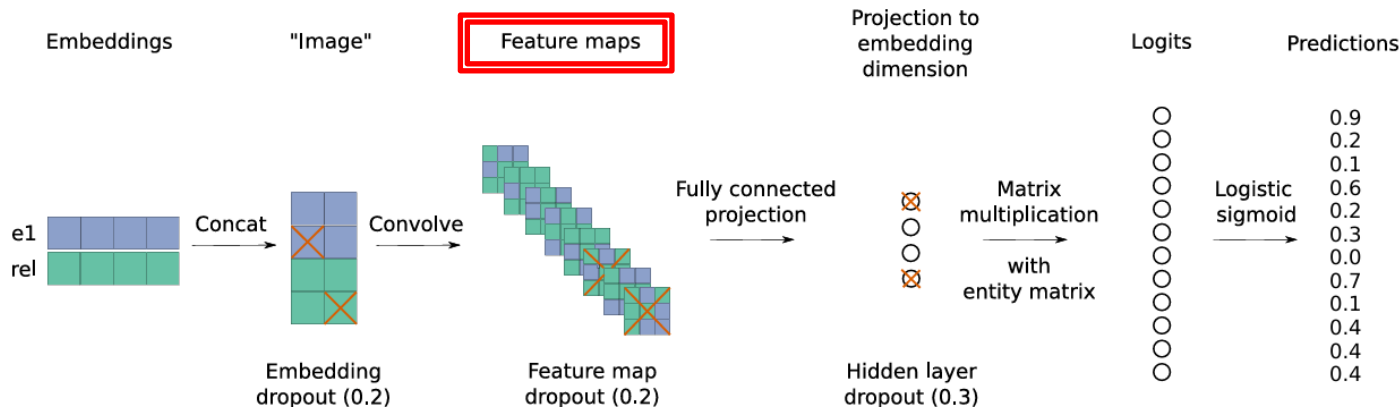


s와 r의 임베딩인 $\mathbf{e}_s, \mathbf{r}_r \in \mathbb{R}^k$ 을 계산한다.

이후 같은 크기의 두 벡터를 image화 시키고 concatenate한다. $\underline{\mathbf{e}}_s, \underline{\mathbf{r}}_r \in \mathbb{R}^{k_w \times k_h} \quad k = k_w k_h$

Architecture

$$\psi_r(\mathbf{e}_s, \mathbf{e}_o) = f(\text{vec}(f([\overline{\mathbf{e}}_s; \overline{\mathbf{r}}_r] * \underline{\omega}))\mathbf{W})\mathbf{e}_o$$



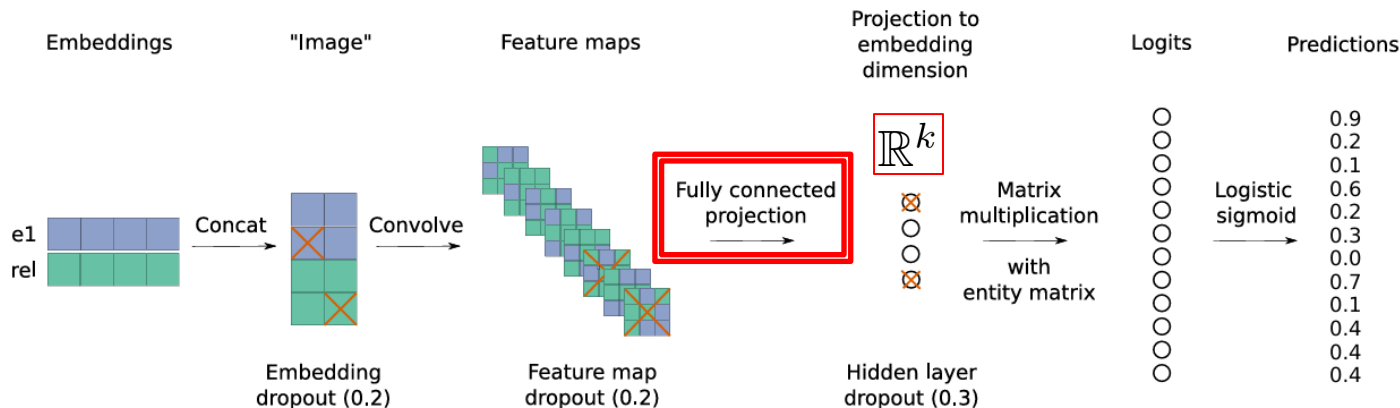
여러개의 필터(ω)를 통해 2D Convolution해서 필터의 개수만큼의 feature map 생성.

이렇게 생성된 feature map tensor를 $\mathcal{T} \in \mathbb{R}^{c \times m \times n}$ 로 표기함(c = number of filters).

Architecture

$$\psi_r(\mathbf{e}_s, \mathbf{e}_o) = f(\text{vec}(f([\overline{\mathbf{e}}_s; \overline{\mathbf{r}}_r] * \omega)) \mathbf{W}) \mathbf{e}_o$$

$$\text{vec}(\mathcal{T}) \in \mathbb{R}^{cmn} \quad \mathbf{W} \in \mathbb{R}^{cmn \times k}$$

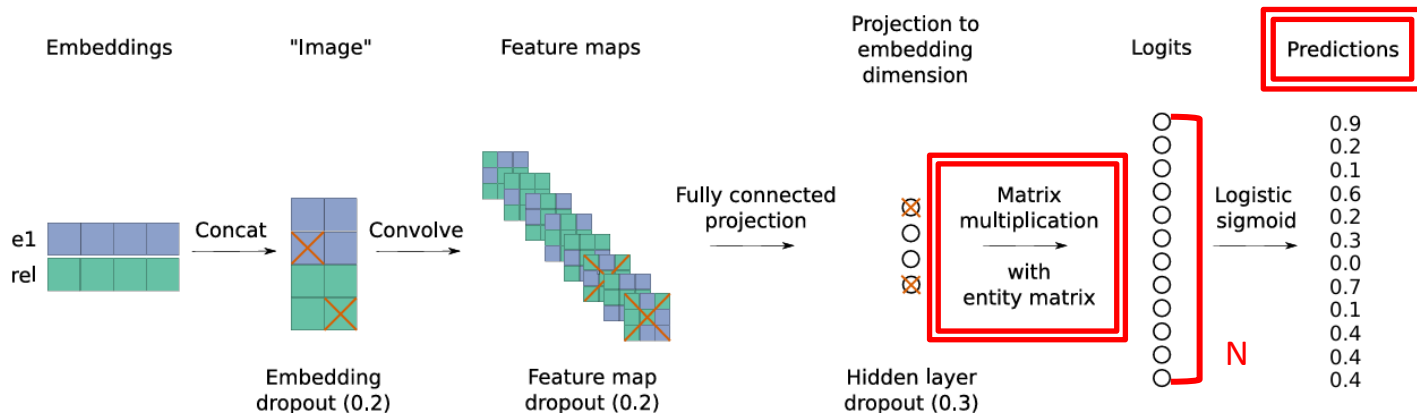


텐서를 하나의 벡터로 압축시키고 fully connected projection을 통해 기존 임베딩들과 같은 차원인 k차원으로 변환시킨다.

동일 차원으로 변환하는 이유는 이후에 object entity와 유사도 측정(\cdot)을 해줘야하기 때문이다.

Architecture

$$\psi_r(\mathbf{e}_s, \mathbf{e}_o) = f(\text{vec}(f([\overline{\mathbf{e}}_s; \overline{\mathbf{r}}_r] * \omega)) \mathbf{W}) \mathbf{e}_o$$



위의 scoring function 수식과는 다르게 연산의 효율성을 극적으로 끌어올리기 위해 여러 object entity와 dot product를 해준다(1-N scoring).

이를 통해 입력에 사용된 subject와 relation에 대한 N 개의 object의 점수가 나오고 이를 sigmoid에 통과시켜주면 normalize된 점수가 나온다.

$$p = \sigma(\psi_r(\mathbf{e}_s, \mathbf{e}_o))$$

binary-cross entropy loss :

$$\mathcal{L}(p, t) = -\frac{1}{N} \sum_i (t_i \cdot \log(p_i) + (1 - t_i) \cdot \log(1 - p_i))$$

$$t \in \mathbb{R}^{1 \times 1} \quad (1-1 \text{ scoring})$$

$$t \in \mathbb{R}^{1 \times N} \quad (1-N \text{ scoring})$$

FB15k에 ConvE를 적용할 때 1-1 scoring은 3시간 이상, 1-N scoring은 1분 이하의 시간이 소요됨
따라서 ConvE는 1-N scoring을 사용함

평가 지표

- **MR**
옳게 예측한 entity의 순위(rank) 의 평균
- **MRR**
예측한 아이템의 순위의 역수
- **Hits@k(%)**
k개의 추천 중 선호 아이템이 있는지 측정(rank 무시)

higher, the better : MRR, Hits

lower, the better : MR

Table 3: Link prediction results for WN18 and FB15k

	WN18					FB15k				
	MR	MRR	@ 10	Hits @ 3	@ 1	MR	MRR	@ 10	Hits @ 3	@ 1
DistMult (Yang et al. 2015)	902	.822	.936	.914	.728	97	.654	.824	.733	.546
ComplEx (Trouillon et al. 2016)	–	.941	.947	.936	.936	–	.692	.840	.759	.599
Gaifman (Niepert 2016)	352	–	.939	–	.761	75	–	.842	–	.692
ANALOGY (Liu, Wu, and Yang 2017)	–	.942	.947	.944	.939	–	.725	.854	.785	.646
R-GCN (Schlichtkrull et al. 2017)	–	.814	.964	.929	.697	–	.696	.842	.760	.601
ConvE	374	.943	.956	.946	.935	51	.657	.831	.723	.558
Inverse Model	740	.963	.964	.964	.953	2501	.660	.660	.659	.658

data leakage가 있는
데이터셋에서 최고의
성능을 못보여줌

Experiments

Table 4: Link prediction results for WN18RR and FB15k-237

	WN18RR					FB15k-237				
	MR	MRR	Hits			MR	MRR	Hits		
			@ 10	@ 3	@ 1			@ 10	@ 3	@ 1
DistMult (Yang et al. 2015)	5110	.43	.49	.44	.39	254	.241	.419	.263	.155
ComplEx (Trouillon et al. 2016)	5261	.44	.51	.46	.41	339	.247	.428	.275	.158
R-GCN (Schlichtkrull et al. 2017)	—	—	—	—	—	—	.248	.417	.258	.153
ConvE	4187	.43	.52	.44	.40	244	.325	.501	.356	.237
Inverse Model	13526	.35	.35	.35	.35	7030	.010	.014	.011	.007

: data leakage가 제거된 데이터셋에서 고성능을 보여줌

Table 2: Parameter scaling of DistMult vs ConvE.

Model	Param. count	Emb. size	MRR	@ 10	Hits @ 3	@ 1
DistMult	1.89M	128	.23	.41	.25	.15
DistMult	0.95M	64	.22	.39	.25	.14
DistMult	0.23M	16	.16	.31	.17	.09
ConvE	5.05M	200	.32	.49	.35	.23
ConvE	1.89M	96	.32	.49	.35	.23
ConvE	0.95M	54	.30	.46	.33	.22
ConvE	0.46M	28	.28	.43	.30	.20
ConvE	0.23M	14	.26	.40	.28	.19

: DistMult와 비교했을 때 parameter 효율성 증대

Problems

- 다양한 관계 학습이 어려움
- 파라미터 개수의 제약으로 KG의 다양한 특성 학습에 제한
- Benchmark dataset 자체의 inverse relation 문제점
- Overfitting

Solutions with ConvE

- 1 layer projection과 2D convolution을 이용해 다채로운 특성 합성 가능
- 적은 수의 파라미터로도 고성능을 보여줌
- Inverse relation 문제를 제거한 데이터셋을 새로 제시하고 향후 연구에서도 사용하기를 권장함
- Drop-out으로 overfitting을 해결함

Evaluation

- Inverse relation이 제거된 데이터셋으로 학습한 결과, 기존의 모델들에 비해 고성능의 link prediction을 수행함
- 파라미터의 개수를 대폭 감소시켜 메모리 효율성을 증진시킴

Networks Are Ubiquitous: Building a Better World through Network Knowledge!

Thank You!



Contact: Sooho Moon (Email: moonwalk725@cau.ac.kr)