Translating Embeddings for Modeling Multi-relational Data

Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran(2013)

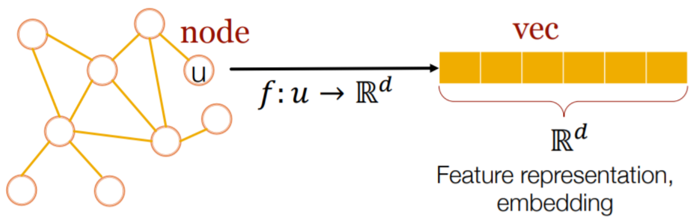
2021 DSAIL Winter Internship

2022.01.18. Daeyoung Kim

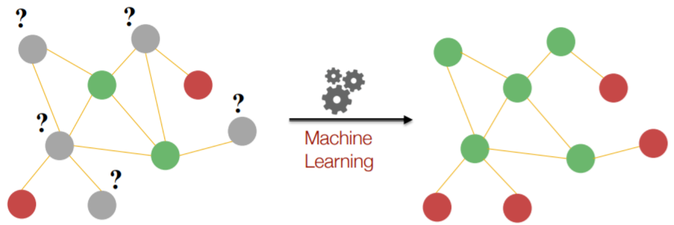
Index

Background

In previous papers…



Feature learning



Label classification

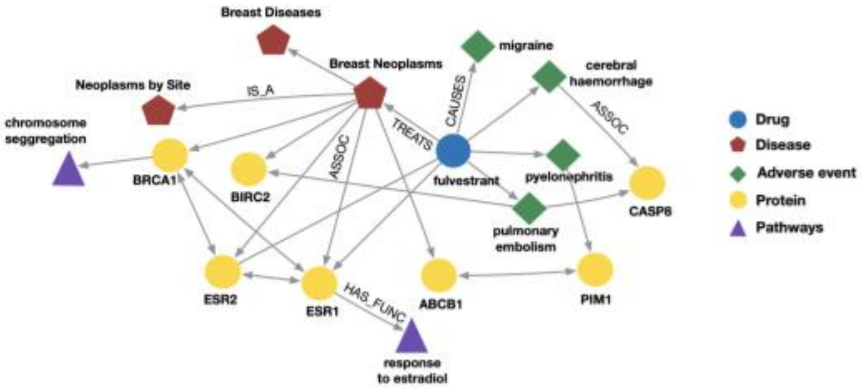
Applied machine learning for **homogeneous** graphs

(ex. DeepWalk, Node2Vec)

Q. How to handle heterogeneous graphs?

Background

Heterogeneous graph



$$G = (V, E, R, T)$$

Nodes($\boldsymbol{V}$) can have **various** types($\boldsymbol{T}$)

Edges($\boldsymbol{E}$) can have types

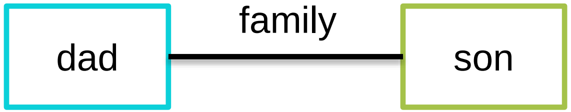
Relations($\boldsymbol{R}$) are defined

Knowledge graph : example of heterogeneous graph

Background

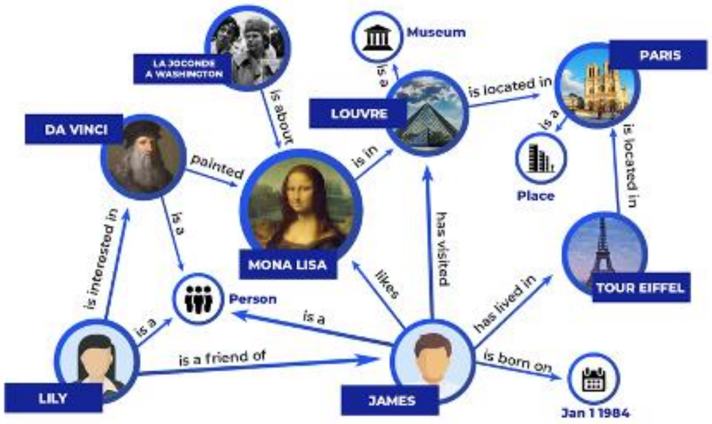
Knowledge graph

Knowledge Base(KB): 지적 활동, 지식, 사실, 규칙 등이 저장되어 있는 데이터 베이스
개체(entity)와 개체 간의 관계(relationship)으로 구성



Knowledge Graph(**KG**): 정보와 지식들을 상호 연결한 그래프
- 그래프의 node와 edge가 각각 entity와 relationship에 대응
- 다양한 정보들의 관계를 나타낼 수 있음

주로 link prediction에 대한 연구 진행



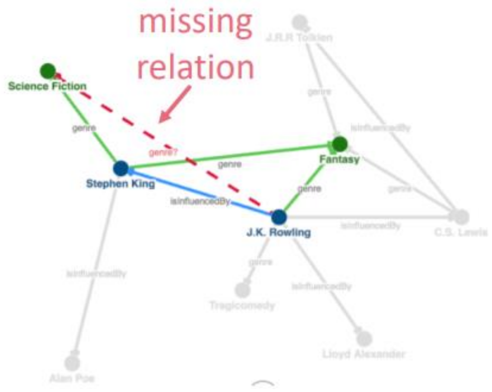
Background

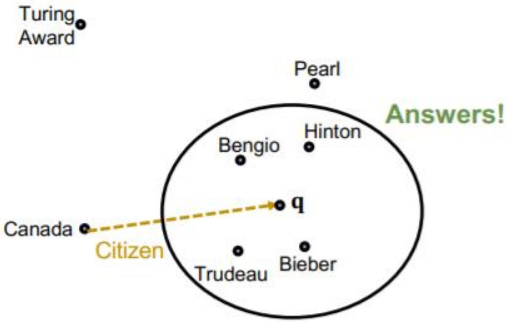
Knowledge graph

지식 그래프의 node와 edge가 많으므로,
저차원 공간에 임베딩 -> 쉽게 표현, 학습

지식 그래프의 local, global 연결 패턴을 학습해서
node와 node 사이의 missing link를 예측

각각의 entity에 대해 relation이 주어지면 **정답**에
해당하는 entity와 연결하는 방식을 이용





Introduction

Abstract

- Problem

  Embedding entities, relationships of multi-relational data in low-dimensional V.S.

- Define relationships as **translations** operating on the low-dimensional embeddings of the entities

- Easy to train, reduced number of parameters, high scalability

Introduction

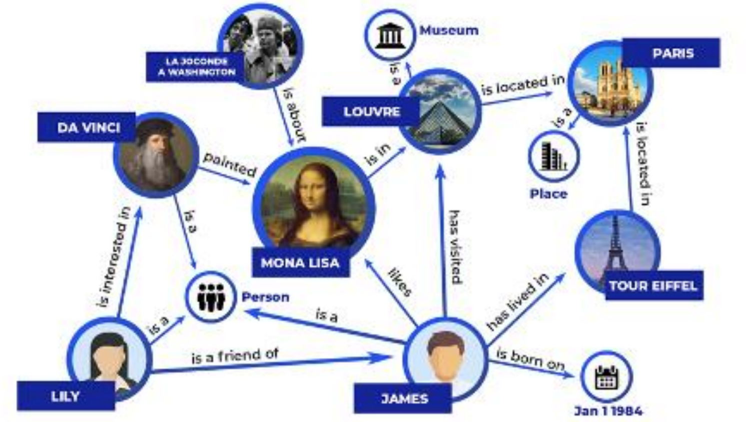
Multi-relational data

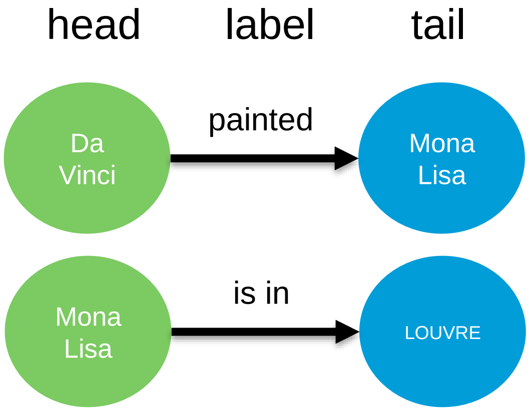
In **Directed graph,** node as entity, edge as relationship

Composed of $(head, label, tail)$

head, tail: entity   label: relationship

Ex)





Used in social network analysis, recommender systems

Goal: automatically adding new facts **without** extra knowledge

Introduction

Modeling multi-relational data

Extract connectivity patterns between entities,

Locality may involve relationships and entities of **different** types at the **same time**
-> requires more **generic** approaches

Most existing methods were based on latent representations of constituents
But higher cost / overfitting / underfitting problem exists.

Simpler, **linear** model provide better trade-offs between accuracy and scalability

Introduction

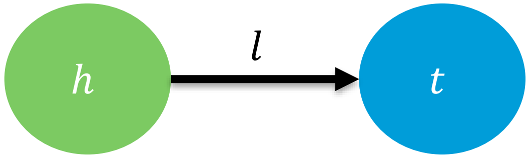
Relationships as translations

In TransE,

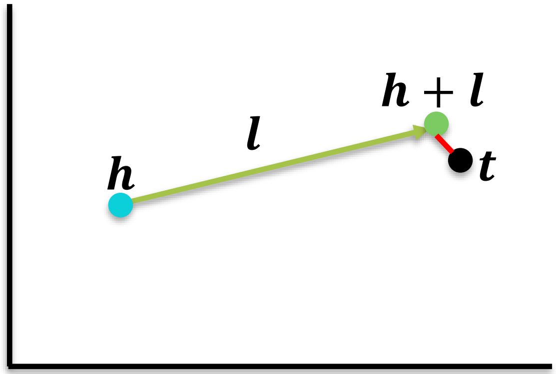
relationships: ***translations*** *in the embedding space*

When (h, l, t) holds, $\boldsymbol{t}$ is close to $\boldsymbol{h} + \boldsymbol{l}$ $(\boldsymbol{h}, \boldsymbol{l}, \boldsymbol{t} \in \mathbb{R}^d)$

Social Network



Embedding Space



Introduction

Motivations

Why use **translations**?

- naturally represent hierarchical relationships(extremely common in KBs)

  ex) parent-child relationship, sibling relationship

- represents 1-to-1 relationships between entities of different types

Model

Definition

Algorithm 1 Learning TransE

input Training set $S = \{(h, \ell, t)\}$, entities and rel. sets E and L, margin γ, embeddings dim. k.

initialize $\boldsymbol{\ell} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $\ell \in L$

$\quad \boldsymbol{\ell} \leftarrow \boldsymbol{\ell} / \|\boldsymbol{\ell}\|$ for each $\ell \in L$

$\quad \mathbf{e} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in E$

loop

$\quad \mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$ for each entity $e \in E$

$\quad S_{batch} \leftarrow \text{sample}(S, b)$ // sample a minibatch of size b

$\quad T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets

$\quad$ **for** $(h, \ell, t) \in S_{batch}$ **do**

$\quad\quad (h', \ell, t') \leftarrow \text{sample}(S'_{(h,\ell,t)})$ // sample a corrupted triplet

$\quad\quad T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$

$\quad$ **end for**

$\quad$ Update embeddings w.r.t. $\displaystyle\sum_{((h,\ell,t),(h',\ell,t')) \in T_{batch}} \nabla \left[\gamma + d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t}) - d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})\right]_+$

end loop

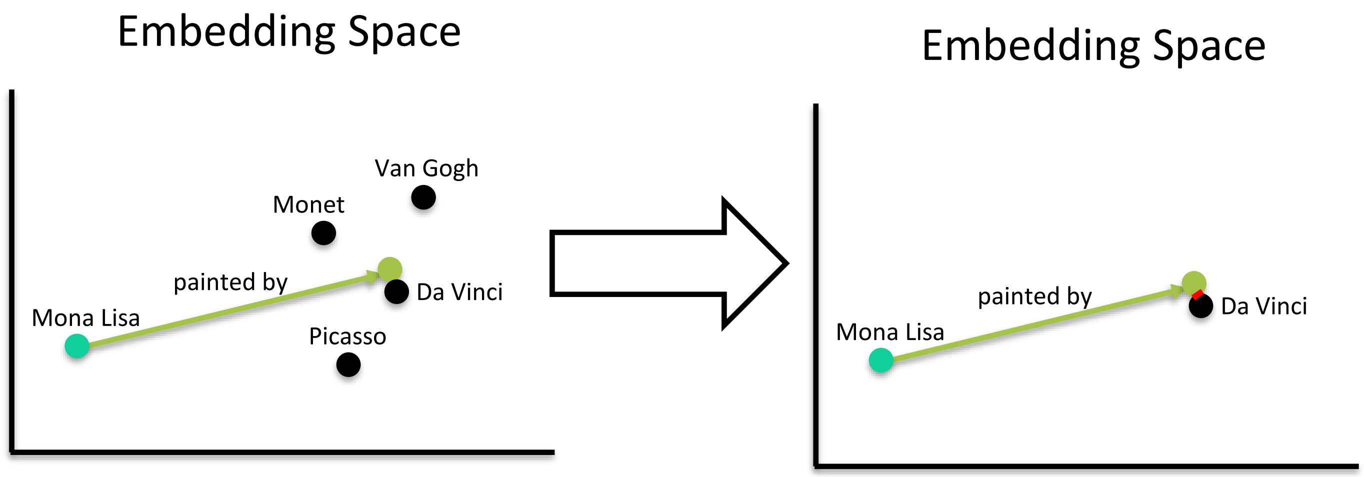
Input	설명
$S = \{(h, l, t)\}$	set of triplets: (head, relation, tail)
$S' = \{(h', l, t')\}$	Corrupted triplets
E	set of entities
L	set of relationships
T	pairs of triplets
k	dimension of embedding space
γ	margin (smallest distance tolerated by model between valid, corrupted triplets)

Model

Idea

When (h, l, t) holds, $\boldsymbol{t}$ is the **nearest neighbor** of $\boldsymbol{h} + \boldsymbol{l}$ $(\boldsymbol{h} + \boldsymbol{l} \approx \boldsymbol{t})$

$d(\boldsymbol{h} + \boldsymbol{l}, \boldsymbol{t})$: dissimilarity of triplet



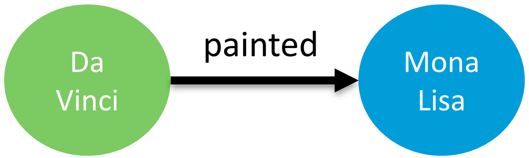
Notations

$\boldsymbol{h}$: Mona Lisa
$\boldsymbol{l}$: painted by
$\boldsymbol{h}$+$\boldsymbol{l}$:
$\boldsymbol{t}$: Da Vinci
$d(\boldsymbol{h} + \boldsymbol{l}, \boldsymbol{t})$:
distance ( )

Model

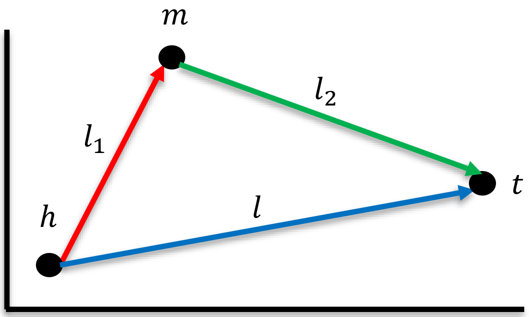
Characteristics

✓ Inverse relation holds



✓ Composition relations

$$(h, l_1, m),\ (m, l_2, t)\ \rightarrow\ (h, l, t)\ \forall\ h, l, t\ (l = l_1 + l_2)$$



Ex) My mother's husband is my father.

Model

Idea

Loss function: $$\mathcal{L} = \sum_{(h,\ell,t)\in S} \sum_{(h',\ell,t')\in S'_{(h,\ell,t)}} \left[\gamma + d(\boldsymbol{h}+\boldsymbol{\ell},\boldsymbol{t}) - d(\boldsymbol{h'}+\boldsymbol{\ell},\boldsymbol{t'})\right]_+$$

Well-defined: lower values of energy for training triplets than corrupted ones

Corrupted triplets:

Replace **either** head or tail with random entity

$$S'_{(h,\ell,t)} = \left\{(h',\ell,t)|h' \in E\right\} \cup \left\{(h,\ell,t')|t' \in E\right\}.$$

Use L_2 normalization for entity embeddings

Model

Algorithm

Algorithm 1 Learning TransE

input Training set $S = \{(h, \ell, t)\}$, entities and rel. sets E and L, margin γ, embeddings dim. k.

initialize $\boldsymbol{\ell} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $\ell \in L$

$\quad \boldsymbol{\ell} \leftarrow \boldsymbol{\ell} / \|\boldsymbol{\ell}\|$ for each $\ell \in L$

$\quad \mathbf{e} \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in E$

loop

$\quad \mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$ for each entity $e \in E$

$\quad S_{batch} \leftarrow \text{sample}(S, b)$ // sample a minibatch of size b

$\quad T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets

$\quad$ **for** $(h, \ell, t) \in S_{batch}$ **do**

$\quad\quad (h', \ell, t') \leftarrow \text{sample}(S'_{(h,\ell,t)})$ // sample a corrupted triplet

$\quad\quad T_{batch} \leftarrow T_{batch} \cup \{((h, \ell, t), (h', \ell, t'))\}$

$\quad$ **end for**

$\quad$ Update embeddings w.r.t. $\boxed{\displaystyle\sum_{((h,\ell,t),(h',\ell,t')) \in T_{batch}} \nabla \left[\gamma + d(\boldsymbol{h} + \boldsymbol{\ell}, \boldsymbol{t}) - d(\boldsymbol{h'} + \boldsymbol{\ell}, \boldsymbol{t'})\right]_+}$

end loop

Lower distance for valid triplets
Higher distance for corrupted triplets

Procedures

Initialize E, L

normalize relationships

Loop

1. Normalize entities
2. Create samples(S_{batch}), initialize T_{batch}
3. Create unseen samples(S'_{batch}), assign both into T_{batch}
4. Update embeddings using SGD

Experiments

Dataset

Wordnet(**WN**): KB composed of senses

Entities: word senses, relationships: lexical relations

Freebase(FB): huge, growing KB of general facts

FB15K: small data set

FB1M: large-scale data to test TransE

DATA SET	WN	FB15K	FB1M
ENTITIES	40,943	14,951	1×10^6
RELATIONSHIPS	18	1,345	23,382
TRAIN. EX.	141,442	483,142	17.5×10^6
VALID EX.	5,000	50,000	50,000
TEST EX.	5,000	59,071	177,404

Experiments

Experiment setup

Implementation

- learning rate(λ): $\{0.001, 0.01, 0.1\}$
- margin(γ): $\{1, 2, 10\}$
- latent dimension(k): $\{20, 50\}$
- dissimilarity measure(d): $\{L_1, L_2\}$
- training time: limited to 1000 epochs

Optimal configurations

- WN: $\{0.01, 2, 20, L_1\}$
- $FB15K$: $\{0.01, 1, 50, L_1\}$
- $FB1M$: $\{0.01, 1, 50, L_2\}$

Experiments

Experiment setup

Ranking Method

1. For each test triplet, eliminate head, replace by each of the entities of dictionary
2. compute dissimilarity of corrupted triplets, sort by ascending order
3. Repeat 1, 2 while removing the tail instead of the head

- Use **mean rank**, **hits@10** as metrics

  mean rank: mean of predicted ranks

  hits@10: proportion of correct entities ranked in the top 10

- Use *filtered rank* as metric

  filtered rank: eliminate corrupted triples in dataset, calculate rank

Experiments

Experiment setup

Baselines

- Unstructured: another version of TransE (mono-relational data, translations=0)
- RESCAL: collective MF model
- SE, SME(linear, bilinear), LFM: energy-based

Method	Nb. of parameters	on FB15k
Unstructured [2]	$O(n_e k)$	0.75
RESCAL [11]	$O(n_e k + n_r k^2)$	87.80
SE [3]	$O(n_e k + 2n_r k^2)$	7.47
SME(linear) [2]	$O(n_e k + n_r k + 4k^2)$	0.82
SME(bilinear) [2]	$O(n_e k + n_r k + 2k^3)$	1.06
LFM [6]	$O(n_e k + n_r k + 10k^2)$	0.84
TransE	$O(n_e k + n_r k)$	0.81

Experiments

Link prediction

Table 3: **Link prediction results.** Test performance of the different methods.

DATASET	WN				FB15K				FB1M	
METRIC	MEAN RANK		HITS@10 (%)		MEAN RANK		HITS@10 (%)		MEAN RANK	HITS@10 (%)
Eval. setting	*Raw*	*Filt.*	*Raw*	*Filt.*	*Raw*	*Filt.*	*Raw*	*Filt.*	*Raw*	*Raw*
Unstructured [2]	315	304	35.3	38.2	1,074	979	4.5	6.3	15,139	2.9
RESCAL [11]	1,180	1,163	37.2	52.8	828	683	28.4	44.1	-	-
SE [3]	1,011	985	68.5	80.5	273	162	28.8	39.8	22,044	17.5
SME(LINEAR) [2]	545	533	65.1	74.1	274	154	30.7	40.8	-	-
SME(BILINEAR) [2]	526	509	54.7	61.3	284	158	31.3	41.3	-	-
LFM [6]	469	456	71.4	81.6	283	164	26.0	33.1	-	-
TransE	**263**	**251**	**75.4**	**89.2**	**243**	**125**	**34.9**	**47.1**	**14,615**	**34.0**

appropriate design + simplicity + translation term

-> high performance

Experiments

Link prediction

Baseline Analysis

SE: more expressive, but higher complexity

SME: insufficient learning to exploit full capability

LFM: insufficient learning + originally designed to predict relationships

Unstructured: cluster co-occurring entities independent of relationships

Experiments

Link prediction

Table 4: **Detailed results by category of relationship.** We compare Hits@10 (in %) on FB15k in the filtered evaluation setting for our model, TransE and baselines. (M. stands for MANY).

TASK	PREDICTING *head*				PREDICTING *tail*			
REL. CATEGORY	1-TO-1	1-TO-M.	M.-TO-1	M.-TO-M.	1-TO-1	1-TO-M.	M.-TO-1	M.-TO-M.
Unstructured [2]	34.5	2.5	6.1	6.6	34.3	4.2	1.9	6.6
SE [3]	35.6	62.6	17.2	37.5	34.9	14.6	68.3	41.3
SME(LINEAR) [2]	35.1	53.7	19.0	40.3	32.7	14.9	61.6	43.3
SME(BILINEAR) [2]	30.9	**69.6**	**19.9**	38.6	28.2	13.1	**76.0**	41.8
TransE	**43.7**	65.7	18.2	**47.2**	**43.7**	**19.7**	66.7	**50.0**

Categorize relationships into 4 classes: 1-1, 1-M, M-1, M-M'

Highest performance at 1-to-1

Experiments

Link prediction

Head와 label이 주어질 때 top predicted tails 표시

True tail, *other true tails* appear **commonly**

(predictions reflect common-sense)

INPUT (HEAD AND LABEL)	PREDICTED TAILS
J. K. Rowling `influenced by`	*G. K. Chesterton*, J. R. R. Tolkien, *C. S. Lewis*, **Lloyd Alexander**, Terry Pratchett, Roald Dahl, Jorge Luis Borges, *Stephen King*, Ian Fleming
Anthony LaPaglia `performed in`	*Lantana*, *Summer of Sam*, *Happy Feet*, *The House of Mirth*, Unfaithful, **Legend of the Guardians**, Naked Lunch, X-Men, The Namesake
Camden County `adjoins`	**Burlington County**, *Atlantic County*, *Gloucester County*, Union County, Essex County, New Jersey, Passaic County, Ocean County, Bucks County
The 40-Year-Old Virgin `nominated for`	*MTV Movie Award for Best Comedic Performance*, *BFCA Critics' Choice Award for Best Comedy*, *MTV Movie Award for Best On-Screen Duo*, MTV Movie Award for Best Breakthrough Performance, **MTV Movie Award for Best Movie**, MTV Movie Award for Best Kiss, D. F. Zanuck Producer of the Year Award in Theatrical Motion Pictures, Screen Actors Guild Award for Best Actor - Motion Picture
Costa Rica football team `has position`	*Forward*, *Defender*, *Midfielder*, **Goalkeepers**, Pitchers, Infielder, Outfielder, Center, Defenseman
Lil Wayne `born in`	**New Orleans**, Atlanta, Austin, St. Louis, Toronto, New York City, Wellington, Dallas, Puerto Rico
WALL-E `has the genre`	Animations, Computer Animation, *Comedy film*, *Adventure film*, *Science Fiction*, **Fantasy**, Stop motion, *Satire*, Drama

Experiments

Relationship prediction

Setup

- Test how well methods could generalize to new facts
- Use FB15k dataset
- Randomly select 40 relationships, then split data into 2 sets

- *FB15k-40rel*: containing all triplets these with 40 relationships

  training: 40000(1000 for each relationship)
- *FB15k-rest*: containing the rest

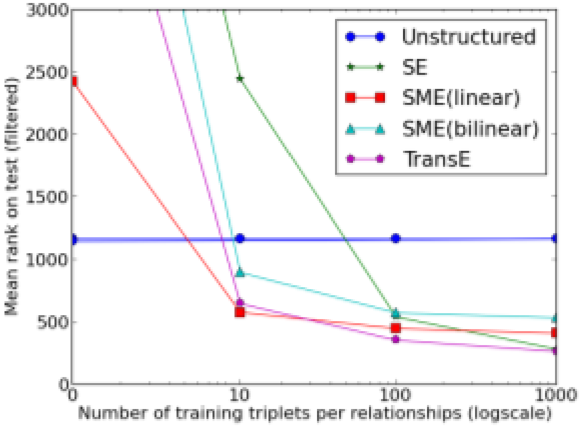
Phases

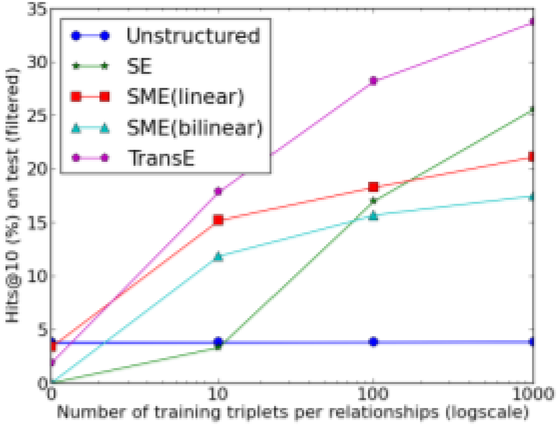
1. Train, select models using ***rest*** train, validation set
2. Train by ***rel*** train set for learn parameters related to 40 relationships
3. Evaluate by ***rel*** test set

Repeat using 0, 10, 100, 1000 examples of each relationships in 2

Experiments

Relationship prediction





TransE: fastest learning model

simplicity: generalize well **without** modify already trained embeddings

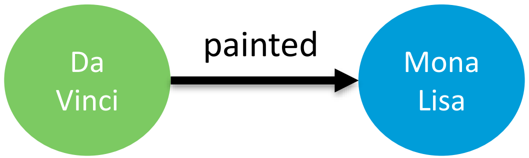
Conclusions

- ✓ TransE: new approach to learn embeddings of KBs
- ✓ Focus on minimal parametrization to represent hierarchical relationships
- ✓ Highly scalable model
- ✓ Remains unclear if all relationship types can be modeled adequately

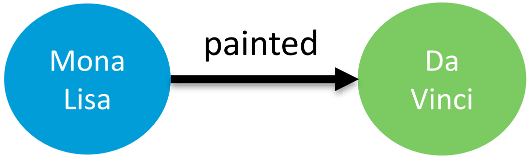
Further Discussion

Limitations

- ✓ Non-symmetric



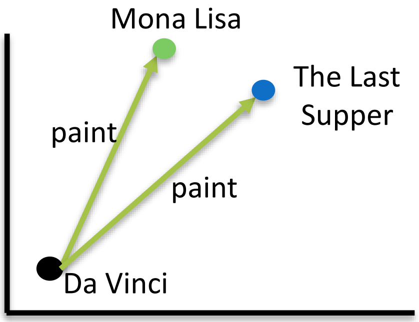
(O)



(X)

- ✓ Not suitable for represent n-ary relations

  Ex) (Da Vinci, paint, Mona Lisa), (Da Vinci, paint, The Last Supper)



Embeddings are equal, but locations of tails are different!
-> **Wrong**!

Further Discussion

Codes

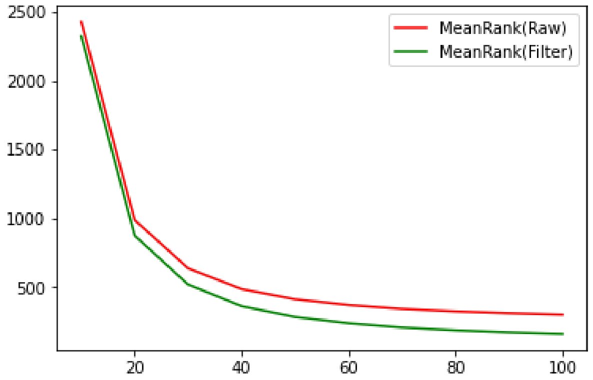
Reproduction of experiment results

Experiment: link prediction

Dataset : FB15k

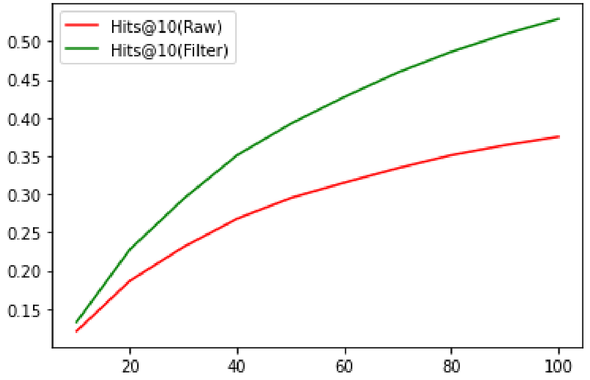
Set epochs = 100, other conditions are same

(computational cost problem)



Results

Experiment	MeanRank (Raw)	MeanRank (Filter)	Hits@10 (Raw)	Hits@10 (Filter)
paper	243	125	34.9	47.1
implementation	299.884	159.487	37.5	52.9



Further Discussion

Implementations

Dataset

NATION: relations between countries

UMLS: biomedical ontology set

Set k=5, using L1-normalization

100 epochs for NATION, 1000 epochs for UMLS

Results

Experiment	MeanRank (Raw)	MeanRank (Filter)	Hits@10 (Raw)	Hits@10 (Filter)
NATION	5.025	2.812	77.5	94.8
UMLS	29.641	23.031	47.3	60.1

THANK YOU