Part I: Models based on RDF triples

TOPICS:

- PROBLEM OVERVIEW
- TRANSE AND ITS VARIATIONS
- RESCAL AND ITS VARIATIONS
- DEEP NEURAL NETWORK ARCHITECTURES
- MODEL TRAINING

Part I: Models based on RDF triples

TOPICS:

- PROBLEM OVERVIEW
- TRANSE AND ITS VARIATIONS
- RESCAL AND ITS VARIATIONS
- DEEP NEURAL NETWORK ARCHITECTURES
- MODEL TRAINING

Problem overview

Given

- A knowledge graph stored as a set of subject-relation-object triples $\mathcal{T}^+ = \{(s, r, o)\}$, where $s, o \in \mathcal{E}$ are entities and $r \in \mathcal{R}$ is a relation
- No entity/relation features or extra information is used

Aim

- To learn distributed representations of entities and relations which can preserve the inherent structure of the original graph

Procedure

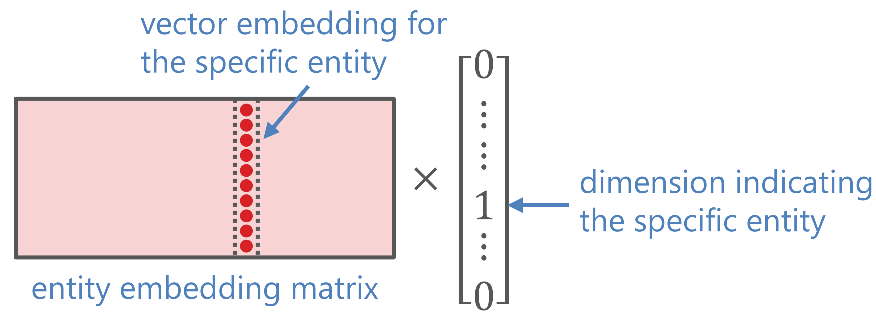
- **Define an encoder** (a mapping from entities/relations to embeddings)
- **Define a triple scoring function** (a measure of validity of triples in the embedding space)
- **Optimize the encoder parameters** (entity/relation embeddings)

Learning distributed representations

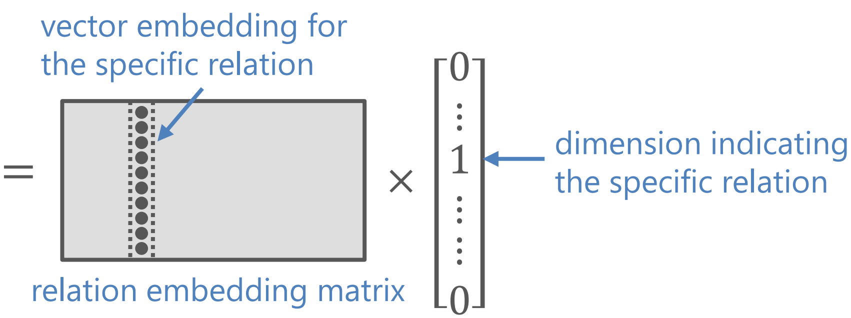
- **Encoder** maps entities and relations to their embeddings **(what to be learned)**
  - Entities are represented as points in the embedding space, i.e., vectors
  - Relations are operations between entities, usually represented as vectors, but sometimes as matrices or tensors

Encoder = Embedding-lookup

$$\text{Encoding}(e) = \mathbf{e} =$$



$$\text{Encoding}(r) = \mathbf{r} =$$



Learning distributed representations (cont.)

- Triple scoring function specifies how the validity of a triple is measured by its entity and relation embeddings
  - TransE and its variations
  - RESCAL and its variations
  - Deep neural network architectures
- Optimize entity and relation embeddings by maximizing total validity of triples observed in the graph

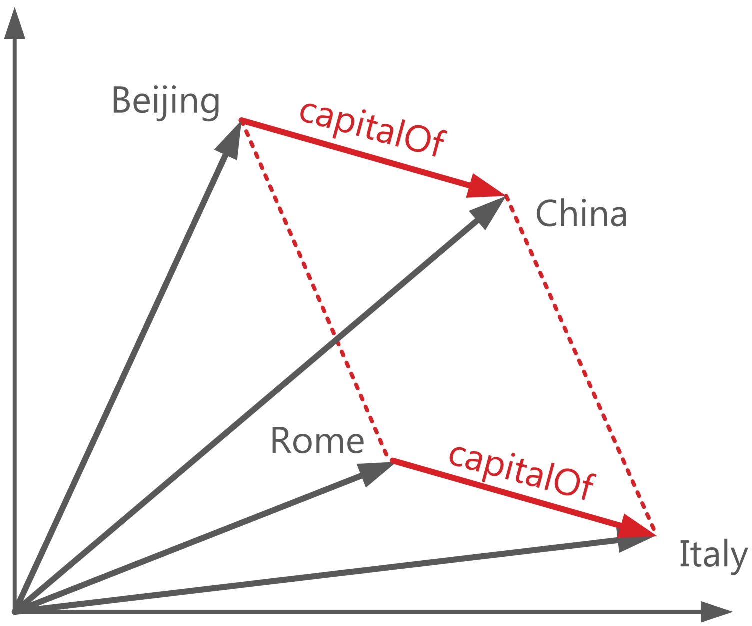
Part I: Models based on RDF triples

TOPICS:

- PROBLEM OVERVIEW
- TRANSE AND ITS VARIATIONS
- RESCAL AND ITS VARIATIONS
- DEEP NEURAL NETWORK ARCHITECTURES
- MODEL TRAINING

TransE (Bordes et al., 2013)

- **Translation assumption:** Relations as translations operating on entity embeddings, i.e., $\mathbf{s} + \mathbf{r} \approx \mathbf{o}$ when (s, r, o) holds
  - $\mathbf{China} - \mathbf{Beijing} = \mathbf{Italy} - \mathbf{Rome} = \mathbf{capitalOf}$
  - $\mathbf{Beijing} + \mathbf{capitalOf} = \mathbf{China}$
  - $\mathbf{Rome} + \mathbf{capitalOf} = \mathbf{Italy}$



Entity embedding: $\mathbf{s}, \mathbf{o} \in \mathbb{R}^d$

Relation embedding: $\mathbf{r} \in \mathbb{R}^d$

Triple scoring function:

$f(s, r, o) = -\|\mathbf{s} + \mathbf{r} - \mathbf{o}\|_{\ell_1/\ell_2}$

$\|\cdot\|_{\ell_1/\ell_2}$ is the ℓ_1 or ℓ_2 norm

Deficiencies in TransE

- Ineffective in dealing with reflexive/many-to-one/one-to-many relations
  - **Reflexive relations:** $(s, r, o) \in \mathcal{T}^+,\ (o, r, s) \in \mathcal{T}^+$

    $$\mathbf{s} + \mathbf{r} = \mathbf{o},\ \mathbf{o} + \mathbf{r} = \mathbf{s} \quad \Rightarrow \quad \mathbf{r} = \mathbf{0},\ \mathbf{s} = \mathbf{o}$$

    (Cristiano Ronaldo, teammates, Sergio Ramos) **teammates = 0**

    (Sergio Ramos, teammates, Cristiano Ronaldo) **Cristiano Ronaldo = Sergio Ramos**
  - **Many-to-one relations:** $\forall i \in \{1, \cdots, n\},\ (s_i, r, o) \in \mathcal{T}^+$

    $$\forall i \in \{1, \cdots, n\},\ \mathbf{s}_i + \mathbf{r} = \mathbf{o} \quad \Rightarrow \quad \mathbf{s}_1 = \cdots = \mathbf{s}_n$$

    (Camas, locatedIn, Spain)

    (Real Madrid, locatedIn, Spain) **Camas = Real Madrid = Barcelona**

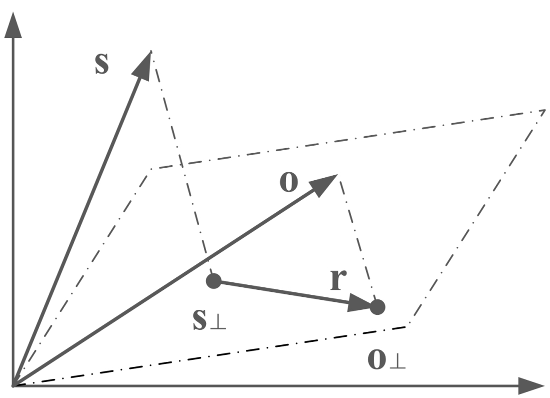
    (Barcelona, locatedIn, Spain)
  - **One-to-many relations:** $\forall j \in \{1, \cdots, m\},\ (s, r, o_j) \in \mathcal{T}^+$

    $$\forall j \in \{1, \cdots, m\},\ \mathbf{s} + \mathbf{r} = \mathbf{o}_j \quad \Rightarrow \quad \mathbf{o}_1 = \cdots = \mathbf{o}_m$$

Improving TransE

- **Introducing relation-specific entity embeddings**
  - TransH (Wang et al., 2014) projects entities into relation-specific hyperplanes
  - TransR (Lin et al., 2015) projects entities into relation-specific spaces

TransH



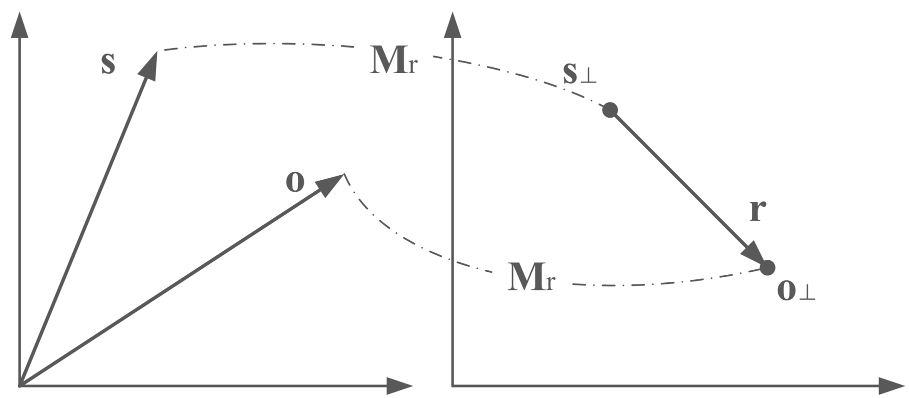
Entity and Relation Space

Projection: $\mathbf{s}_\perp = \mathbf{s} - \mathbf{w}_r^\top \mathbf{s} \mathbf{w}_r$

$\mathbf{o}_\perp = \mathbf{o} - \mathbf{w}_r^\top \mathbf{o} \mathbf{w}_r$

Translation: $\mathbf{s}_\perp + \mathbf{r} \approx \mathbf{o}_\perp$

TransR



Entity Space

Relation Space of r

Projection: $\mathbf{s}_\perp = \mathbf{M}_r \mathbf{s}$

$\mathbf{o}_\perp = \mathbf{M}_r \mathbf{o}$

Translation: $\mathbf{s}_\perp + \mathbf{r} \approx \mathbf{o}_\perp$

Improving TransE (cont.)

Relaxing translation assumption $\mathbf{s} + \mathbf{r} \approx \mathbf{o}$

- TransM (Fan et al., 2014) assigns lower weights to one-to-many/many-to-one/many-to-many relations so that $\mathbf{o}$ can lie far apart from $\mathbf{s} + \mathbf{r}$ in these relations

$$f(s, r, o) = -\theta_r \|\mathbf{s} + \mathbf{r} - \mathbf{o}\|_{\ell_1/\ell_2}$$

- TransF (Feng et al., 2016) only requires $\mathbf{o}$ to lie in the same direction with $\mathbf{s} + \mathbf{r}$, and $\mathbf{s}$ in the same direction with $\mathbf{o} - \mathbf{r}$

$$f(s, r, o) = (\mathbf{s} + \mathbf{r})^\top \mathbf{o} + (\mathbf{o} - \mathbf{r})^\top \mathbf{s}$$

- ManifoldE (Xiao et al., 2016a) allows $\mathbf{o}$ to lie approximately on a manifold, i.e., a hyper-sphere centered at $\mathbf{s} + \mathbf{r}$ with radius θ_r

$$f(s, r, o) = -(\|\mathbf{s} + \mathbf{r} - \mathbf{o}\|^2 - \theta_r^2)^2$$

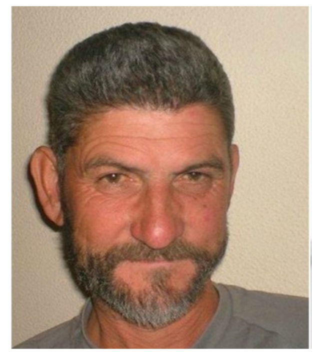
Gaussian embeddings

- To model uncertainties of entities and relations

Uncertainties of entities

bornIn Funchal
nationality Portugal
hasParent Jose Dinis Aveiro
playsFor Real Madrid
teammates Sergio Ramos

Cristiano Ronaldo



hasChild Cristiano Ronaldo

Jose Dinis Aveiro

higher uncertainty

Uncertainties of relations

? **hasParent** Jose Dinis Aveiro

? **playsFor** Real Madrid

? **nationality** Portugal

higher uncertainty

Gaussian embeddings (cont.)

- KG2E (He et al., 2015) represents entities/relations as random vectors drawn from multivariate Gaussian distributions
  - **Entity/relation embeddings**

$$\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}_s, \boldsymbol{\Sigma}_s) \qquad \boldsymbol{\mu}_s, \boldsymbol{\mu}_o, \boldsymbol{\mu}_r \in \mathbb{R}^d : \text{mean vectors}$$

$$\mathbf{o} \sim \mathcal{N}(\boldsymbol{\mu}_o, \boldsymbol{\Sigma}_o) \qquad \boldsymbol{\Sigma}_s, \boldsymbol{\Sigma}_o, \boldsymbol{\Sigma}_r \in \mathbb{R}^{d \times d} : \text{covariance matrices}$$

$$\mathbf{r} \sim \mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r)$$

A larger covariance matrix (determinant or trace) indicates a higher level of uncertainty

  - **Modeling assumption**

$$\mathbf{s} + \mathbf{r} \approx \mathbf{o} \Rightarrow \mathbf{r} \approx \mathbf{o} - \mathbf{s} \Rightarrow \mathcal{N}(\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \approx \mathcal{N}(\boldsymbol{\mu}_o - \boldsymbol{\mu}_s, \boldsymbol{\Sigma}_o + \boldsymbol{\Sigma}_s)$$

  - **Triple scoring function**
    - Kullback-Leibler divergence
    - Probability inner product

Gaussian embeddings (cont.)

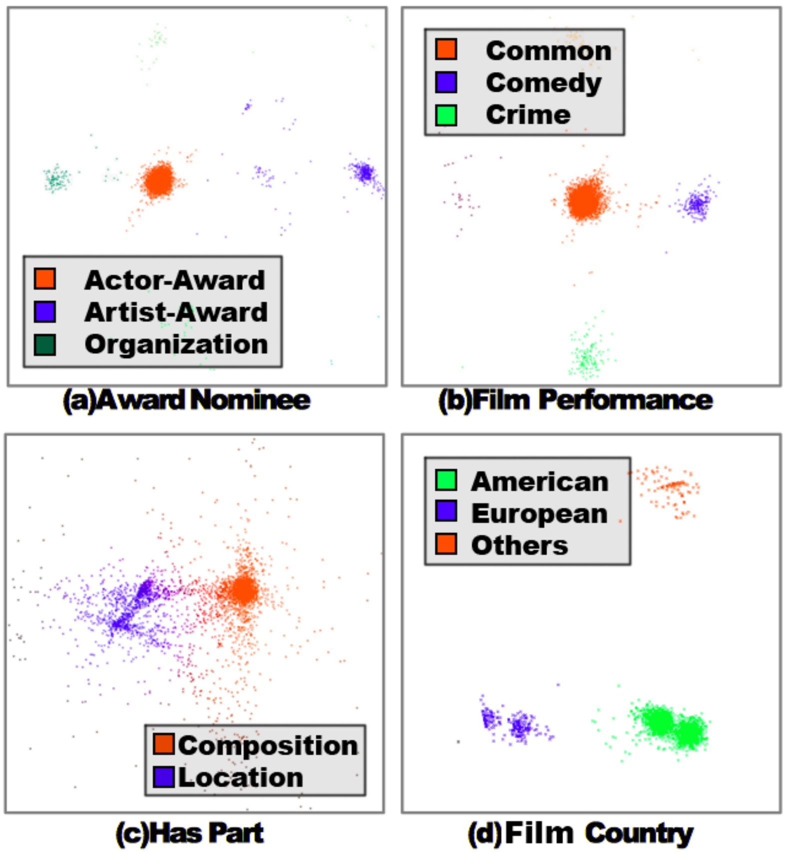
- Covariance matrices and uncertainties of Freebase relations learned by KG2E (He et al., 2015)

Relaion	#Triplet	#Head	#Tail	Type	(log) det	trace
nationality	4198	3755	100	m-1	-4.77	67.90
place_lived	3740	2441	784	m-n	-23.02	53.24
profession	11636	4145	152	m-n	-57.45	23.10
gender	3721	3721	2	m-1	-59.53	21.35
place_of_birth	2468	2468	685	m-1	-63.42	19.41
ethnicity	2030	1610	78	m-1	-69.95	15.00
major	260	217	60	m-1	-69.59	14.62
sibling	131	111	113	1-1	-72.98	14.29
religion	1086	963	45	m-1	-75.07	12.98
spouse	427	395	385	1-1	-77.77	12.24
children	77	69	71	1-1	-76.94	12.14
parents	83	74	76	1-1	-77.20	12.03

uncertainties

Multiple relation semantics

- A relation may have multiple meanings revealed by entity pairs associated with the corresponding triples



- $\mathbf{s} + \mathbf{r} \approx \mathbf{o} \Rightarrow \mathbf{r} \approx \mathbf{o} - \mathbf{s}$
- Visualization of TransE embeddings $\mathbf{o} - \mathbf{s}$ for different relation r, where $(s, r, o) \in \mathcal{T}_+$
- Different clusters indicate different relation semantics, e.g., **composition** and **location** for the **hasPart** relation

Material based on: Xiao et al. (2016). TransG : A generative model for knowledge graph embedding. ACL'16.

Multiple relation semantics (cont.)

- TransG (Xiao et al., 2016b) models multiple relation semantics by mixtures of Gaussians
  - **Entity embeddings:** Random vectors drawn from Gaussian distributions

$$\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}_s, \sigma_s^2 \mathbf{I}), \quad \mathbf{o} \sim \mathcal{N}(\boldsymbol{\mu}_o, \sigma_o^2 \mathbf{I})$$

  - **Relation embeddings:** Mixtures of Gaussians

$$\mathbf{r} = \sum_i \pi_r^i \boldsymbol{\mu}_r^i, \quad \boldsymbol{\mu}_r^i \sim \mathcal{N}\left(\boldsymbol{\mu}_o - \boldsymbol{\mu}_s, (\sigma_s^2 + \sigma_o^2)\mathbf{I}\right)$$

  - **Triple scoring function:** A mixture of translational distances introduced by different semantics of a relation

$$f(s,r,o) = \sum_i \pi_r^i \exp\left(\frac{-\|\boldsymbol{\mu}_s + \boldsymbol{\mu}_r^i - \boldsymbol{\mu}_o\|^2}{\sigma_s^2 + \sigma_o^2}\right)$$

Multiple relation semantics (cont.)

- Different semantics of Freebase/WordNet relations learned by TransG (Xiao et al., 2016b)

Relation	**Cluster**	**Triples (Head, Tail)**
PartOf	Location	(Capital of Utah, Beehive State), (Hindustan, Bharat) ...
	Composition	(Monitor, Television), (Bush, Adult Body), (Cell Organ, Cell)...
Religion	Catholicism	(Cimabue, Catholicism), (St.Catald, Catholicism) ...
	Others	(Michal Czajkowsk, Islam), (Honinbo Sansa, Buddhism) ...
DomainRegion	Abstract	(Computer Science, Security System), (Computer Science, PL)..
	Specific	(Computer Science, Router), (Computer Science, Disk File) ...
Profession	Scientist	(Michael Woodruf, Surgeon), (El Lissitzky, Architect)...
	Businessman	(Enoch Pratt, Entrepreneur), (Charles Tennant, Magnate)...
	Writer	(Vlad. Gardin, Screen Writer), (John Huston, Screen Writer) ...

Summary of the TransE family

- Summary of entity/relation embedding and scoring functions

Method	Ent. embedding	Rel. embedding	Scoring function $f(s, r, o)$	Constraints/Regularization
TransE	$\mathbf{s}, \mathbf{o} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$-\|\mathbf{s}+\mathbf{r}-\mathbf{o}\|_{\ell_1/\ell_2}$	$\|\mathbf{s}\|=1, \|\mathbf{o}\|=1$
TransH	$\mathbf{s}, \mathbf{o} \in \mathbb{R}^d$	$\mathbf{r}, \mathbf{w}_r \in \mathbb{R}^d$	$-\|(\mathbf{s}-\mathbf{w}_r^\top \mathbf{s}\mathbf{w}_r)+\mathbf{r}-(\mathbf{o}-\mathbf{w}_r^\top \mathbf{o}\mathbf{w}_r)\|^2$	$\|\mathbf{s}\| \leq 1, \|\mathbf{o}\| \leq 1$ $\lvert\mathbf{w}_r^\top \mathbf{r}\rvert/\|\mathbf{r}\| \leq \epsilon, \|\mathbf{w}_r\|=1$
TransR	$\mathbf{s}, \mathbf{o} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^k, \mathbf{M}_r \in \mathbb{R}^{k\times d}$	$-\|\mathbf{M}_r\mathbf{s}+\mathbf{r}-\mathbf{M}_r\mathbf{o}\|^2$	$\|\mathbf{s}\| \leq 1, \|\mathbf{o}\| \leq 1, \|\mathbf{r}\| \leq 1$ $\|\mathbf{M}_r\mathbf{s}\| \leq 1, \|\mathbf{M}_r\mathbf{o}\| \leq 1$
TransM	$\mathbf{s}, \mathbf{o} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$-\theta_r\|\mathbf{s}+\mathbf{r}-\mathbf{o}\|_{\ell_1/\ell_2}$	$\|\mathbf{s}\|=1, \|\mathbf{o}\|=1$
TransF	$\mathbf{s}, \mathbf{o} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$(\mathbf{s}+\mathbf{r})^\top\mathbf{o}+(\mathbf{o}-\mathbf{r})^\top\mathbf{s}$	$\|\mathbf{s}\| \leq 1, \|\mathbf{o}\| \leq 1, \|\mathbf{r}\| \leq 1$
ManifoldE	$\mathbf{s}, \mathbf{o} \in \mathbb{R}^d$	$\mathbf{r} \in \mathbb{R}^d$	$-(\|\mathbf{s}+\mathbf{r}-\mathbf{o}\|^2-\theta_r^2)^2$	$\|\mathbf{s}\| \leq 1, \|\mathbf{o}\| \leq 1, \|\mathbf{r}\| \leq 1$
KG2E	$\mathbf{s}\sim\mathcal{N}(\boldsymbol{\mu}_s,\boldsymbol{\Sigma}_s)$ $\mathbf{o}\sim\mathcal{N}(\boldsymbol{\mu}_o,\boldsymbol{\Sigma}_o)$ $\boldsymbol{\mu}_s, \boldsymbol{\mu}_o \in \mathbb{R}^d$ $\boldsymbol{\Sigma}_s, \boldsymbol{\Sigma}_o \in \mathbb{R}^{d\times d}$	$\mathbf{r}\sim\mathcal{N}(\boldsymbol{\mu}_r,\boldsymbol{\Sigma}_r)$ $\boldsymbol{\mu}_r \in \mathbb{R}^d, \boldsymbol{\Sigma}_r \in \mathbb{R}^{d\times d}$	$-\mathrm{tr}(\boldsymbol{\Sigma}_r^{-1}(\boldsymbol{\Sigma}_s+\boldsymbol{\Sigma}_o))-\boldsymbol{\mu}^\top\boldsymbol{\Sigma}_r^{-1}\boldsymbol{\mu}-\ln\frac{\det(\boldsymbol{\Sigma}_r)}{\det(\boldsymbol{\Sigma}_s+\boldsymbol{\Sigma}_o)}$ $-\boldsymbol{\mu}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}-\ln(\det(\boldsymbol{\Sigma}))$ $\boldsymbol{\mu}=\boldsymbol{\mu}_s+\boldsymbol{\mu}_r-\boldsymbol{\mu}_o$ $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_s+\boldsymbol{\Sigma}_r+\boldsymbol{\Sigma}_o$	$\|\boldsymbol{\mu}_s\| \leq 1, \|\boldsymbol{\mu}_o\| \leq 1, \|\boldsymbol{\mu}_r\| \leq 1$ $c_{min}\mathbf{I} \leq \boldsymbol{\Sigma}_s \leq c_{max}\mathbf{I}$ $c_{min}\mathbf{I} \leq \boldsymbol{\Sigma}_o \leq c_{max}\mathbf{I}$ $c_{min}\mathbf{I} \leq \boldsymbol{\Sigma}_r \leq c_{max}\mathbf{I}$
TransG	$\mathbf{s}\sim\mathcal{N}(\boldsymbol{\mu}_s,\sigma_s^2\mathbf{I})$ $\mathbf{o}\sim\mathcal{N}(\boldsymbol{\mu}_o,\sigma_o^2\mathbf{I})$ $\boldsymbol{\mu}_s, \boldsymbol{\mu}_o \in \mathbb{R}^d$	$\boldsymbol{\mu}_r^i\sim\mathcal{N}(\boldsymbol{\mu}_o-\boldsymbol{\mu}_s,(\sigma_s^2+\sigma_o^2)\mathbf{I})$ $\mathbf{r}=\sum_i \pi_r^i\boldsymbol{\mu}_r^i \in \mathbb{R}^d$	$\sum_i \pi_r^i \exp\left(-\frac{\|\boldsymbol{\mu}_s+\boldsymbol{\mu}_r^i-\boldsymbol{\mu}_o\|^2}{\sigma_s^2+\sigma_o^2}\right)$	$\|\boldsymbol{\mu}_s\| \leq 1, \|\boldsymbol{\mu}_o\| \leq 1, \|\boldsymbol{\mu}_r^i\| \leq 1$

Material based on: Wang et al. (2017). Knowledge graph embedding : A survey of approaches and applications. IEEE TKDE.

Summary of the TransE family (cont.)

- Comparison in space and time complexity

Method	Space	Time
TransE	$\mathcal{O}(nd + md)$	$\mathcal{O}(d)$
TransH	$\mathcal{O}(nd + md)$	$\mathcal{O}(d)$
TransR	$\mathcal{O}(nd + mdk)$	$\mathcal{O}(dk)$
TransM	$\mathcal{O}(nd + md)$	$\mathcal{O}(d)$
TransF	$\mathcal{O}(nd + md)$	$\mathcal{O}(d)$
ManifoldE	$\mathcal{O}(nd + md)$	$\mathcal{O}(d)$
KG2E	$\mathcal{O}(nd + md)$	$\mathcal{O}(d)$
TransG	$\mathcal{O}(nd + mdc)$	$\mathcal{O}(dc)$

- TransR introduces for each relation r a projection matrix $\mathbf{M}_r \in \mathbb{R}^{k \times d}$
- TransG models each relation as a mixture of c Gaussian distributions

Material based on: Wang et al. (2017). Knowledge graph embedding : A survey of approaches and applications. IEEE TKDE.

Part I: Models based on RDF triples

TOPICS:

- PROBLEM OVERVIEW
- TRANSE AND ITS VARIATIONS
- RESCAL AND ITS VARIATIONS
- DEEP NEURAL NETWORK ARCHITECTURES
- MODEL TRAINING

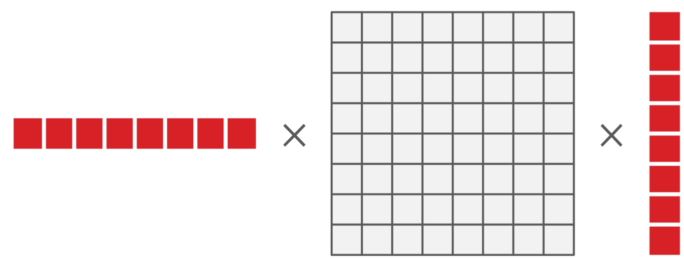
RESCAL (Nickel et al., 2011)

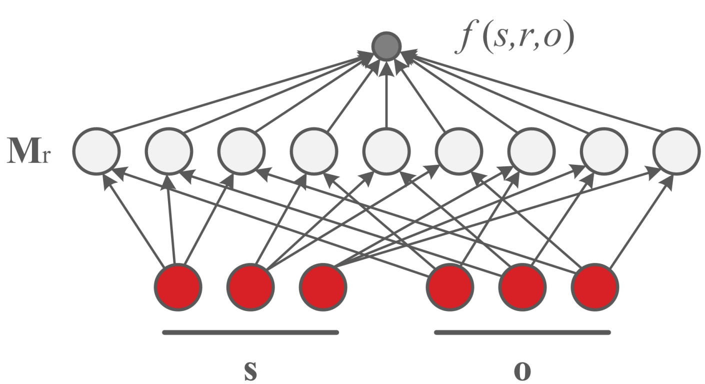
- **Bilinear matching:** Capturing pairwise interactions between all dimensions of subject and object entity embeddings
  - **Entity/relation embeddings**

$$\mathbf{s}, \mathbf{o} \in \mathbb{R}^d, \quad \mathbf{M}_r \in \mathbb{R}^{d \times d}$$

  - **Triple scoring function**

$$f(s, r, o) = \mathbf{s}^\top \mathbf{M}_r \mathbf{o} =$$





- Space complexity: $\mathcal{O}(d^2)$ per relation
- Time complexity: $\mathcal{O}(d^2)$

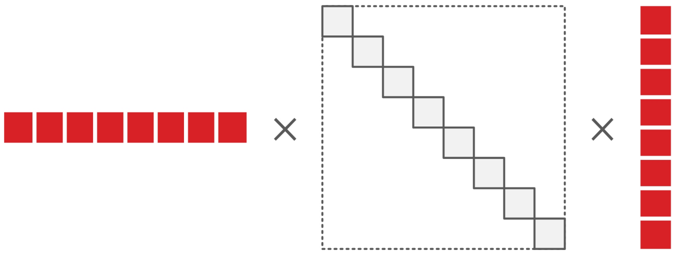
DistMult (Yang et al., 2015)

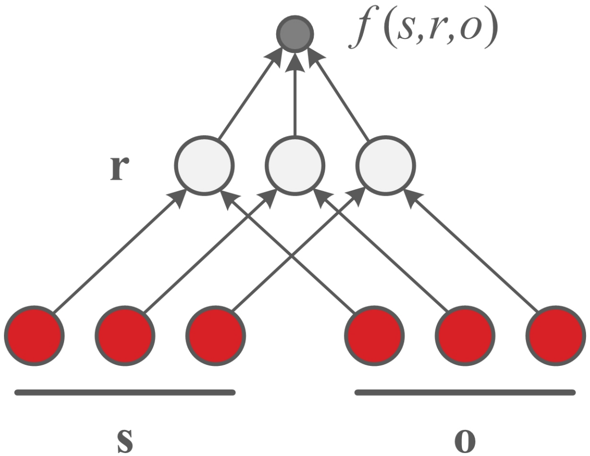
- Simplify RESCAL by restricting $\mathbf{M}_r$ to diagonal matrices
  - **Entity/relation embeddings**

$$\mathbf{s}, \mathbf{o} \in \mathbb{R}^d, \quad \mathbf{r} \in \mathbb{R}^d$$

  - **Triple scoring function**

$$f(s, r, o) = \mathbf{s}^\top \mathrm{diag}(\mathbf{r})\mathbf{o} =$$





- Space complexity: $\mathcal{O}(d)$
- Time complexity: $\mathcal{O}(d)$
- Suitable only for symmetric relations
  $f(s, r, o) = f(o, r, s)$

Holographic embeddings (HolE) (Nickel et al., 2016)

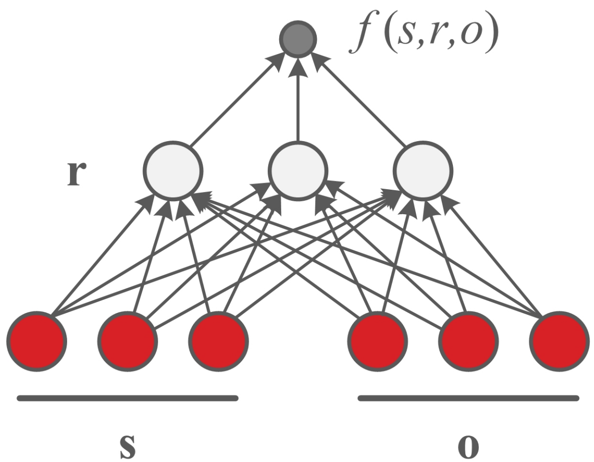
- Combine expressiveness of RESCAL with simplicity of DistMult
  - **Entity/relation embeddings**

$$\mathbf{s}, \mathbf{o} \in \mathbb{R}^d, \quad \mathbf{r} \in \mathbb{R}^d$$

  - **Triple scoring function**

$$f(s, r, o) = \mathbf{r}^\top (\mathbf{s} \star \mathbf{o}), \quad [\mathbf{s} \star \mathbf{o}]_i = \sum_{k=0}^{d-1} s_k o_{(k+i) \bmod d}$$

(circular correlation: $\mathbf{s} \star \mathbf{o} \neq \mathbf{o} \star \mathbf{s}$)



- Space complexity: $\mathcal{O}(d)$
- Time complexity: $\mathcal{O}(d \log d)$
- Suitable for asymmetric relations
  $f(s, r, o) \neq f(o, r, s)$

Complex embeddings (ComplEx) (Trouillon et al., 2016)

- Extend DistMult by introducing complex-valued embeddings so as to better model asymmetric relations
  - **Entity/relation embeddings**

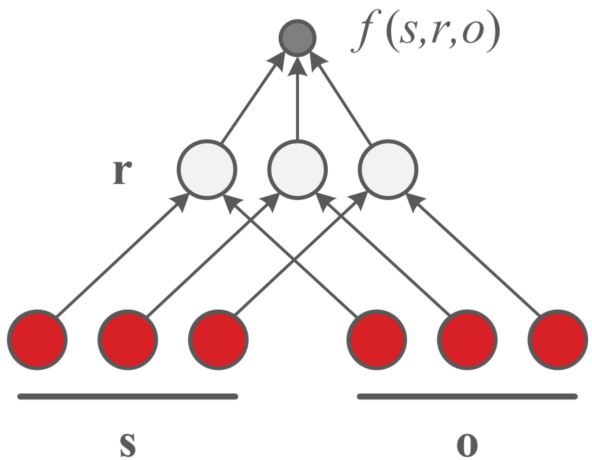
$$\mathbf{s}, \mathbf{o} \in \mathbb{C}^d, \quad \mathbf{r} \in \mathbb{C}^d$$

  - **Triple scoring function**

conjugate of $\mathbf{o}$

$$f(s, r, o) = \mathrm{Re}(\mathbf{s}^\top \mathrm{diag}(\mathbf{r})\bar{\mathbf{o}})$$

real part of a complex value

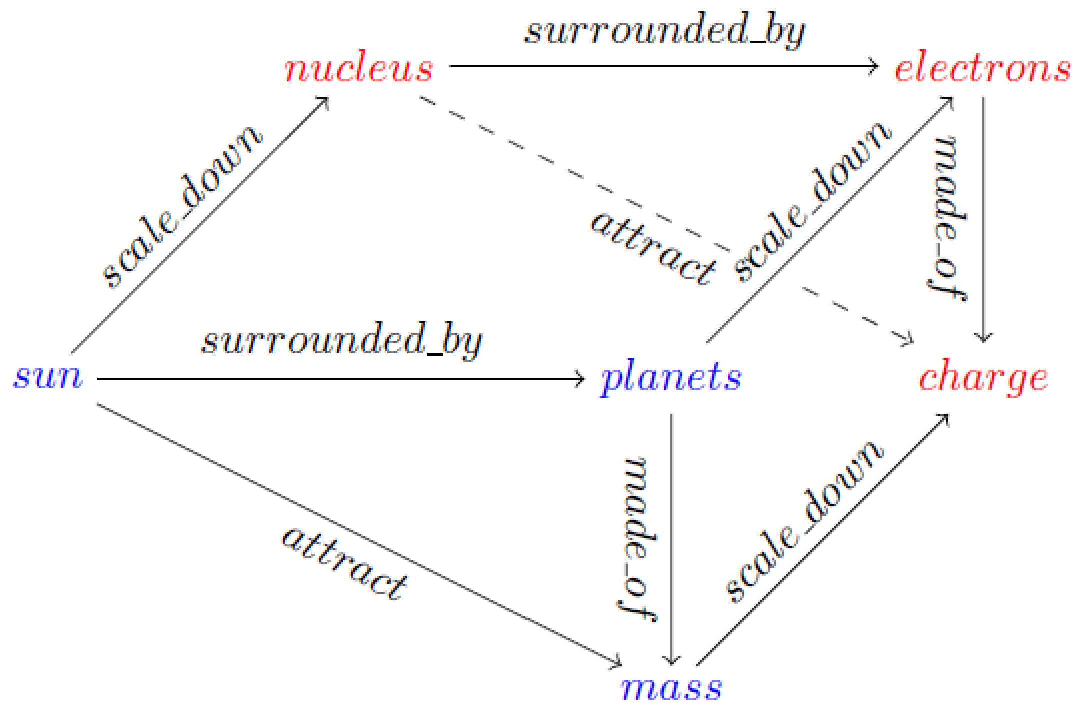


- Space complexity: $\mathcal{O}(d)$
- Time complexity: $\mathcal{O}(d)$
- Suitable for asymmetric relations

  $f(s,r,o) \neq f(o,r,s)$

Analogical inference

- Analogical properties of entities and relations



nucleus is to electrons as sun is to planets

electrons is to charge as planets is to mass ➡ **(nucleus, attract, charge)**

nucleus is to charge as sun is to mass

Material based on: Liu et al. (2017). Analogical inference for multi-relational embeddings. ICML'17.

ANALOGY (Liu et al., 2017)

- Modeling analogical properties

  Analogy: *a* is to *b* as *c* is to *d*

$$a \xrightarrow{r} b, \quad c \xrightarrow{r} d, \quad a \xrightarrow{r'} c, \quad b \xrightarrow{r'} d$$

$$\begin{tikzcd}
\mathrm{a} \arrow[r, "r"] \arrow[d, "r'"] & \mathrm{b} \arrow[d, "r'"] \\
\mathrm{c} \arrow[r, "r"] & \mathrm{d}
\end{tikzcd}$$

  Commutativity: $r \circ r' = r' \circ r$ (compositional equivalence)

- The ANALOGY model

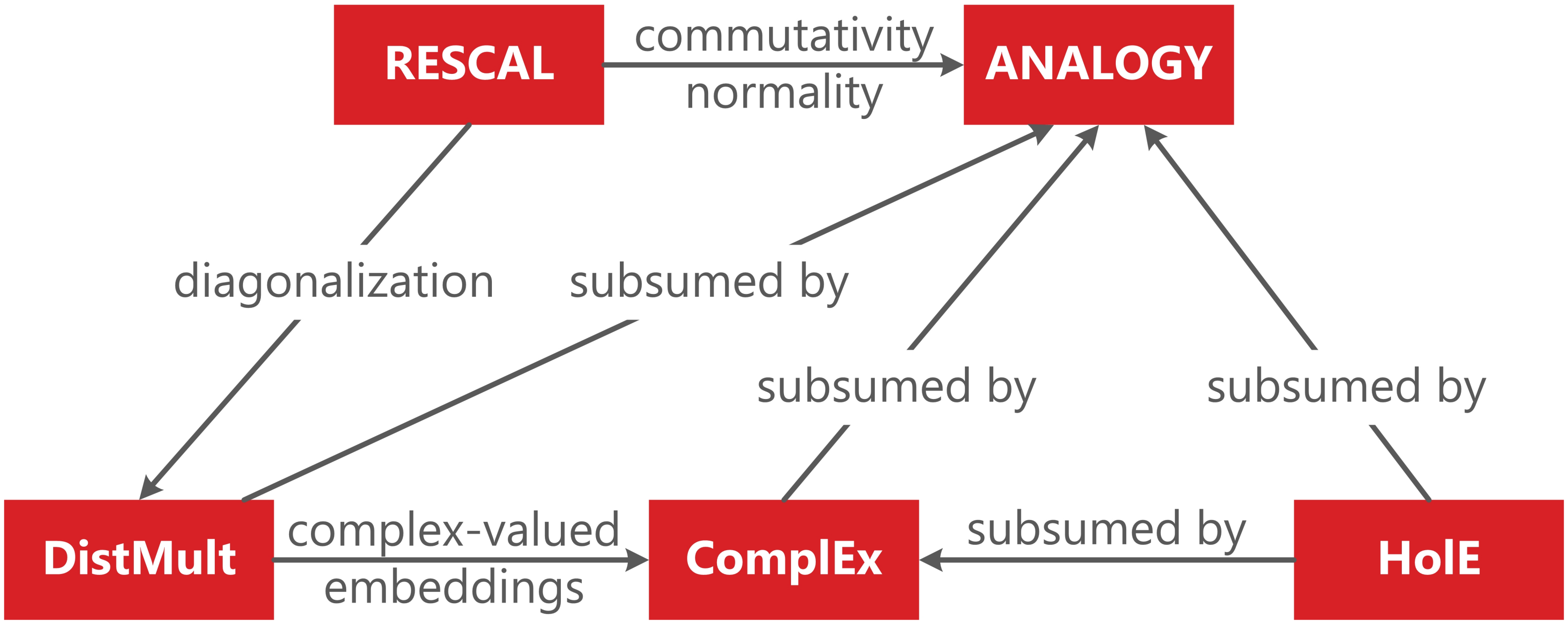
  Bilinear scoring function: $f(s, r, o) = \mathbf{s}^\top \mathbf{M}_r \mathbf{o}$

  Commutativity: $\mathbf{M}_r \mathbf{M}_{r'} = \mathbf{M}_{r'} \mathbf{M}_r, \quad \forall r, r' \in \mathcal{R}$

  Normality: $\mathbf{M}_r \mathbf{M}_r^\top = \mathbf{M}_r^\top \mathbf{M}_r, \quad \forall r \in \mathcal{R}$

Summary of the RESCAL family

- Relationships between different methods



Material based on: Hayashi and Masashi (2017). On the equivalence of holographic and complex embeddings for link prediction. ACL'17.
Liu et al. (2017). Analogical inference for multi-relational embeddings. ICML'17.

Summary of the RESCAL family (cont.)

- Summary of entity/relation embedding and scoring functions

Method	Ent. embedding	Rel. embedding	Scoring function $f(s,r,o)$	Constraints/Regularization
RESCAL	$\mathbf{s},\mathbf{o}\in\mathbb{R}^d$	$\mathbf{M}_r\in\mathbb{R}^{d\times d}$	$\mathbf{s}^\top\mathbf{M}_r\mathbf{o}$	$\\|\mathbf{s}\\|\leq 1, \\|\mathbf{o}\\|\leq 1, \\|\mathbf{M}_r\\|_F\leq 1$
DistMult	$\mathbf{s},\mathbf{o}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$\mathbf{s}^\top\mathrm{diag}(\mathbf{r})\mathbf{o}$	$\\|\mathbf{s}\\|=1, \\|\mathbf{o}\\|=1, \\|\mathbf{r}\\|\leq 1$
HolE	$\mathbf{s},\mathbf{o}\in\mathbb{R}^d$	$\mathbf{r}\in\mathbb{R}^d$	$\mathbf{r}^\top(\mathbf{s}\star\mathbf{o})$	$\\|\mathbf{s}\\|\leq 1, \\|\mathbf{o}\\|\leq 1, \\|\mathbf{r}\\|\leq 1$
ComplEx	$\mathbf{s},\mathbf{o}\in\mathbb{C}^d$	$\mathbf{r}\in\mathbb{C}^d$	$\mathrm{Re}(\mathbf{s}^\top\mathrm{diag}(\mathbf{r})\bar{\mathbf{o}})$	$\\|\mathbf{s}\\|\leq 1, \\|\mathbf{o}\\|\leq 1, \\|\mathbf{r}\\|\leq 1$
ANALOGY	$\mathbf{s},\mathbf{o}\in\mathbb{R}^d$	$\mathbf{M}_r\in\mathbb{R}^{d\times d}$	$\mathbf{s}^\top\mathbf{M}_r\mathbf{o}$	$\\|\mathbf{s}\\|\leq 1, \\|\mathbf{o}\\|\leq 1, \\|\mathbf{M}_r\\|_F\leq 1$ $\mathbf{M}_r\mathbf{M}_{r'}=\mathbf{M}_{r'}\mathbf{M}_r$ $\mathbf{M}_r\mathbf{M}_r^\top=\mathbf{M}_r^\top\mathbf{M}_r$

Material based on: Wang et al. (2017). Knowledge graph embedding : A survey of approaches and applications. IEEE TKDE.

Summary of the RESCAL family (cont.)

- Comparison in space and time complexity

Method	Space	Time
RESCAL	$\mathcal{O}(nd + md^2)$	$\mathcal{O}(d^2)$
DistMult	$\mathcal{O}(nd + md)$	$\mathcal{O}(d)$
HolE	$\mathcal{O}(nd + md)$	$\mathcal{O}(d \log d)$
ComplEx	$\mathcal{O}(nd + md)$	$\mathcal{O}(d)$
ANALOGY	$\mathcal{O}(nd + md)$	$\mathcal{O}(d)$

- RESCAL models each relation r as a projection matrix $\mathbf{M}_r \in \mathbb{R}^{d \times d}$
- Circular correlation is calculated via discrete Fourier transform whose time complexity is $\mathcal{O}(d \log d)$
- Projection matrices can be block-diagonalized into a set of almost-diagonal matrices, each has $\mathcal{O}(d)$ free parameters

Material based on: Wang et al. (2017). Knowledge graph embedding : A survey of approaches and applications. IEEE TKDE.

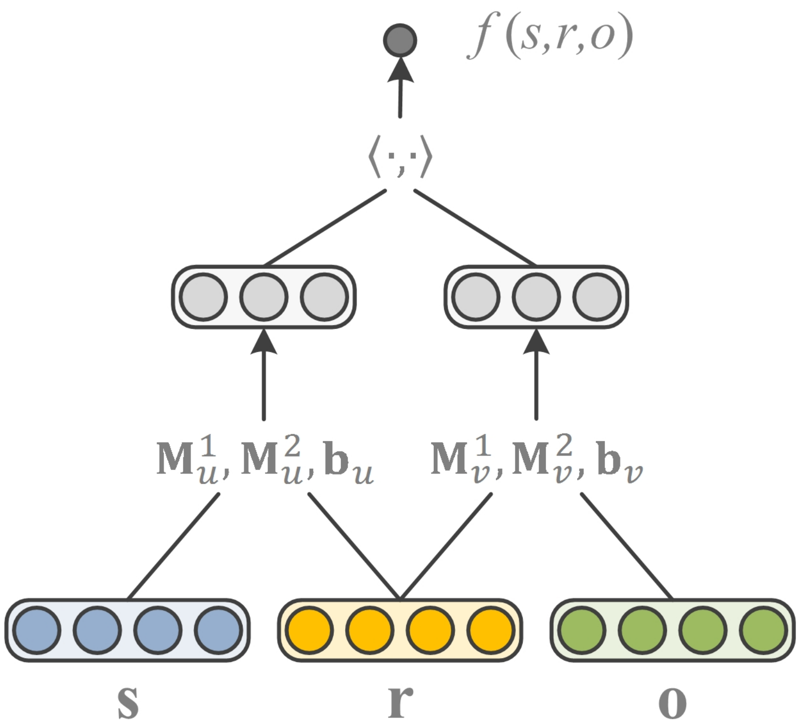
Part I: Models based on RDF triples

TOPICS:

- PROBLEM OVERVIEW
- TRANSE AND ITS VARIATIONS
- RESCAL AND ITS VARIATIONS
- DEEP NEURAL NETWORK ARCHITECTURES
- MODEL TRAINING

Semantic matching energy (SME) (Bordes et al., 2014)

- Combine relation with subject to get g_u, with object to get g_v, and match g_u and g_v by dot product



 # parameters: $\mathcal{O}(nd + md)$

Triple scoring function:

$$f(s,r,o) = g_u(\mathbf{s},\mathbf{r})^\top g_v(\mathbf{o},\mathbf{r})$$

SME (linear):

$$g_u(\mathbf{s},\mathbf{r}) = \mathbf{M}_u^1\mathbf{s} + \mathbf{M}_u^2\mathbf{r} + \mathbf{b}_u$$

$$g_v(\mathbf{o},\mathbf{r}) = \mathbf{M}_v^1\mathbf{o} + \mathbf{M}_v^2\mathbf{r} + \mathbf{b}_v$$

SME (bilinear):

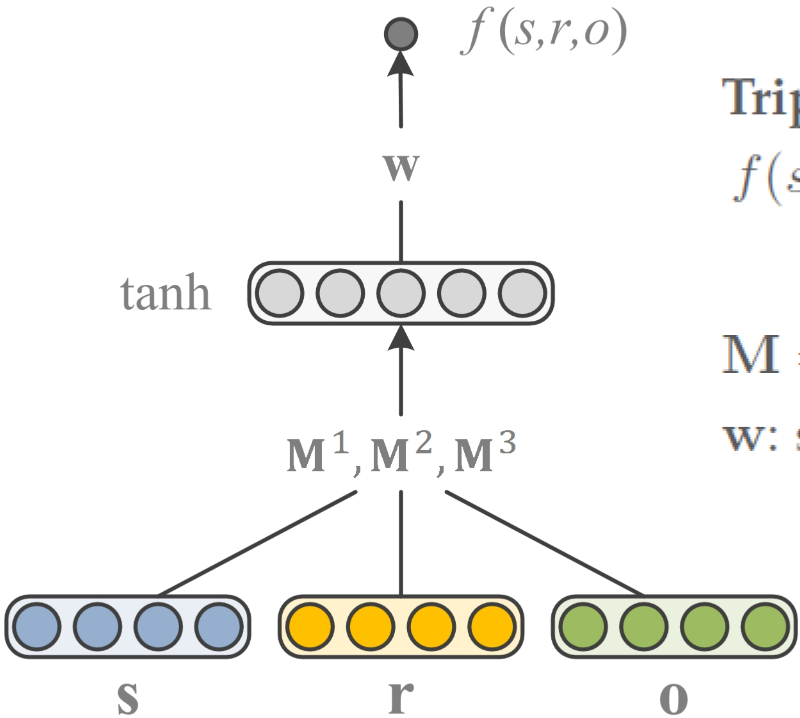
$$g_u(\mathbf{s},\mathbf{r}) = (\mathbf{M}_u^1\mathbf{s}) \odot (\mathbf{M}_u^2\mathbf{r}) + \mathbf{b}_u$$

$$g_v(\mathbf{o},\mathbf{r}) = (\mathbf{M}_v^1\mathbf{o}) \odot (\mathbf{M}_v^2\mathbf{r}) + \mathbf{b}_v$$

$\odot$ is the Hadamard product

Multi-layer perceptron (MLP) (Dong et al., 2014)

- Concatenate subject, relation, object as input, and employ a standard multi-layer perceptron



Triple scoring function:

$$
\begin{aligned}
f(s, r, o) &= \mathbf{w}^\top \tanh\left(\mathbf{M}[\mathbf{s}; \mathbf{r}; \mathbf{o}]\right) \\
&= \mathbf{w}^\top \tanh(\mathbf{M}^1\mathbf{s} + \mathbf{M}^2\mathbf{r} + \mathbf{M}^3\mathbf{o})
\end{aligned}
$$

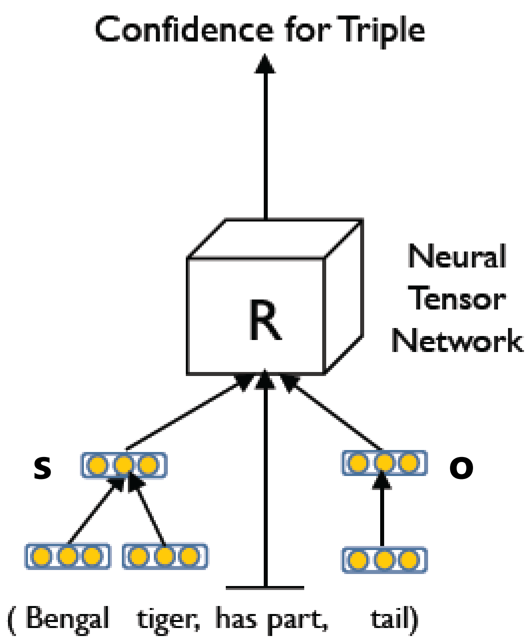
$\mathbf{M} = [\mathbf{M}^1, \mathbf{M}^2, \mathbf{M}^3]$: first layer weights

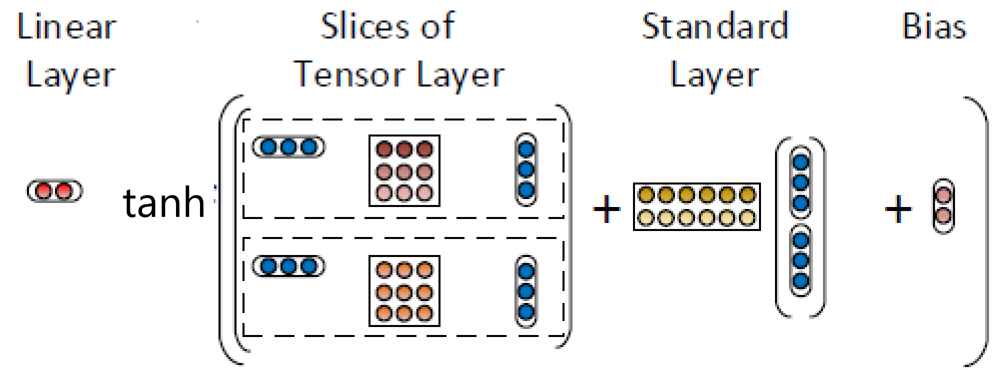
$\mathbf{w}$: second layer weights

 # parameters: $\mathcal{O}(nd + md)$

Neural tensor network (NTN) (Socher et al., 2013)

- An expressive neural network that models relations as tensors (along with matrices and vectors)





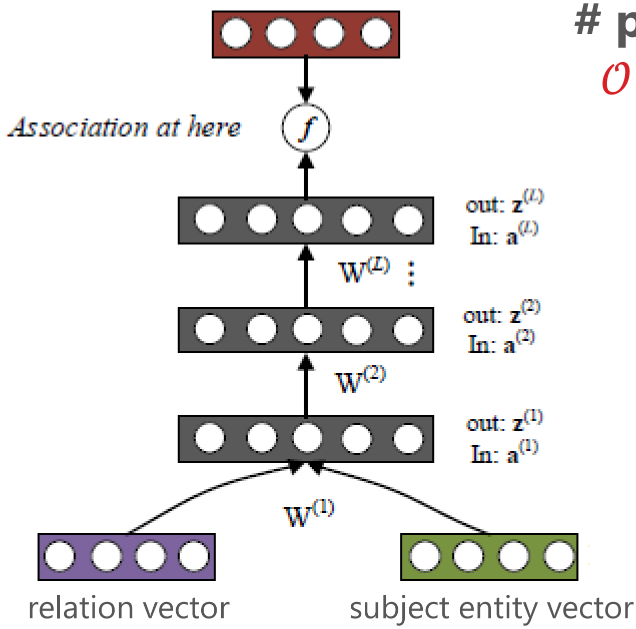
$$f(s,r,o) = \mathbf{r}^\top \tanh(\mathbf{s}^\top \underline{\mathbf{M}}_r \mathbf{o} + \mathbf{M}_r \begin{bmatrix} \mathbf{s} \\ \mathbf{o} \end{bmatrix} + \mathbf{b}_r)$$

$$[\mathbf{s}^\top \underline{\mathbf{M}}_r \mathbf{o}]_i = \mathbf{s}^\top \underline{\mathbf{M}}_r^{[:,:,i]} \mathbf{o}$$

 # parameters: $\mathcal{O}(nd + md^2k)$

Neural association model (NAM) (Liu et al., 2016)

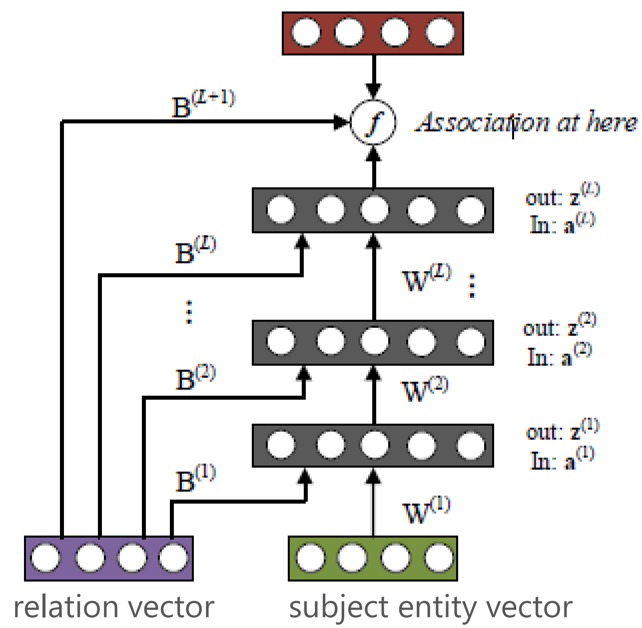
- Model subject-relation-object triples with a deep architecture



$$
\begin{aligned}
&\mathbf{z}^{(0)} = [\mathbf{s}; \mathbf{r}] \\
&\text{For } \ell = 1, 2, \cdots, L \\
&\qquad \mathbf{a}^{(\ell)} = \mathbf{M}^{(\ell)} \mathbf{z}^{(\ell-1)} + \mathbf{b}^{(\ell)} \\
&\qquad \mathbf{z}^{(\ell)} = \mathrm{ReLU}(\mathbf{a}^{(\ell)}) \\
&f(s, r, o) = \mathbf{o}^\top \mathbf{z}^{(L)}
\end{aligned}
$$

 # parameters:

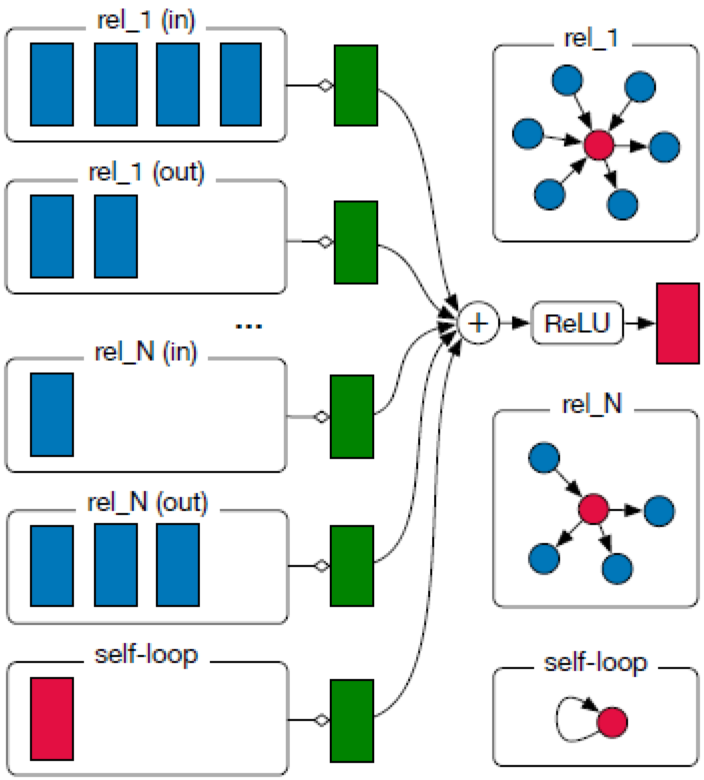
$\mathcal{O}(nd + md)$



$$
\begin{aligned}
&\mathbf{z}^{(0)} = \mathbf{s} \\
&\text{For } \ell = 1, 2, \cdots, L \\
&\qquad \mathbf{a}^{(\ell)} = \mathbf{M}^{(\ell)} \mathbf{z}^{(\ell-1)} + \mathbf{B}^{(\ell)} \mathbf{r} \\
&\qquad \mathbf{z}^{(\ell)} = \mathrm{ReLU}(\mathbf{a}^{(\ell)}) \\
&f(s, r, o) = \mathbf{o}^\top (\mathbf{z}^{(L)} + \mathbf{B}^{(L+1)} \mathbf{r})
\end{aligned}
$$

Relational graph convolutional network (R-GCN) (Schlichtkrull et al., 2018)

- Message passing on knowledge graphs with relation-specific transformations (type + direction)



$$\mathbf{h}_i^{(\ell+1)} = \sigma\left(\sum_{r\in\mathcal{R}}\sum_{j\in\mathcal{N}_i^r}\frac{1}{|\mathcal{N}_i^r|}\mathbf{W}_r^{(\ell)}\mathbf{h}_j^{(\ell)} + \mathbf{W}_0^{(\ell)}\mathbf{h}_i^{(\ell)}\right)$$

$\mathbf{h}_i^{(\ell)}$: hidden state of the i-th entity in the ℓ-th layer

$\mathcal{N}_i^r$: neighbor indices of the i-th entity under relation r

$\mathbf{W}_r^{(\ell)}$: transformation w.r.t. relation r in the ℓ-th layer

$\mathbf{W}_0^{(\ell)}$: transformation w.r.t. self-loop in the ℓ-th layer

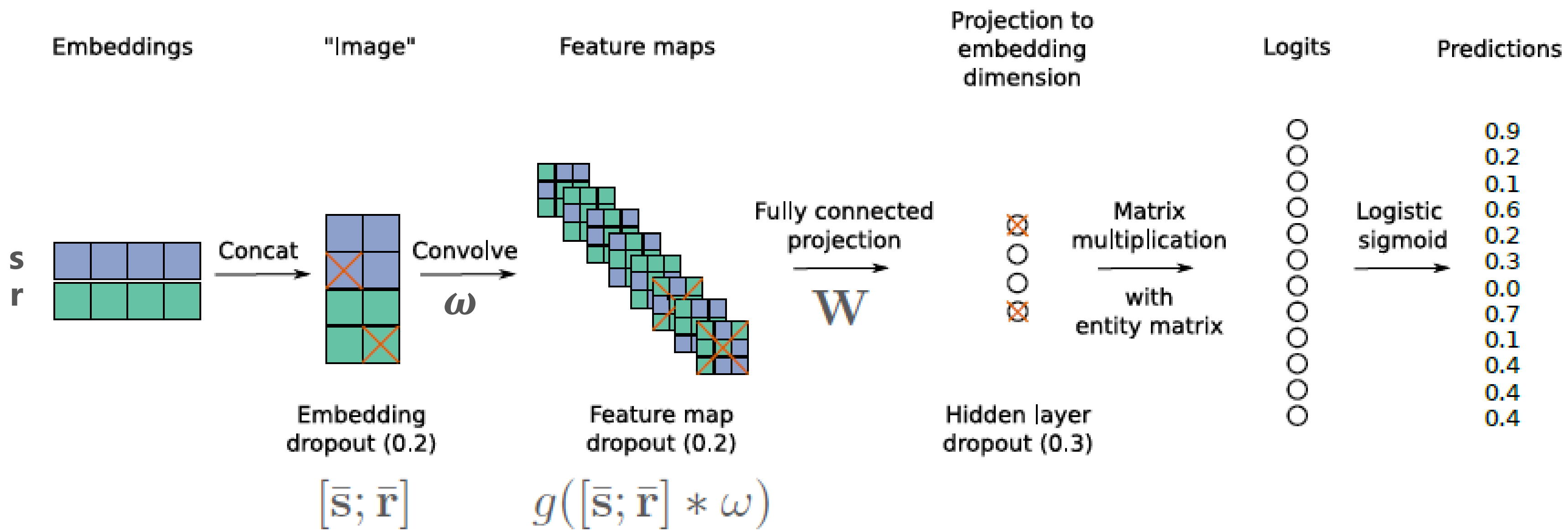
Parameter sharing & enforcing sparsity

$$\mathbf{W}_r^{(\ell)} = \sum_{b=1}^{B} a_{rb}^{(\ell)}\mathbf{V}_b^{(\ell)} \qquad \mathbf{W}_r^{(\ell)} = \bigoplus_{b=1}^{B}\mathbf{Q}_{rb}^{(\ell)}$$

 # parameters: $\mathcal{O}(ndL + mdL)$

Convolutional 2D embeddings (ConvE) (Dettmers et al., 2018)

- Reshape subject/relation vectors into matrices, and apply 2D convolution across concatenated input

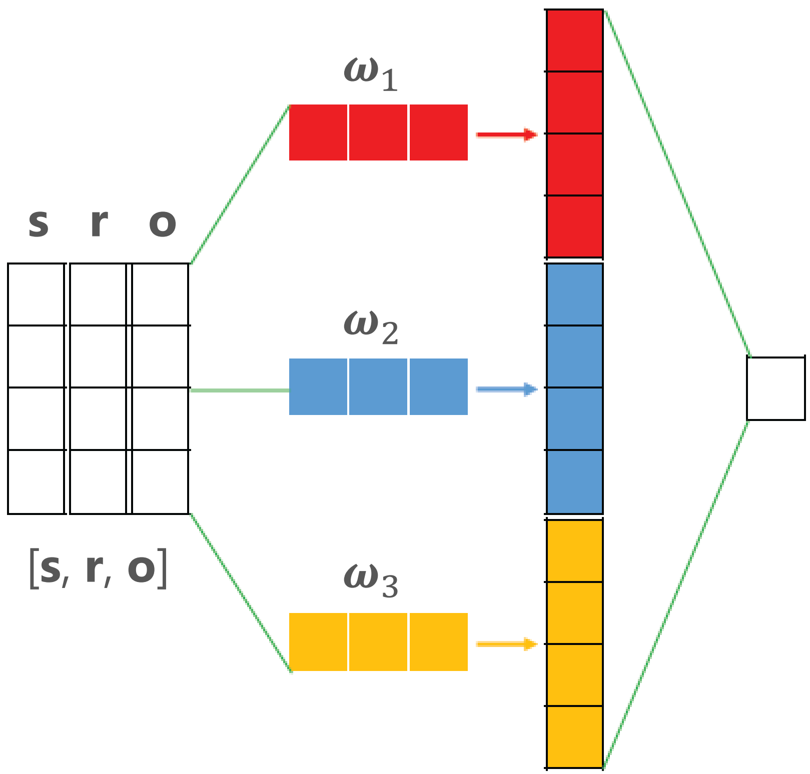


$$f(s, r, o) = \mathbf{o}^\top g\left(\mathbf{W}\mathrm{vec}\left(g([\bar{\mathbf{s}}; \bar{\mathbf{r}}] * \omega)\right)\right), \quad g = \mathrm{ReLU}$$

 # parameters: $\mathcal{O}(nd + md')$

ConvKB (Nguyen et al., 2018)

- Concatenate subject/relation/object vectors into $d \times 3$ matrices, and conduct convolution with multiple 1×3 filters



$$f(s, r, o) = \mathbf{w}^\top \text{vec}\,(g([\mathbf{s}, \mathbf{r}, \mathbf{o}] * \boldsymbol{\Omega}))$$
$$\boldsymbol{\Omega} = \{\omega_1, \omega_2, \omega_3\}, \quad g = \text{ReLU}$$

$$\boldsymbol{\Omega} = \{\omega\}, \quad \omega = [1; 1; -1]$$
$$\mathbf{w} = \mathbf{1}, \quad g(x) = |x| \text{ or } x^2$$

ConvKB = TransE

 # parameters: $\mathcal{O}(nd + md)$

Comparison among different methods

- Link prediction performance (%) on WN18 and FB15k

	WN18					FB15k				
	MRR (filt.)	MRR (raw)	Hits@1 (filt.)	Hits@3 (filt.)	Hits@10 (filt.)	MRR (filt.)	MRR (raw)	Hits@1 (filt.)	Hits@3 (filt.)	Hits@10 (filt.)
TransE[1]	49.5	35.1	11.3	88.8	94.3	46.3	22.2	29.7	57.8	74.9
TransR[1]	60.5	42.7	33.5	87.6	94.0	34.6	19.8	21.8	40.4	58.2
RESCAL[1]	89.0	60.3	84.2	90.4	92.8	35.4	18.9	23.5	40.9	58.7
DistMult[2]	82.2	53.2	72.8	91.4	93.6	65.4	24.2	54.6	73.3	82.4
HolE[1]	93.8	61.6	93.0	94.5	94.9	52.4	23.2	40.2	61.3	73.9
ComplEx[2]	94.1	58.7	93.6	94.5	94.7	69.2	24.2	59.9	75.9	84.0
ANALOGY[2]	94.2	65.7	93.9	94.4	94.7	72.5	25.3	64.6	78.5	85.4
MLP[1]	71.2	52.8	62.6	77.5	86.3	28.8	15.5	17.3	31.7	50.1
R-GCN[3]	81.4	---	69.7	92.9	96.4	69.6	---	60.1	76.0	84.2
ConvE[3]	94.2	---	93.5	94.7	95.5	74.5	---	67.0	80.1	87.3

[1]Results taken from: Nickel et al. (2016). Holographic embeddings of knowledge graphs. AAAI'16.
[2]Results taken from: Liu et al. (2017). Analogical inference for multi-relational embeddings. ICML'17.
[3]Results taken from: Dettmers et al. (2018). Convolutional 2D knowledge graph embeddings. AAAI'18.

Part I: Models based on RDF triples

TOPICS:

- PROBLEM OVERVIEW
- TRANSE AND ITS VARIATIONS
- RESCAL AND ITS VARIATIONS
- DEEP NEURAL NETWORK ARCHITECTURES
- <u>MODEL TRAINING</u>

Training under open world assumption

- Open world assumption (OWA): Missing triples are considered as unobserved data rather as negative examples
- Training data: positive triple set $\mathcal{T}^+$, negative triple set $\mathcal{T}^-$
- Optimization problem
  - Logistic loss

$$\min \sum_{\tau \in \mathcal{T}^+ \cup \mathcal{T}^-} \log\left(1 + \exp(-y_\tau \cdot f(\tau))\right) \quad y_\tau = \begin{cases} +1, & \tau = (s, r, o) \in \mathcal{T}^+ \\ -1, & \tau = (s, r, o) \in \mathcal{T}^- \end{cases}$$

$$=$$

  - Cross-entropy loss

$$\min \sum_{\tau \in \mathcal{T}^+ \cup \mathcal{T}^-} \left[-y_\tau \cdot \log(\sigma(f(\tau))) - (1 - y_\tau) \cdot \log(1 - \sigma(f(\tau)))\right] \quad y_\tau = \begin{cases} 1, & \tau \in \mathcal{T}^+ \\ 0, & \tau \in \mathcal{T}^- \end{cases}$$

  - Pairwise ranking loss

$$\min \sum_{\tau^+ \in \mathcal{T}^+} \sum_{\tau^- \in \mathcal{T}^-} \max\left(0, \gamma - f(\tau^+) + f(\tau^-)\right) \quad \begin{cases} \tau^+ = (s, r, o) \in \mathcal{T}^+ \\ \tau^- = (s', r', o') \in \mathcal{T}^- \end{cases}$$

Generating negative training examples

- Replacing s/o with a random entity sampled uniformly from $\mathcal{E}$

$$\mathcal{T}^- = \{(s',r,o)|s' \in \mathcal{E} \wedge s' \neq s \wedge (s,r,o) \in \mathcal{T}^+\}$$
$$\cup \{(s,r,o')|o' \in \mathcal{E} \wedge o' \neq o \wedge (s,r,o) \in \mathcal{T}^+\}$$

(Cristiano Ronaldo, bornIn, Funchal) → (**Sergio Ramos**, bornIn, Funchal)

(Cristiano Ronaldo, bornIn, **Camas**)

- Replacing r with a random relation sampled uniformly from $\mathcal{R}$

$$\mathcal{T}^- = \{(s',r,o)|s' \in \mathcal{E} \wedge s' \neq s \wedge (s,r,o) \in \mathcal{T}^+\}$$
$$\cup \{(s,r,o')|o' \in \mathcal{E} \wedge o' \neq o \wedge (s,r,o) \in \mathcal{T}^+\}$$
$$\cup \{(s,r',o)|r' \in \mathcal{R} \wedge r' \neq r \wedge (s,r,o) \in \mathcal{T}^+\}$$

(Cristiano Ronaldo, bornIn, Funchal) → (Cristiano Ronaldo, **playsFor**, Funchal)

Generating negative training examples (cont.)

- **Defect of uniformly sampling:** False-negative examples

  (Cristiano Ronaldo, gender, Male) → (**Sergio Ramos**, gender, Male)

- **Solution:** Giving more chance to replace s for one-to-many relations, and o for many-to-one relations (Wang et al., 2014)
  - Probability of replacing s: $ops/(ops + spo)$
  - Probability of replacing o: $spo/(ops + spo)$
  - ops: the average number of object entities per subject
  - spo: the average number of subject entities per object

  (Cristiano Ronaldo, gender, Male) → (**?** , gender, Male) **p=2%**

  (Cristiano Ronaldo, gender, **?**) **p=98%**

Generating negative training examples (cont.)

- **Defect of uniformly sampling:** Too easy negative examples

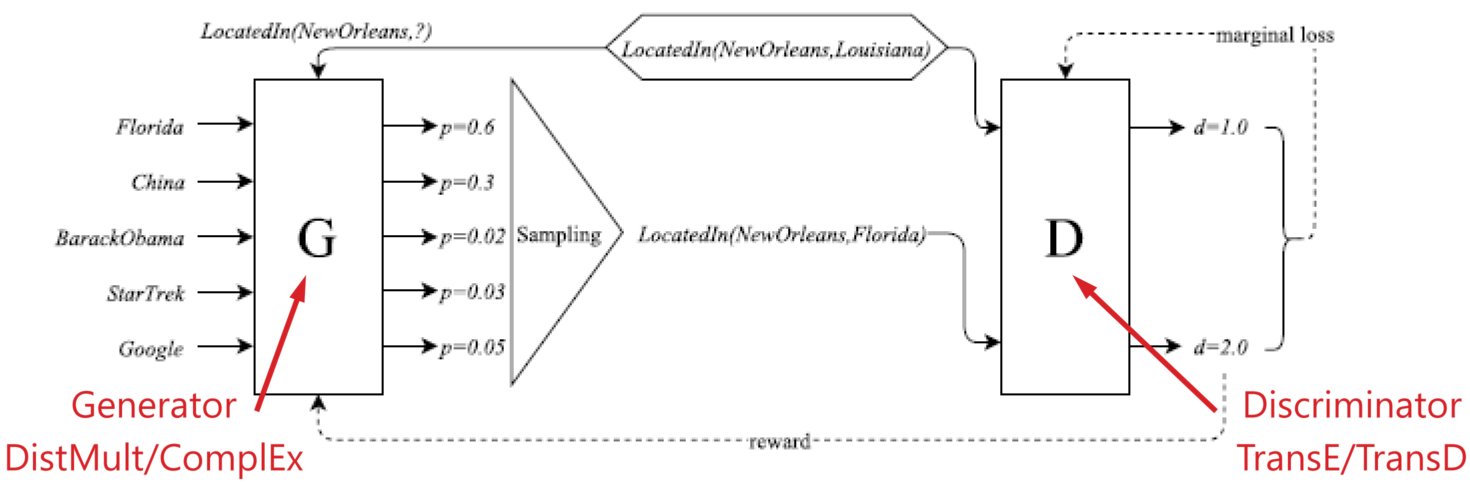
  (Cristiano Ronaldo, gender, Male) → (Cristiano Ronaldo, gender, **Funchal**)

- **Solution I:** Corrupting a position using entities compatible with the position (Krompaß et al., 2015)

  (Cristiano Ronaldo, gender, Male) → (Cristiano Ronaldo, gender, **Female**) √

  (Cristiano Ronaldo, gender, **Funchal**) ✕

- **Solution II:** Adversarial learning (Cai and Wang, 2018)



Generating negative training examples (cont.)

- Negative examples generated by uniformly sampling and adversarial learning (Cai and Wang, 2018)

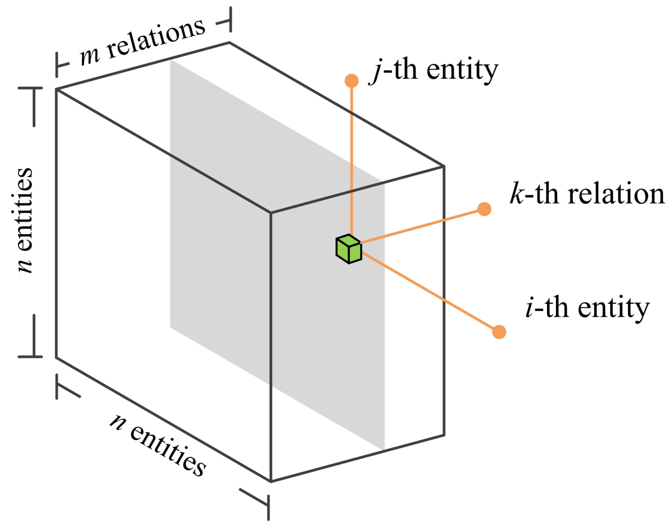
Positive fact	**Uniform random sample**	**Trained generator**
(condensation_NN_2, derivationally_related_form, **distill_VB_4**)	family_arcidae_NN_1 repast_NN_1 beater_NN_2 coverall_NN_1 cash_advance_NN_1	revivification_NN_1 mouthpiece_NN_3 *liquid_body_substance_NN_1* stiffen_VB_2 *hot_up_VB_1*
(**colorado_river_NN_2**, instance_hypernym, river_NN_1)	lunar_calendar_NN_1 umbellularia_californica_NN_1 tonality_NN_1 creepy-crawly_NN_1 moor_VB_3	*idaho_NN_1* *sayan_mountains_NN_1* *lower_saxony_NN_1* order_ciconiiformes_NN_1 jab_NN_3
(**meeting_NN_2**, hypernym, social_gathering_NN_1)	cellular_JJ_1 commercial_activity_NN_1 giant_cane_NN_1 streptomyces_NN_1 tranquillize_VB_1	*attach_VB_1* *bond_NN_6* heavy_spar_NN_1 satellite_NN_1 peep_VB_3

Training under closed world assumption

- Closed world assumption (CWA): All missing triples are taken as negative examples
- Training data: All possible triples $\mathcal{E} \times \mathcal{R} \times \mathcal{E}$
- Optimization problem:
  - Squared loss

$$\min \sum_{(s,r,o) \in \mathcal{E} \times \mathcal{R} \times \mathcal{E}} \left(y_{sro} - f(s,r,o)\right)^2$$

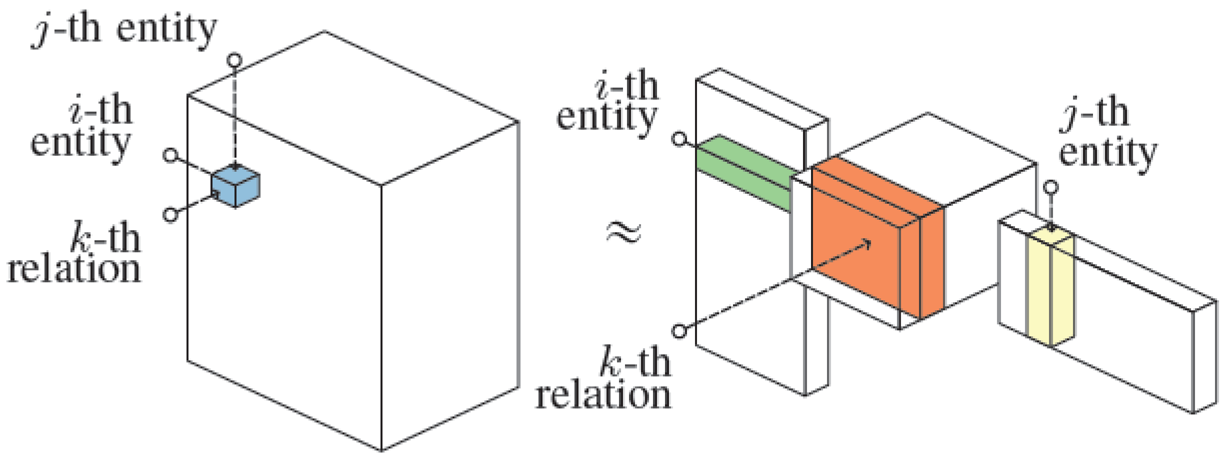
$$y_{sro} = \begin{cases} 1, & (s,r,o) \in \mathcal{T}^+ \\ 0, & \text{otherwise} \end{cases}$$



tensor factorization

RESCAL as tensor factorization

- RESCAL as factorization of tensor $\underline{\mathbf{Y}}$



- Loss in matrix form

$$\min \sum_k \|\underline{\mathbf{Y}}^{[:,:,k]} - \mathbf{E}\mathbf{M}_k\mathbf{E}^\top\|_F^2$$

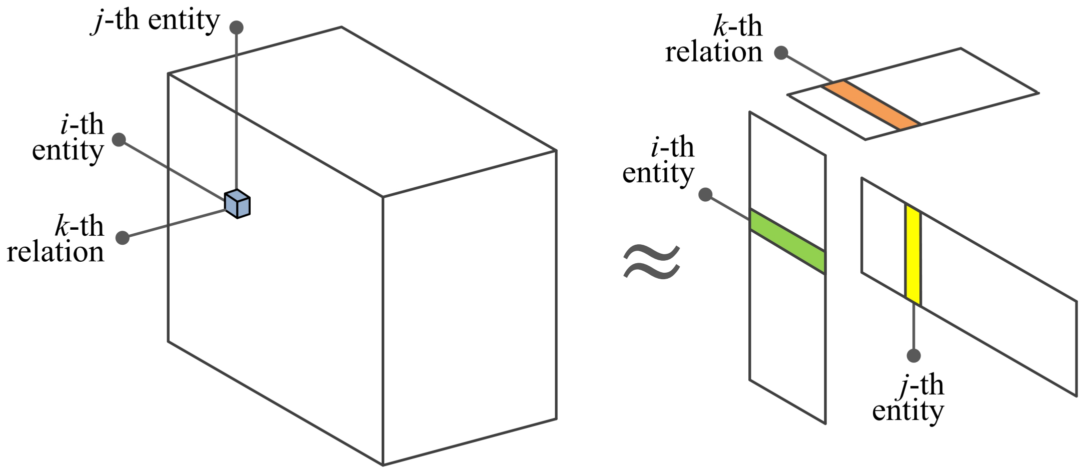
- Loss in element-wise form

$$\min \sum_{i,j,k} \left(y_{ijk} - \mathbf{e}_i^\top \mathbf{M}_k \mathbf{e}_j\right)^2$$

bilinear scoring function

Other tensor factorization models

- The CP tensor decomposition

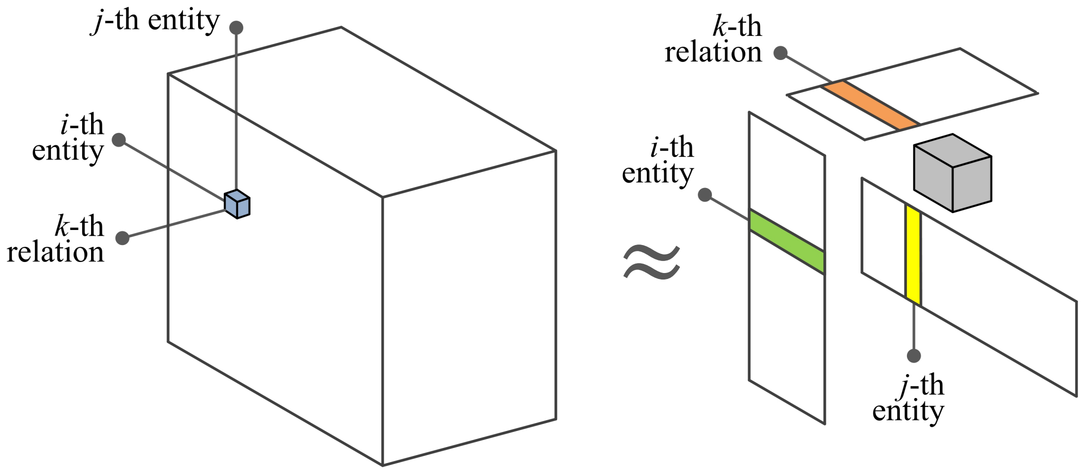


$$y_{ijk} \approx \sum_{\ell=1}^{d} e_{i\ell}^{(s)} \cdot r_{k\ell} \cdot e_{j\ell}^{(o)}$$

- $\mathbf{e}_i^{(s)} \in \mathbb{R}^d$: embedding of the i-th entity when it appears as a subject
- $\mathbf{e}_j^{(o)} \in \mathbb{R}^d$: embedding of the j-th entity when it appears as an object
- $\mathbf{r}_k \in \mathbb{R}^d$: embedding of the k-th relation

Other tensor factorization models (cont.)

- The TUCKER tensor decomposition

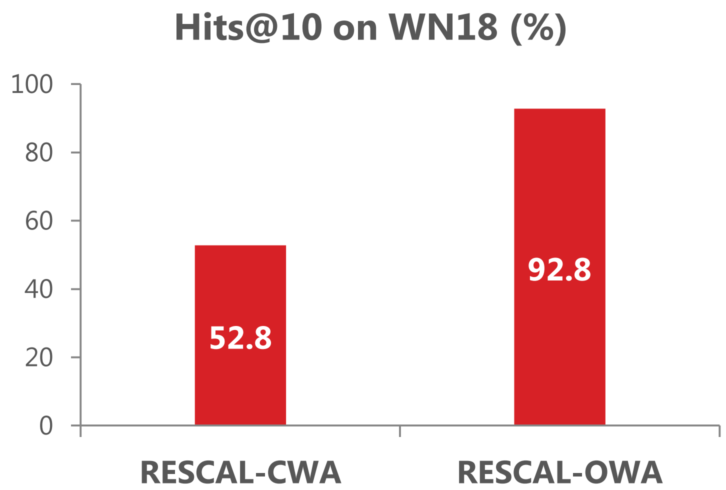


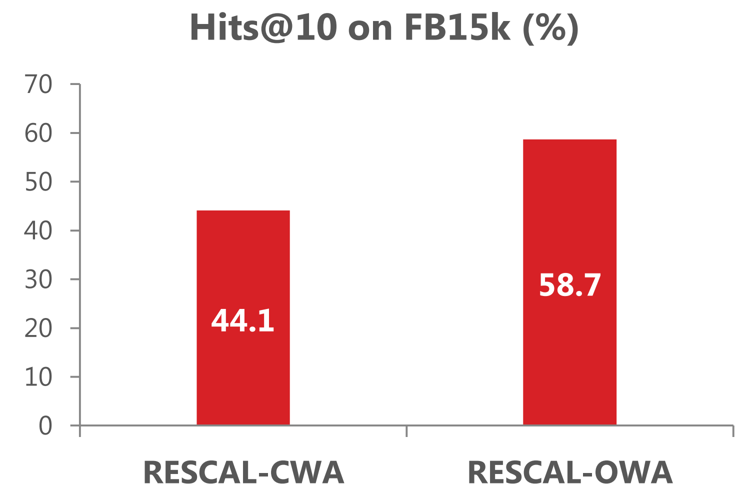
$$y_{ijk} \approx \sum_{a=1}^{d_1} \sum_{b=1}^{d_2} \sum_{c=1}^{d_3} g_{abc} \cdot e_{ia}^{(s)} \cdot r_{kc} \cdot e_{jb}^{(o)}$$

- $\mathbf{e}_i^{(s)} \in \mathbb{R}^{d_1}$: embedding of the i-th entity when it appears as a subject
- $\mathbf{e}_j^{(o)} \in \mathbb{R}^{d_2}$: embedding of the j-th entity when it appears as an object
- $\mathbf{r}_k \in \mathbb{R}^{d_3}$: embedding of the k-th relation; $\underline{\mathbf{G}} \in \mathbb{R}^{d_1 \times d_2 \times d_3}$: core tensor

Defects of closed world assumption

- Unfit for incomplete knowledge graphs (common in real life), usually performing worse than OWA[1]





- Introducing too many negative examples, which may cause scalability issues during model training

[1]Results taken from: Bordes et al. (2013). Translating embeddings for modeling multi-relational data. NIPS'13.
Nickel et al. (2016). Holographic embeddings of knowledge graphs. AAAI'16.

Review

Problem

- To learn distributed representations of entities and relations from subject-relation-object triples

Approaches

- TransE and its variations
- RESCAL and its variations
- Deep neural network architectures

Model training

- Training under open world assumption
- Training under closed world assumption (tensor factorization)

References

[Bordes et al., 2014] Bordes, A., Glorot, X., Weston, J., and Bengio, Y. (2014). A semantic matching energy function for learning with multi-relational data. MACH LEARN.

[Bordes et al., 2013] Bordes, A., Usunier, N., Garc´ıa-Durán, A., Weston, J., and Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In NIPS'13.

[Cai and Wang, 2018] Cai, L. and Wang, W. Y. (2018). KBGAN: Adversarial learning for knowledge graph embeddings. In NAACL'18.

[Dettmers et al., 2018] Dettmers, T., Pasquale, M., Pontus, S., and Riedel, S. (2018). Convolutional 2D knowledge graph embeddings. In AAAI'18.

[Dong et al., 2014] Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., and Zhang, W. (2014). Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In SIGKDD'14.

[Fan et al., 2014] Fan, M., Zhou, Q., Chang, E., and Zheng, T. F. (2014). Transition-based knowledge graph embedding with relational mapping properties. In PACLIC'14.

[Feng et al., 2016] Feng, J., Zhou, M., Hao, Y., Huang, M., and Zhu, X. (2016). Knowledge graph embedding by flexible translation. In KR'16.

[Hayashi and Shimbo, 2017] Hayashi, K. and Shimbo, M. (2017). On the equivalence of holographic and complex embeddings for link prediction. In ACL'17.

[He et al., 2015] He, S., Liu, K., Ji, G., and Zhao, J. (2015). Learning to represent knowledge graphs with Gaussian embedding. In CIKM'15.

References

[Krompaß et al., 2015] Krompaß, D., Baier, S., and Tresp, V. (2015). Type-constrained representation learning in knowledge graphs. In ISWC'15.

[Lin et al., 2015] Lin, Y., Liu, Z., Sun, M., Liu, Y., and Zhu, X. (2015). Learning entity and relation embeddings for knowledge graph completion. In AAAI'15.

[Liu et al., 2017] Liu, H., Wu, Y., and Yang, Y. (2017). Analogical inference for multi-relational embeddings. In ICML'17.

[Liu et al., 2016] Liu, Q., Jiang, H., Evdokimov, A., Ling, Z.-H., Zhu, X., Wei, S., and Hu, Y. (2016). Probabilistic reasoning via deep learning: Neural association models. arXiv.

[Nguyen et al., 2018] Nguyen, D. Q., Nguyen, T. D., Nguyen, D. Q., and Phung, D. (2018). A novel embedding model for knowledge base completion based on convolutional neural network. In NAACL'18.

[Nickel et al., 2016] Nickel, M., Rosasco, L., and Poggio, T. (2016). Holographic embeddings of knowledge graphs. In AAAI'16.

[Nickel et al., 2011] Nickel, M., Tresp, V., and Kriegel, H.-P. (2011). A three-way model for collective learning on multi-relational data. In ICML'11.

[Schlichtkrull et al., 2018] Schlichtkrull, M., Kipf, T., Bloem, P., Berg, R. v. d., Titov, I., and Welling, M. (2018). Modeling relational data with graph convolutional networks. In ESWC'18.

[Socher et al., 2013] Socher, R., Chen, D., Manning, C., and Ng, A. (2013). Reasoning with neural tensor networks for knowledge base completion. In NIPS'13.

References

[Trouillon et al., 2016] Trouillon, T., Welbl, J., Riedel, S., Gaussier, E., and Bouchard, G. (2016). Complex embeddings for simple link prediction. In ICML'16.

[Wang et al., 2017] Wang, Q., Mao, Z., Wang, B., and Guo, L. (2017). Knowledge graph embedding: A survey of approaches and applications. IEEE TKDE.

[Wang et al., 2014] Wang, Z., Zhang, J., Feng, J., and Chen, Z. (2014). Knowledge graph embedding by translating on hyperplanes. In AAAI'14.

[Xiao et al., 2016a] Xiao, H., Huang, M., and Zhu, X. (2016a). From one point to a manifold: Knowledge graph embedding for precise link prediction. In IJCAI'16.

[Xiao et al., 2016b] Xiao, H., Huang, M., and Zhu, X. (2016b). TransG: A generative model for knowledge graph embedding. In ACL'16.

[Yang et al., 2015] Yang, B., Yih, W.-t., He, X., Gao, J., and Deng, L. (2015). Embedding entities and relations for learning and inference in knowledge bases. In ICLR'15.