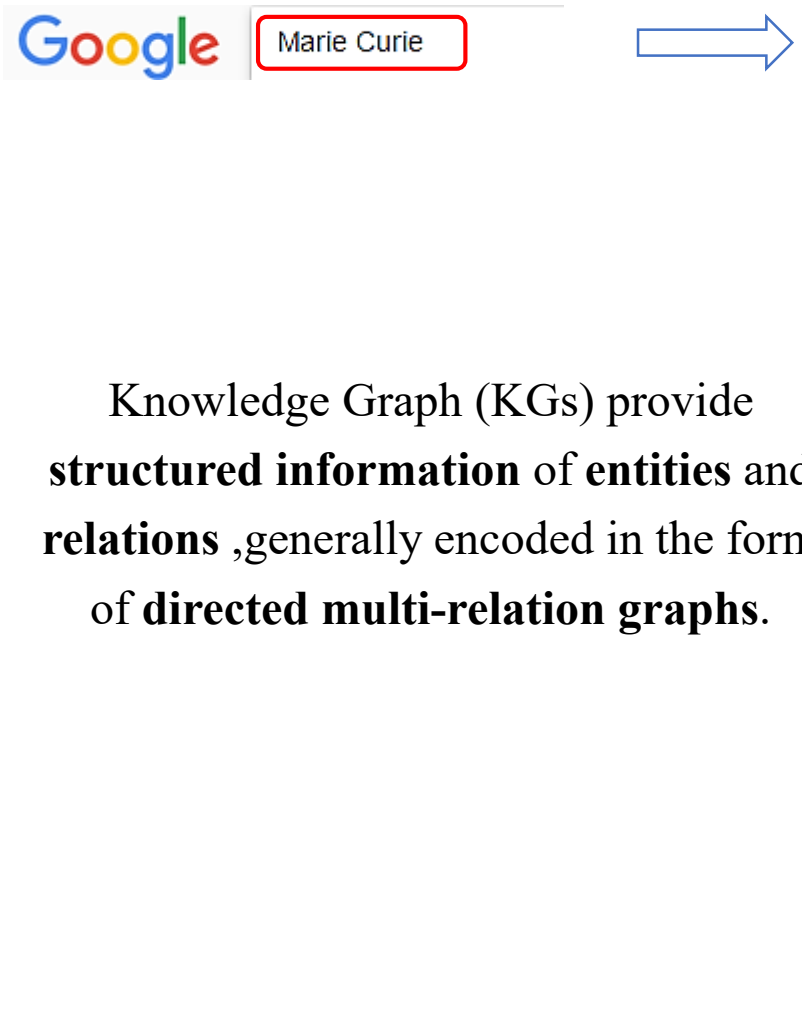
NCLab
nature computation

Convolutional Neural Network for Knowledge Graph Completion

Shao Mingguang

Knowledge Graph



Knowledge Graph (KGs) provide **structured information** of **entities** and **relations** ,generally encoded in the form of **directed multi-relation graphs**.

Application of Knowledge Graph

A large number of KGs has been created, including **Freebase**, **WordNet**, YAGO and Google Knowledge Graph.

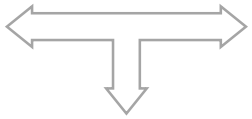
provides general facts about the word

Groups worlds into synonyms and provides lexical relations

Knowledge Graph Construction

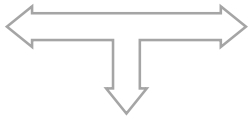
NCLab
nature computation

Corpus



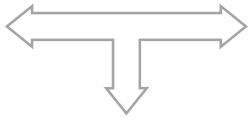
Leonard Nimoy was an actor who played the character Spock in the science-fiction movie Star Trek.

NLP: Entity recognition
Relation recognition



Leonard Nimoy was an *actor* who *played* the *character* *Spock* in the *science-fiction* movie *Star Trek*.

Triple: (head, relation, tail)
Note: Only store true facts



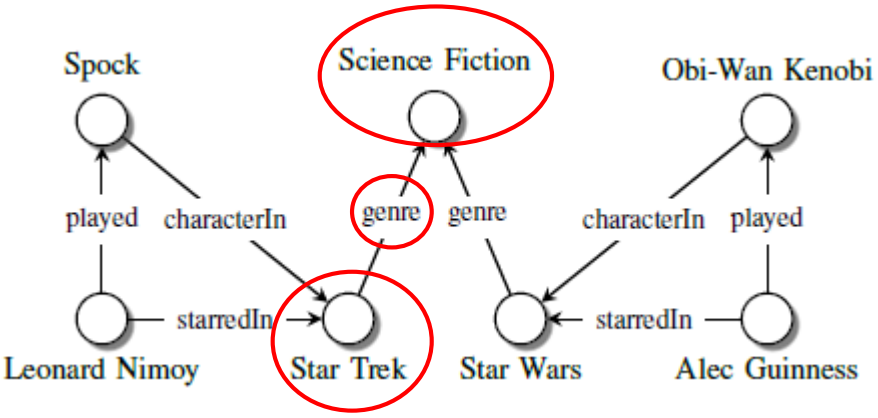
(LeonardNimoy,	*profession,*	*Actor)*
(LeonardNimoy,	*starredIn,*	*StarTrek)*
(LeonardNimoy,	*played,*	*Spock)*
(Spock,	*characterIn,*	*StarTrek)*
(StarTrek,	*genre,*	*ScienceFiction)*

Is that enough?

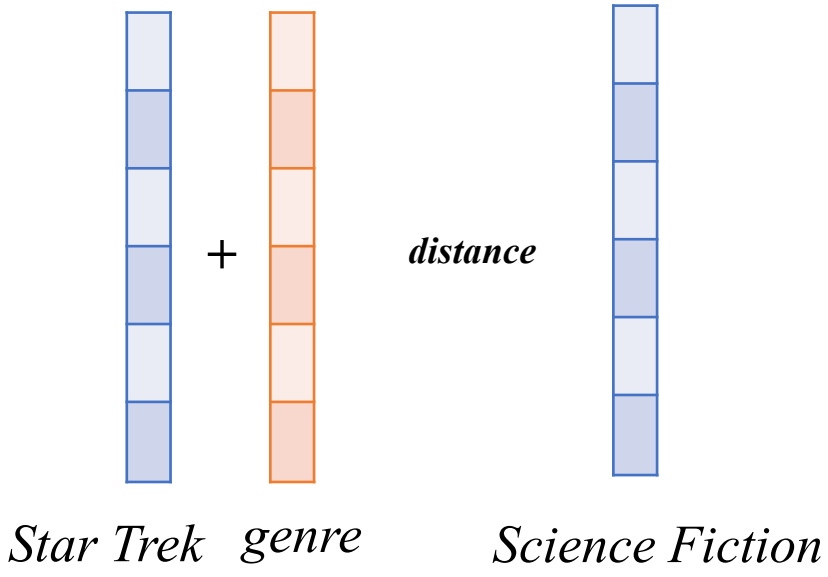
NCLab
nature computation

Problem: We can't use triple for calculations.

For example, calculate distance between *Star Trek + genre* and *Science Fiction*



Embedding to Vector Space



Knowledge Graph Completion

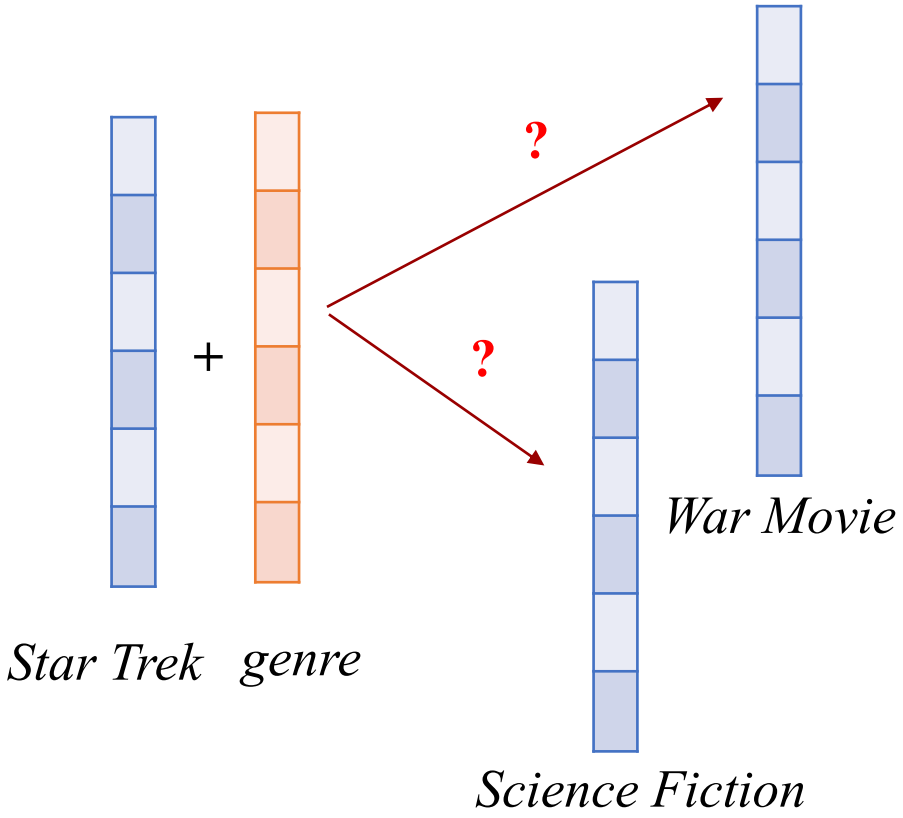
Another Problem: Incomplete knowledge

Existing Knowledge Graphs are often missing many facts.

(Leonard Nimoy ,profession ,?)

(Leonard Nimony ,?, StarTrek)

(Star Trek, genre,?)



Knowledge graph completion(KGC) aims at filling in incomplete triples like *(h,r,?),(?,r,t)* or *(h,?,t)*.

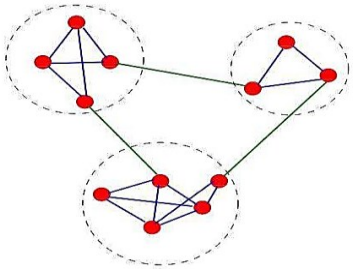
KGC MODEL

Different types of KGC models

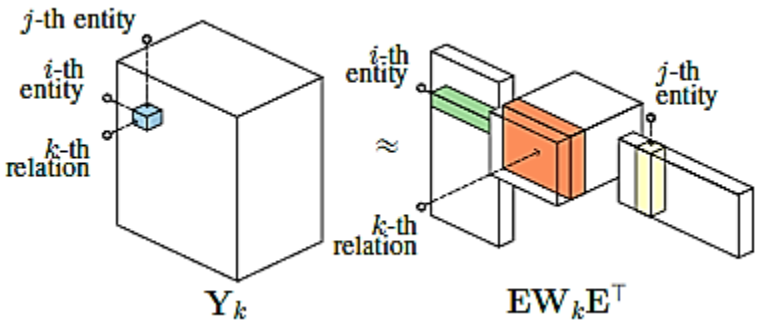
Markov Random Fields

$$F_1 : (x, parentOf, z) \wedge (y, parentOf, z) \Rightarrow (x, marriedTo, y)$$
$$F_2 : (x, marriedTo, y) \Rightarrow \neg(y, parentOf, x)$$

Graph Features Models



Embedding models



Details: A Review of Relation Machine Learning for Knowledge Graphs Maximilian Nickel

NCLab
nature computation

Three key points of a KGC Model: score function, negative sampling, training methods(loss function)

SCORE FUNCTION: the probability of the existence of *(h, r, t)*

TransE: A. Bordes . Translating Embeddings for Modeling Multi-relation Data

Score function:

$$\Pr(h,t,r) = \| h+r,t \|_2^2$$

HolE: Maximilian Nickel. Holographic Embedding of Knowledge Graphs

RESCAL: A Three-Way Model for Collective Learning on Multi-Relation Data

Score function:

$$\Pr(h,t,r) = f(r^T(h \otimes t))$$

ProjE: Baoxu Shi. ProjE: Embedding Projection for Knowledge Graph Completion

Score function:

$$\Pr(h,r) = g(W^c f(D_e h + D_r r + b_c) + b_p)$$

NEGATIVE SAMPLING: generate negative relation instance for training from p, p can be replaced by a (0,1) binomial distribution.

(h, r, t) → *(h', r, t)* *(h, r', t)* *(h, r, t')*

LOSS FUNCTION:

Pair-wise:

$$L = \sum_{(h,r,t)\in S} \sum_{(h',r,t')\in S'} \max\left(0, f_r(h,t) + \gamma - f_r(h',t')\right)$$

Point-wise:

$$\max_{\Theta} \sum_{i\in\{i|y_i=1\}} \log(g(\eta_i)) + \sum_{i\in\{j|y_j=-1\}} \log(1 - g(\eta_j)) + \lambda||\Theta||_2^2 \qquad \eta_i = \mathbf{t}^\top(\mathbf{h}\circ\mathbf{r})$$

List-wise:

$$\max_{\Theta} \log\left(\prod_{i=1}^{m} g(\eta_i)\right) + \lambda||\Theta||_2^2 \qquad \eta_i = \mathbf{t}^\top(\mathbf{h}\circ\mathbf{r})$$

Details： List approach to learning to rank - Theory and Algorithm

Details of Embedding

Leonard Nimoy	*0*
Star Trek	*1*
Science Fiction	*2*
……	

List of entities and index

profession	*0*
starred in	*1*
genre	*2*
……	

List of relations and index

（Leonard Nimoy, profession, Actor）
(Leonard Nimony, starredIn, StarTrek)
(Star Trek, genre, scienceFiction)
……

List of triples(golden triple)

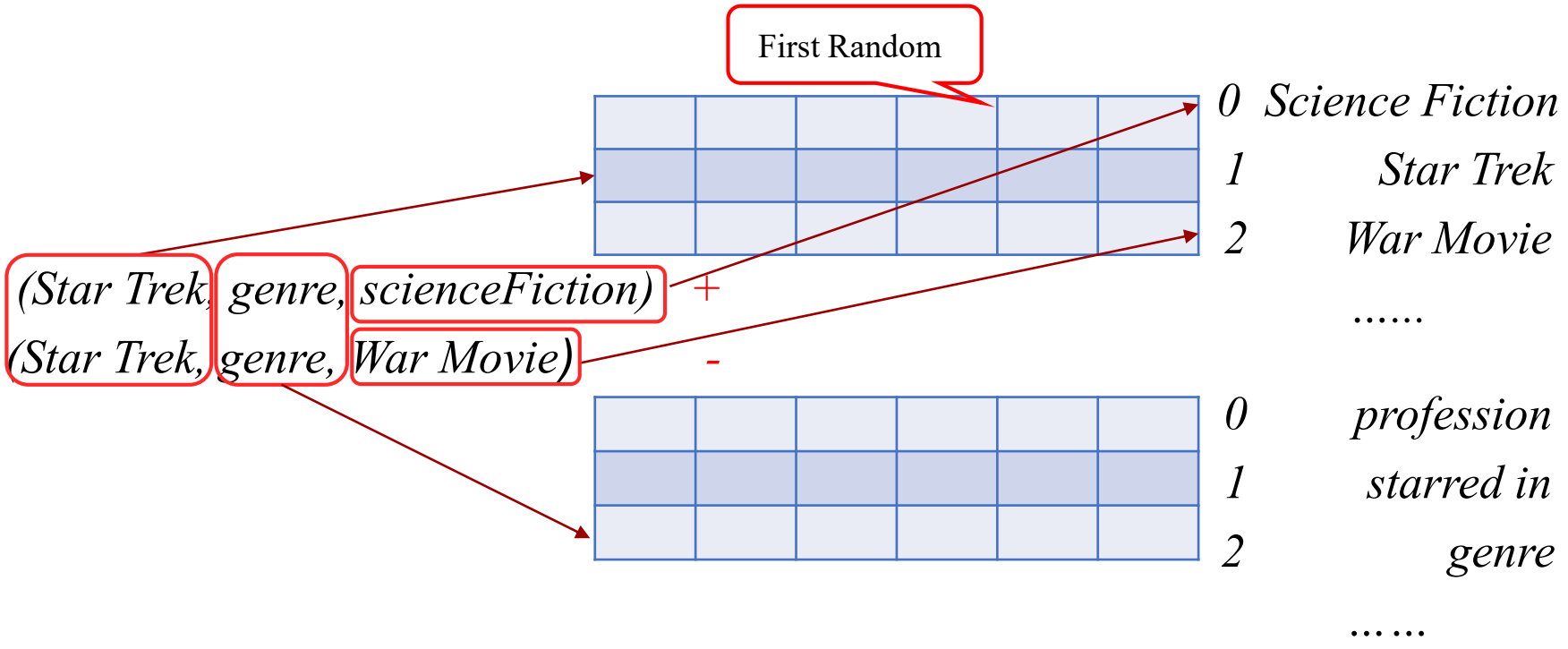
Generate negative instance

（Leonard Nimoy, profession, Actor）
(Leonard Nimony, starredIn, StarTrek)
(Star Trek, genre, scienceFiction)
……

(Star Trek, genre, scienceFiction) +
(Star Trek, genre, WarMovie) -
……

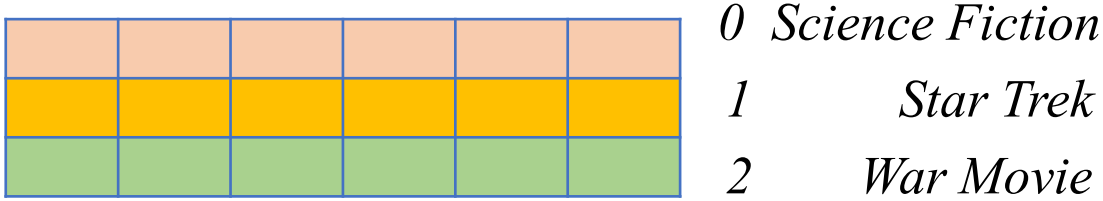
label

Details of Embedding

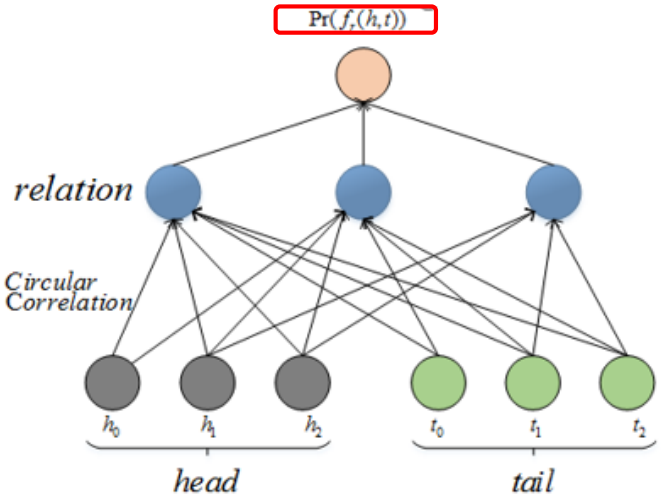


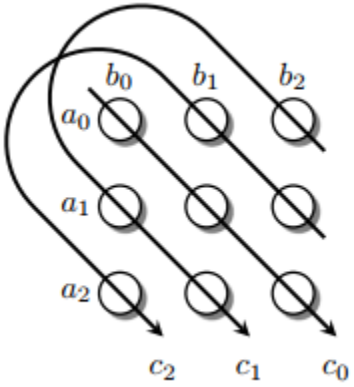
$$min \quad L = \sum_{(h,r,t)\in S} \sum_{(h',r,t')\in S'} \max\left(0, f_r(h,t) + \gamma - f_r(h',t')\right)$$

Then, update *h,t* and *r*



MODEL：RESCAL





$$\boldsymbol{c} = \boldsymbol{a} \star \boldsymbol{b}$$

$$c_0 = a_0 b_0 + a_1 b_1 + a_2 b_2$$
$$c_1 = a_0 b_2 + a_1 b_0 + a_2 b_1$$
$$c_2 = a_0 b_1 + a_1 b_2 + a_2 b_0$$

HolE: Maximilian Nickel. Holographic Embedding of Knowledge Graphs

RESCAL: A Three-Way Model for Collective Learning on Multi-Relation Data

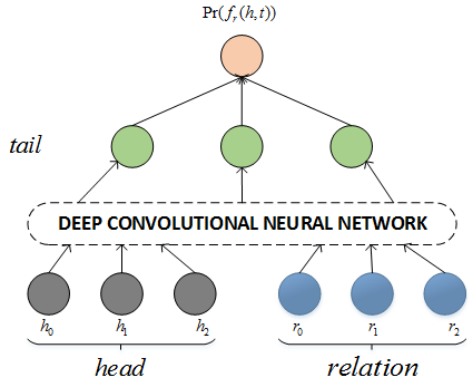
Score function:

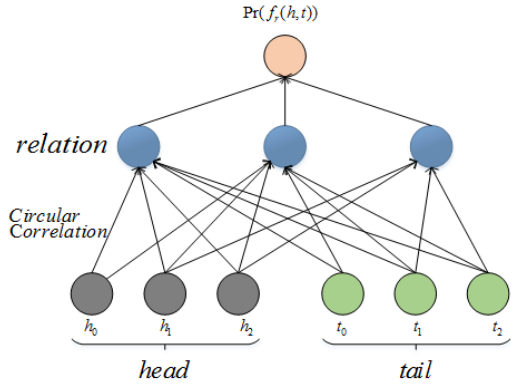
$$\Pr(h,t,r) = f(r^T (h \otimes t))$$

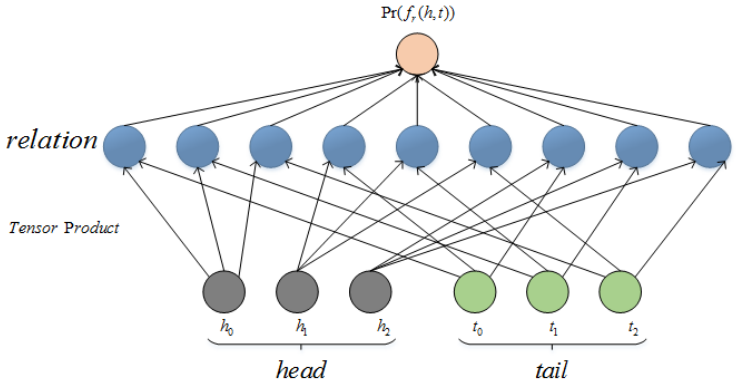
$$L = \sum_{(h,r,t)\in S} \sum_{(h',r,t')\in S'} \max\left(0, f_r(h,t) + \gamma - f_r(h',t')\right)$$

OUR MODEL: CNN-MODEL

We first introduce deep neural network and CNN to KGC:







Leaned by CNN

hand-crafted

HolE: Maximilian Nickel. Holographic Embedding of Knowledge Graphs

RESCAL: A Three-Way Model for Collective Learning on Multi-Relation Data

Score function:

$$\Pr(h,t,r) = f(r^T(h \otimes t))$$

OUR MODEL

Another training method: classification based model:

$$\max_{\Theta} \log(\prod_{i=1}^{m} g(-y_i \eta_i)) + \lambda ||\Theta||_2^2 \qquad \eta_i = \mathbf{t}^\top (\mathbf{h} \circ \mathbf{r})$$

$$\min_{\Theta} \sum_{i=1}^{m} \log(1 + \exp(-y_i \eta_i)) + \lambda ||\Theta||_2^2$$

Evaluation Protocol

- **Mean Rank:** measures the average of rank of correct entities or relations
- **Hits@k:** measures the proportion of valid entities/relations ranked in top k

The "filter" setting protocol remove all instances that appear in the knowledge graph base before ranking.

NOTE: Higher Hits@k or lower mean rank indicates better entity prediction performance.

FUTURE WORK

Negative sampling

- Random
- Corrupted
- LCWA
- PU learning

Score function

- Hand-crafted
- CNN-based

Training method

- pairwise
- pointwise
- List-wise
- classification

Reasoning

- Reinforcement learning

NCLab
nature computation

谢谢各位老师指导！