The Presentation of Entities and Relations and its Application in Inference

Weidi Xu

Peking University, Computational Intelligence Laboratory

June 10, 2015

Overview

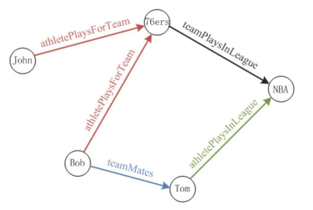
- Knowledge Graph: a directed graph with entiites and relations
- Knowledge Inference: link prediction in knowledge graph
- Representation: efficient and scalable inference methods

Table of Contents

Table of Contents

Knowledge Graph

- Knowledge graph is a directed graph comprised by entities and relations.

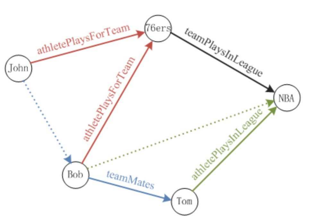


- (John, athletePlaysForTeam, 76ers)
- (Bob, athletePlaysForTeam, 76ers)
- (Tom, athleteMates, Tom)
- (Tom, athletePlaysInLeague, NBA)
- (76ers, teamPlaysInLeague, NBA)

- However the knowledge graph is incomplete: it only covers small part of facts.

Knowledge Inference

- Knowledge inference: link prediction in knowledge graph.



- (John, teamMates, Bob)
- (Bob, athletePlaysInLeague, NBA)
- ... ...

- Inference in Knowledge Graph can increase the coverage.

Methods

- Inference based on representation learning.
  1. Collective Matrix Factoriation (Nickel et al., 2011)
  2. Neural Tensor Networks (Socher et al., 2013)
  3. Translating Embeddings(Bordes et al., 2013)
- Inference based on markov random field.
  1. Markov Logic Networks (Richardson and Domingos, 2006)
  2. Probabilistic Soft Logic (Brocheler et al., 2012)
- Inference based on random walk.
  1. Path Ranking (Lao et al., 2011)

Table of Contents

Framework

- Hidden variable model: modeling data in hidden variable space.

Leanring

- Learning entity and relation embeddings in hidden vector spaces.
- Define a fitness funtion to determine the certainty of entity-relation-entity pairs.
- Define an objective funtion to determine the certainty of entity-relation-entity pairs.

Prediction

- Using the model and objective function to inference.

Framework

- Hidden variable model: modeling data in hidden variable space.

Leanring

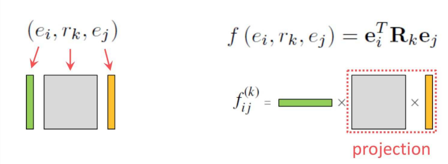
- Learning entity and relation embeddings in hidden vector spaces.
- Define a fitness funtion to determine the certainty of entity-relation-entity pairs.
- Define an objective funtion and use the facts dataset to learn the model parameters.

Objective function

- Reconstruction error
- Ranking loss

RESCAL: reconstruction error

- Representation in hidden vector/matrix space and the corresponding fitness function:



- Learn based on reconstruction error:

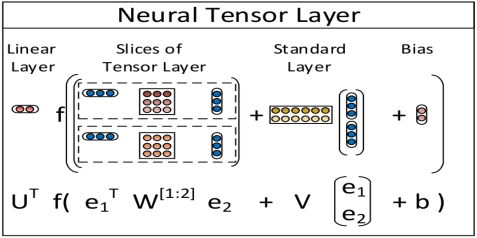
$$\min_{e_i, R_k} \sum_k \sum_i \sum_j (y_{i,j}^k - f(e_i, r_k, e_j))^2 + \lambda R \tag{1}$$

assumption

It is assumed that all pairs not contained in dataset are negative.

Neural Tensor Network: ranking loss

- Representation in hidden vector/matrix space and the corresponding fitness function:



$$f(e_i, r_k, e_j) = U_k^T g(e_i^T W_k^{[1:K]} e_j + V_k[e_i : e_j] + b_k) \qquad (2)$$

- Learn based on ranking loss:

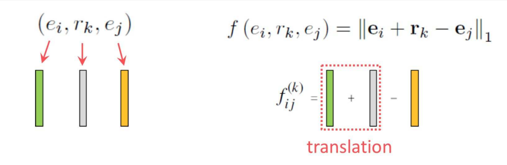
$$\min_{e_i, R_k} \sum_{t^+ \in O} \sum_{t^- \in D} (\lambda + f(e_i, r_k, e_j) - f(e_i', r_k, e_j')) \qquad (3)$$

assumption

It is assumed that all pairs not contained in dataset are negative.

TransE: ranking loss

- Representation in hidden vector space and the corresponding fitness function:



- Learn based on ranking loss:

$$\min_{e_i, R_k} \sum_{t^+ \in O} \sum_{t^- \in D} (\lambda + f(e_i, r_k, e_j) - f(e_i', r_k, e_j')) \tag{4}$$

assumption

It is assumed that pairs not contained in dataset are partially negative.

Table of Contents

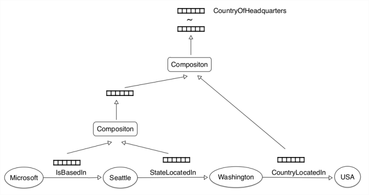
Recent Advance: Consider Long-range Interation

Previous approaches only consider triples information, it's better to take long-range information into consideration.

- Relation between Relations (Bowman et al., 2014)
- Combining Long-Range Interation (Wu et al., 2015)

Relation between Relations (Bowman et al., 2014)

- This model leaves out the representation of entities and concern the relation between relations.



- It tries to minimize the difference between the infered relation $v_r(\pi)$ (w.r.t. path π) and the single relation r.

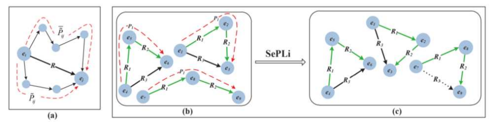
$$v_r(\pi) = RNN(\pi)$$

$$f(\pi_r, r) = sigmoid(v_r(\pi), r)$$

- And negative samples are used to maximize the margin.

Combining Long-Range Interation (Wu et al., 2015)

- This paper jointly utilizes the triple information and long-range interaction in knowledge base.



- Ranking loss for triples:

$$\min_{e_i, R_k} \sum_{t^+ \in O} \sum_{t^- \in D} (\lambda + f(e_i, r_k, e_j) - f(e_i', r_k, e_j'))$$

- And long-range loss:

$$f(\pi_r, r) = L(v_r(\pi), r)$$

Table of Contents

Prons and Cons

- Pros
  1. Hidden space model can capture the complex concepts.
  2. It is computationally efficient.
- Cons
  1. It is not logical: we can capture the co-occurrence information but not logical properties of relation, e.g. transitivity, reflexivity..
  2. It is not precise: data-driven methods will be noisy-prone.

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*, pages 2787–2795.

Bowman, S. R., Potts, C., and Manning, C. D. (2014). Recursive neural networks for learning logical semantics. *arXiv preprint arXiv:1406.1827*.

Brocheler, M., Mihalkova, L., and Getoor, L. (2012). Probabilistic similarity logic. *arXiv preprint arXiv:1203.3469*.

Lao, N., Mitchell, T., and Cohen, W. W. (2011). Random walk inference and learning in a large scale knowledge base. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 529–539. Association for Computational Linguistics.

Nickel, M., Tresp, V., and Kriegel, H.-P. (2011). A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 809–816.

Richardson, M. and Domingos, P. (2006). Markov logic networks. *Machine learning*, 62(1-2):107–136.

Socher, R., Chen, D., Manning, C. D., and Ng, A. (2013). Reasoning with neural tensor networks for knowledge base completion. In *Advances in Neural Information Processing Systems*, pages 926–934.

Wu, F., Song, J., Yang, Y., Li, X., Zhang, Z., and Zhuang, Y. (2015). Structured embedding via pairwise relations and long-range interactions in knowledge base. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.