Deep Learning Project: Combining Knowledge with Mention

Weidi Xu, Haoze Sun

Computational Intelligence Laboratory, Peking University

July 30, 2015

Overview

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

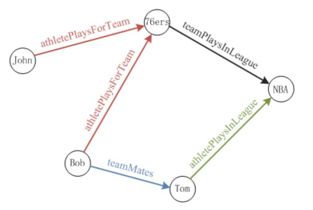
ing

Table of Contents

Table of Contents

Knowledge Graph

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

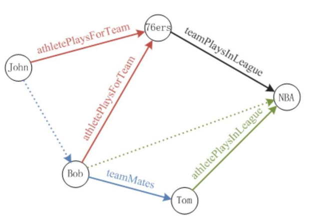


- (John, athletePlaysForTeam, 76ers)
- (Bob, athletePlaysForTeam, 76ers)
- (Tom, athleteMates, Tom)
- (Tom, athletePlaysInLeague, NBA)
- (76ers, teamPlaysInLeague, NBA)

- 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 and use the facts dataset to learn the model parameters.

Prediction

- Using the model and objective function to inference.

Framework

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

Leanring

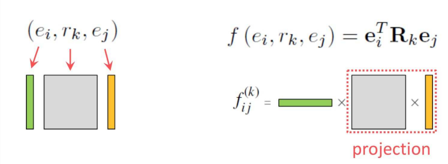
- Learning entity and relation embeddings in hidden vector spaces.
- Define a fitness funtion to determine the certainty of entity-relation-entity pairs.
- Define an objective funtion and use the facts dataset to learn the model parameters.

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:

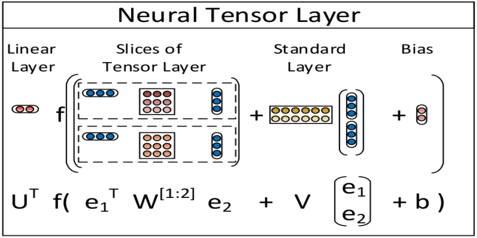
$$\min_{e_i, R_k} \sum_k \sum_i \sum_j (y_{i,j}^k - f(e_i, r_k, e_j))^2 + \lambda R$$

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:



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

- Learn based on ranking loss:

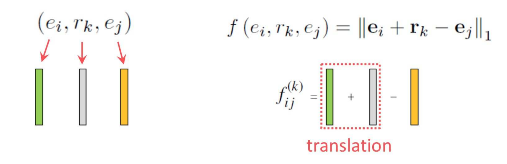
$$\min_{e_i, R_k} \sum_{t^+ \in O} \sum_{t^- \in D} (\lambda + f(e_i, r_k, e_j) - f(e_i', r_k, e_j'))$$

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:

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

assumption

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

Table of Contents

Motivation

- All pages above describes the recent representative learning in knowledge graph field. In this section we briefly introduce our work.
- In the information extraction field, one may have the information both from knowledge graph and text mention. They together can give more confidence in the relation inference.
- Our idea is to build a model which can utilize both knowledege graph and mentions jointly to determine the relation type.

Motivation

- Formally, if given the dataset not only with entity and relation tuple (e_1, r, e_2), but also the text mention m about the relation, can we adopt the information to enhance the inference precision?
- This is the problem that this work is to figure out.

Table of Contents

Data Preparation

- Luckily the NELL dataset contains the evidence when NELL system extract this tuple from the website. We can use the evidence as the relation mention in our work.
- We process the NELL dataset and extract available portion with mention as our experiment dataset. The sample in dataset has a format with (e_1, r, e_2, m).
- The prepared dataset has 34245 entities and 233 relations, each row has word sequance as relation mention.

Table of Contents

Mention Loss

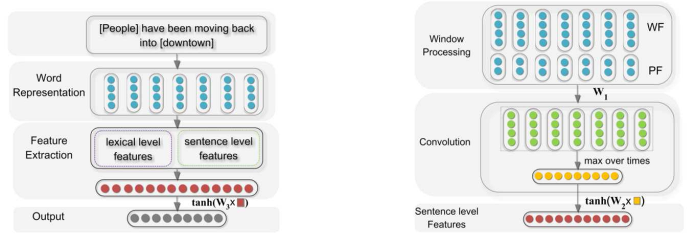
- We use TransE model as our basic model. In addition to TransE loss, we add mention loss function to be minimized. Similarly, we use the hinge loss about the mentions:

$$L_m = [D(g(m), r) + 1 - D(g(m), r')]_+$$

- where g is the function to transfer mention into a vector and $D(\dot{})$ measure the distance between $g(m)$ and r.
- We want the mention to produce the similar vector as r.

Mention Model

- In our implementation the advanced relation extraction CNN model(Zeng et al., 2014) is used.



- The framework keeps the same in his paper but we omit the lexical feature as entity mention is not included in NELL dataset and we modify the last layer to be simple fully connected layer rather than a softmax layer as it is not a classification task.

Table of Contents

Experiment and Result

- We evaluate the performance of TransE with or without mention information to see how much it can improve.
- Here we choose the mean rank measurement in testing. Mean rank measure the ranks of true relation among all relation candidates and average them.

Table : Results of experiments

Case	Train	Valid	Test
TransE	12.4	11.4	13.4
TransE-mention	3.0	2.9	3.3

References I

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.

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.

References II

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.

Zeng, D., Liu, K., Lai, S., Zhou, G., and Zhao, J. (2014). Relation classification via convolutional deep neural network. In *Proceedings of COLING*, pages 2335–2344.

Thanks

Question & Answer?