Reasoning With Neural Tensor Networks for Knowledge Base Completion

Richard Socher*, Danqi Chen*, Christopher D. Manning, Andrew Y. Ng
Computer Science Department, Stanford University
richard@socher.org, {danqi,manning}@stanford.edu

Introduction

• A common problem in knowledge representation and related fields is reasoning over a large joint knowledge graph, represented as triples of a relation between two entities.
• We introduce a model that can accurately predict additional true facts using only an existing database.
• We assess the model by considering the problem of predicting additional true relations between entities given a partial knowledge base. Our model outperforms previous models and can classify unseen relationships in WordNet and FreeBase with an accuracy of 86.2% and 90.0%, respectively.

How can we infer that Francesco Guicciardini is an Italian male person?

Neural Models for Reasoning over Relations

Overview

• Each relation is described by a neural network and pairs of entities are given as input to the model. Each entity has a vector representation, which can be constructed by its word vectors.
• The model returns a high score if they are in that relationship and a low one otherwise. This allows any fact, whether implicitly or explicitly mentioned in the database to be answered with a certainty score.

Neural Tensor Networks

The Neural Tensor Network (NTN) replaces a standard linear neural network layer with a bilinear tensor layer that directly relates the two entity vectors across multiple dimensions. The model computes a score of how likely it is that two entities are in a certain relationship by the following NTN-based function:

\[ g(c_1, R, c_2) = u_h^T f \left( e_1^T W R_{R}^T v_2 + V_0 \left[ e_1^T + b_2 \right] + b_1 \right), \]

where \( f = \tanh \) is a standard nonlinearity applied element-wise, \( W_{R} \in \mathbb{R}^{d \times d \times k} \) is a tensor and the bilinear tensor product \( e_1^T W_R^T e_2 \) results in a vector \( h \in \mathbb{R}^k \). The other parameters for relation \( R \) are the standard form of a neural network: \( V_0 \in \mathbb{R}^{1 \times k} \) and \( U \in \mathbb{R}^k \), \( b_1 \in \mathbb{R}^k \).

Training objective: \( T_{ij} = \langle e_i^T, R_{ij}, e_j \rangle \) is a triplet with a random entity corrupted from a correct triplet \( T_{ij} = \langle e_i^T, R_{ij}, e_j \rangle \).

\[ J(\Theta) = \sum_{c_1, R, c_2} \max \left( 0, 1 - g \left( T(c_1, R, c_2) \right) + g \left( T_{ij} \right) \right) + \lambda \| \Theta \|_{2}^{2}. \]

We use minibatched L-BFGS for training.

Entity Representations Revisited

We propose two further improvements:

• We represent each entity as the average of its word vectors, allowing the sharing of statistical strength between the words describing each entity.
• We can initialize the word vectors with pre-trained unsupervised word vectors, which in general capture some distributional syntactic and semantic information.

Experiments

Datasets

<table>
<thead>
<tr>
<th>Dataset</th>
<th>#R</th>
<th>#Ent.</th>
<th>#Train</th>
<th>#Dev</th>
<th>#Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wordnet</td>
<td>11,38,696</td>
<td>112,581</td>
<td>2,609</td>
<td>10,544</td>
<td></td>
</tr>
<tr>
<td>Freebase</td>
<td>13</td>
<td>75,043</td>
<td>316,232</td>
<td>5,908</td>
<td>23,733</td>
</tr>
</tbody>
</table>

Relation Triplets Classification

We randomly switch entities from correct testing triplets resulting in an equal number of positive and negative examples. We predict the relation \( (c_1, R, c_2) \) holds if \( g(c_1, R, c_2) \geq T_{ij} \) (we use the development set to find \( T_{ij} \)).

Comparison of accuracy of different relations:

Model | WordNet | Freebase
--- | --- | ---
Distance Model | 68.3 | 61.0
Hadamard Model | 80.0 | 68.8
Single Layer Model | 76.0 | 85.3
Bilinear Model | 84.1 | 87.7
Neural Tensor Network | 86.2 | 90.0

Conclusion

We introduced Neural Tensor Networks. Unlike previous models for predicting relationships using entities in knowledge bases, our model allows a direct interaction of entity vectors via a tensor. The model obtains the highest accuracy in terms of predicting unseen relationships between entities through reasoning inside a given knowledge base. We further show that by representing entities through their constituent words and initializing these word representations using unsupervised large corpora, performance of all models improves substantially.

References