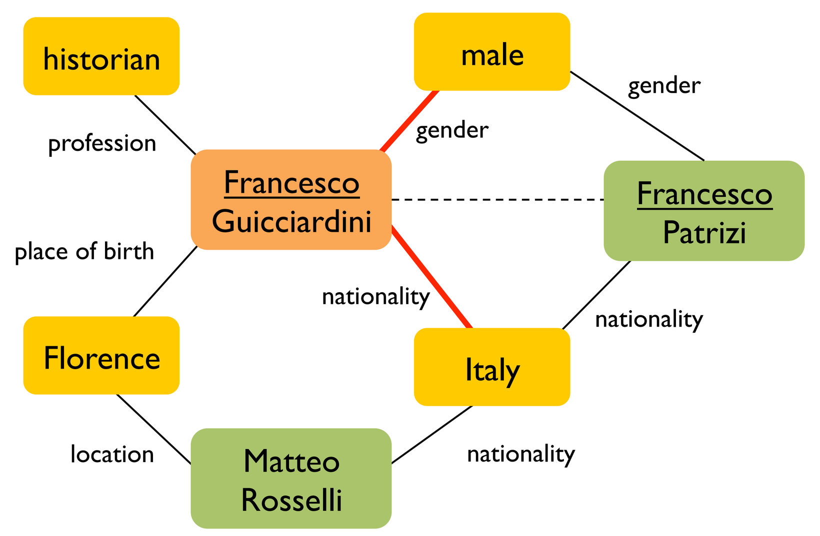
Reasoning With Neural Tensor Networks for Knowledge Base Completion

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

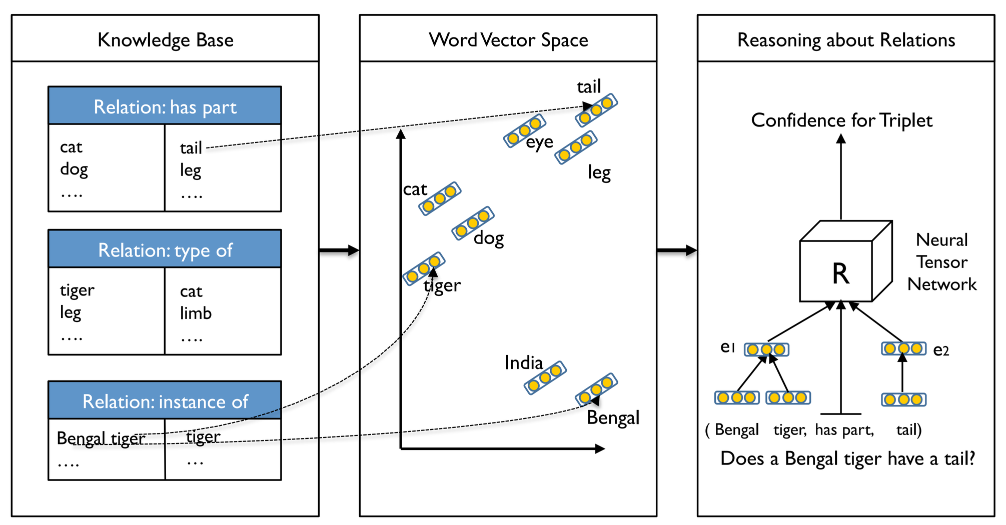


How can we infer that `Francesco Guicciardini` is an Italian male person?

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.

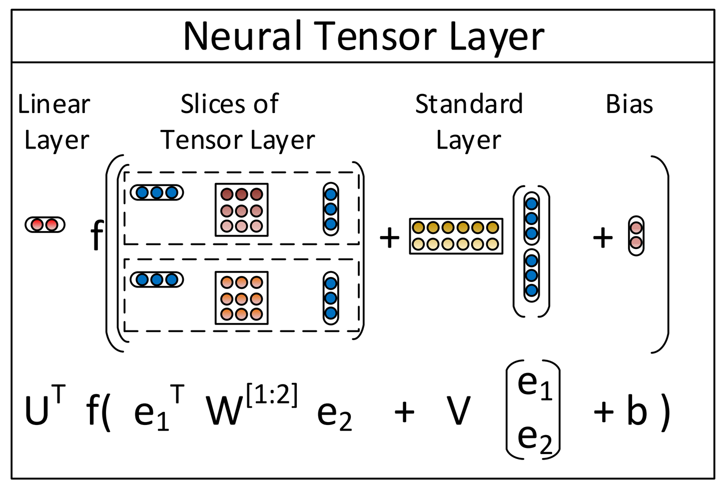


Neural Tensor Networks

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:

$$g(e_1, R, e_2) = u_R^T f\left(e_1^T W_R^{[1:k]} e_2 + V_R \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} + b_R\right),$$

where $f = \tanh$ is a standard nonlinearity applied element-wise, $W_R^{[1:k]} \in \mathbb{R}^{d\times d\times k}$ is a tensor and the bilinear tensor product $e_1^T W_R^{[1:k]} e_2$ results in a vector $h \in \mathbb{R}^k$. The other parameters for relation R are the standard form of a neural network: $V_R \in \mathbb{R}^{k\times 2d}$ and $U \in \mathbb{R}^k, b_R \in \mathbb{R}^k$.



Training objective: $T_c^{(i)} = (e_1^{(i)}, R^{(i)}, e_c)$ is a triplet with a random entity corrupted from a correct triplet $T^{(i)} = (e_1^{(i)}, R^{(i)}, e_2^{(i)})$,

$$J(\Omega) = \sum_{i=1}^{N}\sum_{c=1}^{C} \max\left(0, 1 - g\left(T^{(i)}\right) + g\left(T_c^{(i)}\right)\right) + \lambda\|\Omega\|_2^2,$$

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

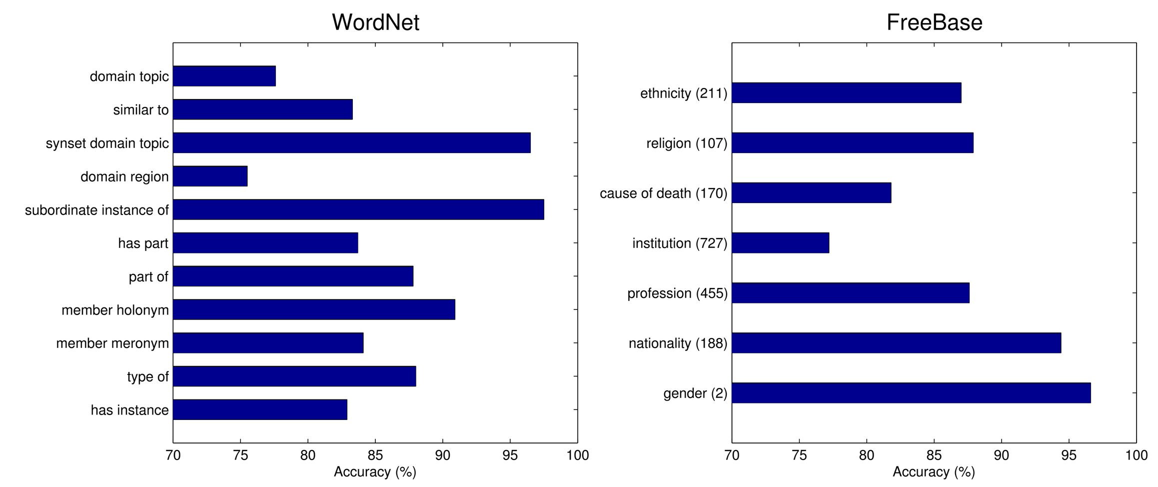
Dataset	#R.	# Ent.	# Train	# Dev	# Test
Wordnet	11	38,696	112,581	2,609	10,544
Freebase	13	75,043	316,232	5,908	23,733

Relation Triplets Classification

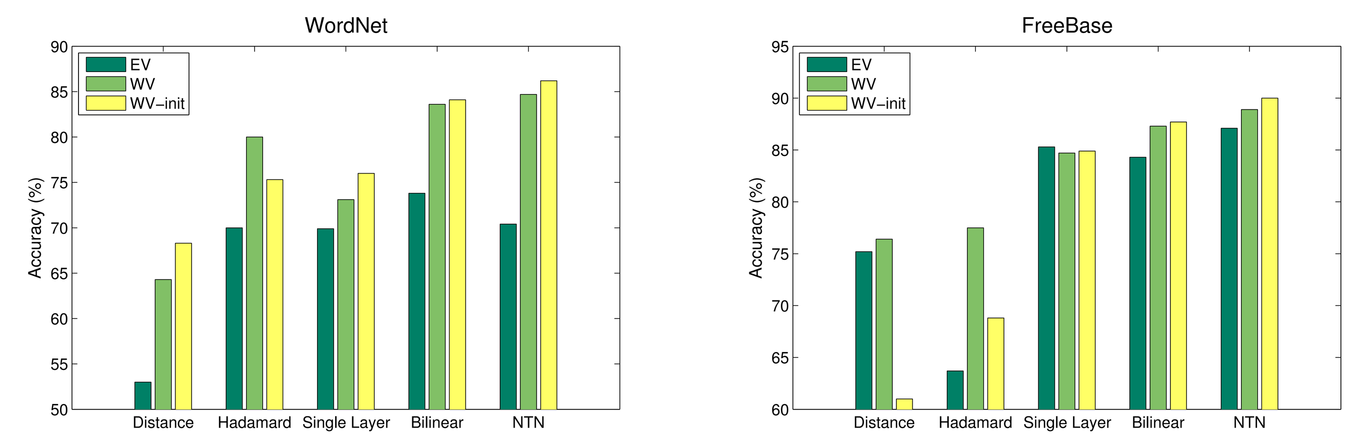
We randomly switch entities from correct testing triplets resulting in an equal number of positive and negative examples. We predict the relation (e_1, R, e_2) holds if $g(e_1, R, e_2) \geq T_R$ (we use the development set to find T_R).

Model	WordNet	Freebase
Distance Model [1]	68.3	61.0
Hadamard Model [2]	80.0	68.8
Single Layer Model	76.0	85.3
Bilinear Model [3]	84.1	87.7
Neural Tensor Network	**86.2**	**90.0**

Comparison of accuracy of different relations:



The influence of entity representations (**EV**: training on entity vectors. **WV**: training on randomly initialized word vectors. **WV-init**: training on word vectors initialized with unsupervised semantic word vectors [4]):



Examples of of relationship predictions by our Neural Tensor Network on WordNet:

Entity e_1	Relationship R	Sorted list of entities likely to be in this relationship
tube	type of	structure; anatomical structure; device; body; body part; organ
creator	type of	individual; adult; worker; man; communicator; instrumentalist
dubrovnik	subordinate instance of	city; town; city district; port; river; region; island
armed forces	domain region	military operation; naval forces; military officier; military court
boldness	has instance	audaciousness; aggro; abductor; interloper; confession;
hispid	similar to	rough; haired; divided; hard; free; chromatic; covered;
people	type of	group; agency; social group; organisation; alphabet; race

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

[1] A. Bordes, J. Weston, R. Collobert, and Y. Bengio. Learning structured embeddings of knowledge bases. In AAAI, 2011.

[2] A. Bordes, X. Glorot, J. Weston, and Y. Bengio. Joint Learning of Words and Meaning Representations for Open-Text Semantic Parsing. AISTATS, 2012.

[3] R. Jenatton, N. Le Roux, A. Bordes, and G. Obozinski. A latent factor model for highly multi-relational data. In NIPS, 2012.

[4] J. Turian, L. Ratinov, and Y. Bengio. Word representations: a simple and general method for semi-supervised learning. In ACL, 2010.