# Translating Embeddings for Modeling Multi-relational Data

Antoine Bordes,* Nicolas Usunier,* Alberto Garciá-Duran,* Jason Weston° & Oksana Yakhnenko°

*Heudiasyc, UMR CNRS 7253, Université de Technologie de Compiègne, France

°Google, 111 8th Avenue, New York, NY. USA

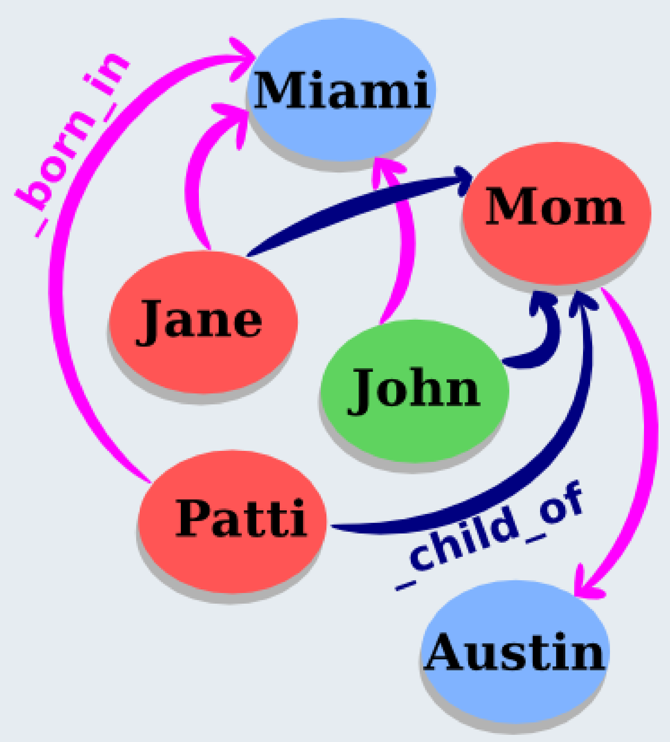
## One Minute Overview

- Knowledge Bases (KBs) are massive amounts of structured data.
- **Main issue:** KBs are hard to manipulate.
  - Very large dimensions: $10^5 - 10^8$ entities, $10^4 - 10^6$ relationships;
  - Noisy/incomplete: missing/wrong relations/entities.
- **Here:** Encode KBs in vector spaces, in which rel. are translations.
  - Simple model with few parameters designed to encode similarities;
  - Easy to train on large-scale databases;
  - Strong results on real-world data.

## Muti-relational Data

**Knowledge Bases:**

- Each node = an entity.
- Each edge = a relation.
- A relation = $(h, r, t)$:
  - $h$ = *head* (or *subject*),
  - $r$ = *relationship*,
  - $t$ = *tail* (or *object*).
- Nodes w/o features.



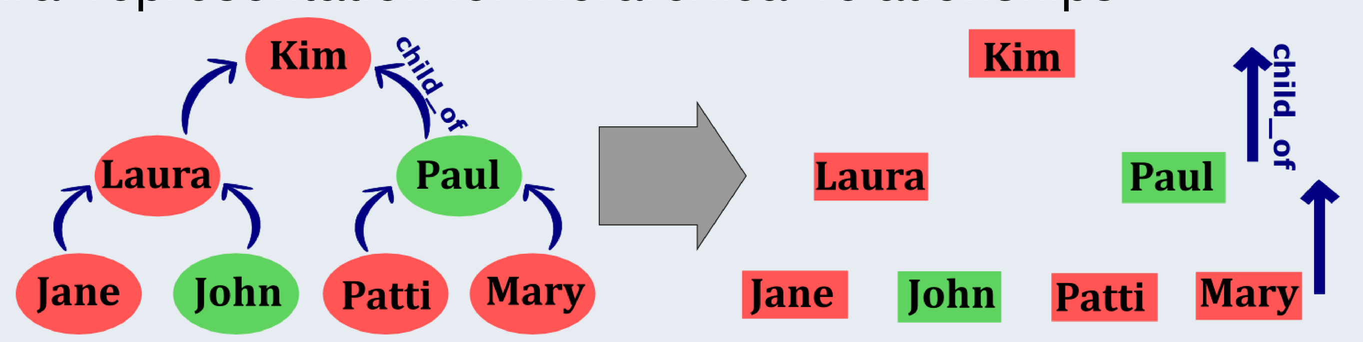

**Examples:**

Freebase, YAGO, IMDB, GeneOntology, UniprotKB, WordNet, etc.

## Embedding-based Framework

**This work: relationships = translations on entity embeddings.**

- Natural representation for hierarchical relationships.




- Recent work on word embeddings (Mikolov et al. 13): there may exist embedding spaces in which relationships are represented by translations.
- Few parameters to encode each relationship.

## Related Work

- Tensor factorization (e.g. (Harshman et al., 94)).
- Explicit modeling of missing data (e.g. (Gao et al., 11))
- Markov-logic Networks (e.g. (Kok et al., 07))
- Extension of SBMs (e.g. (Kemp et al., 06; Sutskever et al., 10)).
- Spectral clustering for undirected graphs (e.g. (Dong et al., 11)).
- Collective matrix factorization (e.g. (Nickel et al., 11)).
- Energy-based learning (e.g. (Bordes et al., 11,13), (Socher et al. 13)).

## Translating Embeddings – TransE

**Learning Representations:**

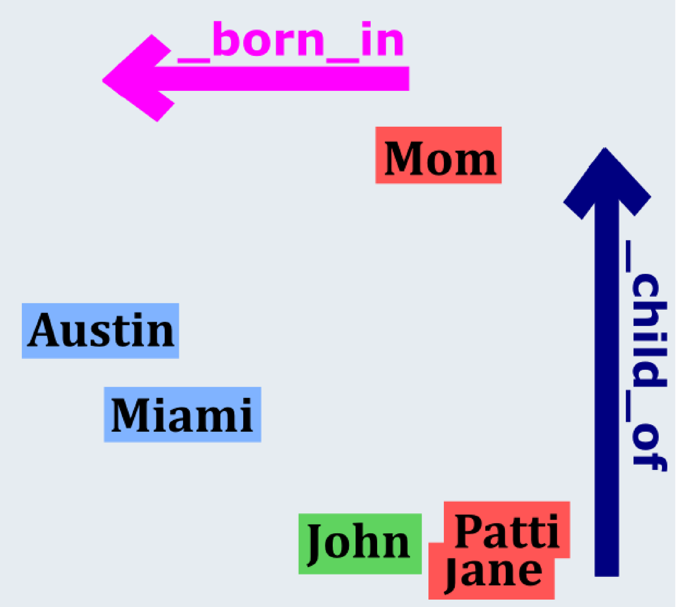
- Entities are represented by embeddings in $\mathbb{R}^k$.
- Relationships = similarity operators between heads/tails.
- We learn $d(h, r, t)$ = dissimilarity measure depending on $r$.

**Relationships as Translations:**

- We would like that $\boldsymbol{h} + \boldsymbol{r} \approx \boldsymbol{t}$.
- We define the dissimilarity measure:

$$d(h, r, t) = ||\boldsymbol{h} + \boldsymbol{r} - \boldsymbol{t}||_2^2$$

Note: $d(.)$ can also use $L_1$ distance instead.




- We learn $\boldsymbol{h}$,$\boldsymbol{r}$ and $\boldsymbol{t}$ using SGD to minimize the ranking loss:

$$\mathcal{L} = \sum_{(h,r,t) \in S} \sum_{(h',r,t') \in S'_{(h,r,t)}} \left[\gamma + d(h, r, t) - d(h', r, t')\right]_+$$

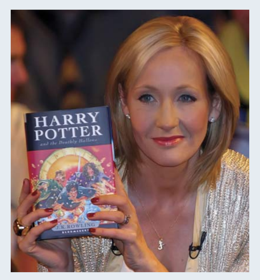
**Training Algorithm:**

**input** Training set $S = \{(h, r, t)\}$, entities and rel. sets $N$ and $L$, margin $\gamma$, embeddings dim. $k$.

**initialize** $\boldsymbol{r} \leftarrow$ uniform$(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $r \in L$
$\boldsymbol{r} \leftarrow \boldsymbol{r} / \|\boldsymbol{r}\|$ for each $r \in L$
$\mathbf{e} \leftarrow$ uniform$(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in N$
**loop**
  $\mathbf{e} \leftarrow \mathbf{e} / \|\mathbf{e}\|$ for each entity $e \in N$
  $S_{batch} \leftarrow$ sample$(S, b)$ // sample a minibatch of size $b$
  $T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets
  **for** $(h, r, t) \in S_{batch}$ **do**
    $(h', r, t') \leftarrow$ sample$(S'_{(h,r,t)})$ // sample a corrupted triplet
    $T_{batch} \leftarrow T_{batch} \cup \{((h, r, t), (h', r, t'))\}$
  **end for**
  Update embeddings w.r.t. $\sum_{((h,r,t),(h',r,t')) \in T_{batch}} \nabla \left[\gamma + d(h, r, t) - d(h', r, t')\right]_+$
**end loop**

## Example Predictions

J. K. Rowling _influenced_by G. K. Chesterton, J. R. R. Tolkien, C. S. Lewis, Lloyd Alexander, Terry Pratchett, Roald Dahl, Jorge Luis Borges, Stephen King



WALL-E _has_genre Animation, Computer animation, Comedy film, Adventure film, Science Fiction, Fantasy, Stop motion, Satire

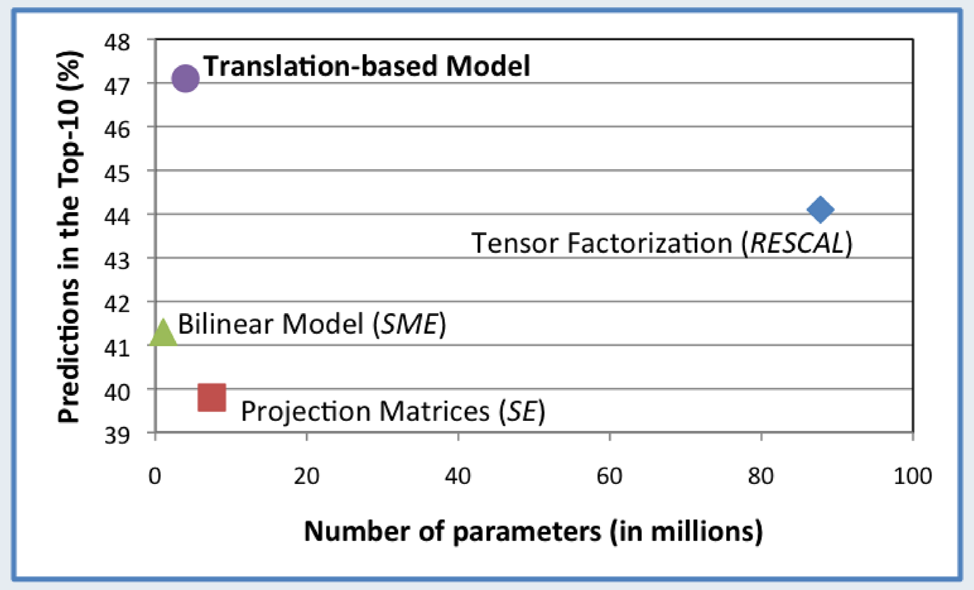

## Experiments

**Data:**

| | Entities | Relationships | Train. Ex. | Valid. Ex. | Test Ex. |
|---|---|---|---|---|---|
| Freebase15k | 14,951 | 1,345 | 483,142 | 50,000 | 59,071 |
| Freebase1M | $1 \times 10^6$ | 23,382 | $17.5 \times 10^6$ | 50,000 | 177,404 |

**Link prediction:** in a ranking evaluation setting.

On Freebase15k:

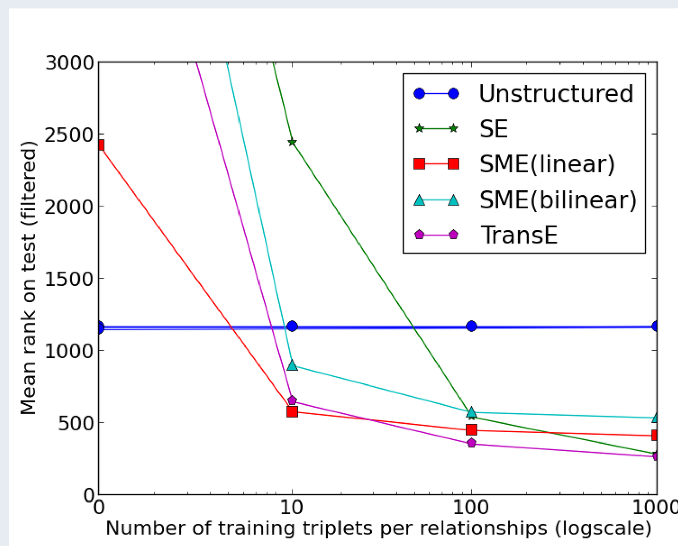


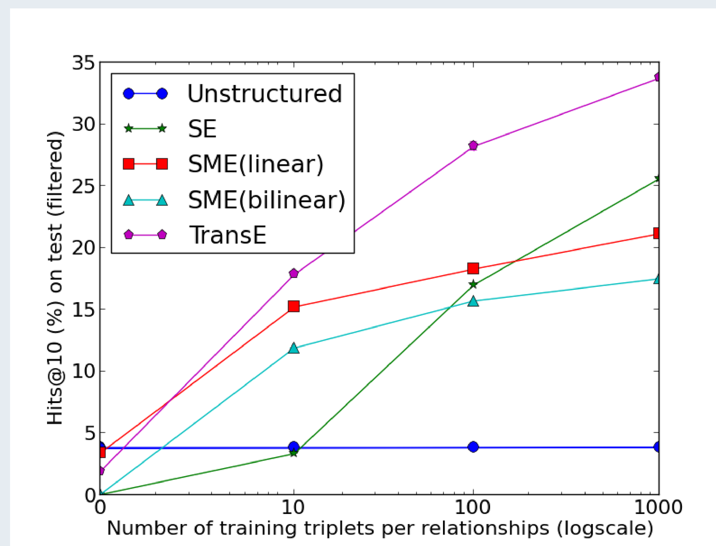

On Freebase1M, TransE predicts 34% in the Top-10.

**Detailed results by category of relationship:**

| Task | Predicting *head* | | | | Predicting *tail* | | | |
|---|---|---|---|---|---|---|---|---|
| Rel. category | 1-to-1 | 1-to-M. | M.-to-1 | M.-to-M. | 1-to-1 | 1-to-M. | M.-to-1 | M.-to-M. |
| Unstructured | 34.5 | 2.5 | 6.1 | 6.6 | 34.3 | 4.2 | 1.9 | 6.6 |
| SE (Bordes et al., 11) | 35.6 | 62.6 | 17.2 | 37.5 | 34.9 | 14.6 | 68.3 | 41.3 |
| SME(linear) (Bordes et al., 13) | 35.1 | 53.7 | 19.0 | 40.3 | 32.7 | 14.9 | 61.6 | 43.3 |
| SME(bilinear) (Bordes et al., 13) | 30.9 | **69.6** | **19.9** | 38.6 | 28.2 | 13.1 | **76.0** | 41.8 |
| TransE | **43.7** | 65.7 | 18.2 | **47.2** | **43.7** | **19.7** | 66.7 | **50.0** |

**Learning new relationships with few examples:**




Performance for learning 40 new relationships.

## Code+Data

Related material is available from `http://goo.gl/0PpKQe`.

## Acknowledgements

This work was carried out within the Labex MS2T (ANR-11-IDEX-0004-02), and funded by the French Nat. Agency for Research (EVEREST-12-JS02-005-01).