# Knowledge Graph Embedding
# A Geometrical Perspective

Han Xiao, **Minlie Huang**, Xiaoyan Zhu

**aihuang@tsinghua.edu.cn**
Tsinghua University

June 12, 2016

# Knowledge Graph Embedding: A Geometrical Perspective

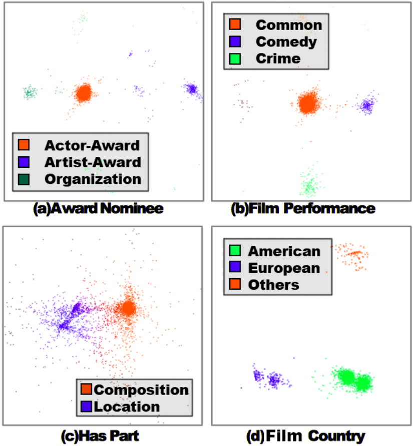
1. From the geometric perspective of **DATA**, we propose *TransG* to model multiple relation semantics.
2. From the geometric perspective of **MODEL**, we propose *ManifoldE* to achieve an algebraic well-posed system and a flexible geometric form.

# Knowledge Graph Embedding: Related Work

- **Translation-based.** TransE (Bordes et al. 2013), TransH (Wang et al. 2014), TransR (cTransR) (Lin et al. 2015a), KG2E (He et al. 2015), PTransE (Lin et al. 2015b), TransD, TransM (Fan et al. 2014), ...
  - **ManifoldE**: manifold-based (Xiao et al. 2016)
- **Structured and Unstructured Embedding.** (Bordes et al. 2011,2012,2014)
- **Neural Network bsed Embedding.** Single Layer Model and Neural Tensor Network (Socher et al. 2013)
- **Factor Models**. Latent factor models (Jenatton et al. 2012) and matrix factorization (Nickel et al. 2011/2012).

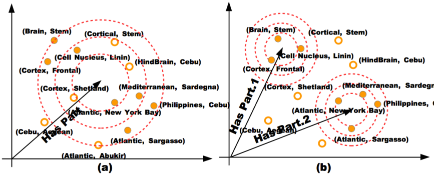
# TransG: A Generative Embedding Model

- **An Interesting Observation.** There are many $< entity, entity >$ clusters in a relation-specific embedding space, representing various semantics, called **Multiple Relation Semantics**.



# TransG: A Generative Embedding Model

- **Multiple Relation Semantics**: A relation may have multiple meanings revealed by the entity pairs associated with the corresponding triples.
  - Composition Related: *(Table, HasPart, Leg)*
  - Location Related: *(Atlantics, HasPart, NewYorkBay)*
- **Reasons.**
  - Artificial Simplification.
  - Nature Of Knowledge.
- **Motivation.** Traditional methods such as TransE, could hardly model this phenomenon, incurring much noise.



# TransG: A Generative Embedding Model

- **Methodology.** TransG leverages a mixture of semantic component vectors for a specific relation. Each component represents a specific latent meaning. By this way, TransG could distinguish multiple relation semantics.
- **Generative Process.**
  1. For an entity $e \in E$:
     1.1 Draw each entity embedding mean vector from a standard normal distribution as a prior: $\mathbf{u_e} \backsim \mathcal{N}(\mathbf{0}, \mathbf{1})$.
  2. For a triple $(h, r, t) \in \Delta$:
     2.1 Draw a semantic component from Chinese Restaurant Process for this relation: $\pi_{r,m} \sim CRP(\beta)$.
     2.2 Draw a head entity embedding vector from a normal distribution: $\mathbf{h} \backsim \mathcal{N}(\mathbf{u_h}, \sigma_h^2 \mathbf{E})$.
     2.3 Draw a tail entity embedding vector from a normal distribution: $\mathbf{t} \backsim \mathcal{N}(\mathbf{u_t}, \sigma_t^2 \mathbf{E})$.
     2.4 Draw a relation embedding vector for this semantics: $\mathbf{u_{r,m}} = \mathbf{t} - \mathbf{h} \backsim \mathcal{N}(\mathbf{u_t} - \mathbf{u_h}, (\sigma_h^2 + \sigma_t^2)\mathbf{E})$.

# TransG: A Generative Embedding Model

- **Score Function**: the probability for generating the triple.

$$\mathbb{P}\{(h,r,t)\} \quad \propto \sum_{m=1}^{M_r} \pi_{r,m} \mathbb{P}(\mathbf{u_{r,m}}|h,t) = \sum_{m=1}^{M_r} \pi_{r,m} e^{-\frac{||\mathbf{u_h}+\mathbf{u_{r,m}}-\mathbf{u_t}||_2^2}{\sigma_h^2+\sigma_t^2}}$$

- **Geometrical Perspective.**
  - TransG generalizes this geometric principle of translation from $\mathbf{h}+\mathbf{r}=\mathbf{t}$ to the selective translation:

$$m^*_{(h,r,t)} = \arg\max_{m=1\ldots M_r} \quad \left(\pi_{r,m} e^{-\frac{||\mathbf{u_h}+\mathbf{u_{r,m}}-\mathbf{u_t}||_2^2}{\sigma_h^2+\sigma_t^2}}\right)$$

$$\mathbf{h} + \mathbf{u}_{\mathbf{r},\mathbf{m}^*_{(\mathbf{h},\mathbf{r},\mathbf{t})}} \approx \mathbf{t}$$

  - Previous studies make translation identically for all the triples of the same relation, but TransG automatically selects the best translation vector according to the specific semantics of a triple.

# TransG: A Generative Embedding Model

- **Training Procedure.**
  - **CRP Part.**

$$\mathbb{P}(m_{r,new}) = \frac{\beta e^{-\frac{||\mathbf{h}-\mathbf{t}||_2^2}{\sigma_h^2+\sigma_t^2+2}}}{\beta e^{-\frac{||\mathbf{h}-\mathbf{t}||_2^2}{\sigma_h^2+\sigma_t^2+2}} + \mathbb{P}\{(h,r,t)\}}$$

  - **Other Parts.**

$$\begin{aligned}\min \quad & -\sum_{(h,r,t)\in\Delta} ln\left(\sum_{m=1}^{M_r} \pi_{r,m} e^{-\frac{||\mathbf{u_h}+\mathbf{u_{r,m}}-\mathbf{u_t}||_2^2}{\sigma_h^2+\sigma_t^2}}\right) + \\ & \sum_{(h',r',t')\in\Delta'} ln\left(\sum_{m=1}^{M_r} \pi_{r',m} e^{-\frac{||\mathbf{u_{h'}}+\mathbf{u_{r',m}}-\mathbf{u_{t'}}||_2^2}{\sigma_{h'}^2+\sigma_{t'}^2}}\right) \\ & + C\left(\sum_{r\in R}\sum_{m=1}^{M_r} ||\mathbf{u_{r,m}}||_2^2 + \sum_{e\in E} ||\mathbf{u_e}||_2^2\right)\end{aligned}$$

# TransG: A Generative Embedding Model

- **Experiments**: Link Prediction.

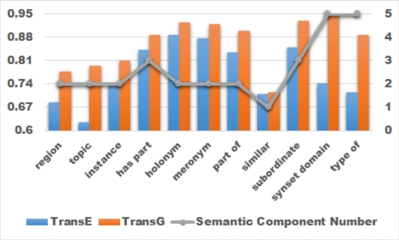
| Datasets | WN18 | | | | FB15K | | | |
|---|---|---|---|---|---|---|---|---|
| Metric | Mean Rank | | HITS@10(%) | | Mean Rank | | HITS@10(%) | |
| | Raw | Filter | Raw | Filter | Raw | Filter | Raw | Filter |
| TransE | 263 | 251 | 75.4 | 89.2 | 243 | 125 | 34.9 | 47.1 |
| TransH | 401 | 388 | 73.0 | 82.3 | 212 | 87 | 45.7 | 64.4 |
| TransR | 238 | 225 | 79.8 | 92.0 | 198 | 77 | 48.2 | 68.7 |
| CTransR | **231** | **218** | 79.4 | 92.3 | 199 | 75 | 48.4 | 70.2 |
| PTransE | N/A | N/A | N/A | N/A | 207 | 58 | 51.4 | 84.6 |
| KG2E | 362 | 348 | 80.5 | 93.2 | 183 | 69 | 47.5 | 71.5 |
| TransG | 357 | 345 | **84.5** | **94.9** | **152** | **50** | **55.9** | **88.2** |

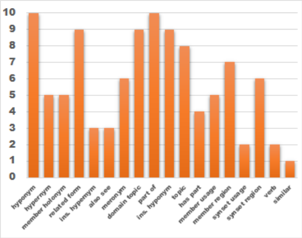
- **Experiments**: Triple Classification.

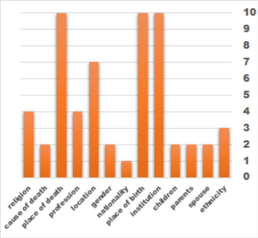
| Methods | WN11 | FB13 | AVG. |
|---|---|---|---|
| NTN | 70.4 | 87.1 | 78.8 |
| TransE | 75.9 | 81.5 | 78.7 |
| TransH | 78.8 | 83.3 | 81.1 |
| TransR | 85.9 | 82.5 | 84.2 |
| CTransR | 85.7 | N/A | N/A |
| KG2E | 85.4 | 85.3 | 85.4 |
| TransG | **87.4** | **87.3** | **87.4** |

# TransG: A Generative Embedding Model

- **Experiments**: Semantic Component Analysis.



# TransG: A Generative Embedding Model

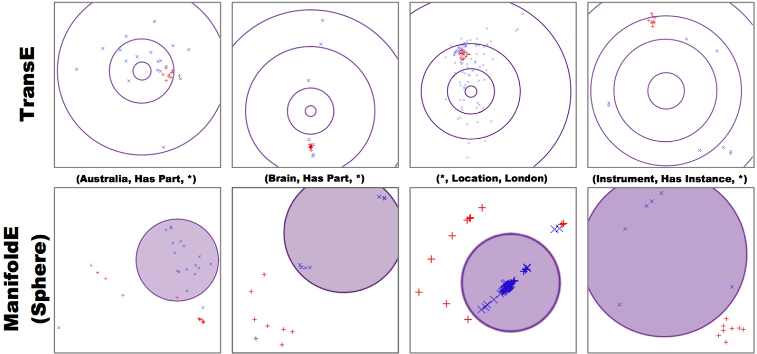
- In this paper, we propose a generative Bayesian non-parametric infinite mixture embedding model, TransG, to address a new issue, multiple relation semantics, which can be commonly seen in knowledge graph.
- TransG can discover the latent semantics of a relation automatically and leverage a mixture of relation components for embedding.
- Extensive experiments show our method achieves substantial improvements against the state-of-the-art baselines.

# Summaries

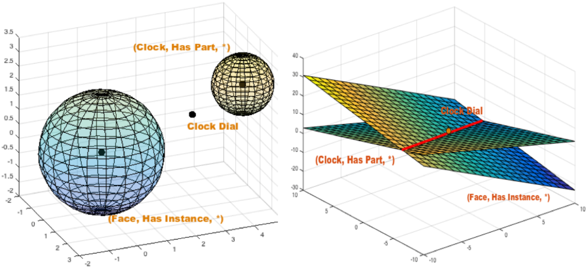
1. From the geometric perspective of **DATA**, we propose *TransG* to model multiple relation semantics.
2. From the geometric perspective of **MODEL**, we propose *ManifoldE* to achieve an algebraic well-posed system and a flexible geometric form.

# ManifoldE: From A Point To A Manifold

- **Precise Link Prediction** attempts to find the exact entity given another entity and the relation.
- **Motivations.**
  - Being *ill-posed algebraic system*.
    - There are $Td$ equations $(h_i + r_i = t_i)$.
    - There are $(E + R)d$ variables.
    - Since $T \gg E + R$, it is *an ill-posed algebraic system*.
  - Adopting *over-strict geometric form*.



# ManifoldE: From A Point To A Manifold




**(a)** **(b)**

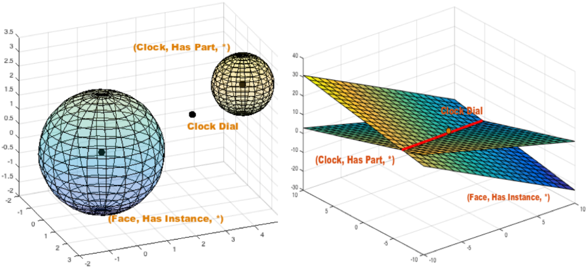
- **Methodology.** To apply the manifold-based principle:

$$\mathcal{M}(\mathbf{h}, \mathbf{r}, \mathbf{t}) = D_r^2$$

When a head entity and a relation are given, the tail entities lay in a high-dimensional manifold.

$$f_r(h, t) = ||\mathcal{M}(h, r, t) - D_r^2||^2$$

# ManifoldE: From A Point To A Manifold




(a) (b)

- **Sphere.**

$$\mathcal{M}(h, r, t) = ||\mathbf{h} + \mathbf{r} - \mathbf{t}||_2^2$$

- **Hyperplane.**

$$\mathcal{M}(h, r, t) = (\mathbf{h} + \mathbf{r}_{\text{head}})^\top (\mathbf{t} + \mathbf{r}_{\text{tail}})$$

# ManifoldE: From A Point To A Manifold

- **Geometric Perspective.**
  - Manifold-Based principle extends one point to a whole manifold, to strengthen the stability.
  - This way would benefit complex relations.
- **Algebraic Perspective.**
  - There are one equation for one triple.
  - Thus, if $d \geq \frac{T}{E+R}$, the system is far away from ill-posed.
- **Training.**

$$\mathcal{L} = \sum_{(h,r,t)\in\Delta} \sum_{(h',r',t')\in\Delta'} [f_{r'}(h',t') - f_r(h,t) + \gamma]_+$$

- **Experiments**: Link Prediction.

| Datasets | WN18 | | | |
|---|---|---|---|---|
| Metric | HITS@10(%) | | HITS@1(%) | Time(s) |
| | Raw | Filter | Filter | One Epos |
| TransE | 75.4 | 89.2 | 29.5 | **0.4** |
| TransH | 73.0 | 82.3 | 31.3 | 1.4 |
| TransR | 79.8 | 92.0 | 33.5 | 9.8 |
| PTransE | - | - | - | - |
| KG2E | 80.2 | 92.8 | 54.1 | 10.7 |
| ManifoldE S. | 80.7 | 92.8 | 55.8 | **0.4** |
| ManifoldE H. | **84.2** | **94.9** | **93.2** | 0.5 |

# ManifoldE: From A Point To A Manifold

- **Experiments**: Link Prediction.

| Datasets | FB15K | | | |
|---|---|---|---|---|
| Metric | HITS@10(%) | | HITS@1(%) | Time(s) |
| | Raw | Filter | Filter | One Epos |
| TransE | 34.9 | 47.1 | 29.4 | **0.7** |
| TransH | 48.2 | 64.4 | 24.8 | 4.8 |
| TransR | 48.4 | 68.7 | 20.0 | 29.1 |
| PTransE | 51.4 | 84.6 | 63.3 | 266.0 |
| KG2E | 48.9 | 74.0 | 40.4 | 44.2 |
| ManifoldE S. | **55.7** | 86.2 | 64.1 | **0.7** |
| ManifoldE H. | 55.2 | **88.1** | **70.5** | 0.8 |

# ManifoldE: From A Point To A Manifold

- **Experiments**: Triple Classification.

| Methods | WN11 | FB13 | AVG. |
|---|---|---|---|
| SE | 53.0 | 75.2 | 64.1 |
| NTN | 70.4 | 87.1 | 78.8 |
| TransE | 75.9 | 81.5 | 78.7 |
| TransH | 78.8 | 83.3 | 81.1 |
| TransR | 85.9 | 82.5 | 84.2 |
| KG2E | 85.4 | 85.3 | 85.4 |
| ManifoldE Sphere | **87.5** | 87.2 | **87.4** |
| ManifoldE Hyperplane | 86.9 | **87.3** | 87.1 |

# ManifoldE: From A Point To A Manifold

- In this paper, we study the precise link prediction problem and attribute two reasons to the problem: ill-posed algebraic system and over-restricted geometric form.
- To alleviate these issues, we propose a novel manifold-based principle and a few models ManifoldE (Sphere/Hyperplane) inspired by the principle. From algebraic perspective, ManifoldE is a nearly well-posed equation system and from a geometric perspective, it expands one point in translation-based principle to a manifold.
- Extensive experiments show our method achieves substantial improvements against the state-of-the-art baselines.

## Summaries

1. From the geometric perspective of **DATA**, we propose *TransG* to model multiple relation semantics.
2. From the geometric perspective of **MODEL**, we propose *ManifoldE* to achieve an algebraic well-posed system and a flexible geometric form.

# Q & A

For more details, please refer to our papers:

- From One Point to A Manifold: Knowledge Graph Embedding For Precise Link Prediction. IJCAI 2016, New York, USA.
- TransG : A Generative Model for Knowledge Graph Embedding. ACL 2016, Berlin, Germany.

**Thanks.**