# Logic Attention Based Neighborhood Aggregation for Inductive Knowledge Graph Embedding

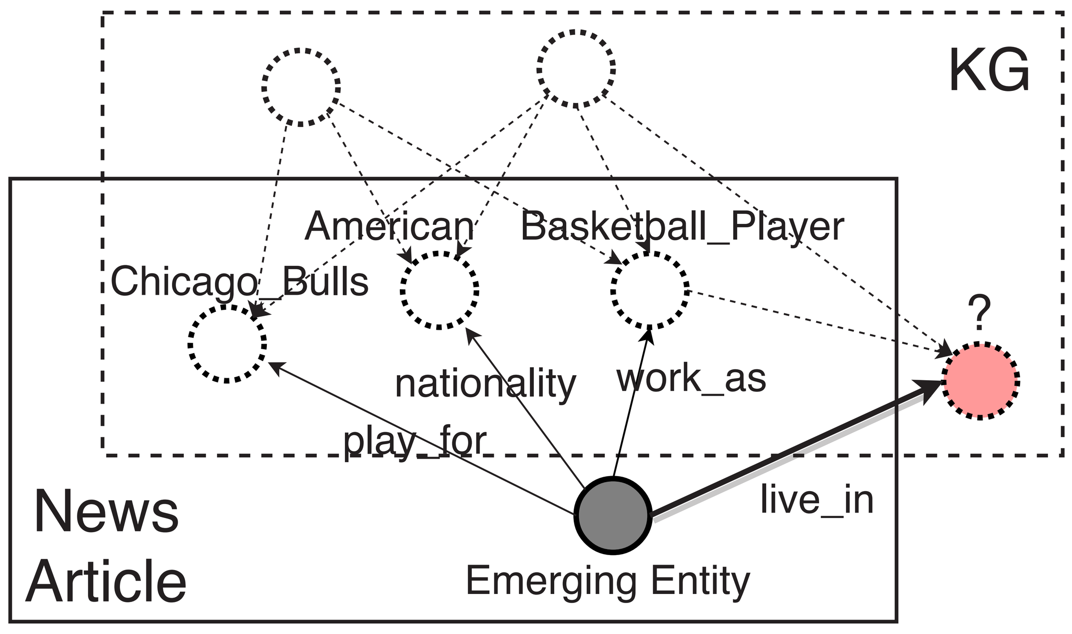
Peifeng Wang[1*], Jialong Han[2], Chenliang Li[3], Rong Pan[1]
[1] *Sun Yat-sen University*, [2] *Tencent AI Lab*, [3] *Wuhan University*
[1] *{wangpf3@mail2., panr@}sysu.edu.cn*, [2] *jialonghan@gmail.com*, [3] *cllee@whu.edu.cn*
[*] *This work was done during the first author's internship at Tencent AI Lab.*

## Motivation and Contributions

The paper aims at embedding new entities from the knowledge graph (KG) inductively by aggregating the neighbors with respect to their unordered and unequal nature.

- We propose three desired properties that decent neighborhood aggregators for KGs should possess.
- We propose a novel aggregator called Logic Attention Network (LAN), to facilitate inductive KG embedding.
- We conduct extensive experiments to validate the superiority of LAN.

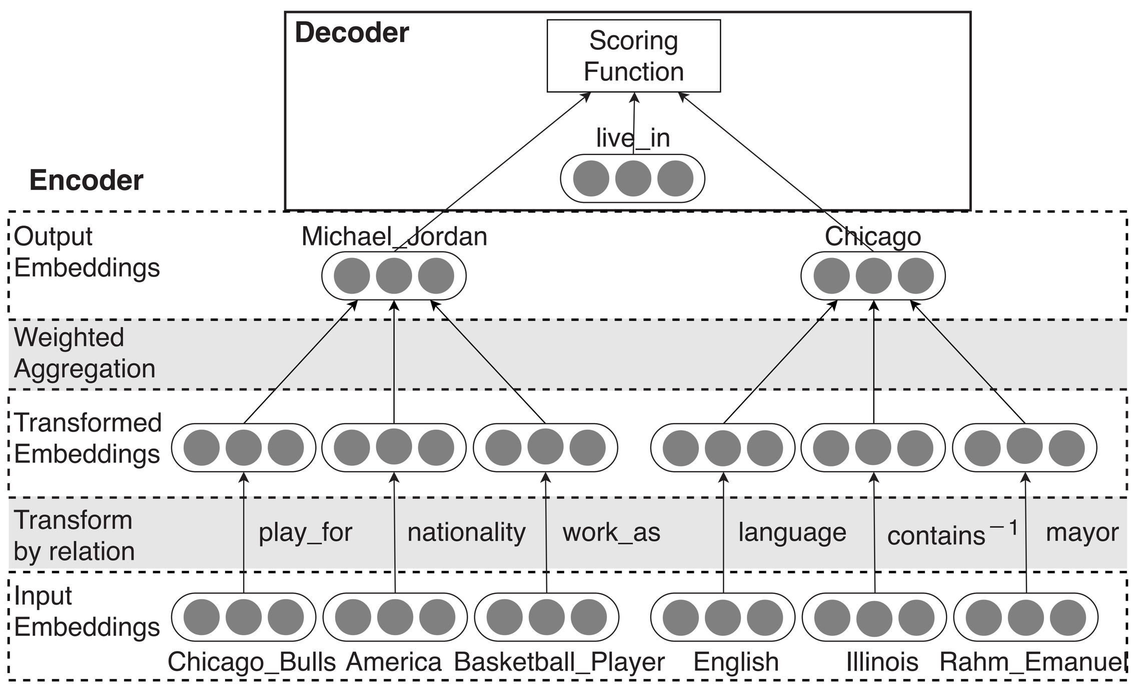
## Background and Problem Setting




- **KG** consists of numerous triplet facts like (s, r, o).
- **KG embedding** represents entities and relations in low-dimension space.
- **Challenge** All the entities are required to be seen during training.

Given a knowledge graph $\mathcal{K}$, we would like to learn a neighborhood aggregator $A$ which uses an entity $e_i$'s neighborhood in $\mathcal{K}$ to embed $e_i$ as a low-dimensional vector $\mathbf{e}_i$.

## Framework




- **Encoder**

1. A relation-specific *transforming function* $T_r(.)$ applied on a neighbor $\mathbf{e}_j^I$:

$$T_r(\mathbf{e}_j^I) = \mathbf{e}_j^I - \mathbf{w}_r^\top \mathbf{e}_j^I \mathbf{w}_r.$$

2. An entity $e_i$'s neighbors are aggregated as:

$$\mathbf{e}_i^O = \sum_{(r,e_j)\in N_{\mathcal{K}}(i)} (\alpha_{j|i,q}^{\text{Logic}} + \alpha_{j|i,q}^{\text{NN}}) T_r(\mathbf{e}_j^I).$$

- **Decoder**

Scoring function on the output embedding:

$$\phi^O(s,q,o) = -|\mathbf{s}^O + \mathbf{q} - \mathbf{o}^O|_{L1}.$$

## Desired Properties of Neighborhood Aggregator for KG

- **Permutation Invariant** - Neighbors of an entity are naturally unordered. The aggregator should be irrelevant to potential permutations of the neighbors.
- **Redundancy Aware** - Facts in KGs tend to depend on each other. It is beneficial to exploit the redundancy in an entity's neighborhood.
- **Query Relation Aware** - An aggregator may take advantage the query relation in concern to concentrate on relevant facts in the neighborhood.

## Logic Rule Mechanism

1. The confidence of the implication $r_1 \Rightarrow r_2$:

$$\mathcal{P}(r_1 \Rightarrow r_2) = \frac{\sum_{e\in\mathcal{E}} \mathbb{1}(r_1 \in N_{\mathcal{R}}(e) \wedge r_2 \in N_{\mathcal{R}}(e))}{\sum_{e\in\mathcal{E}} \mathbb{1}(r_1 \in N_{\mathcal{R}}(e))}.$$

2. The logic rule mechanism of measuring neighbors' importance:

$$\alpha_{j|i,q}^{\text{Logic}} = \frac{\mathcal{P}(r \Rightarrow q)}{\max(\{\mathcal{P}(r' \Rightarrow r) | r' \in N_{\mathcal{R}}(e_i) \wedge r' \neq r\})}.$$

- The *numerator* promotes relations $r$ strongly implying $q$.
- The *denominator* demotes those implied by some other relation in the same neighborhood.

## Neural Network Mechanism

The importance of an entity $e_i$'s neighbor $e_j$ is measured by:

$$\alpha_{j|i,q}^{\text{NN}} = \text{softmax}(\alpha'_{j|i,q}) = \frac{\exp(\alpha'_{j|i,q})}{\sum_{j'\in N_{\mathcal{E}}(i)} \exp(\alpha'_{j'|i,q})}$$

The unnormalized attention weight $\alpha'_{j|i,q}$ is given by an attention neural network as

$$\alpha'_{j|i,q} = \mathbf{u}_a^\top \cdot \tanh(\mathbf{W}_a \cdot [\mathbf{z}_q; T_r(\mathbf{e}_j^I)]),$$

where $\mathbf{u}_a$ and $\mathbf{W}_a \in \mathbb{R}^{d\times 2d}$ are global attention parameters, while $\mathbf{z}_q$ is a relation-specific attention parameter for the query relation $q$.

## Training and Experiments

| | Task 1: Link Prediction | | | | | | | | | | Task 2: Triplet Classification | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Subject-10 | | | | | Object-10 | | | | | Subject | | | Object | | | Both | | |
| Model | MR | MRR | Hits@10 | @3 | @1 | MR | MRR | Hits@10 | @3 | @1 | 1K | 3K | 5K | 1K | 3K | 5K | 1K | 3K | 5K |
| MEAN | 293 | 0.310 | 48.0 | 34.8 | 22.2 | **353** | 0.251 | 41.0 | 28.0 | 17.1 | 87.3 | 84.3 | 83.3 | 84.0 | 75.2 | 69.2 | 83.0 | 73.3 | 68.2 |
| LSTM | 353 | 0.254 | 42.9 | 29.6 | 16.2 | 504 | 0.219 | 37.3 | 24.6 | 14.3 | 87.0 | 83.5 | 81.8 | 82.9 | 71.4 | 63.1 | 78.5 | 71.6 | 65.8 |
| LAN | **263** | **0.394** | **56.6** | **44.6** | **30.2** | 461 | **0.314** | **48.2** | **35.7** | **22.7** | **88.8** | **85.2** | **84.2** | **84.7** | **78.8** | **74.3** | **83.3** | **76.9** | **70.6** |

- **Training**

We apply a margin-based ranking loss on each triplet $(s,q,o)$ as

$$l^O(s,q,o) = [\gamma - \phi^O(s,q,o) + \phi^O(s',q,o')]_+.$$

- **Subtask**:

Scoring function on the input embedding:

$$\phi^I(s,q,o) = -|\mathbf{s}^I + \mathbf{q} - \mathbf{o}^I|_{L1}.$$

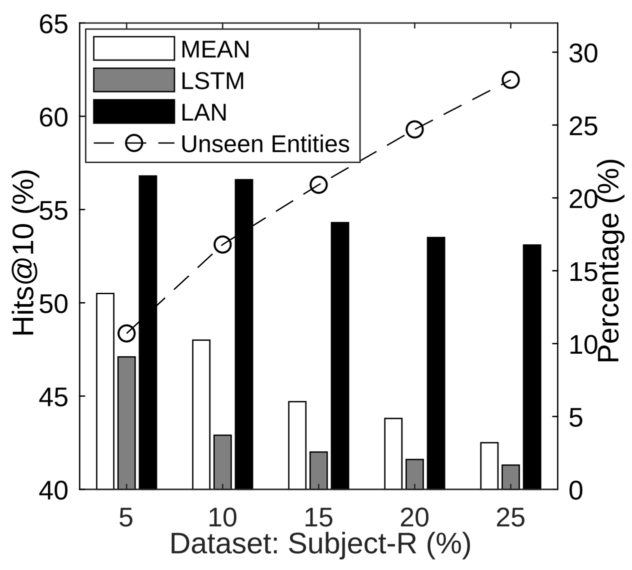
Then a similar margin-based ranking loss $l^I(s,q,o)$ is defined for the subtask.

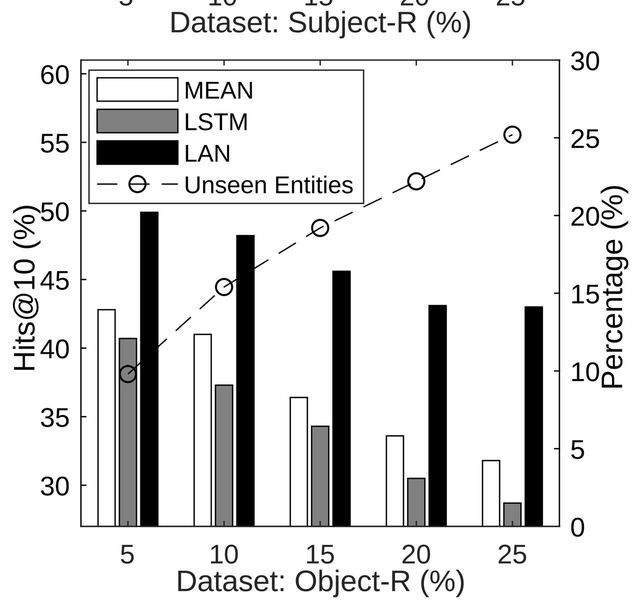
- **Training objective**:

$$\min \sum_{(s,q,o)\in\Delta} \sum_{(s',q,o')\in\Delta'_{(s,q,o)}} [l^O(s,q,o) + l^I(s,q,o)].$$

- **Dataset**

Based on WordNet and Freebase, we construct datasets whose test sets contain new entities unseen during training.




| Model | MRR | Hits@10 | Hit@3 | Hits@1 |
|---|---|---|---|---|
| MEAN | 0.310 | 48.0 | 34.8 | 22.2 |
| Global-Attention | 0.331 | 49.7 | 37.7 | 24.0 |
| Query-Attention | 0.355 | 51.9 | 39.5 | 27.0 |
| Logic Rules Only | 0.375 | 54.7 | 42.7 | 28.0 |
| LAN | **0.394** | **56.6** | **44.6** | **30.2** |

| Subject and Query | Neighbors ranked by LAN | Predicted Object from LAN and MEAN |
|---|---|---|
| Jared_Drake_Bell<br>query: origin | place_lived -> Orange_County<br>breed_origin -> Santa_Ana<br>website_owner -> Universal_Records<br>perform_film -> High_Fidelity<br>friend -> Corbin_Bleu_Reivers<br>gender -> Male | LAN: **Orange_County**, Santa_Ana, Laguna_Beach, City_Orange, Fullerton, Huntington_Beach, Costa_Mesa, Greenwich_Village, Newport_Beach, Anaheim<br>MEAN: Costa_Mesa, Santa_Ana, Southern_California, Berkeley, Oslo, Stuttgart, Newport_Beach, Miami, Surrey, San_Jose |
| Georg_Hegel<br>query: profession | influenced_by -> Aristotle<br>interest -> Metaphysics<br>interest -> Aesthetics<br>interset -> Logic<br>interest -> Epistenmology<br>employment$^{-1}$ -> Humboldt_University<br>ethnicity -> Germans<br>gender -> Male | LAN: **Philosopher**, Economist, Librarian, Psychiatrist, Psychologist, Priest, Scientist, Historian, Pediatrics, Designer<br>MEAN: Physicist, Aristotle, Karl_Marx, Gottfried_Leibniz, John_FRS, Immanuel_Kant, **Philosopher**, Economist, Architect, Plato |
| Stephen_Joseph_Harper<br>query: place_lived | institution -> University_of_Calgary<br>politician -> Conservative_Party_of_Canda<br>appointed_by -> Senate_of_Canada<br>position -> Prime_Minister_of_Canda<br>religion -> Evangelicalism<br>profession -> Economist<br>gender -> Male | LAN: Nunavut, Yukon, Saskatchewan, Alberta, Connecticut, British_Coulumbia, Nova_Scotia, **Calgary**, Oklahoma, Edmonton<br>MEAN: Yukon, Nunavut, Alberta, Prince_Edward_Island, Senate_Of_Canda, Nova_Scotia, British_Columbia, Montana, Quebec, Alaska |