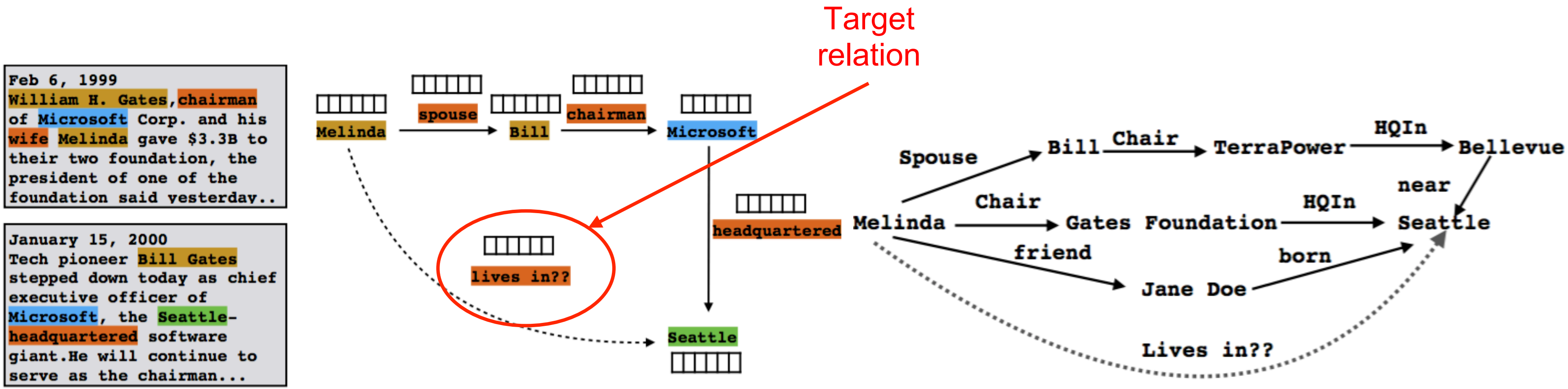
# Chains of Reasoning over Entities, Relations, and Text using RNNs

Authors:
Rajarshi Das, Arvind Neelakantan, David Belanger, Andrew McCallum
Presented by:
Assma Boughoula

# Motivation

- Knowledge Base completion through relation inference
- Infer probability of relation “Lives in” between entities “Melinda” and “Seattle”
- Given the other paths between them in the knowledge graph



# Previous work: Compositional VSM for KB Completion (2015)

**Previously:**

1. No reasoning about entities in path, just relations
2. Reasoning from single path
3. Train a separate model for each relation-type

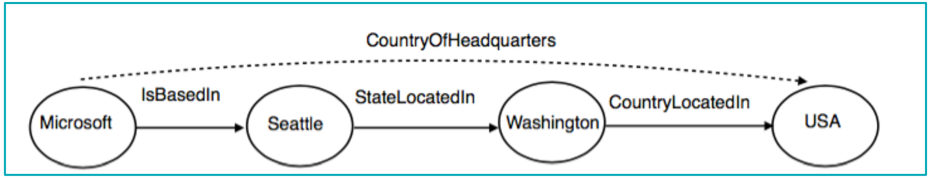
**This work:**

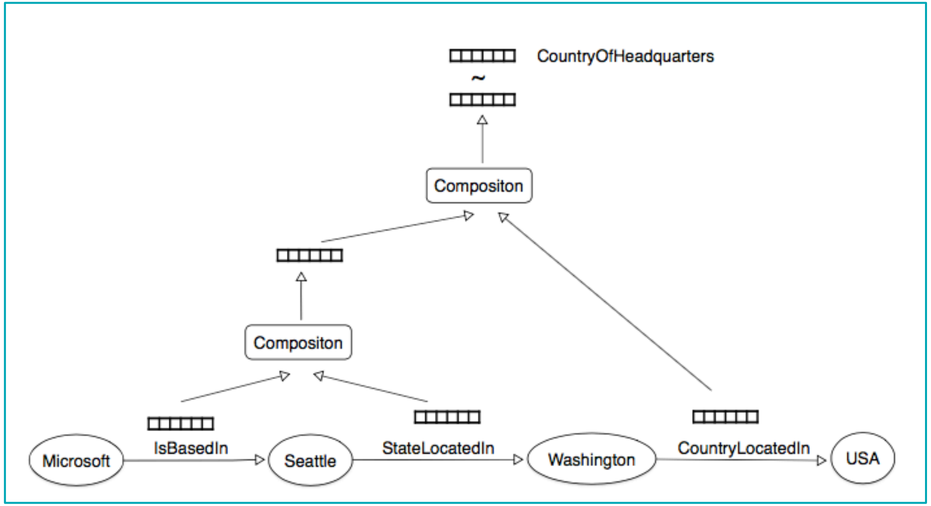
1. Jointly reason about relation-types, entities and entity-types
2. Multiple paths
3. Single RNN that can predict all relation types

# Per-relation Model: Path-RNN

- **Train a separate RNN for each relation type (CountryOfHeadquarters)**

- **Only relation vectors are taken into account**



# Single-Model

**Target Relation: "country of HQ"**
**Start: "Microsoft"**
**Target: "USA"**

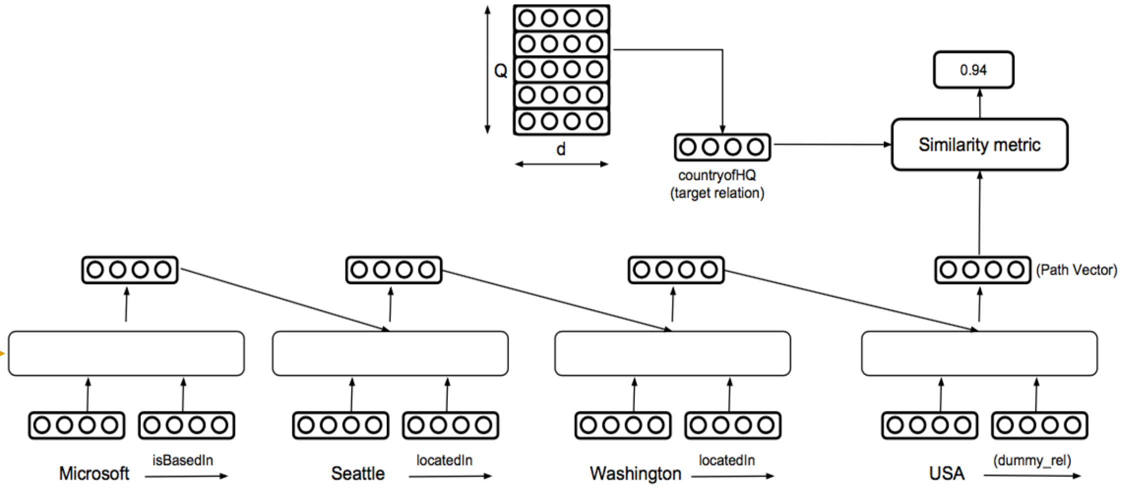
$$Path(e_s \rightarrow e_t): \pi = \{e_s, r_1, e_1, \ldots, r_k, e_t\} \in S$$

$e_j$

$r_k$

Learned during training

Concatenation + RNN step



# Single-Model

**RNN hidden state at step *t* in the path**

**Now**

$$\mathbf{h_t} = f(\mathbf{W_{hh}}\mathbf{h_{t-1}} + \mathbf{W_{ih}}\mathbf{y_{r_t}}).$$

**No dependency on target relation r here!**

**Previous**

$$\mathbf{h_t} = f(\mathbf{W^r_{hh}}\mathbf{h_{t-1}} + \mathbf{W^r_{ih}}\mathbf{y^r_{r_t}}).$$

***f*** = sigmoid function

# Single-Model: incorporating entities

1. Learn Entity Vector Representation

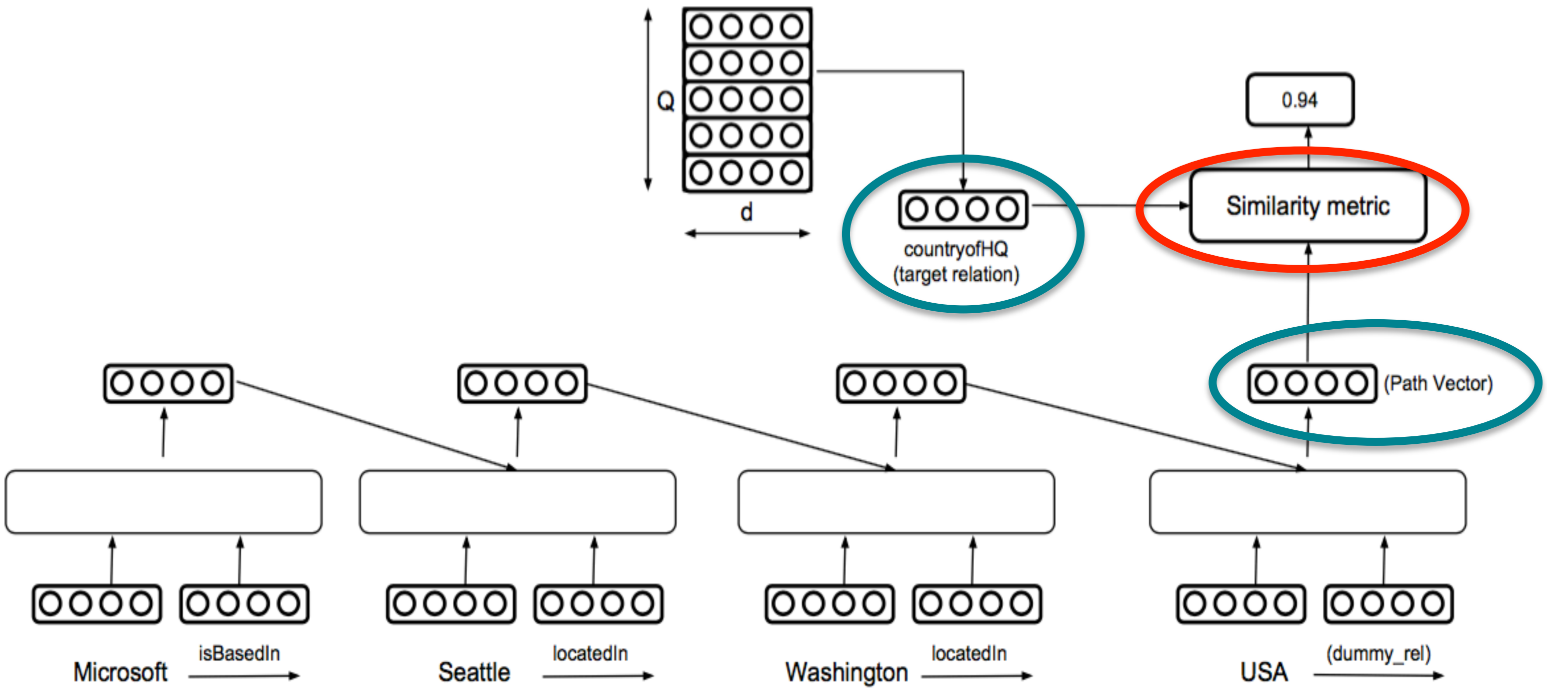
2. Get annotated entity types from FreeBase:
   **“Melinda Gates”**

   **CEO**
   **Duke University Alumni**
   **Philanthropist**
   **American Citezen**

$$\mathbf{h_t} = f(\mathbf{W_{hh}}\mathbf{h_{t-1}} + \mathbf{W_{ih}}\mathbf{y_{r_t}} + \mathbf{W_{eh}}\mathbf{y_{e_t}})$$

**Entity Representation**

# So far …



Q
d
countryofHQ
(target relation)
0.94
Similarity metric
(Path Vector)
Microsoft
isBasedIn
Seattle
locatedIn
Washington
locatedIn
USA
(dummy_rel)

# Single-Model: score pooling

**Top K** $$\mathbb{P}(r|e_s, e_t) = \sigma(\frac{1}{k}\sum_j s_j), \forall j \in \mathcal{K}$$

Assigns 0 weight to some paths

**Average** $$\mathbb{P}(r|e_s, e_t) = \sigma(\frac{1}{N}\sum_{i=1}^{N} s_i)$$

Each path gets same share of gradient regardless of whether it's more/less important

**LogSumExp** $$\mathbb{P}(r|e_1, e_2) = \sigma\left(\log(\sum_i \exp(s_i))\right)$$

Each path gets share of gradient proportional to its score since:

$$\frac{\partial \mathrm{LSE}}{\partial s_i} = \frac{\exp(s_i)}{\sum_i \exp(s_i)}$$

# Single-Model: score pooling

JFK – locatedIn – NYC – locatedIn – NY

YankeeStadium – locatedIn – NYC – locatedIn – NY

Target relation: airport_serves

**Top K** $$\mathbb{P}(r|e_s, e_t) = \sigma(\frac{1}{k}\sum_j s_j), \forall j \in \mathcal{K}$$

Assigns 0 weight to some paths

**Average** $$\mathbb{P}(r|e_s, e_t) = \sigma(\frac{1}{N}\sum_{i=1}^{N} s_i)$$

Each path gets same share of gradient regardless of whether it's more/less important

**LogSumExp** $$\mathbb{P}(r|e_1, e_2) = \sigma\left(\log(\sum_i \exp(s_i))\right)$$

Each path gets share of gradient proportional to its score since:

$$\frac{\partial \mathrm{LSE}}{\partial s_i} = \frac{\exp(s_i)}{\sum_i \exp(s_i)}$$

# Experiments Setup

Dataset:

- Triples: ($e_s$, r, $e_t$ )
- Set of paths S connecting ($e_s$, $e_t$ ) in the knowledge graph

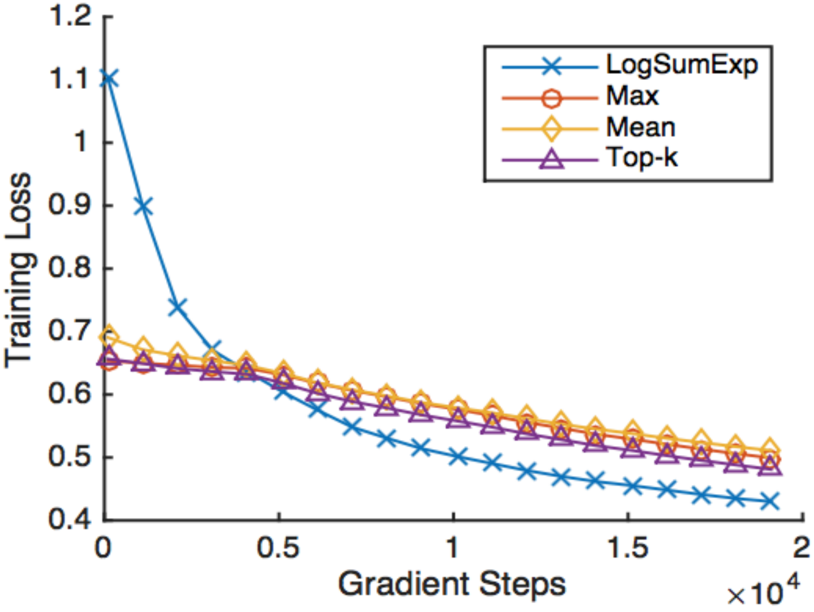
Data is from FreeBase (KB) and ClueWeb(Text)

| Stats | # |
|---|---|
| # Freebase relation types | 27,791 |
| # textual relation types | 23,599 |
| # query relation types | 46 |
| # entity pairs | 3.22M |
| # unique entity types | 2218 |
| Avg. path length | 4.7 |
| Max path length | 7 |
| Total # paths | 191M |

Table 2: Statistics of the dataset.

# Experiments: Effect of Pooling Techniques

- LogSumExp is best
  - Important to include all paths
- Average is worst
  - Important to weigh path scores according to their values



# Experiments Results

| Model | Performance (%MAP) | Pooling |
|---|---|---|
| Single-Model | 68.77 | Max |
| Single-Model | 55.80 | Avg. |
| Single-Model | 68.20 | Top($k$) |
| Single-Model | **70.11** | LogSumExp |
| PRA | 64.43 | n/a |
| PRA + Bigram | 64.93 | n/a |
| Path-RNN | 65.23 | Max |
| Path-RNN | 68.43 | LogSumExp |
| Single-Model | **70.11** | LogSumExp |
| PRA + Types | 64.18 | n/a |
| Single-Model | 70.11 | LogSumExp |
| Single-Model + Entity | 71.74 | LogSumExp |
| Single-Model + Types | **73.26** | LogSumExp |
| Single-Model + Entity + Types | 72.22 | LogSumExp |

Effect of Pooling

Other Multi-Hop Algorithms

Effect Including Entities

# Take-aways:

- **Complete Knowledge Graph by inferring relations between entities using existing paths**
- **Single-Model: trains a single RNN to handle multiple relation types**
- **Incorporating Entity vectors improves results**