

# Generalizing to Unseen Entities and Entity Pairs with Row-less Universal Schema

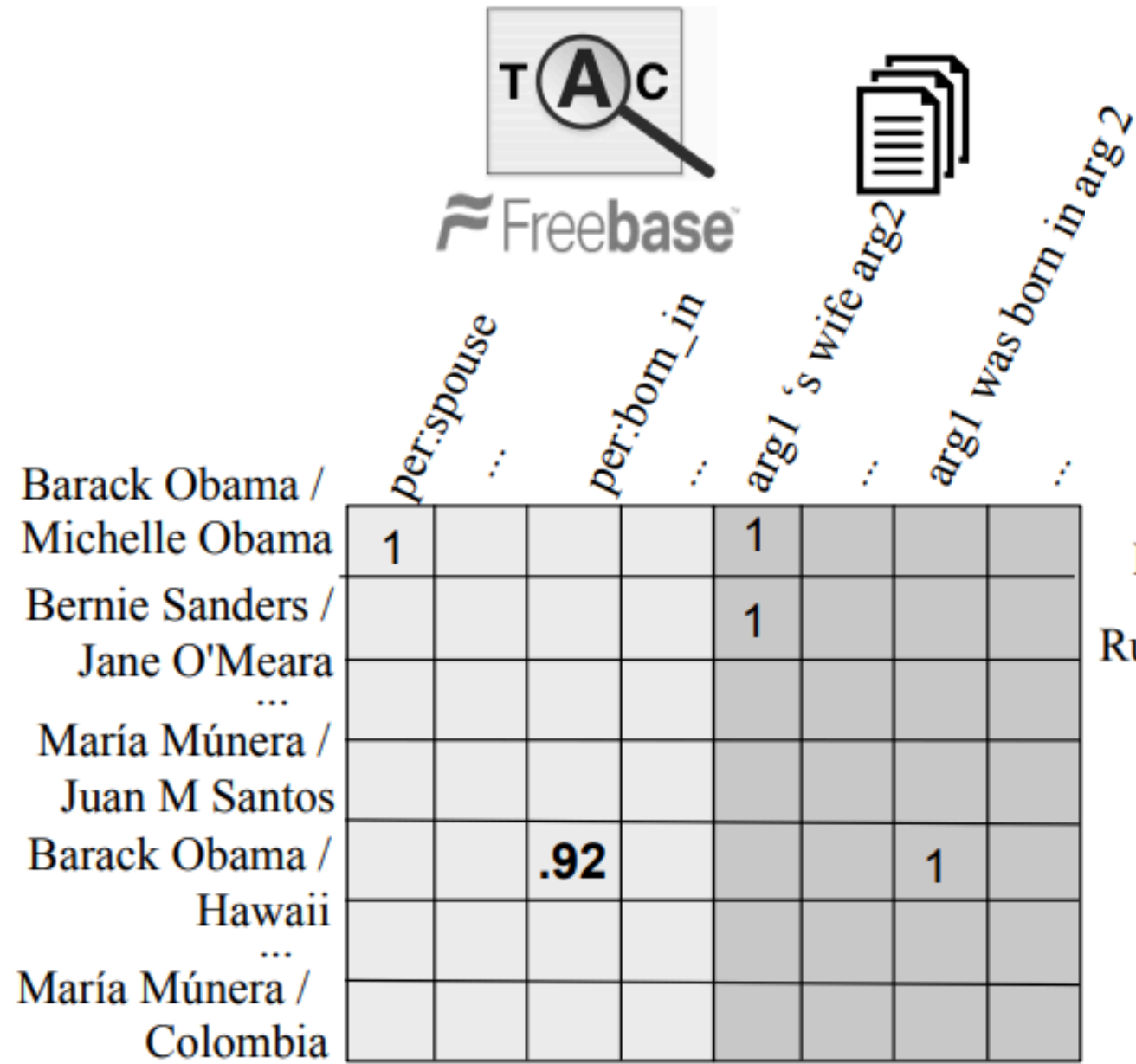
Patrick Verga, Arvind Neelakantan and Andrew McCallum

Present by Ranran Li

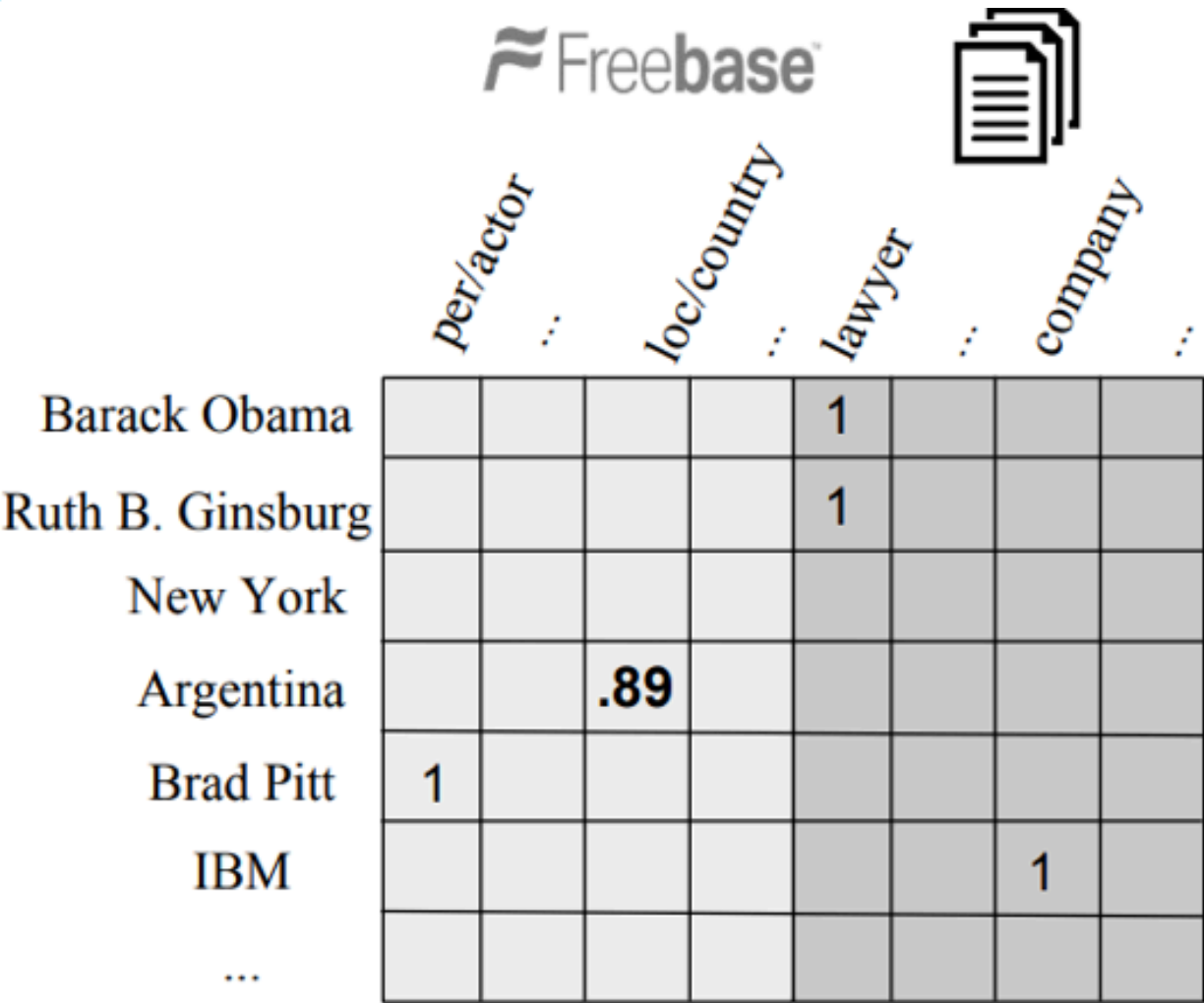
# Task: Automatic Knowledge Base Construction(AKBC)

- Building a structured KB of facts using raw text evidence, and often an initial seed KB to be augmented.
- KB:
  - contain entity type facts
  - *Sundar Pichai **IsA** Person*
  - *contain relation facts:*
  - CEO\_Of(Sundar Pichai, Google)

Relation extraction:

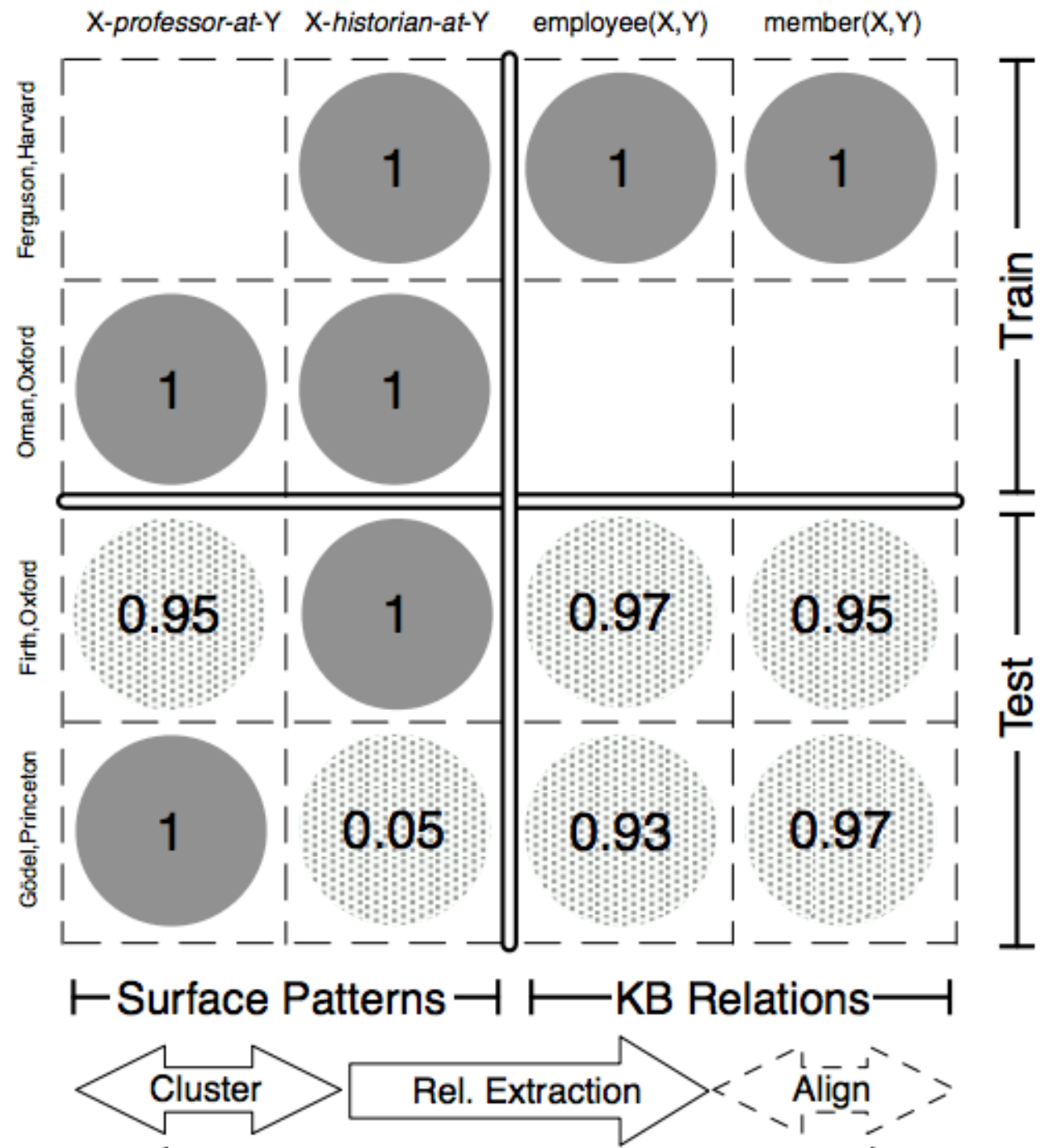


Entity type prediction:



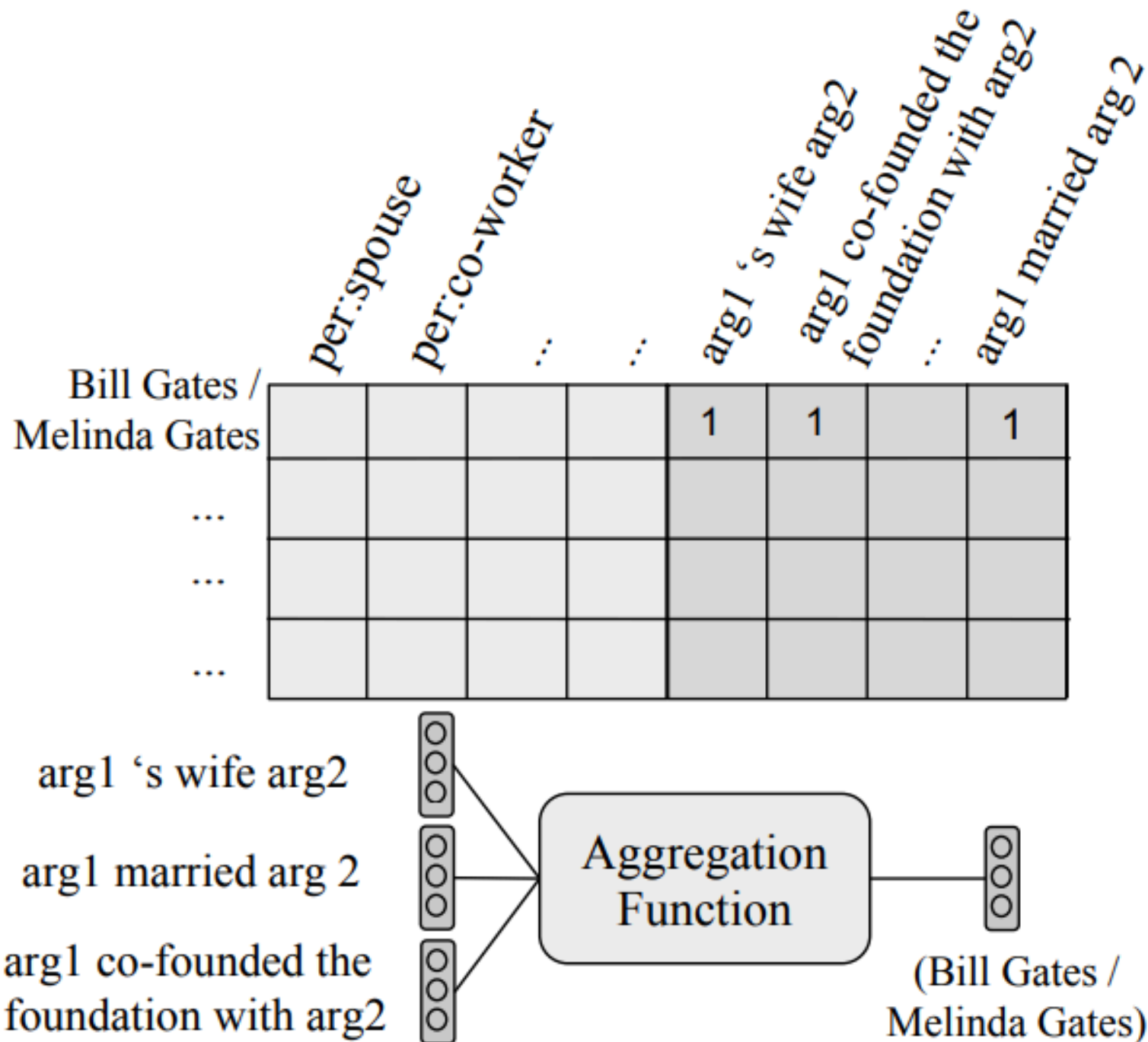
# Background: Universal Schema

- (Riedel et al., 2013)
- relation extraction and entity type prediction is typically modeled as a matrix completion task.



# Motivation:

- **Problem:** Universal schema: **Unseen** rows and columns observed at test time do not have a learned embedding (cold-start problem)
- **Solution:**
  - a 'row-less' extension of universal schema that generalizes to unseen entities and entity pairs
  - (unseen rows).



Encode each entity or entity pair as **aggregate functions** over their observed column entries.

Benefit: when new entities are mentioned in text and subsequently added to KB, we can directly reason on the observed text evidence to infer new binary relations and entity types for the new entities. This avoids re-training the whole model to learn embeddings for the new entities.

# Notations:

- $(r, c)$ : row and column
- Let  $v(r) \in \mathbb{R}^d$  and  $v(c) \in \mathbb{R}^d$  be the embeddings of  $(r, c)$  that are learned during training.
- The embeddings are learned using Bayesian Personalized Ranking (BPR) (Rendle et al., 2009) in which the probability of the observed triples are ranked above unobserved triples.
- To model the probability between row  $r$  and column  $c$ , we consider the set  $V^-(r)$  which contains the set of column entries that are observed with row  $r$  at training time, i.e

$$\forall \bar{c} \in V^-(r), (r, \bar{c}) \in T$$

- The probability of observing the fact is given by:
- $P(y_{r,c} = 1) = \sigma(v(r) \cdot v(c))$
- where  $y_{r,c}$  is a binary random variable that is equal to 1 when  $(r, c)$  is a fact and 0 otherwise

# Query independent Aggregation Functions

- **Mean Pool** creates a single centroid for the row by averaging all of its column vectors, (query independent)

$$v(r) = \sum_{\bar{c} \in V(\bar{r})} v(\bar{c})$$

- **Max Pool** also creates a single representation for the row by taking a dimension-wise max over the observed column vectors:

$$v(r)_i = \max_{\bar{c} \in V(\bar{r})} v(\bar{c})_i, \forall i \in 1, 2, \dots, d$$



# Query specific Aggregation Functions

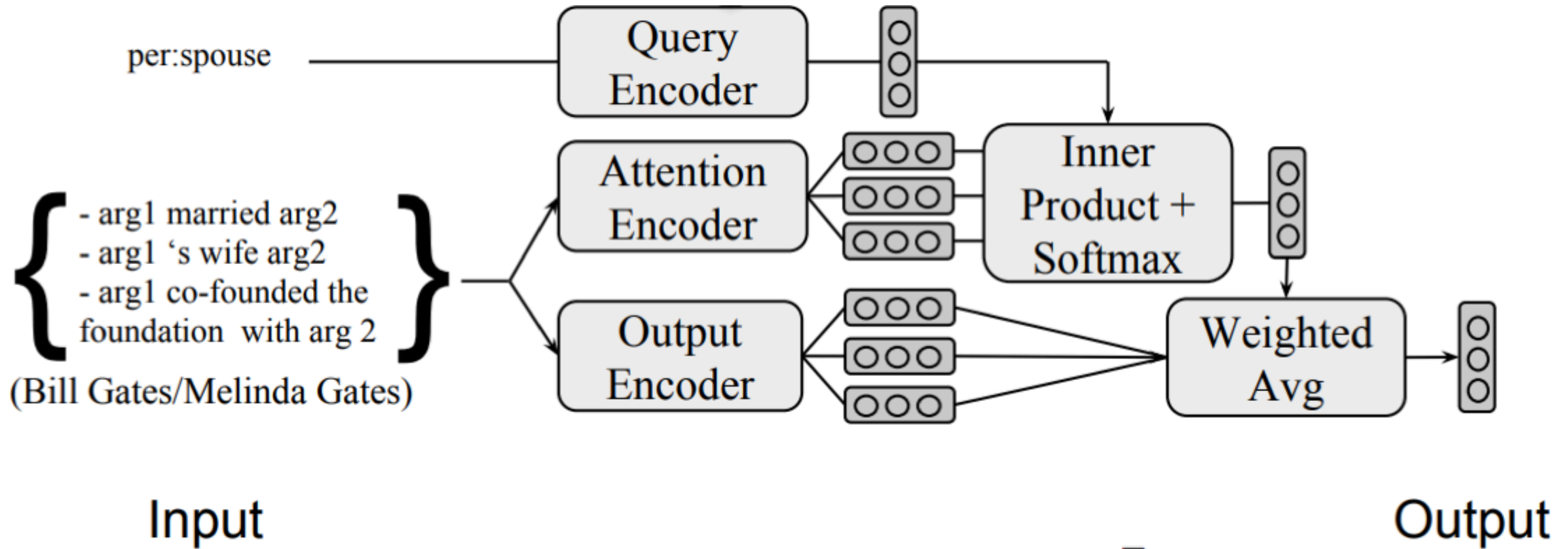
- **Max Relation** aggregation function represents the row as its most similar column to the query vector of interest. Given a query relation

$c$

$$c_{max} = \underset{\bar{c} \in V^-(r)}{\operatorname{argmax}} v(\bar{c}) \cdot v(c)$$
$$v(r) = v(c_{max})$$

- $V^-(r)$  which contains the set of column entries that are observed with row  $r$  at training time

# Attention Aggregation function (Query specific)



$$score_{\bar{c}} = v(c).v(\bar{c}), \forall \bar{c} \in V(\bar{r})$$

$$p_{\bar{c}} = \frac{\exp(score_{\bar{c}})}{\sum_{\hat{c} \in V(\bar{r})} \exp(score_{\hat{c}})}, \forall \bar{c} \in V(\bar{r})$$

$$v(r) = \sum_{\bar{c} \in V(\bar{r})} p_{\bar{c}} \times v(\bar{c})$$

# Training

- Use entity type and relation facts from Freebase (Bollacker et al., 2008) augmented with textual relations and types from Clueweb text (Orr et al., 2013; Gabrilovich et al., 2013).

# Experiment result

- 1. Entity type prediction:
- without unseen

<b>Model</b>	<b>MAP</b>
Entity Embeddings	54.81
Mean Pool	39.47
Max Pool	32.59
Attention	<b>55.66</b>
Max Relation	<b>55.37</b>

(a)

with unseen entities

<b>Model</b>	<b>MAP</b>
Entity Embeddings	3.14
Mean columns	34.77
Max column	43.20
Mean Pool	35.53
Max Pool	30.98
Attention	<b>54.52</b>
Max Relation	<b>54.72</b>

(b)

## 2.Relation Extraction

Query	Observed Columns
/baseball/baseball_player	/sports/pro_athlete, /sports/sports_award_winner, /tv/tv_actor, /people/person, /award/award_winner, /people/person
/architecture/engineer	engineer, /book/author, /projects/project_focus, /people/person, sir
/baseball/baseball_player	baseball_player, /sports/pro_athlete, /people/measured_person, /people/person
/computer/computer_scientist	/education/academic, /music/group_member, /music/artist, /people/person
/business/board_member	/organization/organization_founder, /award/award_winner, /computer/computer_scientist, /people/person, president, scientist
/education/academic	/astronomy/astronomer, /book/author

## 2. Relation extraction

used the FB15k-237 dataset from Toutanova et al. (2015)

Model	MRR	Hits@10
Entity-pair Embeddings	31.85	51.72
Mean Pool	25.89	45.94
Max Pool	29.61	49.93
Attention	31.92	51.67
Max Relation	31.71	51.94

MRR = Mean reciprocal rank scaled by 100

Hits@10 = percentage of positive triples ranked in the top 10 amongst their negatives

Predict entity pairs  
that are not seen at  
train time



Model	MRR	Hits@10
Entity-pair Embeddings	5.23	11.94
Mean Pool	18.10	35.76
Max Pool	20.80	40.25
Attention	<b>29.75</b>	49.69
Max Relation	28.46	48.15



(a)

Without unseen

# column-less version of Universal Schema

- (Toutanova et al., 2015; Verga et al., 2016)
- These models learn compositional pattern encoders to parameterize the column matrix in place of direct embeddings.

# Combine row-less and column-less

Model	MRR	Hits@10
Entity-pair Embeddings	31.85	51.72
Entity-pair Embeddings-LSTM	<b>33.37</b>	54.39
Attention	31.92	51.67
Attention-LSTM	30.00	53.35
Max Relation	31.71	51.94
Max Relation-LSTM	30.77	<b>54.80</b>

Without unseen

(a)

Model	MRR	Hits@10
Entity-pair Embeddings	5.23	11.94
Attention	<b>29.75</b>	49.69
Attention-LSTM	27.95	51.05
Max Relation	28.46	48.15
Max Relation-LSTM	29.61	<b>54.19</b>

Predict entity pairs that are not seen at train time

(b)



# Advantage

- Smaller memory footprint since they do not store explicit row representation

Model	Parameters
Entity Embeddings	3.7 e6
Attention	3.1 e5
Mean Pool/Max Pool/Max Relation	1.5 e5

# Summary

- Proposed a 'row-less' extension to Universal Schema that generalizes to unseen entities and entity pairs.
- Can predict both relations and entity types, with an order of magnitude fewer parameters than traditional universal schema.
- Match the accuracy of traditional model, can predict unseen rows with about the same accuracy as rows available at training time.

# REF: Bayesian Personalized Ranking (BPR) (Rendle et al., 2009)

