JOINT RELATIONAL EMBEDDINGS FOR KNOWLEDGE-BASED QUESTION ANSWERING

MIN-CHUL YANG†, NAN DUAN‡, MING ZHOU‡, AND HAE-CHANG RIM†

MCYANG@NLP.KOREA.AC.KR / †KOREA UNIVERSITY, SEOUL, KOREA / ‡MICROSOFT RESEARCH ASIA, BEIJING, CHINA

Motivation

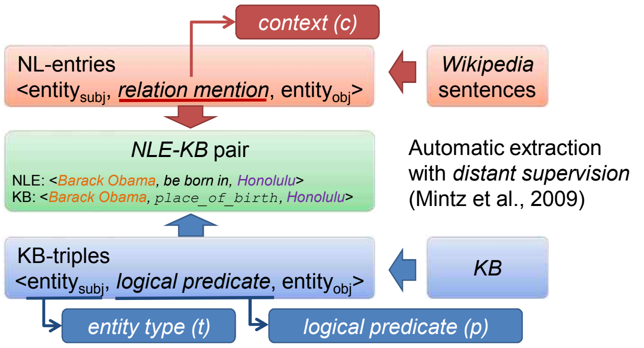
- Fundamental Issues in KB-QA
  - ➢ Given a natural language question,
    1. How to identify entity spans of the question?
    2. How to map the question to its corresponding logical predicate?
- ✓ Solution: Jointly train semantic relations between a question context and logical properties of KB (entities and logical predicates) in the same embedding space.

Relational Components for KB-QA

- Question context (C): represented as n-grams
- Entity type (T): abstract expression of target entities
- Logical predicate (P): canonical form of NL relation phrases

NLE-KB Pair Extraction

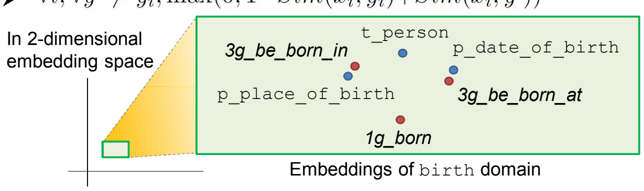
- *NLE-KB* pair: semantically associated tuples for training relational embeddings between NL and KB space
  - ▪ <*Relation Mention*, *Predicate*> pair (*MP*)



- ✓ To get high-quality of tuples
  - ➢ $S(m,p) = \text{PMI}(e_m; e_p) + \text{PMI}(u_m; u_p)$
- ▪ <*Question Pattern*, *Predicate*> pair (*QP*)
  - Frequent lexical patterns starting with *5W1H* words in Web-query logs (Bao et al., 2014)

Joint Relational Embedding Learning

- Construction of training instances
  1. Each *NLE-KB* pair → multiple training triplets
     - ▪ Training triplet $w = [\mathbb{C}, t, p]$ (C: NLE, t and p: KB)
  2. Each training triplet → 3 training pairs
     - ▪ Training pairs: $\mathbb{R} = \{\mathbb{C}\text{-}t, \mathbb{C}\text{-}p, t\text{-}p\}$
  - ✓ In $\mathbb{C}$, the placeholder "<*entity*>" for a target entity is left
- *Ranking loss*-based learning (Weston et al., 2010)
- ✓ Assumption: similarity scores of observed pairs in the training set should be higher than those of any other pairs
  - ➢ $\forall i, \forall y' \neq y_i,\ Sim(x_i, y_i) > 1 + Sim(x_i, y')$
  - ▪ Similarity score: $Sim(a,b) = Sim(r_{ab}) = \mathbb{E}(a)^{\mathsf{T}}\mathbb{E}(b)$
- Embeddings of C, T, and P are trained under the SGD by the above criterion
  - ➢ $\forall i, \forall y' \neq y_i, \max(0, 1 - Sim(x_i, y_i) + Sim(x_i, y'))$



Embeddings of birth domain

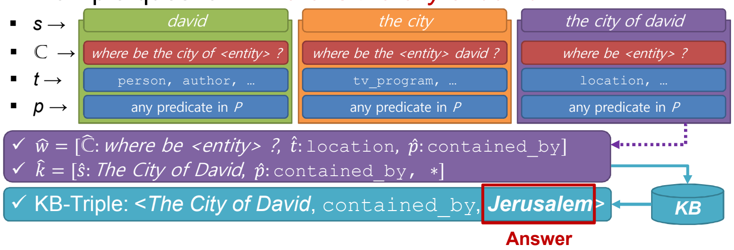
KB-QA using Embedding Models

Given a natural language question q (single-related question),

1. Make all possible decoding triplets W^q, like a training triplet
   - ➢ $\mathbb{C}$: n-grams of q (entity span is replaced with "<*entity*>")
   - ➢ t: one of all available entity types via *Search API* on KB with all string spans in q (candidate entities; s)
   - ➢ p: one of all items in P (candidate logical predicates)
   - ✓ $w_i^q = [\mathbb{C}_i^q, t_i^q, p_i^q]$ is directly linked to KB-query $k_i^q = [s_i^q, p_i^q, *]$, any entities on "$*$" can be potential answers
2. Score W^q through embedding space
   - ➢ Similarities of $\mathbb{R}^q = \{\mathbb{C}_i^q - t_i^q, \mathbb{C}_i^q - p_i^q, t_i^q - p_i^q\}$ are computed
   - ➢ $Sim_{q2k}(q, k^q) = \sum_{r \in \mathbb{R}^q} Z(Sim(r))$ normalization
   - ➢ $\hat{k}(q) = \arg\max_{k \in \mathcal{K}^q} Sim_{q2k}(q, k)$ → corresponding KB-query

- Multi-related question (# target entities = 2)
  - ▪ Heuristic rule: transformed to single-related question
    - ➢ If a pre-defined pair of entity types is detected, they are combined into a *concatenated* entity type
    - ➢ The *concatenated* entity is regarded as one of the candidate entities
  - ❖ *Who plays gandalf in the lord of the rings?*
    `character + film → character-in-film`

❖ Example question: *where is the city of david?*



Experimental Evaluation

- Resource: *Satori* KB / 4.4 M *Wikipedia* articles
- Features: 71,310 n-grams (uni-, bi-, tri-) / 990 entity types / 660 logical predicates (72,960 embeddings)
- Embedding learning: dimension=100, learning rate=0.00001
- Evaluation data: publicly released QA data sets
  - ➢ Free917: 276 QA-pairs (Cai et al., 2013)
  - ➢ WebQuestions: 2,032 QA-pairs (Berant et al., 2013)
- Accuracy on evaluation data

Methods	Free917	WebQ.
Cai and Yates (2013)	59.00%	N/A
Berant et al. (2013)	62.00%	31.40%
Bao et al. (2014)	N/A	37.50%
Our method	71.38%	41.34%

- ▪ Accuracy: average of F_1 scores over all of test questions
- Accuracies of the other methods are from their papers

- Impacts of relationship types

Methods	Free917	WebQ.
Our method	71.38%	41.34%
w/o $\mathcal{T}$-$\mathcal{P}$	70.65%	40.55%
w/o $\mathcal{C}$-$\mathcal{T}$	67.03%	38.44%
w/o $\mathcal{C}$-$\mathcal{P}$	31.16%	19.24%

→ Crucial role in KB-QA

- Problems to be solved
  - ▪ Complex questions requiring multiple stages to detect their target entities
  - ▪ Uncommon questions consisting of rare n-grams