Question Answering

with Subgraph Embeddings

Introduction

Related Work

Task Definition

Embedding Questions and Answers

Experiment

Task

Open-domain question answering

With the rise of large scale structured knowledge bases

Montivation

Don’t use hand-craft lexicons, grammars, and KB schema

Provide a system for open QA able to be trained

Contribution

A more sophisticated inference procedure that is both efficient and can consider longer paths

A richer representation of the answers which encodes the question-answer path and surrounding subgraph of the KB

Introduction

Related Work

Task Definition

Embedding Questions and Answers

Experiment

Information retrieval based

Information retrieval systems first retrieve a broad set of candidate answers by querying the search API of KBs with a transformation of the question into a valid query and then use fine-grained detection heuristics to identify the exact answer

Semantic parsing based

Semantic parsing methods focus on the correct interpretation of the meaning of a question by a semantic parsing system. A correct interpretation converts a question into the exact database query that returns the correct answer

Introduction

Related Work

Task Definition

Embedding Questions and Answers

Experiment

Requirement

A training set of questions paired with answers

A KB providing a structure among answers

Dataset

WEBQUESTIONS	– Train. ex.	2,778
	– Valid. ex.	1,000
	– Test. ex.	2,032
FREEBASE	– Train. ex.	14,790,259
CLUEWEB	– Train. ex.	2,169,033
WIKIANSWERS	– Train. quest.	2,423,185
	– Parap. clust.	349,957
Dictionary	– Words	1,526,768
	– Entities	2,154,345
	– Rel. types	7,210

Introduction

Related Work

Task Definition

Embedding Questions and Answers

Experiment

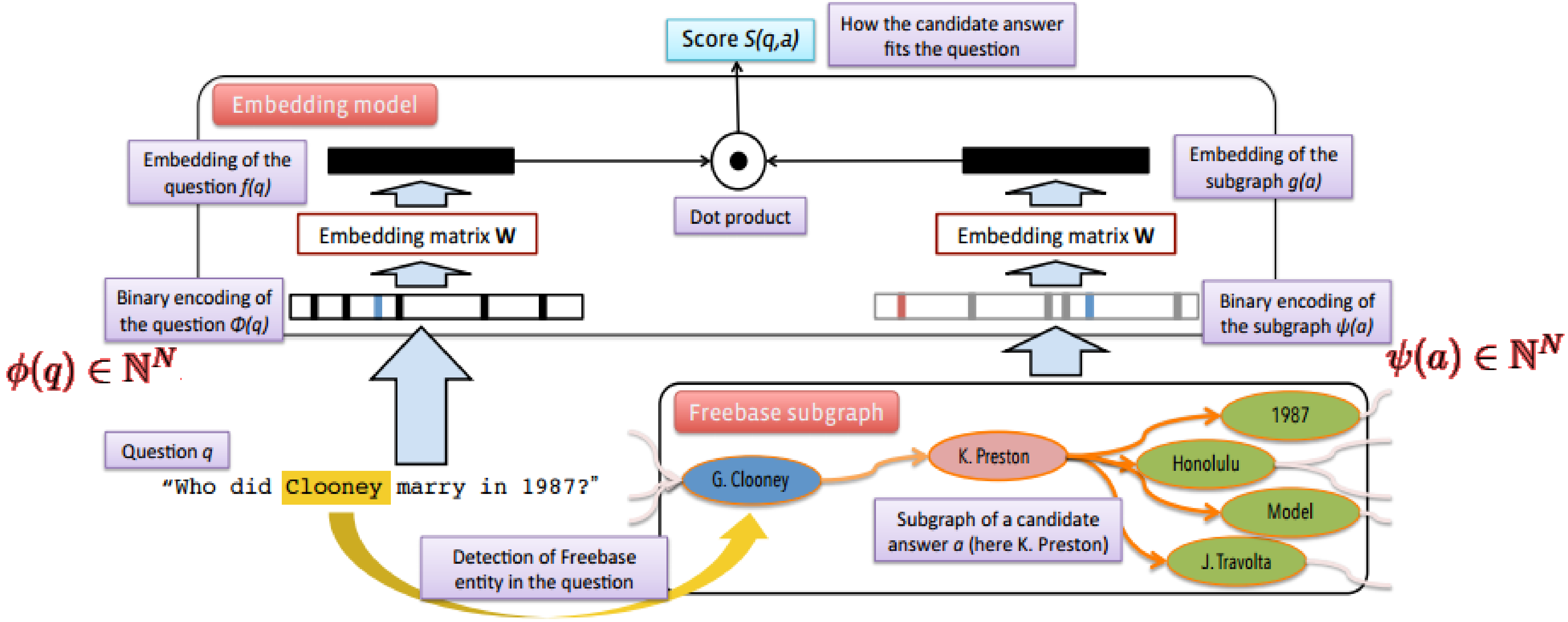
scoring function

$$S(q,a) = f(q)^{\top} g(a).$$

$$N = N_W + N_S$$

N_W the total number of words

N_S the total number of entities and relation types



Representing Candidate Answers

Single Entity

1-of-N_S

Path Representation

1- or 2-hops paths 3-of-N_S or 4-of-N_S

(barack obama, people.person.place of birth, honolulu)

(barack obama, people.person.place of birth, location. location.containedby, hawaii)

Subgraph Representation

C connected entities with D relation types

$3 + C + D$ or $4 + C + D$-of-N_S depending on the path length

Training and Loss Function

Minimize **loss function**

$$\sum_{i=1}^{|\mathcal{D}|} \sum_{\bar{a} \in \bar{A}(a_i)} \max\{0, m - S(q_i, a_i) + S(q_i, \bar{a})\}$$

with the constraint that the columns w_i of $\mathbf{W}$ remain within the unit-ball

i.e., $\forall_i, ||w_i||_2 \leq 1.$

using stochastic gradient descent

Multitask

$$S_{prp}(q_1, q_2) = f(q_1)^\top f(q_2)$$

multi-task the training of our model with the task of paraphrase prediction

Inference

at test time $\hat{a} = \text{argmax}_{a' \in \mathcal{A}(q)} S(q, a')$

where $\mathcal{A}(q)$ is the candidate answer set. This candidate set could be the whole KB but this has both speed and potentially precision issues. Instead, we create a candidate set $\mathcal{A}(q)$ for each question.

C_1 : Answers are directly connected to the questions

C_2 : we can predict its elements in turn using a beam search,
and only add 2-hops candidates to A(q) when these relations appear in their path.
Scores of 1-hop triples are weighted by 1.5

Introduction

Related Work

Task Definition

Embedding Questions and Answers

Experiment

Experiment

Method	P@1 (%)	F1 (Berant)	F1 (Yao)
Baselines			
(Berant et al., 2013) [1]	–	31.4	–
(Bordes et al., 2014) [5]	31.3	29.7	31.8
(Yao and Van Durme, 2014) [14]	–	33.0	42.0
(Berant and Liang, 2014) [2]	–	39.9	43.0
Our approach			
Subgraph & $\mathcal{A}(q) = C_2$	**40.4**	39.2	43.2
Ensemble with (Berant & Liang, 14)	–	**41.8**	**45.7**
Variants			
Without multiple predictions	**40.4**	31.3	34.2
Subgraph & $\mathcal{A}(q)$ = All 2-hops	38.0	37.1	41.4
Subgraph & $\mathcal{A}(q) = C_1$	34.0	32.6	35.1
Path & $\mathcal{A}(q) = C_2$	36.2	35.3	38.5
Single Entity & $\mathcal{A}(q) = C_1$	25.8	16.0	17.8

F1(Yao):questions with no answers are dealt with

Thank You