Simple Question Answering with Subgraph Ranking and Joint-Scoring

Wenbo Zhao[1] Tagyoung Chung[2] Anuj Goyal[2] Angeliki Metallinou[2]
[1]Carnegie Mellon University [2]Amazon Alexa AI

Overview

Task: *knowledge graph* based *simple question* answering (KBSQA)

Knowledge Graph: multi-entity multi-relation directed graph containing fact triples (*subject, relation, object*)

Simple Question: can be answered by a single fact from knowledge graph

Example: "Which Harry Potter series did Rufus Scrimgeour appear in?" *v.s.* (Rufus Scrimgeour, book.book-characters.appears-in-book, Harry Potter and the Deathly Hallows)

Our Method: subgraph ranking + joint scoring model + well-order loss

Result: new state of the art on SimpleQuestions dataset

Motivation

Challenges

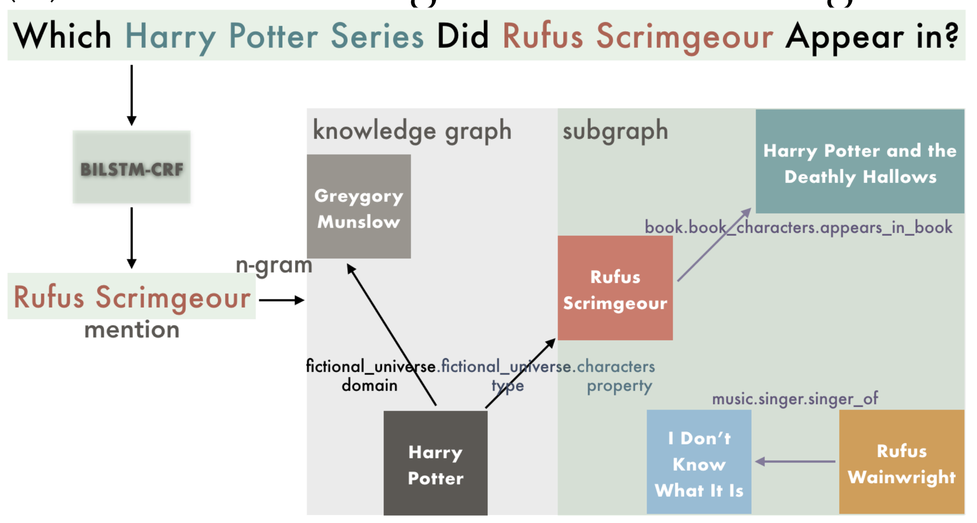
(1) massive size of knowledge graph (billions of facts)
(2) variability of questions in natural language

Two-Step Solution

(1) subgraph selection
(2) fact selection

Conventional Approaches

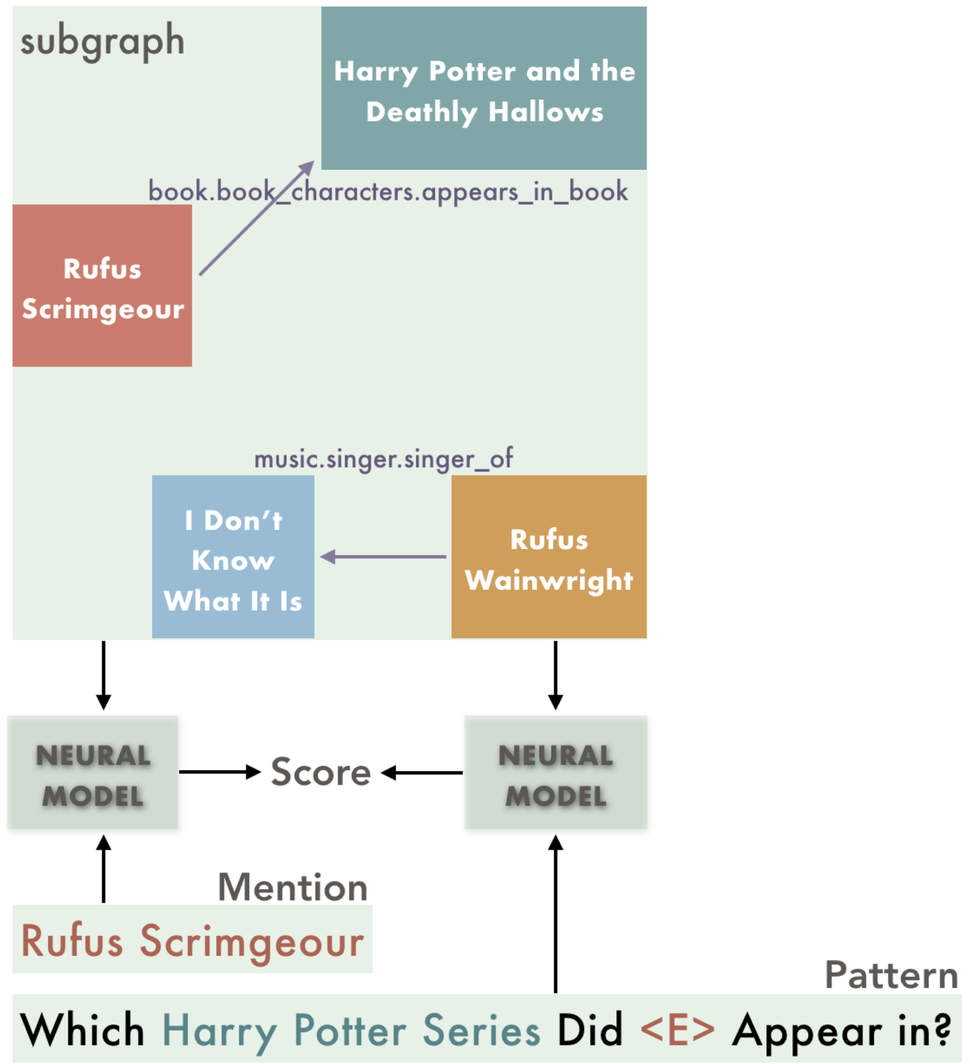
(1) sequence labeling with BiLSTM-CRF + subgraph selection with n-grams
(2) match-scoring model + ranking loss



Problems

(1) subgraph is not ranked by *relevance*
(2) need leverage *dependency* between mention–subjects and pattern–relations
(3) ranking loss is *suboptimal*

$$s^*, r^* \leftarrow \underset{s,r}{\operatorname{argmin}} \left[\text{score}(m, s^-) - \text{score}(m, s^+) + \lambda\right]_+ + \left[\text{score}(p, r^-) - \text{score}(p, r^+) + \lambda\right]_+$$



Proposed Methods

(1) A subgraph ranking method with combined literal and semantic score

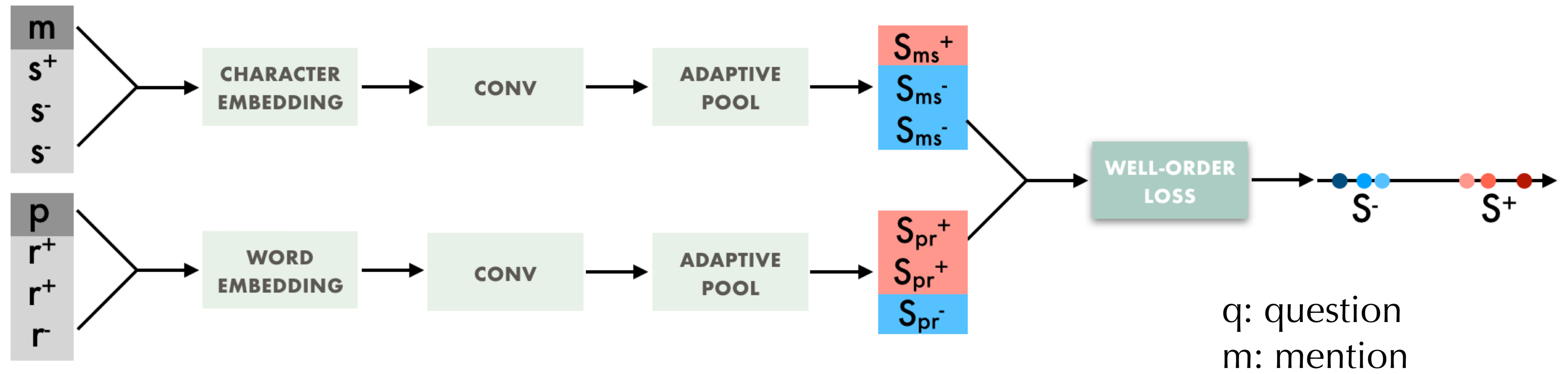
$$\text{score}(s, m) = \tau \underbrace{|\sigma|}_{\text{length of longest common subsequence}}(s, m) + (1 - \tau) \log \mathbb{P}(s, m)$$

$$\begin{aligned}
\mathbb{P}(s, m) &= \mathbb{P}(s|m)\mathbb{P}(m) \\
&= \mathbb{P}(w_1, \dots w_N | \tilde{w}_1, \dots \tilde{w}_M)\mathbb{P}(\tilde{w}_1, \dots \tilde{w}_M) \\
&= \prod_{i=1}^{N} \mathbb{P}(w_i | \tilde{w}_1, \dots \tilde{w}_M)\mathbb{P}(\tilde{w}_1, \dots \tilde{w}_M) \\
&= \prod_{i=1}^{N} \left(\prod_{k=1}^{M} \mathbb{P}(w_i | \tilde{w}_k) \right) \mathbb{P}(\tilde{w}_1, \dots \tilde{w}_M) \\
&= \prod_{i=1}^{N} \left(\prod_{k=1}^{M} \mathbb{P}(w_i | \tilde{w}_k) \right) \prod_{j=1}^{M-1} \mathbb{P}(\tilde{w}_{j+1} | \tilde{w}_j)\mathbb{P}(\tilde{w}_1)
\end{aligned}$$

$\mathbb{P}(w_i | w_j) \approx \exp\{\hat{w}_i^T \hat{w}_j\}$

$\hat{w}$: *pretrained GloVe vector*

(2) A low-complexity joint-scoring CNN model



q: question
m: mention
p: pattern
s: subject
r: relation
S..: score
+/-: positive/negative
I: index set for subjects
J: index set for relations
f/g: character/word CNN
h: scoring map

(3) A well-order loss

$$\min_{q \in \mathcal{Q}, (s,r) \in \mathcal{S}^n_{q\downarrow} \times \mathcal{R}^n_{q\downarrow}} \left[|I^+| \sum_{i^-} h_f(m_q, s^{i^-}) - |I^-| \sum_{i^+} h_f(m_q, s^{i^+}) + |I^+||I^-|\lambda \right]_+ + \left[|J^+| \sum_{j^-} h_g(p_q, r^{j^-}) - |J^-| \sum_{j^+} h_g(p_q, r^{j^+}) + |J^+||J^-|\lambda \right]_+$$

Features

(1) jointly consider both input pairs and their dependency
(2) dependency dynamically adjusted by |I| and |J|
(3) subject mismatch induces larger loss
(4) penalize subject mismatch → prune incorrect relations

Experiments

Dataset

SimpleQuestions: *108,442 questions*
Train/Valid/Test: *75,910/10,845/21,687*

Knowledge Graph

Freebase (FB2M): *2,150,604 entities/6,701 relations/14,180,937 facts*

Results

Table 1. Subgraph Selection Results

Rank Method	Top-N	Recall
Literal: $\|\sigma\|$ + heuristics (Yin et al., 2016)	1	0.736
	5	0.850
	10	0.874
	20	0.888
	50	0.904
	100	0.916
Semantic: $\log \mathbb{P}$	1	0.482
	10	0.753
	20	0.854
	50	0.921
	100	0.848
Joint: $0.9\|\sigma\| + 0.1 \log \mathbb{P}$	1	0.855
	5	0.904
	10	0.920
	20	0.927
	50	0.945
	100	0.928

Table 2. Fact Selection Accuracy

Approach	Object (%)	Subject (%)	Relation (%)
AMPCNN (Yin et al., 2016)	76.4		
BiLSTM (Petrochuk & Zettlemoyer, 2018)	83.57		
AMPCNN + Well-Order Loss	77.69		
Joint Model + Well-Order Loss	81.10	87.44	69.22
Joint Model + Well-Order Loss + Top-50 Subgraph	**85.44**	91.47	76.98
Joint Model + Well-Order Loss + Top-1 Subgraph	79.34	87.97	84.12

Table 3. Error Decomposition (%) (total 3,157 errors)

Incorrect Subject only	8.67
Incorrect Relation only	16.26
Incorrect Subject and Relation	34.50
Other	40.57

Conclusions

(1) our ranking method improves subgraph selection
(2) our joint-scoring model with well-order loss improves fact selection
(3) incorrect subject or relation can still lead to correct answer