Saarland Informatics Campus

UNIVERSITÄT DES SAARLANDES

max planck institut informatik

Question Answering Systems

QA over heterogeneous sources

Rishiraj Saha Roy
Max Planck Institute for Informatics, Germany

9

Saarland University, Summer Semester 2020

Question of the day

s

How can we make question answering sytems work over heterogeneous sources?

KG/KB-QA ∪ Text-QA

You'll find this covered in

- PullNet: Open Domain Question Answering with Iterative Retrieval on Knowledge Bases and Text
  - Sun et al.
  - EMNLP 2019
  - https://www.aclweb.org/anthology/P17-1171.pdf
- Interpretable Question Answering on Knowledge Bases and Text
  - Sydorova et al.
  - ACL 2019
  - https://www.aclweb.org/anthology/P19-1488.pdf

Research paper 1

PullNet: Open Domain Question Answering with Iterative Retrieval on Knowledge Bases and Text

Google Research

PullNet overview

Heterogeneous QA (KG+Text)

- Early fusion
- Late fusion
- Unified representation

KG

KG+Text | Text ×

1 ▪ Works over KGs and text

GRAFT-Net

2 ▪ Based on early fusion philosophy

complex

3 ▪ Focuses on multi-hop questions

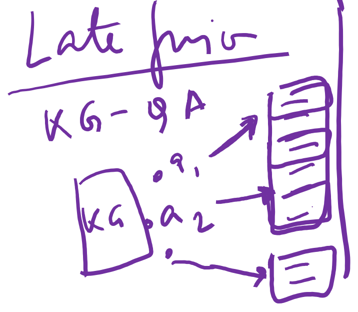
indirection/chain-join | birthplace of director of jur. park

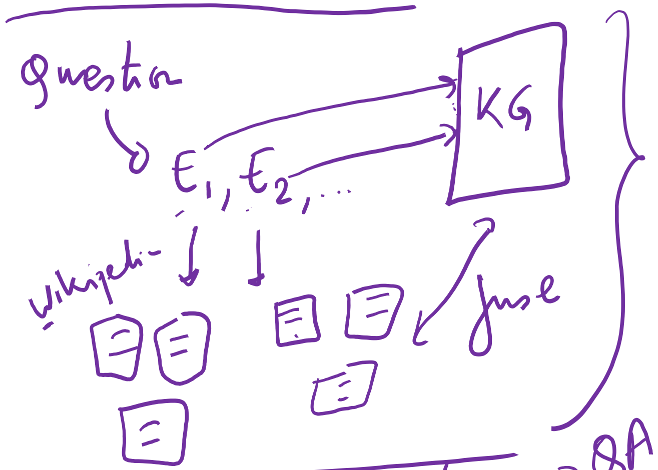
4 ▪ Uses question-focused subgraph

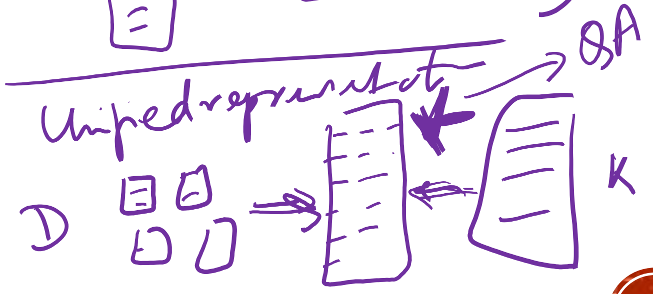
q-nbd/ context graph

5 ▪ Judiciously expands subgraph

▪ Uses classifiers for expansion points and answers





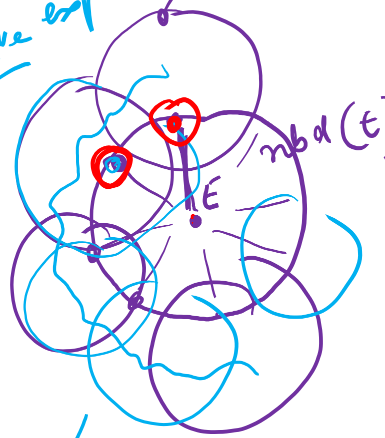


PullNet: Overview

- Works over KGs and text
- Based on early fusion philosophy
- Focuses on multi-hop questions
- Uses question-focused subgraph
- Judiciously expands context subgraph
- Uses 2 classifiers for expansion points and find answers

naive expansion → lp of ss

how many 2 hops ??



nbd(E) → q E

1-hop nbd is good for an simple q.

v.v. large!!

× Complex question!!

Pull facts on expansion pts only

PullNet: Model

1 ▪ Defining the question subgraph – nodes, edges → Heterogeneous graph multiple types of nodes HIN

2 ▪ Iterative subgraph construction

  a ▪ Pull operations

  b ▪ Classify operations

  c ▪ Update operation → context expansion

3 ▪ Training + Answering

The question subgraph

Nodes V — entity v_e 1 □

fact node v_f 2 ○

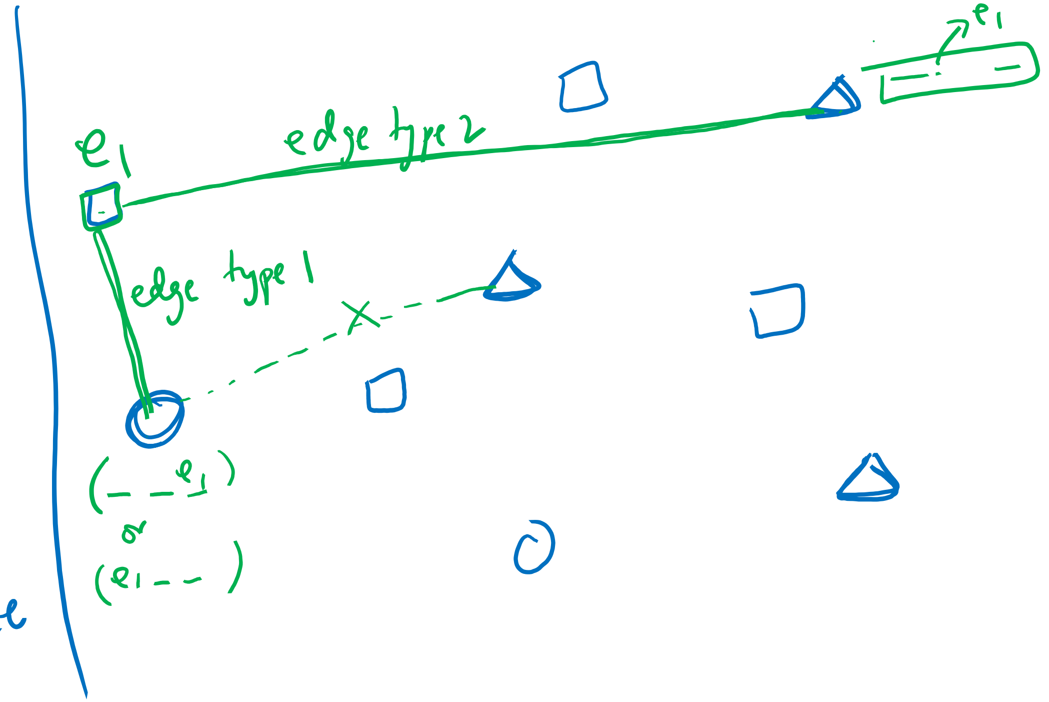
text node v_d 3 △

document → sentence

heterogeneous

triple (v_s, r, v_o)

Edges $\mathcal{E}$



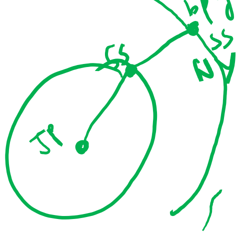
Iterative subgraph construction

Algorithm 1 PullNet

Initialize question graph G_q^0 with question q and question entities, with $\mathcal{V}^0 = \{e_{q_i}\}$ and $\mathcal{E}^0 = \emptyset$.
for $t = 1, \cdots, T$ **do**
    Classify the entity nodes in the graph and select those with probability larger than ϵ → threshold
    $\{v_{e_i}\} = \texttt{classify\_pullnodes}(G_q^t, k)$
    for all v_e in $\{v_{e_i}\}$ **do**
        Perform pull operation on selected entity nodes
        $\{v_{d_i}\} = \texttt{pull\_docs}(v_e, q)$
        $\{v_{f_i}\} = \texttt{pull\_facts}(v_e, q)$
        for all v_d in $\{v_{d_i}\}$ **do**
            Extract entities from new document nodes
            $\{v_{e(d)_i}\} = \texttt{pull\_entities}(v_d)$
        for all v_f in $\{v_{f_i}\}$ **do**
            Extract head and tail of new fact nodes
            $\{v_{e(f)_i}\} = \texttt{pull\_headtail}(v_f)$
    Add new nodes and edges to question graph
    $G_q^{t+1} = \texttt{update}(G_q^t)$
Select entity node in final graph that is the best answer
$v_{\text{ans}} = \texttt{classify\_answer}(G_q^T)$

hop count

out of loop



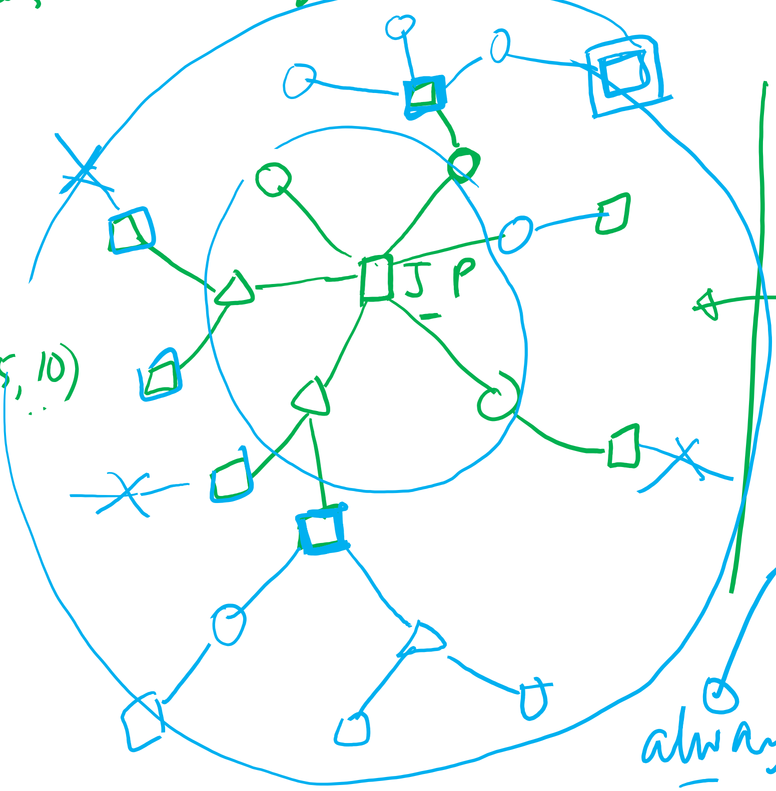
get expansion pth

- top K (1/3, 5, 10)
- E ve
- top-k + E

$\mathcal{V}$ → vertices

$\mathcal{E}$ → edges

q → NED → {e₁, e₂} .. JP



G_q^1 = JP

G_q^2 = JP+

answer classifier

always entity

b of d of JP

nephew → 2 hop (AE)

AE $\xrightarrow{\text{sibl}}$ o $\xrightarrow{\text{child}}$ o

Pull operations

query entity → pull

e_s → E_1, E_2, E_3, E_7, E_4

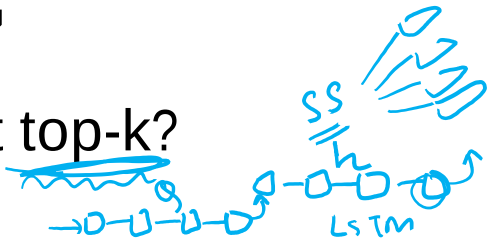
- Pull from documents
  - How to get top-k?

Lucene indexing of sentences

scored top-k

\+ constraint : doc should have query entity

- Pull from KG
  - How to get top-k?



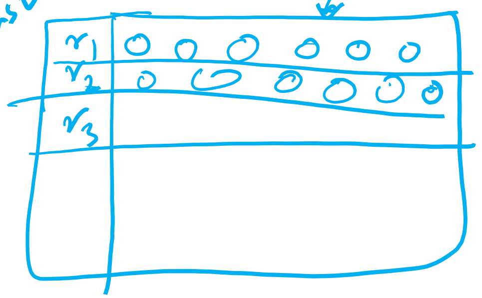
pull-facts (v_e, q)

predicate
relation

h_r → embedding of relation

↳ embed table

w2v / TransE



$$h_q = \mathrm{LSTM}(w_1, \ldots, w_{|q|}) \in \mathbb{R}^n$$

$$S(r, q) = \mathrm{sigmoid}(h_r^T\, h_q)$$

$$S(x) = \frac{1}{1+e^{-x}}$$

- Extract new entities from pulled items

Classify and update operations

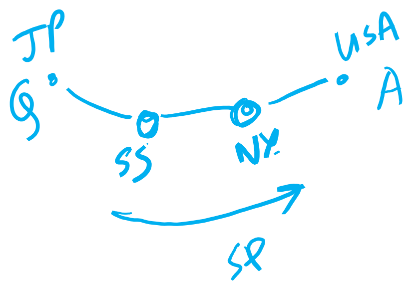
Graphs of Relns. Among Facts and Text Networks

- Classify pull nodes
- Classify answers → end
- Uses CNN-based GRAFT-Net model → EMNLP 2018
- Add newly found entities to subgraph

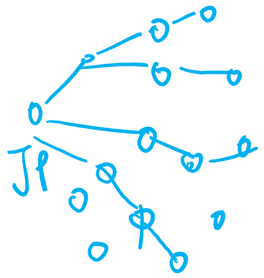
x hops

Training

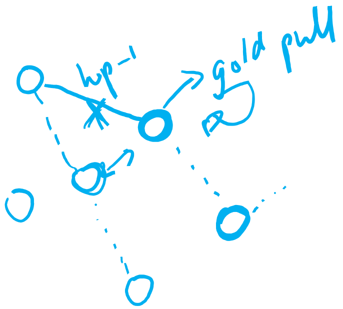
strong supervision: full node annotations



- Distant supervision with QA pairs
- NEQA Uses shortest paths between Q & A in KG => can't run on text alone
- There can be multiple question and answer entities



- Positive and negative sampling
- Uses teacher forcing



- Threshold-based detection ε

Research paper 2

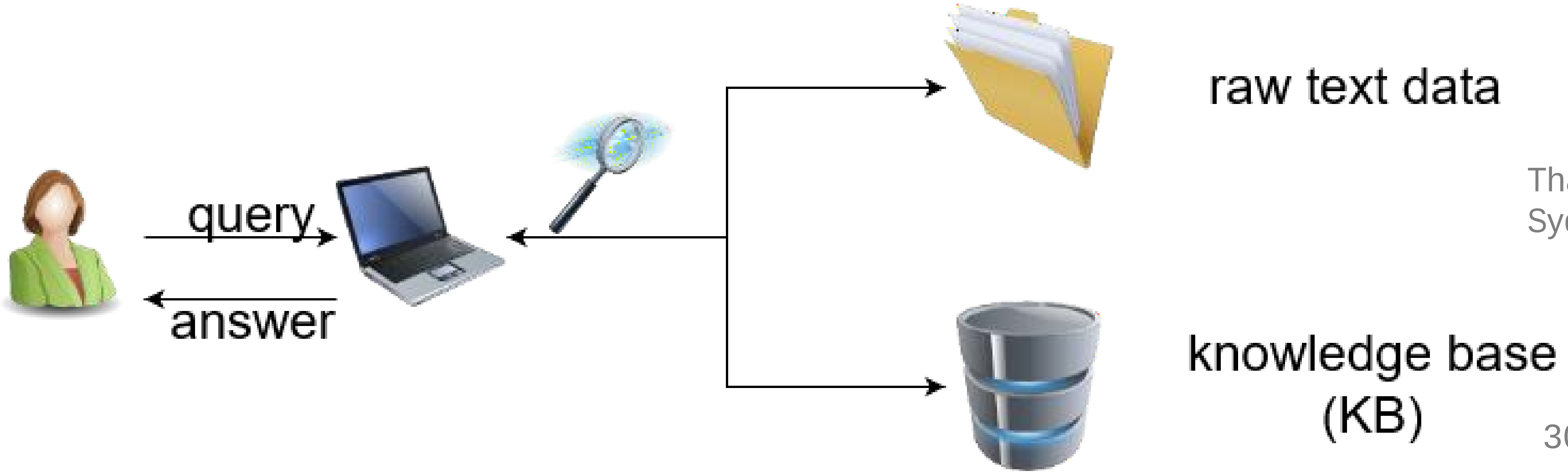
Interpretable Question Answering on

Knowledge Bases and Text

QA on combination of KB and text

1 Information in KB:
- ✓ structured, i.e. clear reasoning chains are possible
- – but often incomplete
- – needs to be maintained and updated

2 Raw text data:
- ✓ immediate access possible
- ✓ versatile data sources
- ✓ context is preserved
- – unstructured / noisy



Thanks to Alona Sydorova for the slides

QA on combination of KB and text

- In many use cases trust and ability to cross-check the model are crucial
- Right to explanation (EU GDPR)

$Q \to \boxed{\text{Sys}} \to A$

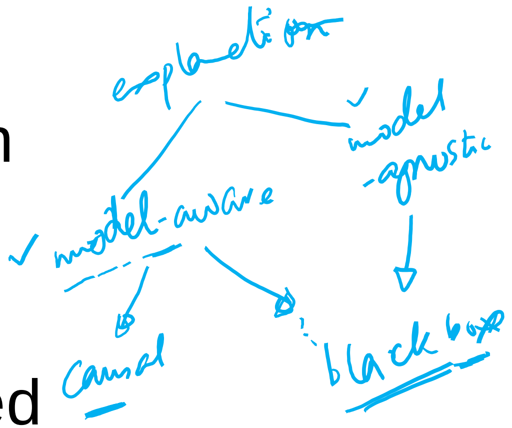
how?

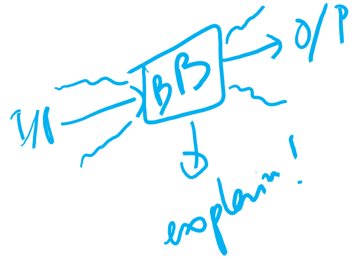
$\Downarrow$

explanation

Contributions

- 1 Adaptation of explanation methods for QA on a combination of KB and text
- 2 A novel automatic evaluation method for explanations, based on ‘fake facts’
- 3 Human evaluation agrees with automatic evaluation





QA model: TextKBQA (Das et al. 2017)

explanation → QA model

1 • A key-value memory network

2 • Attention is used on facts which contain information about entities from the query /question

3 • Questions: SPADES

4 • Fact base: Combination of KB and textual facts

KB

QA model: Task

SPADES

Cloze -QA

Filling in blanks in sentences:

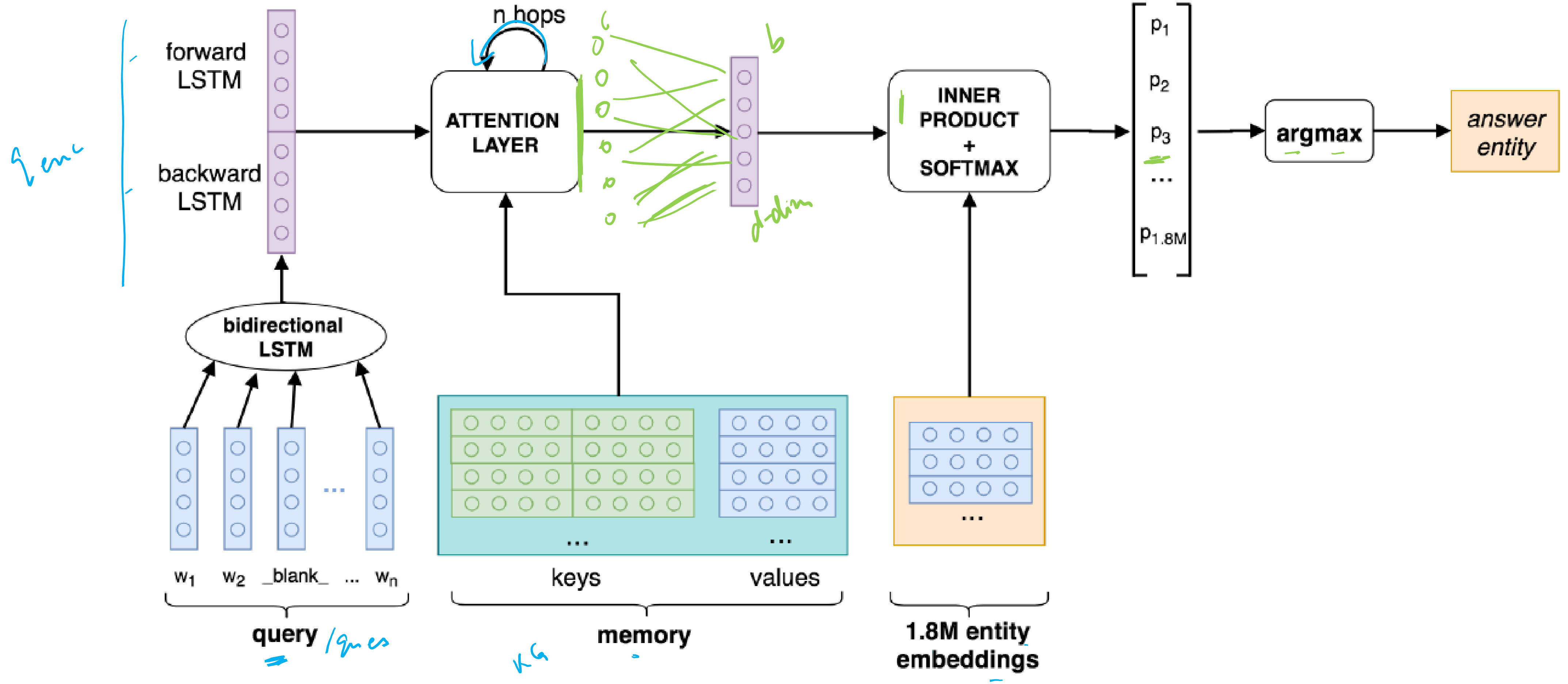
Query

Chicago is the third most populous city in ____.

Answer

the USA

QA model: TextKBQA (Das et al. 2017)



QA model: Notations

$q = [w_1 w_2 \ldots]$ $\vec{e}$

same bi-LSTM

$\to \vec{q} \in \mathbb{R}^{2d}$

KB fact: $(s, r, o) \longrightarrow (\vec{s}, \vec{r}, \vec{o}) \to d$ -> vector

key-value store | key $\vec{k} = [\vec{s}; \vec{r}] \in \mathbb{R}^{2d}$

value $\vec{v} = \vec{o} \in \mathbb{R}^{d}$

Sent: $w_1, w_2 \ldots w_n \to (s, [w_1 \ldots \text{arg1} \ \text{arg2} \ w_n], o)$

$\overrightarrow{LSTM}[w_1, \ldots w_n] ; \to$ concat

$\overleftarrow{LSTM}[w_1, \ldots w_n]$

d $\to \vec{k} \in \mathbb{R}^{2d}$

$\vec{v} = \vec{o} \to d$

$$\mathbf{c}_t = W_t\left(\mathbf{c}_{t-1} + W_p \sum_{(k,v)\in\mathcal{M}} \text{softmax}(\mathbf{c}_{t-1}\cdot\mathbf{k})\mathbf{v}\right)$$

scala

$1 \times 2d$ $\quad 2d \times 1$ $\to d$

$\to d$

$\vec{q}$ rel. pairs (k, v)

context vector:

$\vec{c}_0 = \vec{q}$

$c_1 = \ldots$

$c_2 = \ldots$

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Explanation methods: Notation

$\mathbb{F} = \mathbb{F}_{KB} \cup \mathbb{F}_{text}$

question q $\qquad \mathcal{Y} \subseteq \mathbb{F}$

$\forall f \in \mathbb{F}$: subj. $f \in q$

a_q : answers $=$ TextKBQA $(q, \mathcal{Y})$

capital of Germany?

Germany — —
Germany — —
Germany —

explanation method/function:

$\phi(f, a_q, q, \mathcal{Y}) > \phi(f_2, a_q, q, \mathcal{Y})$

fact $\qquad$ ans $\qquad$ ques

rel. score

Explanation methods

- Task: which facts/sentences were used by the model for answering a question?
- Explanation method assigns relevance scores to KB facts/sentences
- The higher the score, the more relevant is the fact/sentence for the answer

Explanation methods

Question: *<u>Chicago</u> is the third most populous city in _.*

KB

Text

Facts	Relevance
Chicago city.in.state the USA	0.7
Chicago is the most populous city in Illinois.	0.6
New York, Los Angeles and Chicago are the most populous American cities.	0.8
Chicago is a major transportation hub in the United States.	0.2

Explanation methods

- Attention weights
- LIME: Local Interpretable Model-Agnostic Explanations
- IP: Input perturbations

Attention weights

1 $$\phi_{aw}(f, a_q, q, \mathcal{F}) = \text{softmax}(\underline{K_{\mathcal{F}} \cdot \mathbf{q}})_f$$

at hop j

2 $$\phi_{aw_j}(f, a_q, q, \mathcal{F}) = \text{softmax}(K_{\mathcal{F}} \cdot \mathbf{c}_{j-1})_f$$

avg. attn wt.

3 $$\phi_{aw_{avg}}(f, a_q, q, \mathcal{F}) = \frac{1}{h} \sum_{j=1}^{h} \text{softmax}(K_{\mathcal{F}} \cdot \mathbf{c}_{j-1})_f$$

https://en.wikipedia.org/wiki/Softmax_function

Attention weights

Query: *Microsoft Office is a trademark of ____.*

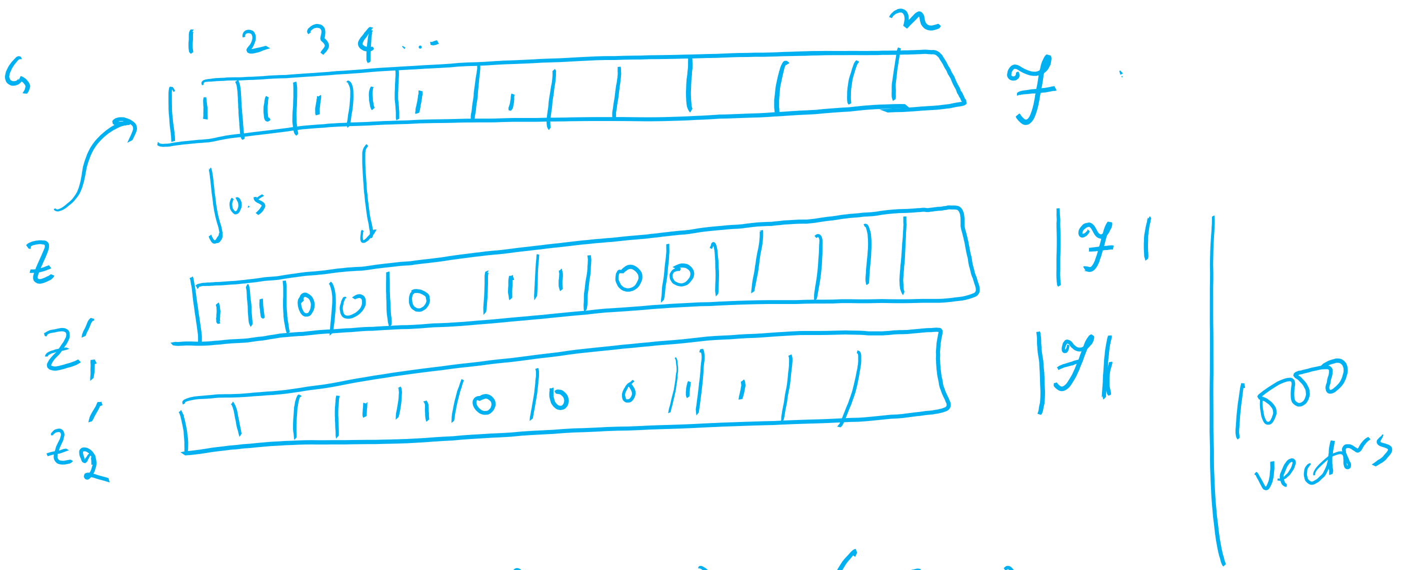
	hop 1	hop 2	hop 3	avg
Microsoft's service pack 2 for Office 97 is 24 Mb.	0.091	0.056	0.000	0.049
For Microsoft this was, and to some extend still is, Office 97.	0.099	0.058	0.000	0.053
Microsoft's Office 97 2004 is the old standard.	0.109	0.078	0.000	0.062
Office 97 is office suite from Microsoft.	0.086	0.049	0.000	0.045
Office 97 is the second-highest revenue generating software for Microsoft, after MS Windows.	0.128	0.076	0.000	0.068
Microsoft had 5 million people sign up to beta-test Office 97.	0.084	0.041	0.000	0.042
Apparently that will now change, as Microsoft attempts to rebrand Office 97.	0.109	0.080	0.000	0.063
Microsoft might then pull Office 97.	0.071	0.039	0.000	0.037
KG Office 97 business.company Microsoft	0.109	0.442	0.998	0.516

LIME

Bag of facts model : n-facts in KG

Sample facts

linear regression



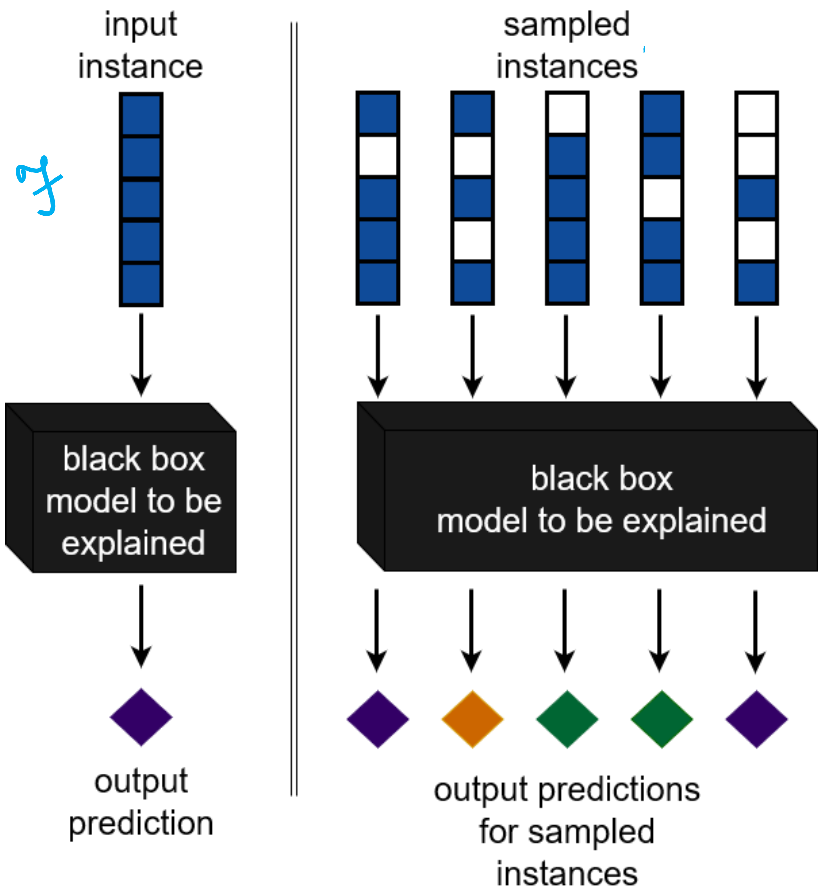
$$\xi(q, \mathcal{F}) = \underset{g \in G}{\operatorname{argmin}} \mathcal{L}(logit, g)$$

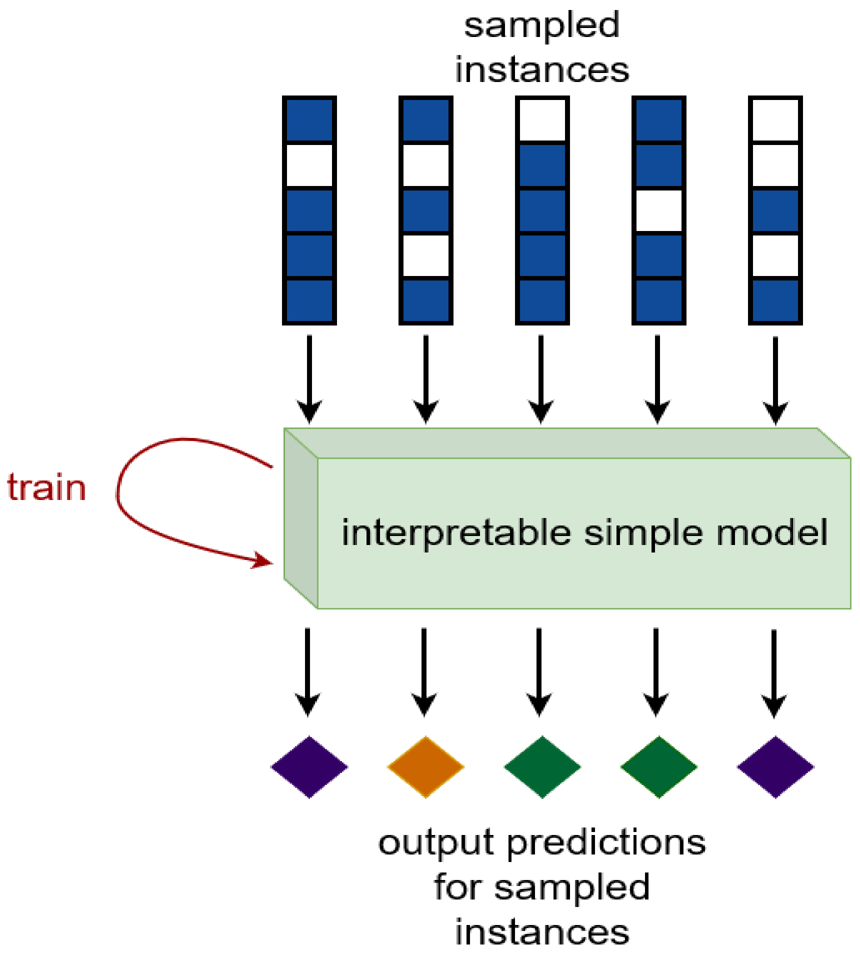
$$\phi_{lime}(f, a_q, q, \mathcal{F}) = w_{g,f}$$

$logit(q, F, a_q) = (E \cdot b) a_q$

$g(z') = w_g \cdot z'$

LIME


input
instance
black box
model to be
explained
output
prediction
sampled
instances
black box
model to be explained
output predictions
for sampled
instances


sampled
instances
train
interpretable simple model
output predictions
for sampled
instances

Input perturbation (IP)

v. bad for good fact

$$\phi_{ip}(f, a_q, q, \mathcal{F}) = \frac{logit(q, \mathcal{F}, a_q) - logit(q, \mathcal{F} \setminus \{f\}, a_q)}{logit(q, \mathcal{F}, a_q)} \quad (7)$$

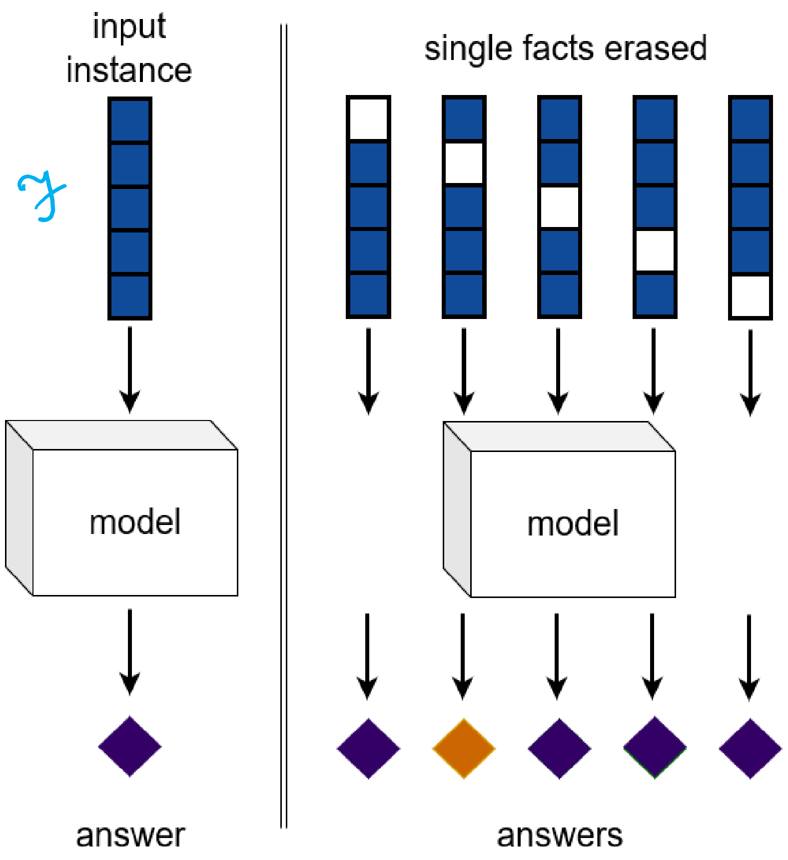
} whole diff → high

f: Germany cap Berlin

- Erase single facts and compare impact on output

- Relevance of a fact corresponds to change in probability for predicting the same answer

Input perturbation



Evaluation

Automatic
User-based

- Based on hybrid document paradigm
- Requires no manual annotation

~ corrupted triples

Hybrid fact sets: for each query, create fake facts by including irrelevant facts and replacing entities in them

Evaluation

Orig query:

Chicago 3rd populous in ___?

Sampled query

Obama studied in ___________________ .

?

Facts

- Obama birthplace Honolulu
- Obama enrolled in Harvard Law School.
- Obama president.of.state the USA
- (etc.)

Facts

- ~~Obama~~ Chicago birthplace Honolulu
- ~~Obama~~ Chicago enrolled in Harvard Law School.
- ~~Obama~~ Chicago president.of.state the USA
- (etc.)

fake

Evaluation: Pointing game

real

$$hit(\phi, q, \hat{\mathcal{F}}) = \begin{cases} 1, & \text{if } rmax(\hat{\mathcal{F}}, q, \phi) \in \mathcal{F}, \\ 0, & \text{if } rmax(\hat{\mathcal{F}}, q, \phi) \in \mathcal{F}' \end{cases}$$

fake

Evaluation: Pointing game

q: *Chicago is the third most populous city in ____.* ?

Facts	Relevance method1
Chicago city.in.state the USA	0.5
Chicago is the most populous city in Illinois.	0.6
New York, Los Angeles and Chicago are the most populous American cities.	0.2
Chicago birthplace Honolulu	0.0
Chicago enrolled in Harward Law School.	0.3
Chicago president.of.state the USA	0.4

real

fake

+1 point

-1

Evaluation: User study

Better "explainer" should pt users to better sys.

Question: And Jacob came into ________.

Answer: Egypt

Which list of facts explains the answer to the query better: facts on the left or facts on the right?

QA model1

Left

- Now Jacob awaked out from Egypt.
- So Jacob went down to Egypt.
- Then Jacob went into Egypt.
- Jacob had to serve through Esau.
- And Jacob went into Egypt.

QA model2

Right

- Jacob people.deceased Ibrahimi Mosque
- Jacob people.deceased Egypt
- Jacob people.marriage.spouse Bilhah
- Jacob people.marriage.spouse Leah
- Jacob people.marriage.spouse Rachel

○ Definitely **left** ○ Rather **left** ○ Difficult to say ○ Rather **right** ○ Definitely **right**

Conclusions

- QA over heterogeneous sources is the way forward !!
- KGs are clean but inherently incomplete , reasoning
- Text corpora are noisy but have more recent information coverage and redundancy
- Graph based methods are powerful models for handling heterogeneous information → KG+Text HIN
- "Explainability" is important but often overlooked