Complex Temporal Question Answering on Knowledge Graphs

Zhen Jia¹, Soumajit Pramanik², Rishiraj Saha Roy³, Gerhard Weikum³

1. Southwest Jiaotong University, 2. IIT Bhilai, 3. Max Planck Institute for Informatics
Temporal questions

- **When** was Obama born?

- Where did Obama live **in 2001**?

- What position was held by Obama **during 9/11**?

- Where did Obama's children study **when he became president**?
Challenges

★ Explicit, implicit or ordinal temporal constraints
★ Multi-hop constraints
★ Identify and reason on time intervals

Where did Obama’s children study when he became president?

- Univ. of Chicago Laboratory Schools
- Sidwell Friends School
- Univ. of Michigan
- Univ. of Harvard

Reason on the time interval of the presidency and the study period
Related Work

★ **Rule-based** framework using question decomposition
  Jia et al. 2018

★ **Benchmark** containing event-centric questions
  Costa et al. 2020

★ Model focusing on *implicit temporal constraints*
  Wu et al. 2020

★ Tool plugging **temporal layer** into existing QA system
  Saquete et al. 2009

★ KG embeddings-based model on **Temporal KGs**
  Saxena et al. 2021
Contributions in EXAQT

★ EXplainable Answering of complex Questions with Temporal intent

⏰ End-to-end system for answering complex temporal questions over KGs

⏰ Fine-tuned BERT models to identify relevant KG facts

⏰ Graph algorithms to compute compact question subgraphs

⏰ Relational graph convolutional networks (R-GCNs) to predict answers with time-enhanced mechanisms

★ TimeQuestions: Benchmark with various types of temporal intents
Temporal question

★ A temporal question is one that contains a temporal expression or a temporal signal, or whose answer is of temporal nature.

<table>
<thead>
<tr>
<th>Category</th>
<th>Question</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXPLICIT</td>
<td>Which movie did Jaco Van Dormael direct in 2009?</td>
<td>OVERLAP</td>
</tr>
<tr>
<td>IMPLICIT</td>
<td>What club did Cristiano Ronaldo play for after Manchester United?</td>
<td>AFTER</td>
</tr>
<tr>
<td></td>
<td>What did Thomas Jefferson do before he was president?</td>
<td>BEFORE</td>
</tr>
<tr>
<td>ORDINAL</td>
<td>What was the first film Julie Andrews starred in?</td>
<td>ORDINAL</td>
</tr>
<tr>
<td>TEMP. ANS.</td>
<td>What year did Lakers win their first championship?</td>
<td>ORDINAL</td>
</tr>
</tbody>
</table>
Temporal fact

★ Main object is a timestamp

![Diagram showing Malia Obama's date of birth: 04-07-1998]

★ Any of the qualifier objects is a timestamp

![Diagram showing Malia Obama's education at Sidwell Friends School: start: 05-01-2009]
Temporal predicate

★ Main object is a timestamp

Binary

Malia Obama \(\rightarrow\) date of birth \(\rightarrow\) 04-07-1998

★ Any of the qualifier objects is a timestamp

N-ary

Malia Obama \(\rightarrow\) educated at \(\rightarrow\) Sidwell Friends School \(\rightarrow\) start \(\rightarrow\) 05-01-2009
Where did Obama’s children study when he became president?

**Approach outline**

⭐ Two-stage approach

**Graph construction**

- Find question-relevant KG facts
- Compute and complete compact subgraphs
- Augment subgraph with temporal facts

**Answer prediction**

- Learn time-aware entity embeddings
- Add temporal category, signal and time encodings
- Integrate attention over temporal predicates

**Output:** Ranked answers

---

**Input:**
- Temporal question
- Knowledge graph

**Recall-oriented**

**Precision-oriented**
Approach outline

★ Two-stage approach

Where did Obama's children study when he became president?

Input: Temporal question + Knowledge graph

Graph construction

- Find question-relevant KG facts
- Compute and complete compact subgraphs
- Augment subgraph with temporal facts

Answer prediction

- Learn time-aware entity embeddings
- Add temporal category, signal and time encodings
- Integrate attention over temporal predicates

Recall-oriented

Precision-oriented

Output: Ranked answers
Stage one: Graph construction

Question: Where did Obama’s children study when he became president?

Use multiple NERD methods to boost answer recall

Find question-relevant KG facts
Compute and complete compact subgraphs
Augment subgraph with temporal facts
Stage one: Graph construction

Question: Where did Obama’s children study when he became president?

Retrieve all KG facts of the question entities.

Find question-relevant KG facts
Compute and complete compact subgraphs
Augment subgraph with temporal facts

zone in KG to start looking for answer
Stage one: Graph construction

Question: Where did Obama’s children study when he became president?

Use distant supervision mechanism to label training set

- <question, fact> pair
- Positive and negative sample
- 1 : 5 ratio

Fine tune BERT model to find relevant KG facts

Barack Obama

President of the US

Malia Obama educated at Sidwell Friends School and start 2009-01-05
Stage one: Graph construction

Question: Where did Obama’s children study when he became president?

Fine-tune BERT model to find relevant KG facts

★ Fine-tune BERT model as a sentence classifier
Stage one: Graph construction

Question: Where did Obama’s children study when he became president?

fine tune BERT model to find relevant KG facts

★ Apply the classifier to sort facts

Rank 1 Barack Obama educated at Harvard Law School and end time ...
Rank 2 Barack Obama educated at State Elementary School Menteng 01 ...
Rank 3 Barack Obama educated at Punahou School start time 1971-01-01 ...
... 
Rank n
Stage one: Graph construction

Question: Where did Obama’s children study when he became president?

Inject connectivity

Diagram:
- Barack Obama
- President of the US
- Find question-relevant KG facts
- Compute and complete compact subgraphs
- Augment subgraph with temporal facts
Stage one: Graph construction

Question: Where did Obama’s children study when he became president?

- Compute **shortest path** between pair of entities
- Add the path with the **highest similarity** to answer graph
  - Get embeddings from BERT
  - Compute cosine similarity
Stage one: Graph construction

Question: Where did Obama’s children study when he became president?

Compute compact subgraph

★ Group Steiner Trees (GSTs)

✔️ Given
Undirected and weighted graph
a subset of terminals in groups

✔️ Find
Minimum cost tree containing at least one terminal from each group
Stage one: Graph construction

Question: Where did Obama’s children study when he became president?

Compute compact subgraph

★ Group Steiner Trees (GSTs)

Given
Undirected and weighted graph
a subset of terminals in groups

Find
Minimum cost tree containing at least one terminal from each group
Stage one: Graph construction

Question: Where did Obama’s children study when he became president?

Compute compact subgraph

★ Group Steiner Trees (GSTs)

Given
Undirected and weighted graph
a subset of terminals in groups

Find
Minimum cost tree containing at least one terminal from each group
Stage one: Graph construction

Question: Where did Obama’s children study when he became president?

Compute compact subgraph

★ Group Steiner Trees (GSTs)

Cost = 1 - score assigned by the classifier of BERT model

Method from Ding et al. 2007

Find question-relevant KG facts

Compute and complete compact subgraphs

Augment subgraph with temporal facts
Stage one: Graph construction

Question: Where did Obama’s children study when he became president?

Complete GST compact subgraphs

Facts linked with the predicate terminals in GSTs
Stage one: Graph construction

Question: Where did Obama’s children study when he became president?

Barack Obama
- start: 20-01-2009
- end: 20-01-2017
- position: President of the US
- Punahou School
  - start: 1971
  - end: 1979
- Harvard Law School
  - start: 1988
  - end: 1991
- Sasha Obama
  - child of: Barack Obama
  - date of birth: 04-07-1998
- Malia Obama
  - child of: Barack Obama
  - date of birth: 04-07-1998
- eduction at Harvard University
  - start: 2017
- eduction at Sidwell Friends School
  - start: 2017
  - end: 10-06-2016
Where did Obama’s children study when he became president?

**Input:**
Temporal question + Knowledge graph

**Graph construction**
- Find question-relevant KG facts
- Compute and complete compact subgraphs
- Augment subgraph with temporal facts

**Answer prediction**
- Learn time-aware entity embeddings
- Add temporal category, signal and time encodings
- Integrate attention over temporal predicates

**Output:**
Ranked answers
Stage two: Answer prediction

★ Build upon the KG-only setting of GRAFT-NET (Sun et al. 2018)
★ R-GCN model with multi-pronged mechanisms for temporal QA

- Temporal Category Encoding (TCE)
- Temporal Signal Encoding (TSE)
- Time Encoding (TE)
- Time-aware Entity Embedding (TEE)
- Attention over Temporal Relation (ATR)
Stage two: Answer prediction
Stage two: Answer prediction

Question initialization

A: Where did Obama’s children study when he became president?

Category encoding

TC: Implicit

TS: Overlap

Signal encoding

Question embedding initialization

★ Temporal category encoding

Label categories
Temporal expression recognition, entity recognition, keywords and POS patterns
Multi-hot encoding

★ Temporal signal encoding

Label signals
dictionary of keywords
Multi-hot encoding
Stage two: Answer prediction

Q: Where did Obama’s children study when he became president?

Use LSTM to model the words in the question as a sequence

Pre-trained word embeddings from Wikipedia2Vec (Yamada et al. 2020)
Stage two: Answer prediction

![Diagram of the answer prediction process]

**Question initialization**

**Question:** Where did Obama’s children study when he became president?

**Initialization Diagram:**
- Category encoding (TC: Implicit)
- Signal encoding (TS: Overlap)
- Concatenate
- LSTM
- Dense layer

**Equations:**

\[ h_q^0 = FFN(TCE(q) \oplus TSE(q) \oplus LSTM(w_1, \ldots, w_{|q|})) \]

\[ h_q^l = FFN(\sum_{e \in NERD(q)} h_q^{l-1}) \]

**Update with the embeddings of entities**

**Question embedding initialization**

**Question embedding update**
Stage two: Answer prediction

$$h^0_e = x_e$$

fixed-size pre-trained embeddings from Wikipedia2Vec

★ Wikipedia2Vec
Yamada et al. 2020

Word-based skip-gram model
Aristotle was a philosopher
The neighboring words of each word are used as contexts

Anchor context model
Aristotle was a philosopher
The neighboring words of a hyperlink pointing to an entity are used as contexts

Link graph model
The neighboring entities of each entity in Wikipedia's link graph are used as contexts
Stage two: Answer prediction

★ Time encoding (similar to position encoding in Vaswani et al. 2017)

- Sinusoidal position encoding
- Provide an unique encoding
- Ensure sequential ordering

\[
TE(k, j) = \begin{cases} 
\sin(k/10000^{\frac{2i}{d}}), & \text{if } j = 2i \\
\cos(k/10000^{\frac{2i}{d}}), & \text{if } j = 2i + 1 
\end{cases}
\]

Entity embedding update
Stage two: Answer prediction

★ Time-aware entity embedding

⦿ An entity $e$ is associated with a set of temporal facts $\{tf(e)\}$
⦿ The temporal facts of $e$ are ordered in a time sequence $\{tf_1(e), tf_2(e), \ldots\}$

Entity embedding update
Stage two: Answer prediction

★ Time-aware entity embedding

Encode tf(e)

Entity embedding update
Stage two: Answer prediction

★ Time-aware entity embedding

Use LSTM to model \{tf_1(e), tf_2(e), \ldots\} as a sequence

\[ h_{TEE(e)}^0 = \text{LSTM}(h_{tf_1}^0, h_{tf_2}^0, \ldots, h_{tf_n}^0) \]
Stage two: Answer prediction

★ Attention over temporal relation

فالع النهاية أن الفصول ذات النطاق نفسه ولكن لهما تمتلكات مختلفة

$$ATR(e, r) = \text{softmax} (x_r \oplus \text{TE}(ts_r)^T h_q^{(l-1)})$$

Entity embedding update
Stage two: Answer prediction

Entity embedding update rule

$$h^l_e = FFN \begin{bmatrix} h^{l-1}_e \\ h^{l-1}_q \\ h^{l-1}_{TEE(e)} \\ \sum_r \sum_{e \in \text{enb}_r(e')} (ATR(e', r) \psi_r(h^{l-1}_e)) \end{bmatrix}$$

(1) Entity representation
(2) Question representation
(3) Time-aware entity representation
(4) Aggregate the states from neighbors
Stage two: Answer prediction

\[ Pr(e \in \{a\}_q \mid RG_q, q) = \sigma(w^T h^l_e + b) \]
Experiment results: Setup

★ Benchmark
   🕒 TimeQuestions

★ Metrics
   🕒 Precision@1
   🕒 Mean Reciprocal Rank
   🕒 Hit@5

★ Baselines
   🕒 UNIQORN (Pramanik et al. 2021)
   🕒 GRAFT-Net (Sun et al. 2018)
   🕒 PullNet (Sun et al. 2019)
Experiment results: Benchmark

★ Benchmark construction

- Collect temporal questions from 8 popular KG-QA benchmarks
- Contain 16181 <question, answer> pairs
- Label temporal categories and signals for each question
- Link answers to Wikidata and Wikipedia
Experiment results: Benchmark

Distribution of question categories by source

- **ORDINAL**
- **TEMP.ANS**
- **IMPLICIT**
- **EXPLICIT**

The diagram shows the distribution of question categories by source, with categories such as Free917, WebQuestions, ComplexQuestions, GraphQuestions, ComplexWebQuestions, ComQA, LC-QuAD, and LC-QuAD 2.0.
Approach outline

★ Two-stage approach

Where did Obama’s children study when he became president?

Input:
Temporal question
+ Knowledge graph

Graph construction

Find question-relevant KG facts
Compute and complete compact subgraphs
Augment subgraph with temporal facts

Answer prediction

Learn time-aware entity embeddings
Add temporal category, signal and time encodings
Integrate attention over temporal predicates

Recall-oriented

Precision-oriented

Output:
Ranked answers
Experiment results: Performance

★ Parameter tuning (S1)

Num-facts \(\text{top-f} = 25\)

Num-gsts \(\text{top-g} = 25\)

Num-temp. facts \(\text{top-t} = 25\)
# Experiment results: Performance

<table>
<thead>
<tr>
<th>Step in EXAQT pipeline</th>
<th>Recall</th>
<th>#Candidates</th>
</tr>
</thead>
<tbody>
<tr>
<td>All KG facts of NERD entities</td>
<td>0.758</td>
<td>2491</td>
</tr>
<tr>
<td>Facts selected by BERT</td>
<td>0.719</td>
<td>48</td>
</tr>
<tr>
<td><strong>Shortest paths</strong> injected for connectivity</td>
<td>0.720</td>
<td>49</td>
</tr>
<tr>
<td>GSTs on largest component</td>
<td>0.613</td>
<td>13</td>
</tr>
<tr>
<td>Union of GSTs from all components</td>
<td>0.640</td>
<td>14</td>
</tr>
<tr>
<td>Completed GSTs from all components</td>
<td>0.671</td>
<td>21</td>
</tr>
<tr>
<td>Temporal facts added by BERT</td>
<td>0.724</td>
<td>67</td>
</tr>
</tbody>
</table>

Understanding the recall-oriented stage one
Experiment results: Performance

EXAQT outperforms others for MRR
Experiment results: Performance

EXAQT outperforms others for P@1
Experiment results: Performance

EXAQT outperforms others for Hit@5
Experiment results: Performance

EXAQT outperforms others in all categories

EXAQT outperforms others in all categories

EXAQT outperforms others in all categories
## Experiment results: Performance

<table>
<thead>
<tr>
<th>Category</th>
<th>Overall</th>
<th>EXPLICIT</th>
<th>IMPLICIT</th>
<th>TEMP. ANS.</th>
<th>ORDINAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXAQT (Full)</td>
<td>0.565</td>
<td>0.568</td>
<td>0.508</td>
<td>0.623</td>
<td>0.420</td>
</tr>
<tr>
<td>EXAQT without TCE</td>
<td>0.545</td>
<td>0.556</td>
<td>0.481</td>
<td>0.590</td>
<td>0.406</td>
</tr>
<tr>
<td>EXAQT without TSE</td>
<td>0.543</td>
<td>0.545</td>
<td>0.465</td>
<td>0.598</td>
<td>0.411</td>
</tr>
<tr>
<td>EXAQT without TEE</td>
<td>0.556</td>
<td>0.564</td>
<td>0.475</td>
<td>0.614</td>
<td>0.413</td>
</tr>
<tr>
<td>EXAQT without TE</td>
<td>0.553</td>
<td>0.556</td>
<td>0.495</td>
<td>0.613</td>
<td>0.398</td>
</tr>
<tr>
<td>EXAQT without ATR</td>
<td>0.534</td>
<td>0.527</td>
<td>0.465</td>
<td>0.594</td>
<td>0.411</td>
</tr>
</tbody>
</table>

### Understanding the precision-oriented Stage two
Conclusion

★★ EXAQT

① Two-stage approach for explainable answering of temporal questions over KGs
① Explainability comes from GSTs, attention and graph visualizations
① Combination of BERT classifiers, GSTs and R-GCNs
① Methods for augmenting components with temporal features

★★ TimeQuestions: benchmark with over 16k temporal questions

Benchmark and demo: https://exaqt.mpi-inf.mpg.de
Code: https://github.com/zhenjia2017/EXAQT

Thank you!