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 Harvard
- Univ. of Michigan
- Obama's presidency

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

 рол End-to-end system for answering complex temporal questions over KGs
 рол Fine-tuned BERT models to identify relevant KG facts
 рол Graph algorithms to compute compact question subgraphs
 рол Relational graph convolutional networks (R-GCNs) to predict answers with time-enhanced mechanisms

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

★ Main object is a timestamp

Binary

- Malia Obama
- date of birth
- 04-07-1998

★ Any of the qualifier objects is a timestamp

N-ary

- Malia Obama
- educated at
- Sidwell Friends School
- start
- 05-01-2009
Temporal predicate

★ Main object is a timestamp

Binary

Malia Obama  date of birth  04-07-1998

★ Any of the qualifier objects is a timestamp

N-ary

Malia Obama  educated at  Sidwell Friends School

Qualifier

start  05-01-2009
Temporal questions

★ 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>
Approach outline

★ Two-stage approach

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

Input: Temporal question + Knowledge graph

Recall-oriented

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

Precision-oriented
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

**Two-stage approach**

**Recall-oriented**
- Approach outline

**Precision-oriented**
Stage one: Graph construction

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

Use multiple NERD methods to boost answer recall
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

Barack Obama

President of the US

zone in KG to start looking for answer

large

noisy

12
Stage one: Graph construction

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

Retrieve all KG facts of the question entities.
Stage one: Graph construction

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

Fine tune BERT model to find relevant KG facts

★ Use distant supervision mechanism to label training set

✧ <question, fact> pair
✧ Positive and negative sample
✧ 1 : 5 ratio
Stage one: Graph construction

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

Use **distant supervision mechanism** to label training set

Verbalize fact as a **natural language sentence**

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

Devlin et al. 2019
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

★ A connected graph is needed for GST and R-GCN algorithms
Stage one: Graph construction

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

Inject connectivity

★ Compute shortest path between pair of entities
★ Add the path with the highest similarity to answer graph

☆ Get embeddings from BERT
☆ Compute cosine similarity

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

President of the US
Barack Obama
Stage one: Graph construction

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

Inject connectivity

- 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)

Terminals are nodes of question entities and predicates matched keywords

terminals

Find question-relevant KG facts

Compute and complete compact subgraphs

Augment subgraph with temporal facts

terminals
Stage one: Graph construction

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

Compute compact subgraph

★ Group Steiner Trees (GSTs)

⑦ Terminals in one group

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 compact subgraph

★ Group Steiner Trees (GSTs)

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

⏲ Method from Ding et al. 2007
Stage one: Graph construction

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

Complete GST compact subgraphs
Stage one: Graph construction

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

Punahou School
- educated at
- start: 1971
- end: 1979

Harvard Law School
- educated at
- start: 1988
- end: 1991

Position
- Barack Obama
- start: 20-01-2009
- end: 20-01-2017

President of the US

1971
1979

Sasha Obama
- child
- date of birth: 04-07-1998

Malia Obama
- child

05-01-2009
10-06-2016

Educated at
1971
1979

Sidwell Friends School
- educated at
- start: 10-06-2016
- end: 05-01-2009

Harvard University
- educated at
- start: 2017

Find question-relevant KG facts

Compute and complete compact subgraphs

Augment subgraph with temporal facts
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 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

Question initialization

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

★ 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

**Question Initialization**

- **TC**: Implicit
- **TS**: Overlap

**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

**Question embedding initialization**

**Question embedding update**

Mathematical equations:

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

\[ h^l_q = FFN\left( \sum_{e \in NERD(q)} h^{l-1}_e \right) \]

Update with the embeddings of entities
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

Entity embedding initialization
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**

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

- Position in time range
- Vector dimension
- Position in vector

Entity embedding update
Stage two: Answer prediction

★ Time-aware entity embedding

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

Time-aware Entity Embedding Initialization

- Malia Obama
- educated at
- start
- Sidwell Friends School
- 05-01-2009

Entity embedding update
Stage two: Answer prediction

⭐ Time-aware entity embedding

_decode tf(e)_
Stage two: Answer prediction

★ Time-aware entity embedding

Use LSTM to model $\{\text{tf}_1(e), \text{tf}_2(e), \ldots\}$ as a sequence

$$h_{TEE(e)}^0 = LSTM(h_{tf_1(e)}^0, h_{tf_2(e)}^0, \ldots, h_{tf_n(e)}^0)$$

Entity embedding update
Stage two: Answer prediction

★ Attention over temporal relation

눗 Distinguish entities with the same relation but having different timestamps

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

Entity embedding update
Stage two: Answer prediction

Entity embedding update rule

\[ h_e^l = FFN \left[ \begin{array}{c}
    h_e^{l-1} \\
    h_q^{l-1} \\
    h_{TEE(e)}^{l-1} \\
    \sum_r \sum_{e \in \text{enbd}_r(e')} (ATR(e', r) \psi_r(h_{e'}^{l-1}))
\end{array} \right] \]

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

\[ Pr(e \in \{a\}_q | RG_q, q) = \sigma(w^T h_e^l + b) \]
Stage two: Answer prediction

A. Question Initialization

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

Word embedding

 LSTM

 Concatenate

 TC: Implicit

 TS: Overlap

 Signal encoding

 Category encoding

 Concatenate

 LSTM

 Dense layer

 $h^0_q$

 B. Time-aware Entity Embedding Initialization

 Malia Obama  educated at  start  Sidwell Friends School  05-01-2009

 Entity embedding

 Relation word embedding

 Time encoding

 Concatenate

 LSTM

 $h^0_{TEE(e)}$

 tf(e)

 tf(e)

 C. Entity Update

 Barack Obama  child  Malia Obama

 educated at

 Attention over Temporal Relations

 $ATR(e', r). \Psi_r(h^{l-1}_e)$

 $h^1_q$

 Question Embedding

 $h^1_{TEE(e)}$

 Time-aware Entity Embedding

 Sidwell Friends School

 $h^1_e$
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, answers> 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**

Legend:
- Free917
- WebQuestions
- ComplexQuestions
- GraphQuestions
- ComplexWebQuestions
- ComQA
- LC-QuAD
- LC-QuAD 2.0
Where did Obama’s children study when he became president?

**Two-stage approach**

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

- **top-f = 25**

**Num-gsts**

- **top-g = 25**

**Num-temp. facts**

- **top-t = 25**

---

Parameter tuning (S1)
### 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 <strong>NERD</strong> entities</td>
<td>0.758</td>
<td>2491</td>
</tr>
<tr>
<td>Facts selected by <strong>BERT</strong></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><strong>GSTs</strong> on largest component</td>
<td>0.613</td>
<td>13</td>
</tr>
<tr>
<td><strong>Union of GSTs</strong> from all components</td>
<td>0.640</td>
<td>14</td>
</tr>
<tr>
<td><strong>Completed GSTs</strong> from all components</td>
<td>0.671</td>
<td>21</td>
</tr>
<tr>
<td><strong>Temporal facts</strong> added by BERT</td>
<td>0.724</td>
<td>67</td>
</tr>
</tbody>
</table>

**Understanding the recall-oriented stage one**
### 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 <strong>NERD</strong> entities</td>
<td>0.758</td>
<td>2491</td>
</tr>
<tr>
<td>Facts selected by <strong>BERT</strong></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><strong>GSTs</strong> on largest component</td>
<td>0.613</td>
<td>13</td>
</tr>
<tr>
<td>Union of <strong>GSTs</strong> from all components</td>
<td>0.640</td>
<td>14</td>
</tr>
<tr>
<td>Completed <strong>GSTs</strong> from all components</td>
<td>0.671</td>
<td>21</td>
</tr>
<tr>
<td><strong>Temporal facts</strong> 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

P@1

EXAQT
PullNet
GRAFT-Net
UNIQORN

EXAQT outperforms others for P@1
Experiment results: Performance

EXAQT outperforms others for Hit@5
Experiment results: Performance

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
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
## 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

② Explainability comes from GSTs, attention and graph visualizations

③ Combination of **BERT** classifiers, **GSTs** and **R-GCNs**

④ Methods for augmenting components with **temporal** features

★ **TimeQuestions**: benchmark with over 16k temporal questions

Benchmark and demo: [https://exaqt.mpi-inf.mpg.de](https://exaqt.mpi-inf.mpg.de)
Code: [https://github.com/zhenjia2017/EXAQT](https://github.com/zhenjia2017/EXAQT)
Thank you!

Complex Temporal Question Answering on Knowledge Graphs