Answering Complex Questions by Joining Multi-Document Evidence with Quasi Knowledge Graphs

XiaoluLu, Soumajit Pramanik, Rishiraj Saha Roy,

Abdalghani Abujabal, Yafang Wang, Gerhard Weikum

Xiaofan Yan

Outline

- Abstract
- Introduction
- System overview
- Graph construction
- Graph Algorithm
- Evaluation setup
- Main results and insights
- Analysis and discussion
- Related work
- Conclusion

Abstract

- QA
  - Answering questions which involve multiple entities and relations
  - Curated KG
    - limited by its inherent incompleteness and potential staleness
- QUEST
  - Answering complex questions from textual sources on-the-fly
  - Unsupervised
  - Builds a noisy quasi KG to compute the answers by an algorithm for Group Steiner Trees(GST)

Introduction

- Motivation
  - Web search
    - Queries have the form of questions
      - Providing direct answers
      - Tapped into textual sources
      - Translating questions into KG/KB/DB
    - Challenge: complex questions that refer to multiple entities and multiple relations

Introduction

- State of the art and the limitation
  - QA over KGs
    - Question-to-query is weak and infeasible
    - Depends on KG’s completeness and freshness
    - Queries is tied to language and style
    - Translation procedure is sensitive to choice KG/KB source
  - QA over textual
    - Match significant terms in the same document
    - Syntactic decomposition patterns break with ungrammatical constructs
    - Deep learning method rely on training data critically
    - Recent work consider a specific set of documents contain the answer

Introduction

- Approach and Contribution
  - First construct a KG by retrieving question-relevant text documents and running OpenIE to produce SPO, compute edges connect potentially synonymous names and phrases
  - Then matching nodes as cornerstones as terminal in GST, non-terminal nodes are candidate answers

System overview

- Don’t consider the complex question like requiring grouping, comparison, and aggregation or involving negations
- Two phases:
  - On-the-fly construction of quasi KG
  - Graph algorithm for computing ranked answers

Graph construction

- Extracting SPO Triples
  - Devise own IE method
  - POS tagging and NER
  - Extra SPO triples

Graph construction

- Building the Quasi KG
  - Nodes
    - S,P,O
    - Types of S, O
  - Edges
    - S to P, P to O
    - Equal or high similarity
    - S or O to their types
  - Node weights: similarity score (compute by word2vec/GloVE/BERT)
  - Edge weights:
    - Triple edges: confidence score
    - Alignment edges: similarity
    - Type edges: 1.0

Graph algorithm

- Computing Group Steiner trees
  - Cornerstones
    - Every node matching a word or phrase with high similarity
  - Group Steiner Trees
    - key idea: candidate nodes should be tightly connected to many cornerstones
  - Algorithm
    - NP-complete problems
  - Relaxation to GST-k

Graph algorithm

- Filtering and Ranking Answers
  - Remove all candidates that are not entities (relation and type nodes)
  - Drop candidates do not have types with similarity above a threshold
- Answer aggregation
  - two answers are merged if
    - one is a subsequence
    - there is an alignment edge between them

Evaluation setup

- Rationale for experiments
  - Compiled a new benchmark of complex questions
  - Focus on unsupervised and distantly supervised methods
- Question benchmarks
  - Complex questions from WikiAnswers (CQ-W)
  - Complex questions from Trends (CQ-T)
- Text corpora and quasi KGs
  - Similarities and thresholds
  - Graph-based algorithms
  - Metrics: Mean Reciprocal Rank (MRR)/Precision@1/Hit@5

Main results and insights

Dataset	#Nodes			#Edges		
	Entity	**Relation**	**Type**	**Triple**	**Alignment**	**Type**
CQ-W	501	466	28	1.2k	5.2k	434
CQ-T	472	375	23	1k	13.2k	436

Table 1: Basic properties of quasi KGs, averaged over all questions.

Method	Metric	ComplexQuestions from WikiAnswers (CQ-W)						ComplexQuestions from Trends (CQ-T)					
		Top	**Strata-1**	**Strata-2**	**Strata-3**	**Strata-4**	**Strata-5**	**Top**	**Strata-1**	**Strata-2**	**Strata-3**	**Strata-4**	**Strata-5**
QUEST	MRR	**0.355***	**0.380***	**0.340***	**0.302***	**0.356***	**0.318***	**0.467***	**0.436***	**0.426***	**0.460***	**0.409***	**0.384***
DrQA [12]	MRR	0.226	0.237	0.257	0.256	0.215	0.248	0.355	0.330	0.356	0.369	0.365	0.380
BFS [38]	MRR	0.249	0.256	0.266	0.212	0.219	0.254	0.287	0.256	0.265	0.259	0.219	0.201
ShortestPaths	MRR	0.240	0.261	0.249	0.237	0.259	0.270	0.266	0.224	0.248	0.219	0.232	0.222
QUEST	P@1	**0.268***	**0.315**	**0.262**	**0.216**	**0.258***	**0.216**	**0.394***	**0.360***	**0.347***	**0.377***	**0.333***	**0.288**
DrQA [12]	P@1	0.184	0.199	0.221	0.215	0.172	0.200	0.286	0.267	0.287	0.300	0.287	0.320
BFS [38]	P@1	0.160	0.167	0.193	0.113	0.100	0.147	0.210	0.170	0.180	0.180	0.140	0.130
ShortestPaths	P@1	0.147	0.173	0.193	0.140	0.147	0.187	0.190	0.140	0.160	0.160	0.150	0.130
QUEST	Hit@5	**0.376**	**0.396**	**0.356**	**0.344**	**0.401**	**0.358**	**0.531**	**0.496***	**0.510**	**0.500**	**0.503**	**0.459**
DrQA [12]	Hit@5	0.313	0.315	0.322	0.322	0.303	0.340	0.453	0.440	0.473	0.487	0.480	0.480
BFS [38]	Hit@5	0.360	0.353	0.347	0.327	0.327	0.360	0.380	0.360	0.370	0.360	0.310	0.320
ShortestPaths	Hit@5	0.347	0.367	0.387	0.327	0.393	0.340	0.350	0.320	0.340	0.310	0.330	0.290

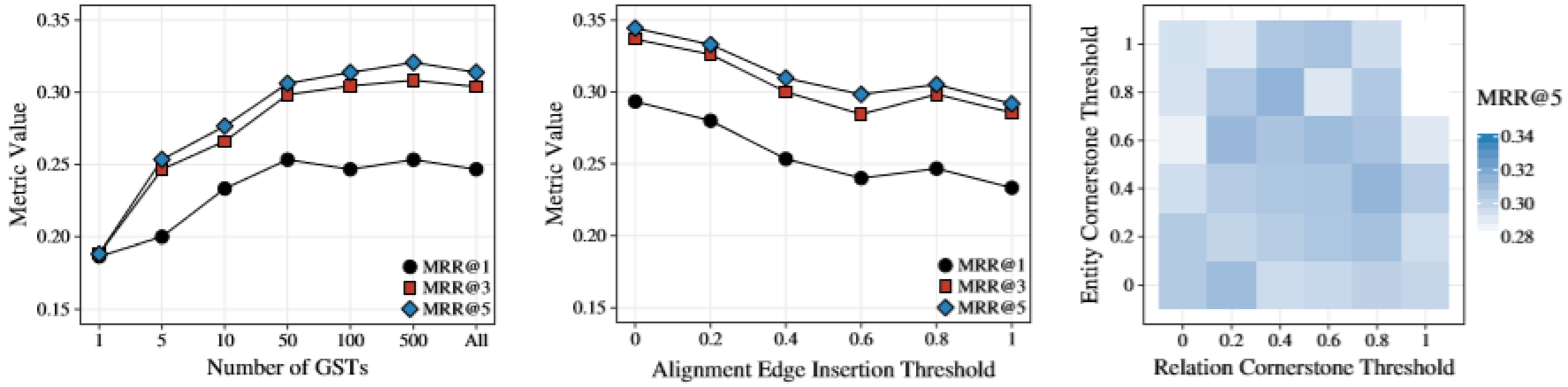
Table 2: Performance comparison of methods on top-10 and stratified search results from the Web. For every metric, the best value per column is in **bold**. "*" denotes statistical significance of QUEST over DrQA, with p-value ≤ 0.05 for a one-tailed paired t-test.

Main results and insights

Benchmark	CQ-W		CQ-T	
GST Ranks	Avg. #Docs in GST	#Q's with A in GST	Avg. #Docs in GST	#Q's with A in GST
01 − 10	2.637	48	3.139	52
11 − 20	2.789	56	3.156	53
21 − 30	2.778	54	3.245	55
31 − 40	2.833	54	3.267	54
41 − 50	2.882	57	3.278	51
#Docs in GST	Avg. Rank of GST	#Q's with A in GST	Avg. Rank of GST	#Q's with A in GST
1	24.525	12	22.124	7
2	27.003	37	24.216	23
3	25.777	53	27.165	50
4	27.064	36	27.069	49
5	29.291	25	26.554	29

Table 3: Effect of multi-document evidence shown via edge contributions by distinct documents to GSTs (on top-10 corpora).

Main results and insights



(a) No. of GSTs (k) **(b)** Alignment edge insertion threshold **(c)** Cornerstone selection thresholds

Figure 2: Robustness of QUEST to various system configuration parameters, on CQ-W with top-10 corpora (similar trends on CQ-T).

Analysis and discussion

- Graph ablation experiments
  - Type nodes and edges are essential
  - informative edge weights and alignment levels
  - Alignment edges
- Answer ranking variants
  - Weighted (with reciprocal of tree cost) sum

Graph configuration	CQ-W	CQ-T	Answer ranking criterion	CQ-W	CQ-T	Error scenario	CQ-W	CQ-T
Full configuration	**0.355**	**0.467**	Wted. sum of GSTs (inv. tree cost sum)	**0.355**	**0.467**	Ans. not in corpus	1%	7%
No types	0.321	0.384*	Wted. sum of GSTs (node wt. sum)	0.318*	0.426*	Ans. in corpus but not in quasi KG	23%	30%
Degenerate edge weights	0.282*	0.377*	Count of GSTs	0.334	0.432*	Ans. in quasi KG but not in top-50 GSTs	10%	6%
No entity alignment	0.329*	0.413*	Wted. dist. to cornerstones	0.321*	0.417*	Ans. in top-50 GSTs but not in candidates	1%	7%
No predicate alignment	0.337	0.403*	Unwted. dist. to cornerstones	0.257*	0.372*	Ans. in candidates but not in top-5	66%	49%

Table 5: Understanding QUEST's mechanism with top-10 corpora. **Left:** Graph ablation (MRR); **Middle:** Answer ranking (MRR); **Right:** Error analysis (Hit@5 = 0). Highest column values in the first two sections in **bold**. Significant drops from these values are shown with *.

Analysis and discussion

- Error analysis
- Effect of OpenIE
- Answering simple questions
- Effect of threshold variations
  - Number of GSTs k,the alignment edge insertion threshold on node-node similarity, and cornerstone selection thresholds on node weight
- Run-time

Related work

- QA over text
- QA over KGs
- QA over hybrid sources
- Reading comprehension

Conclusion

- Presented QUEST, an unsupervised method for QA over dynamically retrieved text corpora based on Group Steiner Trees.