

Complex Sequential Question Answering: Towards Learning to Converse Over Linked Question Answer Pairs with a Knowledge Graph

Amrita Saha¹

amrsaha4@in.ibm.com

Vardaan Pahuja^{3*}

vardaanpahuja@gmail.com

Mitesh M. Khapra²

miteshk@cse.iitm.ac.in

Karthik Sankaranarayanan¹

kartsank@in.ibm.com

Sarath Chandar³

apsarathchandar@gmail.com

¹IBM Research AI

²Indian Institute of Technology Madras, India,

³MILA, Universite de Montreal



*Work done while at IBM Research AI

Outline




- Introduce a new dataset for Complex Sequential QA over a large-scale Knowledge Base
- Motivate the need for such a dataset
- Highlight the poor performance of the state-of-the-art models on this new dataset
- Encourage the research community to develop models for such complex QA tasks

Existing KB based QA/Conversation Datasets




- Restaurant reservation (Bordes and Weston 2016)

-  • Size of KB (Knowledge Base) is toyish (< 10 cuisines, locations, ambience, etc.)
-  • Very few states in dialog

- SimpleQuestions Dataset (Bordes 2015)

-  • Q/A over a Large KB of millions of entities
-  • Consisting of only simple questions requiring single tuple lookup in the KB
-  • Not in a dialog setting

- Sequential Question Answering (SQA, 2016)

-  • Complex QA pairs are linked as in a dialog
-  • Q/A over small tables and not a KB
-  • Only 17 K questions

... and a few other datasets

Wishlist for a new KB based Sequential QA Dataset

- KB based Challenges
 - Need for a realistic scale Knowledge Base (of at least few millions of entities)
 - Go beyond simple questions, which are answerable from a single KB tuple, to more complex questions
 - Need for sequences of different inferencing (logical/comparative/quantitative) over larger subgraphs of the KB
- Conversational Challenges
 - Use conversation context to resolve co-references and ellipsis in utterances
 - Ask for clarifications for ambiguous queries

Wishlist for a new KB based Sequential QA Dataset

- KB based Ch

- Need for a (at least few millions of entities)
- Go beyond simple queries (able from a single KB tuple, to more complex queries)
- Need for simple queries (syntactical/comparative/quantitative) over larger subgraphs of the KB



.... *In our new dataset on Complex Sequential Question Answering over KB (CSQA)*

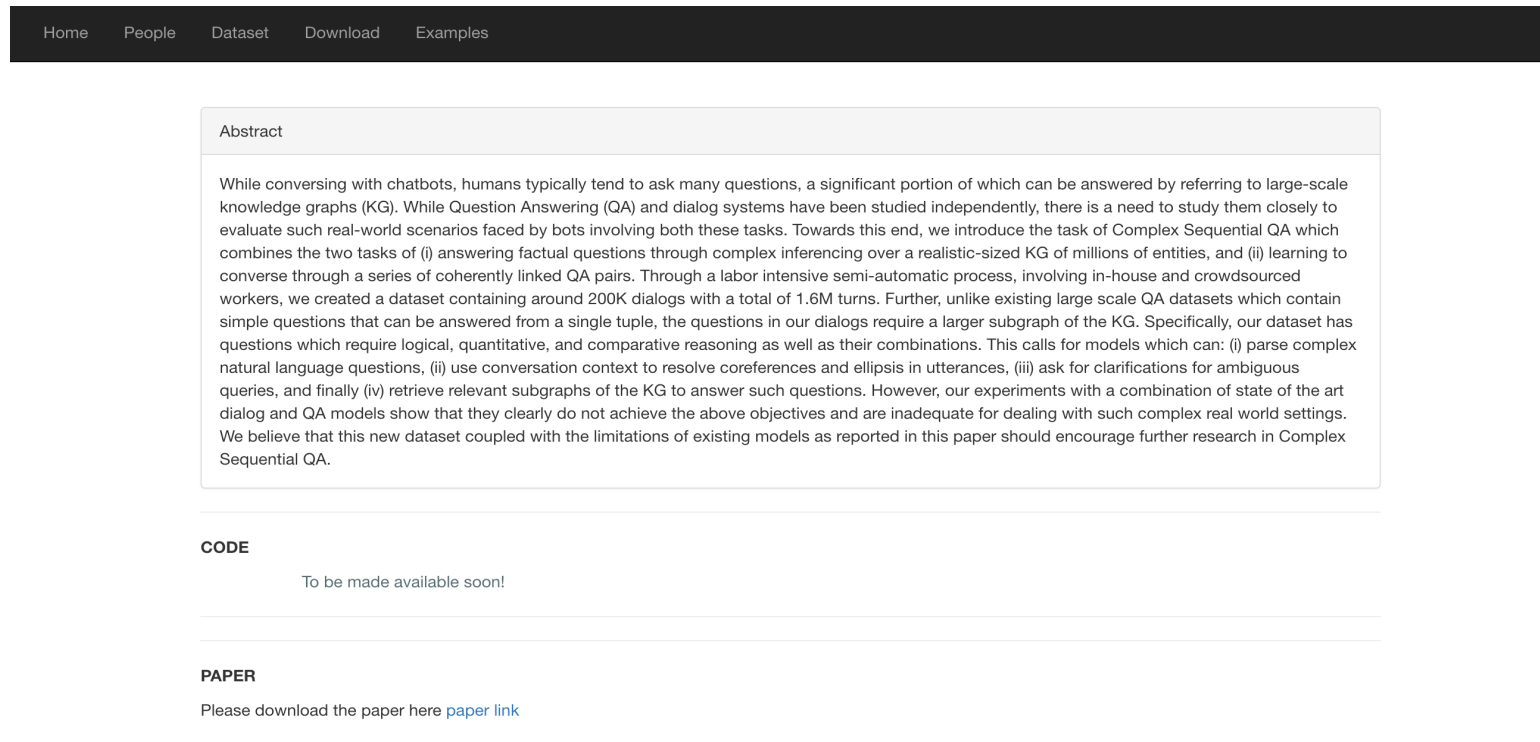
- Ask for clarifications for ambiguous queries

Highlights of the CSQA Dataset

- Question Answering is done over WikiData, an open-domain KB, having 13 Million entities and 21 Million facts
- With the help of domain experts, we designed 19 dialog states each comprising of simple or complex types of questions answerable from subgraphs of the KB
- Further designed an automata over the dialog states to create non goal-oriented dialogs
- Instantiated the automata to create 200K such dialogs with a total of 1.6 M QA turns.

Link to Download Dataset

- Dataset is available at <https://amritasaha1812.github.io/CSQA/>



The screenshot shows a website with a dark navigation bar containing links for Home, People, Dataset, Download, and Examples. Below the navigation bar is a white content area with a light gray header for the 'Abstract' section. The abstract text describes the CSQA dataset, its creation process, and its characteristics. Below the abstract, there is a 'CODE' section with the text 'To be made available soon!'. At the bottom, there is a 'PAPER' section with the text 'Please download the paper here [paper link](#)'.

Home People Dataset Download Examples

Abstract

While conversing with chatbots, humans typically tend to ask many questions, a significant portion of which can be answered by referring to large-scale knowledge graphs (KG). While Question Answering (QA) and dialog systems have been studied independently, there is a need to study them closely to evaluate such real-world scenarios faced by bots involving both these tasks. Towards this end, we introduce the task of Complex Sequential QA which combines the two tasks of (i) answering factual questions through complex inferencing over a realistic-sized KG of millions of entities, and (ii) learning to converse through a series of coherently linked QA pairs. Through a labor intensive semi-automatic process, involving in-house and crowdsourced workers, we created a dataset containing around 200K dialogs with a total of 1.6M turns. Further, unlike existing large scale QA datasets which contain simple questions that can be answered from a single tuple, the questions in our dialogs require a larger subgraph of the KG. Specifically, our dataset has questions which require logical, quantitative, and comparative reasoning as well as their combinations. This calls for models which can: (i) parse complex natural language questions, (ii) use conversation context to resolve coreferences and ellipsis in utterances, (iii) ask for clarifications for ambiguous queries, and finally (iv) retrieve relevant subgraphs of the KG to answer such questions. However, our experiments with a combination of state of the art dialog and QA models show that they clearly do not achieve the above objectives and are inadequate for dealing with such complex real world settings. We believe that this new dataset coupled with the limitations of existing models as reported in this paper should encourage further research in Complex Sequential QA.

CODE

To be made available soon!

PAPER

Please download the paper here [paper link](#)

Updated version of the paper and results are in <https://arxiv.org/abs/1801.10314>

Challenges in Complex Sequential QA over KB

User:	Can you tell me which cities border Verderio Inferiore?
System:	Cornate d'Adda, Bernareggio, Robbiate
User:	And which cities flank that one ?
System:	Did you mean Robbiate ?
User:	No, I meant Cornate d'Adda
System:	Bottanuco, Busnago, Trezzo sull'Adda
User:	Which ones share border with Robbiate and are not part of Azerbaijan ?
System:	Trezzo sull'Adda, Busnago, Bottanuco
User:	Which city is the capital of minimum number of french administrative divisions ?
System:	Riga
User:	How many political territories have strategic relation with around 22 other countries ?
System:	11
User:	Which countries have lesser number of political relations with others than Australia ?
System:	Denmark, Canada, Grenada

User:	Can you tell me which cities border Verderio Inferiore?	<i>Simple Question</i>
System:	Cornate d'Adda, Bernareggio, Robbiate	
User:	And which cities flank that one ?	
System:	Did you mean Robbiate ?	
User:	No, I meant Cornate d'Adda	
System:	Bottanuco, Busnago, Trezzo sull'Adda	
User:	Which ones share border with Robbiate and are not part of Azerbaijan ?	
System:	Trezzo sull'Adda, Busnago, Bottanuco	
User:	Which city is the capital of minimum number of french administrative divisions ?	
System:	Riga	
User:	How many political territories have strategic relation with around 22 other countries ?	
System:	11	
User:	Which countries have lesser number of political relations with others than Australia ?	
System:	Denmark, Canada, Grenada	

User:	Can you tell me which cities border Verderio Inferiore?
System:	Cornate d'Adda, Bernareggio, Robbiate
User:	And which cities flank that one ?
System:	Did you mean Robbiate ?
User:	No, I meant Cornate d'Adda
System:	Bottanuco, Busnago, Trezzo sull'Adda
User:	Which ones share border with Robbiate and are not part of Azerbaijan ?
System:	Trezzo sull'Adda, Busnago, Bottanuco
User:	Which city is the capital of minimum number of french administrative divisions ?
System:	Riga
User:	How many political territories have strategic relation with around 22 other countries ?
System:	11
User:	Which countries have lesser number of political relations with others than Australia ?
System:	Denmark, Canada, Grenada

Ellipses and coreference resolution required

Clarification required for ambiguous question

User:	Can you tell me which cities border Verderio Inferiore?
System:	Cornate d'Adda, Bernareggio, Robbiate
User:	And which cities flank that one ?
System:	Did you mean Robbiate ?
User:	No, I meant Cornate d'Adda
System:	Bottanuco, Busnago, Trezzo sull'Adda
User:	Which ones share border with Robbiate and are not part of Azerbaijan ?
System:	Trezzo sull'Adda. Busnago. Bottanuco
User:	Which city is the capital of minimum number of french administrative divisions ?
System:	Riga
User:	How many political territories have strategic relation with around 22 other countries ?
System:	11
User:	Which countries have lesser number of political relations with others than Australia ?
System:	Denmark, Canada, Grenada

Requires Logical inferencing e.g. set difference operation over two subgraphs of the KB

User:	Can you tell me which cities border Verderio Inferiore?
System:	Cornate d'Adda, Bernareggio, Robbiate
User:	And which cities flank that one ?
System:	Did you mean Robbiate ?
User:	No, I meant Cornate d'Adda
System:	Bottanuco, Busnago, Trezzo sull'Adda
User:	Which ones share border with Robbiate and are not part of Azerbaijan ?
System:	Trezzo sull'Adda, Busnago, Bottanuco
User:	Which city is the capital of minimum number of french administrative divisions ?
System:	Riga
User:	How many political territories have strategic relation with around 22 other countries ?
System:	11
User:	Which countries have lesser number of political relations with others than Australia ?
System:	Denmark, Canada, Grenada

***Quantitative Reasoning
(e.g. minimum) over a
collection of subgraphs
of the KB***

User:	Can you tell me which cities border Verderio Inferiore?
System:	Cornate d'Adda, Bernareggio, Robbiate
User:	And which cities flank that one ?
System:	Did you mean Robbiate ?
User:	No, I meant Cornate d'Adda
System:	Bottanuco, Busnago, Trezzo sull'Adda
User:	Which ones share border with Robbiate and are not part of Azerbaijan ?
System:	Trezzo sull'Adda, Busnago, Bottanuco
User:	Which city is the capital of minimum number of french administrative divisions ?
System:	Riga <i>Quantitative reasoning over multiple subgraph of the KB</i>
User:	How many political territories have strategic relation with around 22 other countries ?
System:	11
User:	Which countries have lesser number of political relations with others than Australia ?
System:	Denmark, Canada, Grenada

User:	Can you tell me which cities border Verderio Inferiore?
System:	Cornate d'Adda, Bernareggio, Robbiate
User:	And which cities flank that one ?
System:	Did you mean Robbiate ?
User:	No, I meant Cornate d'Adda
System:	Bottanuco, Busnago, Trezzo sull'Adda
User:	Which ones share border with Robbiate and are not part of Azerbaijan ?
System:	Trezzo sull'Adda, Busnago, Bottanuco
User:	Which city is the capital of minimum number of french administrative divisions ?
System:	Riga
User:	How many political territories have strategic relation with around 22 other countries ?
System:	11 <i>Comparative Reasoning between multiple subgraphs of the KB</i>
User:	Which countries have lesser number of political relations with others than Australia ?
System:	Denmark, Canada, Grenada

State-of-the-art models for KB
based QA

State-of-the-art* performance on CSQA

Question Type	Recall (%)	Precision (%)
Overall	18.4	6.3
Simple Question (Direct)	33.3	8.58
Simple Question (Co-referenced)	12.67	5.09
Simple Question (Ellipsis)	17.3	6.98
Logical Reasoning (All)	15.11	4.75
Quantitative Reasoning (All)	0.913	1.01
Comparative Reasoning (All)	2.11	4.97
Clarification	25.09	12.13
Question Type	Accuracy (%)	
Verification (Boolean) (All)	21.04	
Quantitative Reasoning (Count) (All)	12.13	
Comparing Reasoning (Count) (All)	8.67	

*System used here is a state-of-the-art Neural model based on Key Value Memory Network [Miller et al. 2016] with a hierarchical encoder for encoding a dialog context

State-of-the-art* performance on CSQA

Question Type	Recall (%)	Precision (%)
Overall	18.4	6.3
Simple Question (Direct)	33.3	8.58
Simple Question (Co-referenced)	12.67	5.09
Simple Question (Ellipsis)	17.3	6.98
Logical Reasoning (All)	15.11	4.75
Quantitative Reasoning (All)	0.913	1.01
Comparative Reasoning (All)	2.11	4.97
Clarification	25.09	12.13
Question Type	Accuracy (%)	
Verification (Boolean) (All)	21.04	
Quantitative Reasoning (Count) (All)	12.13	
Comparing Reasoning (Count) (All)	8.67	

Answer is a set of KB entities

Answer is a set of booleans or integers

State-of-the-art* performance on CSQA

Question Type	Recall (%)	Precision (%)
Overall	18.4	6.3
Simple Question (Direct)	33.3	8.58
Simple Question (Co-referenced)	12.67	5.09
Simple Question (Ellipsis)	17.3	6.98
Logical Reasoning (All)	15.11	4.75
Quantitative Reasoning (All)	0.913	1.01
Comparative Reasoning (All)	2.11	4.97
Clarification	25.09	12.13
Question Type	Accuracy (%)	
Verification (Boolean) (All)	21.04	
Quantitative Reasoning (Count) (All)	12.13	
Comparing Reasoning (Count) (All)	8.67	

Questions with co-reference and ellipsis is significantly harder to answer than direct questions

State-of-the-art performance on CSQA

Question Type	Recall (%)	Precision (%)
Overall	18.4	6.3
Simple Question (Direct)	33.3	8.58
Simple Question (Co-referenced)	12.67	5.09
Simple Question (Ellipsis)	17.3	6.98
Logical Reasoning (All)	15.11	4.75
Quantitative Reasoning (All)	0.913	1.01
Comparative Reasoning (All)	2.11	4.97
Clarification	25.09	12.13
Question Type	Accuracy (%)	
Verification (Boolean) (All)	21.04	
Quantitative Reasoning (Count) (All)	12.13	
Comparing Reasoning (Count) (All)	8.67	

State-of-the-art models are not appropriate for modeling complex question answering

State-of-the-art performance on CSQA

Question Type	Recall (%)	Precision (%)
Overall	18.4	6.3
Simple Question (Direct)	33.3	8.58
Simple Question (Co-referenced)	12.67	5.09
Simple Question (Ellipsis)	17.3	6.98
Logical Reasoning (All)	15.11	4.75
Quantitative Reasoning (All)	0.913	1.01
Comparative Reasoning (All)	2.11	4.97
Clarification	25.09	12.13
Question Type	Accuracy (%)	
Verification (Boolean) (All)	21.04	
Quantitative Reasoning (Count) (All)	12.13	
Comparing Reasoning (Count) (All)	8.67	

These models cannot perform quantitative reasoning, they treat integers also as vocabulary words

State-of-the-art performance on CSQA

Question Type	Recall (%)	Precision (%)
Overall	18.4	6.3
Simple Question (Direct)	33.3	8.58
Simple Question (Co-referenced)	12.67	5.09
Simple Question (Ellipsis)	17.3	6.98
Logical Reasoning (All)	15.11	4.75
Quantitative Reasoning (All)	0.913	1.01
Comparative Reasoning (All)	2.11	4.97
Clarification	25.09	12.13
Question Type	Accuracy (%)	
Verification (Boolean) (All)	21.04	
Quantitative Reasoning (Count) (All)	12.13	
Comparing Reasoning (Count) (All)	8.67	

If memory based models are used, complex question answering would require large memory (>100K tuples)

Conclusion

- Introduced a dataset of 200K dialogs of over 1.6M Question-Answer pairs covering 19 different simple and complex Question Types
- Showed how each of the complex questions need sequences of logical, quantitative and comparative reasoning over subgraphs of the million-sized open-domain KB Wikidata
- Highlighted the limitations of current neural models in handling complex QA over such large scale KBs
- With this, we encourage research into learning to converse over complex KB based question-answering

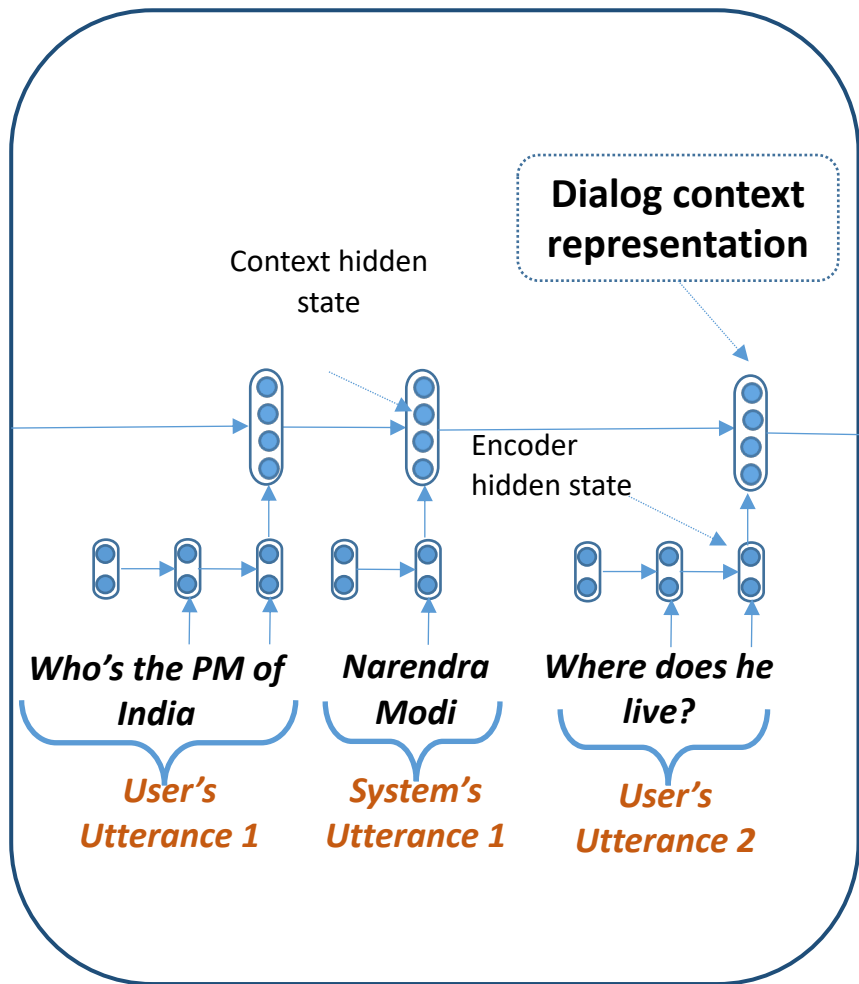
THANK YOU

Backup Slides

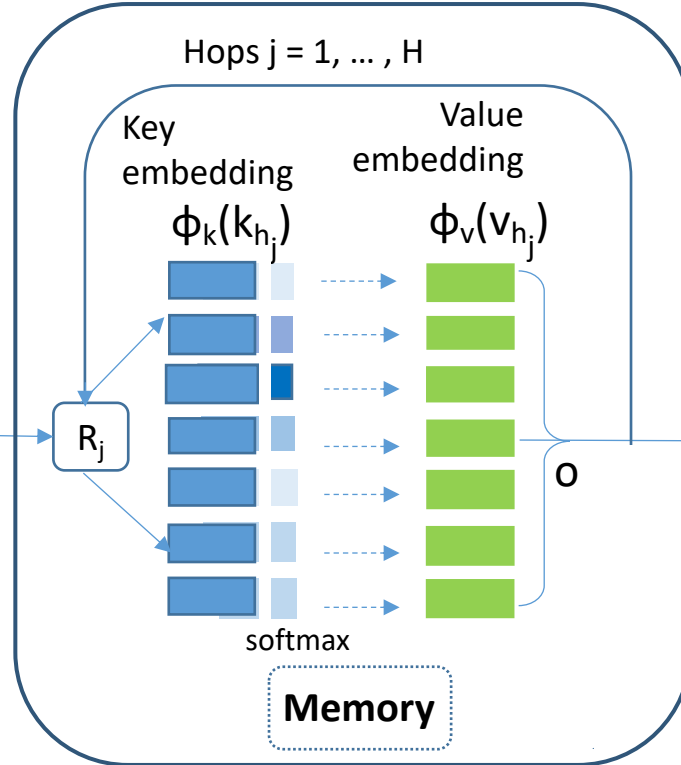
Question Types

Reasoning	Type	Containing	Example
Logical	Union/ Intersection/ Difference	Single Relation	Which rivers flow through India and/or/but not China?
	Any of above	Multiple Relations	Which river flows through India but does not originate in Himalayas?
Verification	Boolean	Single/Multiple Entities	Does Ganga flow through India ?
Quantitative	Count	Single/Multiple Entity Type	How many rivers (and lakes) flow through India ?
		Logical Operators	How many rivers flow through India and/or/but not China?
	Min/Max	Single/Multiple Entity Type	Which country has maximum number of rivers (and lakes)?
	Atleast/ Atmost/ Approx/ Equal	Single/Multiple Entity Type	Which country has at least N rivers (and lakes) ?
	Count over Atleast /Atmost /Approx /Equal	Single/Multiple Entity Type	How many countries have at least N rivers (and lakes)?
Comparative	More/Less	Single/Multiple Entity Type	Which countries have more rivers (and lakes) than India ?
	Count over More/Less	Single/Multiple Entity Type	How many countries have more number of rivers (and lakes) than India ?

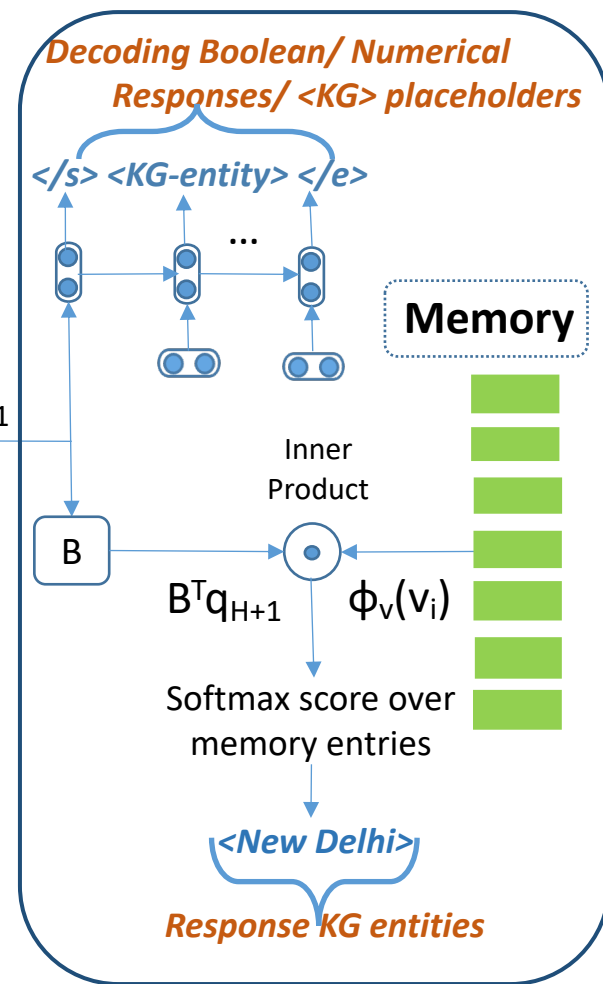
Hierarchical Encoder



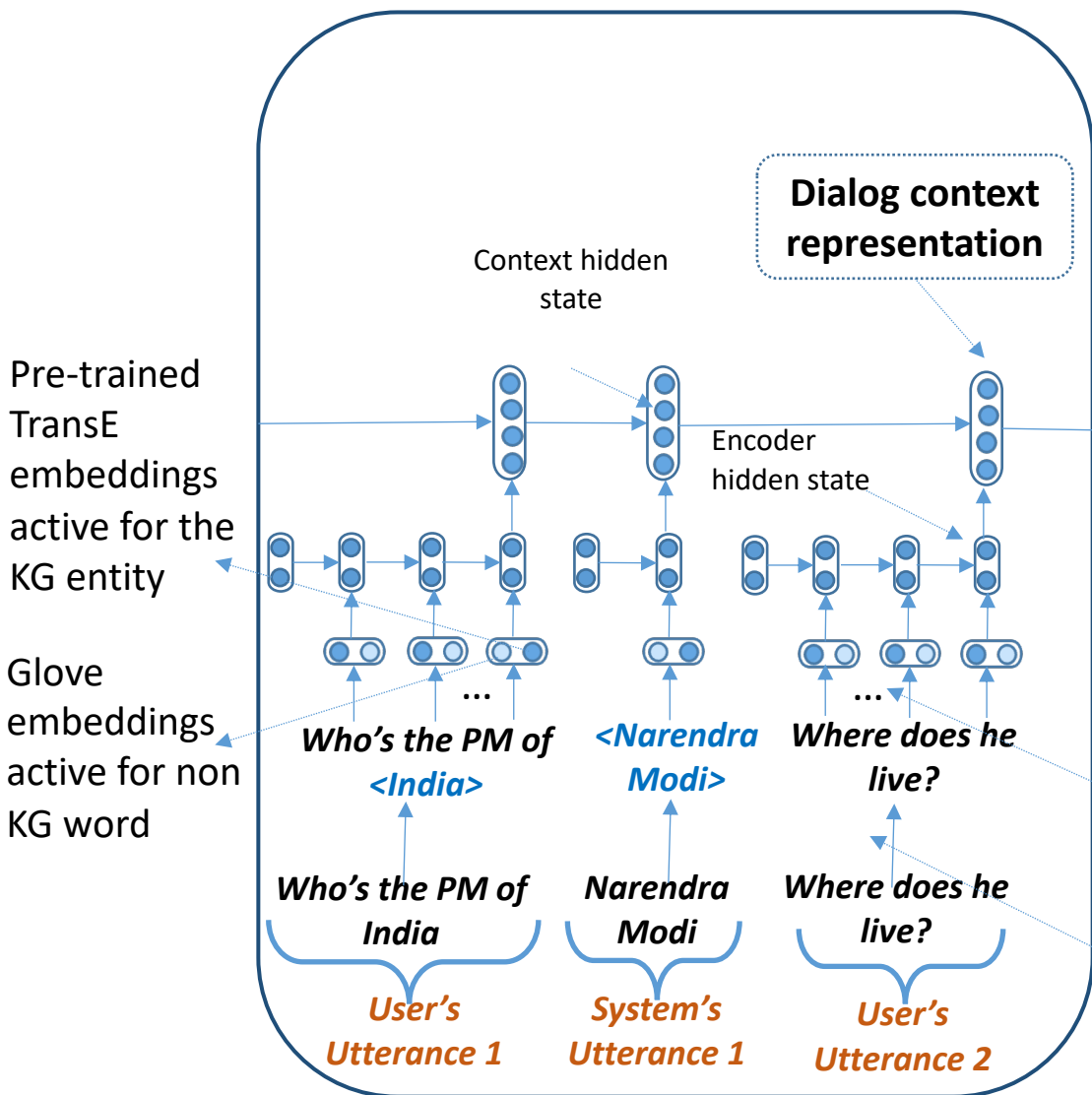
Key-Value Memory Network



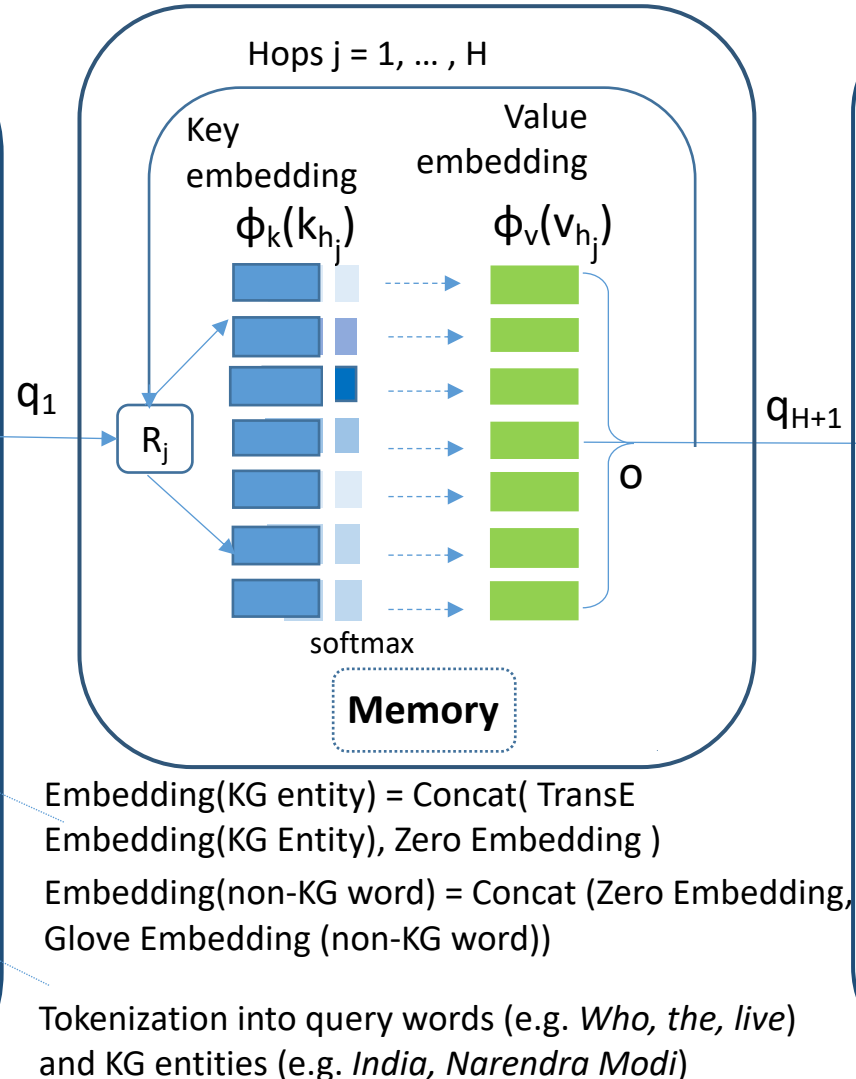
Decoder



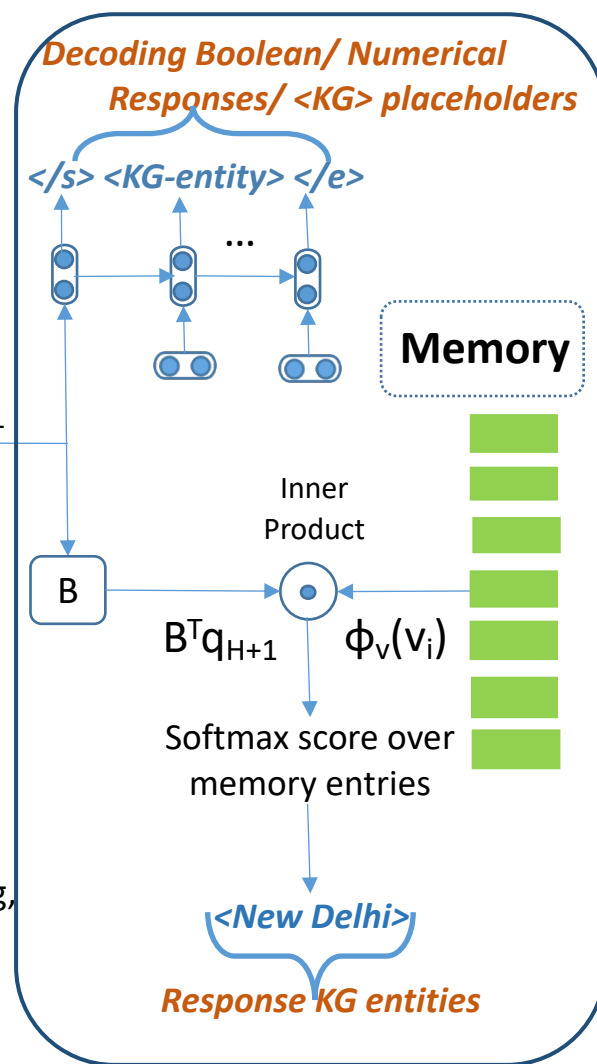
Hierarchical Encoder



Key-Value Memory Network



Decoder



Dataset statistics

- 200,000 dialogs
- 3.2 million utterances (1.6 M turns)
- Wikidata as KG
 - 12.8 M entities
 - 330 unique relations
 - 21.2 M tuples
 - 642 entity types

Dataset Statistics	Train	Valid	Test
Total No. of Dialogs(chat sessions)	152391	16413	27797
Avg. No. of Utterances per dialog	15.9	15.65	19.44
Total No. of Utterances having Question/Answer	1.2M	.13M	.27M
Length of user's question (in words)	9.7	9.68	10.28
Length of system's response (in words)	4.74	4.67	4.37
Avg. No. of Dialog states per dialog	3.89	3.84	4.53
Vocab size (freq>=10)	0.1M	-	-

Example dialog (see more in git repo)

USER: Which people were casted in The 36th Chamber of Shaolin ?

SPEAKER: Gordon Liu, Lo Lieh

USER: Which works of art stars Gordon Liu and Lo Lieh as actor ?

SPEAKER: Kill Bill Volume 1, Flying Swords of Dragon Gate, The Man with the Iron Fists

USER: Which language was used originally for the creation of that work of art ?

SPEAKER: Did you mean Kill Bill Volume 1 ?

USER: No, I meant The Man with the Iron Fists. Could you tell me the answer for that?

SPEAKER: English

USER: How many notations are used as the scripts for approximately the same number of languages as Coptic alphabet ?

SPEAKER: 4

USER: And what about Bamum script?

SPEAKER: 4

USER: Which languages are English or Latin script the writing system of ?

SPEAKER: The answer count is 33. Do you want to see all possibilities?

USER: No, show only a few of them

SPEAKER: Norn, Wailaki language, Slavoserbian, Polabian, Sierra Puebla Nahuatl, Basque, Gaulish, Central Nahuatl, Pangasinan language, Nynorsk

USER: Which languages have writing system as English and Latin script ?

SPEAKER: Pangasinan language, Wailaki language, Gaulish, Norn, Central Nahuatl, Slavoserbian, Basque, Sierra Puebla Nahuatl, Polabian, Nynorsk

USER: How many notations are the writing systems used for approximately 1 language ?

SPEAKER: 4

USER: How many classification schemes are English a component of ?

SPEAKER: 2

USER: And also tell me about German submarine U-42?

SPEAKER: 1