

GameOfThronesQA: Answer-Aware Question-Answer Pairs for TV Series

Aritra Kumar Lahiri and Qinmin Vivian Hu

Ryerson University, Toronto, Canada

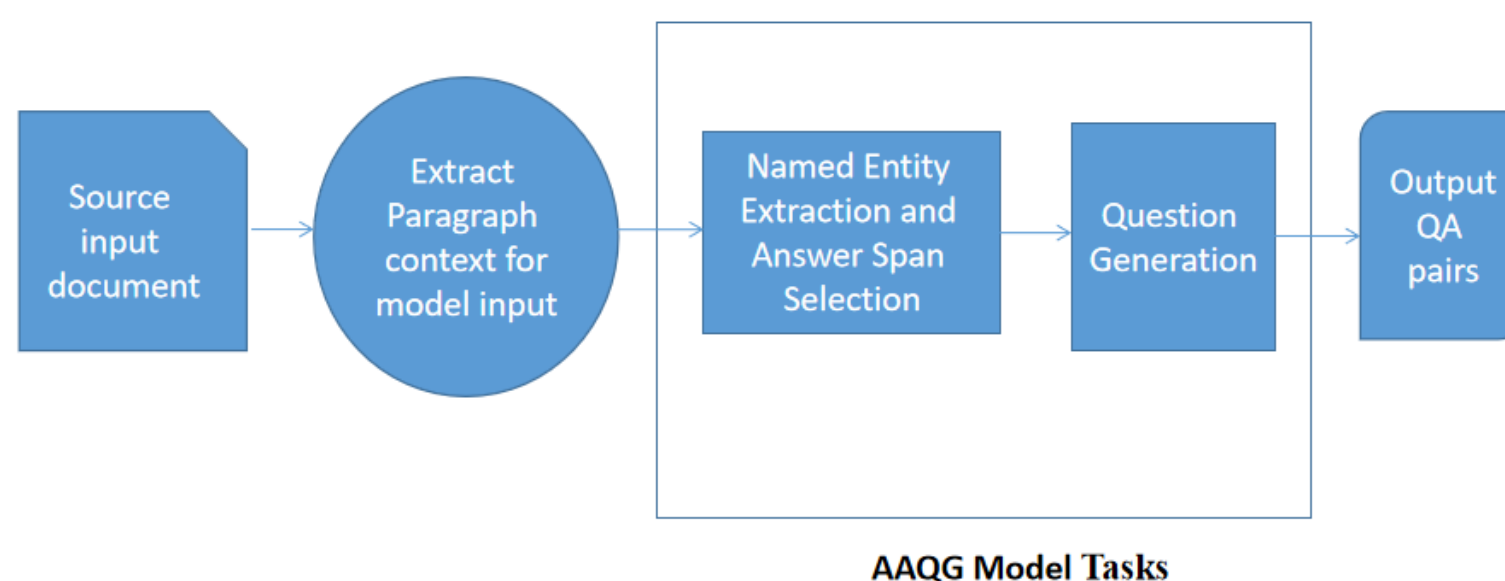
Abstract

In this paper, we offer a corpus of question answer pairs related to the TV series generated from paragraph contexts. The data set called GameofThronesQA V1.0 contains 5237 unique question answer pairs from the Game Of Thrones TV series across the eight seasons. We provide a pipeline approach for answer aware question generation, where the answers are extracted based on the named entities from the TV series.

Objectives

Steps to generate the answer-aware question-answer pairs –

- Approach for selecting answer spans from the named entities.
- Approach to generate questions based on answers, including a sequence-level classification, a span level prediction, and a token-level prediction.
- Combined pipeline approach to output the question answer pairs.



Methodology

Answer Span Selection: Defined by using two token formats - Unique Prepend format [SEP] token and Unique Highlight format is noted as the [HL] token

Answer Aware Question Generation (AAQG) Model: Works on the principle of sequential question generation by fine-tuning adapted BERT-HLSQG

Question Answer Pair Output: Three Tasks - i) **qg** - single question generation , ii) **multi-qa-qg** for multiple QA pairs generation and iii) **e2e-qg** for end-to-end QA pair generation.

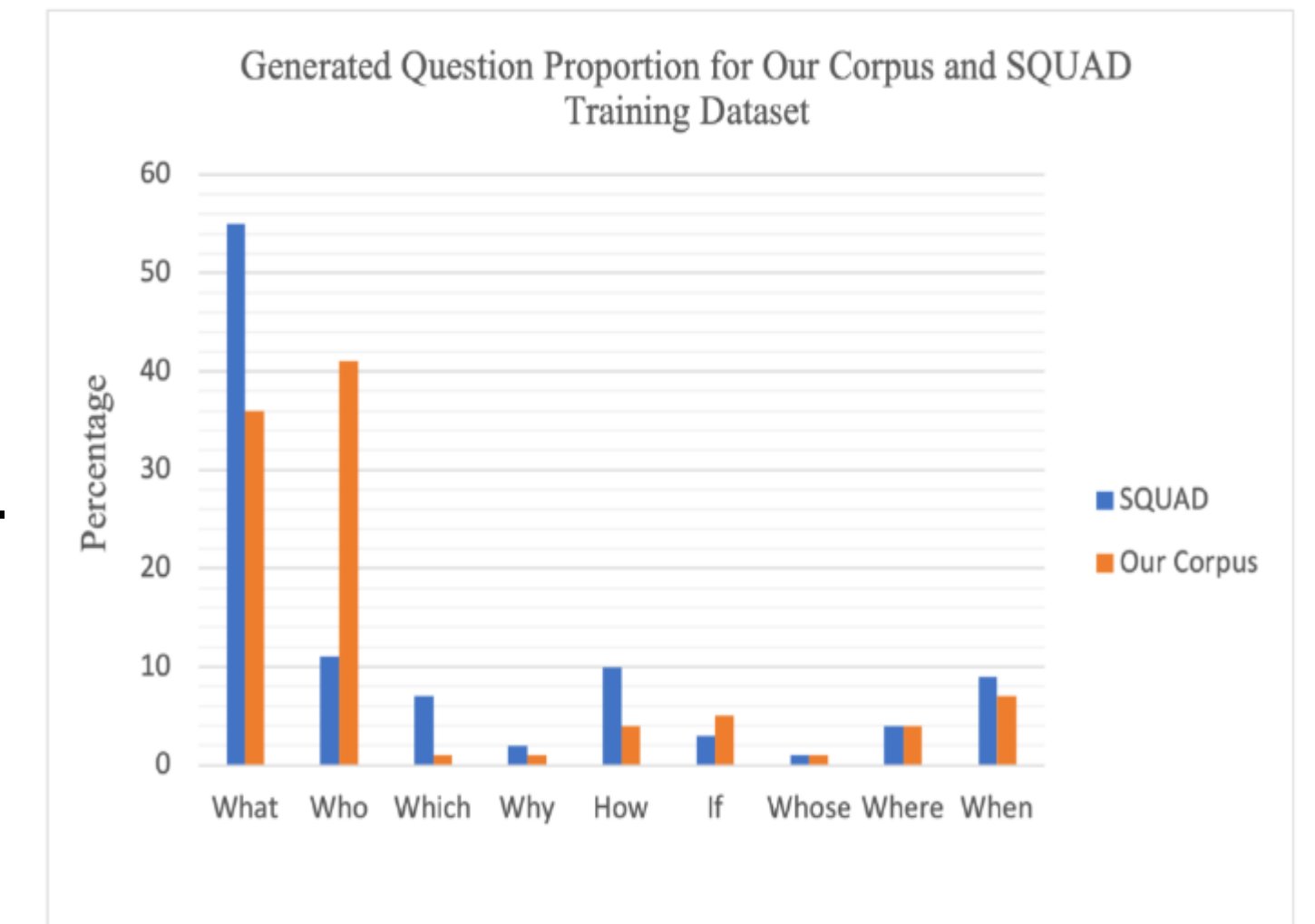
Sample GameOfThronesQA dataset json

```
{
  "answer": "Sandor",
  "question": "Who convinces Arya to abandon her quest for vengeance?"
},
{
  "answer": "Arya's younger brother Bran",
  "question": "Who is chosen as king in Westeros?"
},
{
  "answer": "Arya decides to leave Westeros and sail west to discover what lies beyond where the maps of the known world end.",
  "question": "What does Arya decide to leave Westeros after bidding farewell to her siblings?"
},
{
  "answer": "Arya Stark",
  "question": "Who is the youngest daughter of Lady Catelyn and Lord Ned Stark?"
},
{
  "answer": "Lady Catelyn and Lord Ned Stark",
  "question": "Who are Arya Stark's daughters?"
},
{
  "answer": "Winterfell",
  "question": "Where was Arya born and raised?"
},
{
  "answer": "Jon Snow",
  "question": "Who is Arya's bastard half-brother?"
},
{
  "answer": "Sansa",
  "question": "Who is Arya's older sister?"
},
{
  "answer": "Arya",
  "question": "Who rejects the notion that she must become a lady and marry for influence and power?"
},
}
```

Comparison with Baseline Datasets

Dataset	Distribution Percentage							
	What	Who	How	Where	Why	Which	When	Others
TweetQA	42.33	29.36	7.79	7.00	2.61	2.43	2.16	6.32
GameOfThronesQA	36	41	4	7	5	1	1	5

Class	WIKI QA	GameOfThrones QA
Location	373(11)	884(14.8)
Human	494 (16)	989(17.3)
Numeric	658 (22)	234(3)
Abbreviation	16 (1)	2(0.1)
Entity	419 (14)	3605 (60.74)
Description	1087 (36)	221(3.7)



Conclusion

- Provided question answering dataset called GameofThronesQA whose content comes from the Game of Thrones Wiki Web page.
- Investigated data set with other state-of-the-art question answering sets such as TriviaQA, TweetQA, WikiQA and SQuAD.
- Observed that our data set (1) has good match and F1 scores on the QA pairs, (2) has a different focus on question types which reflects in the "who" and "what" questions because of named entities in answer spans.

Acknowledgements

This study was supported in part by the Discovery and CREATE grants from the Natural Sciences and Engineering Research Council (NSERC) of Canada.