IEEE ICASSP 2022

A QUESTION-ORIENTED PROPAGATION NETWORK FOR NEWS READING COMPREHENSION

Liang Wen, Houfeng Wang, Dehong Ma, Jun Fan, Yingwei Luo, Xiaolin Wang, Daiting Shi, Zhicong Cheng, Dawei Yin

News Reading Comprehension

- **Question:**

  Where is Brittanee Drexel from?

- **Article:**

  The mother of a 17-year-old Rochester, New York, high school student who vanished over the weekend on spring break in Myrtle Beach, South Carolina, says she did not give her daughter permission to go on the trip… Brittanee Marie Drexel's mom says she thought she was at the beach in New York, not South …

- **Answer:**

  Rochester, New York

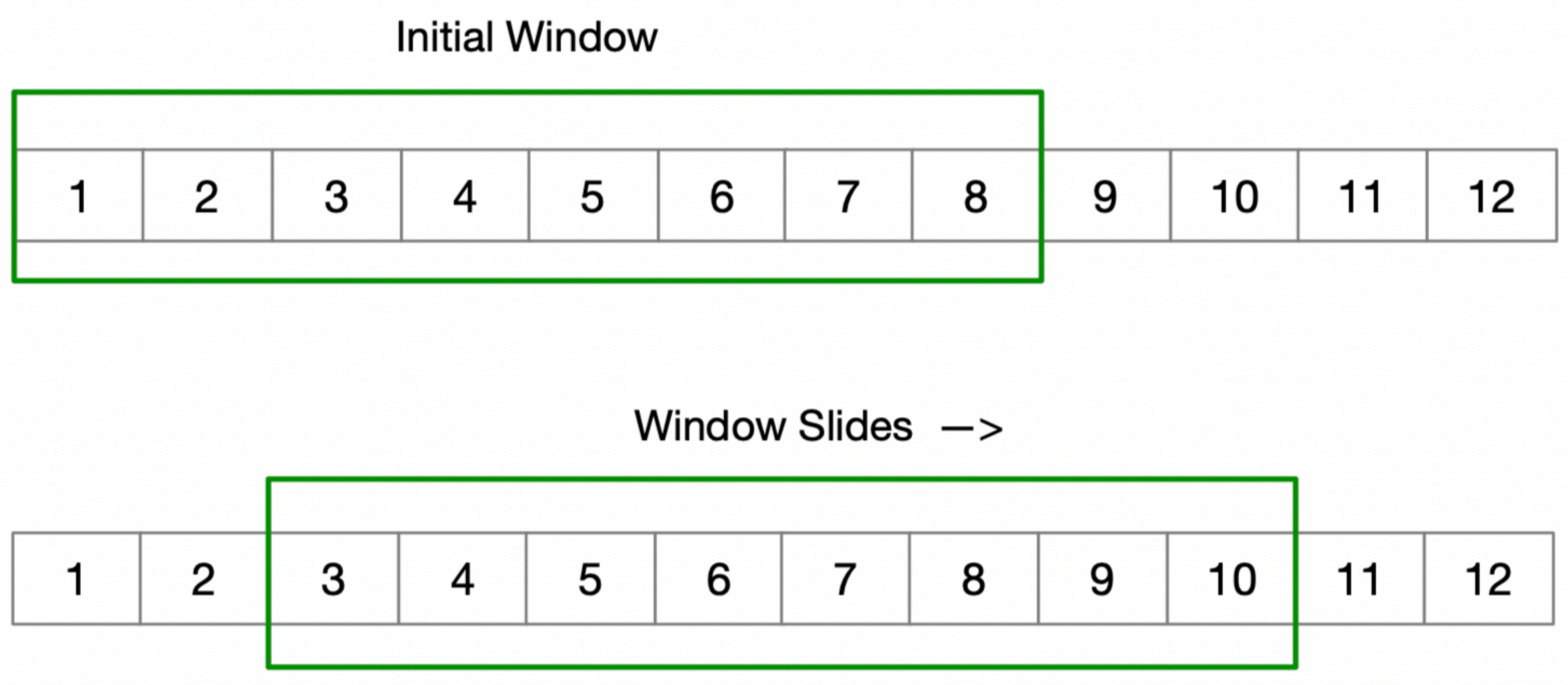
Challenges

- 1) News articles usually are long while the maximum input length of state of the art question answering (QA) models such as BERT [2] and RoBERTa [3] is limited to 512.
- 2) To answer a question, one need to synthesize information across different parts of an article [4, 5].

Outline

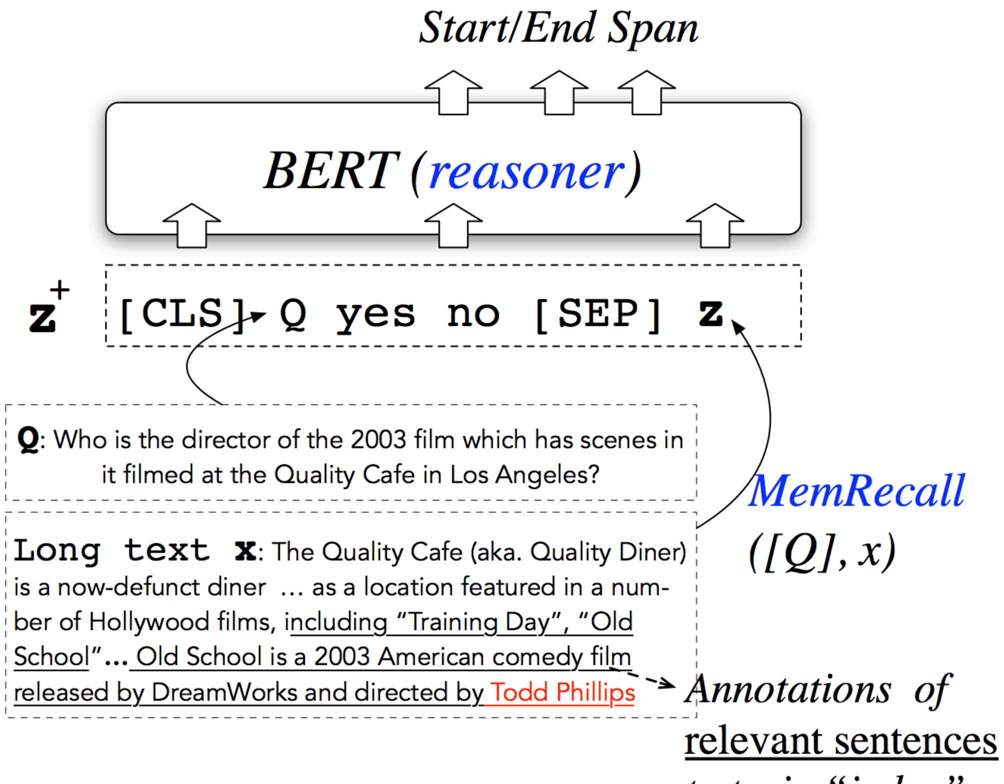
- Related works & Motivation
- Model
- Experiments
- Analysis

Related works: Sliding Window Technique



Sliding Window Technique

Related works: Coarse-to-Fine Paradigm



(a) Span Extraction Tasks
e.g. question answering

(Ding et al., 2020)

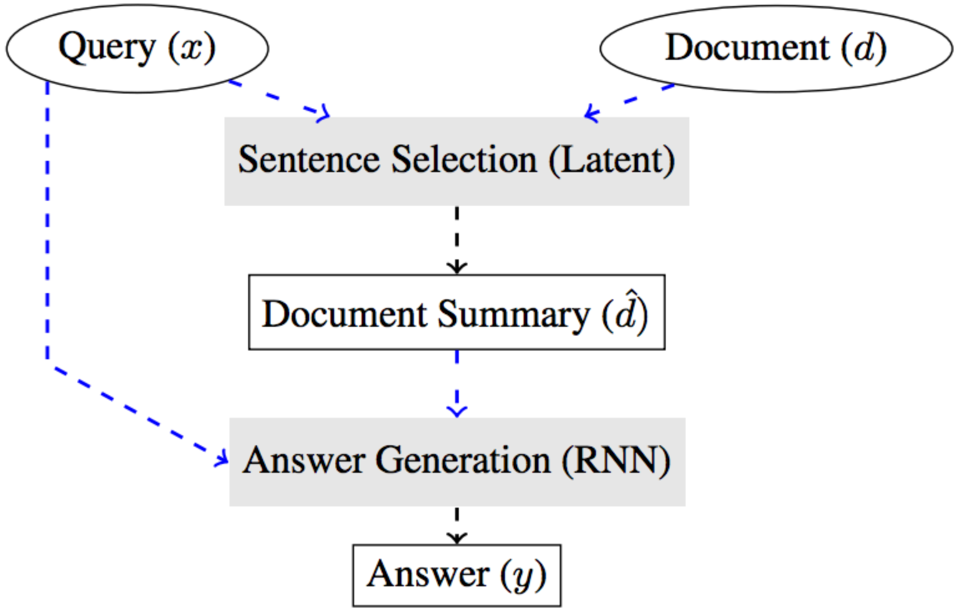
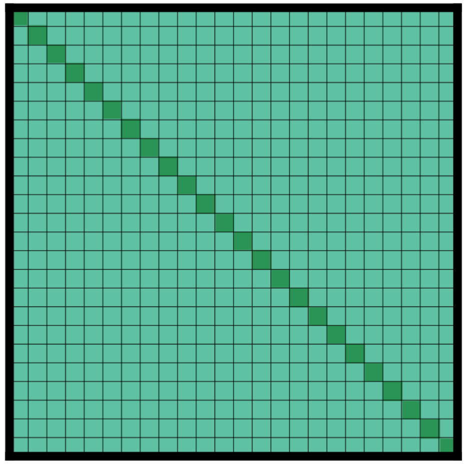


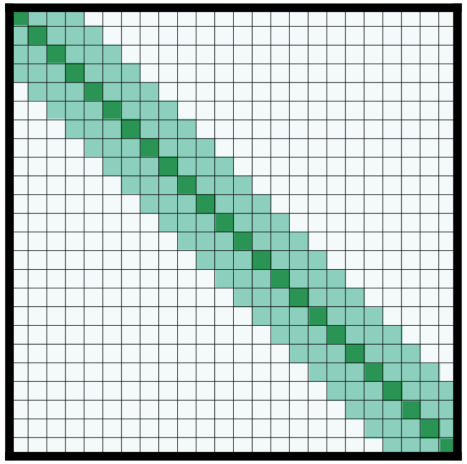
Figure 1: Hierarchical question answering: the model first selects relevant sentences that produce a document summary ($\hat{d}$) for the given query (x), and then generates an answer (y) based on the summary ($\hat{d}$) and the query x.

(Choi et al., 2017)

Related Works: Sparse Attention Mechanism

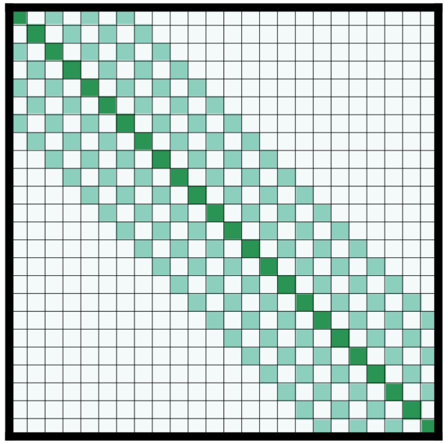


(a) Full n^2 attention

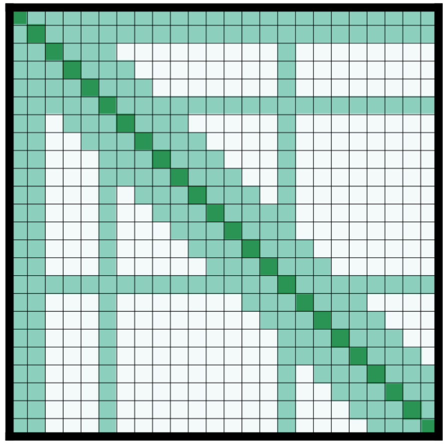


(b) Sliding window attention

(Beltagy, Peters and Cohan, 2020)



(c) Dilated sliding window



(d) Global+sliding window

How People Read Long-Form Article?

- 1) Read the long article segment by segment.
- 2) Focus on question aspects only and measures to which extent question aspects are covered by each segment.
- 3) Re-considerate the question and the question-related context to decide the answer.

Model: Question-Oriented Propagation Network

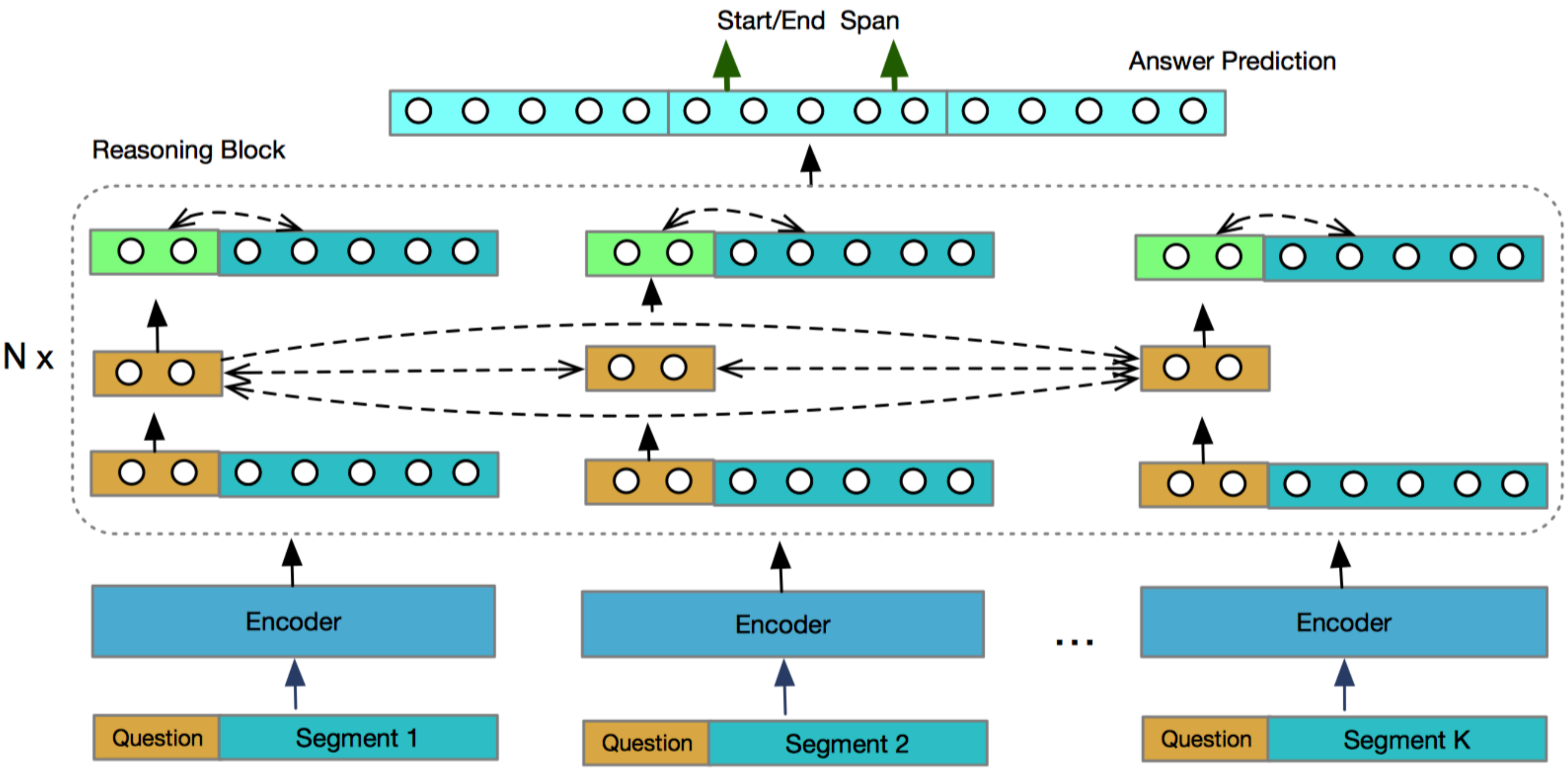


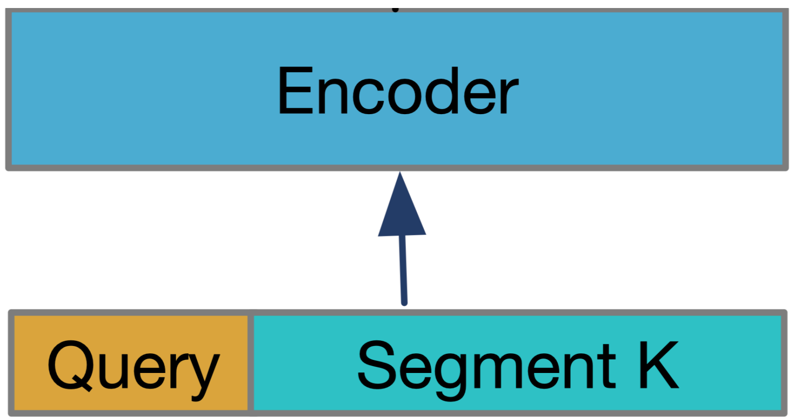
Fig. 1. The architecture of the proposed QOPN model.

Model

- Context Encoding Module
- Multi-step Reasoning Module
- Answer Prediction Module

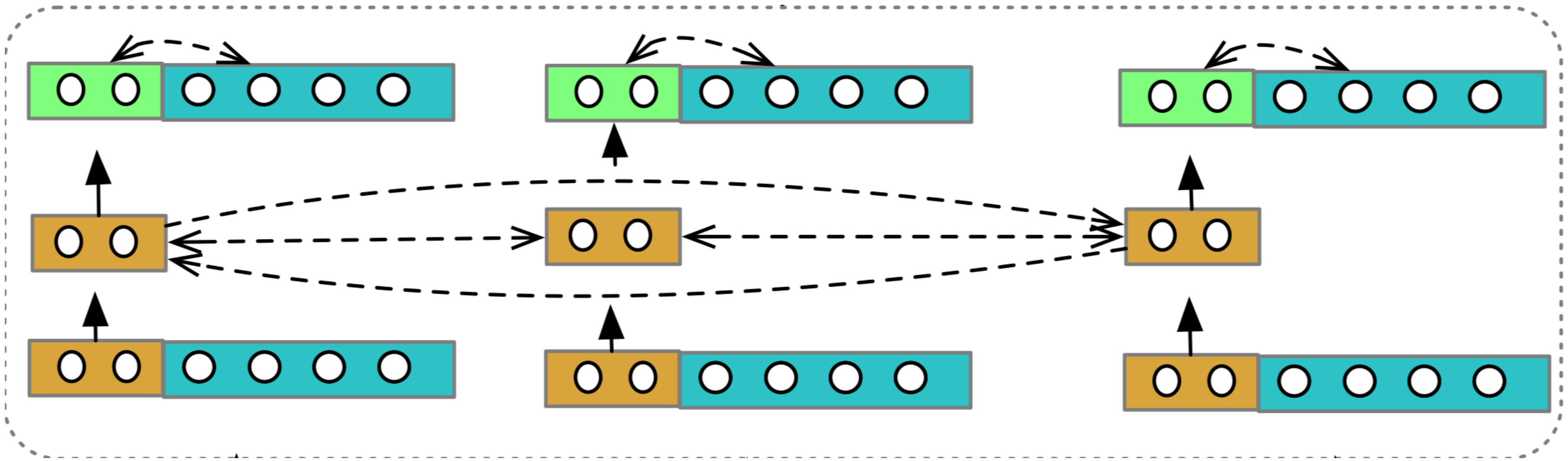
Context Encoding Module

- Given a question Q and a article P, we first split the article into small non-overlapping K segments. Then,



$$\mathbf{H}^k = \mathrm{RoBERTa}\left([CLS], Q, [SEP], S^k, [SEP]\right)$$

Multi-step Reasoning



- Question-Oriented Information Interaction

  Adopt a token-wise multi-head self-attention mechanism

- Gate-Based Information Fusion

  Use a gating mechanism to selectively incorporate the global question-focused information representations

- Question-Guided Information Propagation

  Spread the gated fusion representations to the corresponding question-aware segment representations.

Answer Prediction Module

The probabilities of the start and end positions of an answer span:

$$\mathbf{p}^s = \mathrm{softmax}\left(\mathbf{w}_s \mathbf{Z}\right), \quad \mathbf{p}^e = \mathrm{softmax}\left(\mathbf{w}_e \mathbf{Z}\right)$$

The training loss function:

$$\mathcal{L}(\theta) = -\frac{1}{n} \sum_{i}^{n} \log\left(\mathbf{p}^s_{y_i^s}\right) + \log\left(\mathbf{p}^e_{y_i^e}\right)$$

Advantages:

- 1. Take the whole article into consideration and jointly learn to find question-related clues and make inference over them implicitly.

- 2. Does not rely on hand-designed patterns and directly aims at question-focused information.

Experimental Datasets & Evaluation Metrics

Datasets	#Samples	#Doc.	Avg #Words
NewsQA	100k	13k	616
NLQuAD	31k	13k	877

Table 1. Statistics of MRC datasets. #Samples and #Doc. are the number of samples and documents respectively. Avg #Words denotes the average number of words per document.

Evaluation Metrics: EM, F1 and IOU

Experimental Results

Model	EM(%)	F1(%)
FastQAExt [18]	42.8	56.1
AMANDA [19]	48.4	63.7
MINIMAL [20]	50.1	63.2
DECAPROP [21]	53.1	66.3
RoBERTa-large [3] (sliding window)	49.6	66.3
CogLTX [9]	55.2	70.1
QOPN (RoBERTa-base)	**61.2**	**75.1**
QOPN	**65.5**	**79.8**

Table 2. Performance on the NewsQA test set.

Experimental Results

Model	**EM(%)**	**F1(%)**	**IoU(%)**
BERT-base [2]	25.0	64.0	53.8
BERT-large [2]	30.3	67.9	58.4
RoBERTa-base [3]	29.1	67.2	57.7
RoBERTa-large [3]	33.4	71.1	62.4
Longformer [12]	50.3	81.4	73.6
QOPN	**54.0**	**82.9**	**75.8**

Table 3. Performance on the NLQuAD test set.

Results on much longer-context QA benchmark

Dataset	#train	#test	med	max
Quasar-T	29k	3k	2.8k	8.2k

Statistics of Quasar-T dataset (Dhingra et al., 2017)

Model	EM（%）	F1（%）
DrQA (Chen et al., 2017)	37.7	44.5
R3 (Wang et al., 2018)	35.3	41.7
DECAPROP (Tay et al., 2018)	38.6	46.9
DS-QA (Lin et al., 2018)	42.2	49.3
Multi-passage BERT (Wang et al., 2019)	51.1	59.1
Sliding Window	52.9	62.8
Sparse Attention (Child et al., 2019)	52.1	62.0
Locality-Sensitive Hashing (Kitaev et al., 2020)	53.2	62.9
Cluster-Former(#C=512) (Wang et al., 2021)	54.0	63.9
QOPN	**55.7**	**64.6**

Performance on the Quasar-T test set.

Analysis

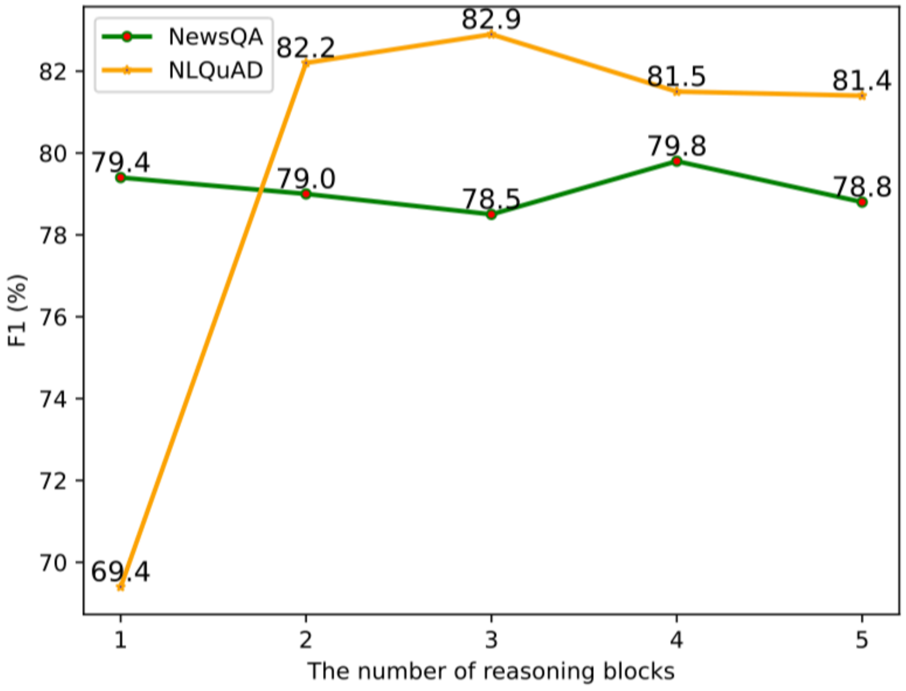


Fig. 2. The effect of number of reasoning blocks.

Model	**EM**	**F1**	**IoU**
QOPN(full model)	**54.0**	**82.9**	**75.8**
- Question-oriented interaction	47.0	79.4	71.2
- Gate-based fusion	52.8	82.3	75.1
- Question-guided propagation	51.2	80.9	73.3

Table 4. Ablation studies of QOPN on the NLQuAD dataset.

Conclusion

- Summary

  1) Propose a question-oriented propagation network model

  2) Achieve new state-of-the-art performances on challenging machine comprehension datasets

- Future work

  1) Query-focused Multi-document Summarization (either extractive or abstract)

  2) Long-text Matching


2022 IEEE International Conference on
Acoustics, Speech, and Signal Processing
human-centric signal processing
icassp 2022
Singapore, China, Virtual
May 22 - 27, 2022 - In-Person
@ Marina Bay Sands Expo and Convention Centre
May 22 - 27, 2022 - In-Person
@ The Chinese University of HongKong, Shenzhen
May 7 - 13, 2022 - Virtual for All Paper Presentations
IEEE

Thank You !