



# Denoising Distantly Supervised Open-Domain Question Answering



Yankai Lin, Haozhe Ji, Zhiyuan Liu, Maosong Sun  
Tsinghua University

## Overview

**Goal:** build a denoising distantly supervised open-domain question answering system.

**Motivation:** How human being read?

Fast Skimming + Careful Reading + Summarization

**Question:** What's the capital of Ireland?



**Paragraphs:**

$p_1$ : As the capital of Ireland, Dublin is ...  
 $p_2$ : Ireland is an island in the North Atlantic...  
 $p_3$ : Dublin is the capital of Ireland. Besides, Ottawa is one of famous tourist cities in Ireland and ...

Summarizing  $P(a|q,P) = \sum_{p_i \in P} P(a|q,p_i)P(p_i|q,P)$

### Fast Skimming

$p_1$ : As the capital of Ireland, Dublin is ...  
 $p_3$ : Dublin is the capital of Ireland. Besides, Ottawa is one of famous tourist cities in Ireland and ...

Paragraph Selector

$P(p_i|q,P)$

### Careful Reading

$p_1$ : As the capital of Ireland, Dublin is ...  
 $p_3$ : Dublin is the capital of Ireland. Besides, Dublin is one of famous tourist cities in Ireland and ...

Paragraph Reader

$P(a|q,p_i)$

## Experimental Settings

### Datasets

- Distantly Supervised data
  - (Q, A)  $\rightarrow$  (Q, P, A) P is retrieved using traditional IR
- Quasar-T (Dhingra et al., 2017)
- SearchQA (Dunn et al., 2017)
- TriviaQA (Joshi et al., 2017)

### Evaluation Metrics

- Exact Match (EM)
- F1

## Overall Results

Improve consistently and significantly!

- Our+AVG: consider more information
- Our+FULL: filter out noise information

Datasets	Quasar-T		SearchQA		TriviaQA	
	EM	F1	EM	F1	EM	F1
Models						
GA	26.4	26.4	-	-	-	-
BiDAF	25.9	28.5	28.6	34.6	-	-
AQA	-	-	40.5	47.4	-	-
R <sup>3</sup>	35.3	41.7	49.0	55.3	47.3	53.7
<b>Our + AVG</b>	<b>38.5</b>	<b>45.7</b>	<b>55.6</b>	<b>61.0</b>	<b>42.7</b>	<b>48.2</b>
<b>+ FULL</b>	<b>42.2</b>	<b>49.3</b>	<b>58.8</b>	<b>64.5</b>	<b>48.7</b>	<b>56.3</b>

## Potential Improvement

Our model has much potential!

- Top-5 vs Top-1: 14~18 points improvement
- Can be further improved by re-ranking Approach

Datasets	Quasar-T			SearchQA		
	Models	Top-k	EM	F1	EM	F1
R <sup>3</sup>		1	35.5	41.6	51.2	57.3
		3	46.2	53.5	63.9	68.9
		5	51.0	58.9	69.1	73.9
Our+FULL		1	42.2	49.3	58.8	67.4
		3	53.1	62.0	72.9	77.4
		5	56.4	66.4	76.9	81.0

## Case Study

Needs information from all paragraphs!

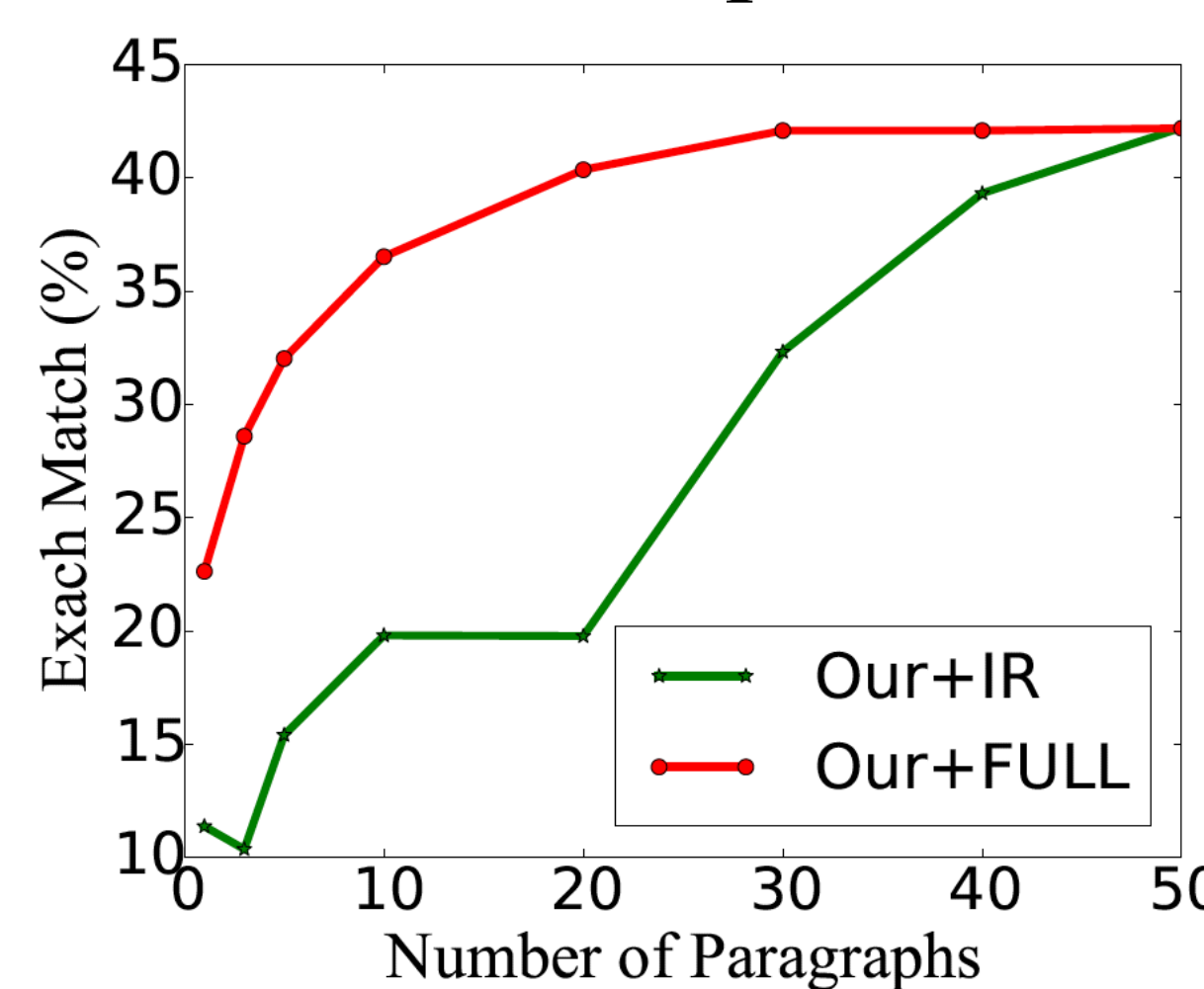
- Three clues (artist, left hand, right hand) from three paragraphs

Question	What famous artist could write with both his left and right hand at the same time
Ground Truth	Leonardo Da Vinci
P1	Leonardo Da Vinci was and is best known as an artist,...
P2	... the reason Leonardo da Vinci used his left hand exclusively was that his right hand was paralyzed.
P3	... forced me to use my right-hand,... beat my left-hand fingers with ... so that I use the right hand.

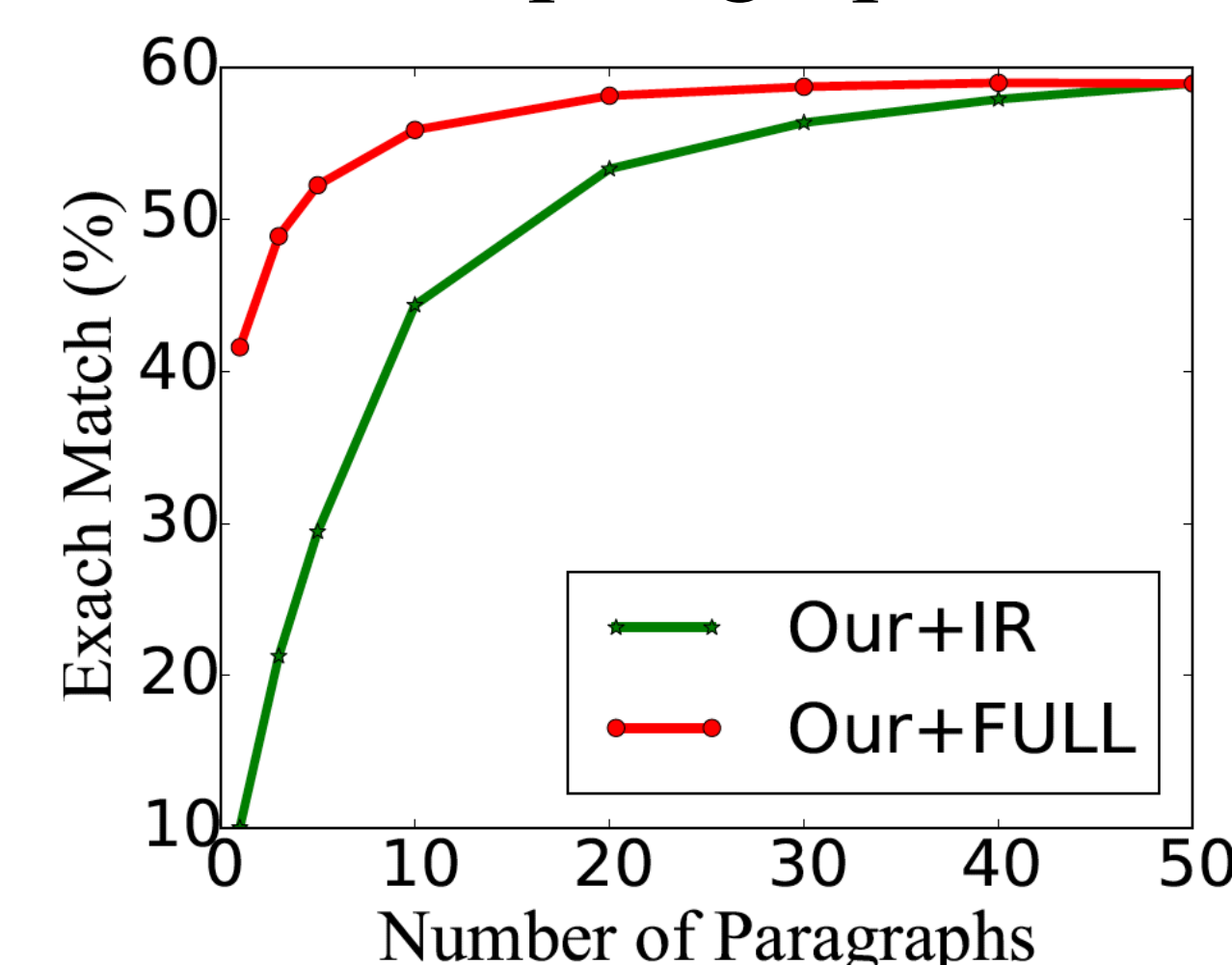
## Different Numbers of Paragraphs

Our model can better select informative paragraphs!

- Achieve comparable results with half paragraphs



Left: Quasar-T



Right: SearchQA

## Contact Me

Email: [linyk14@mails.tsinghua.edu.cn](mailto:linyk14@mails.tsinghua.edu.cn)

Paper:



Code:

