

Understanding Dataset Design Choices for Multi-hop Reasoning



Jifan Chen and Greg Durrett

The University of Texas at Austin



Multi-hop Reasoning

Question: The Oberoi family is part of a hotel company that has a head office in what city?

Doc 1 The Oberoi family is an Indian family that is famous for its involvement in hotels, namely through The Oberoi Group ...

Doc 2 The Oberoi Group is a hotel company with its head office in Delhi.
...



Multi-hop Reasoning

Question: The Oberoi family is part of a hotel company that has a head office in what city?

Doc 1 The Oberoi family is an Indian family that is famous for its involvement in hotels, namely through The Oberoi Group ...

Doc 2 The Oberoi Group is a hotel company with its head office in Delhi.
...



Multi-hop Reasoning

Question: The Oberoi family is part of a hotel company that has a head office in what city?

Same entity

Doc 1

The Oberoi family is an Indian family that is famous for its involvement in hotels, namely through The Oberoi Group ...

Doc 2

The Oberoi Group is a hotel company with its head office in Delhi.

...



Multi-hop Reasoning

Question: The Oberoi family is part of a hotel company that has a head office in what city?

Same entity

Doc 1

The Oberoi family is an Indian family that is famous for its involvement in hotels, namely through The Oberoi Group ...

Doc 2

The Oberoi Group is a hotel company with its head office in Delhi.

...



Multi-hop Reasoning

Question: The Oberoi family is part of a hotel company that has a head office in what city?

Same entity

Doc 1

The Oberoi family is an Indian family that is famous for its involvement in hotels, namely through The Oberoi Group ...

Same entity

Doc 2

The Oberoi Group is a hotel company with its head office in Delhi.
...



Multi-hop Reasoning

Question: The Oberoi family is part of a hotel company that has a head office in what city?

Same entity

Doc 1

The Oberoi family is an Indian family that is famous for its involvement in hotels, namely through The Oberoi Group ...

Same entity

Doc 2

The Oberoi Group is a hotel company with its head office in Delhi.

...



Multi-hop Reasoning

Question: The Oberoi family is part of a hotel company that has a head office in what city?

Same entity

Doc 1

The Oberoi family is an Indian family that is famous for its involvement in hotels, namely through The Oberoi Group ...

Same entity

Doc 2

The Oberoi Group is a hotel company with its head office in Delhi.

...

This is an idealized version of multi-hop reasoning. Do models **need** to do this to do well on this task?



Multi-hop Reasoning

Question: The Oberoi family is part of a hotel company that has a head office in what city?

Doc 1

The Oberoi family is an Indian family that is famous for its involvement in hotels, namely through The Oberoi Group ...

Doc 2

The Oberoi Group is a hotel company with its head office in Delhi.

...



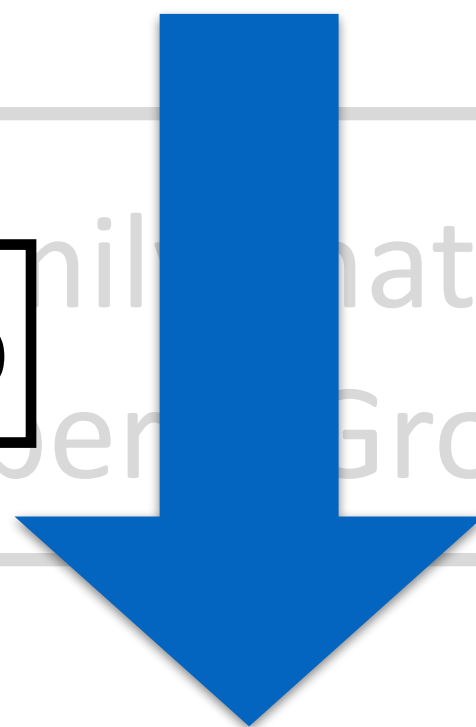
Multi-hop Reasoning

Question: The Oberoi family is part of a hotel company that has a head office in what city?

Doc 1

The Oberoi family is a prominent family that is famous for its involvement in hotels, namely through the Oberoi Group ...

High lexical overlap



Doc 2

The Oberoi Group is a hotel company with its head office in Delhi.

...

Ignore the bridging entity and directly jump to the answer!



Multi-hop Reasoning

Question: What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell ?



Multi-hop Reasoning

Question: What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell ?

Doc 1 Shirley Temple Black was an American actress, businesswoman, and singer ...
As an adult, she served as Chief of Protocol of the United States
...

Doc 2 Kiss and Tell is a comedy film in which 17-year-old Shirley Temple acts as
Corliss Archer .
...

Doc 3 Meet Corliss Archer is an American television sitcom that aired on CBS ...



Multi-hop Reasoning

Question: What government position was held by the woman who portrayed **Corliss Archer** in the film Kiss and Tell ?

Doc 1 Shirley Temple Black was an American actress, businesswoman, and singer ...
As an adult, she served as Chief of Protocol of the United States
...

Doc 2 Kiss and Tell is a comedy film in which 17-year-old Shirley Temple acts as
Corliss Archer .
...

Doc 3 Meet Corliss Archer is an American television sitcom that aired on CBS ...



Multi-hop Reasoning

Question: What government position was held by the woman who portrayed **Corliss Archer** in the film Kiss and Tell ?

Doc 1 Shirley Temple Black was an American actress, businesswoman, and singer ...
As an adult, she served as Chief of Protocol of the United States
...

Same entity

Doc 2 Kiss and Tell is a comedy film in which 17-year-old Shirley Temple acts as
Corliss Archer .
...

Doc 3 Meet Corliss Archer is an American television sitcom that aired on CBS ...



Multi-hop Reasoning

Question: What government position was held by the woman who portrayed **Corliss Archer** in the film Kiss and Tell ?

Doc 1 Shirley Temple Black was an American actress, businesswoman, and singer ...
As an adult, she served as Chief of Protocol of the United States
...

Same entity

Doc 2 Kiss and Tell is a comedy film in which 17-year-old **Shirley Temple** acts as
Corliss Archer .
...

Doc 3 Meet Corliss Archer is an American television sitcom that aired on CBS ...



Multi-hop Reasoning

Question: What government position was held by the woman who portrayed **Corliss Archer** in the film Kiss and Tell ?

Doc 1 **Shirley Temple** Black was an American actress, businesswoman, and singer ...
As an adult, she served as Chief of Protocol of the United States
...
Same entity

Doc 2 Kiss and Tell is a comedy film in which 17-year-old **Shirley Temple** acts as
Corliss Archer .
...

Doc 3 Meet Corliss Archer is an American television sitcom that aired on CBS ...



Multi-hop Reasoning

Question: What government position was held by the woman who portrayed **Corliss Archer** in the film Kiss and Tell ?

Doc 1 Shirley Temple Black was an American actress, businesswoman, and singer ...
As an adult, she served as Chief of Protocol of the United States ...

Same entity

Same entity

Doc 2 Kiss and Tell is a comedy film in which 17-year-old Shirley Temple acts as
Corliss Archer

Doc 3 Meet Corliss Archer is an American television sitcom that aired on CBS ...



Multi-hop Reasoning

Question: What government position was held by the woman who portrayed **Corliss Archer** in the film Kiss and Tell ?

Doc 1 Shirley Temple Black was an American actress, businesswoman, and singer ...
As an adult, she served as **Chief of Protocol** of the United States

Same entity

...

Same entity

Doc 2 Kiss and Tell is a comedy film in which 17-year-old **Shirley Temple** acts as **Corliss Archer** .

...

Doc 3 Meet Corliss Archer is an American television sitcom that aired on CBS ...

No simple lexical overlap.



Multi-hop Reasoning

Question: What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell ?

Doc 1 Shirley Temple Black was an American actress, businesswoman, and singer ...
As an adult, she served as Chief of Protocol of the United States
...

Doc 2 Kiss and Tell is a comedy film in which 17-year-old Shirley Temple acts as
Corliss Archer .
...

Doc 3 Meet Corliss Archer is an American television sitcom that aired on CBS ...



Multi-hop Reasoning

Question: What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell ?

Doc 1 Shirley Temple Black was an American actress, businesswoman, and singer ...
As an adult, she served as Chief of Protocol of the United States
...

Doc 2 Kiss and Tell is a comedy film in which 17-year-old Shirley Temple acts as
Corliss Archer .
...

Doc 3 Meet Corliss Archer is an American television sitcom that aired on CBS ...



Multi-hop Reasoning

Question: What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell ?

Doc 1 Shirley Temple Black was an American actress, businesswoman, and singer ...
As an adult, she served as Chief of Protocol of the United States
...

Doc 2 Kiss and Tell is a comedy film in which 17-year-old Shirley Temple acts as
Corliss Archer .
...

Doc 3 Meet Corliss Archer is an American television sitcom that aired on CBS ...

Only one government position appears in the context!



Multi-hop Reasoning Datasets

How much multi-hop reasoning is really needed in the recently proposed multi-hop oriented datasets?

WikiHop (Welbl et al., 2018)

HotpotQA (Yang et al., 2018)



Multi-hop Reasoning Datasets

How much multi-hop reasoning is really needed in the recently proposed multi-hop oriented datasets?

WikiHop (Welbl et al., 2018)

- ▶ Multiple-choice

HotpotQA (Yang et al., 2018)

- ▶ Span-based



Multi-hop Reasoning Datasets

How much multi-hop reasoning is really needed in the recently proposed multi-hop oriented datasets?

WikiHop (Welbl et al., 2018)

- ▶ Multiple-choice
- ▶ Final answer is given as supervision

HotpotQA (Yang et al., 2018)

- ▶ Span-based
- ▶ Answer and supporting facts



Research Questions



Research Questions

Can a model identify the answer with only a set of candidates?

Government position → Chief of Protocol, actress, singer



Research Questions

Can a model identify the answer with only a set of candidates?

Government position → Chief of Protocol, actress, singer

Can a model identify where the answer is in a single hop?

Oberoi Family → Delhi



Research Questions

Can a model identify the answer with only a set of candidates?

Government position → Chief of Protocol, actress, singer

Can a model identify where the answer is in a single hop?

Oberoi Family → Delhi

Does the type of supervision matter?

Span-based vs. Multiple-choice



Research Questions

Can a model identify the answer with only a set of candidates?

Government position → Chief of Protocol, actress, singer

Can a model identify where the answer is in a single hop?

Oberoi Family → Delhi

Does the type of supervision matter?

Span-based vs. Multiple-choice



Finding the answer directly

Question: What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell ?

Doc 1 Shirley Temple Black was an American actress, businesswoman, and singer ...
As an adult, she served as Chief of Protocol of the United States

Doc 2 Kiss and Tell is a comedy film in which 17-year-old Shirley Temple acts as
Corliss Archer .

Doc 3 Meet Corliss Archer is an American television sitcom that aired on CBS ...



Chief of Protocol

businesswoman

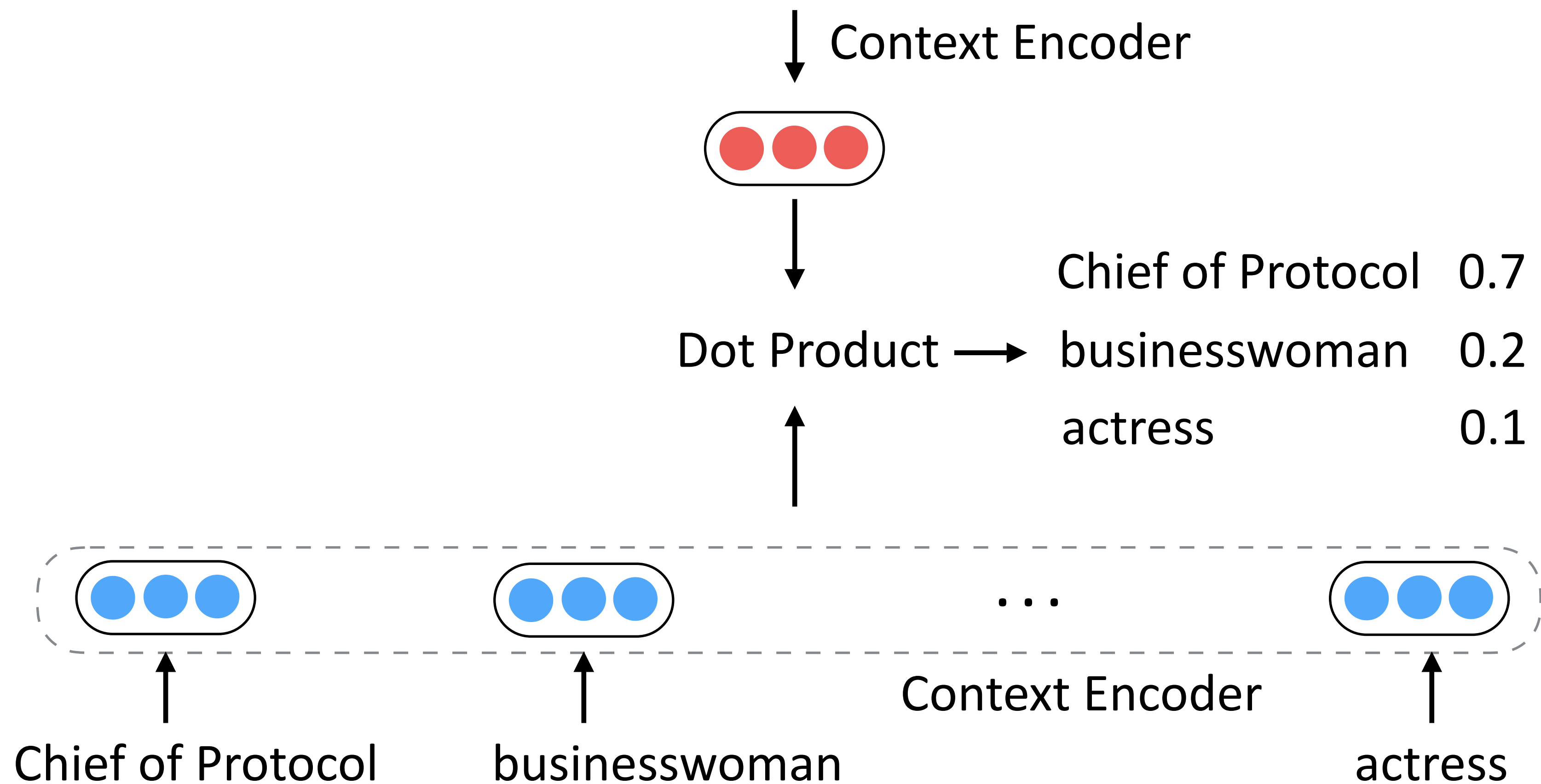
...

actress



No Context Baseline

Question: What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell ?

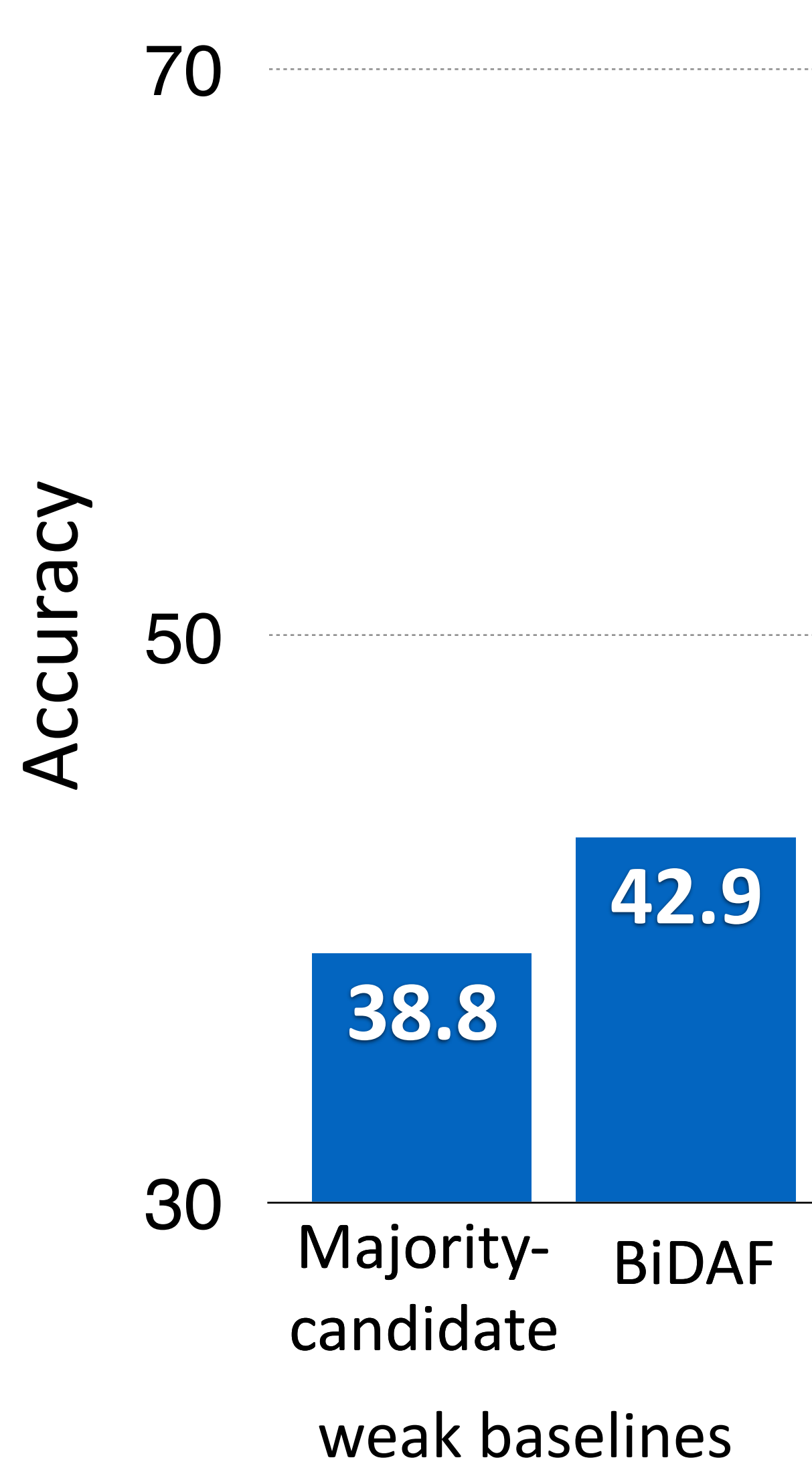




Results on WikiHop

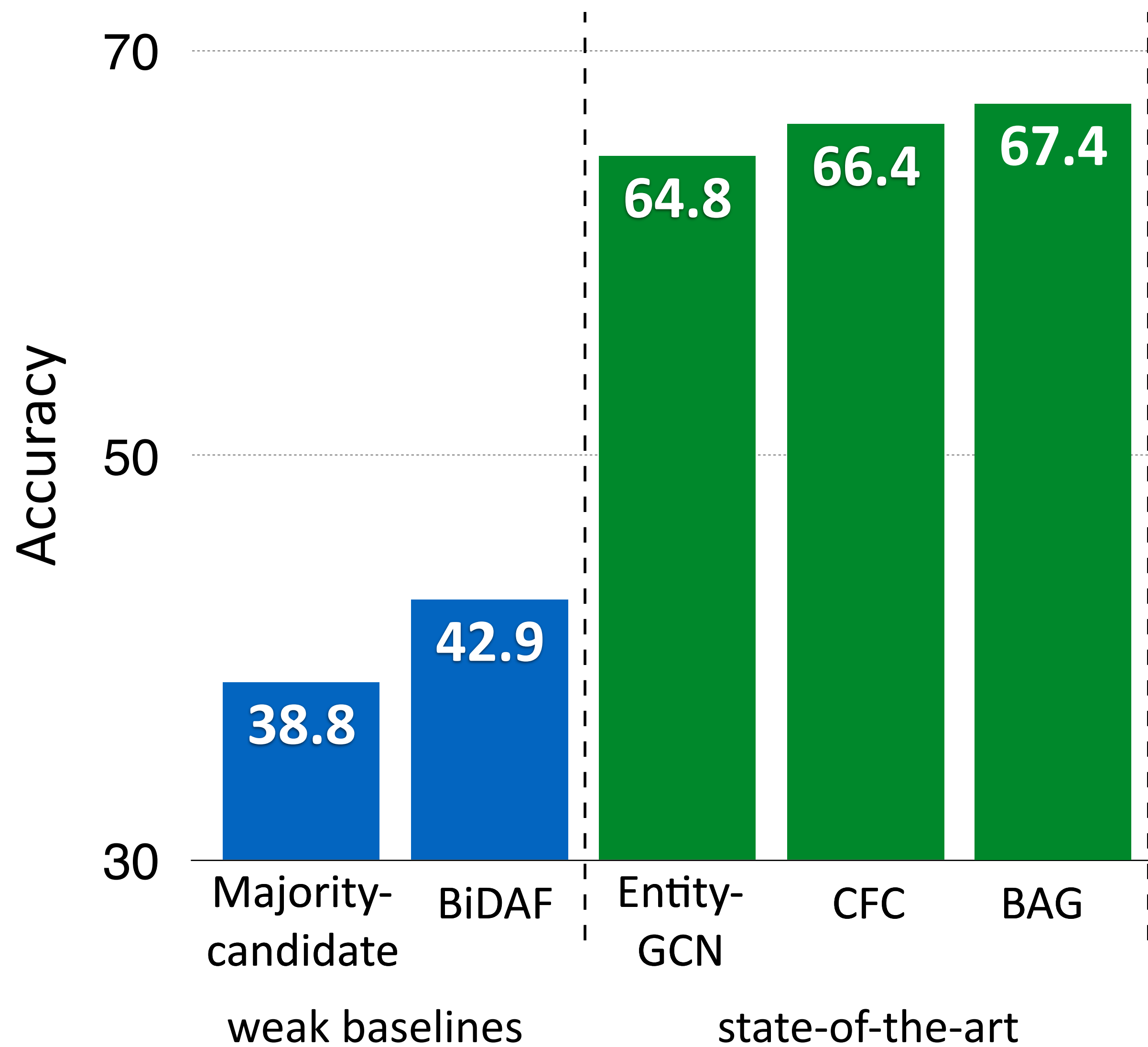


Results on WikiHop



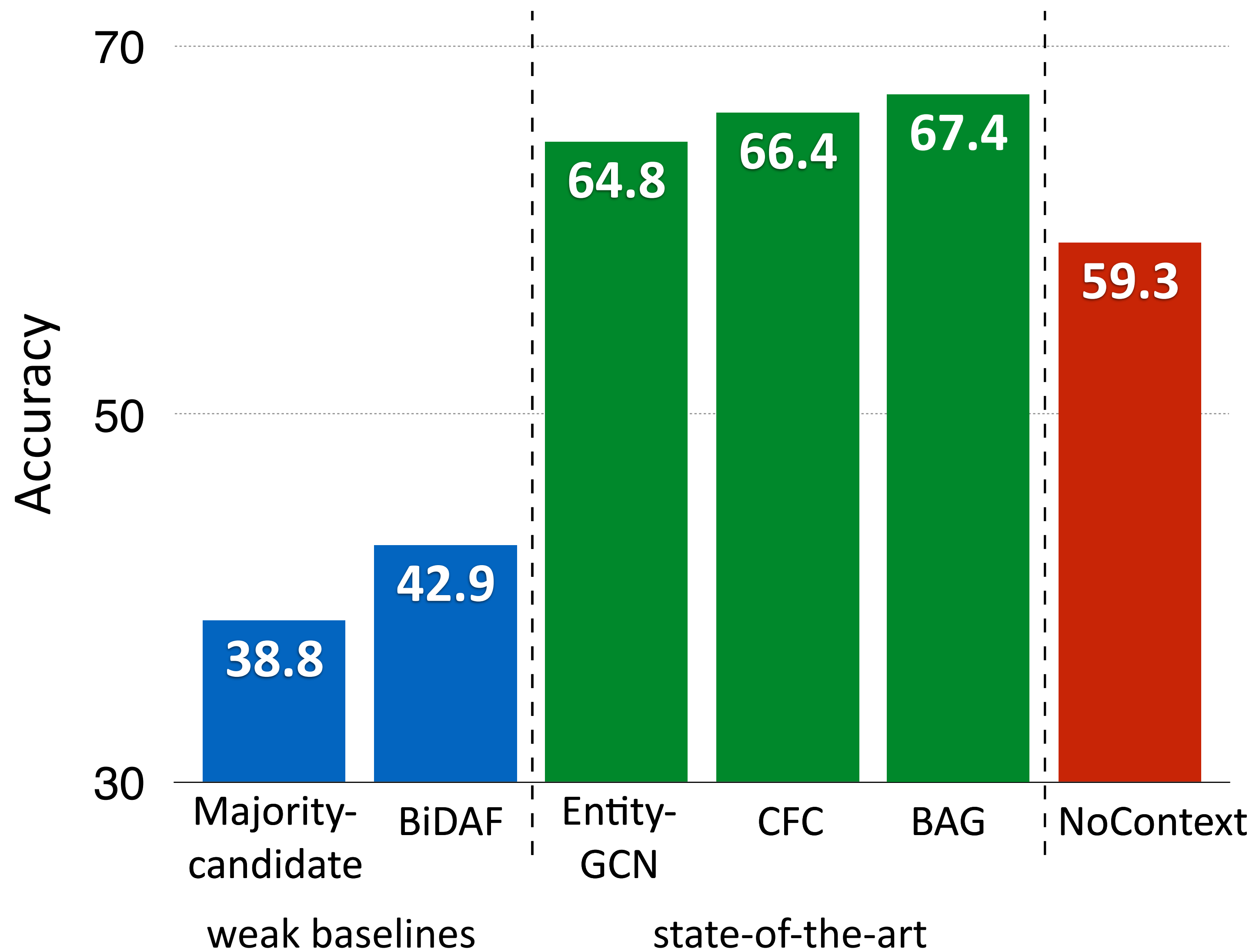


Results on WikiHop



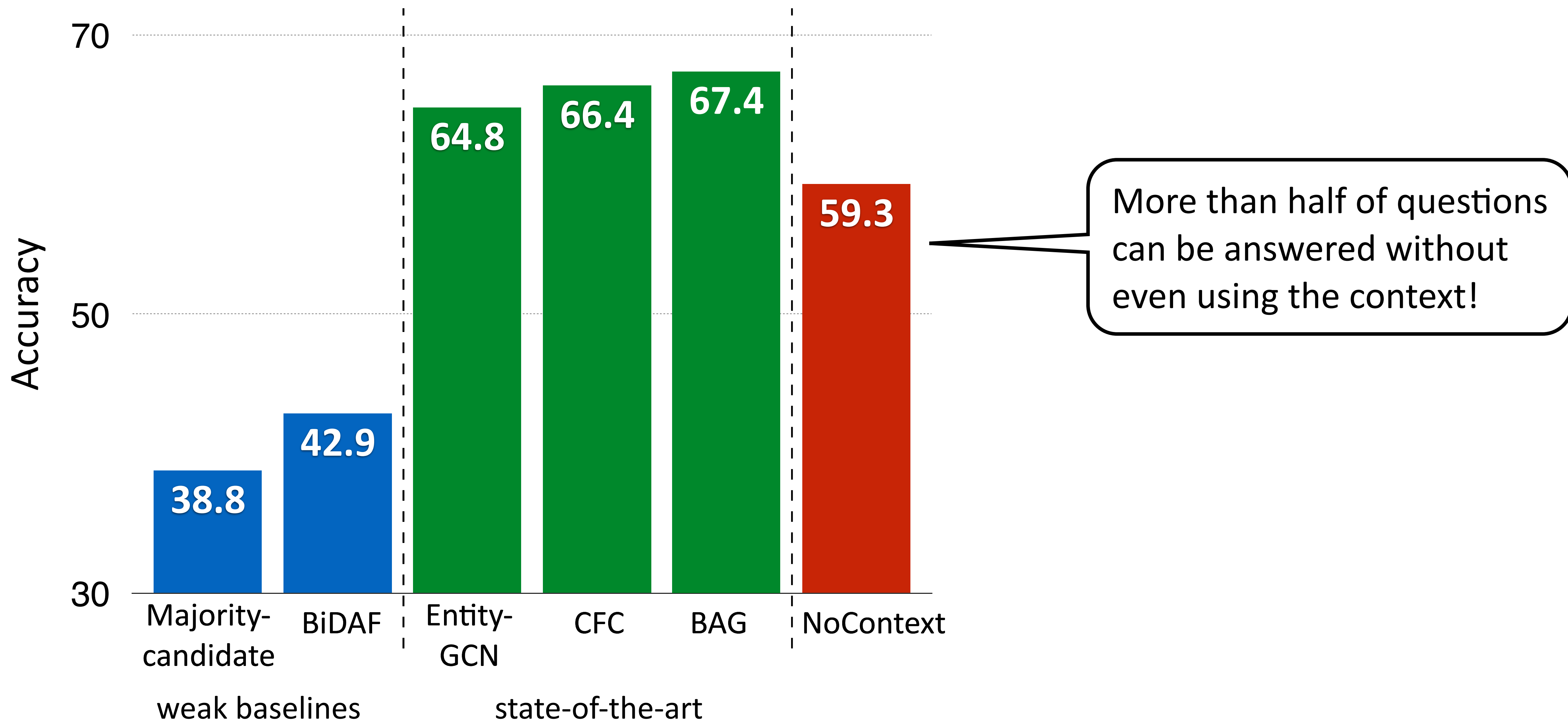


Results on WikiHop



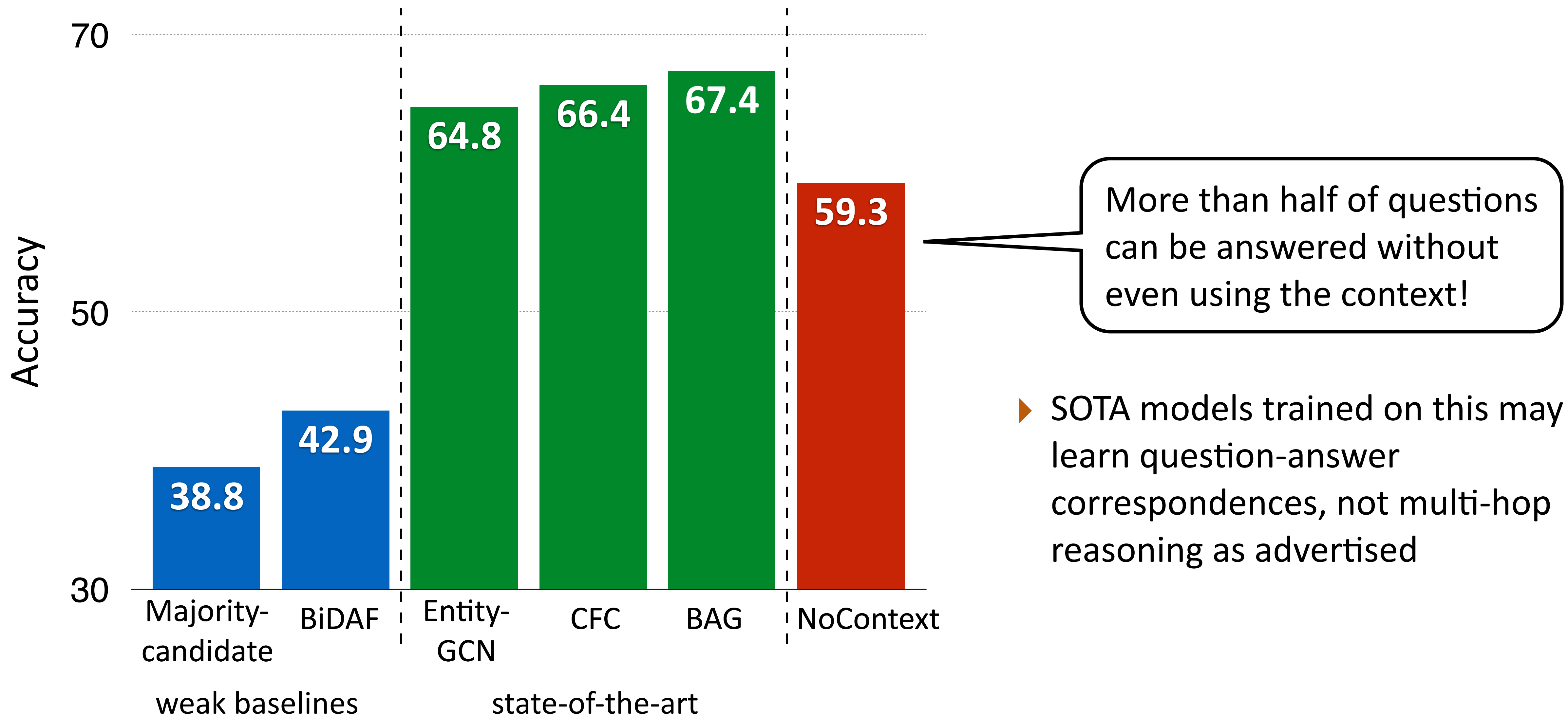


Results on WikiHop





Results on WikiHop





Research Questions

Can a model identify the answer with only a set of candidates?

Government position → Chief of Protocol, actress, singer

Can a model identify where the answer is in a single hop?

Oberoi Family → Dehli

Does the type of supervision matter?

Span-based vs. Multiple-choice



Finding the answer directly

Question: What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell ?

Doc 1 Shirley Temple Black was an American actress, businesswoman, and singer ...
As an adult, she served as Chief of Protocol of the United States

Doc 2 Kiss and Tell is a comedy film in which 17-year-old Shirley Temple acts as
Corliss Archer .

Doc 3 Meet Corliss Archer is an American television sitcom that aired on CBS ...

A: Chief of Protocol

B: businesswoman

C: Actress



Finding the answer directly

Question: What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell ?

Doc 1 Shirley Temple Black was an American actress, businesswoman, and singer ...
As an adult, she s

There is no candidate set on HotpotQA

Doc 2 Kiss and Tell is a comedy film in wh 17-year-old Shirley Temple acts as
Corliss Archer .

Doc 3 Meet Corliss Archer is an American television sitcom that aired on CBS ...

A: Chief of Protocol

B: businesswoman

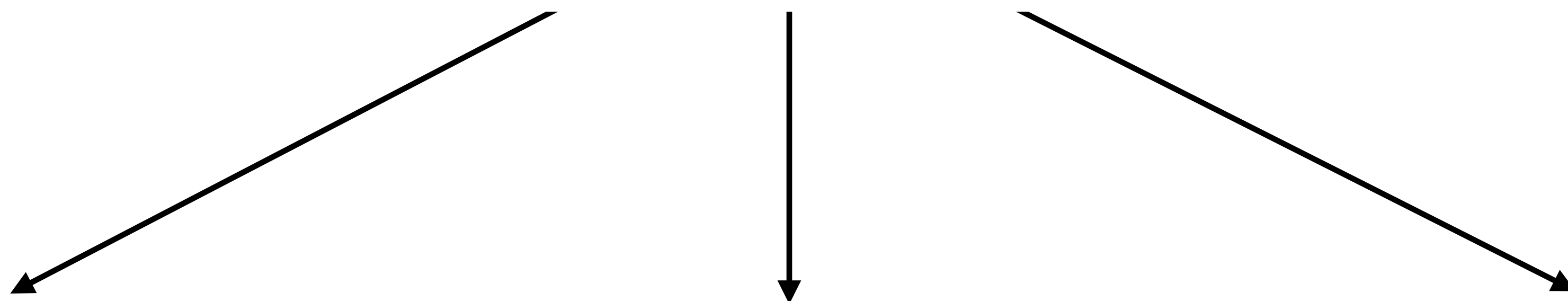
C: Actress



Sentence Factored Model

Find the answer by comparing each sentence with the question **separately!**

Question: The Oberoi family is part of a hotel company that has a head office in what city?



Doc 1

The Oberoi family is an Indian family that is ...

Doc 2

The Oberoi Group is a hotel company with its head office in Delhi.

Doc 3

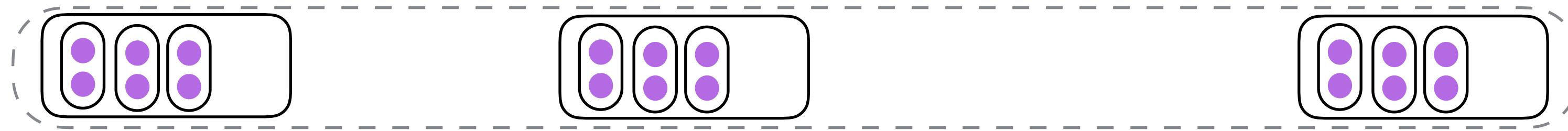
Future Fibre Technologies a fiber technologies company ...



Sentence Factored Model

Answer prediction:

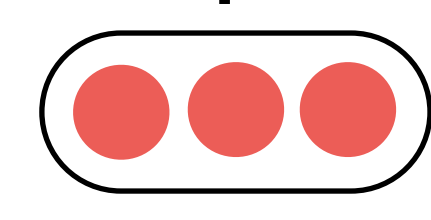
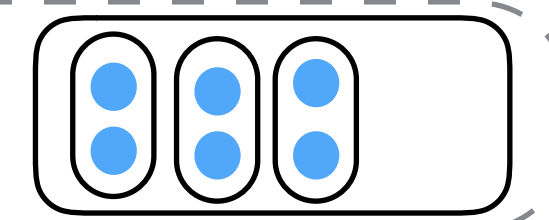
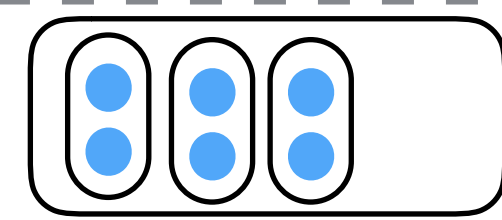
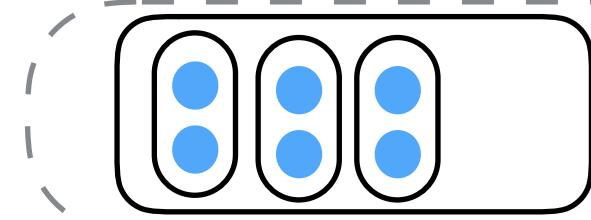
Delhi



BiDAF

BiDAF

BiDAF



The Oberoi
Group ... in Delhi.

Future Fibre
Technologies
is a fibre...

The Oberoi
family ...

The Oberoi family
... what city?

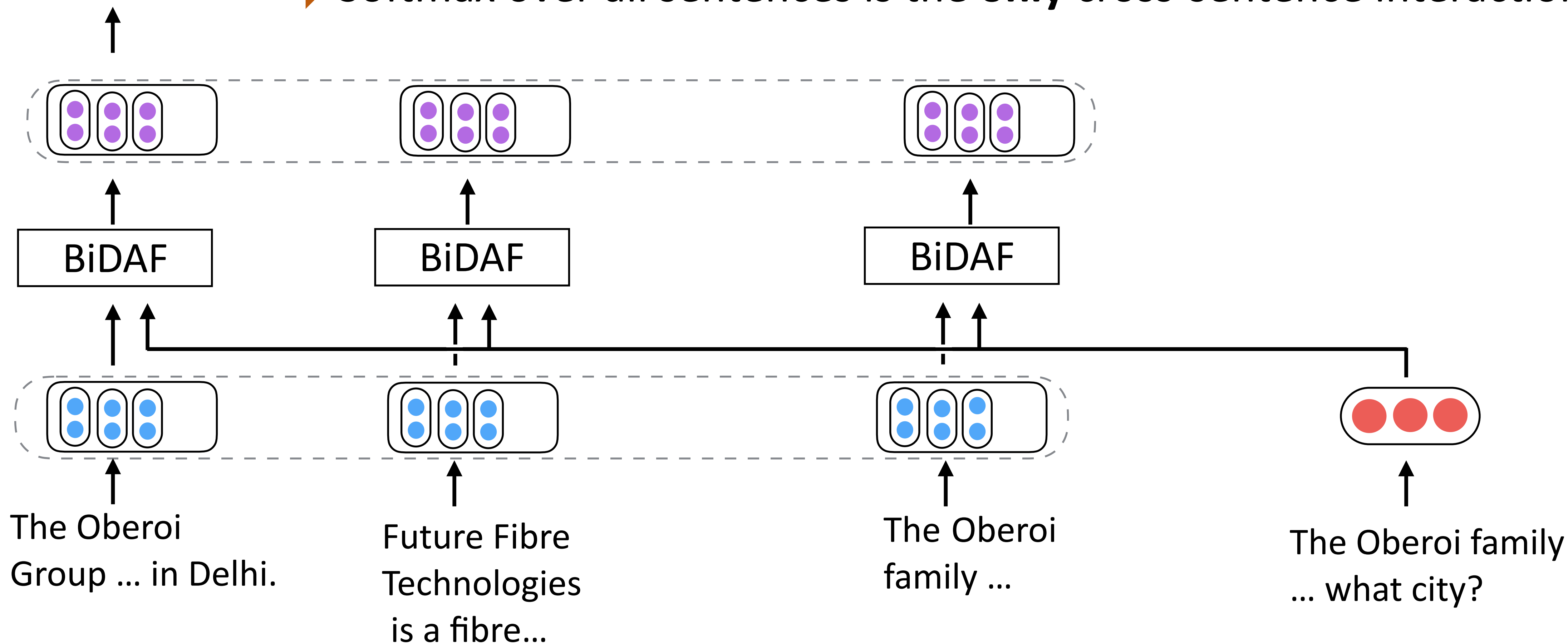


Sentence Factored Model

Answer prediction:

Delhi

► Softmax over all sentences is the **only** cross-sentence interaction

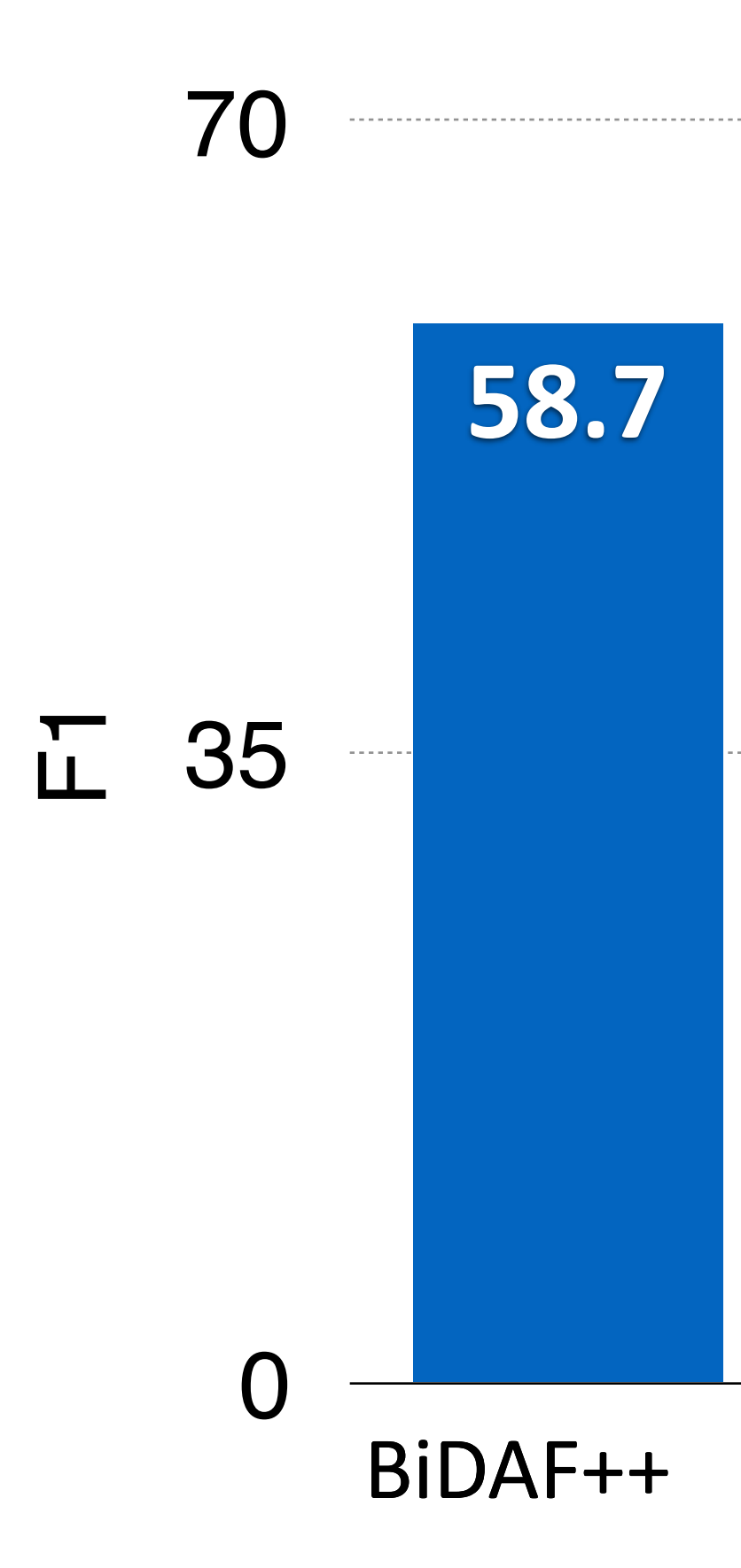




Results on HotpotQA

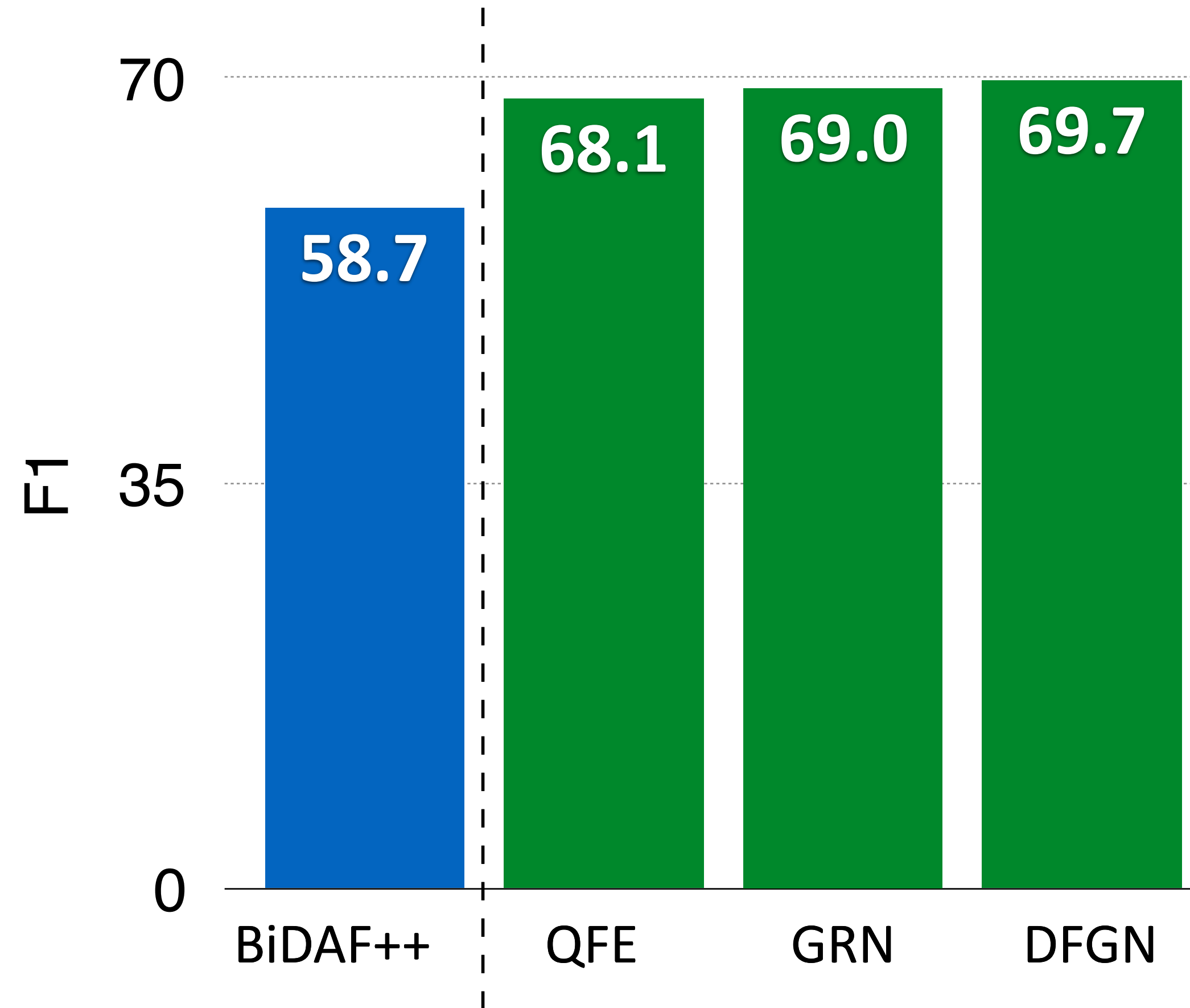


Results on HotpotQA



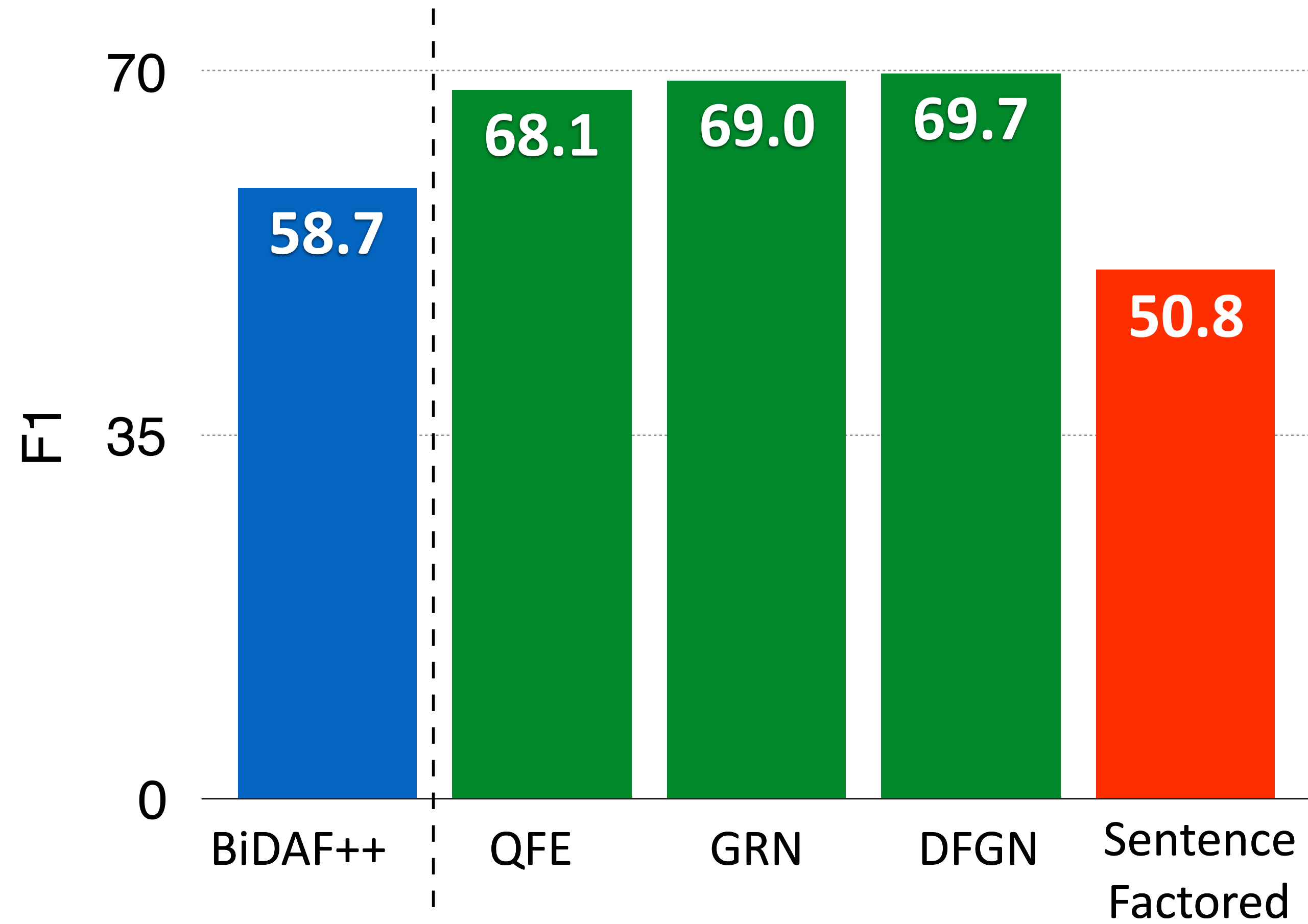


Results on HotpotQA



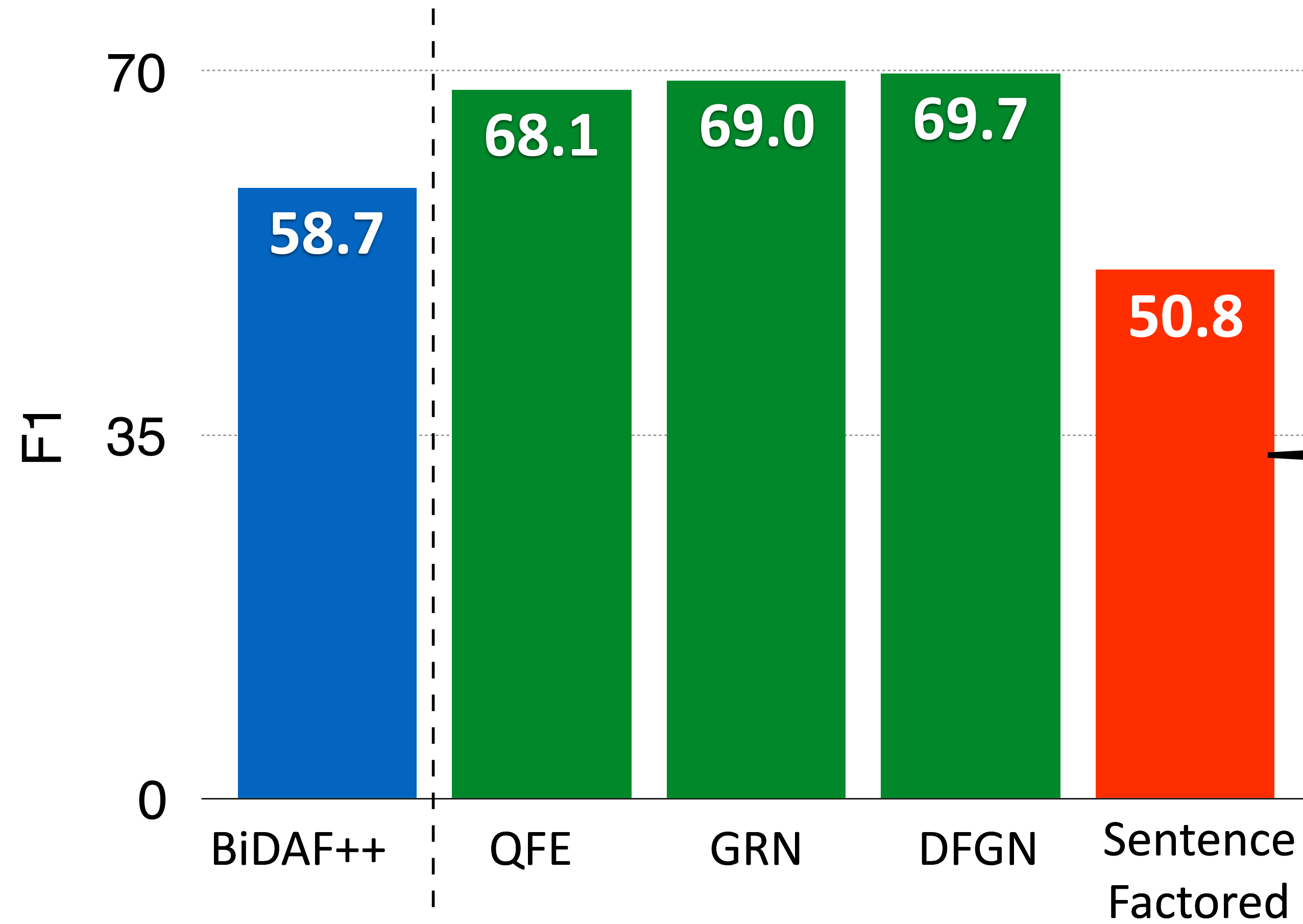


Results on HotpotQA





Results on HotpotQA



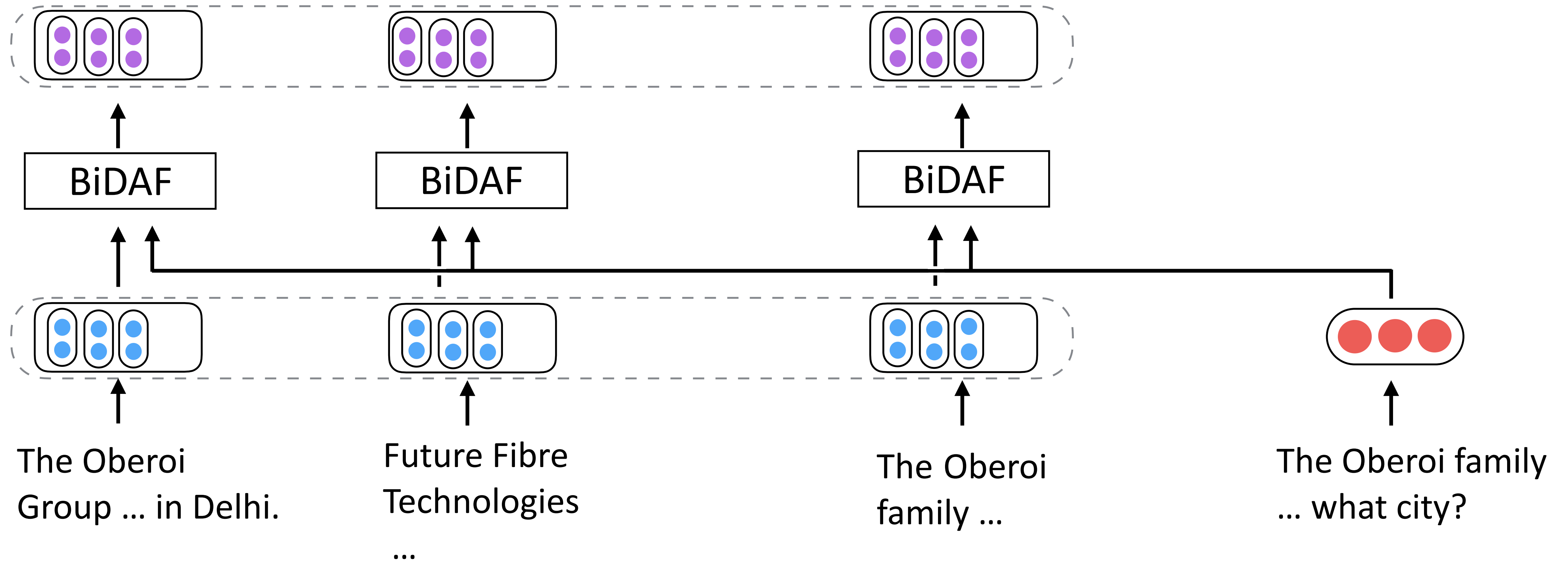
A simple single sentence reasoning model could solve more than half questions on HotpotQA.



Sentence Factored Model

Containing answer:

1



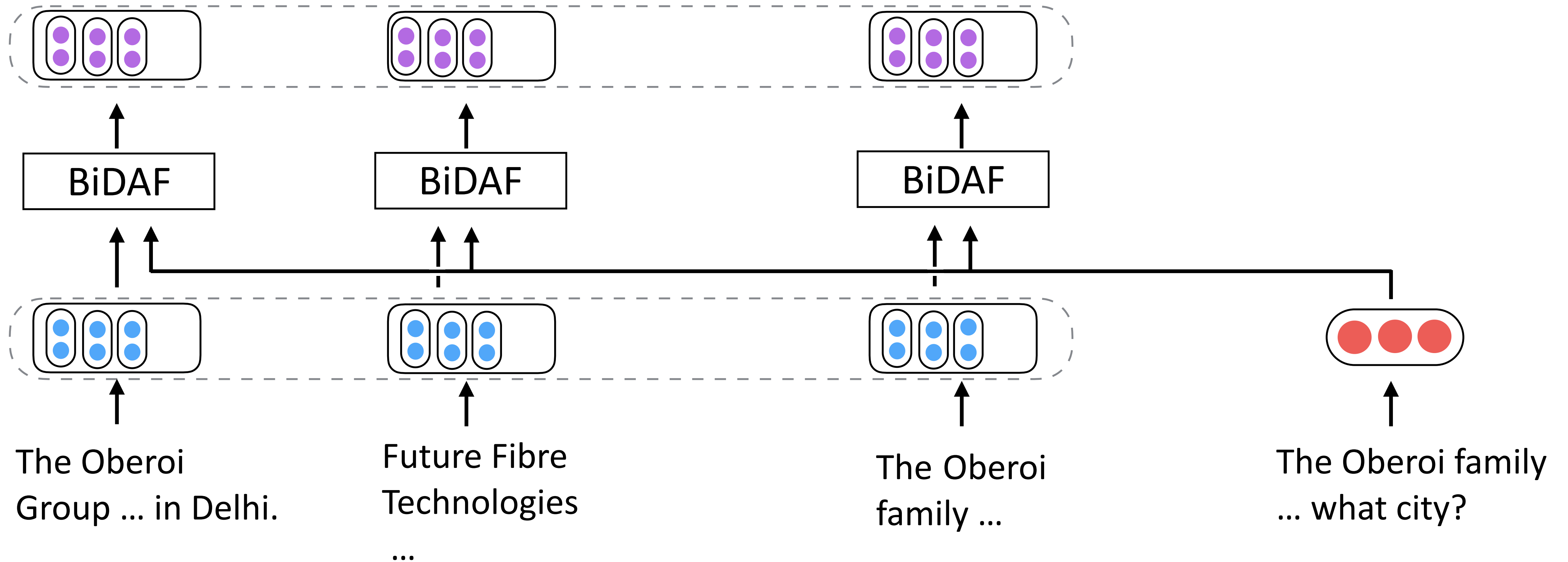


Sentence Factored Model

► Predict which sentence **contains** the answer

Containing answer:

1





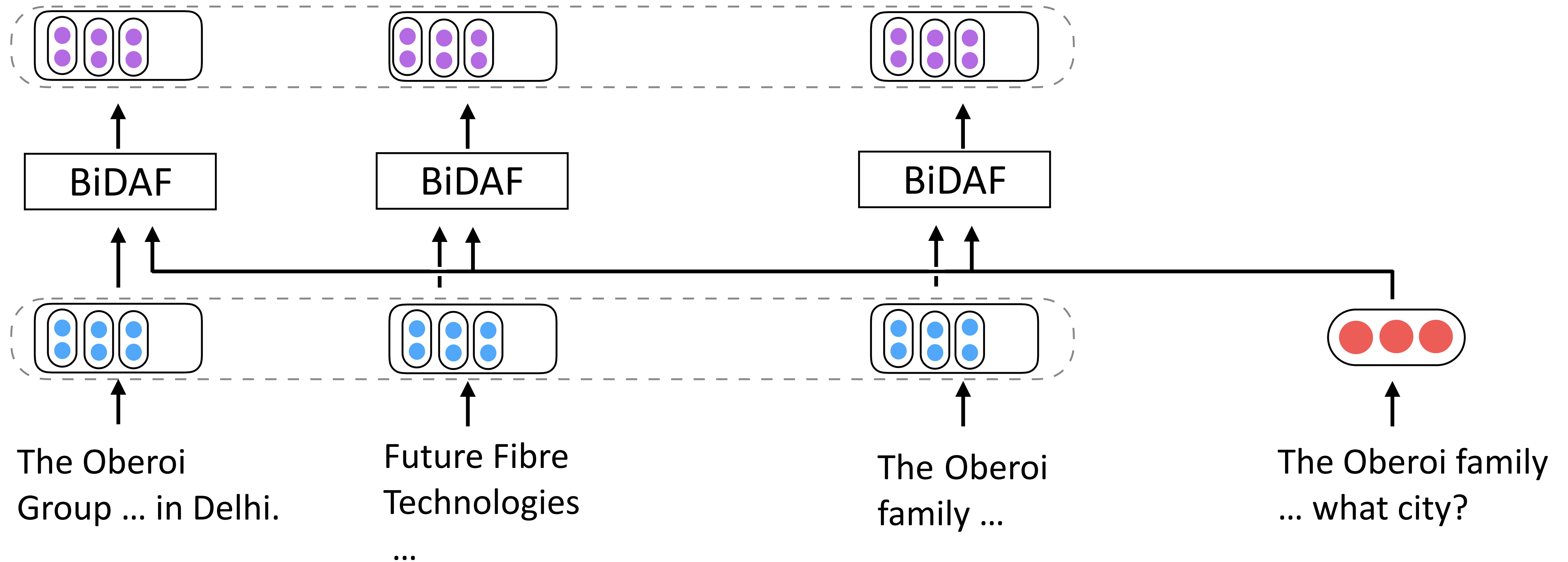
Sentence Factored Model

Containing answer:

1

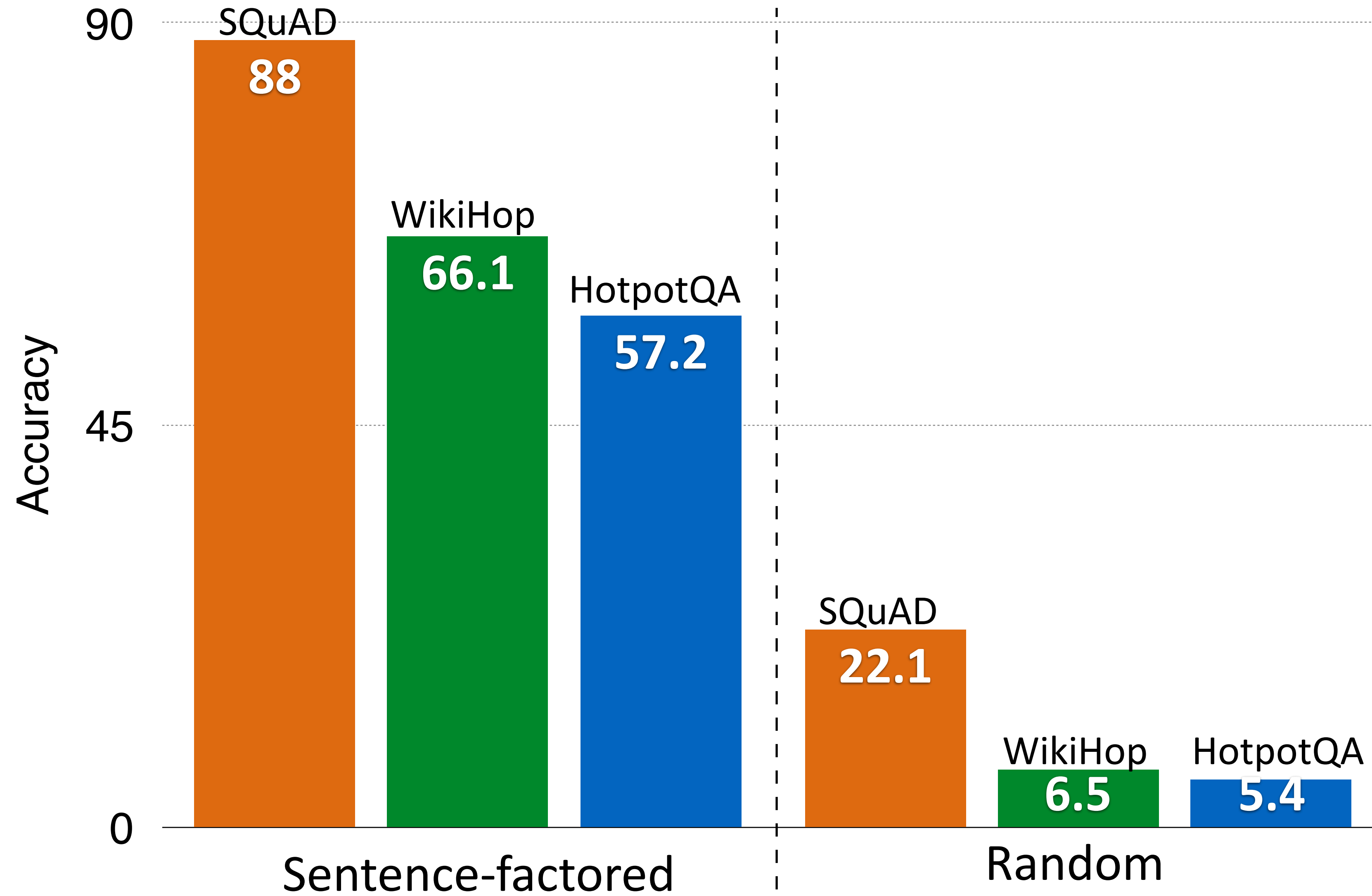


- ▶ Predict which sentence **contains** the answer
- ▶ A simpler way to test how much multi-hop reasoning is needed



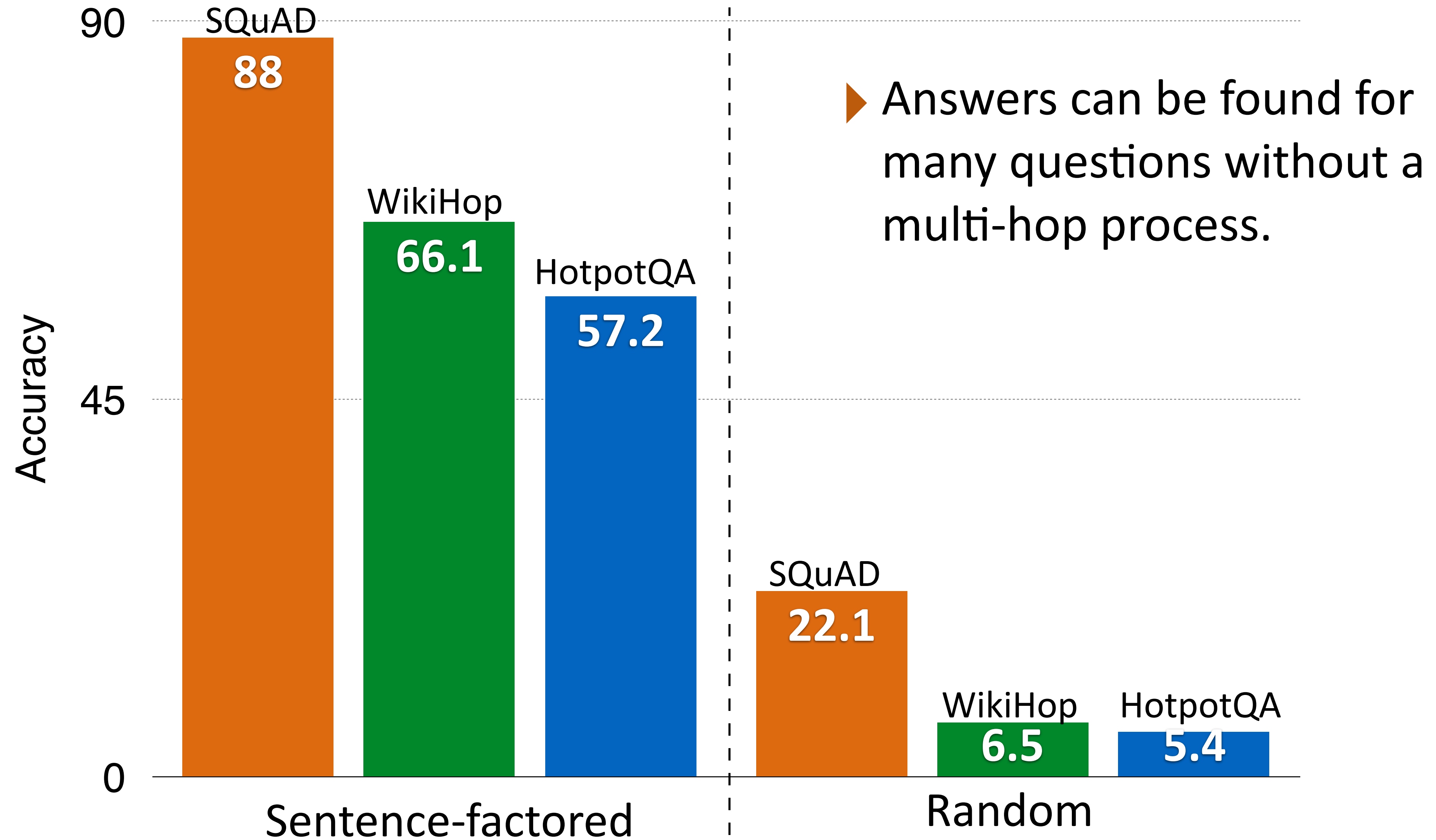


Answer Sentence Prediction Results





Answer Sentence Prediction Results

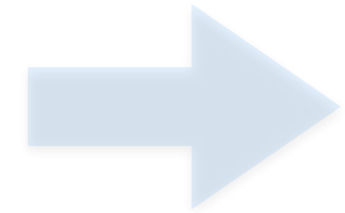




Research Questions

Can a model identify the answer with only a set of candidates?

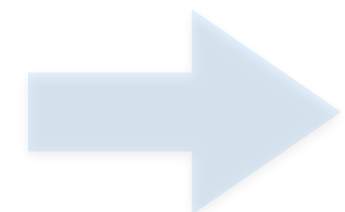
Government position



Chief of Protocol, actress, singer

Can a model identify where the answer is in a single hop?

Oberoi Family



Delhi

Does the type of supervision matter?

Span-based vs. Multiple-choice



Making HotpotQA Multiple-Choice

Question: The Oberoi family is part of a hotel company that has a head office in what city?

Doc 1

The Oberoi family is an Indian family that is famous for its involvement in hotels, namely through The Oberoi Group ...

Doc 2

The Oberoi Group is a hotel company with its head office in **Delhi**.
...

Doc 3

Future Fibre Technologies a fiber technologies company with its **US** head office in **Mountain View**, **Middle East** head office in **Dubai** ...



Making HotpotQA Multiple-Choice

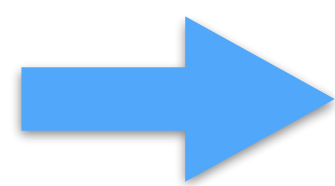
Question: The Oberoi family is part of a hotel company that has a head office in what city?

Doc 1 **The Oberoi family** is an Indian family that is famous for its involvement in hotels, namely through The Oberoi Group ...

Doc 2 **The Oberoi Group** is a hotel company with its head office in **Delhi**.
...

Doc 3 **Future Fibre Technologies** a fiber technologies company with its **US** head office in **Mountain View**, **Middle East** head office in **Dubai** ...

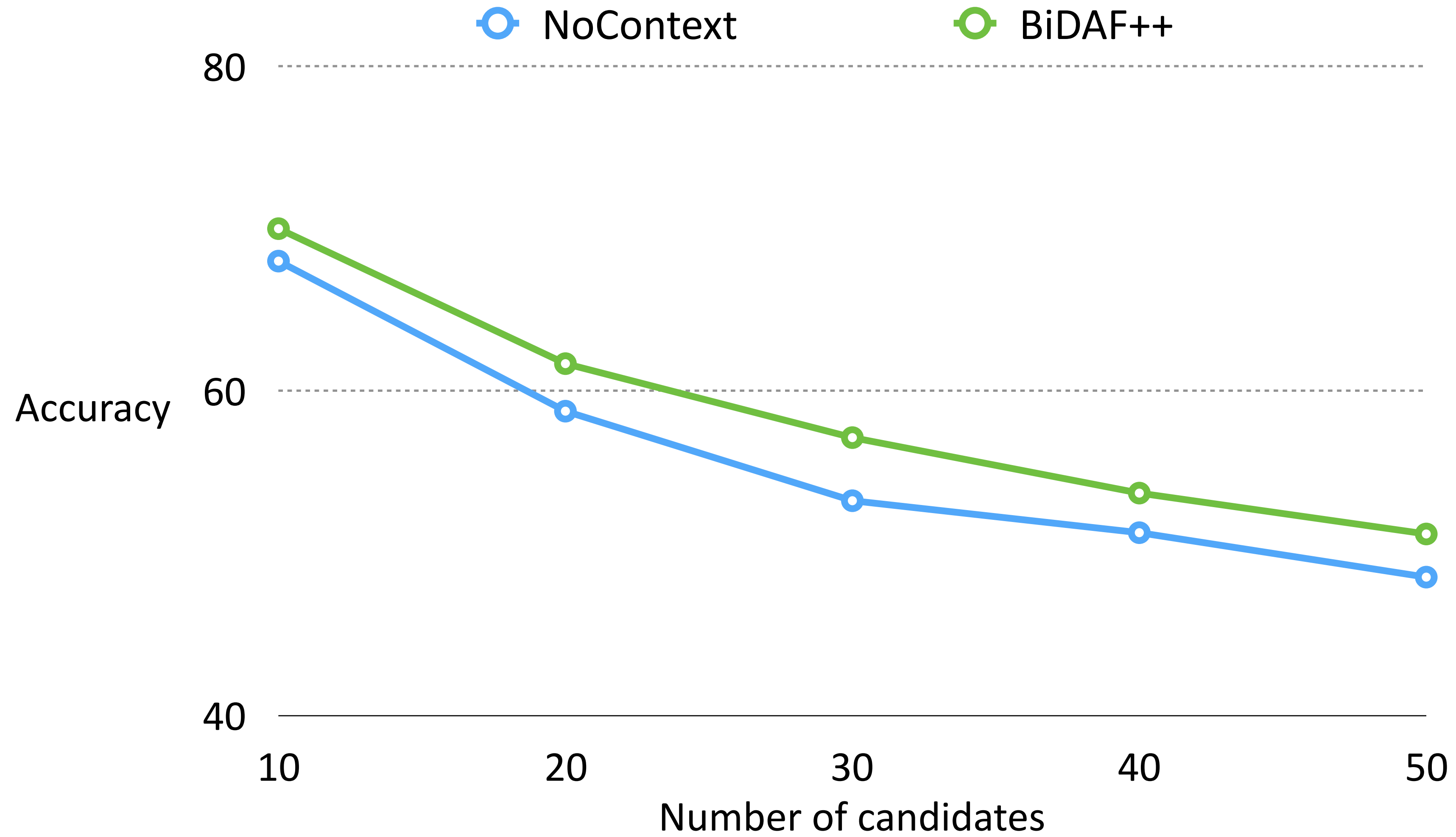
► Randomly choose entities in the documents as candidates.



The Oberoi family, The Oberoi Group, Delhi, Future Fibre Technologies, US, Mountain View, Middle East, Dubai

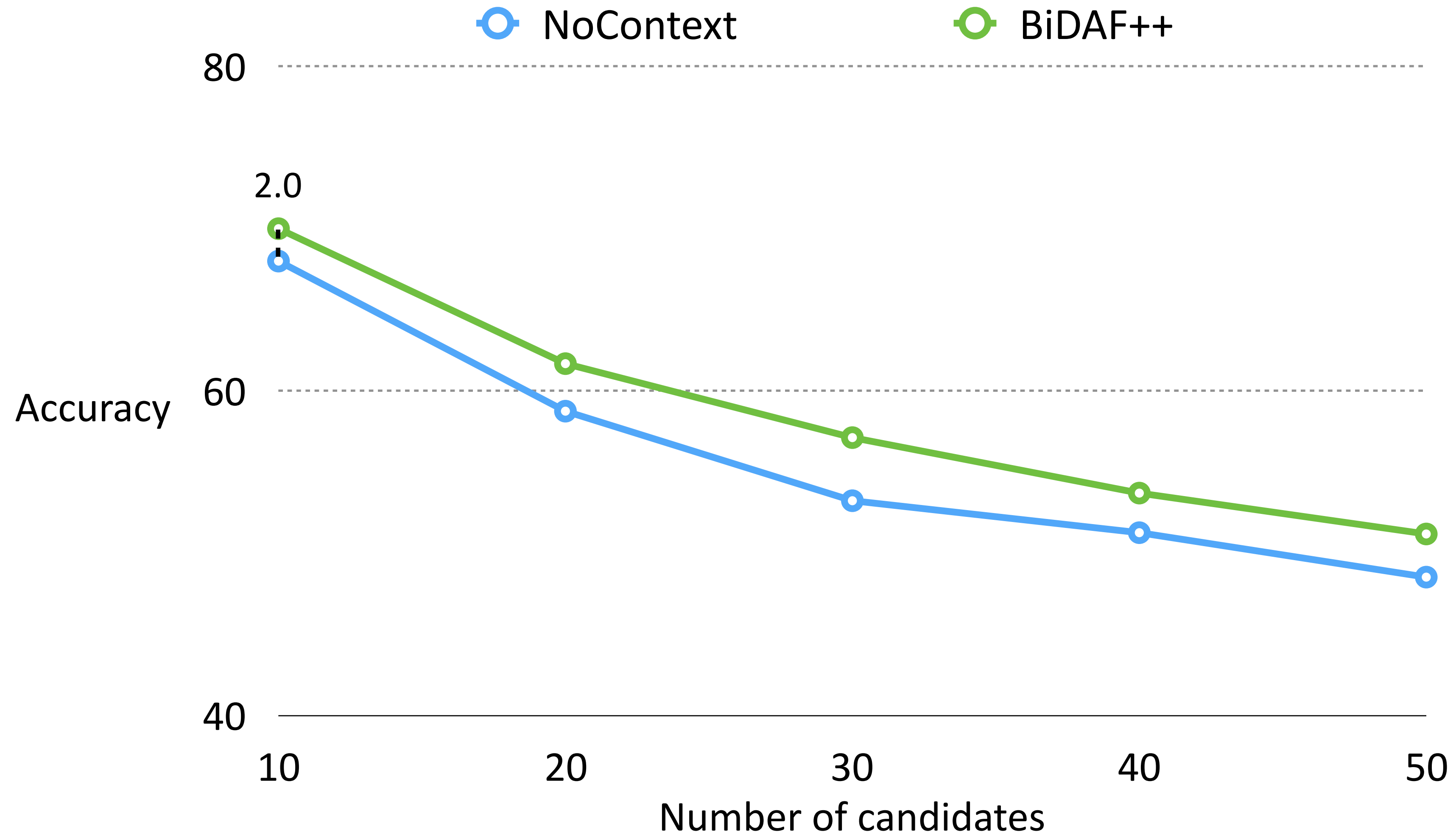


Results of HotpotQA Multiple-Choice



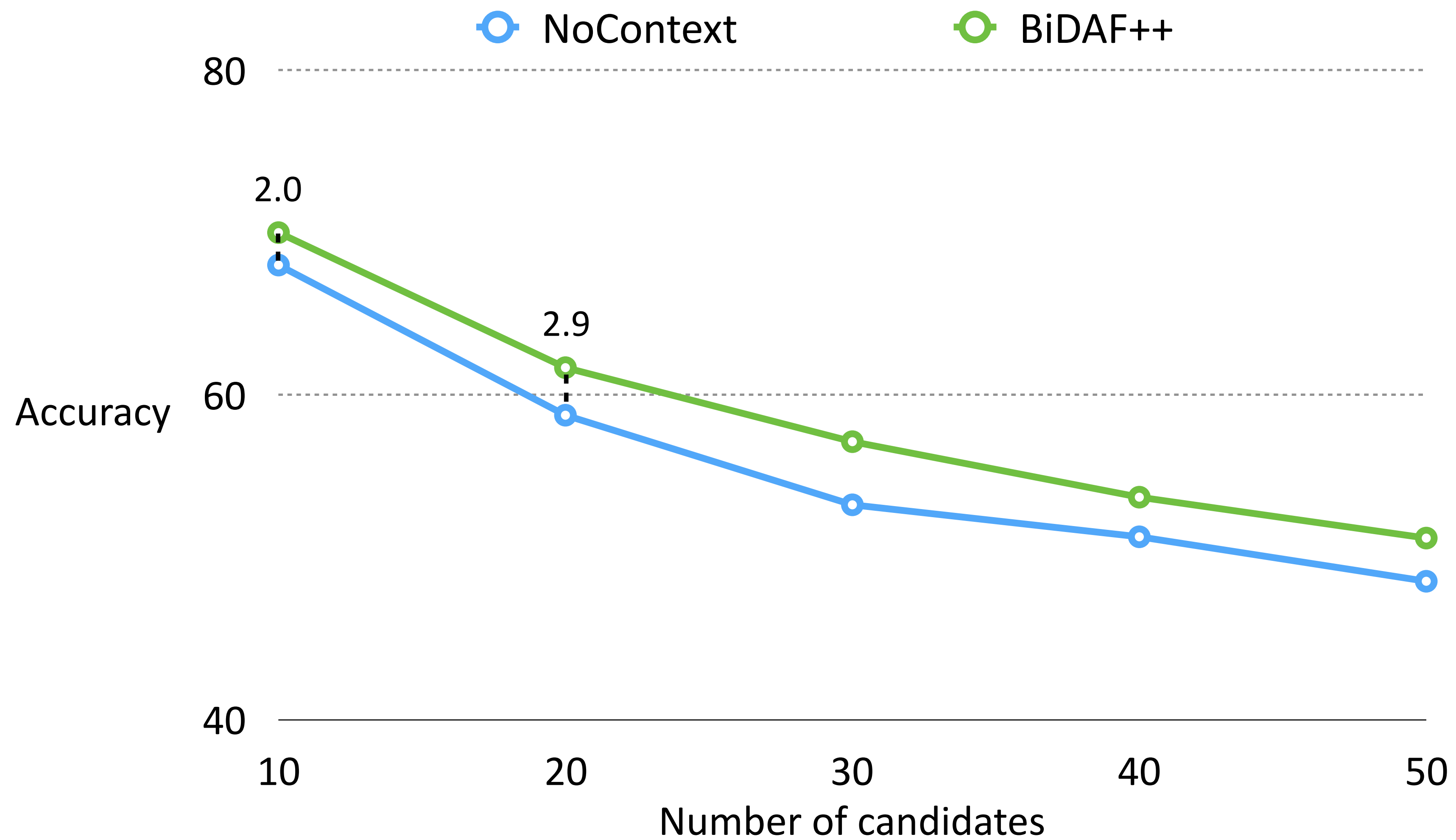


Results of HotpotQA Multiple-Choice



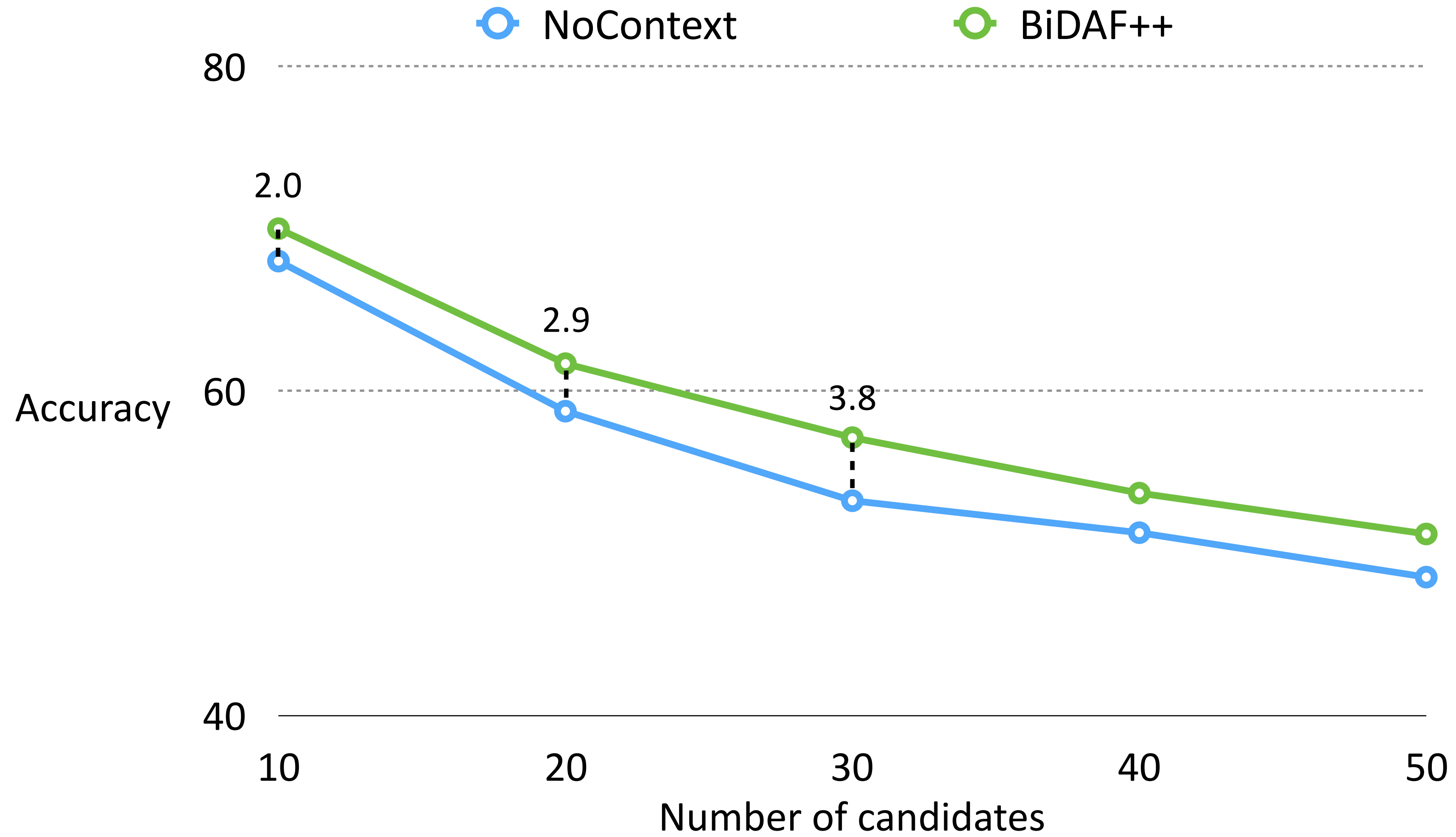


Results of HotpotQA Multiple-Choice



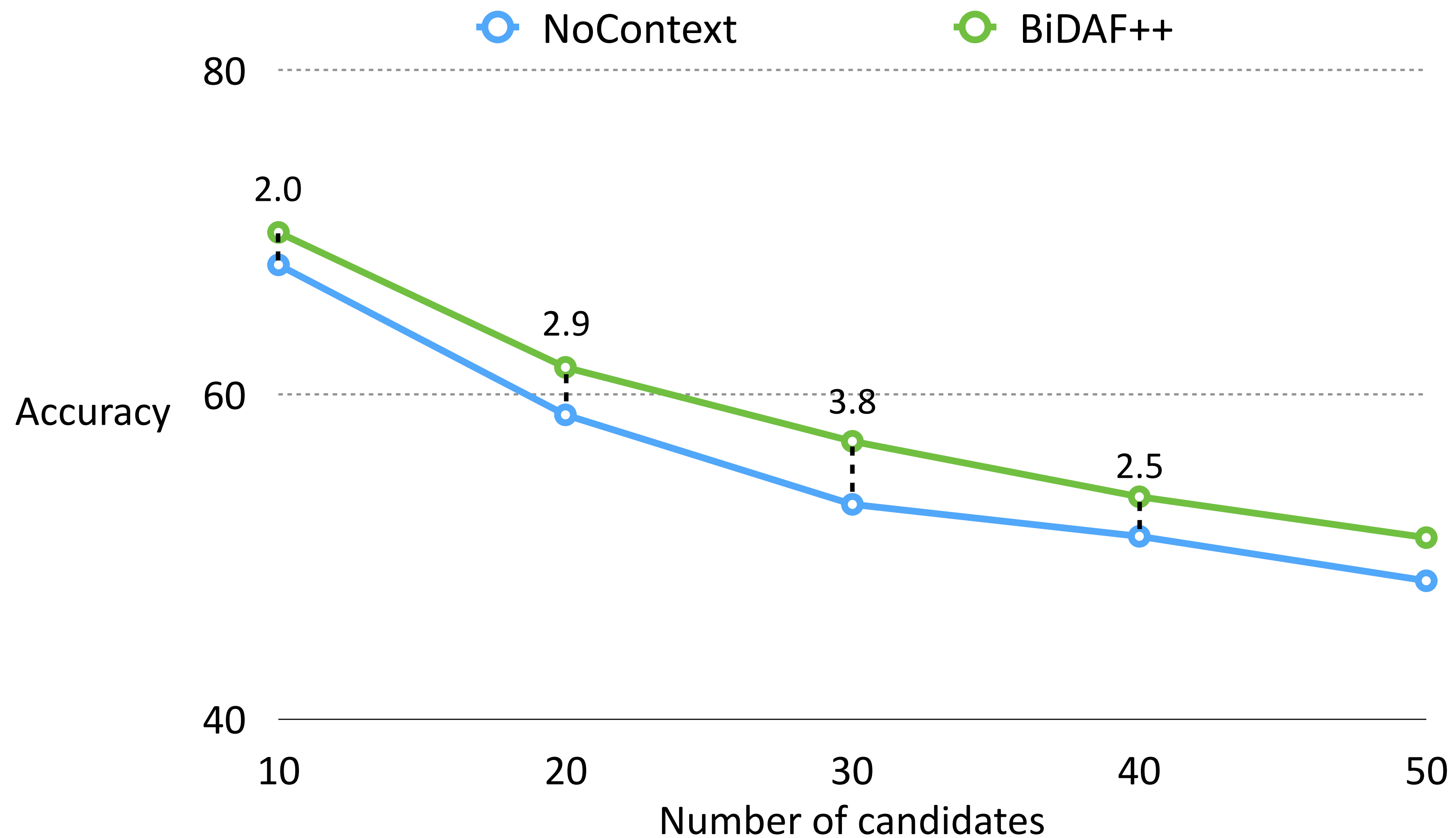


Results of HotpotQA Multiple-Choice



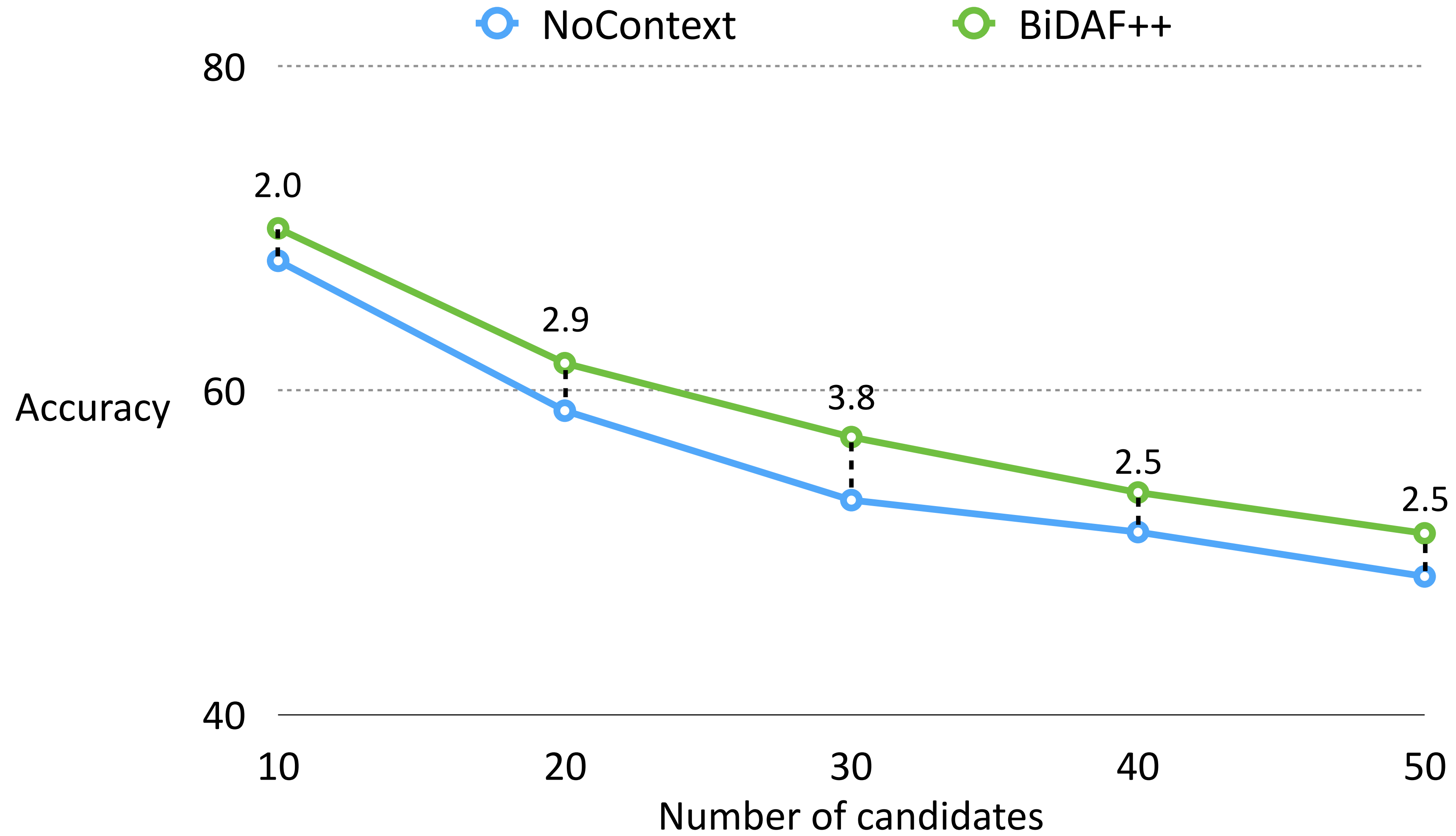


Results of HotpotQA Multiple-Choice



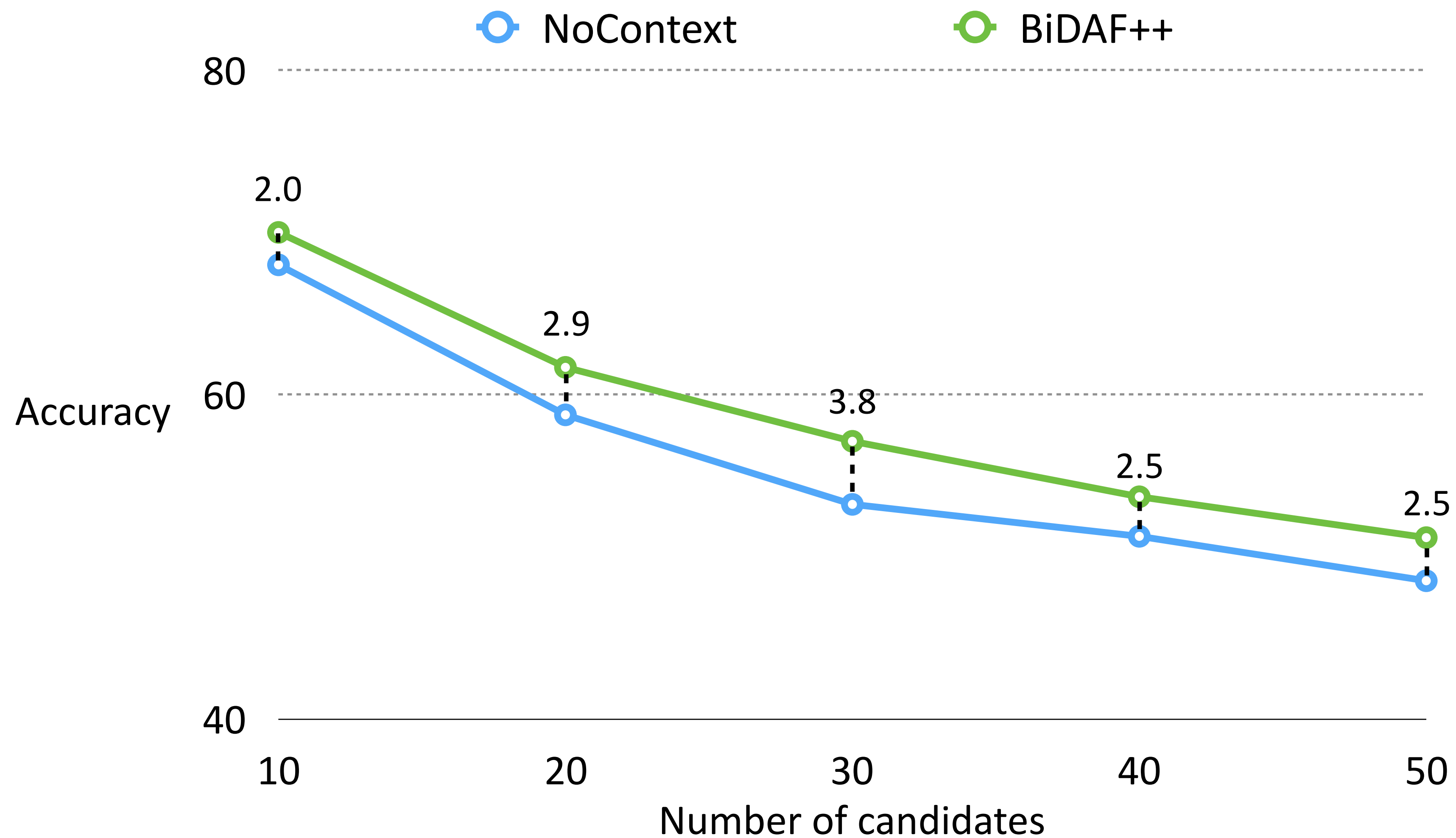


Results of HotpotQA Multiple-Choice





Results of HotpotQA Multiple-Choice



More candidates does make the prediction harder, but BiDAF++ does not learn more than the NoContext baseline



Conclusions

Can a model identify the answer with only a set of candidates?

Can a model identify where the answer is in a single hop?

Does the type of supervision matter?



Conclusions

Can a model identify the answer with only a set of candidates?

- ▶ Models can achieve high performance by “cheating” through extra candidate information.

Can a model identify where the answer is in a single hop?

Does the type of supervision matter?



Conclusions

Can a model identify the answer with only a set of candidates?

- ▶ Models can achieve high performance by “cheating” through extra candidate information.

Can a model identify where the answer is in a single hop?

- ▶ More than half of the questions in WikiHop and HotpotQA can be answered in a single hop.

Does the type of supervision matter?



Conclusions

Can a model identify the answer with only a set of candidates?

- ▶ Models can achieve high performance by “cheating” through extra candidate information.

Can a model identify where the answer is in a single hop?

- ▶ More than half of the questions in WikiHop and HotpotQA can be answered in a single hop.

Does the type of supervision matter?

- ▶ Span-based datasets may be less hackable.



Conclusions

Can a model identify the answer with only a set of candidates?

- ▶ Models can achieve high performance by “cheating” through extra candidate information.

Can a model identify where the answer is in a single hop?

- ▶ More than half of the questions in WikiHop and HotpotQA can be answered in a single hop.

Does the type of supervision matter?

- ▶ Span-based datasets may be less hackable.

*No results for WikiHop-Masked or HotpotQA fullwiki settings



Takeaways

When constructing multi-hop based dataset:



Takeaways

When constructing multi-hop based dataset:

- ▶ If using multiple-choice supervision, check the no-context baseline to see whether there are strong correlations between question and candidates.



Takeaways

When constructing multi-hop based dataset:

- ▶ If using multiple-choice supervision, check the no-context baseline to see whether there are strong correlations between question and candidates.
- ▶ Do an adversarial test using a sentence-factored model.



Takeaways

When constructing multi-hop based dataset:

- ▶ If using multiple-choice supervision, check the no-context baseline to see whether there are strong correlations between question and candidates.
- ▶ Do an adversarial test using a sentence-factored model.

Thank you!

Q & A