A Multi-Agent Communication Framework for Question- Worthy Phrase Extraction and Question Generation

AAAI 2019

Siyuan Wang[1], Zhongyu Wei[1], Zhihao Fan[1], Yang Liu[2], Xuanjing Huang[3]
1 School of Data Science, Fudan University, China
2 LAIX Inc., China
3 School of Computer Science, Fudan University, China

Question Generation (for text)

- Given a sentence or paragraph, construct questions automatically

Sentence:
CBS provided digital streams of the game via *CBSSports.com*, and the CBS Sports apps on tablets, Windows *10*, *Xbox One* and other digital media players (such as Chromecast and Roku).

Questions:
- What CBS website provided a stream?
- What version of Windows supported the CBS sports app?
- On what game console was the CBS Sports app available?

Introduction

- Two subtasks in Question Generation
  - What to say: determine the targets that should be asked
  - How to say: produce the surface-form of the question

This paper focus on

Most papers focus on

Motivation

Sentence:
CBS provided digital streams of the game via *CBSSports.com*, and the CBS Sports apps on tablets, Windows *10*, *Xbox One* and other digital media players (such as Chromecast and Roku).

Questions:

- What CBS website provided a stream?
- What version of Windows supported the CBS sports app?
- On what game console was the CBS Sports app available?

Auxiliary Task: question-worthy phrases extraction

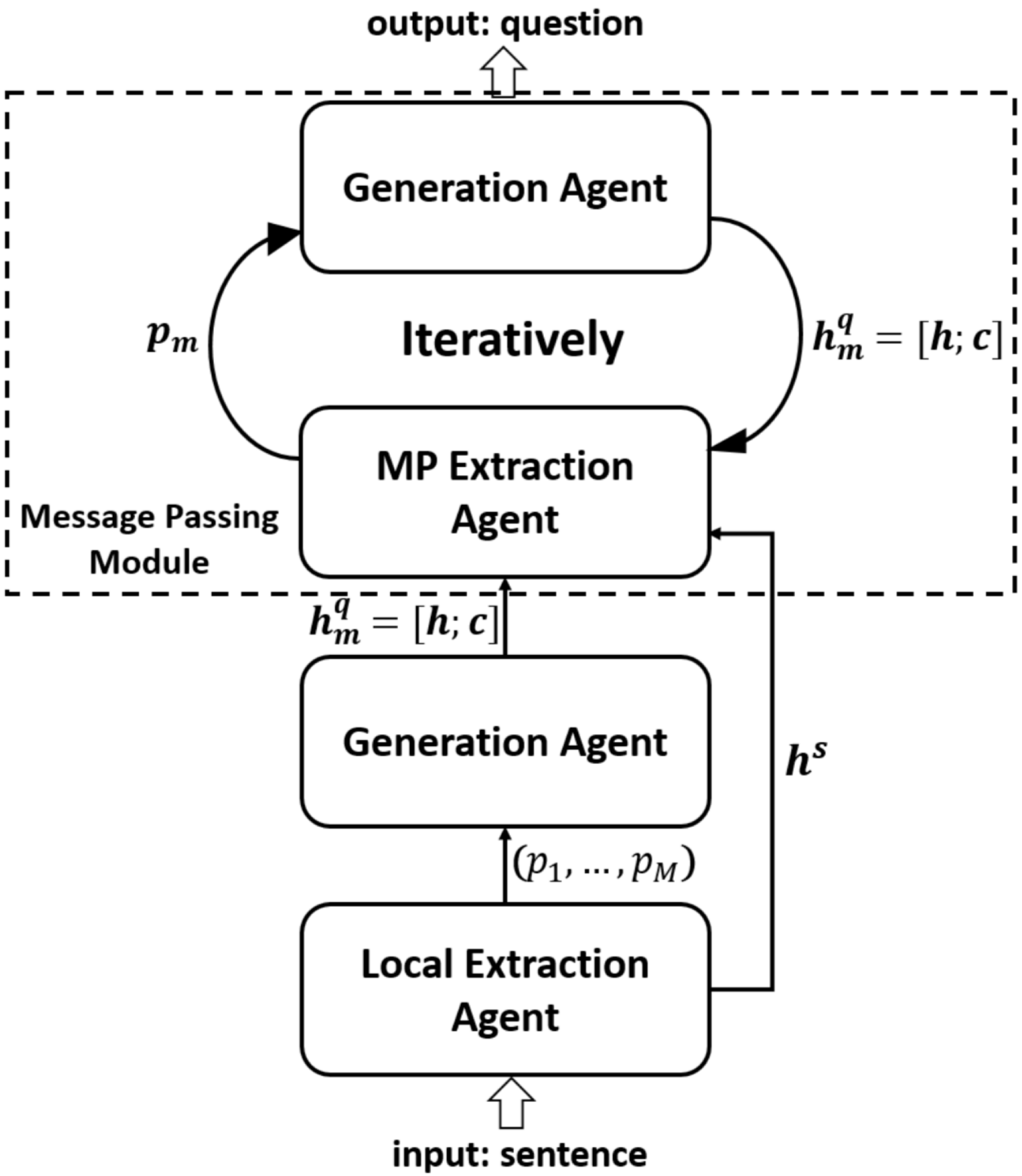
Main Task: question generation

- Not all content pieces in the input are significant, the authors propose to use **question-worthy phrases** to identify which phrases are worthwhile to be asked about.
- Moreover, if there are several focuses in an input and we extract question-worthy phrases, we can then generate **various questions**

Contributions

- Propose to generate multiple questions given input sentence **without ground-truth answers**.
- Extract **question-worthy phrases** from the input sentence and generate questions based on such information.
- Developing a **multi-agents communication framework** to learn the two tasks simultaneously.

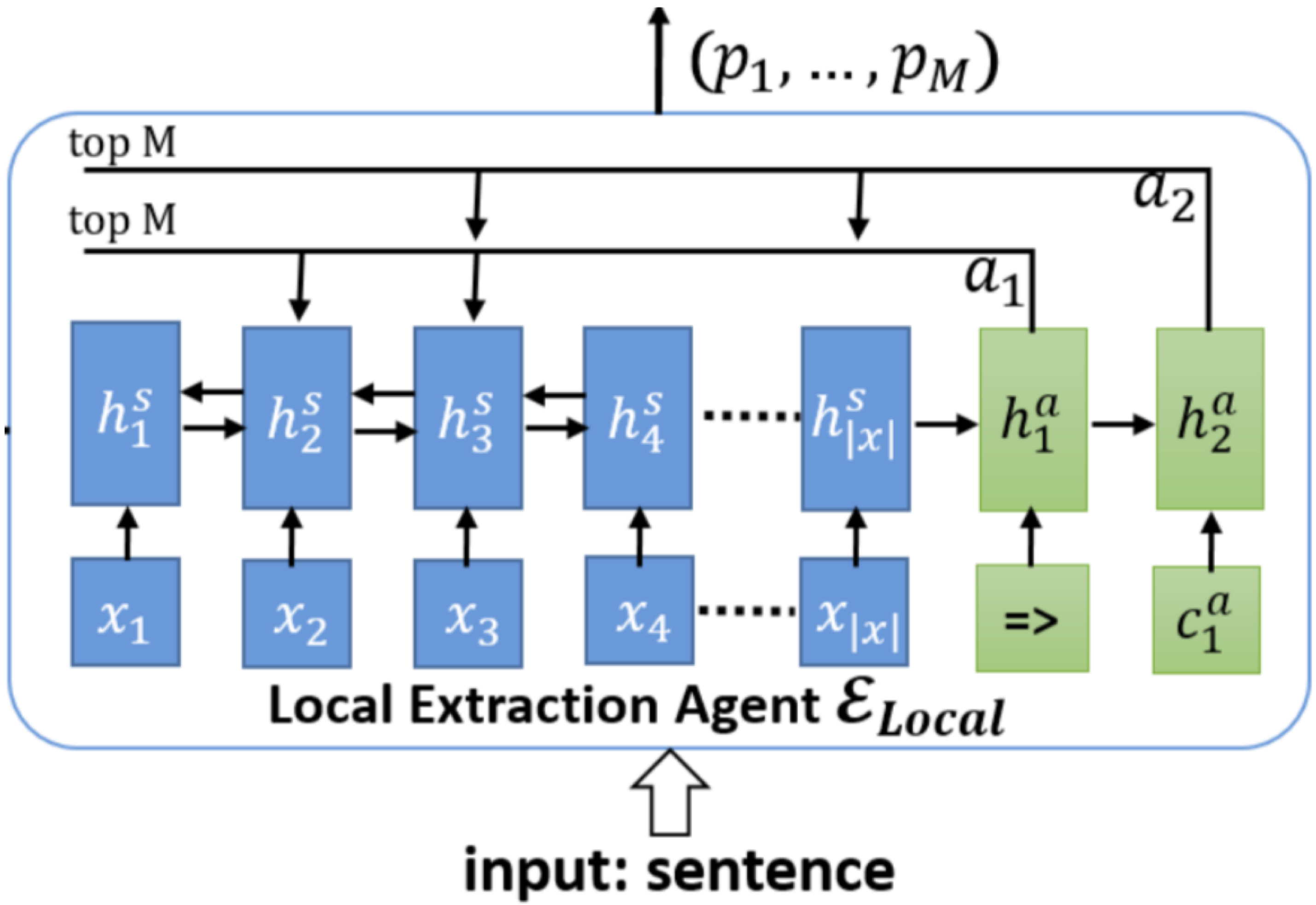
Framework



output: question
Generation Agent
p_m
Iteratively
$h_m^q = [h; c]$
MP Extraction Agent
Message Passing Module
$h_m^q = [h; c]$
Generation Agent
h^s
$(p_1, \ldots, p_M)$
Local Extraction Agent
input: sentence

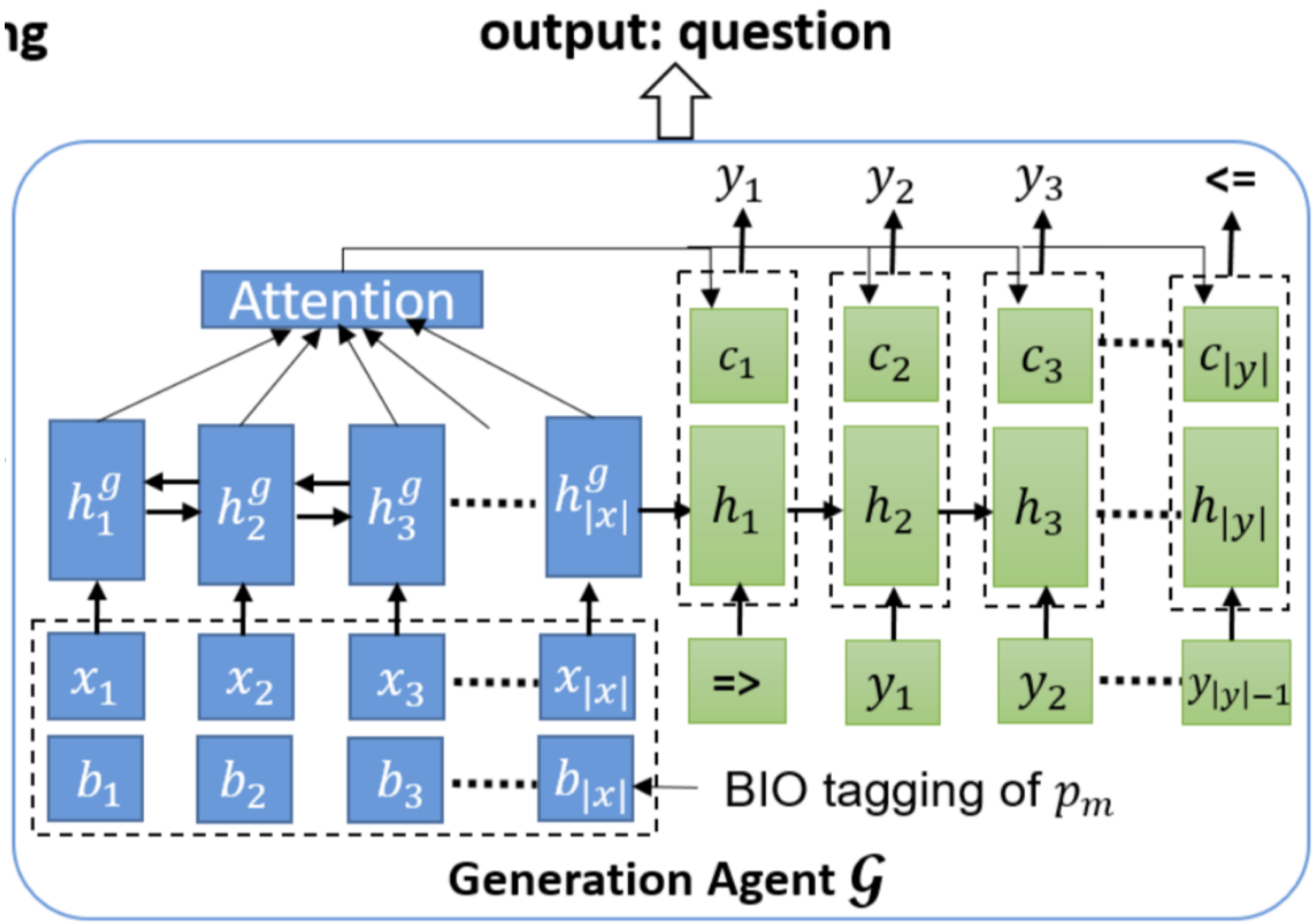
Local extraction agent

- The Boundary Model of Pointer Network
- Pick M pairs of start and end index of phrases



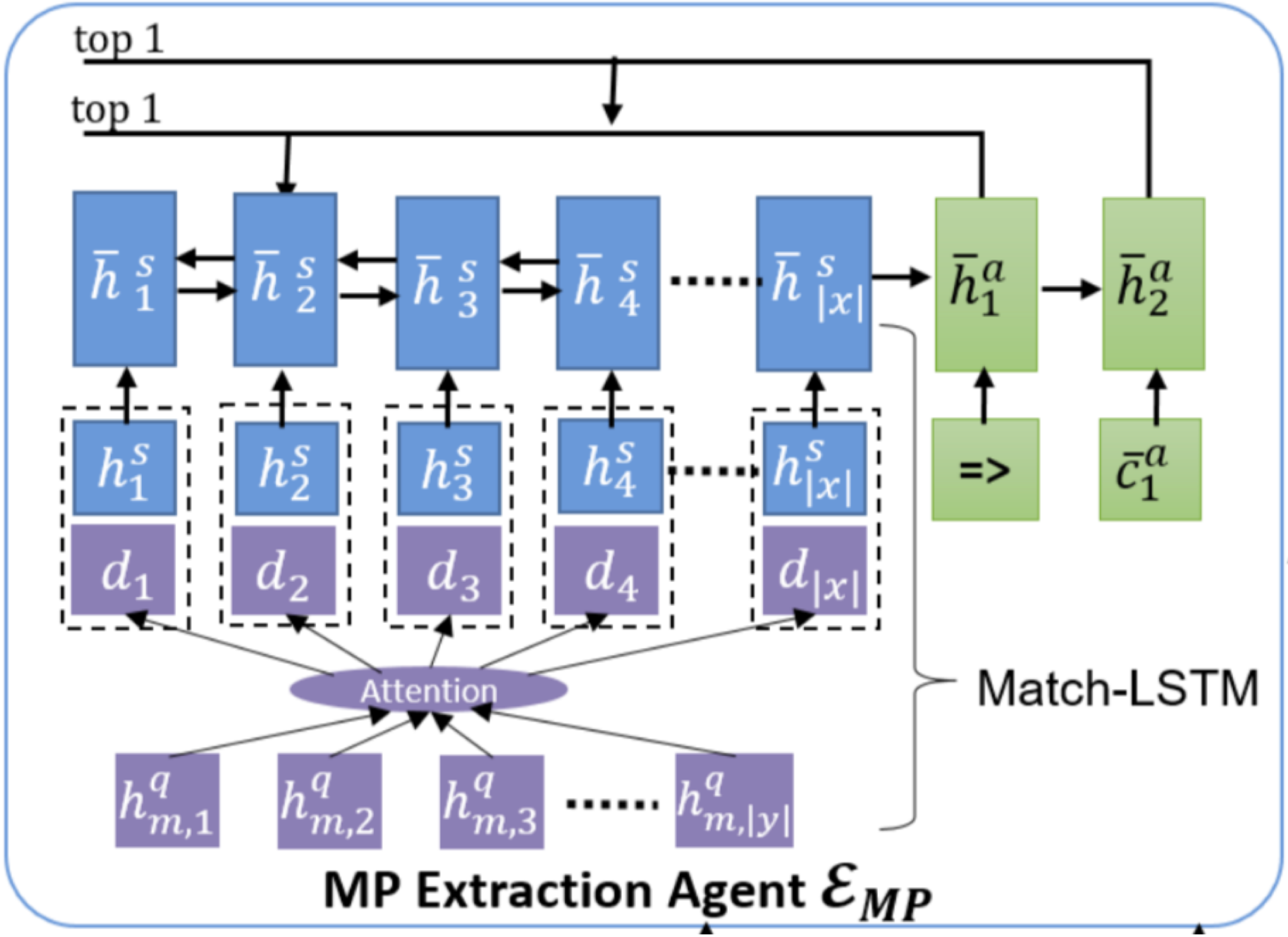
Generation agent

- Seq2Seq Model + Attention Mechanism
- Each time we take both the sentence and an extracted phrase as input, to generate a question

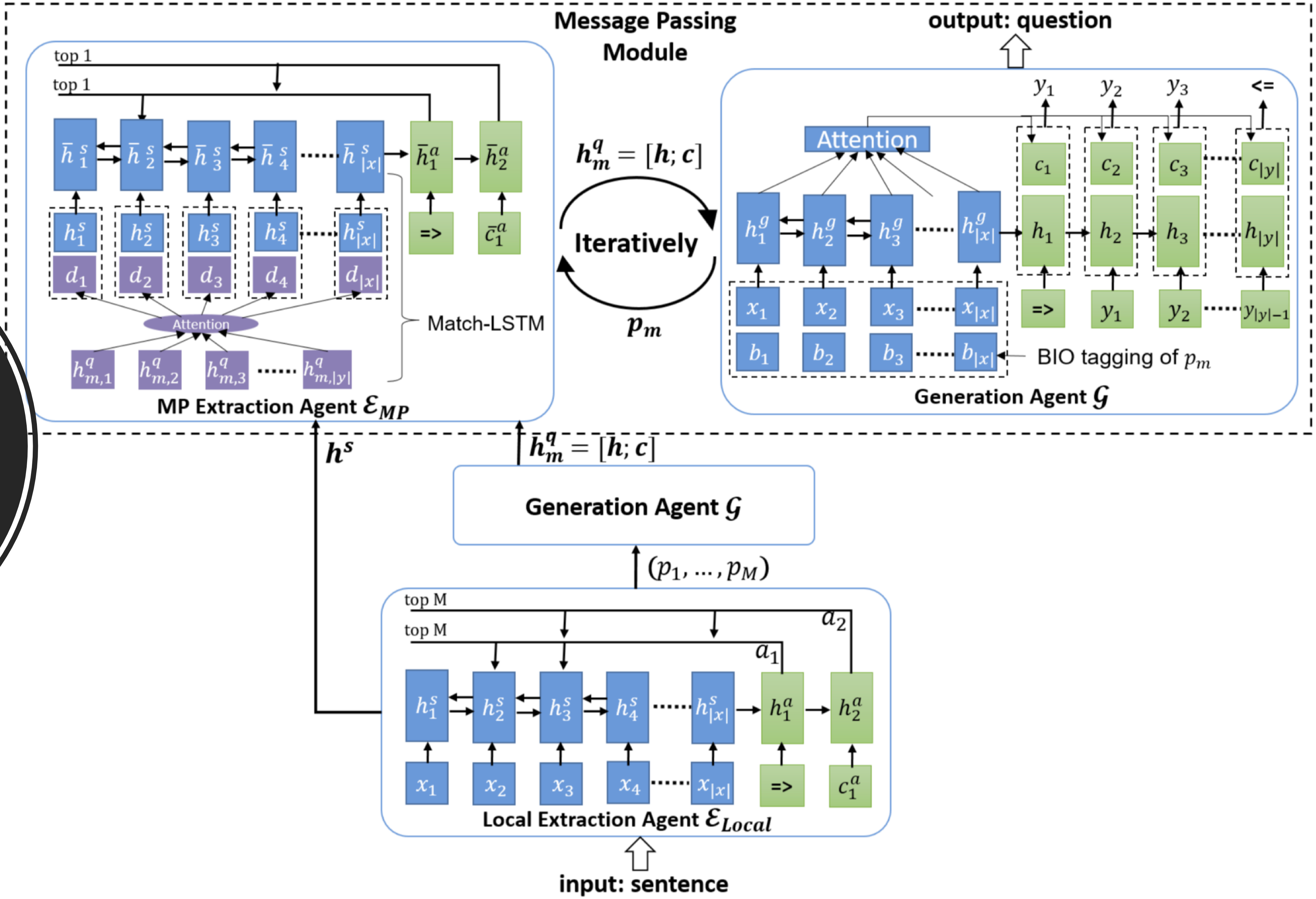


Message Passing (MP) Extraction agent

- Match-LSTM + Pointer Network
- Use representations of input tokens from generation agent as auxiliary information



Framework



Message Passing Module
output: question
top 1
top 1
Match-LSTM
Attention
MP Extraction Agent $\mathcal{E}_{MP}$
$h_m^q = [h; c]$
Iteratively
p_m
Attention
BIO tagging of p_m
Generation Agent $\mathcal{G}$
h^s
$h_m^q = [h; c]$
Generation Agent $\mathcal{G}$
$(p_1, \ldots, p_M)$
top M
top M
a_2
a_1
Local Extraction Agent $\mathcal{E}_{Local}$
input: sentence

Experiments

- SQuAD Dataset
  - Answers, extractive from sentences, are treated as target question-worthy phrases
  - More than 30% sentences have multiple questions (One-to-Many here)

# of questions	# of sentences	percentage
1	41,356	67.11%
2	14,499	23.53%
3	3,921	6.36%
4	1,198	1.95%
$\geqslant 5$	649	1.05%
in total	61,623	100%

Table 1: Distribution of number of questions per sentence in our dataset.

Comparison of Extraction Models

- $\mathcal{E}_{NER}$: take recognized name entities as question-worthy phrases
- $\mathcal{E}_{Local}$: the extraction agent in local layer
- $\mathcal{E}_{MP}$: the extraction agent in message passing layer

Model	EM	F1	avg. num
$\mathbf{E}_{\mathbf{NER}}$	13.12%	17.33	0.86
$\mathbf{E}_{\mathbf{Local}}$	24.27%	38.63	1.43
$\mathbf{E}_{\mathbf{MP}}$	$\underline{\mathbf{35.77\%}}$	$\underline{\mathbf{46.71}}$	1.38

Table 2: Evaluation results of different phrase extraction models. (underline: diff. with both comparison models ($\mathbf{E}_{\mathbf{NER}}$, $\mathbf{E}_{\mathbf{Local}}$) $p < 0.01$; **Bold**: the best performance in the column)

Comparison of Generation Models

Model	BLEU 1	BLEU 2	BLEU 3	BLEU 4	METEOR	ROUGE$_L$
NQG$_{\mathbf{Rule}}$	38.15	21.03	14.15	9.98	13.38	29.00
NQG$_{\mathbf{Base}}$	43.83	23.80	14.46	9.05	14.63	36.50
NQG$_{\mathbf{NER}}$	44.00	23.79	14.52	9.22	14.89	36.32
NQG$_{\mathbf{Local}}$	44.36	24.58	15.23	9.76	15.15	37.00
NQG$_{\mathbf{MP}}$	**45.70***	**25.87***	**16.33***	**10.56***	**15.76***	**38.09***
NQG$_{\mathbf{G-t}}$	47.49	27.81	17.9	11.81	16.84	40.23

Table 3: Evaluation results of different question generation models in terms of BLEU 1-4, METEOR and ROUGE$_L$. (underline: diff. with all the comparison models (**NQG$_{\mathbf{Rule}}$**, **NQG$_{\mathbf{Base}}$**, **NQG$_{\mathbf{NER}}$**, **NQG$_{\mathbf{Local}}$**) $p < 0.01$; *: $p < 0.05$; **Bold**: the best performance for each column)

Case Study

Sample 1

Input: the panthers finished the regular season with a 15 – 1 record , and quarterback cam newton was named the nfl most valuable player (mvp) .

Phrases

Ground-truth: 15 – 1, quarterback cam newton.
NER: panthers, ⟨*blank*⟩.
E_{MP}: 15, quarterback cam newton.

Questions

Ground-truth: what was the ratio in 2015 for the carolina panthers during their regular season ? which carolina panthers player was named most valuable player ?
$\mathrm{NQG}_{\mathrm{NER}}$: who wons the regular season ? what was the regular session in the afl ?
$\mathrm{NQG}_{\mathrm{MP}}$: how many wins did the panthers win during the regular season ? who was named the nlf most valuable player?

Sample 2

Input: next to the main building is the basilica of the sacred heart .

Phrases

Answers: main building.
NER: sacred heart
E_{MP}: next to the main building.

Questions

Ground Truth: the basilica of the sacred heart at notre dame is beside to which structure ?
$\mathrm{NQG}_{\mathrm{NER}}$: what is next to main building ?
$\mathrm{NQG}_{\mathrm{MP}}$: where is the basilica of prayer ?