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

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

AAAI 2019

Outlines

- **Introduction and Motivation**
- Related Work
- Framework
- Experiments
- Future Work

Introduction

- Question Generation for text:
  - Given a sentence or paragraph, ask natural questions

Oxygen is used in cellular respiration and released by photosynthesis, which uses the energy of sunlight to produce oxygen from water.

- What life process produces oxygen in the presence of light?

- Photosynthesis uses which energy to form oxygen from water?

- From what does photosynthesis get oxygen?

Introduction

- Why question generation?
  - **For advance AI:** initializing a conversation for chatbot
  - **For education:** play as a tutor, or generate questions for reading comprehension
  - **For research:** construct corpus for the NLP task of QA

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?

- Sentence-level generated question usually focuses on a specific fragment of sentence that is deemed to contain significant information.
- However not all content pieces in the input are significant, so we propose a concept: ***question-worthy phrases*** to identify which phrases are worthwhile to be asked about.

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?

- We believe a question generation system can be **improved** by taking those phrases as auxiliary information.
- Moreover, if there are several focuses in an input and we extract multiple phrases, we can then generate **various questions**.

Task

- **Main task** of sentence-level question generation with an **auxiliary task** of question-worthy phrase extraction.
    - Given a sentence, we first extract one or more question-worthy phrases;
    - Then according to various extracted phrases, we generate corresponding questions for such a sentence.

Outlines

- Introduction and Motivation
- **Related Work**
- Framework
- Experiments
- Future Work

Phrases Extraction

- Rule-based approaches, such as using lexical features, POS tagging and NER tagging *(Liu et al., 2011; Yang et al., 2017)*
- Recurrent Neural Network
  - BiLSTM+CRF *(Huang et al., 2015)*
  - Pointer Network *(Vinyals et al., 2015)*

Question Generation

- Rule-based approaches
  - Key aspects extraction and template filling *(Heilman and Smith, 2010)*
- Seq2Seq Model
  - Concept: *question-worthy sentence (Du and Cardie, 2017)*
  - **Answer-aware** question generation *(Zhou et al., 2017)*
- Joint Models: learn QA and QG simultaneously *(Tang* et al., 2017; Wang et al., 2017*)*
  - *Ground truth information is required for its counter-part task when testing, etc. true answers for QG and true questions for QA.*

Contributions

- Propose to generate multiple questions given input sentence **without ground-truth answers.**

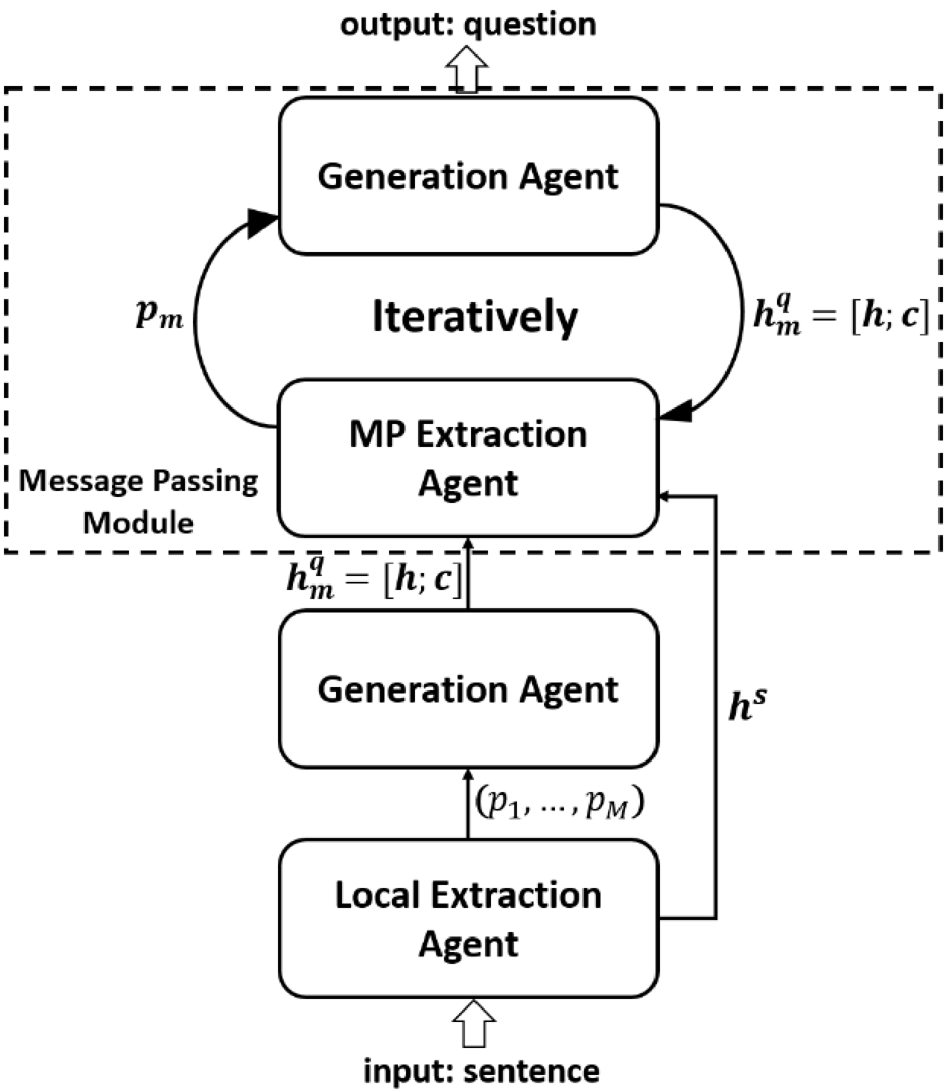
- Extract **question-worthy phrases** *from the input sentence and generate questions based on such information.*

- Developing a **multi-agents communication framework** to learn two tasks simultaneously

Outlines

- Introduction and Motivation
- Related Work
- **Framework**
- Experiments
- Future Work

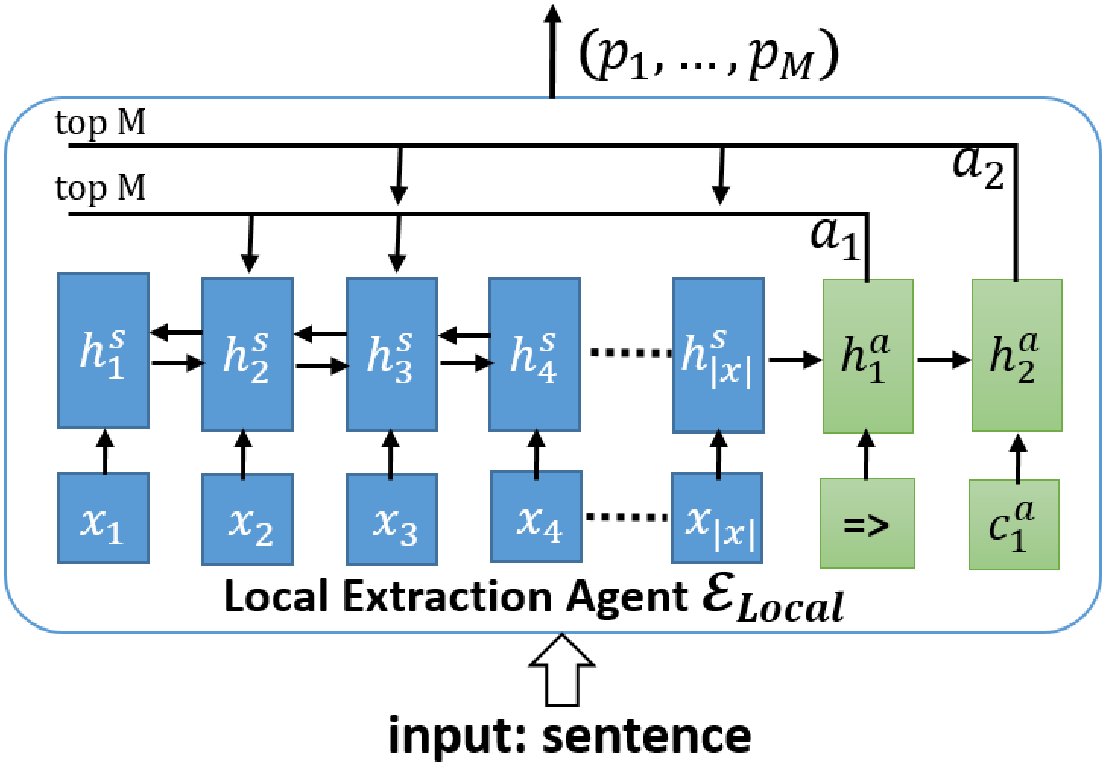
Framework



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

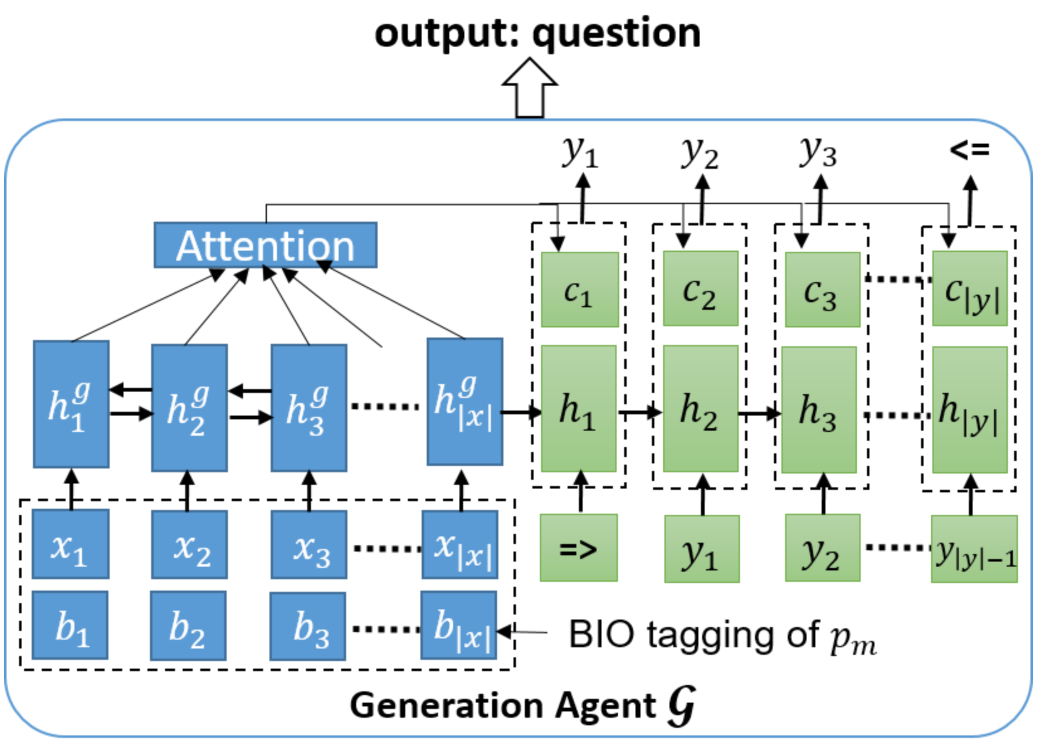
Local extraction agent ($\mathcal{E}_{Local}$)

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



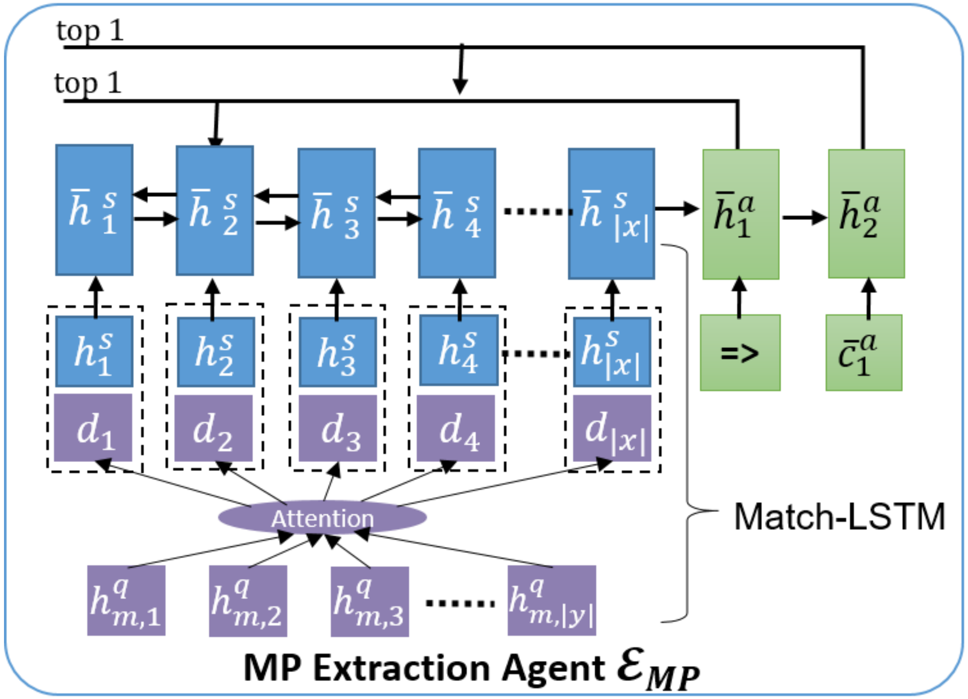
Generation agent ($\mathcal{G}$)

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

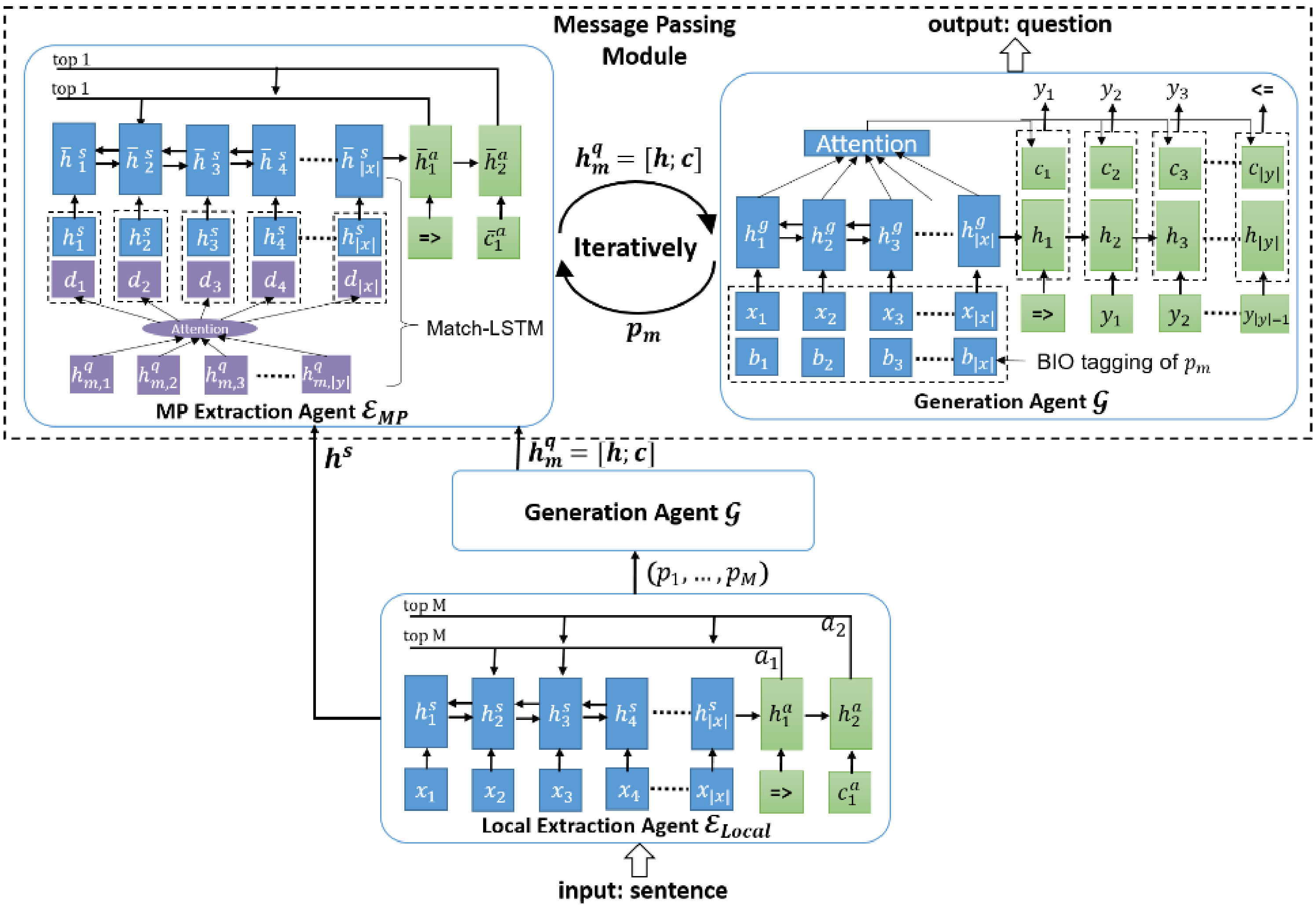


MP Extraction agent ($\mathcal{E}_{MP}$)

- Match-LSTM + Pointer Network
- Use question representations from generation agent as auxiliary information



Framework



Message Passing Module
output: question
$h_m^q = [h; c]$
Iteratively
p_m
Match-LSTM
Attention
MP Extraction Agent $\mathcal{E}_{MP}$
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_1
a_2
Local Extraction Agent $\mathcal{E}_{Local}$
input: sentence

Outlines

- Introduction and Motivation
- Related Work
- Framework
- **Experiments**
- Future Work

Dataset - SQuAD (Rajpurkar et al., 2016)

- Answers, extractive from sentences, are treated as target question-worthy phrases
- More than 30% sentences have multiple questions

Question Number	Sentence Quantity
1	41,356
2	14,499
3	3,921
4	1,198
≥ 5	649
in total	61,623

Data Set	Data Size
Training set	68,704
Validation Set	10,313
Test Set	11,665

Comparison of Extraction Models

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

Model	EM	F1
$\mathcal{E}_{\mathbf{NER}}$	13.12%	17.33
$\mathcal{E}_{\mathbf{Local}}$	24.27%	38.63
$\mathcal{E}_{\mathbf{MP}}$	**35.77%**	**46.71**

Model	Avg.# of phrases
$\mathcal{E}_{\mathbf{NER}}$	0.86
$\mathcal{E}_{\mathbf{Local}}$	1.43
$\mathcal{E}_{\mathbf{MP}}$	1.38

- EM: ExactMatch , F1 score for evaluation
- average numbers of extracted phrases
- significance test: underline $p < 0.01$

Comparison of Generation Models

- $\mathbf{NQG_{Rule}}$: a rule-based model, using an overgenerate-and-rank approach
- $\mathbf{NQG_{Pure}}$: a pure version of question generation, using Seq2Seq Model with Attention
- $\mathbf{NQG_{NER}}$: take phrases extracted from $\mathcal{E}_{\mathbf{NER}}$ as additional input to generate questions
- $\mathbf{NQG_{Local}}$: question generation using phrases extracted from $\mathcal{E}_{\mathbf{Local}}$
- $\mathbf{NQG_{MP}}$: question generation using phrases extracted from $\mathcal{E}_{\mathbf{MP}}$

- $\mathbf{NQG_{Answers}}$: answer-aware, use the ground truth of answers to generate questions, **upper bound**

Comparison of Generation Models

- Evaluation Metrics: BLEU 1-4, METEOR, ROUGE_L

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

- significance test

$\underline{\text{underline}}$: $p < 0.01$; *: $p < 0.05$

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*⟩.
$\textbf{E}_{\text{MP}}$**:** 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 ?
$\textbf{NQG}_{\text{NER}}$**:** who wons the regular season ? what was the regular session in the afl ?
$\textbf{NQG}_{\text{MP}}$**:** 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
$\textbf{E}_{\text{MP}}$**:** next to the main building.
Questions
Ground Truth: the basilica of the sacred heart at notre dame is beside to which structure ?
$\textbf{NQG}_{\text{NER}}$**:** what is next to main building ?
$\textbf{NQG}_{\text{MP}}$**:** where is the basilica of prayer ?

Outlines

- Introduction and Motivation
- Related Work
- Framework
- Experiments
- **Future Work**

Future Work

- Improve **phrase extraction**, especially for multiple phrases extraction
- Explore better **evaluation** schemes for QG

References

- P. Piwek, H. Hernault, H. Prendinger, and M. Ishizuka, “Intelligent Virtual Agents,” vol. 4722, 2007.
- M. Heilman, V. Aleven, W. W. Cohen, D. J. Litman, and N. A. Smith, “Automatic Factual Question Generation from Text.”
- Z. Liu, X. Chen, Y. Zheng, and M. Sun, “Automatic Keyphrase Extraction by Bridging Vocabulary Gap *,” pp. 135–144, 2011.
- V.-W. Soo, S.-Y. Lin, S.-Y. Yang, S.-N. Lin, and S.-L. Cheng, “A cooperative multi-agent platform for invention based on patent document analysis and ontology,” Expert Syst. Appl., vol. 31, no. 4, pp. 766–775, Nov. 2006.
- Z. Huang, B. Research, W. Xu, and K. Yu, “Bidirectional LSTM-CRF Models for Sequence Tagging.”
- O. Vinyals, M. Fortunato, and N. Jaitly, “Pointer Networks,” pp. 1–9, 2015.

References

- M. Heilman and N. A. Smith, “Good question! Statistical ranking for question generation,” NAACL HLT 2010 - Hum. Lang. Technol. 2010 Annu. Conf. North Am. Chapter Assoc. Comput. Linguist. Proc. Main Conf., no. June, pp. 609–617, 2010.
- X. Du, J. Shao, and C. Cardie, “Learning to Ask: Neural Question Generation for Reading Comprehension,” 2017.
- Q. Zhou, N. Yang, F. Wei, C. Tan, H. Bao, and M. Zhou, “Neural Question Generation from Text: A Preliminary Study,” 2017.
- D. Tang, N. Duan, T. Qin, Z. Yan, and M. Zhou, “Question Answering and Question Generation as Dual Tasks,” 2017.
- T. Wang, X. Yuan, and A. Trischler, “A Joint Model for Question Answering and Question Generation,” 2017.
- P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, “SQuAD: 100,000+ Questions for Machine Comprehension of Text,” pp. 2383–2392, 2016.

Thank you！