HAS-QA: **H**ierarchical **A**nswer **S**pans Model for Open-domain Question Answering

Liang Pang [†], Yanyan Lan [†*], Jiafeng Guo [†], Jun Xu [†], Lixin Su [†], Xueqi Cheng [†]

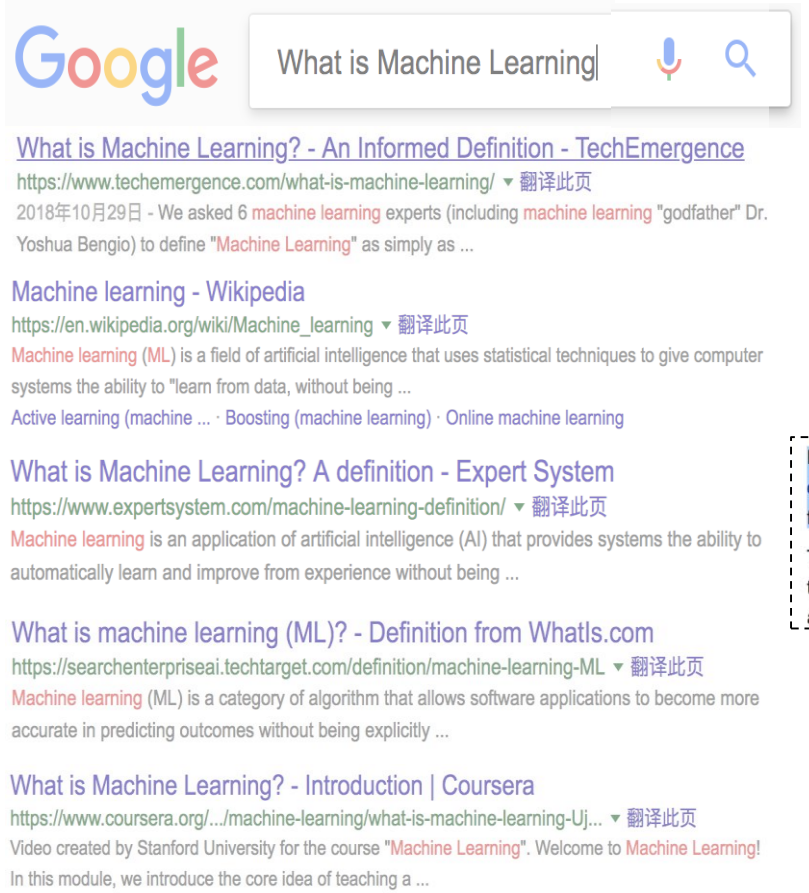
† CAS Key Lab of Network Data Science and Technology Institute of Computing Technology,
Chinese Academy of Sciences, Beijing, China

† University of Chinese Academy of Sciences, Beijing, China

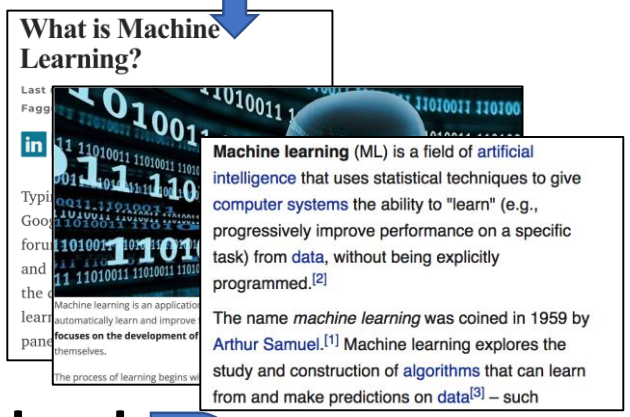
∗ Department of Statistics, University of California, Berkeley

I Background

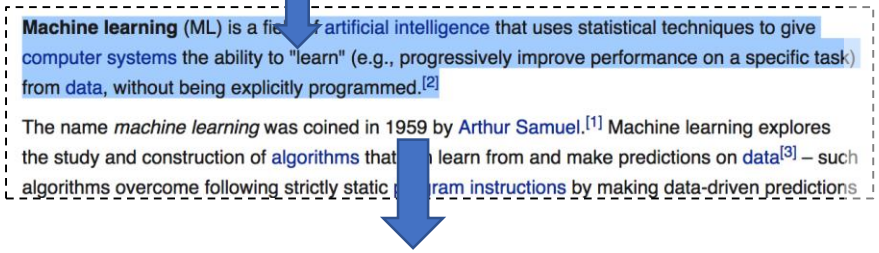
Search Engine



Search

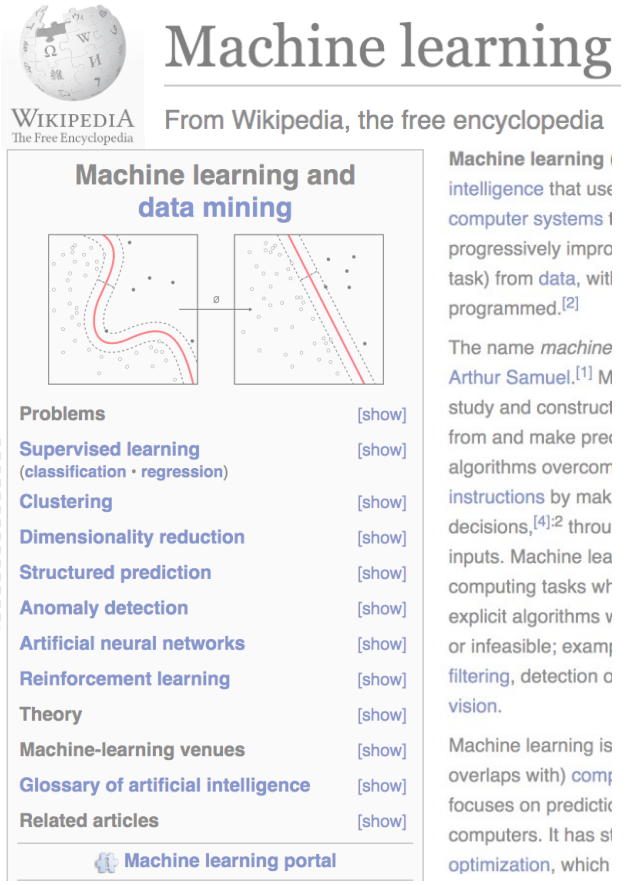


Extract



Machine learning (ML) is a field of artificial intelligence that uses statistical techniques to give computer systems the ability to "learn" from data, without being explicitly programmed.

Knowledge Base



High Question Diversity

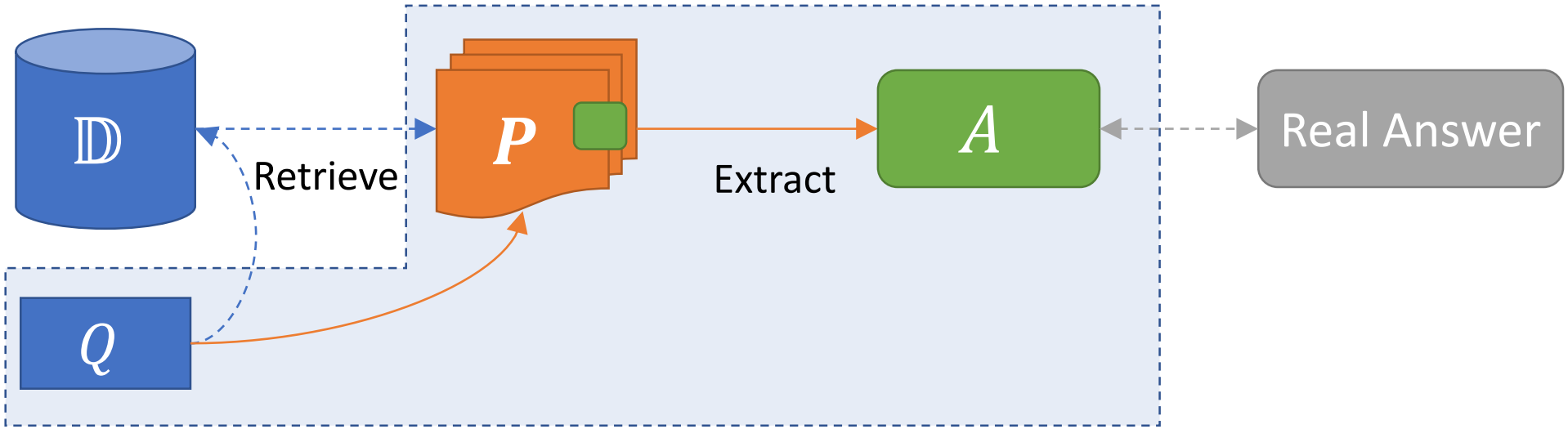
Low Information Entropy

I Outline

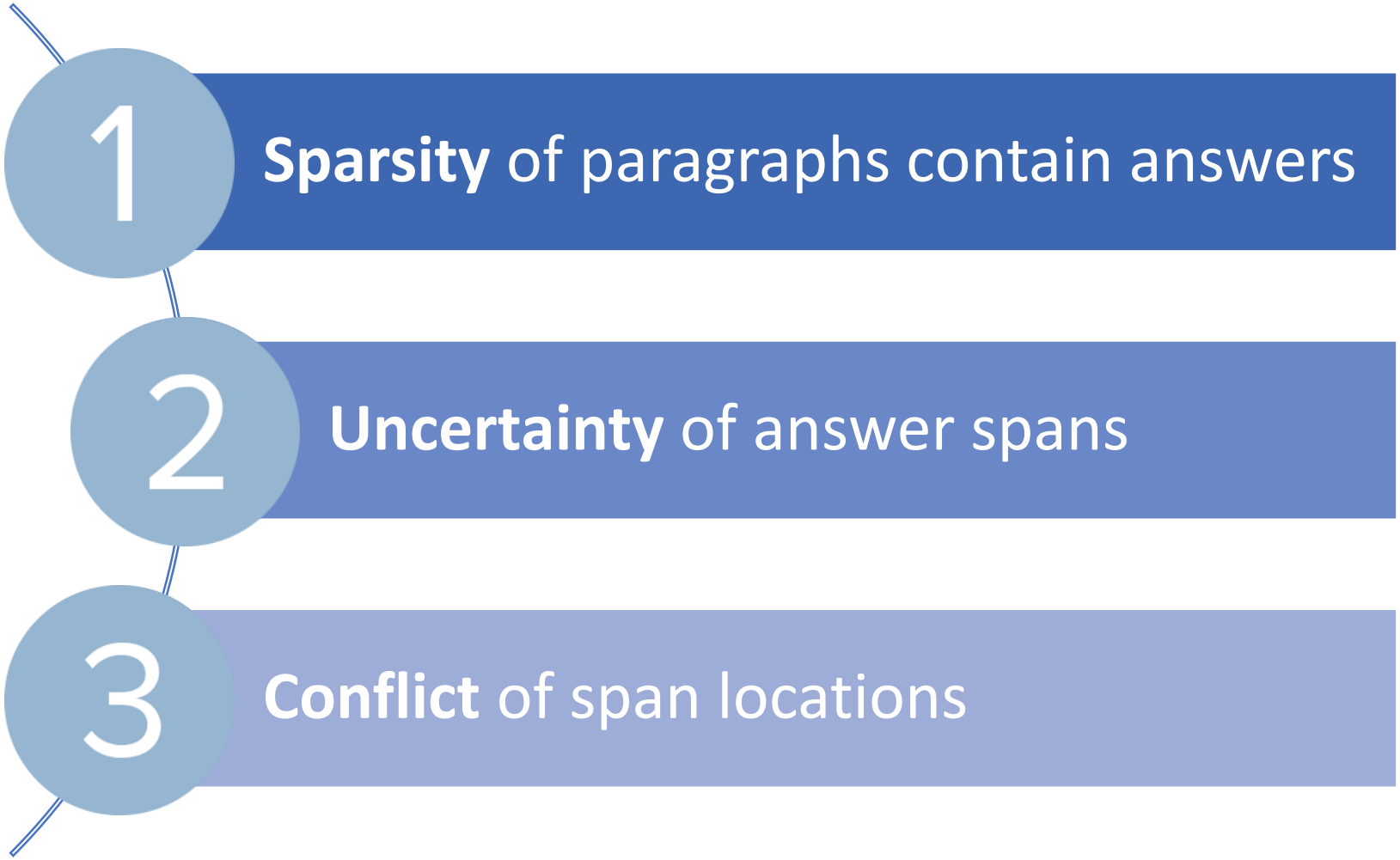
- Open-domain Question Answering
- Formulation & HAS-QA
- Experiments
- Conclusion & Going Forward

I OpenQA - Open-domain Question Answering

- **Input:**
  - Q: real world question.
  - $\mathbb{D}$: huge *unstructured* text dataset, such as **web page corpus** or **Wikipedia.**
- **Output:**
  - A: answer text from text dataset, which can answer the input question



I Challenges



Challenges - Sparsity

新版五元人民币是什么花

百度一下

第五套人民币上面的花卉是些什么花?_百度知道

7个回答 - 回答时间: 2018年9月27日 - 40人觉得有用

最佳答案: 1、1元纸币——兰花 象征意义 风姿绰约 幽香远溢 ... 中四君子之一,四季名花之司春使者,在...

更多关于新版五元人民币是什么花的问题>>

https://zhidao.baidu.com/quest... - 百度快照

这张5元人民币,一定要记清楚,千万不要随便花!

2018年5月22日 - 2002年11月18日中国人民银行发行了99... 民币,虽然并不是1999年发行的,但是确定是属于99版人民... 中...

baijiahao.baidu.com/s?... - 百度快照

最新版人民币的正面分别是什么花?背面分别是哪些风景?_百... 道

2个回答 - 回答时间: 2016年6月23日 - 39人觉得有用

最佳答案: 最新版的 100:人民大会堂 50:布达拉宫 20:桂林山水 10:长江三峡 5:泰山... 西湖十景之三潭印月 参考 第四套,即新版人民币,正面图案一是体现我国...

更多关于新版五元人民币是什么花的问题>>

https://zhidao.baidu.com/quest... - 百度快照

五元人民币上的水印花是什么花?_百度知道

1个回答 - 回答时间: 2016年6月16日

[专业] 答案:面额:5元 水印:水仙花 背景图案:泰山 发行时间:2002.11.18 背... 色:紫色... 版:2005.08.31 正面图案:毛泽东头像 规 格:116×56 正面主色:...

更多关于新版五元人民币是什么花的问题>>

https://zhidao.baidu.com/quest... - 百度快照

人民币从一元钱到一百上面的图案都有花请问这些都是什么花?

1个回答 - 回答时间: 2008年04月26日

最佳答案: 第五套人民币 1元:水印和正面花团中的花都是兰花； 5元:水印和正面花团中的... 是水仙； 10元:水印和正面花团中的花都是月季； 20元:水印和正面花团中... 都是...

https://iask.sina.com.cn/b/124... - 百度快照

5元人民币有两个版本,出现这种号码,一定不要花掉!

2018年8月9日 - 第五套人民币是在1999年开始发行的,而面... 元的的人民币也是在这个时期发行的,但是在第五套人民币中...

https://baijiahao.baidu.com/s?... - 百度快照

新版5元人民币背面是什么图??? 爱问知识人

举报 新版5元人民币 背面是什么图??? 自-浮生未 分享: 全部答案 ...安迪出生的地... 小县城,安迪的妈妈是当地有名的花癫,因为疯了以后喜欢穿...

https://iask.sina.com.cn/b/iR2... - 百度快照

Challenges - Uncertainty

新版五元人民币是什么花 百度一下

第五套人民币上面的花卉是些什么花?_百度知道

7个回答 - 回答时间: 2018年9月27日 - 40人觉得有用

最佳答案: 1、1元纸币——兰花 象征意义 风姿绰约 幽香远溢 ... 中

四君子之一,四季名花之司春使者,在...

更多关于新版五元人民币是什么花的问题>>

https://zhidao.baidu.com/quest... - 百度快照

第五套人民币上面的花卉是些什么花?

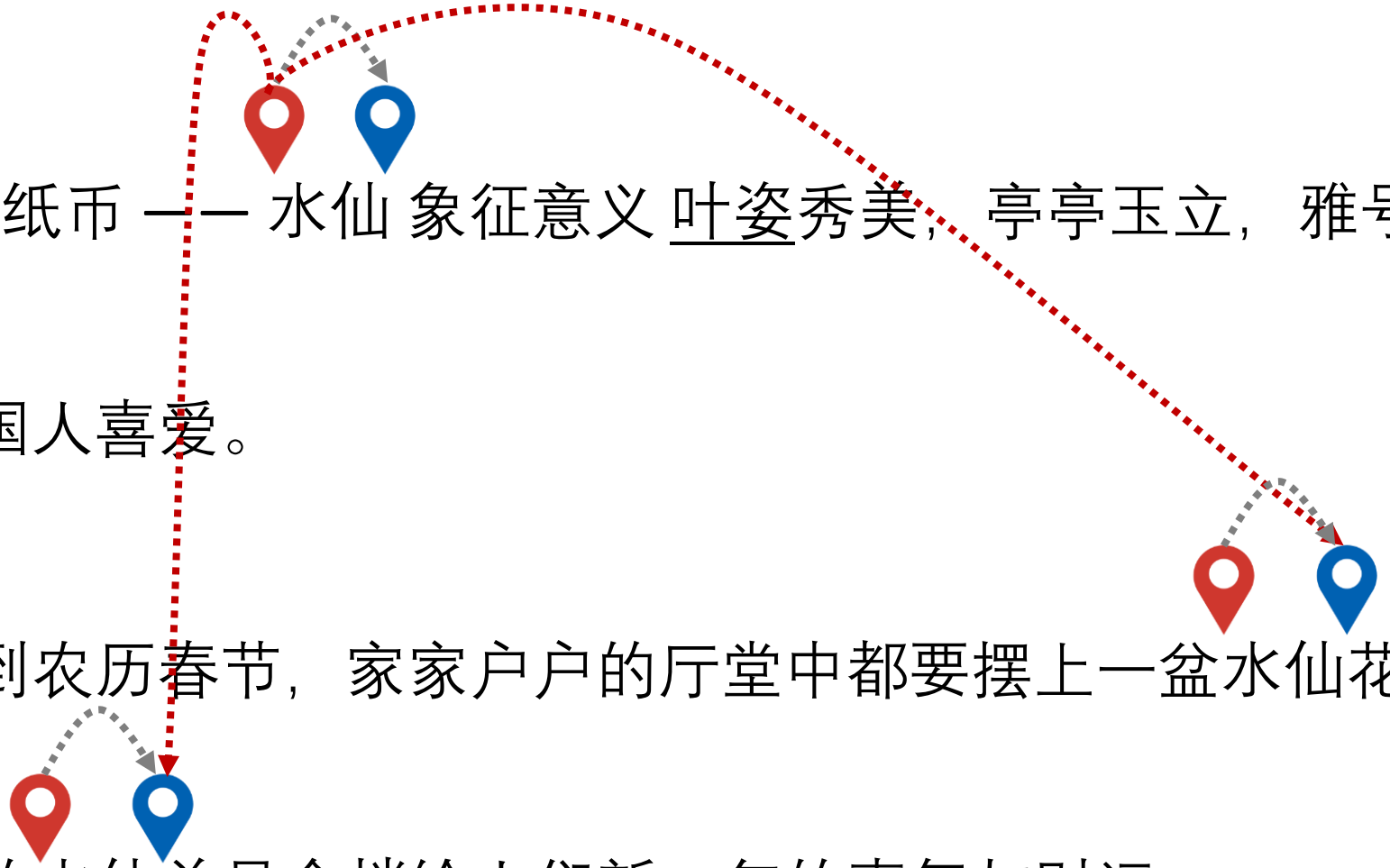
我来答

2、5元纸币——水仙

象征意义

叶姿秀美，亭亭玉立，雅号“凌波仙子”，深受国人喜爱。每到农历春节，家家户户的厅堂中都要摆上一盆水仙花，黄蕊白被如金银的水仙总是会捎给人们新一年的喜气与财运。

I Challenges - Conflict



5元纸币 —— 水仙 象征意义 叶姿秀美，亭亭玉立，雅号“凌波仙子”，深受国人喜爱。

每到农历春节，家家户户的厅堂中都要摆上一盆水仙花，黄蕊白被如金银的水仙总是会捎给人们新一年的喜气与财运。

I Outline

- Open-domain Question Answering
- Formulation & HAS-QA
- Experiments
- Conclusion & Going Forward

I Probabilistic Model – Target

- **Target:**

Answer Probability: $P(A|Q,\mathbf{P})$

- Answer Text
- Question
- Retrieved Paragraphs

- **Interpret:**

  - The probability of the text A is the answer text, given the question Q and retrieved paragraphs $\boldsymbol{P}$.

- **Implement:** HAS-QA Model

Probabilistic Model – Challenge 1

- **Challenge 1:**

  Many paragraphs without the answer span are included in the data collection.

  Answer Probability: $P(A|Q,\mathbf{P}) = \sum_{i=1}^{K} P(P_i|Q,\mathbf{P}) P(A|Q,P_i);$

  Conditional Answer Probability

  Paragraph Probability

- **Interpret:**
  - **Paragraph Probability**: The probability of the paragraph P_i that contains answer text, given the question Q and retrieved paragraphs $\boldsymbol{P}$.
  - **Conditional Answer Probability**: The probability of the text A is the answer text, given the question Q and paragraph P_i.

- **Implement:** Paragraph Quality Estimator

I Probabilistic Model – Challenge 2

- **Challenge 2:**

  Multiple answer spans may exist within one given paragraph.

Conditional Answer Probability:

$$P(A|Q,P_i) := \mathcal{F}(\{P(L_j(A)|Q,P_i)\}_j),$$
$$j \in [1, |\mathcal{L}(A,P_i)|];$$

Span Probability

- **Interpret:**
  - **Span Probability**: The probability of the text span $L_j(A)$ is the answer span, given the question Q and paragraph P_i.
  - $\mathcal{F}$ is an aggregation function, such as MAX, SUM.
  - There exist $|\mathcal{L}(A,P_i)|$ answer spans in paragraph P_i.

- **Implement:** Multiple Span Aggregator

Probabilistic Model – Challenge 3

- **Challenge 3:**

  The end position of an answer span is dependent with the start position.

Span Probability: $P(L_j(A)|Q,P_i) = P(L_j^s(A)|Q,P_i) \cdot P(L_j^e(A)|Q,P_i,L_j^s(A)).$

$P(L_j^s(A)|Q,P_i)$ → Location Start Probability

$P(L_j^e(A)|Q,P_i,L_j^s(A))$ → Location End Probability

- **Interpret:**
  - **Location Start Probability**: The probability of the text span start location $L_j^s(A)$ is the answer span location.
  - **Location End Probability**: The probability of the text span end location $L_j^e(A)$ is the answer span location, given start location $L_j^s(A)$.

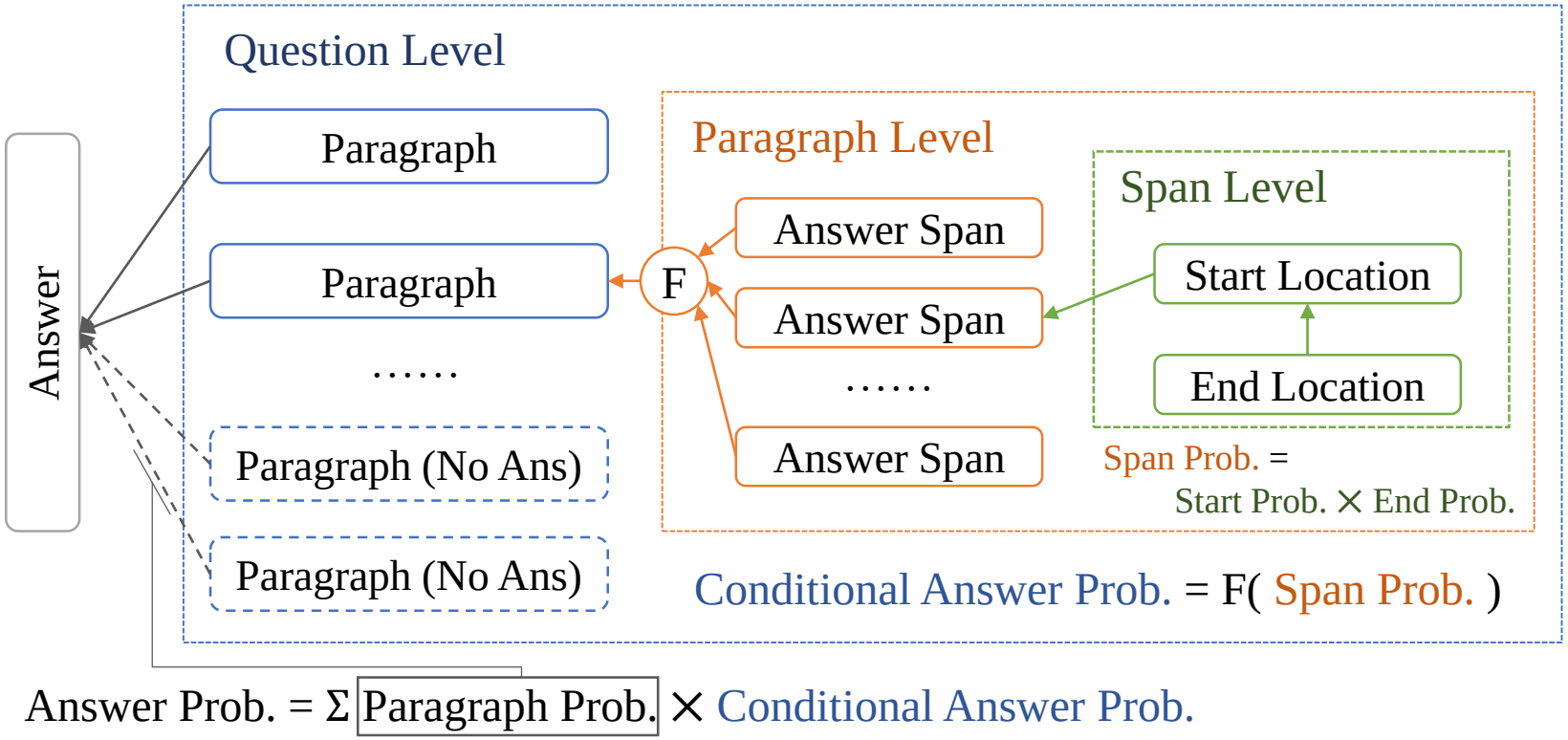
- **Implement:** Conditional Span Predictor

HAS-QA Model – OpenQA Task

- **Target:**

Answer Probability: $P(A|Q,\mathbf{P})$

- **Model:**



HAS-QA Model – RC Task

For a **Reading Comprehension Task**

- **Object:** $P(A|Q, P^+)$

- **Model:**

$$P(A|Q, P^+) := P(L(A)|Q, P^+);$$

$$P(L(A)|Q, P^+) = P(L^s(A)|Q, P^+) \cdot P(L^e(A)|Q, P^+).$$

RC ⊆ OpenQA

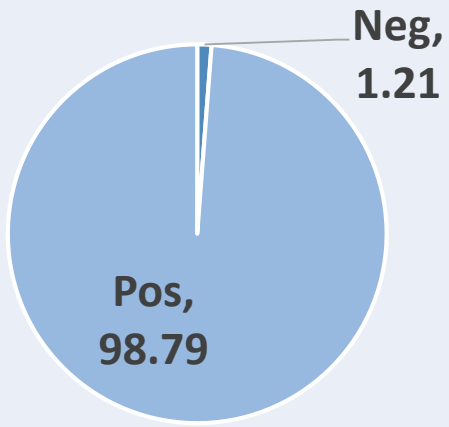
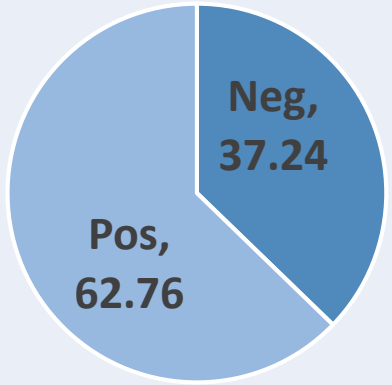
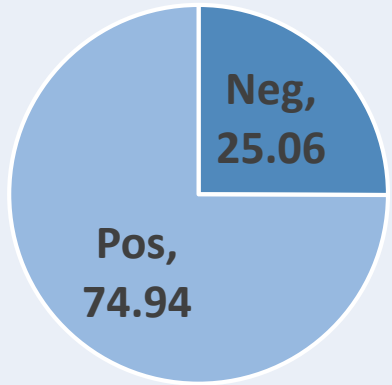
1. One paragraph
2. One answer span
3. Independence start and end location

I Outline

- Open-domain Question Answering
- Formulation & HAS-QA
- Experiments
- Conclusion & Going Forward

I Dataset

Three OpenQA Dataset

	QuasarT	TriviaQA	SearchQA
Question Count	43K	95K	140K
Negative Paragraph Ratio			
Average Answer Span Count	SPAN SPAN SPAN SPAN SPAN	SPAN SPAN SPAN SPAN	SPAN SPAN SPAN SPAN SPAN SPAN

Large Dataset

Noisy Paragraphs

Multiple Answer Spans

Experiments

Model	QuasarT EM	QuasarT F1	TriviaQA EM	TriviaQA F1	SearchQA EM	SearchQA F1
GA (Dhingra et al., 2017a)	0.264	0.264	-	-	-	-
BiDAF (Seo et al., 2016)	0.259	0.285	0.411	0.474	0.286	0.346
AQA (Buck et al., 2017)	-	-	-	-	0.387	0.456
DrQA (Chen et al., 2017)	0.377	0.445	0.323	0.383	0.419	0.487
R^3 (Wang et al., 2017a)	0.353	0.417	0.473	0.537	0.490	0.553
Shared-Norm (Clark and Gardner, 2017)	0.386	0.454	0.613	0.672	0.598	0.671
HAS-QA (MAX Ans. Span)	**0.432**	**0.489**	**0.636**	**0.689**	**0.627**	**0.687**

1) HAS-QA **outperforms** traditional RC baselines with a large gap, such as GA, BiDAF, AQA listed in the first part.

2) HAS-QA **outperforms** recent OpenQA baselines, such as DrQA, R3 and Shared-Norm listed in the second part.

I Outline

- Open-domain Question Answering
- Formulation & HAS-QA
- Experiments
- Conclusion & Going Forward

I Conclusion

- A new probabilistic formulation of **OpenQA**, based on a three-level hierarchical structure, i.e., the question level, the paragraph level and the answer span level.

- HAS-QA Model
  - 1) a paragraph quality estimator makes it robust for the paragraphs without answer spans
  - 2) a multiple span aggregator points out that it is necessary to combine the contributions of multiple answer spans in a paragraph
  - 3) a conditional span predictor is proposed to model the dependence between the start and end positions of each answer span.

I Going Forward

Single Answer Spans Extraction

Multiple Answer Spans Extraction

Well-formed Answers Generation

Thanks  Q & A

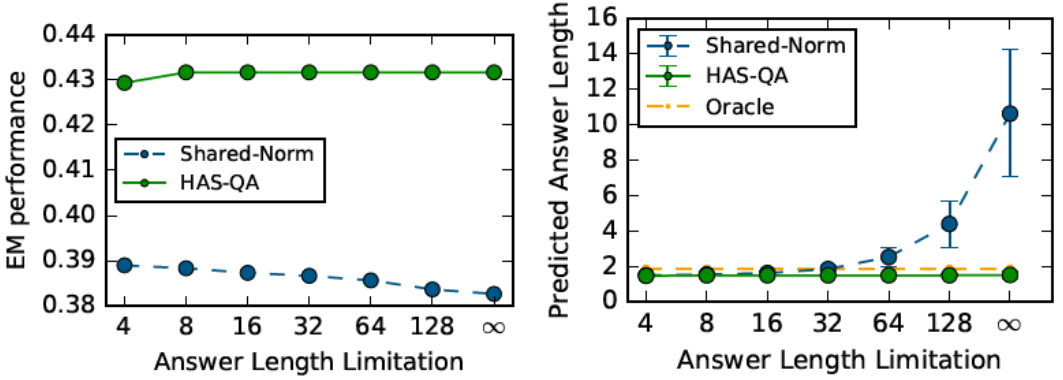
Name: Liang Pang | Email: pangliang@ict.ac.cn

Experimental Analysis

- Effects of Conditional Pointer Networks

1. The performance of Shared-Norm **decreases** when removing the answer length limitation, while the performance of HAS-QA first **increases** then becomes stable.

2. The average predicted answer length **increases** in Shared-Norm when removing the answer length limitation. However, HAS-QA stably **keeps** the about 1.8 average words, where the oracle average answer length is about 1.9 words.



Example:

About Celebrating the contributions of Louis Braille January 5th , 2009
On the 200th anniversary of Louis Braille ' s birth , people around the world are saluting a man whose tactile alphabet has provided a lifeline to people with impaired vision .

Shared-Norm HAS-QA

Experimental Analysis

- Effects of Multiple Spans Aggregation

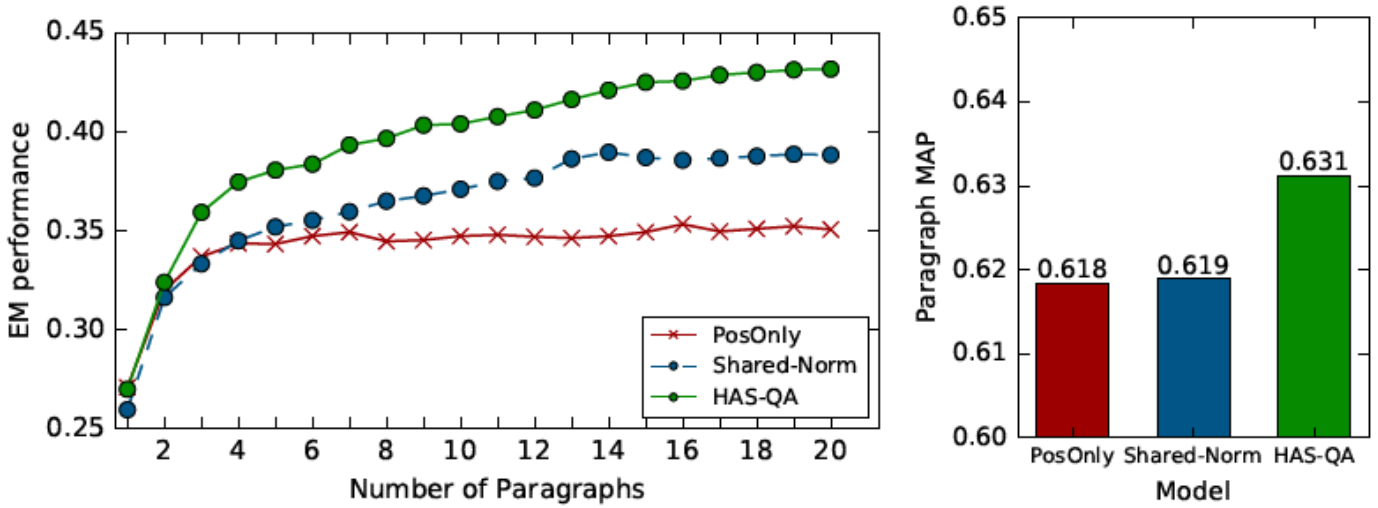
Model	EM	F1
HAS-QA (HEAD Ans. Span)	0.372	0.425
HAS-QA (RAND Ans. Span)	0.341	0.394
HAS-QA (SUM Ans. Span)	0.423	0.484
HAS-QA (MAX Ans. Span)	0.432	0.489

1. **SUM and MAX operations**. They take advantages of using multiple answer spans for training and improve about 6% - 10% in EM comparing to the **HEAD operation**.

2. The failure of **RAND operation**, mainly comes down to the conflicting training samples.

Experimental Analysis

- Effects of Paragraph Quality



1. With the increasing number of given paragraphs which ordered by the rank of search engine, EM performance of HAS-QA **sustainably grows**.

2. The Mean Average Precision (MAP) score between the predicted scores and the label whether a paragraph contains answer spans, shows that HAS-QA can rank **the high quality paragraphs** in the front of the given paragraph list.