

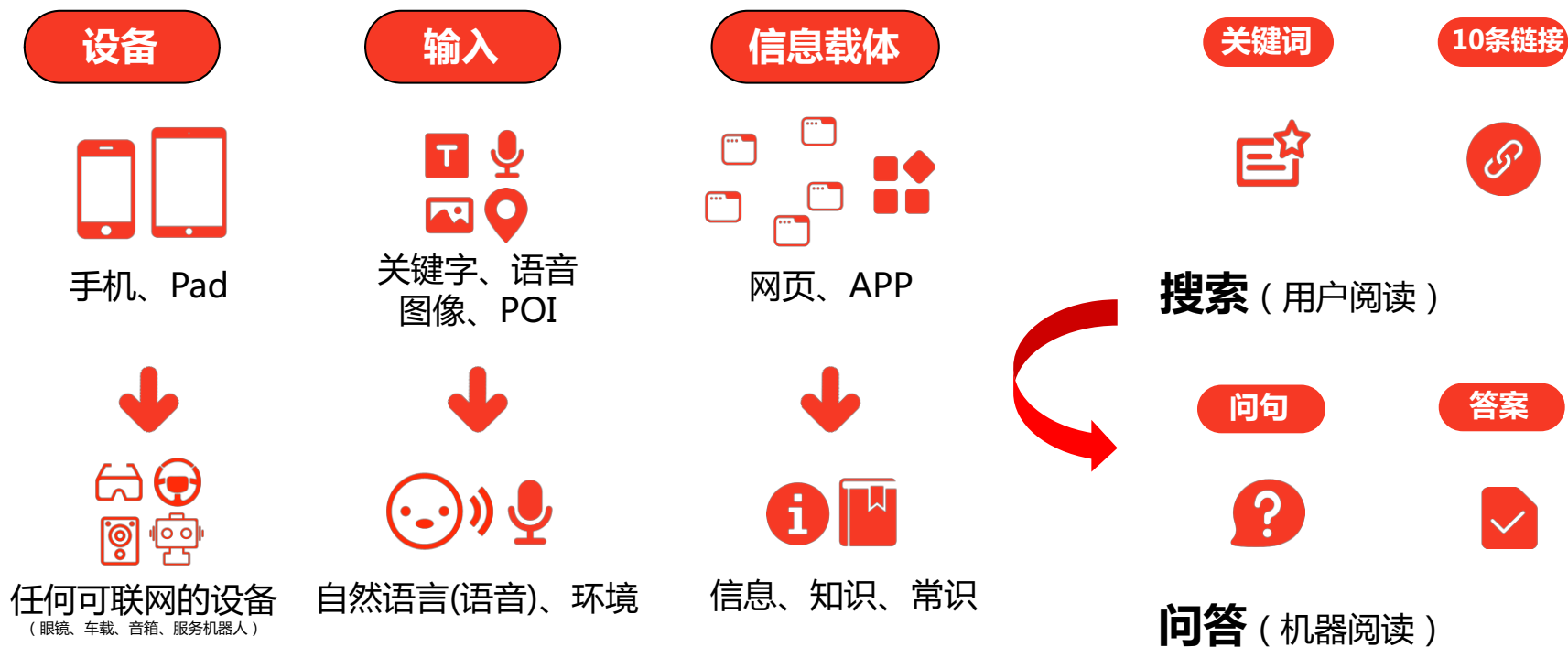


# 搜索引擎中的智能问答

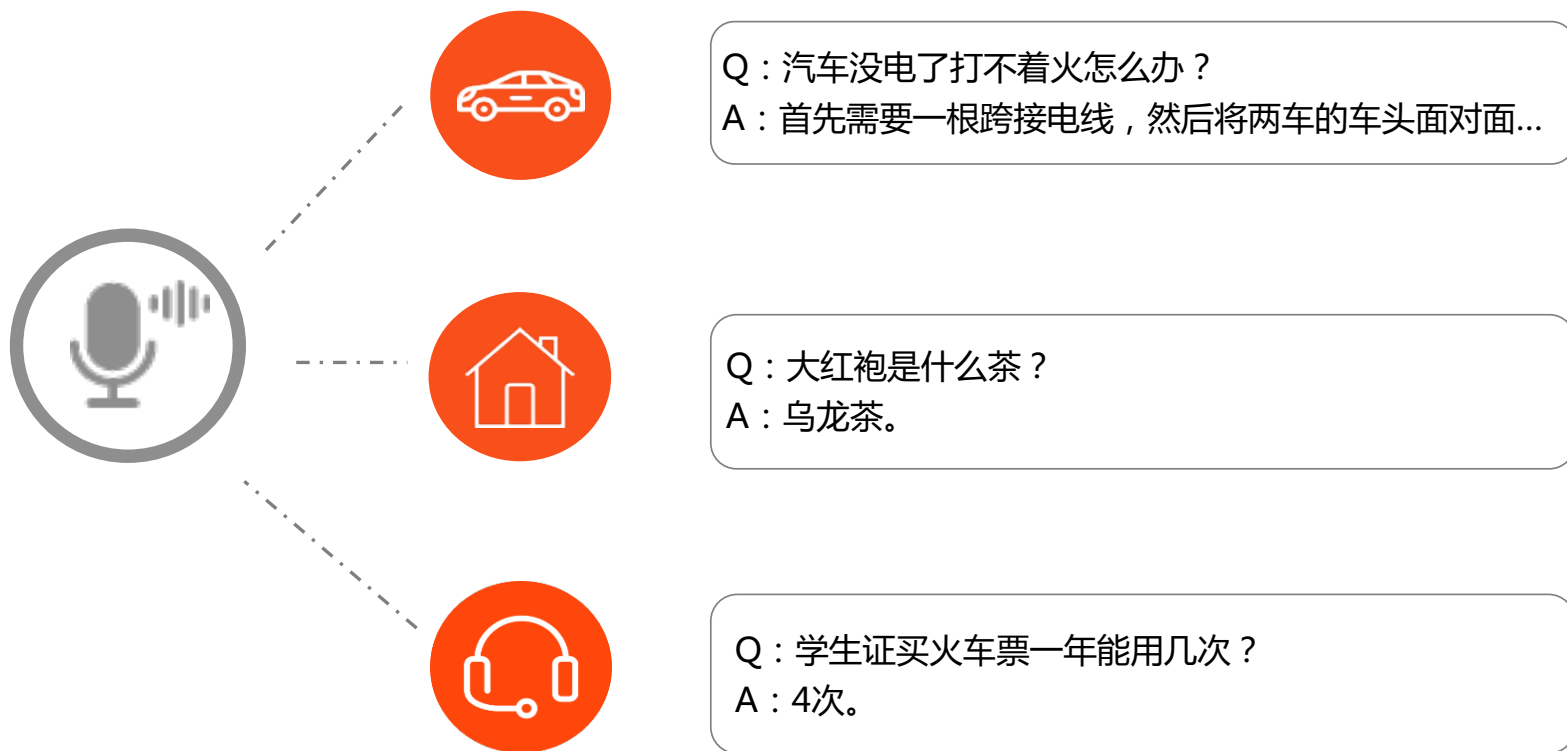
张奇

复旦大学

# 智能问答 -- 从PC到移动



# 智能问答 -- 无处不在的问答引擎



# 智能问答 -- 无处不在的问答引擎

中国移动 晚上8:28

中国最长的隧道

网页 微信 新闻 图片 视频 明医 英

### 中国最长隧道 立知

 **秦岭终南山公路隧道**  
秦岭终南山特长公路隧道是中国国道主干线包头至北海段在陕西境内的西康高速公路北段,同是...

搜狗百科 纠错

---

### 中国最长的隧道——秦岭终南山公路隧道,总投资286亿元 - 知乎

 建设规模世界第一,是中国公路隧道之最,隧道之最 秦岭终南山公路隧道创造了高速公路隧道建设史上的六项之最.一、是世界上...

知乎专栏 - zhuanian.zhihu.com

---

### 中国最长的公路隧道排名

这是中国自行设计施工的世界最长的双洞单向公路隧道 北起西安市长安区沓岔,南至商洛市所辖的柞水

< > 🔍 🏠 ☰

中国移动 晚上8:29

美国上任总统是谁

网页 微信 新闻 图片 视频 明医 英

### 美国上任总统是谁 立知

 **奥巴马**  
2008年11月4日,美国各大电视网公布的初步统计结果显示,美国民主党总统候选人、伊利诺伊州...

搜狗百科 纠错

---

### 当总统到底有多累?10位美国总统任职前后惊人对比[1]

2016年8月25日 - 看看美国这10位总统任职前后的颜值对比你就不难看出来了. **奥巴马** 2008-2016 阅读 电话:8610-84883645 传真:86...

英语 中国日报网 - m.lt.chinadaily.com.cn

---

### 美国有多少任总统?他们是谁?- 百度知道

截止2018年,美国共有45任总统,他们的名字如下: 1、

< > 🔍 🏠 ☰

中国移动 晚上8:32

如何办理护照

网页 微信 新闻 图片 视频 明医 英

 **如何办理护照 立知**

“护照是公民在国际间通行所使用的身份证和国籍证明,也是一国政府为其提供外交保护的重要依据。为此,居民出国旅游,需要申请办理护照。具体程序是: 1、携带本人身份”

 宋正东 - 专职律师 - 上海市锦天城 (成都) 律师...

搜狗律师

---

### 护照(证明公民国籍和身份的合法证件)\_ 搜狗百科

 国民政府公布了公布《护照条例》和办理华侨登记.此两措施,皆为将中华民国护照法制化之准备.

< > 🔍 🏠 ☰

中国移动 晚上8:33

苹果手机如何重启

 **苹果手机死机怎么重启**

**苹果手机死机怎么重启 立知**

查看一下iPhone是否能正常关机,长按电源键,直到屏幕上出现滑动关机的滑块,向右滑动滑块,要是毫无反应才考虑强制重启, iPhone6S和更早的设备: ...

云骑士

---

### iPhone7怎么重启?iPhone7重启按哪个键 - 265G iPhone频道

 现在的许多苹果手机用户都是依靠Home键和电源键来实现手机的重启功能,但最新推... 一直按住两个按键直到屏幕黑屏,并持续按...

265G苹果站 - iphone.265g.com

---

### iphone手机如何快速重启 - 百度经验

< > 🔍 🏠 ☰



# 智能问答 -- 来自搜索的问题分布

用户需求强，难度大

问题类型多

猕猴桃和奇异果有什么区别

腰肌劳损挂什么科

眼镜蛇和眼镜王蛇的区别

蜂蛹泡酒的功效

人有多少颗牙齿

什么是幽门螺旋杆菌

① 约1/4用 没有房产证的酒店拆迁如何补偿

手机怎么注册淘宝账号

② 更长尾，更偏向自然语言描述

低烧可以进藏吗

落地签证是什么意思

事实类，非事实类，观点类

| 查询类别 | 问题             | 搜索请求占比 |
|------|----------------|--------|
| 问答   | 影响发电量的三大因素     | 24%    |
| 医疗   | esp是什么         | 9%     |
| 小说   | 神经绷紧怎么回事       | 7%     |
| 教育   | 顶的网络意思         | 4.5%   |
| 人物   | 大米生虫子了还能吃吗     | 4.5%   |
| 知识   | 网很卡是什么原因       | 2%     |
| 购物   | 折耳兔怎么养         | 2%     |
|      | 石家庄市银监局投诉电话是多少 | 2%     |
|      | 充电器可以托运吗       | 20%    |
|      | 苹果手机信号不好怎么办    | 2%     |
|      | 烫伤的水泡会自然消吗     | 2%     |
|      | 网速突然变慢的原因      | 2%     |
|      | 在家原地跑步能减肥吗     | 2%     |
|      | 交首付款注意事项       |        |
|      | 故乡鲁迅发生的变化原因    |        |
|      | 疝气会影响生育吗       |        |

# 智能问答 -- 解决方案

基于社区的问题回答

Community-based Question Answering

基于知识图谱的问题回答

Knowledge-based Question Answering

阅读理解

Machine Reading Comprehension

# 智能问答 -- 解决方案

基于社区的问题回答

Community-based Question Answering

基于知识图谱的问题回答

Knowledge-based Question Answering

阅读理解

Machine Reading Comprehension

# 智能问答 -- 基于社区的问题回答



问： ios12.0.1系统为什么升级后信号变差？

如题 昨天升级后明显信号变得很差 许多APP点开都不能刷新

发布日期 2018年10月13日 上午1:25

回复

我也有此问题 (1)



化学 化工 化学元素 元素周期表 化学专业

## 集齐元素周期表所有元素，大概需要多少钱？

突然开的脑洞，不需要剂量特别多，想收藏全套。

谢谢大家提醒，考虑到半衰期，稀有度和放射

关注问题

写回答

邀请回答

44 条评论

分享

举报

...

查看全部 293 个回答



氧化氩氩

27,008 人赞同了该回答

嗯...一群巨佬已经把非放射性元素收的差不多了.....我就大约上两张图，稍微估个价吧。

先在最前面说一嘴，价格是要可以以此流通的，所以很多半衰期很短的元素不在我的讨论中（完全没发流通对不对...）。

会做到评论90%回复，剩下10%有的可能不知道回点啥.....有点也是没来得及回复。私信可以100%回复，欢迎提问。

哇哇哇，一晚上赞数15个诶，好激动，在最后补一个铯融化的视频。

以后说不好还会上稀有气体的辉光视频。

开心，不到一天快百赞了，第一次回答很开心，慢慢发稀有气体的辉光视频，有审核，所以可能有些慢。（视频已更完）

# 智能问答 - 语义匹配

## Apple 官方支持：



如果您的 iPhone、iPad 或 iPod touch 开不了机或死机怎么办

如果设备屏幕死机、在您触摸后没有响应或者设备在开机时卡住，请了解该怎么做。

## 用户：

iPhone 死机了怎么办

**苹果8**突然卡死，主界面不能动，在线急等

我的**肾7**突然无法操作，这种情况怎么解决？

用户输入与标准问法差别很大

# 智能问答 -- 语义匹配

脑袋测得出的东西叫智商，

小孩子发烧**38**脑袋测不出的东西叫智慧；

尧**41**度怎么办

北京大学 ← 眼睛看得到的地方叫视线，

盐酸**氯**丙嗪 ← 眼睛看不到的地方叫视野；

**定金** ↔ **订金** 耳朵听得到的动静是声音，

耳朵听不到的动静是声誉；

嘴里说得出来的话叫内容，

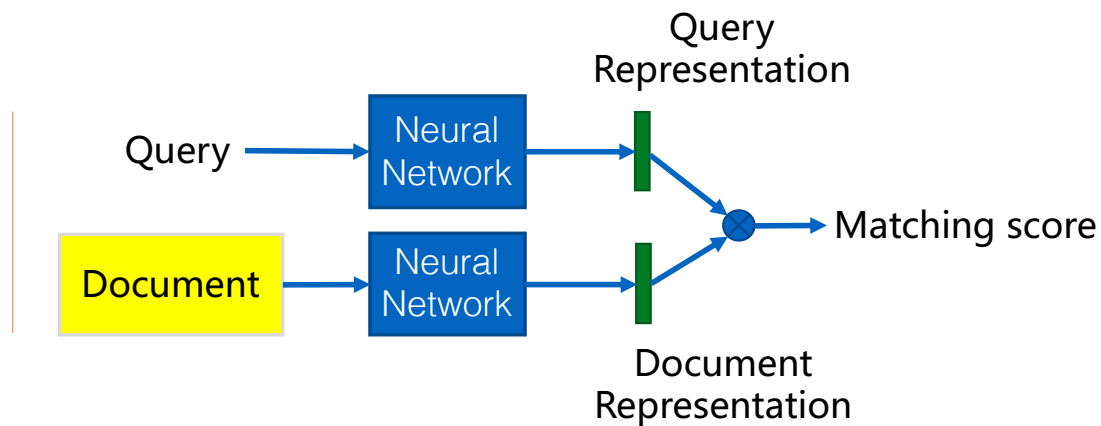
… …

嘴里说不出来的话叫内涵；

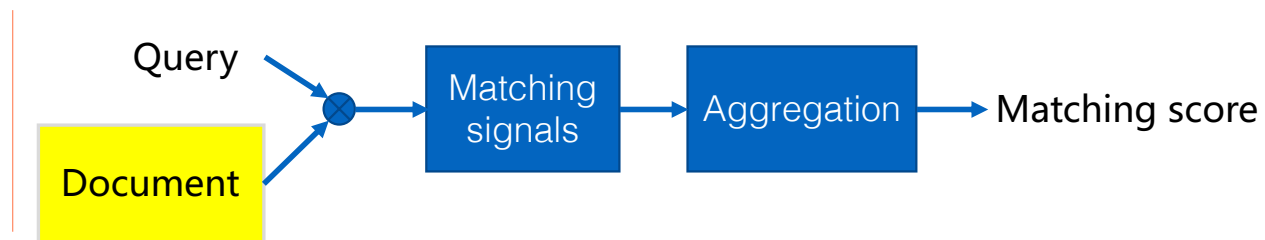
**失之毫厘，谬以千里！**

# 智能问答 -- 语义匹配

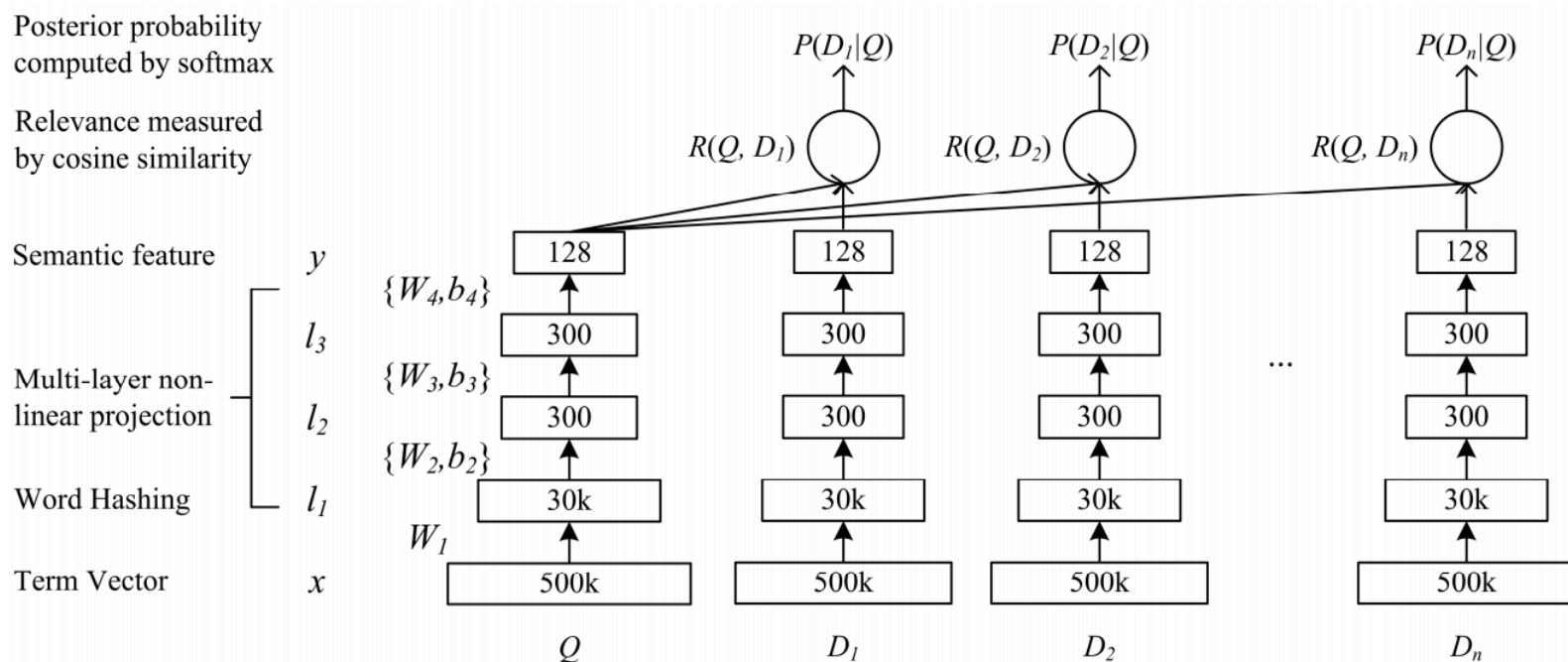
## 基于句子表示的方法



## 基于交互关系的方法



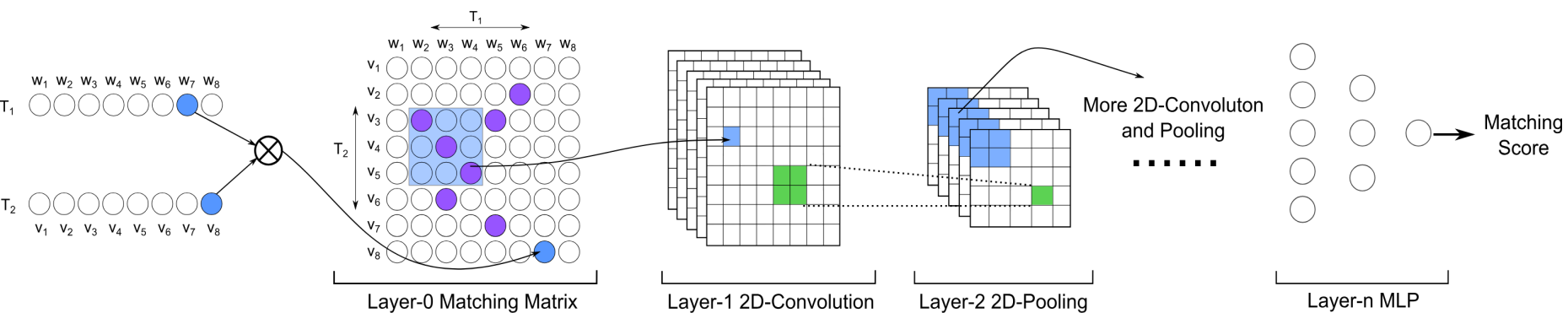
# 智能问答 -- 语义匹配



DSSM: Learning Deep Structured Semantic Models for Web Search using Click-through Data (Huang et al., CIKM'13)

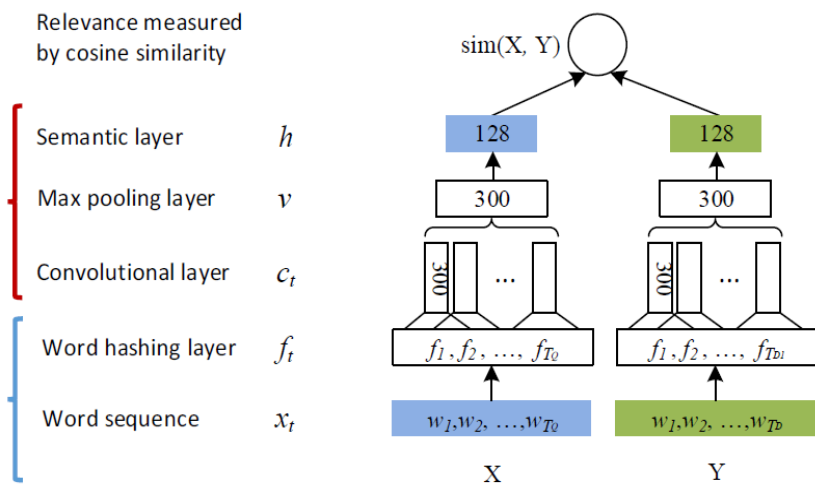


# 智能问答 -- 语义匹配

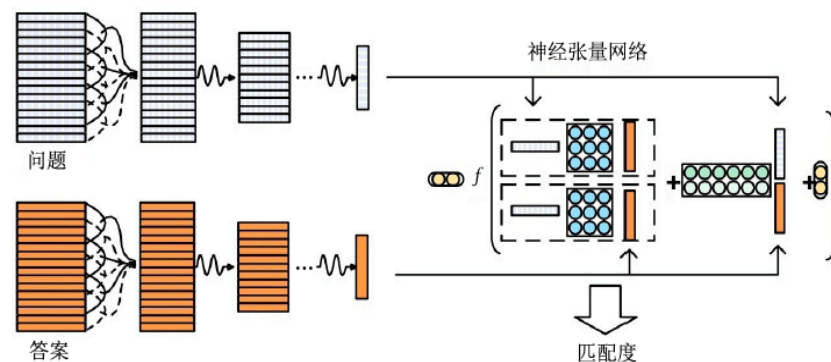


MatchPyramid

# 智能问答 -- 语义匹配



CNN-DSSM



CNTN

# 智能问答 -- 语义匹配

**Case one:**

**Q1:** What can I do for **IAS**?

**A1:** **UPSC** exam is also called **IAS** with a low pass rate.  
You must take a lot energy.

**Q2:** How do I start preparation for **UPSC**?

**A2:** To prepare for the **IAS exam**, or more precisely, the **UPSC** Civil Services Exam , self-discipline is the first.

**Case two:**

**Q1:** What should I do if I **have** a slight **fever**?

**A1:** Wiping with **alcohol** may help you, and it is my usual practice.

**Q2:** What items can **remove oil stains**?

**A2:** **Alcohol** may be a good choice. Try it out.

Figure 1: Two examples from Quora. In case one, the corresponding answers explain that the IAS is equivalent to UPSC, which is crucial for determining the relationship between the both questions. In case two, the questions have distinct semantics, but their paired answers are semantically similar.

短文本蕴含信息不足，通过自适应注意力机制从外部知识中抽取信息，增强文本表示

# 智能问答 -- 语义匹配

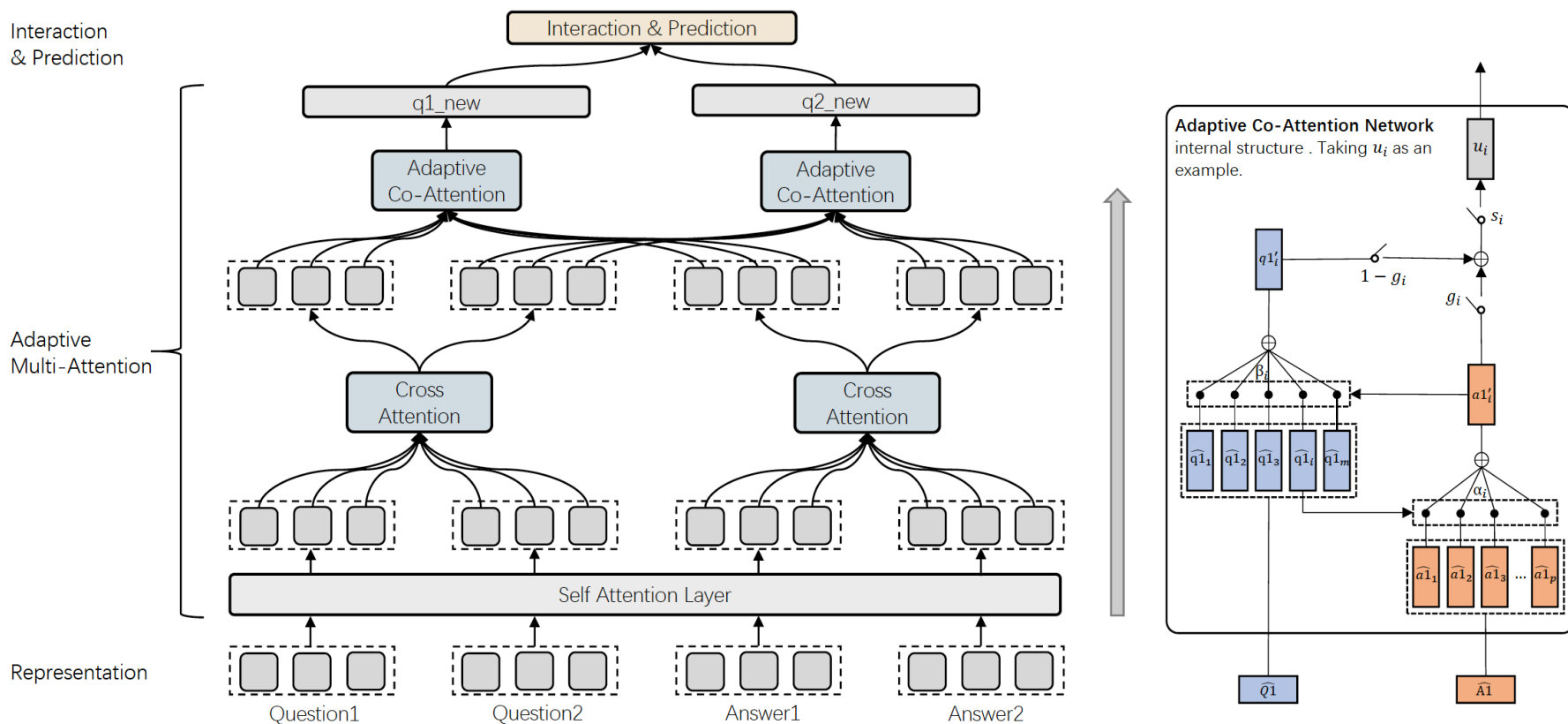


Figure 2: The overall view of our model. The left part is the main framework of this work. The right part is the detailed structure of the adaptive co-attention network. The thick arrow indicates that information flows from bottom to top.

# 智能问答 -- 语义匹配

Table 3: Overall results on AeQQP.

| <b>Model</b>                | <b>Accuracy</b> |
|-----------------------------|-----------------|
| InferSent [6]               | 84.00           |
| SSE [24]                    | 86.62           |
| PWIM [11]                   | 72.59           |
| Multi-Perspective-CNN [37]  | 78.98           |
| Multi-Perspective-LSTM [37] | 79.12           |
| BiMPM [37]                  | 87.32           |
| pt-DECATT [35]              | 86.43           |
| ESIM [3]                    | 84.35           |
| AF-DMN [8]                  | 87.61           |
| DIIN [4]                    | 88.20           |
| <b>AMAN(ours)</b>           | <b>90.07</b>    |

# 智能问答 -- 语义匹配

Table 4: Overall results on CQADupStack.

| MODEL                       | Precision    | Recall       | F1           | Accuracy     |
|-----------------------------|--------------|--------------|--------------|--------------|
| Multi-Perspective-CNN [37]  | 82.81        | 92.23        | 87.13        | 90.12        |
| Multi-Perspective-LSTM [37] | 83.64        | 94.08        | 87.98        | 90.15        |
| BiMPM [37]                  | 84.78        | 97.21        | 90.04        | 94.14        |
| ESIM [3]                    | 87.83        | 95.20        | 90.81        | 93.85        |
| AF-DMN [8]                  | 89.22        | 93.66        | 90.92        | 94.72        |
| DIIN [4]                    | 89.46        | 94.60        | 91.36        | 94.73        |
| <b>AMAN(ours)</b>           | <b>90.52</b> | <b>97.87</b> | <b>94.05</b> | <b>96.28</b> |

# 智能问答 -- 语义匹配

---

**Premise:** A dog is jumping for a Frisbee in the snow.

**Hypothesis:** An animal is playing with a plastic toy.

**Label:** Entailment

---

**Premise:** He was crying like his mother had just walloped him.

**Hypothesis:** He was crying like his mother hit him with a spoon.

**Label:** Neutral

---

**Premise:** Several men in front of a white building.

**Hypothesis:** Several people in front of a gray building.

**Label:** contradiction

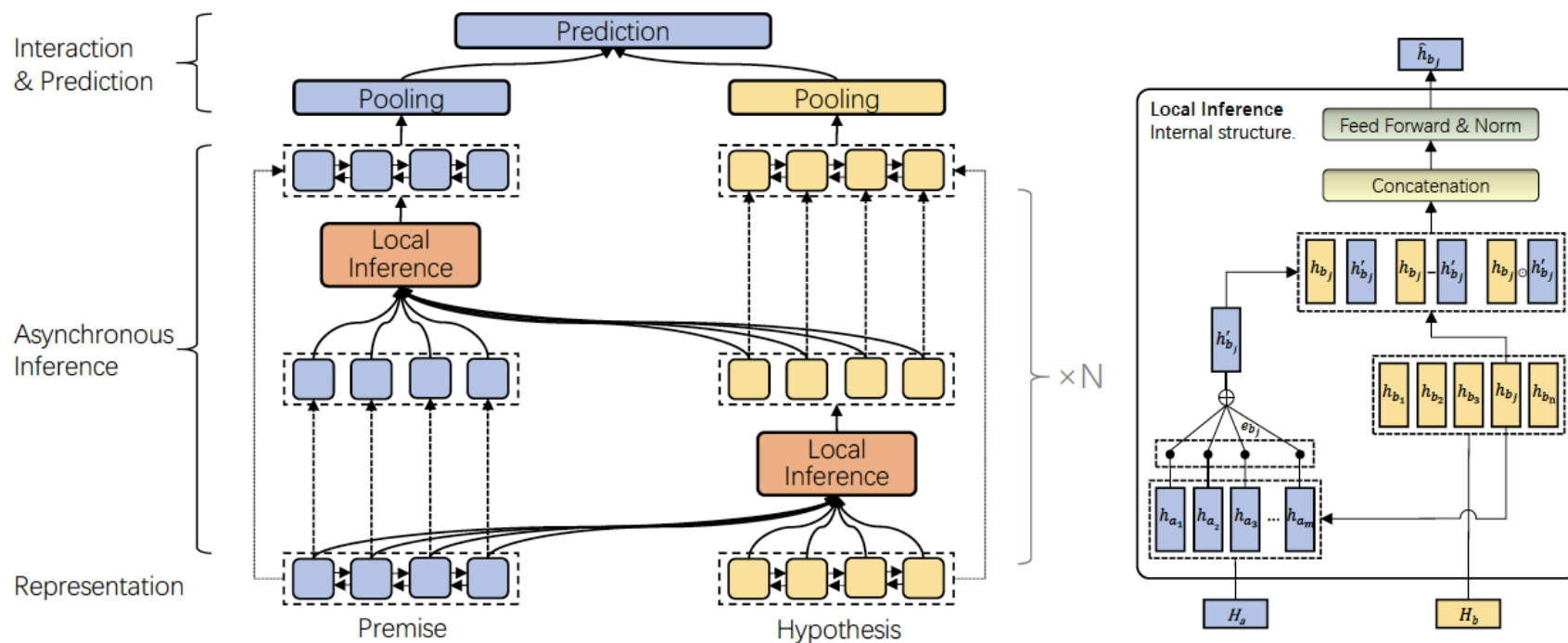
---

Existing methods typically have framed the reasoning problem as a semantic matching task.

The both sentences are encoded and interacted symmetrically and in parallel.

The role of the two sentences is obviously different.

# 智能问答 -- 语义匹配





# 智能问答 -- 语义匹配

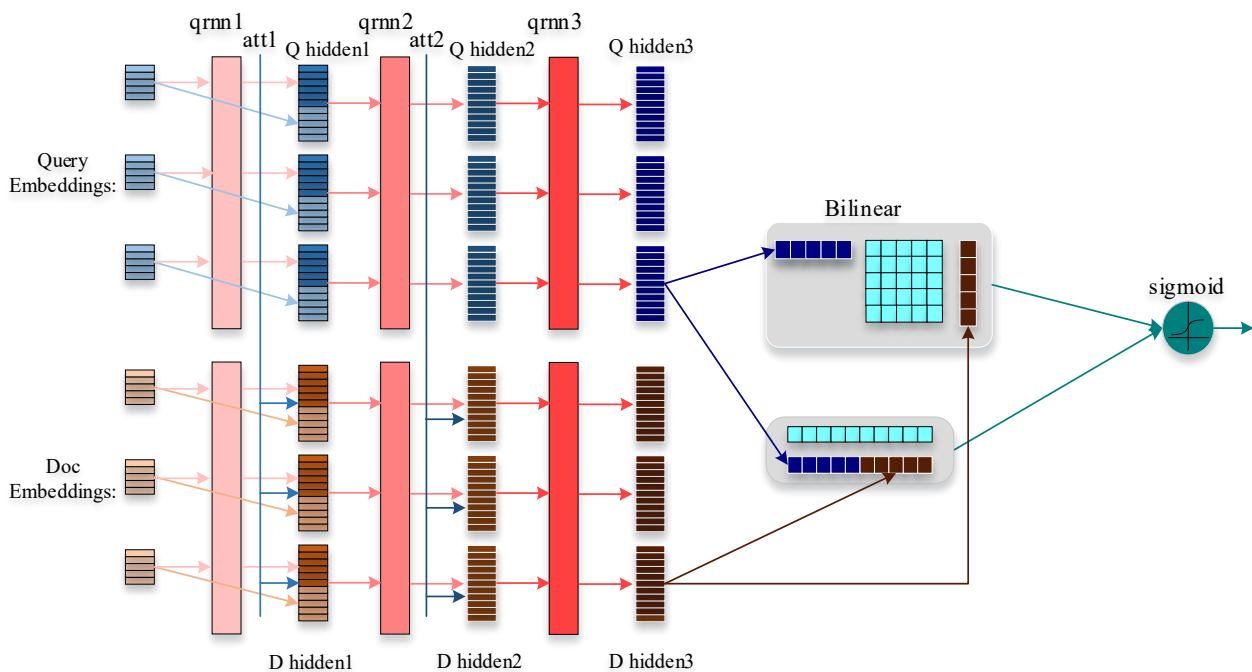
| Model                                   | train | test        |
|---|-------|-------------|
| <b>Single Models</b>                    |       |             |
| 200D DecompAtt<br>(Parikh et al., 2016) | 90.5  | 86.8        |
| 600D ESIM (Chen et al., 2016)           | 92.6  | 88.0        |
| BiMPM (Wang et al., 2017)               | 90.9  | 87.5        |
| 448D DIIN (Gong et al., 2017)           | 91.2  | 88.0        |
| 300D MwAN (Tan et al., 2018)            | 94.5  | 88.3        |
| 300D CAFE (Tay et al., 2018)            | 89.8  | 88.5        |
| 300D AF-DMN<br>(Duan et al., 2018)      | 94.5  | 88.6        |
| KIM (Chen et al., 2018)                 | 94.1  | 88.6        |
| <b>ADIN (ours)</b>                      | 93.6  | <b>88.8</b> |
| <b>Ensemble Models</b>                  |       |             |
| 600D ESIM (Chen et al., 2016)           | 93.5  | 88.6        |
| BiMPM (Wang et al., 2017)               | 93.2  | 88.8        |
| 448D DIIN (Gong et al., 2017)           | 92.3  | 88.9        |
| 300D AF-DMN<br>(Duan et al., 2018)      | 94.9  | 89.0        |
| 300D CAFE (Tay et al., 2018)            | 92.5  | 89.3        |
| <b>ADIN (ours)</b>                      | 95.6  | <b>89.5</b> |

Table 3: Comparison with previous models on the SNLI dataset.

| Model                         | Test Accuracy |             |
|-------------------------------|---------------|-------------|
|                               | Matched       | Mismatched  |
| <b>Single Models</b>          |               |             |
| ESIM<br>(Chen et al., 2016)   | 72.3          | 72.1        |
| DIIN<br>(Gong et al., 2017)   | 78.8          | 77.8        |
| AF-DMN<br>(Duan et al., 2018) | 76.9          | 76.3        |
| CAFE<br>(Tay et al., 2018)    | 78.7          | 77.9        |
| MwAN<br>(Tan et al., 2018)    | 78.5          | 77.7        |
| <b>ADIN (ours)</b>            | <b>78.8</b>   | <b>77.9</b> |
| <b>Ensemble Models</b>        |               |             |
| DIIN<br>(Gong et al., 2017)   | 80.0          | 78.7        |
| CAFE<br>(Tay et al., 2018)    | 80.2          | 79.0        |
| MwAN<br>(Tan et al., 2018)    | 79.8          | 79.4        |
| <b>ADIN (ours)</b>            | <b>80.3</b>   | <b>79.6</b> |

Table 4: Comparison with previous models on the MultiNLI dataset.

# 智能问答 -- 语义匹配

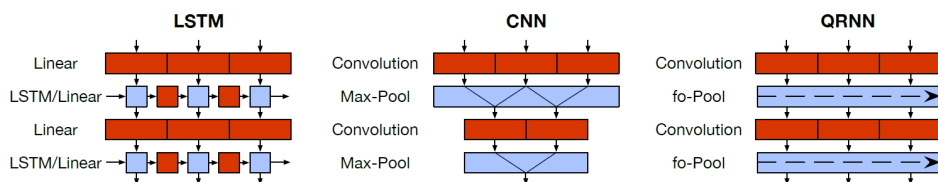


多层表示学习 QRNN-ATT

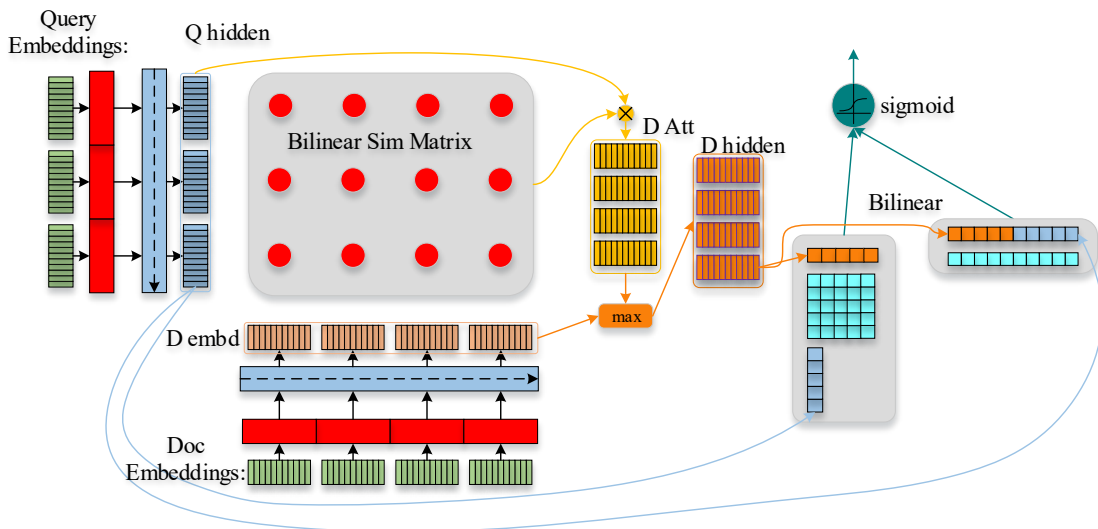
Query和Doc的多层交互，三层QRNN提供了由浅到深的语义编码能力

Attention机制建模Query和Doc的语义一致性，利用Query重构Doc上下文表示的能力，语义信息层层递进

# 智能问答 -- 语义匹配



基于QRNN和注意力机制的表示学习  
序列表示，3倍加速于LSTM  
注意力机制加强Query-Doc的交互



单层QRNN-ATT (Sogou)

# 智能问答 -- 语义匹配

## Query-问题匹配 精度损失

(实际为不相同, 模型判定为相同)

|                    |                          |
|--------------------|--------------------------|
| Chrome浏览器怎么删除      | Chrome浏览器 <b>历史</b> 怎么删除 |
| 哪些精致的游戏            | 游戏的细节能精致到什么程度            |
| 男生自己理头发            | 男生 <b>多久</b> 理一次头发合适     |
| 中国的美食 <b>有哪些</b>   | 美食中国                     |
| 国庆节手抄报 <b>怎么</b> 画 | 国庆手抄报的画                  |
| 公安和狱警哪个好           | 公安跟狱警哪个好 <b>考</b> 些      |
| 港澳通行证如何续签          | <b>卡式</b> 往来港澳通行证怎么续签    |
| 简单的赚钱方法            | <b>网上</b> 最简单的赚钱方法       |

# 智能问答 -- 语义匹配

## Query-问题匹配 召回损失

(实际为相同，模型判定为不相同)

|            |                              |
|------------|------------------------------|
| 手机为什么冲不进电  | 为何手机充不进去电该怎么解决               |
| 华为手机怎样截长屏  | 华为手机怎么长截图三种方式实现长图截屏          |
| 司法考自学      | 如何通过自学参加司法考试                 |
| 洗衣液瓶子废物利用  | 洗衣液瓶子别扔了这样改造放在卫生间太实用了看到的人都夸好 |
| 显腿长的穿法     | 除了高跟鞋这些显腿长的穿法你知道几个           |
| 蚂蚁会员有几个    | 蚂蚁会员一共有几个等级                  |
| 小程序可以有朋友圈吗 | 微信小程序怎么分享朋友圈                 |
| 打工和创业的区别   | 创业和打工的差别有多大看完此文就明白为什么和老板想的不同 |

# 智能问答 -- 语义匹配

## 答案质量问题

路由器重启后电脑无法联网怎么办-太平洋IT百科

路由器重启后电脑无法联网的解决方法: 1、此类故障的停网一般不会出现全部电脑无法... 2、选择网络连接选项。 3、在本..

product.pconline.com.cn

如果是电脑宽带上不了网,可...

1. 路由器问题: (1)查看路由器灯是否正常,如果是光猫,一般除了LOS灯不亮外,其他灯是亮的。可关闭路由器和猫电源,过5分钟后再重新打开。(2)如光猫灯出现异常情况,可查看光猫光接口...
2. 电脑系统配置问题: (1)重置网络:在桌面按WIN+R输入CMD点击确定打开命令提示符窗口,win10按win+X直接打开命令提示符,在命令提示符中输入:netsh winsock reset,稍后,会有成功...

我重启了路由器 电脑没网络 怎么办...  
zhidao.baidu.com



# 智能问答 -- 语义匹配

## 答案非所问



### 如何提高孩子的专注力 立知

不同年龄阶段的孩子专注力的“时长”并非一样，比如2岁以下的宝宝是以被动注意为主，2-3岁的宝宝是7-9分钟，4-6岁的孩子是15-30分钟，7岁以上的孩子是35分钟以上。所以家长掌握孩子注意力特征后，要合理判断孩子专注力发展，不可以盲目对孩子施加压力，造成孩子内心恐慌。同时，孩子在表现不同情绪时，专注力的“时长”也会不同。此时，训练孩子专注力第一件事，是安抚孩子情绪，在平静祥和的状态进行，效果最佳。

养娃视频

纠错

### 如何提高新闻稿写作能力

答：如何提高新闻写作能力,以下文字，仅供参考：1、充分认识新闻宣传的重要性新闻宣传是喉舌，是要将各项工作、所取得的各种成绩进行广泛的宣传，使人认识了解，树立良好的外部形象，赢得好的声誉。同时还要热爱新闻宣传报道这项工作，增强责...

2017-01-17 回答者: 绿日证书人才 2个回答 4

### 床垫有凹陷怎么办?

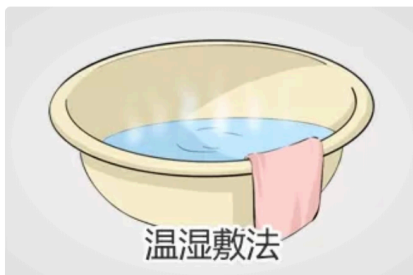
答：梦神床垫温馨提醒：建议三到四个月左右，将床垫定期对调和翻身，可以上下翻转或首尾对调（功能型床垫不建议上下翻转，一般家庭3至6个月换位一次即可）使垫面均匀受力，避免凹陷，延长使用寿命；不要经常坐在床的边缘，因为床垫的4个角最为脆弱，...

2019-09-26 回答者: olhdjhp616 2个回答

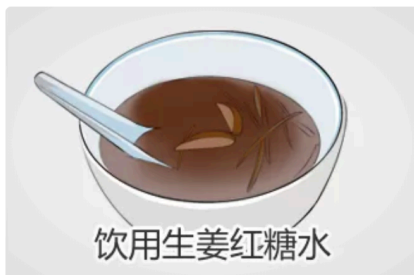
# 智能问答 -- 语义匹配

## 答案权威性问题的

### 发烧退烧土办法



温湿敷法



饮用生姜红糖水

发烧后除了选择物理降温 and 药物降温外，日常生活中也有一些土方法有利于降温。常用的退烧土方法包括以下几种：

- 温湿敷法：将毛巾浸入热水中，水温以不烫手为度，浸湿之后不断擦洗发烧患者的额头、胸、背、四肢，擦完后再将湿毛巾敷于患者的胸部。温湿敷可以使皮肤血...
- 饮用生姜红糖水：生姜性温、味辛，具有散寒发汗、解表祛风作用，尤其适合风寒感冒发烧者食用。一般用生姜及红糖适量冲泡饮用，发汗后可退烧；
- 食用葱白大米粥：葱白性温、味辛，具有调节体温，促进汗腺正常排汗的作用，可减少和预防伤风感冒的发生。受凉感冒后，可以用葱白和小米煮成稀粥频频食用；

更多内容 >



梁连春 | 主任医师

首都医科大学附属北京佑安医院 [三级甲等](#)



中国医师协会医学科学普及分会



### 成年人发烧了怎么办? - 快速问医生

当体温在38.5°C以下首先应给予多饮温水,同时可采用冰敷防止大血管的部位如:颈部两侧和腹股沟的部位和腋下,另外建议用约50白酒加在温水里面混匀后擦拭除腹部背部胸部的其他位置,最好是大血管分布的部位,四肢头部都是可以的啊.干了又擦干了又擦.直到温度下降至正常啊.当然如...



许能文 | 主治医师

丽水市中心医院 [三级甲等](#)

快速问医生 - m.120ask.com

### 发烧39度怎么办? - 快速问医生

- 1、**发热**是机体受到有害刺激后的一种反应形式,一定程度的**发热**可以唤醒机体的抵抗力,提高自身的抗病能力,对疾病的康复是有利的,导致**发烧**的原因一般考虑与感染有关,需要抗炎和退烧治疗.
- 2、**发热**应予紧急处理,采用冷敷和拭浴的方法,尽快降温.可用冷湿毛巾敷额,还可用不漏水的塑...



马建平 | 副主任医师

宁夏回族自治区石嘴山市第三人民医院

快速问医生 - m.120ask.com



# 智能问答 -- 语义匹配

## 答案时效性问题

### 现在北京大学校长是谁？ - 搜狗问问

1个答案 - 提问时间: 2008-12-27 - 9个赞

**最佳答案** 周其凤 个人简历 男，汉族，1947年10月8日（农历）出生，湖南浏阳人，中共党员，理学博士。中国科学院院士，教授，博士生导师，2008年11月开始担任北京大学校...

搜狗问问 - [wenwen.sogou.com/z...](http://wenwen.sogou.com/z...) - 2008-12-27 - 快照

### 北大现任的校长是谁？

3个答案 - 提问时间: 2014-02-20

**最佳答案** 1、北大现任的校长是林建华。2、林建华，男，汉族，1955年10月出生，山东高密人，中共党员，1973年11月参加工作，1986年北京大學化学系毕业，研究生学历，理学...

搜狗问问 - [wenwen.sogou.com/z...](http://wenwen.sogou.com/z...) - 2014-2-21 - 快照

# 智能问答 -- 语义匹配

## 语料库怎么构建

The screenshot shows the JD.com Help Center interface. At the top, there is a navigation bar with links for '请登录', '免费注册', '我的订单', '我的京东', '京东会员', '企业采购', '手机京东', '关注京东', '客户服务', and '网站导航'. Below this is the '京东 帮助中心' header with a search bar and a '搜索' button. A sidebar on the left lists '常见问题分类' including '购物指南', '订单百事通', '配送方式', '支付问题', '发票问题', '售后服务', '售后政策 (自营)', '售后政策 (三方)', '退换货/维修申请', '售后服务单查询与修改', '售后商品返回', '售后退款说明', '售后到家查询', '售后保障服务', and '售后常见问题'. The main content area shows a breadcrumb trail: '售后服务 > 售后服务单查询与修改 > 售后服务单审核时效'. The title of the article is '我提交的售后返修单何时能为我审核?'. The content includes two numbered points: '1. 京东自营商品: 当天18:00前提交申请, 一般可在当天22:00前审核完毕; 当天18:00后提交申请, 一般可在次日12:00前审核完毕;' and '2. 第三方卖家商品: 申请后48小时内审核完毕 (部分商家双休, 节假日不工作, 审核会有延迟);'. Below the text is a poll asking '这条帮助是否解决了您的问题?' with '已解决' and '未解决' buttons. At the bottom, there is a section for '猜你喜欢的问题' with a sub-heading '售后返修单已提交成功, 为什么没有给我安排上门取件?'.



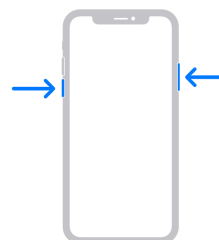
## Restart your iPhone

Learn how to turn your iPhone off, then back on.



## How to restart your iPhone X or 11

1. Press and hold either volume button and the side button until the power off slider appears.



# 智能问答 -- 语义匹配

## 弱监督数据如何使用

| uid                      | time                | action                                 | query          | url  |
|--------------------------|---------------------|--|----------------|--|
| 00a7daa77545146095dc54f4 | 2018-07-16 14:39:27 | 主动输入 "去黑头的产品"                          | 去黑头的产品         | null   |
| 00a7daa77545146095dc54f4 | 2018-07-16 14:40:40 | 点击 第 2 条 "结构化"                         | 去黑头的产品         | http://jingyan.baidu.com/m/article/f3e34a128bb3a1f5eb653522.html   |
| 00a9537359d40999fd2afd5c | 2018-07-16 13:43:59 | 主动输入 "联通邮箱格式"                          | 联通邮箱格式         | null   |
| 00a9537359d40999fd2afd5c | 2018-07-16 13:44:07 | 点击 第 2 条 "聚合"                          | 联通邮箱格式         | https://zhidao.baidu.com/question/65005370852099205?g_f=11301026   |
| 00419442d69c8876bc3d616  | 2018-07-16 01:02:03 | 主动输入 "法国比克罗地亚的赔率多少"                    | 法国比克罗地亚的赔率多少   | null   |
| 00419442d69c8876bc3d616  | 2018-07-16 01:02:06 | 点击 第 1 条 "结构化"                         | 法国比克罗地亚的赔率多少   | http://m.yiwan.com/news/437226.html  |
| 00419442d69c8876bc3d616  | 2018-07-16 01:02:28 | 点击 第 2 条 "结构化"                         | 法国比克罗地亚的赔率多少   | http://m.yiwan.com/news/437168.html  |
| 00419442d69c8876bc3d616  | 2018-07-16 01:03:16 | 点击 第 3 条 "结构化"                         | 法国比克罗地亚的赔率多少   | http://m.candou.com/news/843100.shtml  |
| 00419442d69c8876bc3d616  | 2018-07-16 01:03:28 | 点击 第 5 条 "结构化"                         | 法国比克罗地亚的赔率多少   | http://hz.edushi.com/bang/3g/info/6-132-n4452291.html  |
| 00419442d69c8876bc3d616  | 2018-07-16 01:04:11 | 点击 第 10 条 "结构化"                        | 法国比克罗地亚的赔率多少   | http://m.redshu.com/article/1399359.html   |
| 00419442d69c8876bc3d616  | 2018-07-16 01:04:22 | 点击 第 11 条 "结构化"                        | 法国比克罗地亚的赔率多少   | http://m.shhuoche.com/a-205-43825-0.html   |
| 00419442d69c8876bc3d616  | 2018-07-16 01:04:51 | 点击 第 13 条 "结构化"                        | 法国比克罗地亚的赔率多少   | http://m.12349.net/app/421175.html   |
| 00419442d69c8876bc3d616  | 2018-07-16 01:04:57 | 点击 第 13 条 "普通"                         | 法国比克罗地亚的赔率多少   | https://m.sogou.com/web/searchList.jsp?keyword=%E6%B3%95%E5%9B%BDvs%E5%85%8B%E7%BD%97%E5%9C%B0%E               |
| 00419442d69c8876bc3d616  | 2018-07-16 01:04:58 | 主动修改 "法国vs克罗地亚赔率"                      | 法国vs克罗地亚赔率     | null   |
| 00419442d69c8876bc3d616  | 2018-07-16 01:05:01 | 点击 第 1 条 "聚合"                          | 法国vs克罗地亚赔率     | http://m.game234.com/yxrd/20180715/3300798.html  |
| 00419442d69c8876bc3d616  | 2018-07-16 01:05:29 | 点击 第 3 条 "结构化"                         | 法国vs克罗地亚赔率     | http://m.candou.com/news/843060.shtml  |
| 00419442d69c8876bc3d616  | 2018-07-16 01:05:43 | 点击 第 4 条 "结构化"                         | 法国vs克罗地亚赔率     | http://m.candou.com/news/843100.shtml  |
| 00419442d69c8876bc3d616  | 2018-07-16 01:06:30 | 点击 第 6 条 "普通"                          | 法国vs克罗地亚赔率     | http://m.candou.com/news/842256.shtml  |
| 00419442d69c8876bc3d616  | 2018-07-16 01:06:43 | 点击 第 9 条 "结构化"                         | 法国vs克罗地亚赔率     | http://m.yiwan.com/news/437168.html  |
| 00419442d69c8876bc3d616  | 2018-07-16 01:07:14 | 点击 第 11 条 "结构化"                        | 法国vs克罗地亚赔率     | http://msports.eastday.com/a/180713193655518000000.html  |
| 00419442d69c8876bc3d616  | 2018-07-16 09:30:36 | 主动修改 "看任何电影的网站7080"                    | 看任何电影的网站7080   | null   |
| 00419442d69c8876bc3d616  | 2018-07-16 09:30:41 | 点击 第 2 条 "普通"                          | 看任何电影的网站7080   | http://m.yiwan.com/soft/481911.html  |
| 00419442d69c8876bc3d616  | 2018-07-16 09:30:50 | 点击 第 1 条 "结构化"                         | 看任何电影的网站7080   | http://m.quxiu.com/p/273341_1.html   |
| 00419442d69c8876bc3d616  | 2018-07-16 09:31:04 | 点击 第 5 条 "结构化"                         | 看任何电影的网站7080   | https://m.sogou.com/web/searchList.jsp?keyword=%E7%A5%9E%E9%A9%AC%E7%94%B5%E5%BD%B17080&uID=LPxb5ji            |
| 00419442d69c8876bc3d616  | 2018-07-16 09:31:05 | 从 结果页中间hint 切入 "神马电影7080"              | 神马电影7080       | null   |
| 00419442d69c8876bc3d616  | 2018-07-16 09:31:06 | 点击 第 1 条 "普通"                          | 神马电影7080       | http://www.cnhitao.cn/114368.html  |
| 00419442d69c8876bc3d616  | 2018-07-16 09:31:22 | 点击 第 2 条 "结构化"                         | 神马电影7080       | http://tv.07dyy.com/?m=vod-play-id-2833-src-1-num-1.html   |
| 00419442d69c8876bc3d616  | 2018-07-16 09:31:43 | 点击 第 3 条 "结构化"                         | 神马电影7080       | http://www.7k7k.tv/?1343193137   |
| 00419442d69c8876bc3d616  | 2018-07-16 09:32:53 | 点击 第 7 条 "普通"                          | 神马电影7080       | http://su3j.club/  |
| 00419442d69c8876bc3d616  | 2018-07-16 09:33:37 | 点击 第 8 条 "普通"                          | 神马电影7080       | http://www.moxgwz2dyit73cywhmhvve.news44.4i6cw0.cn/  |
| 00419442d69c8876bc3d616  | 2018-07-16 09:34:57 | 主动修改 "看任何电影的网站7080"                    | 看任何电影的网站7080   | null   |
| 00419442d69c8876bc3d616  | 2018-07-16 09:35:01 | 点击 第 2 条 "普通"                          | 看任何电影的网站7080   | http://m.yiwan.com/soft/481911.html  |
| 00419442d69c8876bc3d616  | 2018-07-16 09:37:58 | 点击 第 2 条 "普通"                          | 看任何电影的网站7080   | http://m.yiwan.com/soft/481911.html  |
| 00419442d69c8876bc3d616  | 2018-07-16 09:38:13 | 点击 第 4 条 "结构化"                         | 看任何电影的网站7080   | http://m.9553.com/zt/7080yysjbxzz/   |
| 00419442d69c8876bc3d616  | 2018-07-16 09:38:24 | 点击 第 6 条 "结构化"                         | 看任何电影的网站7080   | https://www.douban.com/event/14496487/   |
| 00419442d69c8876bc3d616  | 2018-07-16 09:38:38 | 主动修改 "快播电影天堂"                          | 快播电影天堂         | null   |
| 00419442d69c8876bc3d616  | 2018-07-16 09:38:38 | 点击 第 9 条 "VR"                          | 看任何电影的网站7080   | https://m.sogou.com/web/searchList.jsp?s_from=kmap_recommend&pid=sogou-mobb-3b3dda5d26f6b954&dp=1&keyword=%l   |
| 00419442d69c8876bc3d616  | 2018-07-16 09:38:41 | 点击 第 1 条 "VR"                          | 快播电影天堂         | https://wapv.sogou.com/qdetal/x7wlfjn47j3bthmztbq.html   |
| 00419442d69c8876bc3d616  | 2018-07-16 09:39:57 | 主动修改 "看任何电影的网站7080"                    | 看任何电影的网站7080   | null   |
| 00419442d69c8876bc3d616  | 2018-07-16 09:40:04 | 点击 第 9 条 "VR"                          | 看任何电影的网站7080   | https://m.sogou.com/web/searchList.jsp?s_from=kmap_recommend&pid=sogou-mobb-3b3dda5d26f6b954&dp=1&keyword=%l   |
| 00419442d69c8876bc3d616  | 2018-07-16 09:40:04 | 主动修改 "快播电影天堂"                          | 快播电影天堂         | null   |
| 00419442d69c8876bc3d616  | 2018-07-16 09:40:12 | 点击 第 7 条 "普通"                          | 快播电影天堂         | https://agame.qq.com/bbs_joy/topic.shtml?tid=8124590218117568&frompcbbs=1&page_no=1&rank_way=1&look_lz=0&tpl=0 |
| 00419442d69c8876bc3d616  | 2018-07-16 09:40:32 | 从 结果页底部hint和横屏hint 切入 "7080电 7080电影在线" | 7080电 7080电影在线 | null   |

# 智能问答 – 阅读理解

周杰伦 [编辑词条](#)

[+](#) 添加义项 | [日](#) 同义词 | [★](#) 收藏 | [分享](#)



周杰伦，1979年1月18日出生于台湾省新北市，中国台湾流行乐男歌手、音乐人、演员、导演、编剧、监制、商人。2000年被吴宗宪发掘并推出首张专辑《Jay》。2005年涉足影视，以电影《头文字D》获第42届台湾电影金马奖及第25届香港电影金像奖“最佳新人”奖。2007年成立JVR有限公司；同年，自编自导自演电影《不能说的秘密》，获得台湾电影金马奖年度台湾杰出电影奖。2008年凭借歌曲《青花瓷》获得第19届金曲奖最佳作曲人奖。2011年凭借专

2013年2月3日，周杰伦蜡像亮相上海杜莎夫人蜡像馆，成为首位拿到杜莎夫人蜡像馆亚洲地区所有分馆入住权的唯一华人男歌手（上海，香港，曼谷）；5月17日，“魔天伦”世界巡回演唱会在上海奔驰文化中心启动；6月1日，周杰伦配音《十万个冷笑话》第9、10集太乙真人；7月11日，周杰伦信主后首部自导自演的新片《天台爱情》首映。

周杰伦蜡像什么时间亮相上海杜莎夫人蜡像馆？

文章

+

问题



答案

# 智能问答 – 阅读理解

2015年之前的模型：单词匹配、逻辑回归、分类等

2015年之后的模型：**神经网络**

文章

+

问题



答案开始位置

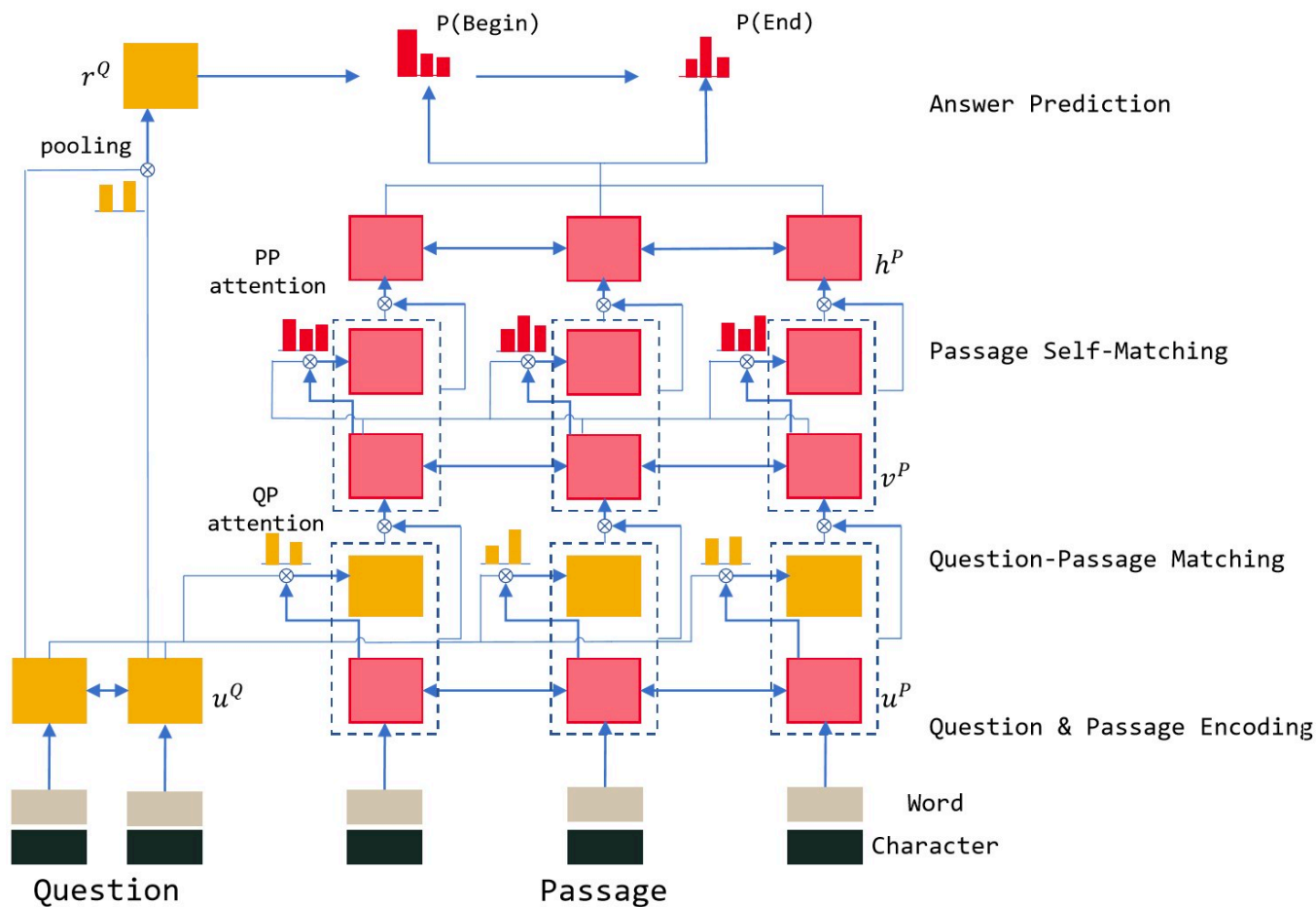
答案结束位置

# 智能问答 -- 阅读理解的挑战

面临复杂的互联网文档，文档质量、文本长度千差万别，机器阅读理解难度更大，更具挑战性。

|         | Squad1.0/2.0              | 搜索问答  |
|---------|---------------------------|---|
| 文档类型    | Wiki的段落                   | 网页文本  |
| 文档质量    | 高质量                       | 页面种类多，质量差异大<br>UGC内容低质                      |
| 文档长度    | 平均长度138个词，300词<br>以下占比98% | 平均长度约460字（视频等非文本页内容少<br>除外）<br>600字以上占比约1/3 |
| 文档是否有答案 | 1.0一定存在答案<br>2.0不一定存在答案   | 不一定存在答案<br>单个搜索结果答案可能错误<br>搜索中的答案可能客观错误     |

# 智能问答 - 阅读理解



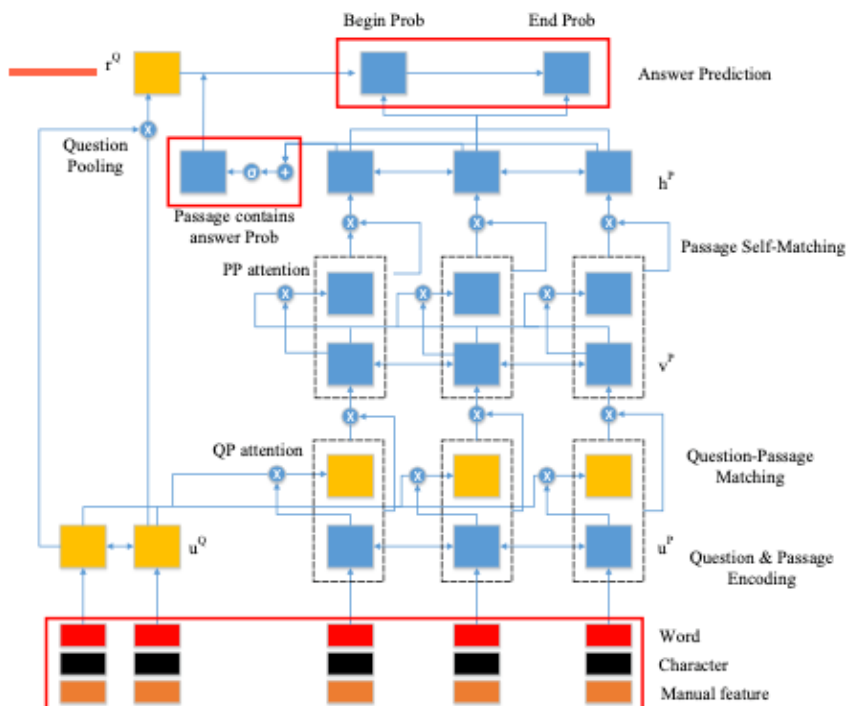
# 智能问答 - 阅读理解

## 问题

- ① 模型假设文本中存在正确答案，而真实场景面临不存在答案的情况；
- ② 模型容易出现常识性错误，对于涉及推理的问题效果不好；

## 思路

- ① 对答案存在性进行判断，避免错答；
- ② 充分利用外部知识，不仅有利于处理复杂推理问题，也有助于提高答案抽取的精度和置信度；



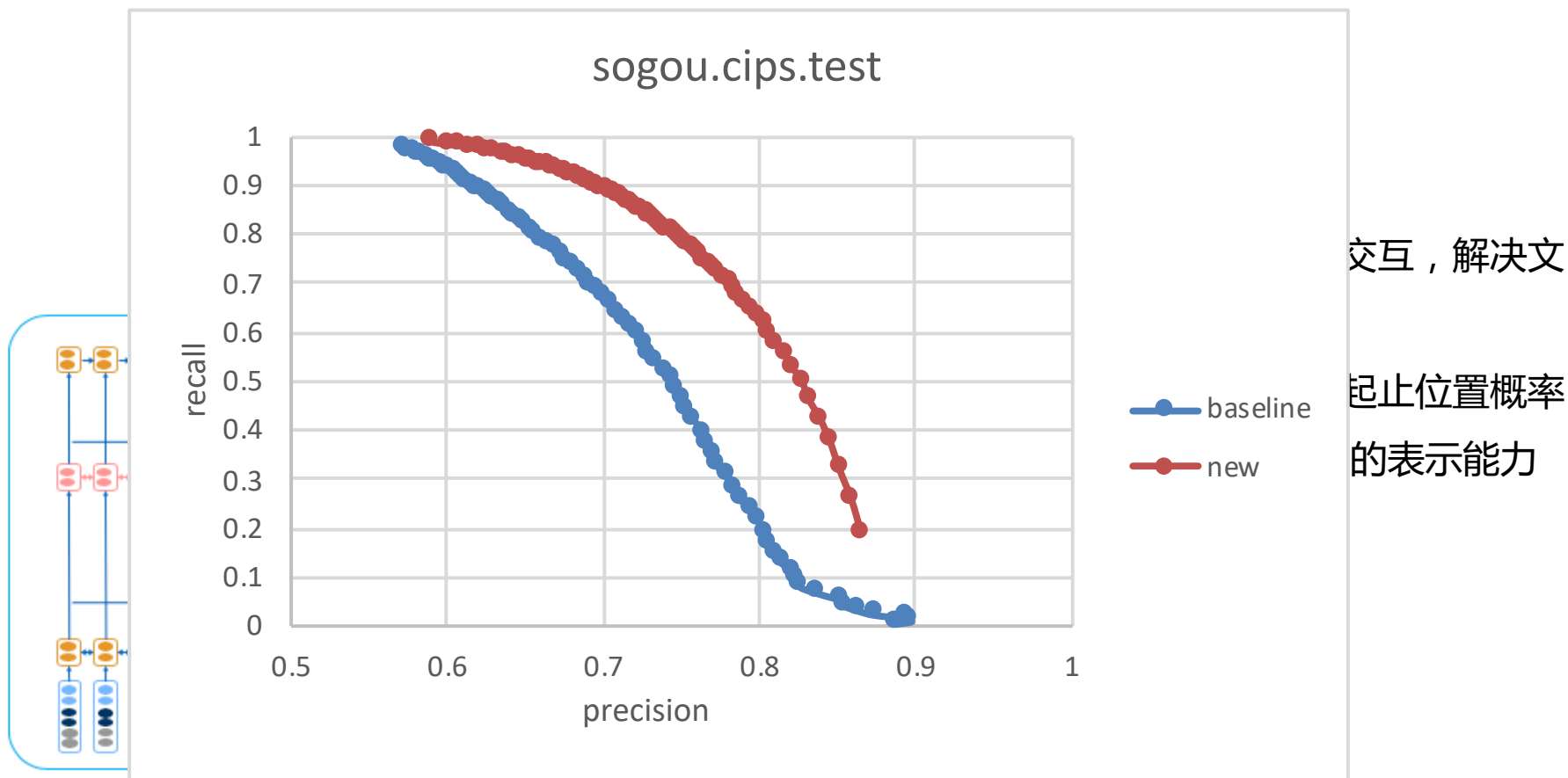
- 加入答案存在性判断
- 在字、词表示基础上加入外部知识表示 ( 全局/局部词频、LexicalAnswerType、Entity类型、POS、网页质量特征等 )



# 智能问答 – 阅读理解

**问题：**网页通常包含长文本，模型往往因为长距离依赖导致丢失重要信息

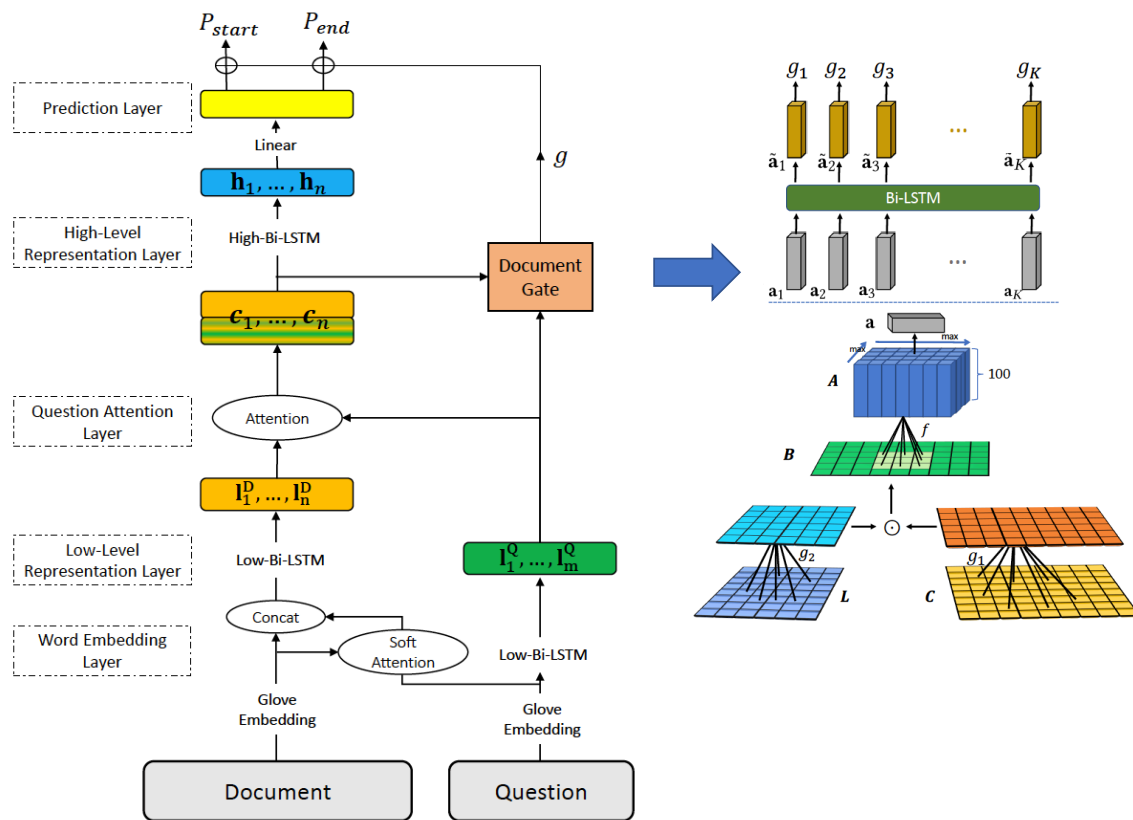
**思路：**以词为单位的表示和注意力机制基础上，融入以句子为单位的长距离信息传递和注意力机制



# 智能问答 – 阅读理解

**问题：**采用Distantly Supervised 方法标注的数据噪声影响大

**思路：**更新Loss Function考虑全部文档的结果以及文档之间的相关性，并引入基于Bootstrapping的数据增强机制



# 智能问答 – 阅读理解

**问题：**采用Distantly Supervised 方法标注的数据噪声影响大

**思路：**更新Loss Function考虑全部文档的结果以及文档之间的相关性，并引入基于Bootstrapping的数据增强机制

|             |    | BiDAF | DrQA | DrQA <sub>MTL</sub> | SR <sup>2</sup> | R <sup>3</sup> | DS-QA | w/o BD | DGR <sub>ppl</sub> | w/o DG | B-Mean | DGR  | ↑ Δ   |
|-------------|----|-------|------|---------------------|-----------------|----------------|-------|--------|--------------------|--------|--------|------|-------|
| SQuAD       | EM | 23.7  | 28.4 | 29.8                | 27.2            | 29.1           | -     | 31.5   | 31.8               | 31.1   | 31.9   | 33.9 | 13.7% |
|             | F1 | 30.9  | -    | -                   | 35.8            | 37.5           | -     | 40.6   | 39.7               | 39.0   | 41.1   | 43.6 | 16.3% |
| SearchQA    | EM | 28.6  | -    | -                   | -               | 49.0           | 58.8  | 53.3   | 53.8               | 55.2   | 56.9   | 61.4 | 4.4%  |
|             | F1 | 34.6  | -    | -                   | -               | 55.3           | 64.5  | 60.2   | 59.9               | 59.0   | 61.3   | 65.7 | 1.9%  |
| WebQuestion | EM | 10.9  | 19.5 | 20.7                | 15.6            | 17.1           | 18.5  | 21.7   | 18.9               | 18.2   | 20.9   | 22.6 | 9.2%  |
|             | F1 | 19.8  | -    | -                   | 22.5            | 24.5           | 25.6  | 25.3   | 24.8               | 21.7   | 24.2   | 27.3 | 6.6%  |
| WikiMoviews | EM | 30.6  | 34.3 | 36.5                | 38.1            | 38.8           | -     | 32.5   | 33.4               | 34.1   | 35.0   | 36.2 | -     |
|             | F1 | 36.4  | -    | -                   | 39.3            | 39.9           | -     | 37.1   | 39.2               | 38.3   | 40.6   | 42.5 | 6.5%  |

Table 2: The result of different models on open domain QA. The last column is the relative improvement of DGR over previous state-of-the-arts. All experiments have been gone through the significant test, i.e., one-tailed paired t-test with a default 95% significance level is used here. It needs to mention that for SQuAD some questions are document dependent (i.e. *Who was the university's 5th president?*), so the result is poor compared with the original machine reading settings.

# 智能问答 – 阅读理解

## ReCO: A Large Scale Chinese Reading Comprehension Dataset on Opinion

### Question:

吃完钙片可以吃香蕉吗？

(Can I eat bananas after taking calcium tablets?)

### Paragraph:

...我们说香蕉可充粮食，但不能空腹大量地吃。香蕉中含有大量的钾、磷、镁，对于正常的人，大量摄入钾和镁可使体内的钠、钙失去平衡，对健康不利。所以不可空腹食过多的香蕉。胃酸过多不宜食用过多香蕉或空腹吃香蕉，否则会引起肠胃功能紊乱。此外，提醒肾功能不全的人也不宜空腹食用香蕉或多吃，因为此类人群需要控制钾的摄入量。香蕉含有比较多的果胶，有可能会和钙反应，出现沉淀，导致腹痛等症状，也容易出现结石。咱们中国人有一个非常好的传统，那就是一些宝贵的生活常识都是从老辈开始一代代传下来的，所以，无论年龄大小，每个人都有着丰富的生活经验…

### Evidence:

香蕉含有比较多的果胶，有可能会和钙反应，出现沉淀，导致腹痛等症状，也容易出现结石

(Bananas are high in pectin, and the pectin is likely to precipitate when react with calcium which will cause abdominal pain, lithiasis and other symptoms)

### Candidate Answers:

可以 | 不可以 | 无法确定

(Yes, you can | No, you can't | uncertain)

Factoid Problem

Comprehension → Retrieval

# 智能问答 – 阅读理解

| dataset          | type  | question source  | passage source                | answer source    | datasize | question type |
|------------------|-------|------------------|-------------------------------|------------------|----------|---------------|
| SQuAD            | OQA   | human generated  | wiki passage                  | human extracted  | 100k     | F             |
| SearchQA         | OQA   | J! Archive       | search result                 | J! Archive       | 100k     | F             |
| MCTest           | MC    | human generated  | story                         | human generated  | 2,000    | NF            |
| CNN/Daily Mail   | CLOZE | abstract summary | news                          | entities         | 1.4m     | F             |
| CBT              | CLOZE | children's book  | children's book               | entities         | 680k     | NF            |
| MARCO            | OQA   | query logs       | search result                 | human generated  | 100k     | F/NF          |
| NarrativeQA      | OQA   | human generated  | books and movie scripts       | human generated  | 44k      | NF            |
| NaturalQuestions | OQA   | queries logs     | wiki document                 | human extracted  | 350k     | F             |
| DRCDD            | OQA   | human generated  | wiki document                 | human extracted  | 30k      | F             |
| DuReader         | OQA   | queries logs     | search result                 | human extracted  | 200k     | F/NF          |
| ReCO             | MC    | queries logs     | extraction from search result | human summarized | 300k     | F/NF          |

Table 1: Different MRC datasets. ‘OQA’, ‘MC’ refers to *open question answering* and *multiple-choice* respectively. *datasize* is the whole data size regardless the train/dev/test. F and NF denotes whether the question is factoid or non-factoid.

# 智能问答 – 阅读理解

| Reasoning Type                    | Example  | SQuAD | MARCO | NewsQA      | DuReader | ReCO        |
|-----------------------------------|--|-------|-------|-------------|----------|-------------|
| 1) Lexical Knowledge $\diamond$   | Q:榴莲籽可以吃吗 P:榴莲的种子富含蛋白质<br>Q: Could I eat the seed of durian<br>P: The germ of durian is high in protein.   | 70.0  | 63.0  | 84.0        | 32.5     | 26.5        |
| 2) Syntactic Knowledge $\diamond$ | Q:脑梗可以吃牛肉吗 P:牛肉在得了脑梗后可以吃<br>Q: Could I eat beef after stroke?<br>P: Beef is edible after stroking.   |       |       |             | 31.0     | 32.0        |
| 3) Coreference $\clubsuit$        | Q:白色和黄色是一样的味道么?<br>P:...白色是淡香精, 黄色是香精, 它们的味道...<br>Q:the yellow one and white one have same smell?<br>white ..., yellow, their smell is ...                  | 13.0  | 15.0  | <b>24.0</b> | 21.5     | 10.0        |
| 4) Casual Inference $\clubsuit$   | Q:环球黑卡能不能刷卡?<br>P:环球黑卡是标识卡, 标识卡不能刷卡...<br>Q:Could I pay by Global black card?P:Global black card is ID card, ID card couldn't be used for paying...          | 0.0   | 0.0   | 4.0         | 17.0     | <b>35.5</b> |
| 5) Ellipsis $\clubsuit$           | Q:上海的网吧可以抽烟么?<br>P:抽烟是命令禁止的, 网吧内一般不可以<br>Q:Could I smoking in the Internet bar of Shanghai?<br>P:Smoking is forbidden in anywhere, Internet bar is included. | 3.0   | 2.0   | 15.0        | 21.0     | <b>29.0</b> |
| 6) Logical Reasoning $\clubsuit$  | Q:所有的茶都是碱性的吗?P:乌龙茶、红茶呈弱酸性<br>Q:Are all the teas alkaline?<br>P:Oolong and black tea is acidic.   | 0.0   | 1.0   | 2.0         | 4.0      | <b>18.5</b> |
| 7) Specific Knowledge $\clubsuit$ | Q:石家庄属于北方吗? P:秦岭淮河以北的就是北方<br>Q: Is Shijiazhuang located in North China?:<br>P:North of the Huaihe River is the northern China.                               | 26.0  | 14    | <b>29.0</b> | 24.0     | 21.5        |

Table 4: Frequencies (%) of the shallow ( $\diamond$ ) or deep ( $\clubsuit$ ) reasoning skills required for MRC datasets. The results of SQuAD, MARCO and NewsQA are borrowed from Sugawara et al. (2017). We exclude the ‘undefined’ questions because the evidence may not relate to the question. Note that some question requires more than one skill.

# 智能问答 – 阅读理解

| Metric | SQuAD | SNLI  | MARCO   | DuReader | ReCO-paragraph | ReCO        |
|--------|-------|-------|---------|----------|----------------|-------------|
|        | F1    | ACC   | Rouge-L | Rouge-L  | ACC            | ACC         |
| Random | 0*    | 33.3  | 0*      | 0*       | 33.2           | 33.2        |
| BiDAF  | 77.2  | 86.7  | 23.9    | 39.0     | 55.8           | 68.3        |
| BiDAF* | 81.1  | 88.7  | 43.6    | 52.9     | 58.9           | 70.9        |
| BERT_b | 88.5  | 90.2  | 48.2    | 53.4     | 61.1           | 73.4        |
| BERT_l | 91.8  | 90.8  | -       | -        | 65.3           | <b>77.0</b> |
| Human  | 91.2  | 87.7* | 53.9    | 57.4     | 88.0           | 91.5        |

Table 5: Result of different models on ReCO and other MRC datasets. \* means the estimated results. ReCO-para denotes we use the original paragraph as the context. BERT<sub>b|l</sub> denotes BERT base or large model. The results of SQuAD, SNLI, DuReader and MARCO were derived from their leaderbord or paper.

# 智能问答 – 阅读理解

## Question Generation → CQA 基于隐含答案信息的问题生成方法

**问题：**所生成的问题是否可以被回答没有考虑

---

**Paragraph:** In accordance with his father’s wishes, **Luther** enrolled in law school at the same university that year but dropped out almost immediately, believing that law represented uncertainty. Luther sought assurances about life and was drawn to theology and philosophy.

---

**Predicted Answer:** **Luther**

Answer-Aware: **Predicted Question:** Who enrolled in law school accord with his father’s wishes?

---

Answer-Agnostic: **Predicted Question:** Why does Luther sought assurances about life?

---

训练：阅读理解语料集合包含答案  
应用：只有文档-问题信息

答案信息当做归纳偏置

Figure 1: The bad case of answer-aware and answer-agnostic NQG. In answer-aware question generation, the generated answer *Luther* is just a named entity that could be trivially inferred by subsequent text, and without much value to be asked. In answer-agnostic case the generated question is fluent but could not be answered by the paragraph.



# 智能问答 – 阅读理解

## Paragraph Encoder

$$\begin{aligned}\mathbf{q}^l, \mathbf{k}^l, \mathbf{v}^l &= \mathbf{W}_q^l \mathbf{h}^{l-1}, \mathbf{W}_k^l \mathbf{h}^{l-1}, \mathbf{W}_v^l \mathbf{h}^{l-1}, \\ \mathbf{h}^l &= \text{MultiHeadAttention}(\mathbf{q}^l, \mathbf{k}^l, \mathbf{v}^l)\end{aligned}\quad (1)$$

$$\text{MultiHeadAttention}(q, k, v) = \text{softmax}\left(\frac{\mathbf{q}^T \mathbf{k}}{\sqrt{d}}\right) \mathbf{v} \quad (2)$$

## Pivot Answer Predictor

$$\begin{aligned}\mathbf{z}_i &= \text{MLP}(\mathbf{h}_i), \\ g_i &= \sigma(\mathbf{z}_i^T \mathbf{w}_z), \\ z_i &= \begin{cases} 1, & g_i > 0.5, \\ 0, & g_i \leq 0.5 \end{cases}\end{aligned}$$

# 智能问答 – 阅读理解

## Training

$$\mathcal{L}_{\text{MLE}} = -\mathbb{E}_q \sum_i \log p(q_i | q_{<i}, \mathbf{C})$$

### Auxiliary Pivot Prediction

$$\bar{\mathbf{H}} = [\mathbf{H}; \mathbf{S}]$$

$$\tilde{\mathbf{H}} = \text{TransformerEncoder}(\bar{\mathbf{H}})$$

$$\mathbf{b} = \sigma(\mathbf{w}_h^T \tilde{\mathbf{H}}_{[0:n]}) \quad (10)$$

$$\mathcal{L}_{\text{AP}} = -\sum_i z_i \log b_i + (1 - z_i) \log(1 - b_i)$$

### Supervised Training with Golden Answers

$$\mathcal{L}_S = -\sum_i a_i \log g_i + (1 - a_i) \log(1 - g_i) \quad (12)$$

$$\mathcal{L} = \mathcal{L}_{\text{MLE}} + \lambda_1 \mathcal{L}_{\text{AP}} + \lambda_2 \mathcal{L}_S \quad (13)$$

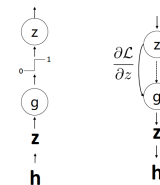


Figure 2: The forward and backward pass in our model. We use the straight through estimator to estimate the gradient of the scalar values  $z$  during backward phase.

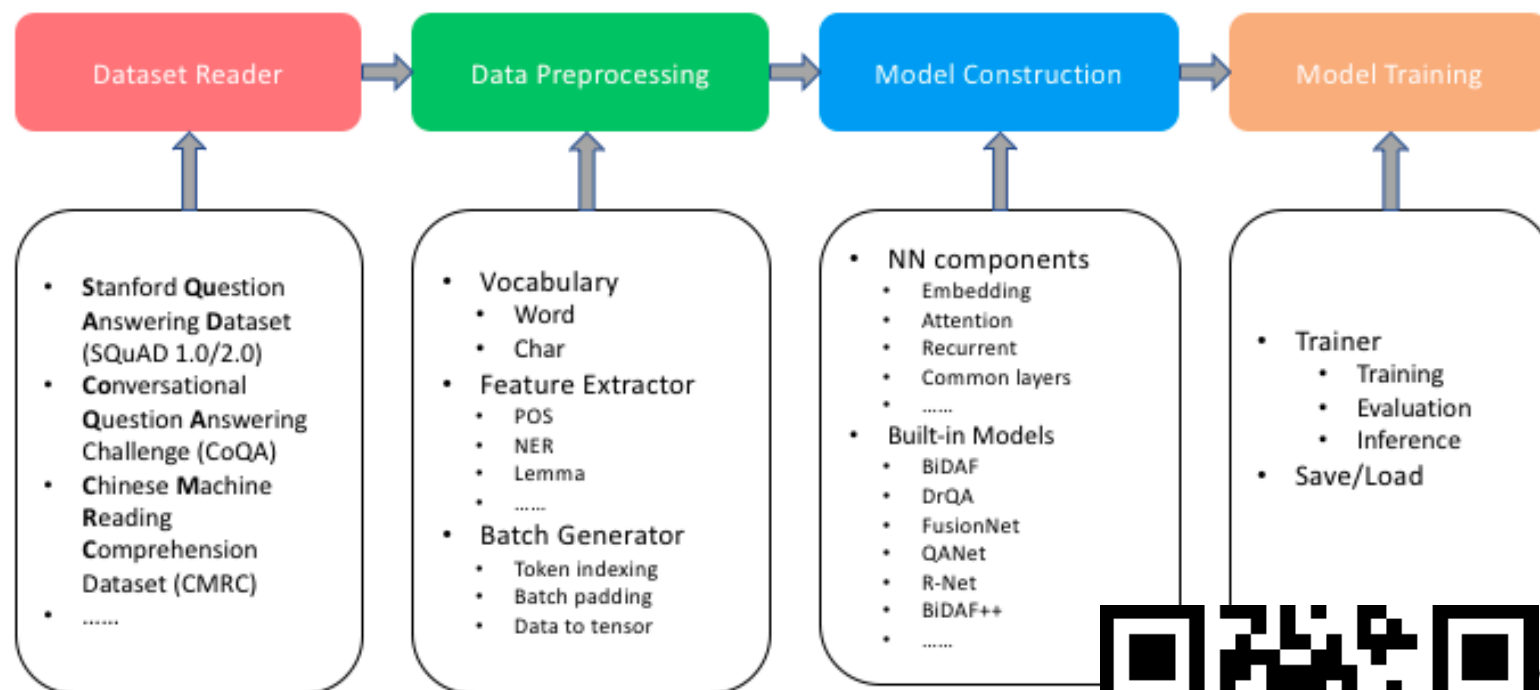
### Straight-Through-Estimator

# 智能问答 – 阅读理解

|                        | BLEU-1                             | BLEU-2                            | BLEU-3                            | BLEU-4                             | Meteor                             | Rouge-L                           |
|------------------------|------------------------------------|-----------------------------------|-----------------------------------|------------------------------------|------------------------------------|-----------------------------------|
| <i>answer-agnostic</i> |                                    |                                   |                                   |                                    |                                    |                                   |
| L2A                    | 43.09                              | 25.96                             | 17.50                             | 12.28                              | 16.62                              | 39.75                             |
| SAQG                   | 43.33                              | 26.27                             | 18.32                             | 13.23                              | -                                  | 40.22                             |
| Our model w/o AP       | 43.18 $\pm$ 0.57                   | 25.99 $\pm$ 0.36                  | 18.67 $\pm$ 0.25                  | 13.09 $\pm$ 0.18                   | 17.24 $\pm$ 0.24                   | 40.98 $\pm$ 0.37                  |
| Our model              | <b>47.24 <math>\pm</math> 0.67</b> | <b>28.03<math>\pm</math> 0.42</b> | <b>20.96<math>\pm</math> 0.29</b> | <b>14.78 <math>\pm</math> 0.17</b> | <b>18.61<math>\pm</math> 0.28</b>  | <b>41.87<math>\pm</math> 0.27</b> |
| <i>answer-aware</i>    |                                    |                                   |                                   |                                    |                                    |                                   |
| NQG++                  | 44.82                              | 26.06                             | 18.28                             | 13.02                              | 16.68                              | 41.22                             |
| SynNet                 | 47.98                              | 28.08                             | 20.45                             | 15.27                              | 17.53                              | 41.91                             |
| KPG                    | 46.23                              | 27.78                             | 20.95                             | 14.93                              | 17.24                              | 40.48                             |
| ASs2s                  | 47.09                              | 28.62                             | 19.04                             | 15.02                              | 17.02                              | 41.89                             |
| Our model              | <b>48.26 <math>\pm</math> 0.51</b> | <b>29.23<math>\pm</math> 0.30</b> | <b>22.37<math>\pm</math> 0.19</b> | <b>16.42 <math>\pm</math> 0.13</b> | <b>18.95 <math>\pm</math> 0.21</b> | <b>43.07<math>\pm</math> 0.19</b> |

Table 1: Experimental results on answer-agnostic answer-aware setting. *Our model w/o AP* refers to our model without the answer pivot layer. In answer-aware setting the ground truth answers are only provided during training. The performances of our models are reported as the mean and standard derivation values run on five different initiation.

# 智能问答 – 阅读理解



<https://github.com/sogou/SMRCToolkit>



# 智能问答 -- 挑战

- 问答是验证自然语言处理技术发展的重要途径
- 目前面临的挑战：
  1. 如何利用知识和常识
  2. 如何进行推理
  3. 如何快速低成本进行领域切换
  4. 如何构造有效的置信度机制

谢谢！