

# QQ浏览器搜索智能问答的探索与实践

腾讯 QQ浏览器搜索团队  
常景冬

## 搜索引擎的演化历程



### 第一代 人工分类

1994年第一代互联网搜索引擎Lycos诞生，它以人工分类目录为主，代表厂商是Yahoo



### 第二代 文本检索

用户输入关键字查询，信息检索模型判断网页重要性，返回给用户相关程度高的信息，最其代表性的是Google



### 第三代 整合分析

将用户输入关键字，反馈回来的海量信息，智能整合成一个门户网站式的界面，Google首次使用这种模式并且大获成功



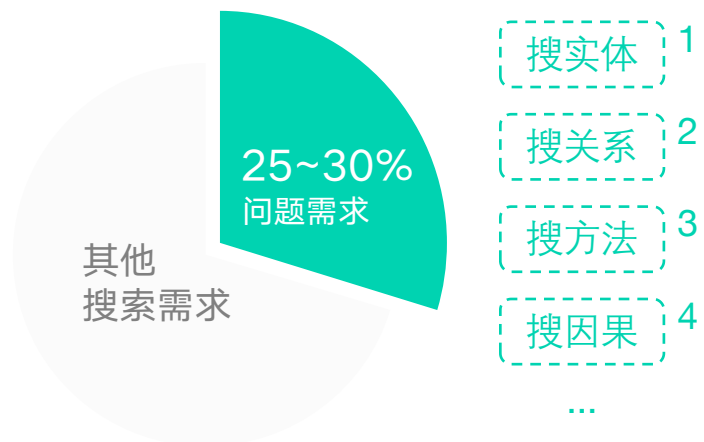
### 第四代 智能搜索

第四代搜索引擎采用特征提取和文本智能化等策略，给用户呈现数据全面、更新及时、分类细致的主题搜索结果

## 搜索中的问题需求



信息类包含直接型、间接型、建议型、定位型、列表型



## 搜索中的问答形态

Q 天空为什么是蓝色的 搜索

综合 视频 资讯 用户



是大气对太阳光散射作用，蓝色是短波中能量最大的，被大气中的微粒散射，造成了天空是蓝色的现象。

天空为什么是蓝色的 >

爱问扒拉小程序

Q 搜索结果

Q 万里悲秋常作客 搜索

综合 视频 资讯 用户

万里悲秋常作客 - 古诗文网 >

登高

[唐代] 杜甫

风急天高猿啸哀，渚清沙白鸟飞回。  
无边落木萧萧下，不尽长江滚滚来。  
万里悲秋常作客，百年多病独登台。  
艰难苦恨繁霜鬓，潦倒新停浊酒杯。

注释 译文 解析

Q 搜索结果

Q 公积金能在网上提取吗 搜索

综合 视频 资讯 用户

- 1. 打开浏览器，在浏览器中搜索你所在市的住房公积金管理中心的网站，打开主页后，点击网上办事大...
- 2. 进入登陆页面，输入相关公积金账号、密码、验证码，然后点击登录。
- 3. 登录之后，点击个人提取，选择自己对应的提取形式，这里以支付本市范围内住房租金提取为例。
- 4. 依次按照要求填写相关的提取信息，填写完成后，点击提交。
- 5. 点击提交后弹出的确认对话框，点击确定即可。

怎样在网上提取住房公积金 >

看点精准问答

Q 搜索结果

Q 刘德华的老婆 搜索

综合 视频 资讯 用户

刘德华 / 妻子



朱丽倩

朱丽倩（别名朱丽卿，1966年4月6日—），出生于马来西亚檳城，马来西亚女模特。...

查看详情 >

看点精准问答

Q 搜索结果

Q 糖尿病需要做什么检查 搜索

综合 视频 资讯 用户

权威医生解答

### 糖尿病需要做什么检查？

糖尿病患者平时在内分泌科就诊需要做哪些检查呢？首先是空腹血糖及餐后两小时血糖、糖化血红蛋白、血脂、肝功能、血常规、眼底检查、心电图、颈部及下肢血管彩超、下肢神经检查、颅脑磁共振及血管检查、心脏检查、24小时微量蛋白...

姜静 副主任医师 鹰潭市人民医院 三甲

民福康

< Q 珠穆朗玛峰多高 搜索



8848.86米

珠穆朗玛峰海拔高度

搜狗问问·微信小程序

Q 珠穆朗玛峰多高 <

Q 珠穆朗玛峰多高有多少米 <

Q 珠穆朗玛峰多高英语 <

Q 珠穆朗玛峰多高海拔多少 <

Q 珠穆朗玛非多高的 <

---

## 搜索中问答的技术线条

### KBQA

基于知识图谱的  
推理问答

”

### DeepQA

基于文本内容挖掘的  
机器阅读理解问答

”

### IRQA

基于FAQ问答库的  
检索式问答

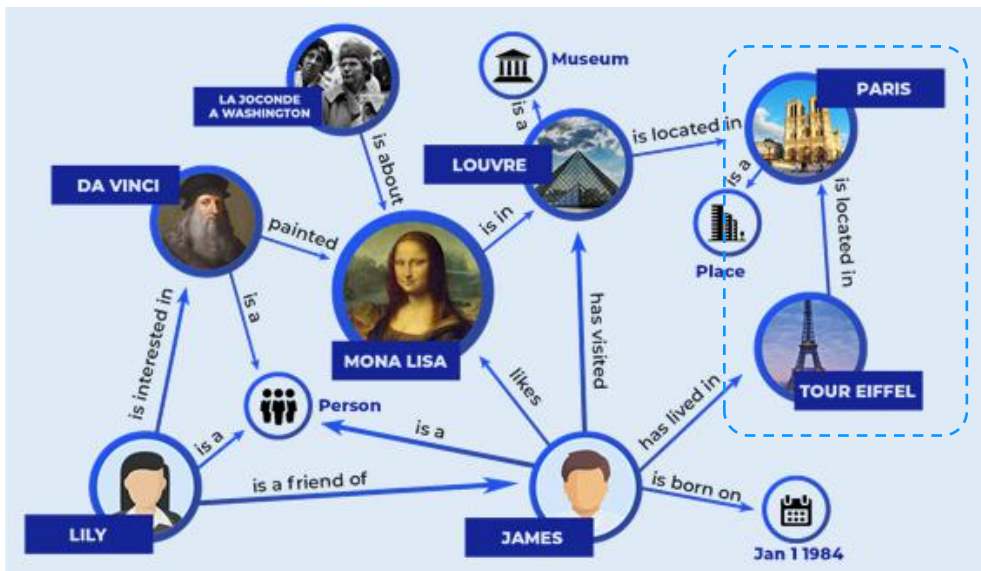
”

# KBQA

Knowledge-Based QA

## 什么是KBQA

### 知识图谱



2012年，谷歌第一次搭建起以维基百科作为背书的知识库，更名为知识图谱  
—— 使用实体、关系组成的结构化知识本体

### KBQA

以知识库作为知识来源，进行计算和推理，获取答案



问: 艾菲尔铁塔在哪里

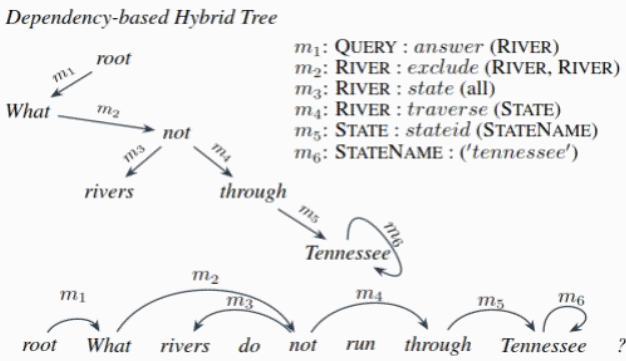
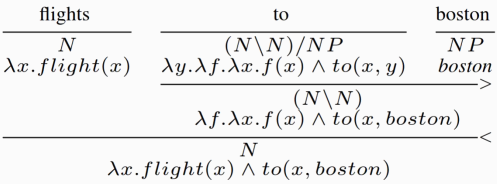


答: 巴黎

## 解决方案

### 方案一：结构化推理

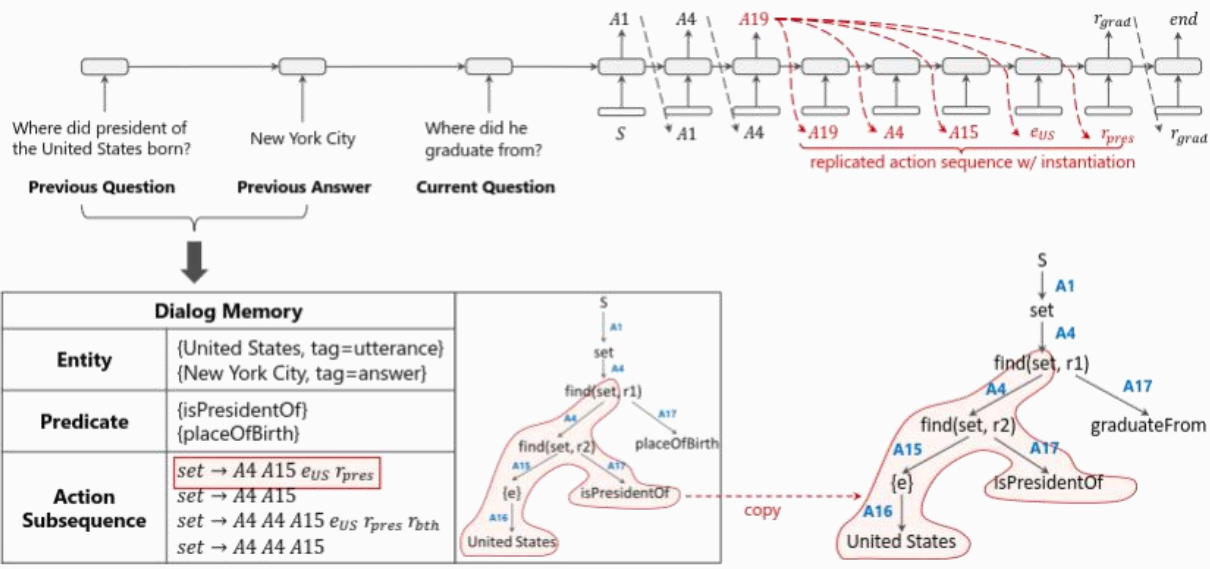
基于组合范畴语法 (CCG) 或句法依存树解析Query结构，并将其以固定的规则转换为图引擎表达式



\*期间可能使用模型辅助决策

### 方案二：结果端到端

基于神经网络，一站式完成从原始文本的输入到图引擎表达式/查询语句的输出





## 方案选择

### 腾讯QQ浏览器知识图谱



覆盖人物、影视、体育、音乐、文学、医疗等数十个大领域



亿万级  
实体

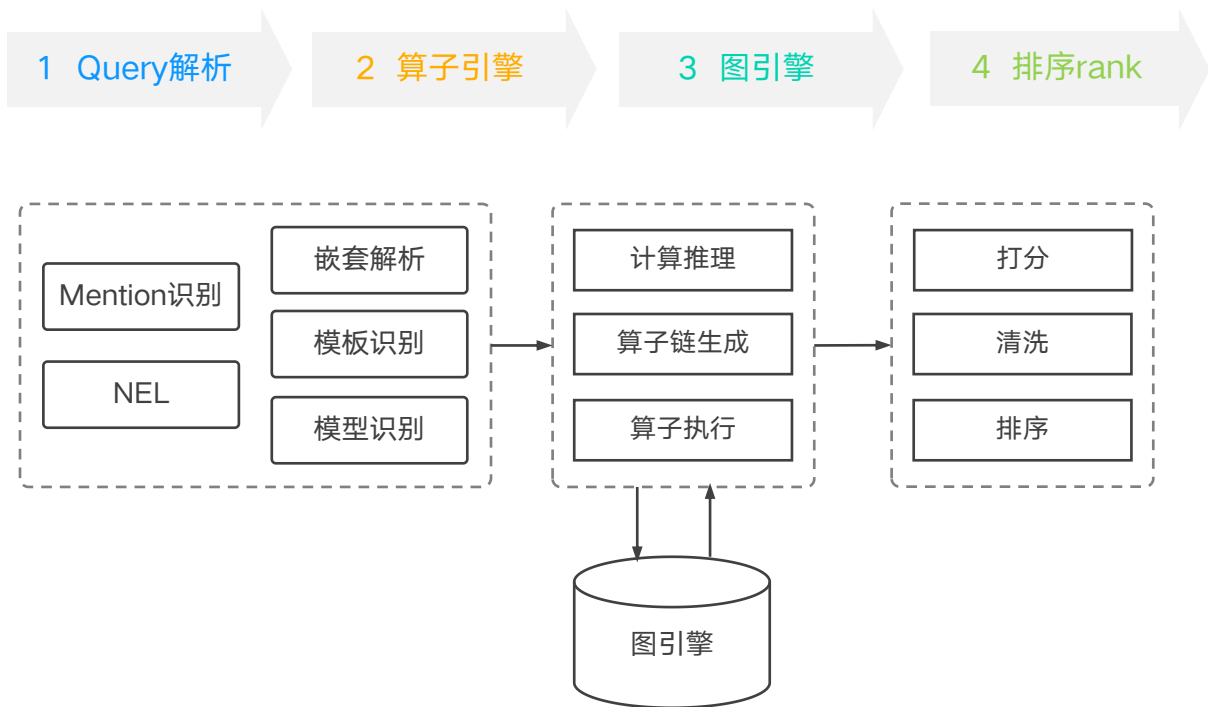


十亿级  
三元组

### 当前搜索场景的一些特点？

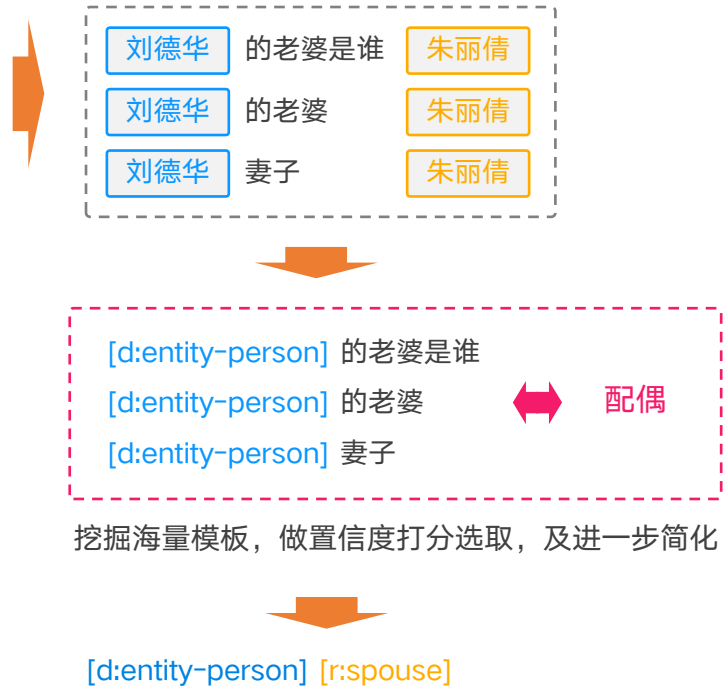
- ✓ 表达较短，同时有很多简单词堆砌
- ✓ 形式简单，多为简单SPO查询
- ✓ 长尾化严重，需要考虑长尾品类拓展成本
- ✓ 需要灵活多变，可解释性要求较强，遇到问题需要批量收敛

### 结构化推理方案



## Query解析：模板挖掘

基于种子SPO，从多方数据来源挖掘模板



## Query解析：层次化模板匹配

简单识别，与多层次嵌套模板识别

e.g. 特朗普大女儿

[d:entity-person] [r:first\_daughter]  
特朗普 大女儿

e.g. 刘德华张家辉合作作品

[d:entity-person] [d:entity-person] [r:cooperative\_works]  
刘德华 张家辉 合作作品

e.g. 1991年10月出生是什么星座

[d:entity-year] [d:entity-month] [x:birth] [r:constellation]  
1991年 10月 出生 是什么星座

e.g. 爸爸的姐姐的儿子怎么称呼

[w:\*] [r:kinship]  
↓ 怎么称呼

[w:\*] [r:son]  
↓ 的儿子

[d:entity-person] [r:elder\_sister]  
爸爸 的姐姐

e.g. 今天是农历几月几号

[w:\*] [x:is] [r:lunar\_day]  
↓ 是 农历几月几号

[d:entity-relative\_date]  
今天

## Query解析：模型预测

使用轻量且精准的模型兜住模板未能覆盖的10%的用户表达

e.g. 刘德华老婆是谁

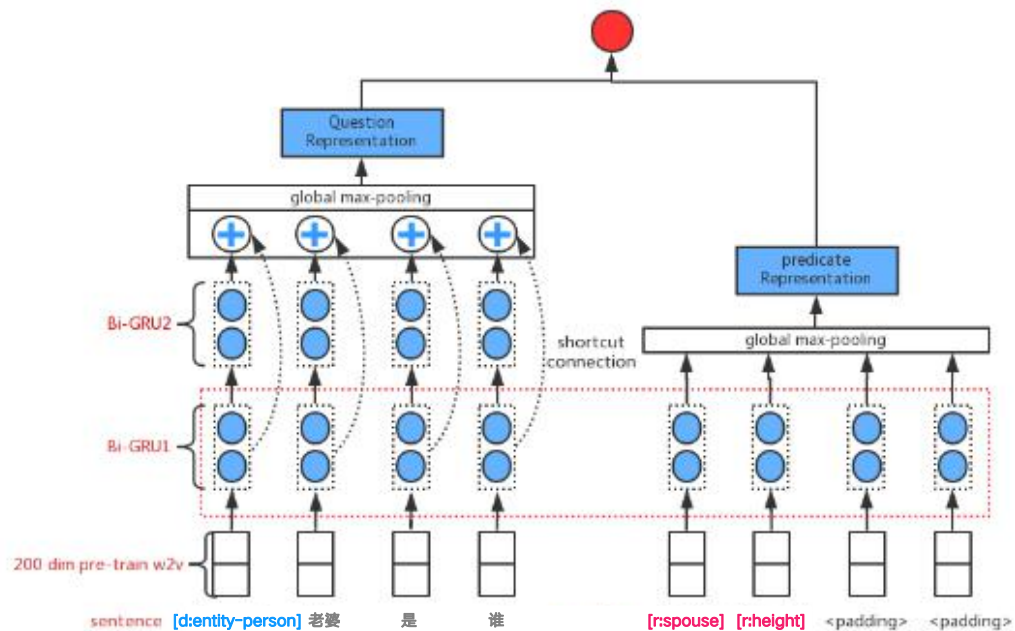
[d:entity-person] 老婆是谁



HRRNN-Match



[d:entity-person] [r:spouse]



## 结构化推理：算子引擎

算子的无限种组合，可以解决搜索场景下绝大多数复杂问题

**verify**  
验证正误

e.g. 张国立离婚了吗

**inter**  
取交集

e.g. 张学友梁朝伟合作作品

**union**  
取并集

e.g. 乔丹科比多高

**indexSort**  
排序取值

e.g. 特朗普二女儿叫什么

**compare**  
比较

e.g. 广州和深圳GDP对比

**count**  
计数

e.g. 广东有几个地级市

**calTime**  
时间换算

e.g. 去年是什么生肖

e.g. 田亮女儿多大了

[w:\*] [r:age]

↓ 多大了

[d:entity-person] [r:daughter]  
田亮 女儿

e.g. 爸爸的姐姐的儿子怎么称呼

[w:\*] [r:kinship]

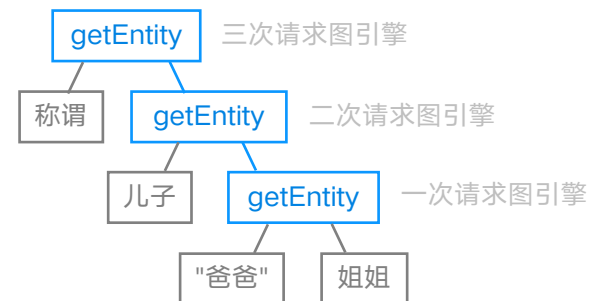
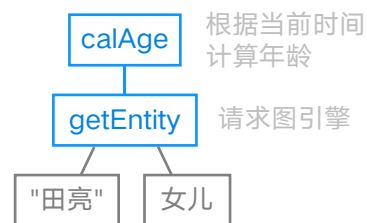
↓ 怎么称呼

[w:\*] [r:son]

↓ 的儿子

[d:entity-person] [r:elder\_sister]  
爸爸 的姐姐

\*自上而下，解析Query



\*自下而上，执行算子

# DeepQA

Machine Reading Comprehension

# DQA解决什么问题

- 背景

DeepQA是指深度挖掘自由文本内容，结合QP匹配和MRC等技术，从自由文本中抽取文本片段作为答案，从而产出高质量问答对，扩充线上FAQ库，提升问答整体覆盖率。
- 特点

✓ 来源简单

✓ 数据丰富

✓ 广泛适配多数场景
- 难点

如何从中抽取正确的答案？

Q: 白雪公主出讲的是什么故事

A: 讲述了白雪公主受到继母皇后（格林兄弟最初手稿中为生母）的虐待，逃到森林里，遇到七个小矮人的故事

Q: 白雪公主出自哪部童话？

A: 格林童话

白雪公主是广泛流行于欧洲的一个童话故事中的**人物**，其中最著名的故事版本见于德国1812年的《格林童话》。讲述了白雪公主受到继母皇后（格林兄弟最初手稿中为生母）的虐待，逃到森林里，遇到七个小矮人的故事。

历史学家巴特尔思据称白雪公主的历史原型是1725年生于德国西部美茵河畔洛尔城的玛利亚·索菲亚·冯·埃尔塔尔。说是白雪公主时，王后就伪装成巫。

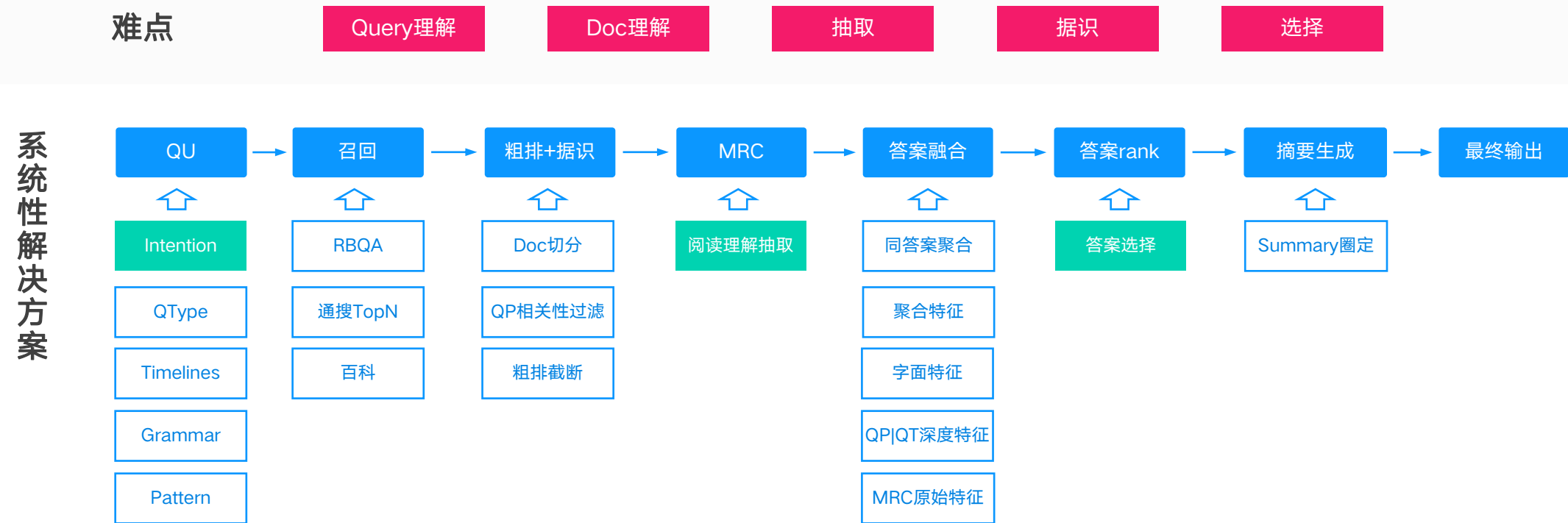
Q: 玫瑰花象征什么？

A:玫瑰花象征爱情和真挚纯洁的爱，爱情、和平、友谊、勇气和献身精神的化身

玫瑰花语在古希腊神话中，玫瑰集爱与美于一身，既是美神的化身，又溶进了爱神的血液。每到情人节，玫瑰更是身价倍增，是恋人、情侣之间的宠物。玫瑰代表爱情，不同颜色、朵数的玫瑰还另有吉意。

玫瑰花象征爱情和真挚纯洁的爱，爱情、和平、友谊、勇气和献身精神的化身。人们多把它作为爱情的信物。

# 搜索场景下的DQA





# 问答Query的理解

## 对Query的识别与拒识

### 无问题意图

赞美祖国  
快递单号查询  
表情包大全

### 语意不清

读大专要多少  
怎样与新同学  
他害怕吗？

### 缺乏主成分

为什么语音识别不出来  
怎么修改id  
今日头条怎样申请  
为什么被限制登录

### 计算

### 寻址

...

据识掉不合规Q

## Query的时效

### 强时效

哪只牛股可以买  
苹果公司财报营收  
iphone12上市

### 弱时效

黄晓明主演的什么电影  
五一放几天

### 无时效

中国最长河流  
异性相吸同性相斥原理  
iphone6s屏幕多大

事件、实体时效相关问题，  
统一由图谱结构化来统一解决

强时效过滤  
弱时效排序侧加重时间因子权重

## Query的类型

Who 人物/人名

Where 地点/位置信息

Numeric 数字/计量

When 时间

How 步骤/方法

Why 原因/起因

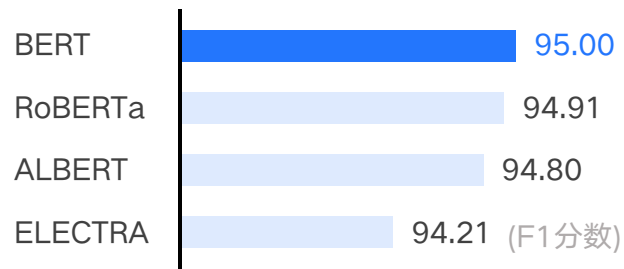
Definition 名词/定义/解释

Desc 观点/作用/特点/描述

通过type限定长、短、列表答案

## Query意图识别模型

### 优化一: BERT-Series Model



### 优化二: IDPT+ITPT

在监督学习之前, 使用领域数据/训练数据进行无监督预训练, 使模型适应领域知识 (+0.3~1.0)

PT	100.0	54.1	34.5	27.3	19.2
News	54.1	100.0	40.0	24.9	17.3
Reviews	34.5	40.0	100.0	18.3	12.7
BioMed	27.3	24.9	18.3	100.0	21.4
CS	19.2	17.3	12.7	21.4	100.0
	PT	News	Reviews	BioMed	CS

交叉词汇分布 (新闻, 影评, 生物医疗等)

### 优化三: Standard Tricks

- 特征矫正
  - Drop Pooler (+0.1~0.3)
  - Weight Reinitialization (-0.1)
- 规避灾难性遗忘
  - Layer-wise LR Decay (+0.4~0.9)
  - Pre-trained Weight Decay (+0.2)
- Attention 细粒度优化
  - 禁用 Attention Dropout (+0.1~0.4)
  - 关闭 CLS 自注意力通道 (+0.1~0.3)
- 优化
  - 随机扰动 (+0.2)
  - 对抗式训练 (+0.3~0.8) (见下一节)

### 优化四: 边界样本增强

$$Uncertainty = \frac{\sum_{i=1}^N p_s(i) \log p_s(i)}{\log \frac{1}{N}}$$

简单却行之有效的策略, 在与下游分布一致的海量无监督语料上进行推理

- 从 Uncertainty 最高的 1% 样本中抽取 2,000+ 加入到训练集中 (+0.5)
- 从 Uncertainty 最低的 20% 样本中抽检 1,000+ 错误样本进行标注, 加入到训练集中 (+0.3)

## Query意图识别

### 优化五：对抗式训练

通过在训练样本上添加正梯度方向的扰动，使模型参数更为平滑，增强模型的鲁棒性和泛化能力。当模型上下游分布差异较大及模型极易过拟合时，对抗式训练的收益达到最大

$$\min_{\theta} \mathbb{E}_{(\mathbf{Z}, y) \sim \mathcal{D}} \left[ \max_{\|\delta\| \leq \epsilon} L(f_{\theta}(\mathbf{X} + \delta), y) \right]$$

#### FGM ICLR2017 (+0.3)

沿着梯度的正方向添加扰动，取梯度的平均值更新参数

$$\mathbf{r}_{adv} = \epsilon \mathbf{g} / \|\mathbf{g}\|_2$$

#### PGD ICLR2018 (+0.2)

在 FGM 的基础上多次迭代，找到更为合理的扰动点

$$\mathbf{r}_{adv|t+1} = \alpha \mathbf{g}_t / \|\mathbf{g}_t\|_2$$

#### FreeAT ICLR2019 (+0.1)

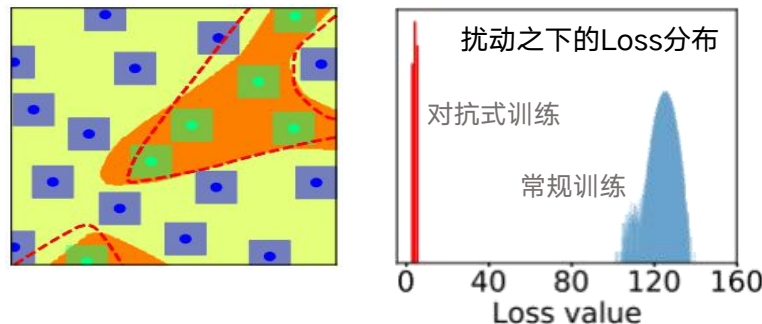
迭代计算扰动的同时更新参数，加速对抗式训练进程

$$\mathbf{r}_{t+1} = \mathbf{r}_t + \epsilon \cdot \text{sign}(\mathbf{g})$$

#### FreeLB NIPS2019 (+0.4~0.6)

采用均匀分布初始化扰动，并取滑动平均梯度更新参数

$$\delta_{t+1} = \Pi_{\|\delta\|_F \leq \epsilon} (\delta_t + \alpha g(\delta_t) / \|g(\delta_t)\|_F)$$



#### SMART ACL2020 (+0.5~0.8)

采用 Symmetrized KL-Divergence 对添加扰动前后的输出分布进行约束，提高模型抵抗扰动的能力；同时，通过同样的方法约束每轮迭代前后的梯度更新，避免 aggressive update

$$\mathcal{R}_s(\theta) = \frac{1}{n} \sum_{i=1}^n \max_{\|\tilde{x}_i - x_i\|_p \leq \epsilon} \ell_s(f(\tilde{x}_i; \theta), f(x_i; \theta))$$

Smoothness-inducing Adversarial Regularization

$$\mathcal{D}_{\text{Breg}}(\theta, \theta_t) = \frac{1}{n} \sum_{i=1}^n \ell_s(f(x_i; \theta), f(x_i; \theta_t))$$

+ Bregman Proximal Point Optimization

BERT	95.00
RoBERTa-large	94.91
ALBERT-xxlarge	94.80
ELECTRA	94.21
BERT-IDPT	95.96
BERT-ITPT	96.41
BERT-ITPT-D	96.58
BERT-ITPT-Lr	97.12
BERT-ITPT-A+	96.72
BERT-ITPT-A++	96.68
BERT-ITPT-Pr	96.80
BERT-ITPT-StandardTricks	97.23
BERT-ITPT-StandardTricks	97.90
BERT-ITPT-StandardTricks-FGM	97.95
(在这之后经历了测评集调整)	
BERT-ITPT-StandardTricks-FGM	95.25
BERT-ITPT-StandardTricks-PGD	95.31
BERT-ITPT-StandardTricks-FreeAT	95.11
BERT-ITPT-StandardTricks-FreeLB	95.44
BERT-ITPT-StandardTricks-SMART	95.45

## MRC模型优化

### QQ浏览器大规模搜索领域预训练模型-Motian

预训练模型：借助腾讯QQ浏览器搜索团队自研的预训练模型 Motian，针对问答MRC场景做finetune

### 额外信息引入

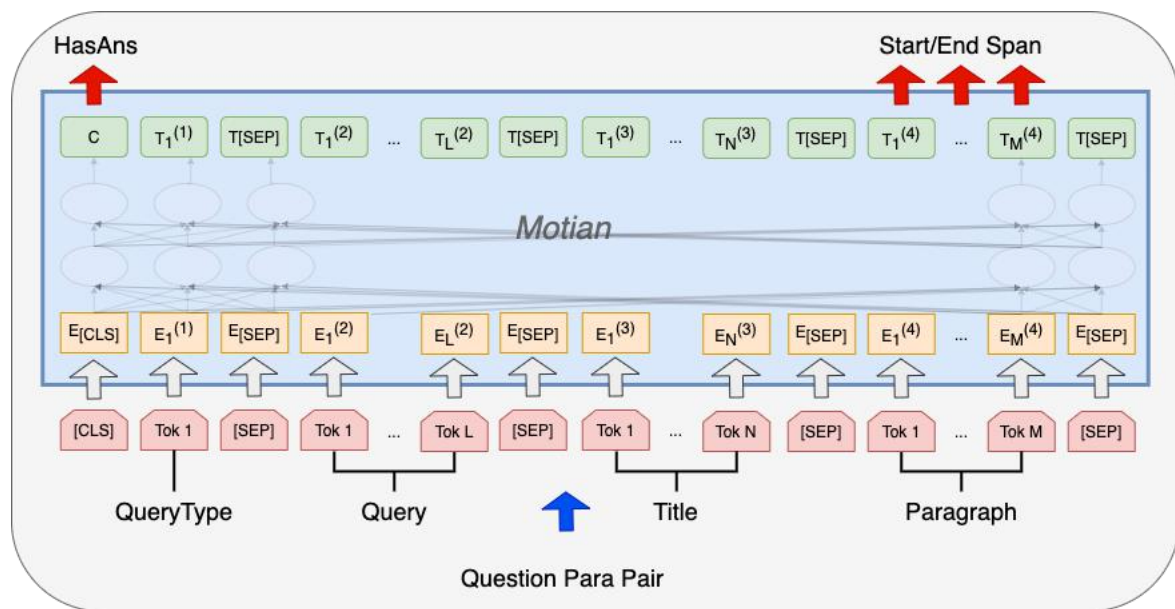
引入QType信息，避免模型跨类别抽取错误

短答案类

when where who  
numeric entity...

其他类

definition desc why  
how multi top...



## MRC数据强化

### 数据引入

**外部数据：**引入CMRC、LIC、WebQA等公开数据集，增强数据体量，做一阶段抽取范式学习

**精标数据：**在已有企鹅号、百科等基础精标MRC数据基础上，做二阶段微调，任务适配

### 数据增强

**正样本增强：**同类Entity替换正例增强

**正样例增强：**答案远窗口非支撑片段近似替换

**负样例增强：**答案短窗口支撑片段近似替换

**主动学习：**人工标记负样例+边界样例

Query: 八字方针是哪个会议提出的

Para: 董少鹏今年的中央经济工作会议在部署“推动高质量发展”时，强调要坚持巩固、增强、提升、畅通的方针，以创新驱动和改革开放为两个轮子，全面提高经济整体竞争力，加快现代化经济体系建设。“**巩固、增强、提升、畅通**”八字方针是2018年12月份**中央经济工作会议**提出的，2019年经济工作之所以在十分困难的情况下实现**稳中有进、稳中有变、稳中向好**，与深入落实八字方针息息相关。今年重申八字方针，体现了思路的连续性、政策...

↑  
支撑片段

↑  
答案片段

↑  
非支撑片段

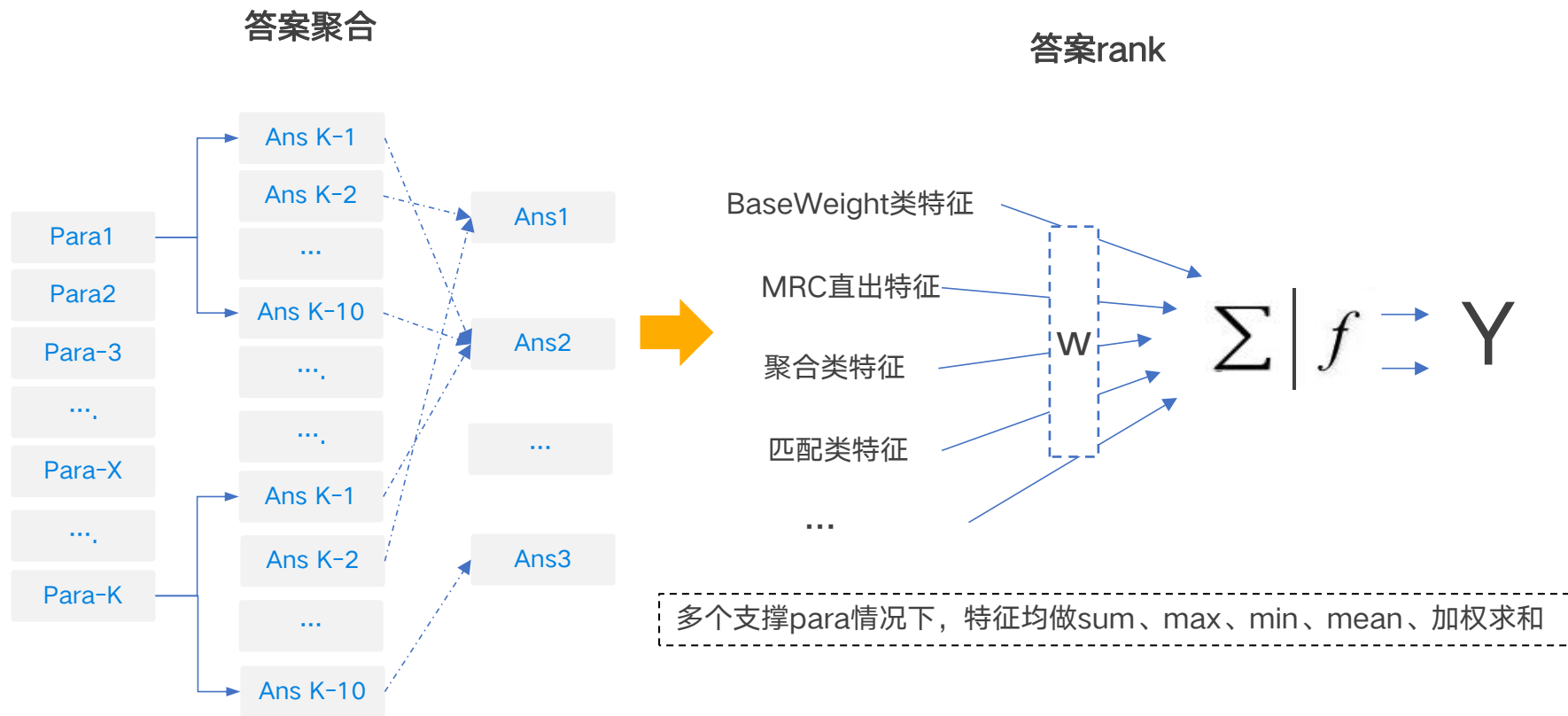
## 答案选择与排序

### 答案选择排序

- **答案聚合：**多来源多文档下，同答案聚合，提升信号强度
- **答案排序：**选取基础Q-T-P-A四元组基础字面匹配特征、MRC模型特征、语义类特征等，做多维信号融合打分排序

### 据识后置

- **MRC据识：**MRC模型面临大量负样本，过召回问题收敛难度大，过度收敛会导致召回变低
- **后置据识：**通过后置多维度融合后的分数，对答案进行进一步据识，降低过召回情况



# 方案的效果

Motian预训练 + MRC + rank 取得 CLUE-CMRC 榜单 Top1

CLUE总排行榜\*记最佳得分(10.22后中文原版数据集OCNLI替代CMNLI。bert\_base初始化分数，可重新跑OCNLI并提交)

排行	模型	研究机构	测评时间	Score	认证	AFQMC	TNEWS	IFLYTEK	CMNLI	OCNLI_50K	CLUEWSC 2020	CSL	CMRC2018	CHID	C3
1	HUMAN	CLUE	19-12-01	85.610	已认证	81.000	71.000	80.300	76.000	90.3	98.000	84.000	92.400	87.100	96.000
2	Motian	QQ浏览器搜索	21-06-25	84.055	待认证	78.29...	73.18	65.46...	85.4374...	84.967	94.8275...	90.16...	85.300	94.425	88.489
3	BERTSG	Sogou Search	21-06-25	83.824	待认证	79.84...	74.15	64.53...	85.2990...	85.933	95.1724...	89	83.800	93.059	87.436
4	Mengzi	TBH	21-06-21	83.052	待认证	79.51...	73.35...	64.38...	84.9566...	81.167	95.8620...	88.5	82.250	92.481	88.052
5	Pangu	华为云-循环智能	21-04-23	83.046	待认证	78.114	72.070	65.192	85.190	83.300	95.517	87.733	84.450	93.253	85.637
6	MT-BERTs	Meituan NLP	21-03-10	81.064	待认证	77.363	70.030	64.308	85.139	83.467	89.655	87.400	83.200	89.788	80.293
7	LICHEE	腾讯看点	21-01-08	80.508	待认证	76.975	70.500	64.154	84.541	81.300	90.690	87.400	79.800	87.505	82.220
8	roberta_selfrun	OPPO小布助手	21-02-05	80.238	待认证	77.881	69.370	63.923	82.939	80.400	93.103	87.267	80.100	90.107	77.287
9	BERTs	BERTs	20-12-24	80.220	待认证	76.768	69.940	63.923	84.483	82.900	88.966	86.767	80.500	89.513	78.443
10	UER-ensemble	TencentPretrain & TI...	20-11-28	80.087	待认证	76.819	72.200	64.000	84.090	80.800	90.345	85.833	79.150	86.031	81.603
11	Archer-24E-SINGLE	search-nlp	20-12-24	79.794	待认证	77.260	69.540	62.269	85.226	83.567	90.000	85.733	75.650	85.656	83.042
12	BI-ALBERT	It's me.	21-03-04	79.569	待认证	76.042	68.270	63.808	83.245	82.167	87.931	86.767	81.250	88.000	78.212

# IRQA

Information Retrieval QA



## 互联网中的问答数据

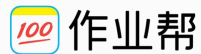
### 综合类问答内容

以UGC用户生产的各大社区



### 垂直类问答内容

以PGC专家生产的垂类网站



数据能否直接使用，存在哪些问题？

1. 准确性

2. 时效性

3. 安全性

4. 结构性

# IRQA需要做哪些事情

在线



离线



# 匹配：相关性计算模型

## 字面特征+XGB

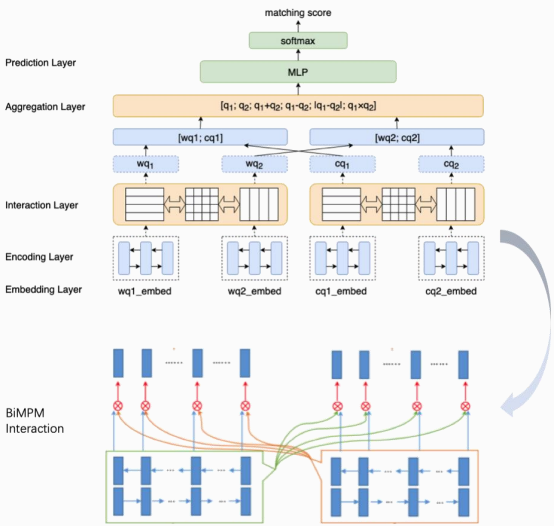
采用433 维特征，包括：  
 400 维预训练词向量  
 33 维其他手工特征

XGB- 文本结构化特征		
句子1切词长度	编辑距离	300维句子1向量
句子2切词长度	曼哈顿距离	300维句子1向量
句子切词差异长度	欧式距离	
句子切词差异长度比值	杰拉德距离	
句子1字符长度	兰氏距离	
句子2字符长度	Tf-idf	
句子字符差异长度	...	
句子字符差异长度比值		
共现词个数		
共现词比例		
去重后共现词个数		
共现字个数		
共现字比例		
去重后共现字个数		
去停用词后词共现		
...		

## 多粒度 Interaction Model

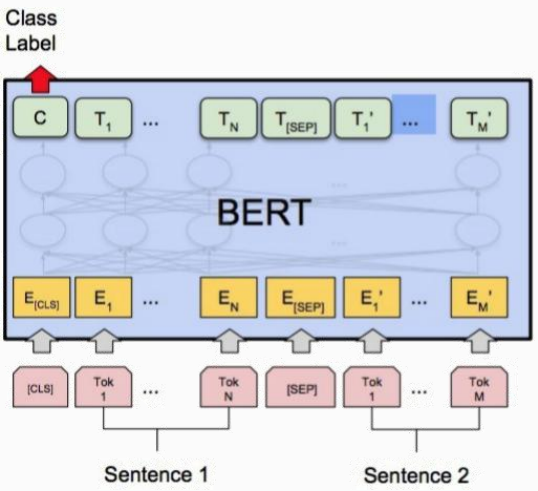
使用 W+C 双维度表示层解决 OOV  
 使用 BiLSTM 编码层进行 Encoder  
 使用 BIMPM 进行信息交互  
 使用 LocalAggregate 增强表示

### Interaction Model



## 预训练 BERT

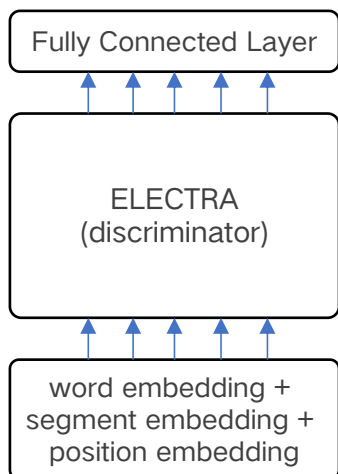
使用基础Pret-train BERT 进行双句拼接训练



## 预训练模型应用

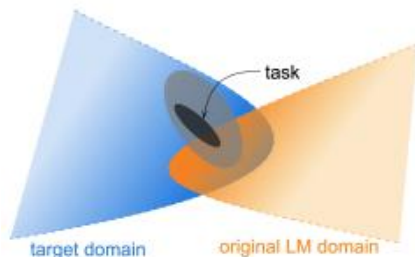
<sup>1</sup>ELECTRA + <sup>2</sup>领域知识预训练 + <sup>3</sup>Standard Tricks + <sup>4</sup>SMART 对抗式训练 + <sup>5</sup>数据增强 = 现役模型

1. ELECTRA 采用 MLM+RTD 联合训练，在各大自然语言处理榜单上取得最佳成绩



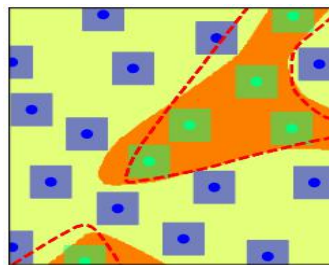
丢弃手工特征，使用纯文本

2. 基于海量离线问答对进行无监督预训练，从 token 维度最大程度地提取领域语义信息，融入语言模型中



3. 深度学习通用训练技巧，提高模型表现和稳定性  
e.g. 分层学习率

4. SOTA 对抗式训练算法，通过在 embedding space 添加扰动，提高模型的鲁棒性和泛化能力



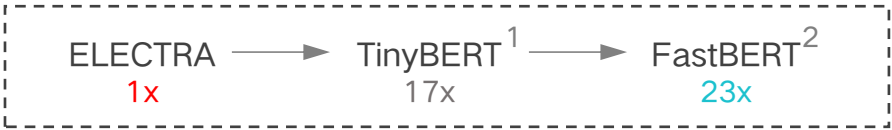
$$\mathcal{R}_s(\theta) = \frac{1}{n} \sum_{i=1}^n \max_{\|\tilde{x}_i - x_i\|_p \leq \epsilon} \ell_s(f(\tilde{x}_i; \theta), f(x_i; \theta))$$

(Smoothness-inducing Adversarial Regularizer)

$$\mathcal{D}_{\text{Breg}}(\theta, \theta_t) = \frac{1}{n} \sum_{i=1}^n \ell_s(f(x_i; \theta), f(x_i; \theta_t))$$

(Bregman Point Optimization)

# 大模型加速



1. 模型压缩，全方位减少参数规模

2. 动态推理，易区分的样本提前离开模型

$$\mathcal{L}_{\text{embd}} = \text{MSE}(\mathbf{E}^S \mathbf{W}_e, \mathbf{E}^T) \quad \mathcal{L}_{\text{attn}} = \frac{1}{h} \sum_{i=1}^h \text{MSE}(\mathbf{A}_i^S, \mathbf{A}_i^T)$$

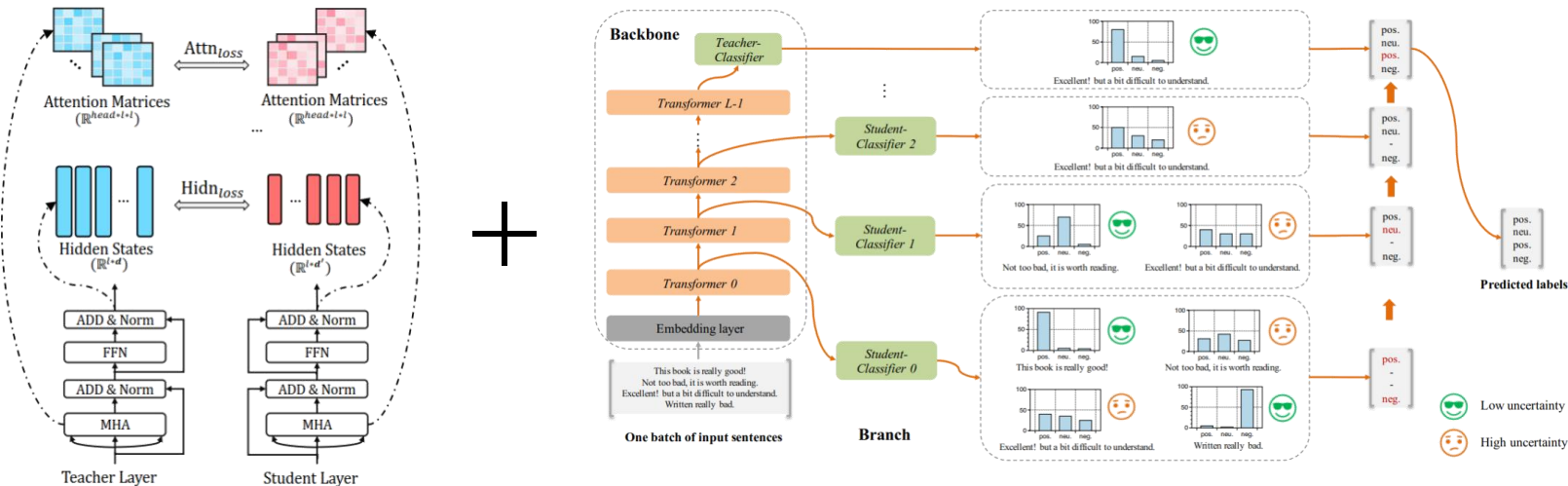
$$\mathcal{L}_{\text{hidn}} = \text{MSE}(\mathbf{H}^S \mathbf{W}_h, \mathbf{H}^T)$$

$$\mathcal{L}_{\text{pred}} = -\text{softmax}(\mathbf{z}^T) \cdot \log\text{-softmax}(\mathbf{z}^S / t)$$

$$\text{Uncertainty} = \frac{\sum_{i=1}^N p_s(i) \log p_s(i)}{\log \frac{1}{N}}$$

$$\text{Loss}(p_{s_0}, \dots, p_{s_{L-2}}, p_t) = \sum_{i=0}^{L-2} D_{KL}(p_{s_i}, p_t)$$

\*相对于论文，调整了子分类器架构，并实行了子分类器剪枝



## 一些思考

- ？ IRQA对内容生态的依赖、TOP1&SEO的互利结合
- ？ DQA事实的支撑、 KB完备性、KB的联合
- ？ 更全面的应用

# 欢迎加入

## QQ浏览器搜索中心简介

以QQ浏览器为主要阵地，打造基于腾讯特色内容生态的搜索引擎，并在技术沉淀的基础上联合腾讯视频等共建搜索技术中台，打磨高并发、高可用、高时效的百亿级检索系统，为PCG各个业务的搜索场景提供一套基础的搜索引擎服务。当前服务已经支持QQ浏览器、QQ看点、快报、腾讯视频、腾讯新闻、微视等PCG主要平台。用户在各个平台内可享受到文章、图片、小说、短小视频、长视频等众多内容的搜索服务；从NLP、深度神经网络、语义向量计算、用户行为反馈、相关性计算模型、百亿级规模数据挖掘、用户行为交互，用户需求分析等搜索技术，搜索内容和服务生态、富媒体内容理解、知识图谱、搜索产品、搜索推荐、用户行为研究等方向，吸引了大量业界一流的人才。基于年轻用户群喜好，共同打造一款年轻人爱用、爱玩的搜索引擎产品，并高效赋能商业增长，高效吸引用户，使之成为新一代搜索引擎的代表。



THANKS