



2018 JD Dialog Challenge

京东多轮对话系统挑战赛

团 队: feel free

成 员: 曾斌 (队长)、高峰、周丽芳

演讲人: 周丽芳



目录



- 1、赛题介绍
- 2、数据及预处理
- 3、初赛解决方案
- 4、决赛解决方案
- 5、可落地性分析
- 6、参赛经验总结
- 7、对话系统的发展趋势



1 - 赛题介绍



- 初赛：根据 $Q_1A_1Q_2A_2Q_3$ 信息预测对应的 A_3
- 决赛：在初赛的基础上，加入了多轮评测，即根据每一轮次的 Q_n 预测对应的 A_n

- 初，决赛客观评价
 - deltaBLEU

- 决赛主观评价
 - 人工评审：给定10个业务主题，客服使用给定的主题与系统进行对话，即模拟人机对话。
 - 人工评审的评价方法：对话满意度 * 50% + 任务完成率 * 30% + 任务完成效率 * 20%



2 - 数据及预处理



- 主数据：用户(Q)与客服(A)的对话历史记录
- 辅助数据：商品品类、订单信息等

bfea0048adc	USERID_10074093	0	0	0	可以吗?有什么条件
bfea0048adc	USERID_10074093	1	0	0	小妹正飞奔着为您查询,还请您稍等一下呢
bfea0048adc	USERID_10074093	1	0	0	亲,商品是什么问题呢
bfea0048adc	USERID_10074093	0	0	0	颜色不喜欢
bfea0048adc	USERID_10074093	0	0	0	想问下是否可以更换颜色
bfea0048adc	USERID_10074093	1	0	0	亲,不喜欢需要申请退货
bfea0048adc	USERID_10074093	0	0	0	申请了
bfea0048adc	USERID_10074093	1	0	0	退货后重新买其他颜色的喔
bfea0048adc	USERID_10074093	0	0	0	哦
bfea0048adc	USERID_10074093	1	0	0	#E-s[数字x]#E-s[数字x]
bfea0048adc	USERID_10074093	0	0	0	退货是吗
bfea0048adc	USERID_10074093	1	0	0	请问还有其他还可以帮到您的吗?
bfea0048adc	USERID_10074093	0	0	0	是无条件退货吗
bfea0048adc	USERID_10074093	1	0	0	是的呢
bfea0048adc	USERID_10074093	1	0	0	无理由退货

2 - 数据及预处理



- 按session拆分对话历史记录
- 合并同一角色连续说话内容（以逗号分隔）

用户(Q): 这个订单我想换货
用户(Q): 可以吗?有什么条件
客服(A): 小妹正飞奔着为您查询
客服(A): 亲, 商品是什么问题呢
用户(Q): 颜色不喜欢
用户(Q): 想问下是否可以更换颜色
客服(A): 亲, 不喜欢需要申请退货
用户(Q): 申请了
客服(A): 退货后重新买其他颜色的喔



用户(Q): 这个订单我想换货, 可以吗?
有什么条件
客服(A): 小妹正飞奔着为您查询, 亲,
商品是什么问题呢
用户(Q): 颜色不喜欢, 想问下是否可以
更换颜色
客服(A): 亲, 不喜欢需要申请退货
用户(Q): 申请了
客服(A): 退货后重新买其他颜色的喔

- 删除对话轮数小于3的session
- 中文分词, 停用词去除, 特殊字符处理

3 - 初赛解决方案



■ 初赛数据集构建：使用前三轮对话记录构造 $Q_1A_1Q_2A_2Q_3 + A_3$

用户(Q): 这个订单我想换货, 可以吗?
客服(A): 小妹正飞奔着为您查询, 还请您稍等一下呢, 亲, 商品是什么问题呢
用户(Q): 颜色不喜欢, 想问下是否可以更换颜色
客服(A): 亲, 不喜欢需要申请退货, 重新买其他颜色
用户(Q): 申请了
客服(A): 请问还有其他还可以帮到您的吗?
用户(Q): 没有了
客服(A): ~~#E-s[数字x]#E-s[数字x]~~
用户(Q): ~~是无条件退货吗~~
客服(A): ~~是的呢, 无理由退货~~



$Q_1A_1Q_2A_2Q_3$:

这个订单我想换货, 可以吗?<s>小妹正飞奔着为您查询, 还请您稍等一下呢, 亲, 商品是什么问题呢<s>颜色不喜欢, 想问下是否可以更换颜色<s>亲, 不喜欢需要申请退货, 重新买其他颜色<s>申请了

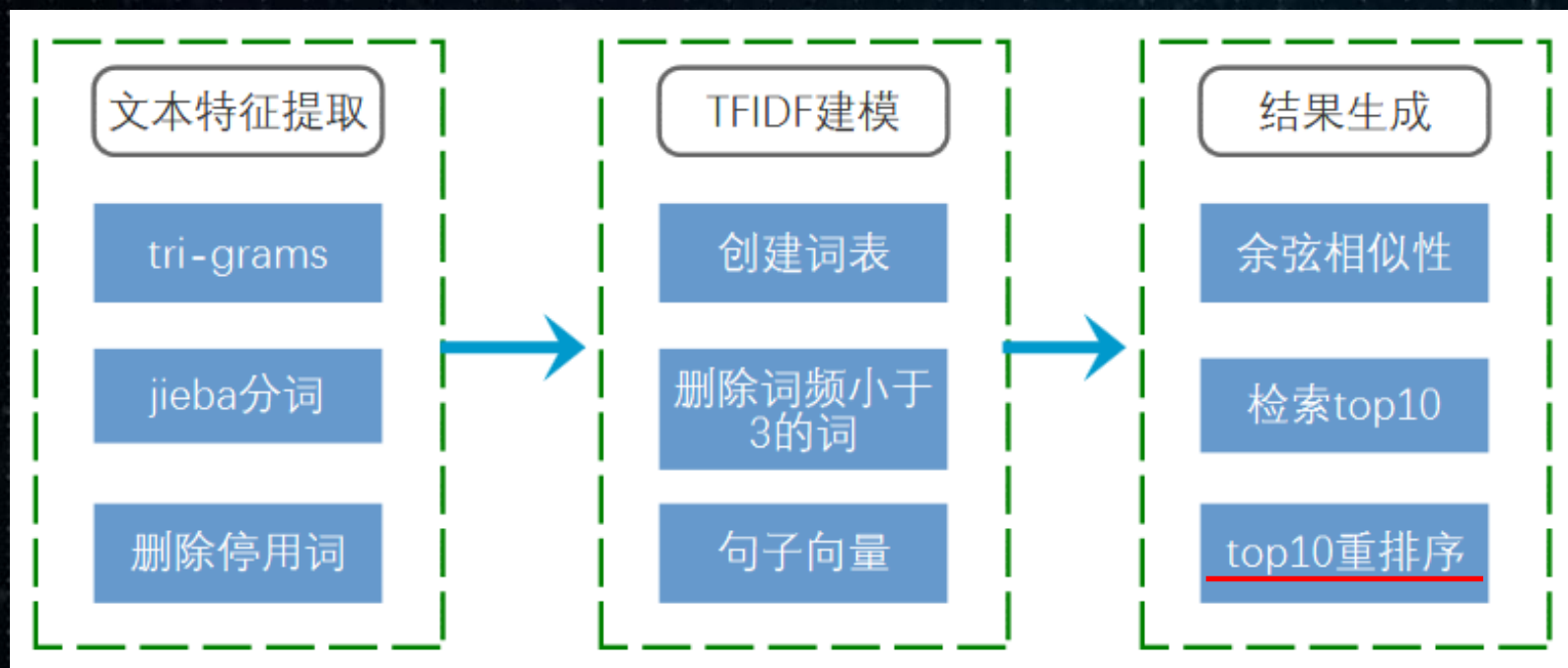


A_3 : 请问还有其他还可以帮到您的吗?

3 - 初赛解决方案



■ 基于TFIDF的检索式方案



● **技术创新:** 采用以answer长度为依据的top10重排序方案生成结果

● 评测得分0.25, 曾获得排行榜第2名, 最终排名为第17名

3 - 初赛解决方案



■ 为什么采用以answer长度为依据的top10重排序方案生成结果?

a) 相似性top10的结果召回率更高

b) 观察数据集可以发现，客服(A)的回复内容通常都要比用户(Q)说的内容更长，即客服(A)倾向于回复更长的内容

```
[('0', '这款电脑能用不\t写的是车载'),
 ('1', '小妹正在火速为您查询，还请您稍等一下呢，谢谢#E-s[数字x]\t 亲爱的 还辛苦您先稍等下的哈\t这边咨询下下 供应商的哈'),
 ('0', '好的'),
 ('1', '#E-s[数字x]#E-s[数字x]\t 亲爱的 感谢您的理解的哈\t请问还有什么需要为您效劳的呢?#E-s[数字x]'),
 ('0', '没事'),
 ('1', '[ORDERID_10187709]\t电脑 都是支持使用的哦\t 亲爱的这个是供应商的回复的哈'),
 ('0', '这个是车载的 在电脑上会影响读写速度不?'),
 ('1', ' 亲爱的 辛苦您稍等下的哈'),
 ('0', '好的'),
 ('1', '不影响的\t 亲爱的'),
 ('0', '好的 谢谢\t希望不要有划痕'),
 ('1',
 '亲爱的您太客气啦 也是没有帮上您什么忙的哈#E-s[数字x]这是我的分内之事的哈#E-s[数字x]有什么服务不周的还请您谅解的哈 我这边也是会积极改进的呢#E-s[数字x]\t遇到像您这样宽容的客户真是对我莫大的支持呢~#E-s[数字x]祝老板您天天日进斗金~腰缠万贯~#E-s[数字x]\t请问还有什么需要为您效劳的呢?#E-s[数字x]')]
```


4 - 决赛解决方案



■ 当前轮次Context引入方案优化

方案一：使用当前轮次之前用户(Q)和客服(A)对话的所有内容作为context



方案二：使用当前轮次之前用户(Q)说的所有内容作为context



方案三：仅使用用户(Q)上一轮次说的内容作为context

4 - 决赛解决方案



■ 决赛赛数据集构建: $Q_{n-1}Q_n + A_n$

用户(Q): 这个订单我想换货, 可以吗?
客服(A): 小妹正飞奔着为您查询, 还请您稍等一下呢, 亲, 商品是什么问题呢
用户(Q): 颜色不喜欢, 想问下是否可以更换颜色
客服(A): 亲, 不喜欢需要申请退货, 重新买其他颜色
用户(Q): 申请了
客服(A): 请问还有其他还可以帮到您的吗?
用户(Q): 没有了
客服(A): #E-s[数字x]#E-s[数字x]
用户(Q): 无条件退货吗
客服(A): 是的呢, 无理由退货

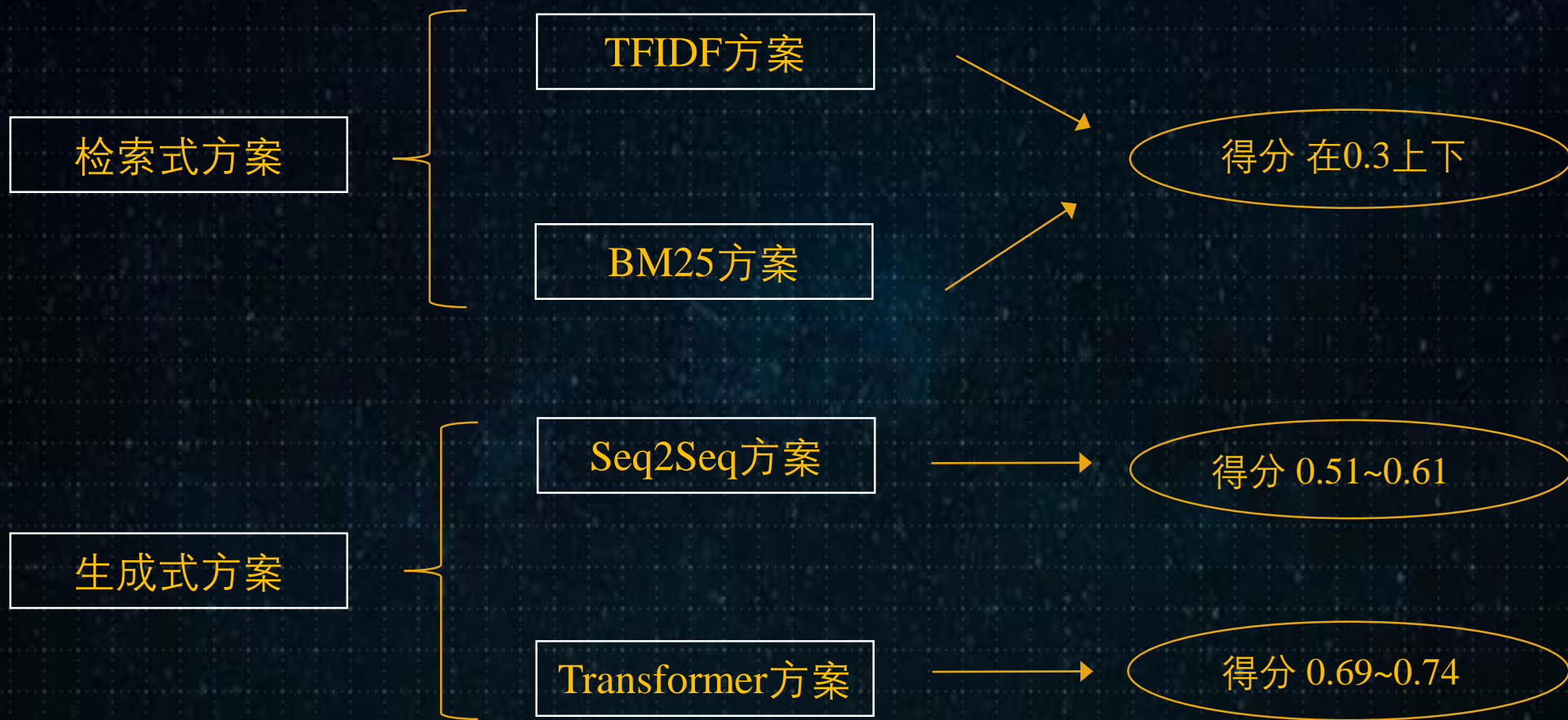


轮次	$Q_{n-1}Q_n$	A_n
1	这个订单我想换货, 可以吗?	小妹正飞奔着为您查询, 还请您稍等一下呢, 亲, 商品是什么问题呢
2	这个订单我想换货, 可以吗?<s>颜色不喜欢, 想问下是否可以更换颜色	亲, 不喜欢需要申请退货, 重新买其他颜色
3	颜色不喜欢, 想问下是否可以更换颜色<s>申请了	请问还有其他还可以帮到您的吗?
4	申请了<s>没有了	#E-s[数字x]#E-s[数字x]
5	没有了<s>无条件退货吗	是的呢, 无理由退货

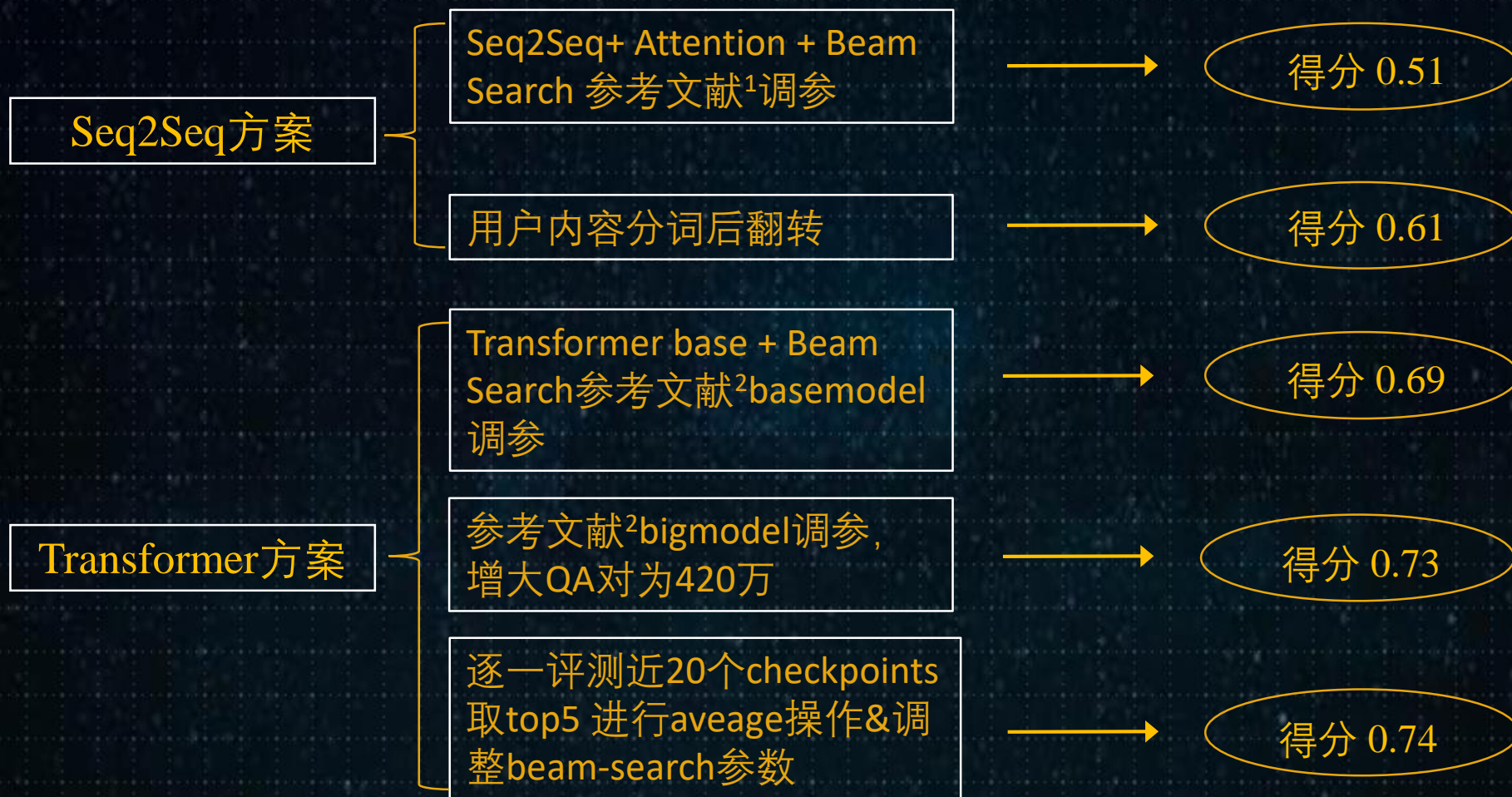
注: 仅保留A的长度范围在[3, 200]内的QQA Pairs

为了兼顾决赛中的单轮评测, 随机选取了10%的session仅保留 $Q_2Q_3 + A_3$ 数据对

4 - 决赛解决方案



4 - 决赛解决方案



[1] SUTSKEVER I, VINYALS O, LE Q V. Sequence to Sequence Learning with Neural Networks[J]. arXiv:1409.3215 [cs], 2014.

[2] VASWANI A, SHAZEER N, PARMAR N等. Attention Is All You Need[J]. arXiv:1706.03762 [cs], 2017.

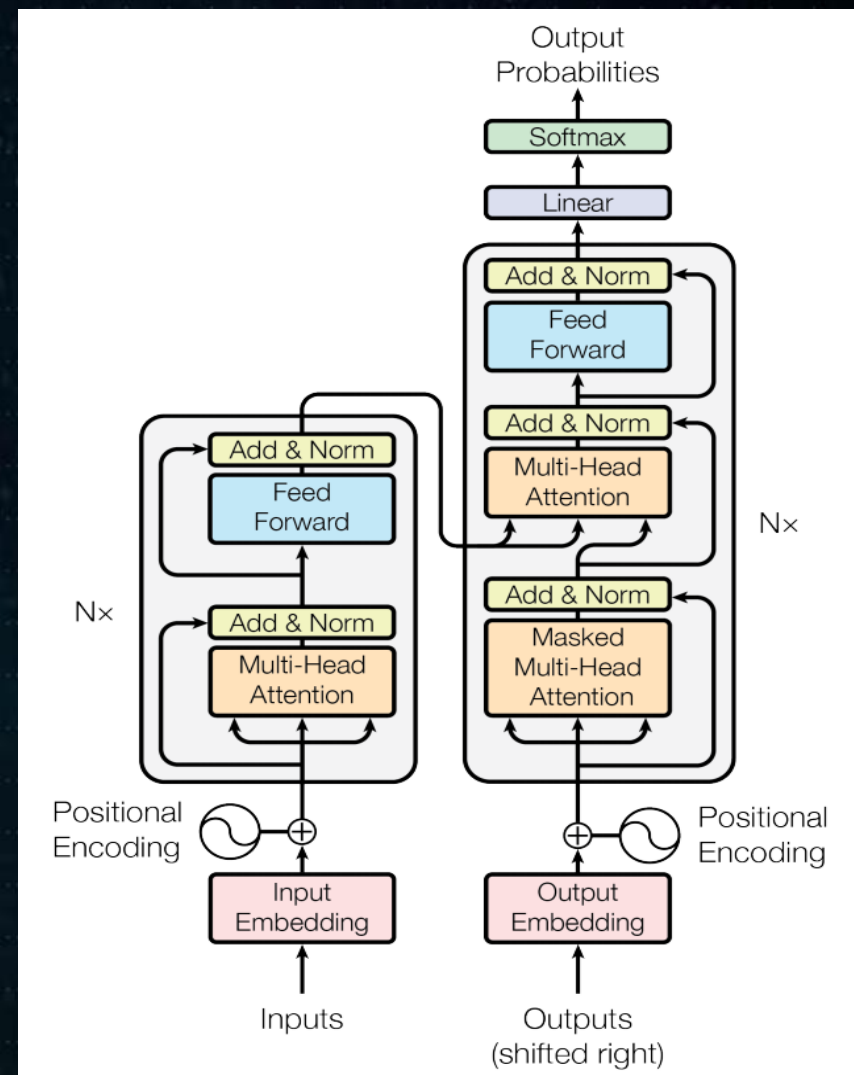
4 - 决赛解决方案



Transformer模型

Transformer	N	d_{model}	d_{ff}	h	d_k	d_v	P_{drop}	ϵ_{ls}
base	6	512	2048	8	64	64	0.1	0.1
big	6	1024	4096	16	64	64	0.3	0.1

- stack层数N不变
- 每一层的输出维度调整为1024
- Feed Forward Layer的隐层神经元数量调整为4096
- Multi-head Attention的h参数调整为16
- residual drop参数调整为0.3



4 - 决赛解决方案



■ 总结

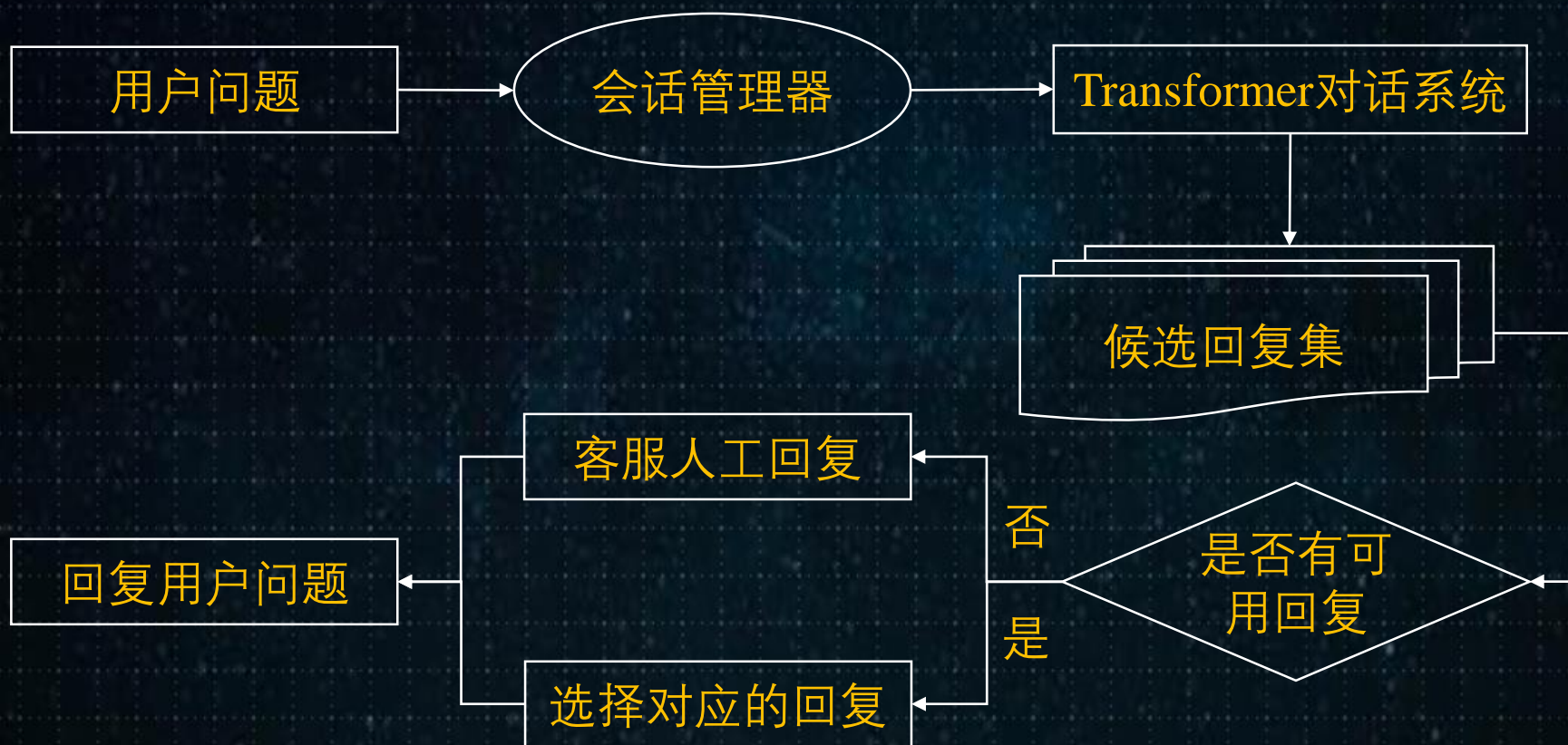
- 重视数据分析
- context 的引入方案优化
- 生成式方案的优化

序号	解决方案	测评得分
1	QQA:1M + BM25 + Top10 Rerank	0.307353
2	QQA:1M + Seq2Seq:4L + Attention + Beam Search:4	0.510164
3	QQA:2M + Seq2Seq:5L + Attention + Beam Search:4	0.613008
4	QQA:1M + Transformer base + Beam Search:4	0.69883
5	QQA:4.2M + Transformer big + Beam Search:4	0.73342
6	QQA:4.2M + Transformer big + Beam Search:6	0.744771

5 - 可落地性分析



■ 辅助人工客服，提升其工作效率



5 - 可落地性分析



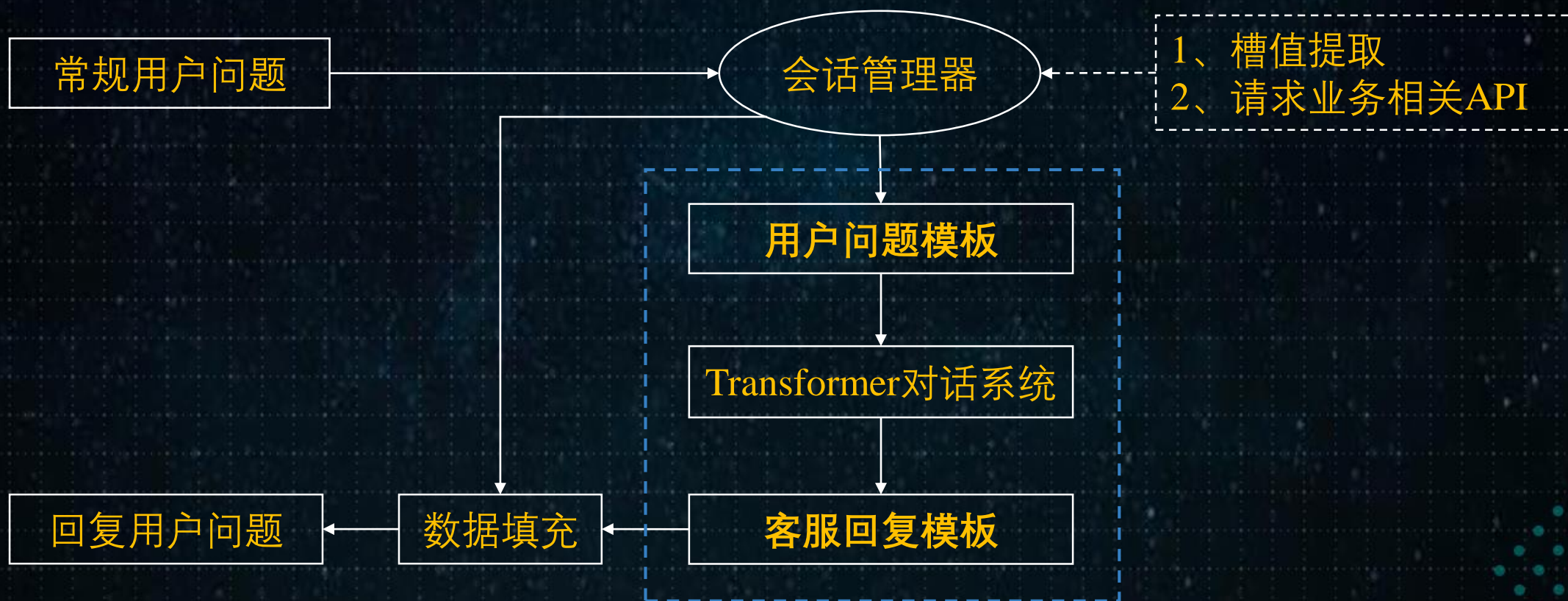
■ 结合Slot Filling技术，直接应用于常规用户问题回复

- 常规用户问题：用户询问一些常见业务主题的问题，如：开发票、查询物流信息、退货处理等；
- 前置一个业务主题分类器，非常规业务主题的相关问题转至人工客服；
- 使用Transformer对话系统完成“用户问题模板”至“客服回复模板”的翻译，模板是指 Slot Data 用 Slot Type 替代后的句子，如：
“我想查询<订单>的物流信息。” => “亲，您查询的订单物流信息如下：<订单物流信息>。”

5 - 可落地性分析



- 结合Slot Filling技术，直接应用于常规用户问题回复



5 - 可落地性分析



■ 其他

- 没有使用多个模型进行融合，利于快速响应
- 使用Tensorflow Serving部署模型，方便快捷

6 - 参赛经验总结



重视baseline及其优化，而不是盲目的尝试各种模型

引入多轮对话中的context信息应该尽可能简单，过于复杂往往意味着无效

调参是个无底洞，当性能达到一个还不错的状态时，期望通过调参来继续提升性能是不现实的

针对数据集的一些调整，也许会有意想不到的性能提升

选取几个效果最好的checkpoint进行average是提高模型性能的一个有效思路

7 - 对话系统发展趋势



技术发展

- 亟待稳定、可靠的生成式方案
- 与知识图谱结合，提升NLU性能

应用前景

- 作为语音操作形式的底层架构，前景广阔
- 任意功能与用户仅一句话之隔，足够便利



2018 JD Dialog Challenge

THANK YOU

