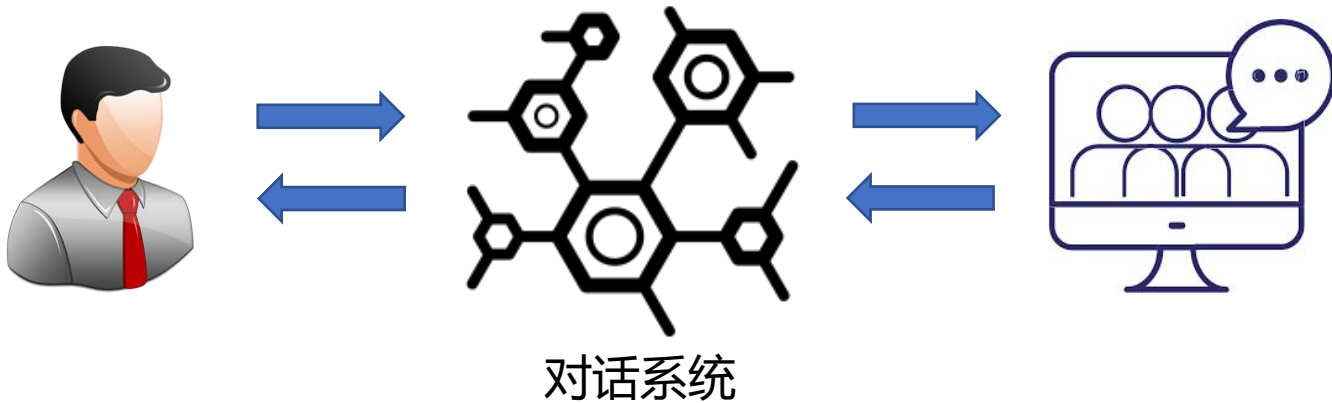
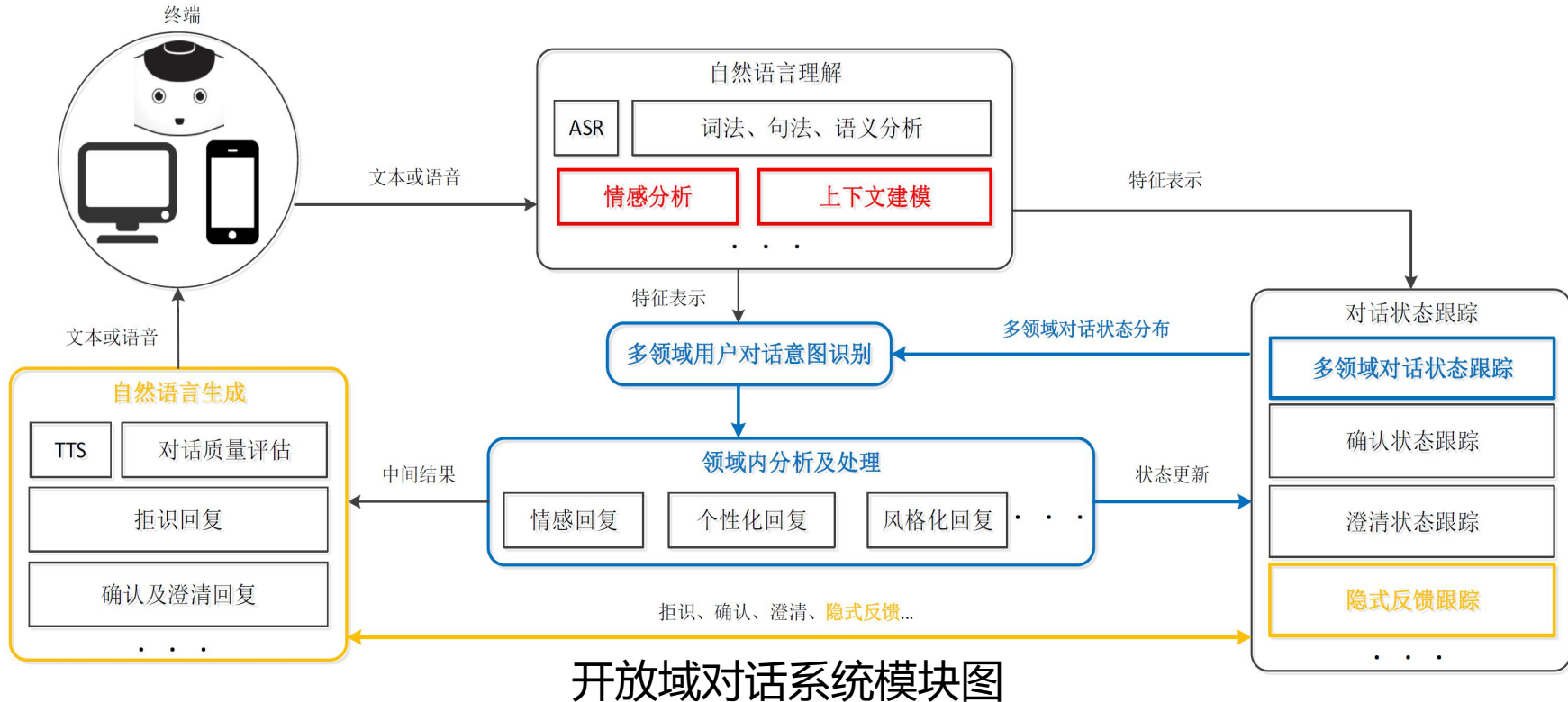


对话系统背景介绍



对话系统



图片摘自：张伟男《人机对话技术前沿及动态》，CCL2018

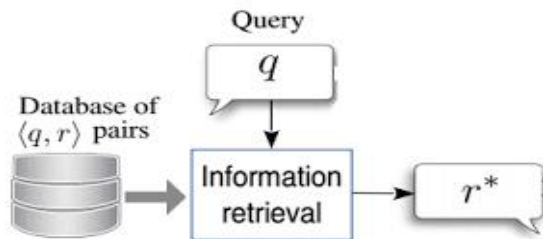
2020-10-13

中国科学院计算技术研究所

对话系统技术演化历程

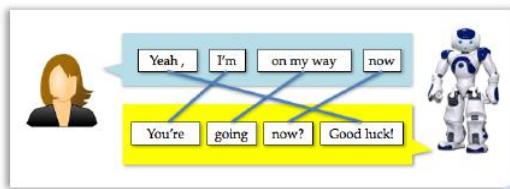
匹配

基于检索的方法

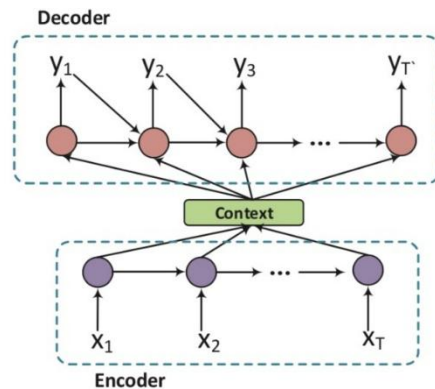


生成

基于翻译的技术

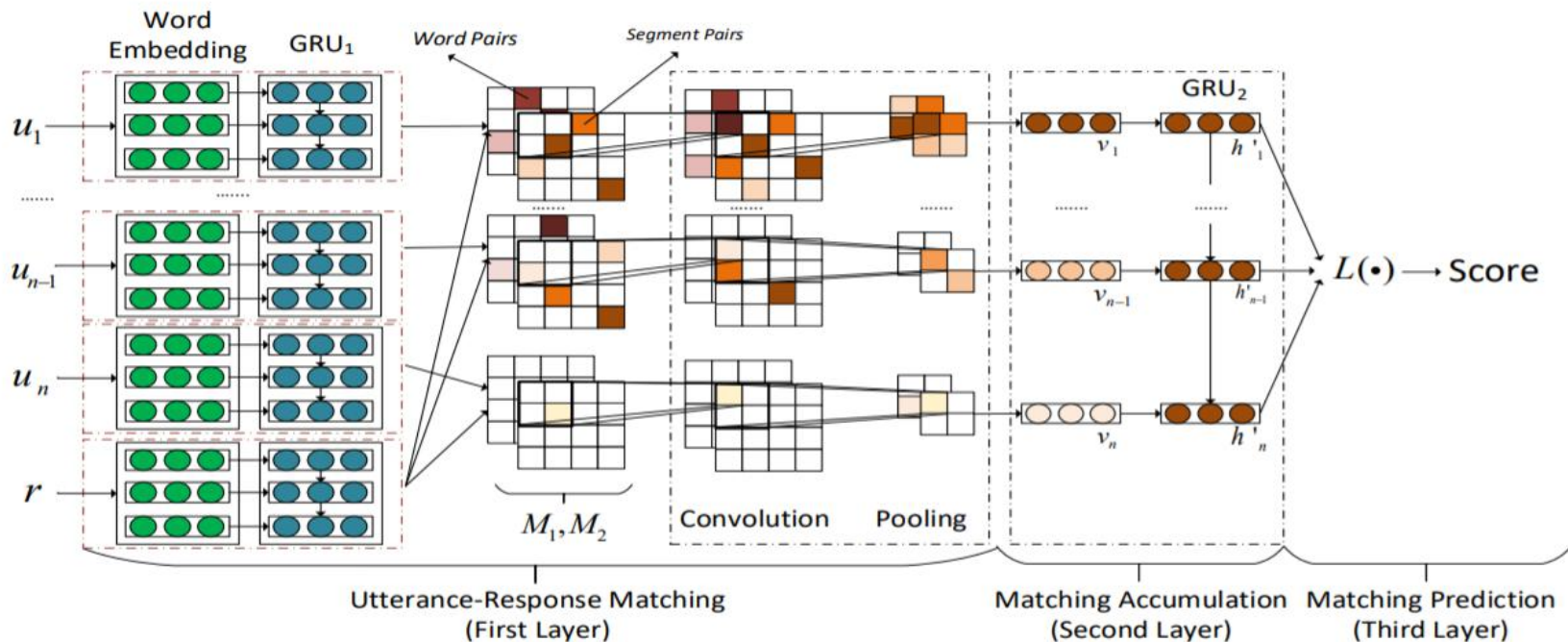


Seq2seq、GAN



检索式对话系统

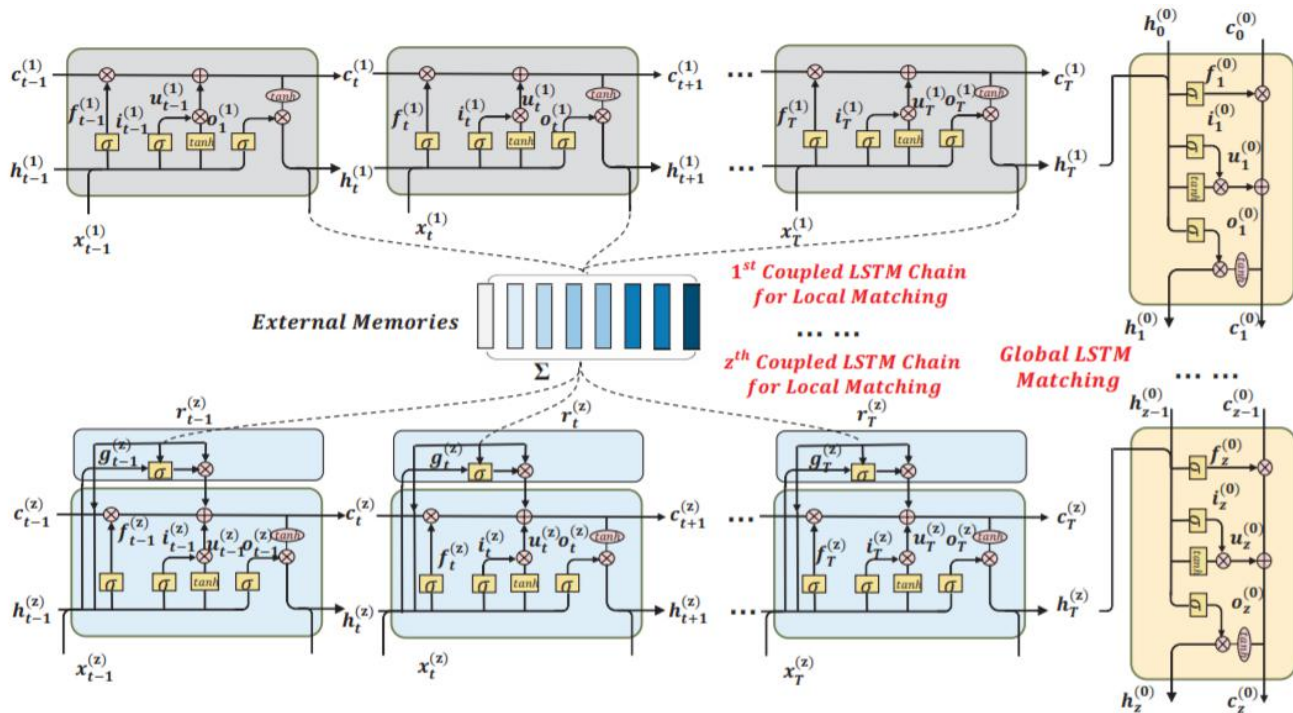
微软-多轮检索式对话



Y. Wu et al., Sequential Matching Network: A New Architecture for Multi-turn Response Selection in Retrieval-Based Chatbots, In ACL2017.

检索式对话系统

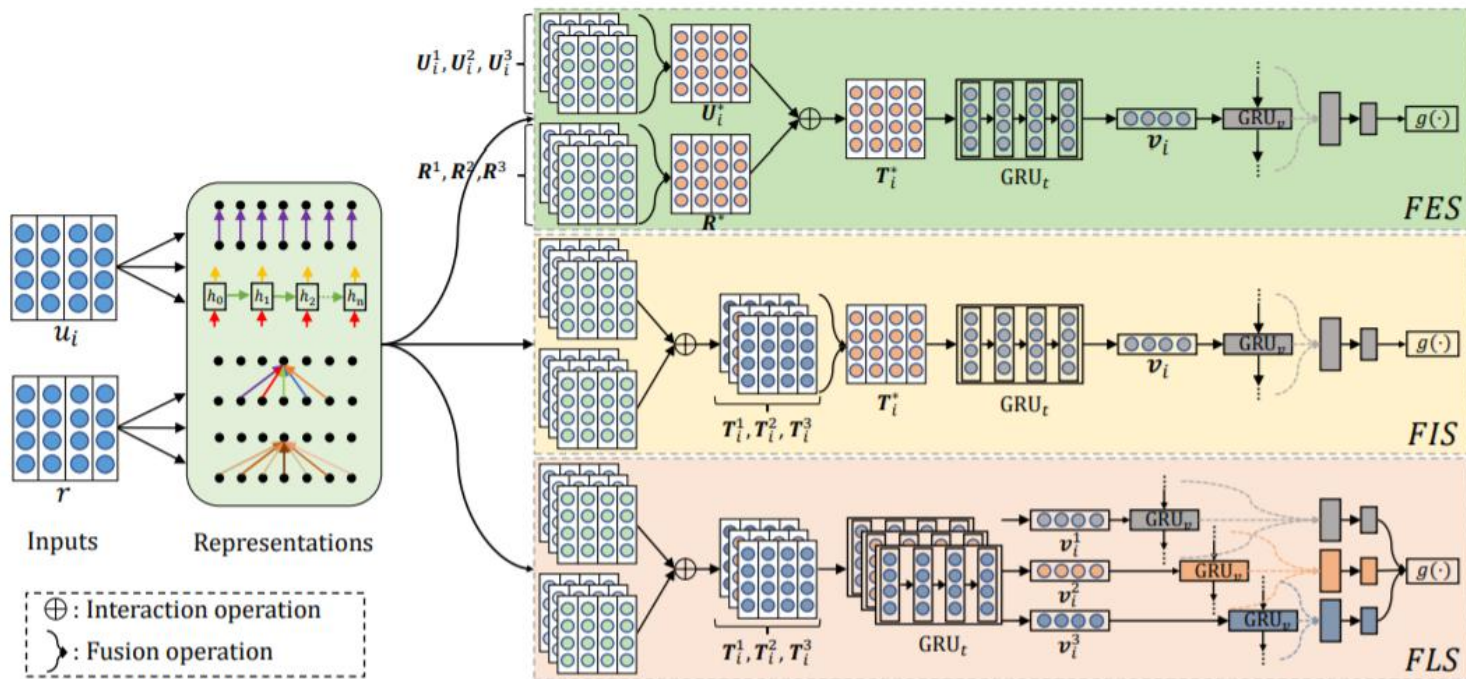
北大-局部+全局匹配的检索式对话



R. Yan et al. Coupled Context Modeling for Deep Chit-Chat: Towards Conversations between Human and Computer, In KDD2018.

检索式对话系统

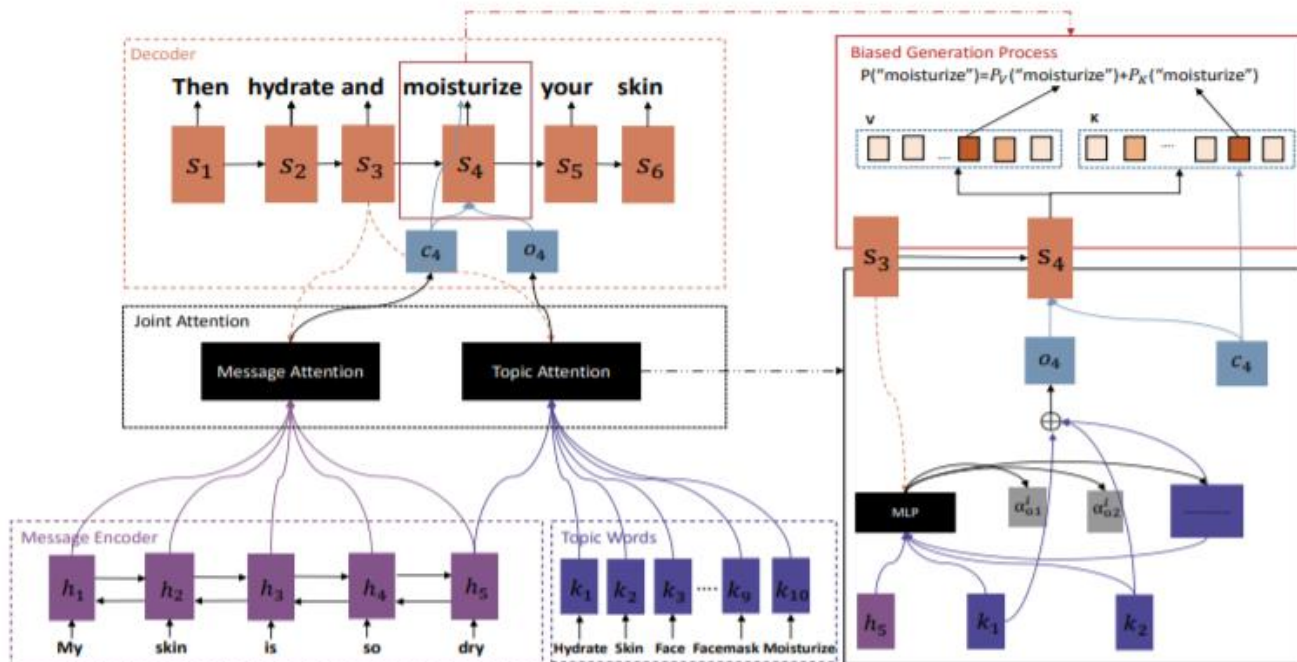
微软+北大-多轮检索式对话



C. Tao et al., Multi-Representation Fusion Network for Multi-turn Response Selection in Retrieval-based Chatbots, In WSDM2019

生成式对话系统

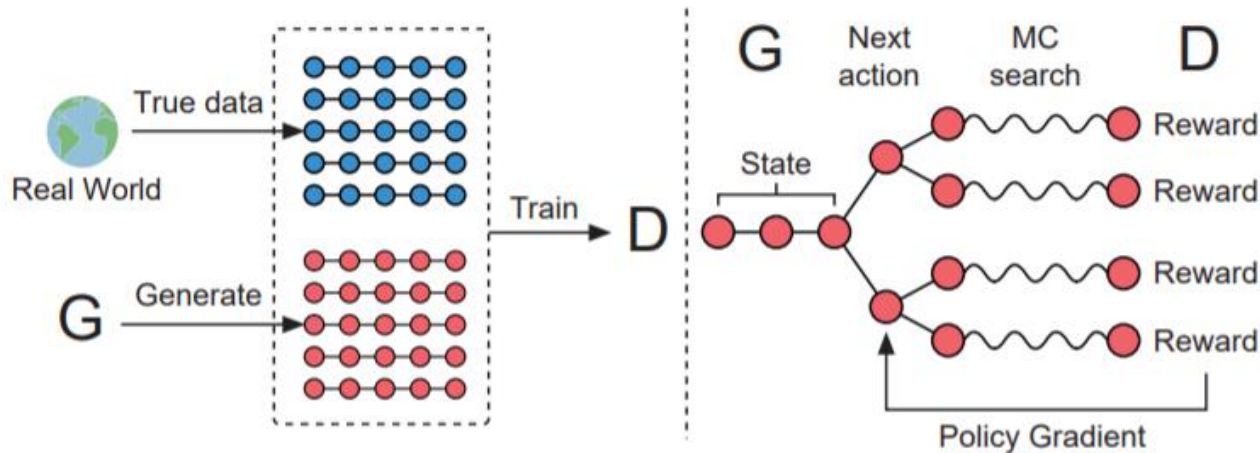
微软-话题敏感的生成式对话



Xing et al., Topic Aware Neural Response Generation. In AAAI2017

生成式对话系统

上交-基于GAN的语言生成



对抗训练+强化学习（reward是生成序列的质量得分）

对话系统多机制建模

- 人们在会话时，对**同一个问题**，可以有**各种各样不同的回答风格、侧重点**
- 对同一个问题，对话语料往往对应大量不同答案
 - 例如：“你好笨”有六十多种不同回答
 - 不笨、你才笨、为什么笨、别这样说...
- 这使得从机器翻译借鉴建模方法有失妥当
- 不同的回答会指向不同的对话方向和对话目的，从而影响后续的对话内容**
 - 对于客服场景，可能影响用户的**满意度**

Men: 这个保险如何支付保费？



亲，支付宝支持的



您可一定要使用支付宝支付！

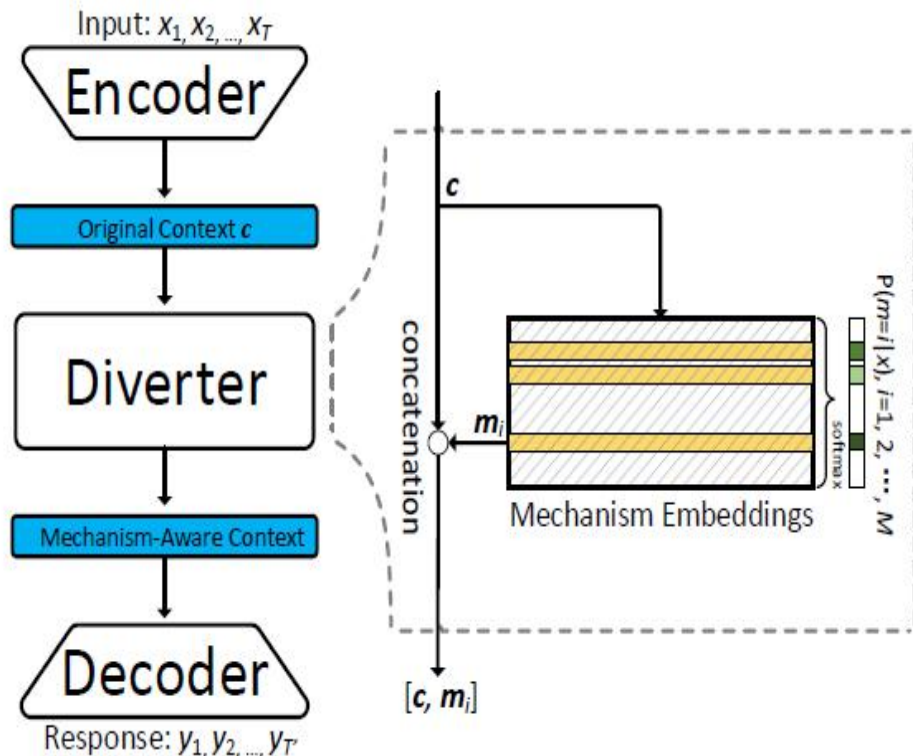


您想怎么付？支付宝有吗？



推荐支付宝，它比刷卡支付更安全

建模固定数量的机制



$$p(\mathbf{y}|\mathbf{x}) = \sum_{i=1}^M p(\mathbf{y}, m_i | \mathbf{x})$$

$$= \sum_{i=1}^M p(m_i | \mathbf{x}) \prod_{j=1}^{T'} p(y_j | m_i, \mathbf{x}, y_{<j})$$

学习mechanism的embedding:

$$p(m_i | \mathbf{x}) = \frac{\exp g(\mathbf{m}_i, \mathbf{c})}{\sum_{k=1}^M \exp g(\mathbf{m}_k, \mathbf{c})}.$$

使用maxout函数防止过拟合:

$$g(\mathbf{m}_i, \mathbf{c}) = \mathbf{m}_i^T \mathbf{W}_t \mathbf{t}_i,$$

$$\mathbf{t}_i = [\max\{\tilde{t}_{i,2j-1}, \tilde{t}_{i,2j}\}]_{j=1,2,\dots,l_c}^T,$$

$$\tilde{\mathbf{t}}_i = \mathbf{W}_c \mathbf{c}.$$

建模固定数量的机制

Questions	Mechanism 1	Mechanism 2	Mechanism 3	Mechanism 4
我要坐地铁回家 I take subway back home	好的，路上注意安全 Ok, be safe on the subway	真是个不错的选择 That's a good choice	坐车去哪里 Where are you going on the subway	坐车一定要注意安全 Be careful on the subway
生活中怎样感到开心快乐 How do I feel happy in my life	开心就来和我聊天 Have a talk with me if you are happy	看来你今天心情很不错 You seem to be happy today	有什么开心的事情 What are the happy things you have experieced	和我聊天就很开心 Talk with me and you will be happy
明天考试、求安慰 I have a test tomorrw, and I need some solace	好好复习就好了 Just work hard on the reviews	我会为你加油的 I will stand by you	你要去哪里试试 Where are you taking the test	我会一直安慰你的 I will always be on your side
上班好累啊 The work is really hard	那就好好休息 Take a good rest	要懂得劳逸结合 You need to balance work and life	你是不是累了 Are you tired	上班就是这样 That is work

效果

将多样性建模引入到Seq2seq神经网络框架中，回答的多样性增长了1.7倍，准确率提高了9.8%

建模固定数量的机制

Keyword in m_1		Keyword in m_2		Keyword in m_3		Keyword in m_4	
Chinese	English	Chinese	English	Chinese	English	Chinese	English
看	look	个	a	哪里	where	不会	cannot
好好	ok	看来	seem	?	?	一直	always
还是	still	不错	nice	呀	oh	可是	however
那	that	小	little	么	why	一定	must
注意	attention	还	yet	怎么	how	都	all
自己	self	这样	this	什么	what	陪	company
知道	know	人	person	在	stay	也	also
就	at once	可	may	样子	appearance	会	can
可以	can	微	tiny	想	think	认错	admit
去	go to	对	yes	说	say	很	very

效果

将多样性建模引入到Seq2seq神经网络框架中，**回答的多样性增长了1.7倍，准确率提高了9.8%**

自动确定机制数目



支持退款吗?

- 支持
- 不支持
- 得看情况

封闭问题，回答多样性较**低**



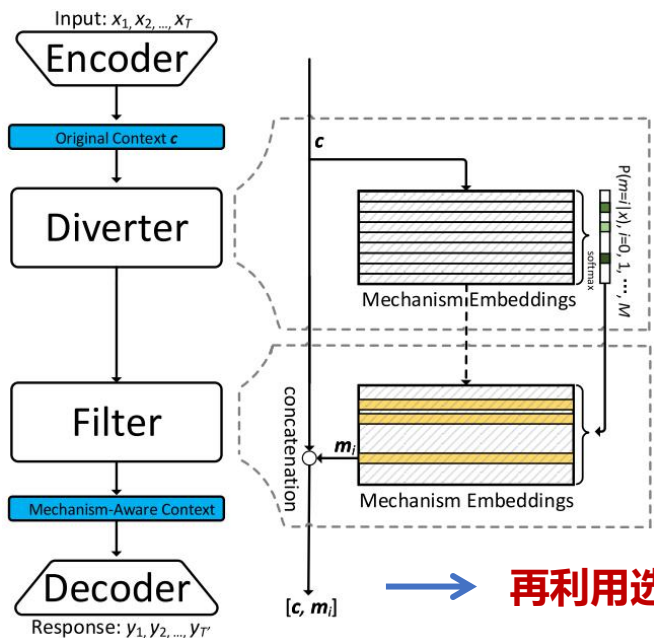
我账户登录不成功

- 亲，系统什么提示呢?
- 您别着急
- 您那边网络连接OK吗?
- 稍等，正在检查
- 用户名密码告我，后台查一下
- 凉凉

开放问题，回答多样性较**高**

自动确定机制数目

首先学习得到大量机制embedding，然后用强化学习自动选择合适的回答机制数量



将 $p(m_i|x)$ 降序排列所有的机制，记为 $m_{i_1}, m_{i_2}, \dots, m_{i_M}$ ，扫描这 M 个机制，并选出最好的 K 个，即 $m_{i_1}, m_{i_2}, \dots, m_{i_K}$ 用于生成答句。

Action: 继续扫描 a_c 和终止 a_t

State: $\mathcal{S}_x^{(k)} = \{m_{i_j}\}_{j=1}^k$

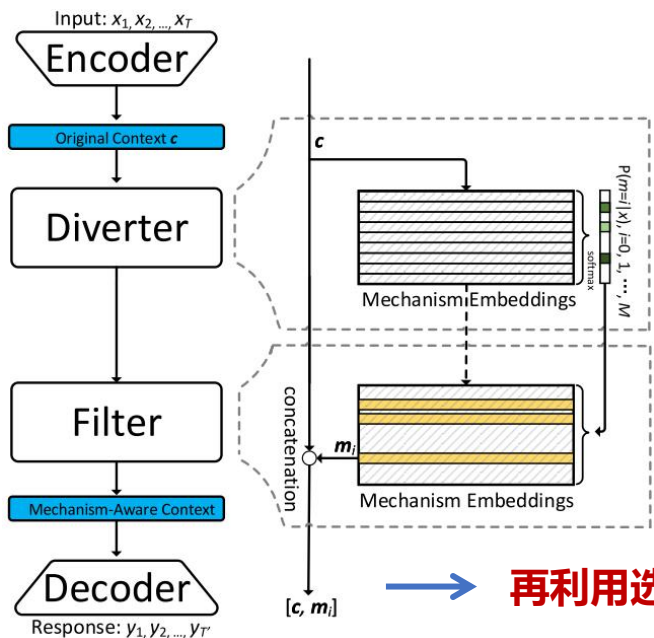
Reward: $\log p(y|x; \mathcal{S}_x^{(K)})$

强化学习：先选出用哪些机制回答更合适

再利用选出的机制集合生成答案

自动确定机制数目

首先学习得到大量机制embedding，然后用强化学习自动选择合适的回答机制数量



Policy: 随机策略 $\pi(\cdot | \mathcal{S}_x^{(k)})$

$$\pi(a_t | \mathcal{S}_x^{(k)}) = \frac{p(m_0 | \mathbf{x})}{1 - \sum_{m \in \mathcal{S}_x^{(k)}} p(m | \mathbf{x})} \quad \pi(a_c | \mathcal{S}_x^{(k)}) = 1 - \pi(a_t | \mathcal{S}_x^{(k)})$$

Action: 继续扫描 a_c 和终止 a_t

$$\text{State: } \mathcal{S}_x^{(k)} = \{m_{ij}\}_{j=1}^k$$

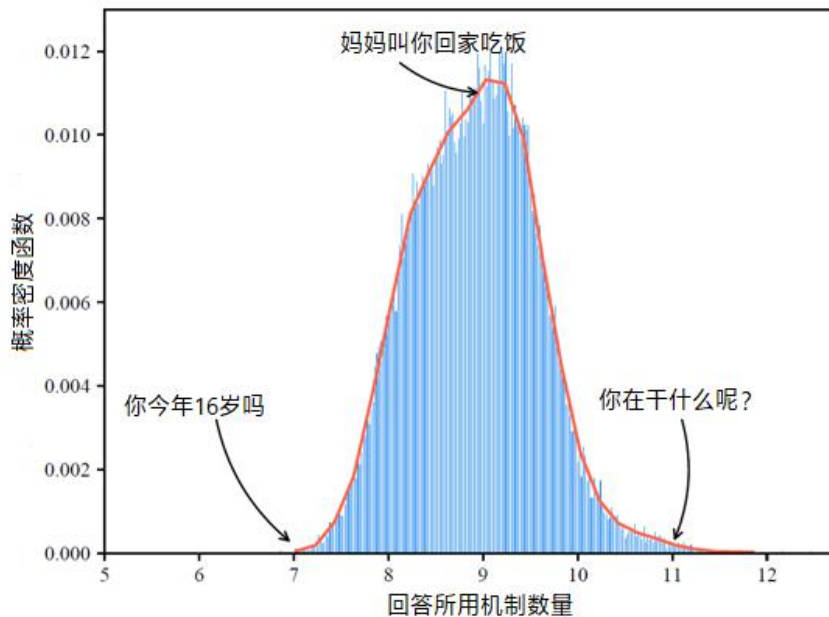
$$\text{Reward: } \log p(y | x; \mathcal{S}_x^{(K)})$$

强化学习：先选出用哪些机制回答更合适

再利用选出的机制集合生成答案

自动确定多样性数目

Post	我早上工作感觉死死的是怎么回事 What's wrong with me when I go to work this morning? I feel bad.	听说过外星生命吗 Do you hear alien life?
Mechanism 1	那你要去医院看看了 You shall go to the hospital to check it.	当然有啦 Of course.
Mechanism 2	那你要去医院看看 You shall go to hospital to check it.	我是机器人 I am a robot.
Mechanism 3	你可以去网上查查啊。 You could check it through internet.	主人，快从火星上回来吧 Master, please come back from the Mars soon.
Mechanism 4	你继续把话说完啊。 Go on, tell what you want to tell.	一直都是浮云 Nothing matters.
Mechanism 5	你要去医院么？ Do you go to the hospital?	你在说谁呀？ Who are you talking about?
Mechanism 6	你在干嘛呢呀 What are you doing?	当然没有 Of course not.
Mechanism 7	你去过什么事情？ What happen?	当然有 Of course.
Mechanism 8	夸奖 Praise	机器人都是浮云 Robot is nothing.
Mechanism 9	那你要去医院看看哦 You shall go to hospital to check it.	我可没去过 I have not been to there.



效果

较国际同类工作，提升答句生成的**多样性**，人工评测：F1值提高了7.41%