

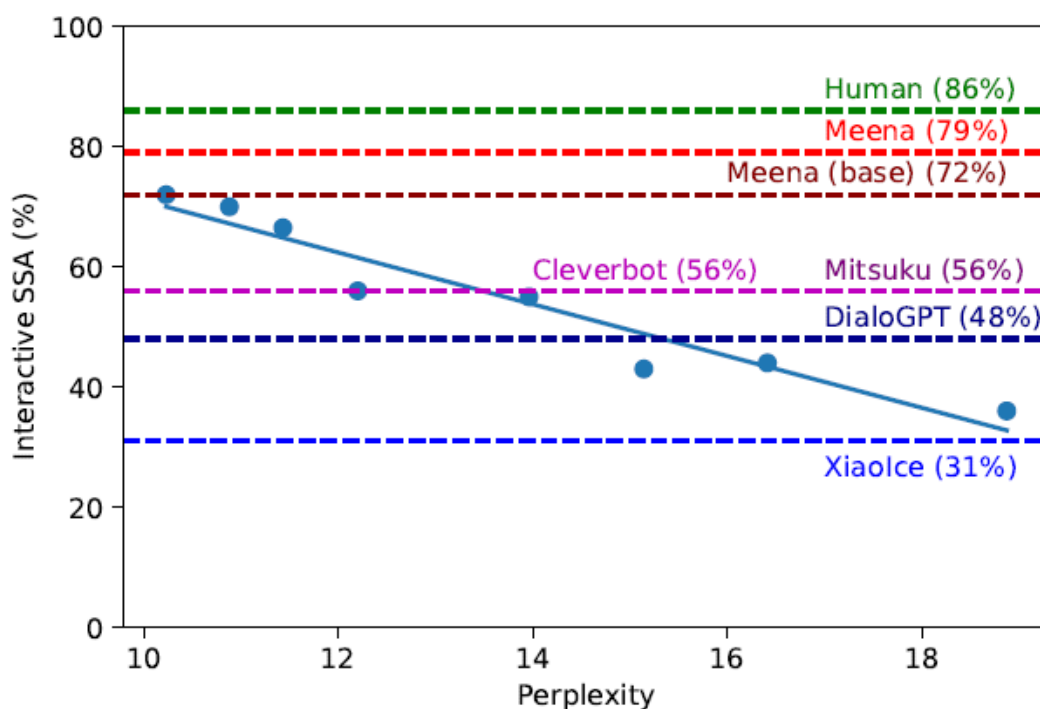
Towards a Human-like Open-Domain Chatbot

摘要：

1. 提出Meena，一个端到端训练的多轮开域聊天机器人，数据来源于公开的社交媒体对话
2. 提出人类评价标准SSA，更好的捕捉人类在多轮对话中的要素，SSA和目前通用的混淆度PPL有高相关性。
3. Meena在SSA上达到79%，人类水平86%，比其他存在的chatbots在SSA上高23%

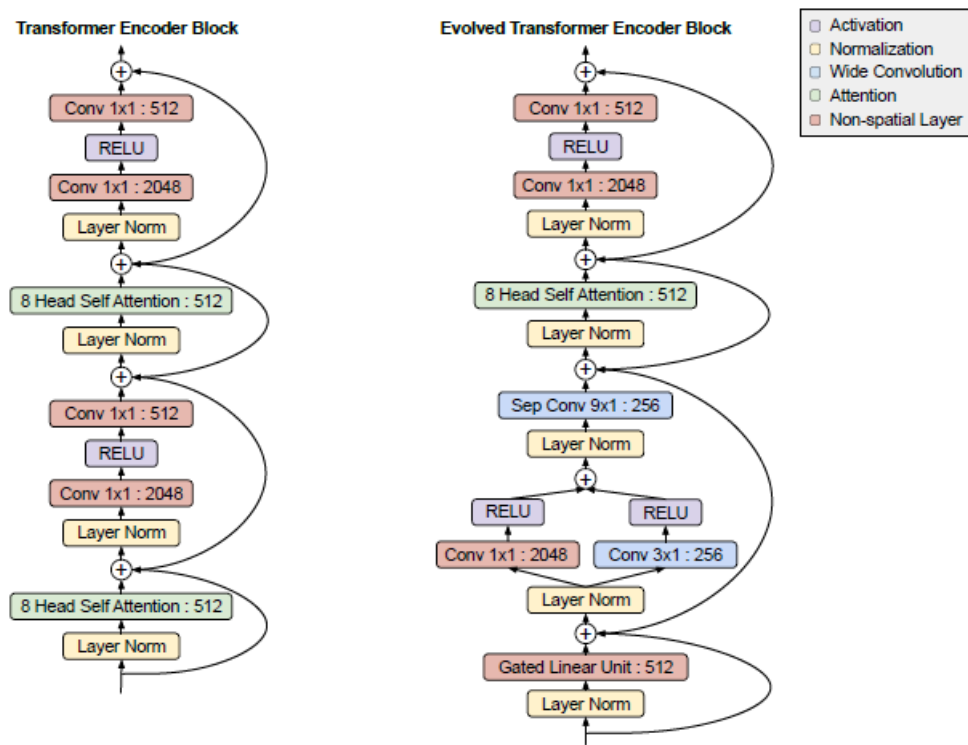
介绍：

1. 闭域聊天机器人只是对关键词进行响应，开域问题则需要机器人适应任意的对话主题。传统方法依赖于检索，规则等构建的复杂对话管理系统，如MILABOT，CLeverbot等；端到端是突破方向。
2. SSA评价标准考察两个方面，话是否说得通（有意义），说的话是否具体（生动）。模型采用静态和动态交互的两种测试方式；
3. 评价各类聊天机器人：首先定义了人类对话的要素，提出SSA评价标准；基于该标准评估了Cleverbot，DialogPT以及Mitsuku和XiaoIce模型；探讨论文提出的SSA标准和主流常用的自动评测标准PPL的关系
4. 验证了SSA标准与混淆度PPL标准是高度负相关的，即PPL越低，SSA就越高；[未成为行业共识]

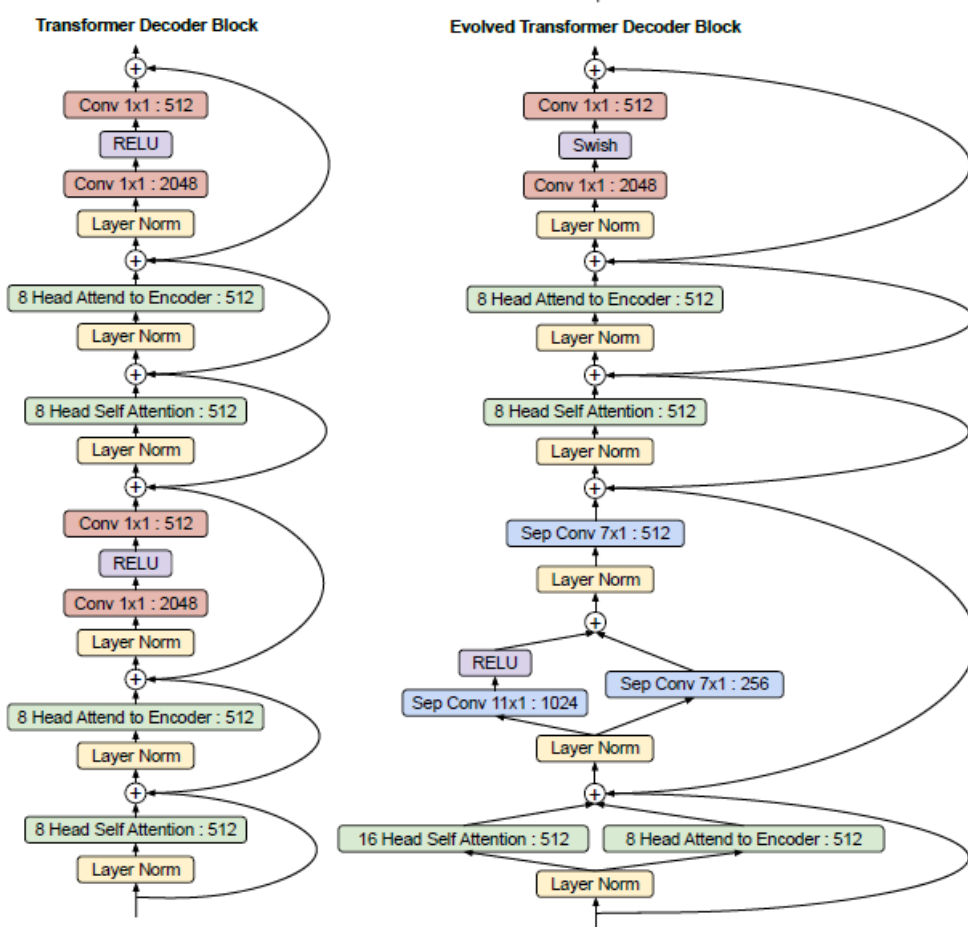


Meena 模型结构

论文在模型方面没有重大创新，使用基于Transformer的Seq2Seq模型。相比于DialogPT采用标准的GPT-2，Meena采用的是google通过网络架构搜索NAS方法得到的进化版的Transformer模型（ET）；Meena整体结构由13个ET编码器和1个ET解码器构成；Meena主要是模型规模增大了很多；采用了巨大的隐含层（2560个隐层单元，32个attention heads），总参数达到2.6B，相比最大规模的DialogPT参数量是762M。



上下图分别是ET和标准Transformer编解码的对比，两者结构上差别不大，ET论文中提到四个值得注意的架构变化：1) wide向的通道可分离卷积；2) 门控线形单元；3) 采用分支结构；4) swish激活函数。根据论文结论在小参数模型时ET比标准的Transformer性能更好，大模型参数时相差不多。



Meena数据

1. Meena训练数据来自于公开的社交媒体对话，总词数是40B，文本大小341G；做了大量的数据清洗工作，主要是删除太长和太短的，重复词语多的文本，过滤含有URL和不常见词等，采用(context; response)形式构成训练数据，使用2048个TPU v3核训练了30天，造价昂贵，验证了google大力出奇迹的一贯思路

Meena Conversation 2	
Human:	<i>Hi!</i>
Meena:	<i>Hello! How are you? :)</i>
Human:	<i>I'm doing okay. How about you?</i>
Meena:	<i>I'm doing pretty good, thanks for asking! :)</i>
Human:	<i>What's your name?</i>
Meena:	<i>It's Kimberly! What's your name?</i>
Human:	<i>That's a nice name. My name is Paul.</i>
Meena:	<i>Nice to meet you Paul! :)</i>
Human:	<i>Nice to meet you as well Kimberly</i>
Meena:	<i>So, what do you like to do?</i>
Human:	<i>I like going to the movies</i>
Meena:	<i>Cool! What's your favorite movie?</i>
Human:	<i>Pulp Fiction</i>
Meena:	<i>Nice! I love Reservoir Dogs. :)</i>
Human:	<i>Yes, that movie is good as well</i>
Meena:	<i>Who's your favorite character in Reservoir Dogs?</i>

核心认知

- 提出了评估多轮对话效果的指标SSA；
- PPL和SSA高度负相关，所以可用PPL自动评估模型效果；
- 足够大的端到端模型可以打败复杂架构的对话系统。

2020相关进展

- Facebook的Blender: Recipes for building an open-domain chatbot
- 微软的DialogPT: Large-Scale Generative Pre-training for Conversational Response Generation
- the Evolved Transformer: <https://arxiv.org/pdf/1710.05941.pdf>