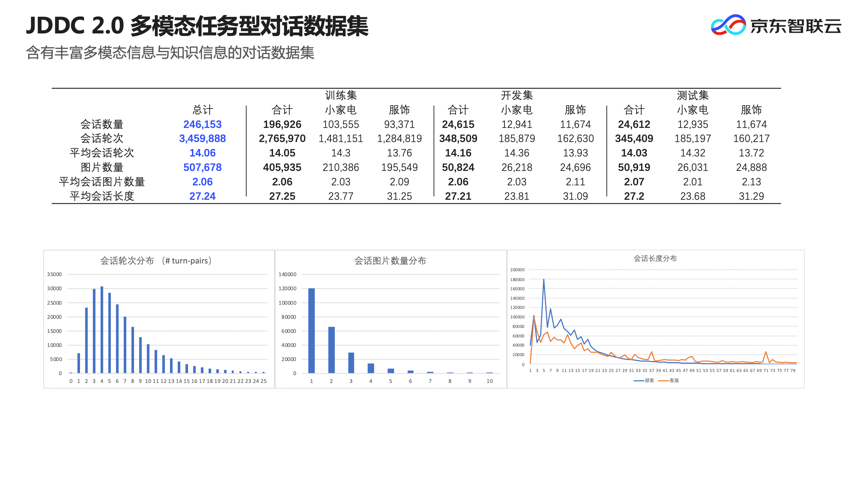
人机对话系统是自然语言理解领域重要的研究方向之一，图灵测试的核心就是人机多轮交互。近些年随着语音助手、虚拟数字人、智能服务机器人的普及，智能对话系统在零售、旅游等领域变得越来越重要。为推动跨模态智能对话与人机交互技术发展，智源-京东跨媒体对话智能联合实验室在2018年首届任务导向型对话挑战赛、2019年基于知识增强的任务导向型对话挑战赛的基础上，共同举办2020年多模态多轮任务导向型对话挑战赛。今年的大赛聚焦大规模复杂零售场景下多模态人机交互问题，通过打造多模态话系统提升人机交互的自然度和体验。10月31日中国计算语言学大会（CCL 2020）技术评测研讨会智源-京东多模态对话挑战大赛任务研讨会在线召开，并通过视频直播对外开放。CCL是中国中文信息学会(CIPSC)的重要会议，是中国最大的自然语言处理学者和专家的社区。经过二十多年的发展，CCL被广泛认为是最权威的，全国最具影响力、规模最大的NLP会议。

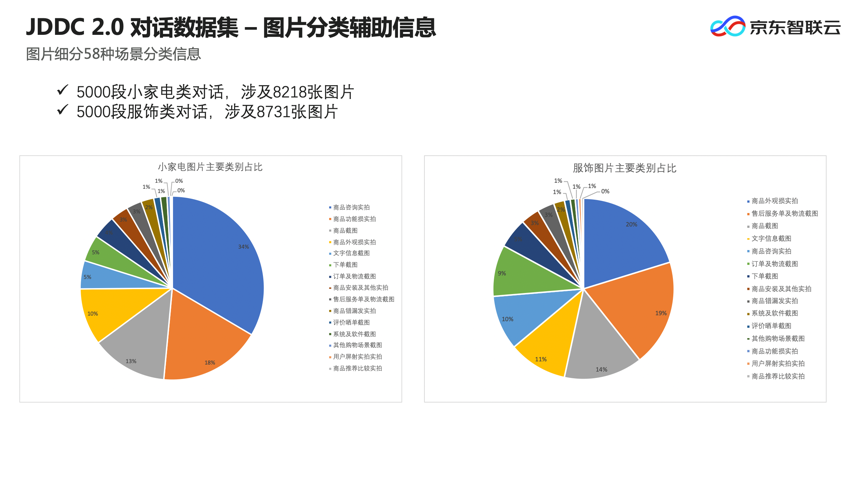
此次研讨会上，主办方单位京东AI研究院对本年的对话大赛进行了全方位的回顾与总结。

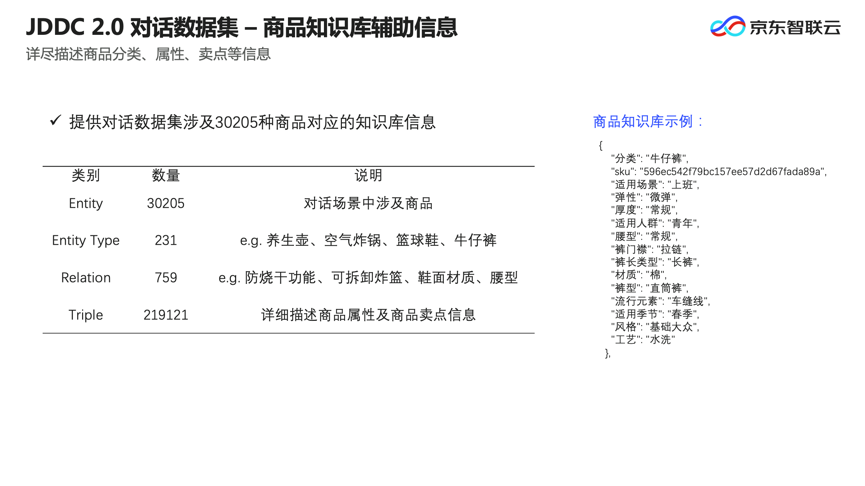
为了挑战真实场景下多模态任务型对话这个难题，这次对话大赛构建了JDDC 2.0 (Jing Dong Dialogue Corpus 2.0) 数据集，数据集由服饰品类和小家电品类共17家店铺线上金牌客服的含有多模态信息的对话日志组成。在目前真实线上服务场景中，用户发送的是多模态的图文信息，客服一般回复的都是文本信息。所以，本次大赛的数据集和大赛考察的重点是多模态的上下文语义理解，单模态的文本应答这样一个任务场景。一个好的数据集就是需要反映真实世界的状况，来自线上电商线上客服的对话数据JDDC 2.0，在语言表达、会话话题转换与情感交流等方面，相比目前普遍采用众包方式产生的数据集，更加自然而真实。本数据集也是首个中文多模态对话数据集，其中包含多模态对话24.6万段，平均会话长度为14轮。该数据集在今年大赛结束后，也在对话大赛官网（www.jddc.com）向公众公开发布，届时欢迎大家下载引用。



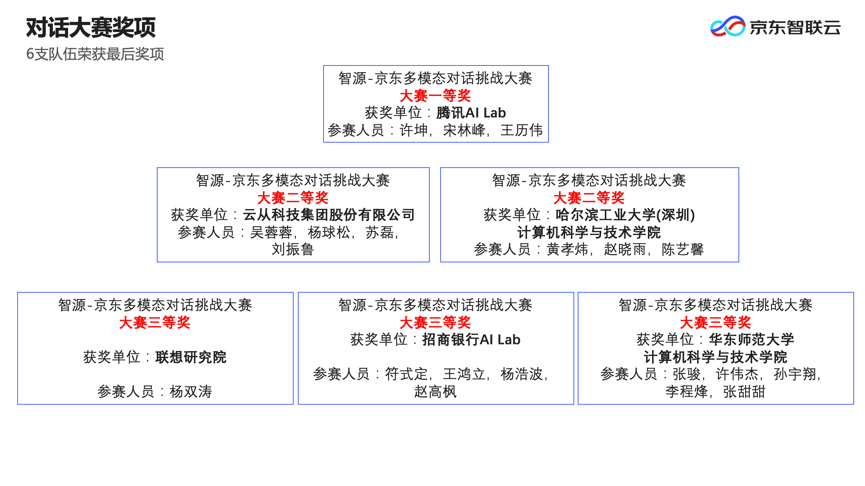


这次大赛为了使参赛者更方便的使用对话中的图片信息，数据集选取了5000段小家电类对话和5000段服饰类对话，对其中所包含的图片进行了人工分类打标，共提供了1.69万张图片的58分类标签。此外，数据集提供了对话中所涉及的商品知识库信息，整个知识库一共包含有商品知识三元组21.9万条，其中含商品实体共3w多个，分别属于231种商品三级品类，共包含759种商品属性关系。商品属性关系准确详尽，不仅包含商品基本属性信息，还包含商品卖点信息，可应用于商品属性应答、商品推荐等场景。

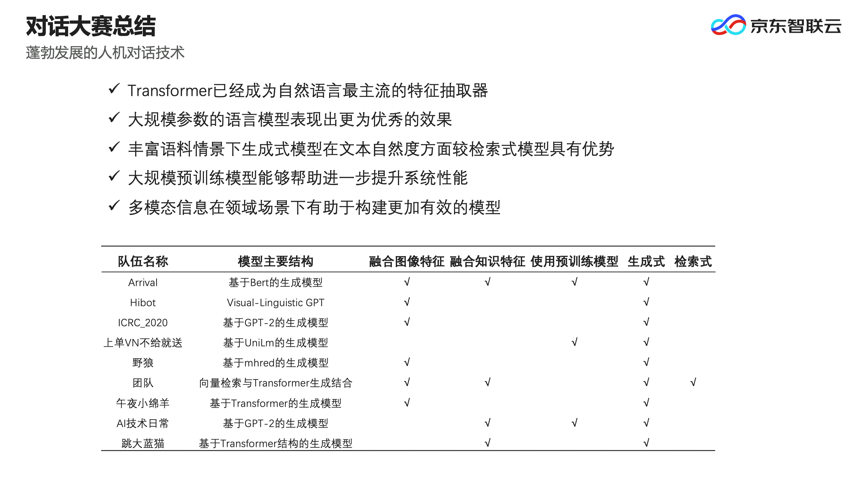




本年度对话大赛从5月25日开赛，到9月15日截止模型提交，总共历时17周，报名参赛选手共734人。京东云还为各位参赛选手提供了免费的P40 GPU计算资源，累计帮助大赛选手提供GPU算力超过8万小时。最终，通过初试评测与复赛评测两个阶段，采用自动评测、人工评测与技术方案评价相结合的方式，评选出本年获得一二三等奖的队伍。CCL 2020大会评测委员会和京东云共同为此次优胜队伍颁发获奖证书，此外京东云还向获奖队伍提供奖金奖励，其中一等奖奖金5万元，二等奖奖金2万元，三等奖奖金1万元。



通过此次大赛优胜团队的参赛模型，可以看出目前蓬勃发展的人机对话技术的一些趋势。首先从各团队的模型主要结构就可以看出，要么是使用Transformer based语言模型，要么是直接使用Transformer构建编解码器，可见Transformer已经取代RNN成为自然语言处理最主流的特征抽取器。Transformer在语义特征提取能力，长距离特征捕获能力，并行计算能力及运行效率等方面都表现出明显的优势。在最终排名前几名的模型中，普遍看到了Bert，GPT，UniLM这样一些大规模参数的语言模型，更多的模型参数能够带来更好的模型性能，从而能够产生出更好的对话效果。今年对话大赛的优胜队伍几乎一边倒的使用了生成式模型，这也说明在语料充足的情况下，各种以Transformer为基础的生成模型，在对话生成的语言流程程度、应答相关性等方面表现与检索式模型没有区别，甚至会更有优势。另外，比赛中的第一名和第四名都使用了大规模数据预训练的BERT模型作为基础，这两个模型在人工单项评分中应答的满意率也略高一些，可见大规模预训练模型有助于进一步提升系统性能。此外，通过比赛也可以看出，恰当融合多模态知识，在某些场景下才能够泥补单模态的信息缺失，能够提供更加满意的答案。



从此次比赛的人工评分可以看出，比赛中最好的模型也就能回答好45%的问题，这说明真实场景下的对话任务还有很多挑战仍待解决。例如，大多数情况下顾客都是问电磁炉怎么调节功率这样的问题，对于这些常见问题，我们的参赛模型都能给出正确完整的回答，但是偶尔有几个顾客问怎么在电磁炉上设置温度的问题，我们的模型就不能给出正确而又自然的应答了。这就是我们遇到的一个挑战，要求我们模型不仅能回答好高频问题，对于长尾问题也能给出满意答案。另外，在购物场景中对话轮次相对较长，且前后内容相关性很强，只有有效建模上下文，才能准确表达上下文中的顾客所中意的款式这样的细节信息。比赛中我们模型有不少回答，仅看当前或近几轮的交互，貌似回答正确流畅，但再往上回溯几轮对话的话，你就会发现产生的答案在逻辑细节上是存在矛盾的。

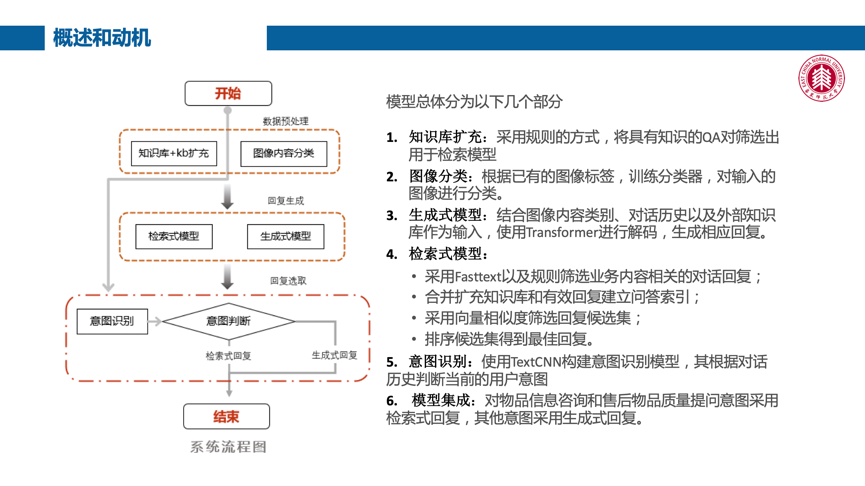


在多模态交互中，只有读懂了图，才能知道客户的问题所在。在我们线上零售客服这个场景下怎么才能够更为恰当使用到这些多模态信息，我们是用整张图片提取特征，还是用图片的ROI提取特征，还是用图片细分类或者OCR提取文字作为特征等等，这些提取到的多模态特征又如何与文本模态特征进行融合呢，我们这次比赛模型已经做了一些尝试，但这些问题都还需要进一步的深入研究。此外，在真实场景中有时顾客感觉受了委屈会比较生气，这就需要我们的模型在解决问题的同时还要能够与客户有比较好的情感交流，提供更贴心的服务。另外，有些商品不可避免的有些小瑕疵，也不影响使用，一般人工客服都会选择对顾客进行小额经济补偿，但如果我们的模型选择每单都让顾客退货的话，这样其实也会造成一定的社会资源浪费，因为退货商品是不可以进行二次销售的。这就涉及到如何在对话中进行决策的问题了。总之，真实对话场景的还有很多待研究解决的问题。

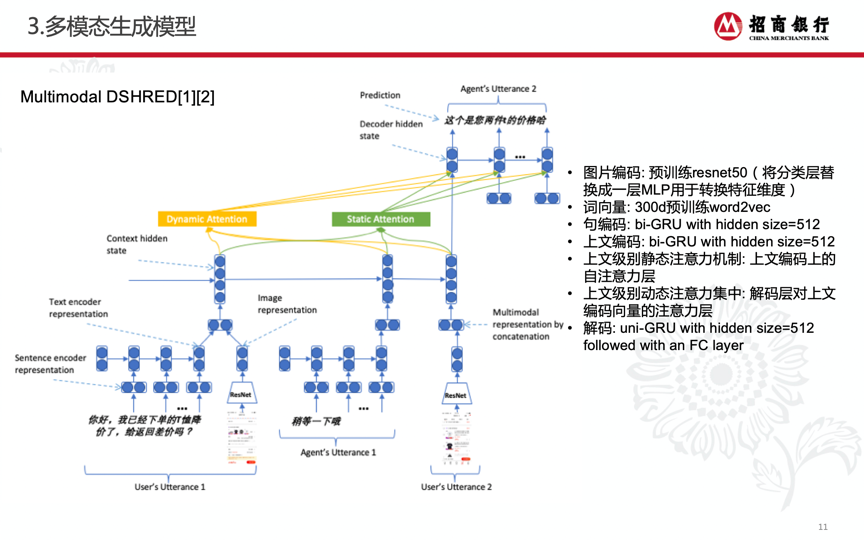


研讨会上，六支获奖团队受邀详细分享了此次大赛的比赛方案，对比赛中的得失进行了总结与分析，为广大人机对话系统研究者带来精彩的学术和技术盛宴。

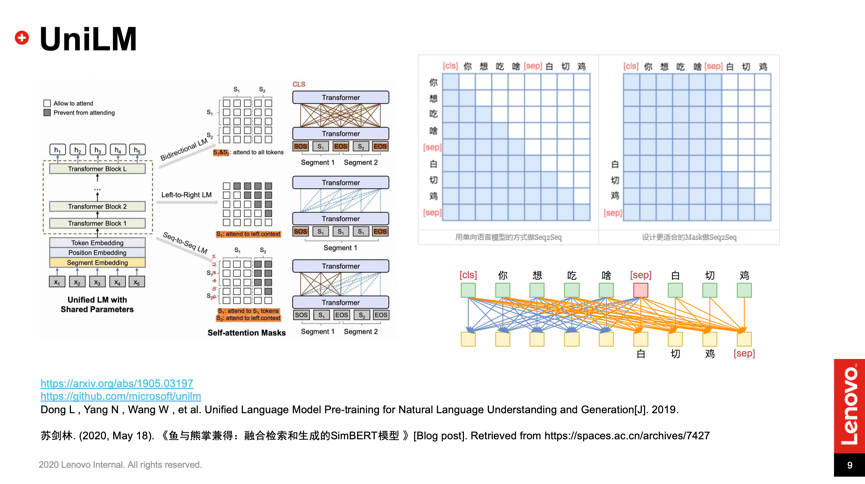
本次大赛的第六名来自华东师范大学的团队，报告内容是“基于用户意图判断的多模态对话集成系统探索”，介绍了本次比赛中采用的检索式模型与生成式模型相结合的方案。检索式模型，采用规则的方式扩充知识库，将具有知识的QA对筛选出用于问题检索；生成模型，结合图像内容类别、对话历史以及外部知识库作为输入，使用Transformer进行解码，生成相应回复。最后结合二者模型的优势，对物品信息咨询和售后物品质量提问意图采用检索式回复，其他意图采用生成式回复。



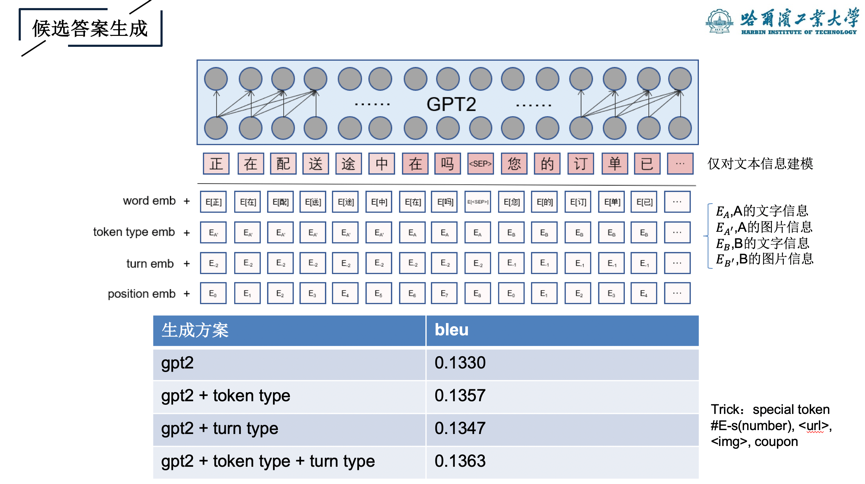
本次大赛的第五名是来自招商银行AI Lab的团队，报告内容是“生成式模型在多模态对话任务上的探索”，介绍了基于GRU的分层编码、解码以及Resnet做图片特征抽取的多模态生成模型。方案中分别对文本和图片编码后再拼接进行进一步的上文信息联合编码；针对上文级别信息的筛选，加入了两种注意力机制层，静态注意力机制对上文信息自身进行关注，动态注意力机制在每一步解码目标回复时对上文信息进行筛选。最终目标回复通过解码层对上文信息进行整合解码得到。

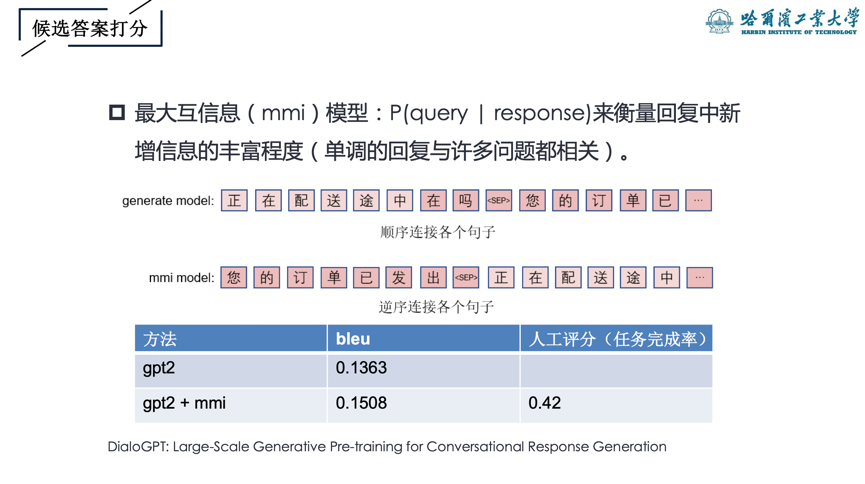


本次大赛的四名是来自联想研究院的团队，报告内容是“UniLM模型在多轮对话中的应用和实践”，介绍了大赛过程中的使用UniLm模型的经验。此次大赛中，使用JDAI-BERT大规模预训练Bert为初始化权重，以国产有效开源实现bert4keras为基础，采用UniLm的方案构建复杂场景对话生成模型，并取得了良好的效果。

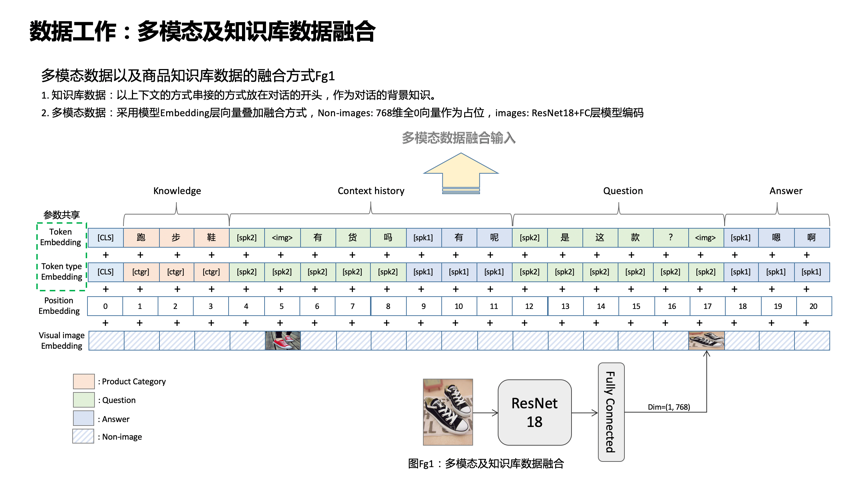


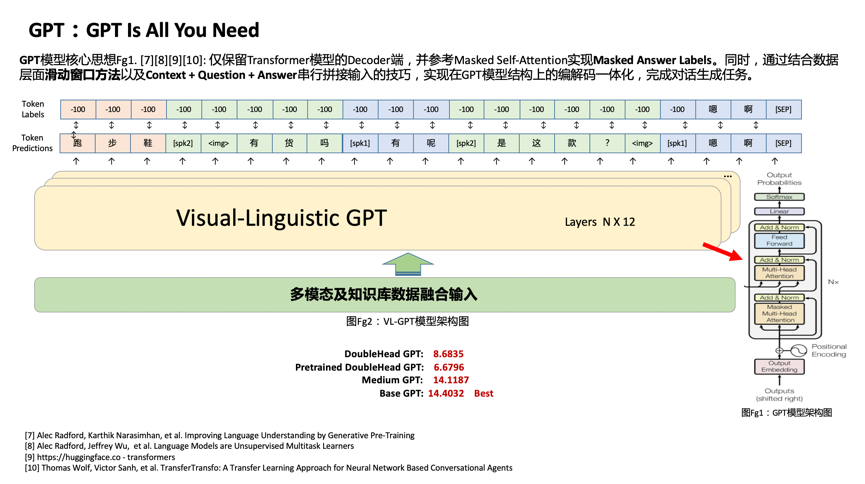
本次大赛的第三名是来自哈尔滨工业大学（深圳）的团队，报告内容是“融合图片信息和对话模式的GPT-2对话应用”，介绍了将图片进行OCR处理提取文字信息并用于对话生成的方案。方案GPT2作为语言模型，在给定图片信息和文本信息的条件下，对文本信息的概率分布进行建模；并使用最大互信息（mmi）模型P(query | response)来衡量回复中新增信息的丰富程度，从而给出最优答案。



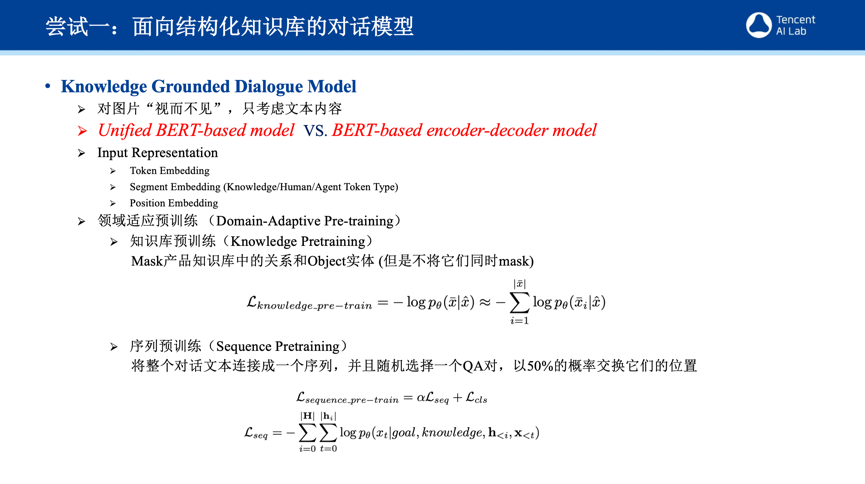


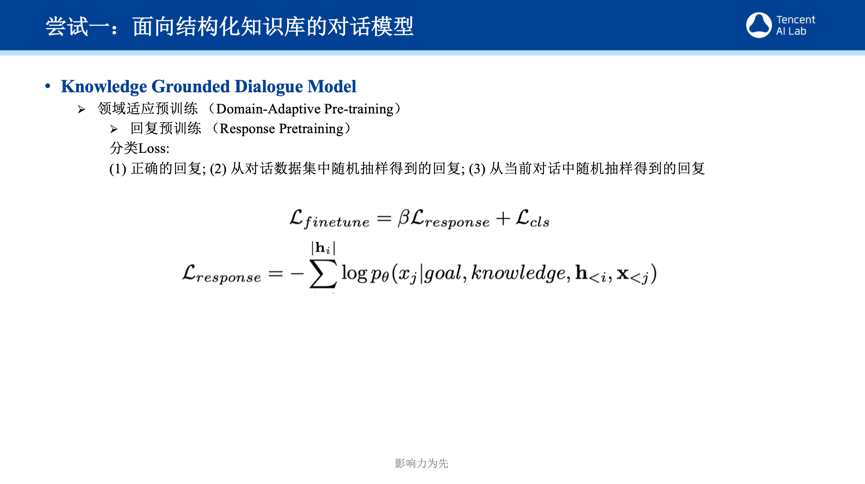
本次大赛的第二名是来自云从科技的团队，报告内容是“基于GPT模型的多模态融合方法及系统”，介绍了大赛中使用的多模态方式打造GPT对话模型的实践经验。方案中模型采用上下文的串接的方式将对话涉及的商品知识三元组放在对话的开头，作为对话的背景知识；多模态图片信息则采用ResNet模型提取特征后，采用模型Embedding层向量叠加与文本特征相融合，并最终输入GPT结构为核心实现的编解码一体化模型，完成对话生成任务。

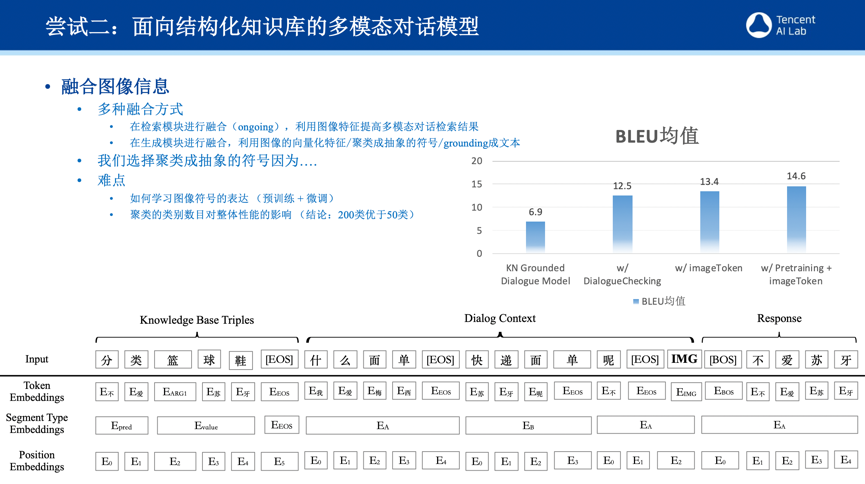




本次大赛的第一名是来自腾讯AI Lab的团队，报告的内容是“基于预训练语言模型和结构化知识库的多模态对话生成模型”，方案充分利用对话中多模态信息与知识信息构建基于预训练BERT的对话生成模型取得良好效果。方案中通过大量对比实验选择出最优解决方案，首先通过领域适应预训练，借助于Mask产品知识库中的关系和Object实体然后进行实体预测训练和应答回复训练，训练出一个面向结构化知识库的对话模型；然后，使用ResNet模型抽取图片特征并通过K-means实现图片聚类，将图片抽象化为200类token信息，并融入先前训练好的面向结构化知识库的对话模型中，通过训练产生能够同时支持多模态信息和知识信息的生成模型。







今年的多模态对话大赛已经完美落幕，为解决真实场景对话的各种挑战，京东AI已经连续举办了三届对话大赛，明年也会在同个时间周期举办2021年的对话大赛。无论明年的赛题怎么变换，但是京东AI的初衷不会变，就是要为解决真实场景下人机交流障碍问题，实现拟人化的人机交互服务。欢迎大家继续参加京东AI明年的对话大赛。