

基于强化学习的跨媒体导航任务

11821012 常嘉乐 导师：庄越挺

摘要：

跨媒体导航是一个培养具有第一人称视角代理在现实世界中执行自然语言指令的任务。代理通过分析视觉场景并遵循自然语言说明选择相应动作向目标移动。这与其他一些视觉语言任务不同的是，在那些任务中，视觉感知和自然语言的输入通常是固定的（例如视觉问答），而在跨媒体导航任务中，代理可以与实际环境进行交互，并且它所感知到的像素在其移动时正在发生变化。因此，代理必须学会根据其对世界的感知和对自然语言教学的理解，将视觉输入映射到正确的行动中。

本文从室内级别和室外城市级别的导航任务入手，不仅仅分析了在这两种场景中基本的视觉语义导航，还分析了在这之上添加对话或者问答的复杂任务导航，最后还提出了一些小小的个人见解。

关键词：强化学习 跨媒体导航

正文：

跨媒体导航任务：

注：这里的跨媒体导航任务不同于 Slam 技术，是没有地图的导航任务。

(1) Grounded Language Learning in a Simulated 3D World

我们的生活越来越依赖 AI，它会帮助我们做决定，小到推荐哪些新闻事件，大到该买什么股票收益较大。甚至有的时候，它还会直接替我们采取行动。因此，一个越来越迫切的需求也随之到来：人需要和 AI 在真实环境中进行沟通，同时指挥和引导它们。

2017 年 DeepMind 发表了一篇论文，在摘要中 DeepMind 用两个词来形容他们想让 agent 学习人类语言学到什么程度：

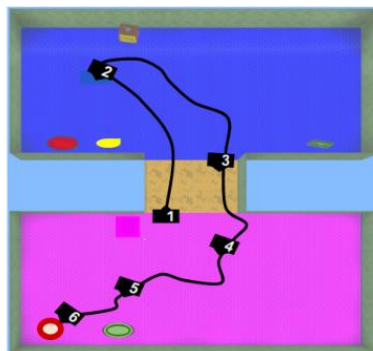
Grounded，有基础、接地气的，可以把语言中的词和 agent 在环境中遇到的物品、经历的行为联系起来；

Embodied，即可以具体表达出来的。

这两个词汇是跨媒体导航任务的中心思想。

在这篇论文中，agent 身处一个虚拟的 3D 迷宫，人类会给 agent 一个需要找到的目标，同时会提供目标物体的描述以及其周围环境的描述，比如有一条指令是这样的：“把绿色物体旁边的红色物体捡起来”。如果 agent 能够成功并正确地执行了这条用人类语言描述的指令，那么它就可以获得奖励。

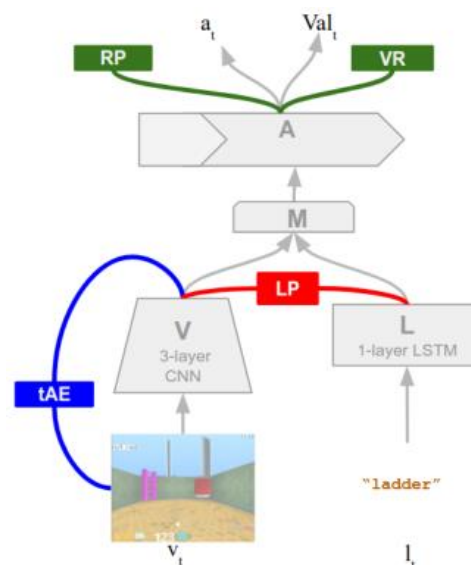
Top down view



我们可以看到：**agent** 在收到指令之前处于位置 1，收到指令后它把两个房间逛了一遍，具体的去查看了每一个房间里的物体及其相对位置，最后在位置 6 找到了指令中的目标物体。

Agent 这种探索、选择的行为，并没有预先建模（没有先验知识），完全是借助强化学习的激励机制学会的，**agent** 可以通过将语言符号和它周围物理环境中出现的感知表征与行为序列关联起来，从而学习人类的语言。这样的探索训练有数十万种变体，**agent** 会遇到不同的房间布局、不同的物体摆放位置等等。

由于没有先验知识，**agent** 在一个完全陌生的环境中，也可以正确执行新的指令，这种泛化和自我扩展语义知识的能力，说明这种方法可以让 **agent** 理解模糊的自然语言与复杂的实体世界之间的关联。

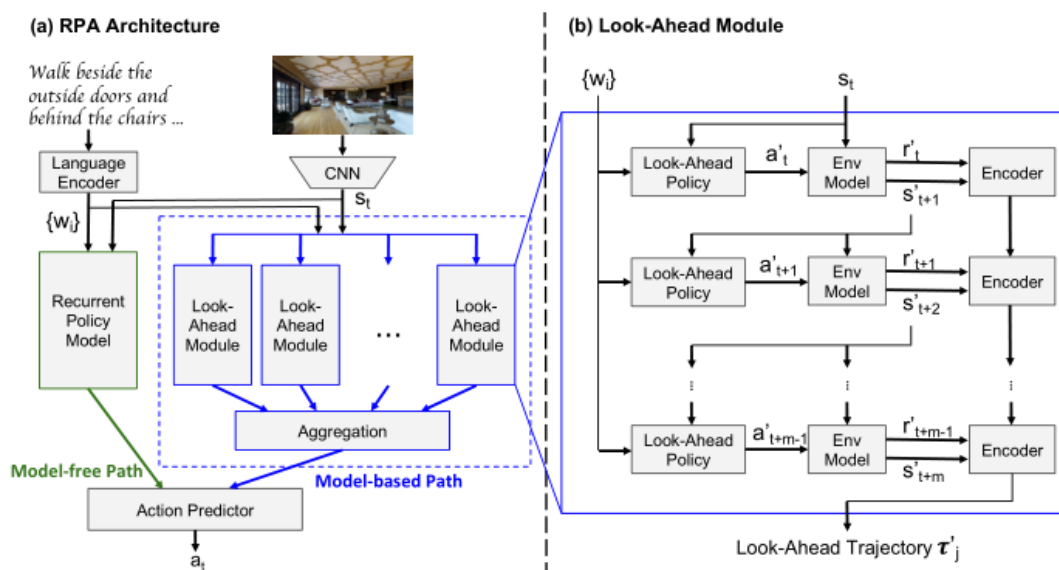


上图就是具体的网络结构：图中浅颜色的部分是模型的基础部分。我们可以看到：模型将 **agent** 观测到的图像通过一个 tAE（Temporal autoencoding）来进行编码，而人类给定的指令则通过一个 LSTM 来进行编码，最后将视觉环境与语言指令的嵌入表示进行结合（M 模块的作用）。双向 LSTM（A 模块）中的隐藏状态通过一个策略函数，在可能的动作序列和状态估计上计算概率分布，从而预测即将采用的动作以及下一状态的价值。

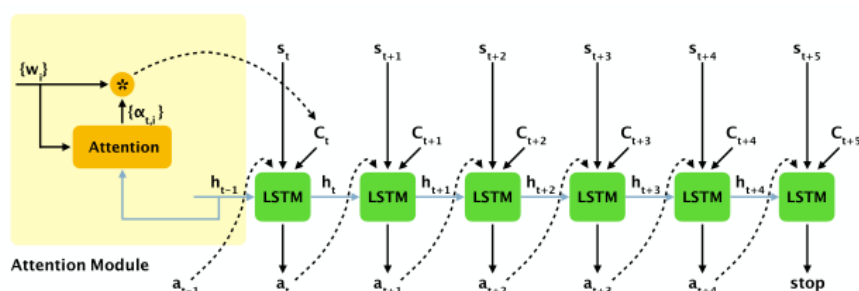
另外，论文作者还使用了一些小的技巧——蓝色部分代表模型将使用当前观测到的图像以及当前的动作来预测下一步的图像；红色部分是语言模型和观察图像之间的对应，它的输入是图像，输出是单词，可以这样解释：**agent** 可以发出它认为能最好地描述它目前正在观察的东西的单词；绿色部分则表示着 reward prediction (RP)和 value replay (VR)，这些技巧都可以使得模型更加的鲁棒。

（2）Look Before You Leap: Bridging Model-Free and Model-Based Reinforcement Learning for Planned-Ahead Vision-and-Language Navigation

这篇是 UCSB 王威廉组的论文，发表于 2018 年的 ECCV 上，我们可以仔细解读一下他的网络结构：



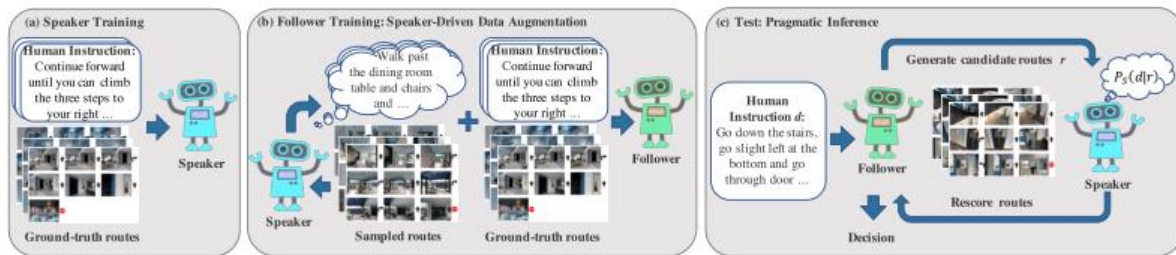
一些关于强化学习以及特征提取的基础知识在这里就不详细介绍了，本文最大的创新点是第一次在跨媒体导航任务中将基于模型的强化学习算法和免模型的强化学习算法结合到一起。如上图所示，最下方的当前状态动作预测模块，其拥有两个输入，左边的是我们常见到的 LSTM 编码器，右边就是基于模型的强化学习算法计算得到的结果。如果单看左边的 LSTM 和下面的动作预测模块，那么这就是一个很显然的 policy-based 的方法。



右边的模块中，在做 Aggregation 之前有许多的 Look-ahead 模块，这些就是基于模型中的“模型”，也就是基于环境的建模。虽然基于模型的强化学习算法早已出现，然而，是 Google Brain 在一篇名为《World Models》的论文中将深度学习和传统方法结合为强化学习的发展开辟了新的道路。这里的模型，输入是当前的状态和接下来的动作，输出是得到的奖励以及下一步的状态，我们可以理解为 AI 在做动作之前会在大脑中进行预想，这也是论文标题中 Look-ahead 的真正含义，当然，预想几步以及预想动作的数量是超参数，是自己指定的。论文里，这两个参数分别为 2 和 5。

(3) Speaker-Follower Models for Vision-and-Language Navigation

这篇论文训练了两个 agent，一个名叫 speaker，另一个名叫 follower，如下图所示，这两个 agent 共同组成了一个跨媒体导航系统：



在该系统中，首先，我们利用真实的导航路线和对应的指令训练一个 **speaker** 模型。在训练 **follower** 之前，我们会对环境新采样一些路线同时为之提供合成的导航描述（生成指令），然后将其用作对 **follower** 的额外监督信号，我们可以理解为数据增广。

在模型测试的时候，**follower** 生成 K 种可能的路径作为对人类给定指令和起始上下文的解释，此时 **speaker** 需要务实地对这些可能的路径进行排名，选择一个在上下文中对人类指令提供最好解释的路线，排名的过程就是使用了基于模型的强化学习算法，需要向前看几步然后才能选择出最合适的路径。

(4) Reinforced Cross-Modal Matching and Self-Supervised Imitation Learning for Vision-Language Navigation

这篇也是 UCSB 王威廉组的论文，发表于 2019 的 CVPR 上，在审稿期间获得了满分。这篇论文将强化学习和模仿学习结合，提出了新型强化跨模态匹配 (Reinforced Cross-Modal Matching, RCM) 模型，通过强化学习方法将看得到的局部环境和看不见的全局场景联系到了一起。下图是 RCM 模型的框架：

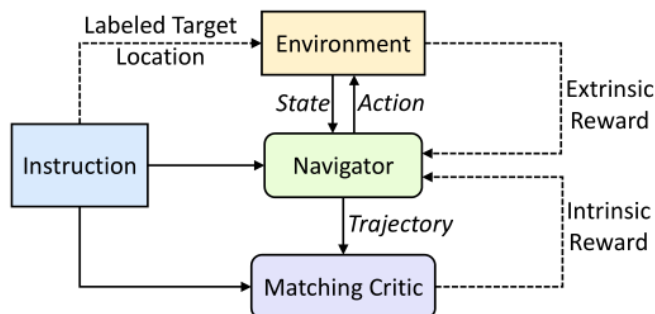


Figure 2: Overview of our RCM framework.

从上图中可以看出：在 RCM 模型中，Navigator 扮演着一个中心角色。它通过学习文本指令和局部视觉图像中跨模态的场景，让智能体推断潜在指令，明白到底应该向哪看，向哪里前进。

当然，Matching Critic 模块也非常重要，它是用来评估从原始指令中重建场景执行情况，并且设置了循环重建奖励，即图中的 Intrinsic Reward。

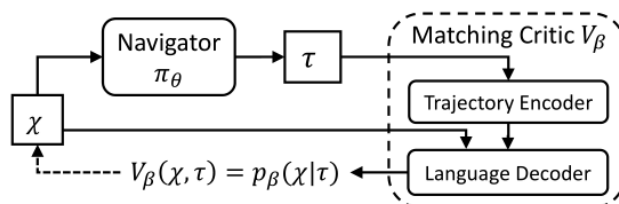


Figure 4: Cross-modal matching critic that provides the cycle-reconstruction intrinsic reward.

这样的思想和 Speaker-Follower Models 中 speaker 对环境中新采样一些路线生成合理的导航描述是有着异曲同工之妙的，只不过 Speaker-Follower Models 将之用在了数据增广上，意图解决数据量不够大的问题，而本文是将之定义为奖励函数的一部分，与模型的整体进行端到端的联合训练。

接下来我们看一下论文对奖励函数的定义以及训练的公式：

Extrinsic Reward:

$$r(s_t, a_t) = \mathcal{D}_{target}(s_t) - \mathcal{D}_{target}(s_{t+1}), \quad t < T$$

$$r(s_T, a_T) = \mathbb{1}(\mathcal{D}_{target}(s_T) \leq d)$$

$$R_{extr}(s_t, a_t) = \underbrace{r(s_t, a_t)}_{\text{immediate reward}} + \underbrace{\sum_{t'=t+1}^T \gamma^{t'-t} r(s_{t'}, a_{t'})}_{\text{discounted future reward}}$$

Intrinsic Reward:

$$R_{intr} = p_{\beta}(\mathcal{X}|\pi_{\theta}(\mathcal{X})) = p_{\beta}(\mathcal{X}|\tau)$$

Learning:

$$A_t = R_{extr} + \delta R_{intr} - b_t.$$

$$\nabla_{\theta} L_{rl} = -A_t \nabla_{\theta} \log \pi_{\theta}(a_t|s_t)$$

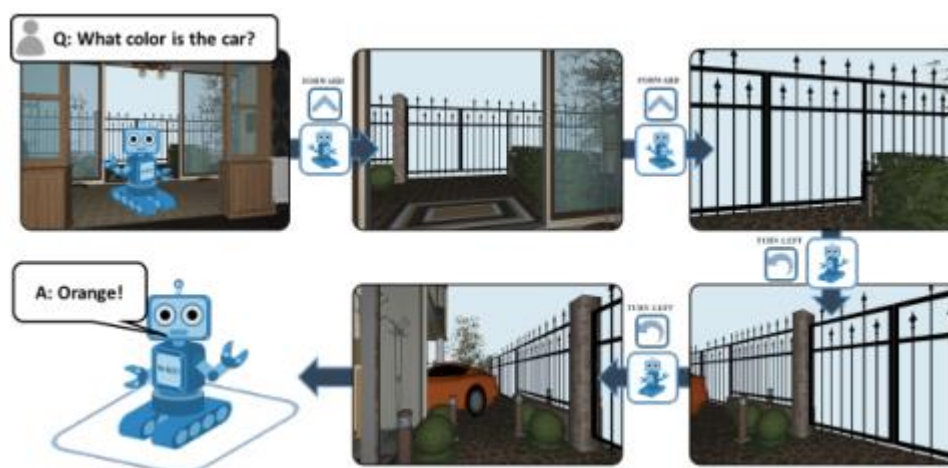
当然，为了让智能体能够在杂七杂八的现实物体中快速专注于有用的场景，论文还提出了一种自监督的模仿学习，帮助智能体探索未知场景中没有标注的数据，让智能体学会利用走过的老路来获取经验。

简单来说就是 Navigator 去执行多种 Roll-outs，将其中评估出的比较好的轨迹存储在缓冲区中，方便 Navigator 在后面的路径中模仿。这样，Navigator 在行进中就能逐步接近最好的路径，规划出最佳决策。

(5) Embodied Question Answering

为了给那些对人类微观管理依赖程度较低且更具通用性的系统铺平道路，Facebook 人工智能研究院 (FAIR) 开发了一系列用于训练和测试自主智能体的虚拟环境以及能够学习智能探索那些环境的新型智能体。

同时，FAIR 提出了一个多步骤任务，称为 Embodied-QA。与聊天机器人或者智能音箱相比，此智能体必须在物理环境中去学习和运行，当智能体被问及单一问题时，例如“汽车是什么颜色的？”或“我的钥匙落在了哪个房间里？”，智能体必须能够理解书面语言，然后使用相机并用第一人称视角去感知周围环境，探索 3D 的室内环境并找到答案。并且为了使得智能体的自主性更加完善，它还将通过自然语言的方式回复该答案以完成自己的使命。下图就是一个具体的例子：



为了训练和评估这些智能体，所需的虚拟环境不仅仅要具备交互功能，还要具备多样化和数量充足的特性，以避免智能体在相同环境中反复运行，这对自主智能体的发展而言是一个更加巨大的挑战。FAIR 的解决方案是一个数据集——House3D，它是由 45000 个手动创建的模拟室内环境所组成的（包括房间的名称以及相应物体的位置）。目前已经开源，可以在 GitHub 上获取。

为了在 House3D 中每次都能进行独特的探索活动，智能体必须在完成任务的过程中学习一系列核心能力——从对室内物体的识别（沙发、椅子等）到对问题的语言理解。

第一种习得的能力是主动感知（Active perception）。一开始目标对象不太可能会刚好位于智能体的视线范围内。因此智能体不是被动地感知问题中提及的对象，而是需要通过探索周围环境主动将目标对象寻找出来。

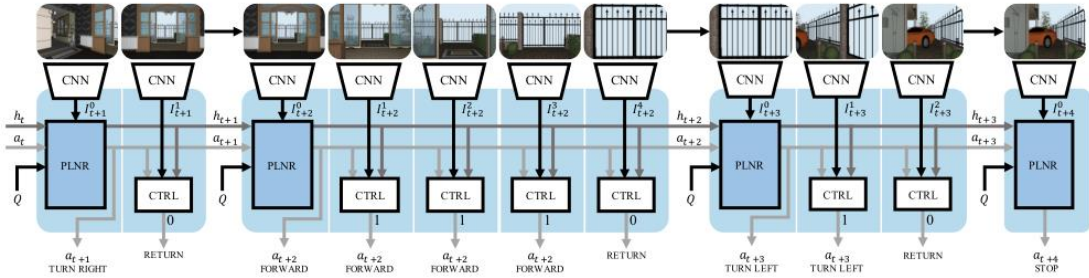
该论文的创新点在于实现了导航的模块化方法，其将导航任务划分成了两个部分——规划者（Planner）和控制器（Controller）。其中，规划者负责选择移动的方向，例如向左移动等等，而控制器则负责确定在该方向需要移动多远（这里使用了 Graves 提出的 Adaptive Computation Time (ACT) RNNs）。这种策略避免了智能体在移动前创建长远而又详细的路径规划，因为这么干会导致更多的错误和更弱的适应性。此外规划者—控制器这种设置也更适合于强化学习，在这里智能体将根据积极或者消极的反馈来动态调整其导航。

下一步，智能体必须学会常识推理，然后在一个新的但是却并非完全陌生的环境中去发挥作用。尽管人们可能知道车库往往位于房屋的周围，因此可以通过外部门进入，但是 AI 系统却需要自己去学习这些知识。随着在不同模拟家庭环境中成功达到目标，智能体必须发展出这种常识，以缩短寻找指定对象然后回答相关问题所花费的时间。作为常识积累的后续阶段，智能体还要学习语言基础，或者是学习如何把单词与特定的动作联系起来。

最后，因为 FAIR 的目标是超越繁琐、逐步的人为监督，从微弱和远期的目标驱动奖励中学习，所以智能体必须学会的最重要的能力是“功劳分配”，这意味着智能体能够知道自己一路上所做的行为哪些是正确的而哪些又是错误的。例如，如果询问“房间里有多少张椅子？”，然后智能体并没有被明确告知需要检查每一个房间，甚至没有被告知要从家中含有椅子的区域开始计算。智能体能够凭借自己的力量去探索和回答问题，而不需要借助人类提供的详细的逐步计划。因此，从它是否正确回答的一个微弱信号来看，它必须自行学习其数百项相互依赖的行动中是哪一项导致了成功。为了增强传统强化学习的效果，FAIR 使用模

仿学习（比较智能体的运动与到达目标的可能最短路径）和奖励塑造（通过“变得更近”和“变得更远”信号随时间变化而改进智能体的表现）来允许智能体把它真正使命拼凑起来，即便该使命在开始时并不明晰。

论文使用了如下图所示的网络结构：

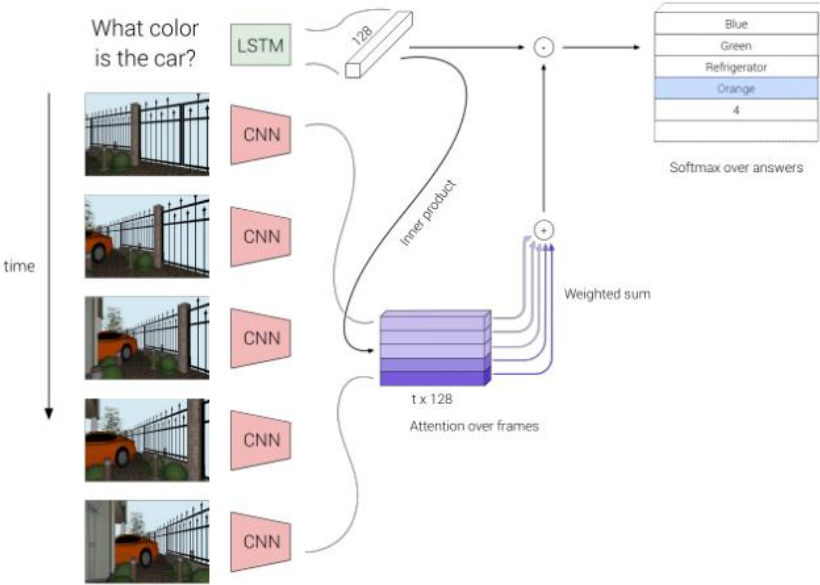


模型可以分为四大模块——视觉、语言、导航以及问答模块。

强大的视觉模块编码了以下几种特征——有关对象属性（颜色和纹理）、语义（对象类别）和环境几何（即深度）的信息。因此，我们在多任务的像素到像素的预测框架下对 CNN 进行预训练并将其视为编码器网络。我们还要训练多个解码网络来解码（1）原始 RGB 值、（2）语义类和（3）每个像素的深度（可以从 House3D 渲染器中获得）。

针对语言模块。我们需要注意的是，我们为导航和问答模块各自学习一个单独的 LSTM 编码器，这是因为每个编码器可能需要关注问题的不同部分。例如，在“厨房里的椅子是什么颜色的”这个问题中，“颜色”对导航模块无关，而“厨房”对问答模块来说也不重要。

导航模块在之前已经介绍过了，即引入 ACT 将导航分解为“规划器”以及“控制器”。这有点类似于分层强化学习的思想。



如上图所示，在导航帧和问题的条件下，问题回答模块需要计算在到达目标位置视觉输入的最后五帧的特征以及问题特征的点积，使用注意力机制对两种编码进行加权组合，以预测最终的答案。

上面几篇论文都是基于室内场景的，接下来我们关注一下跨媒体导航任务

在室外场景、真实城市环境中的应用。

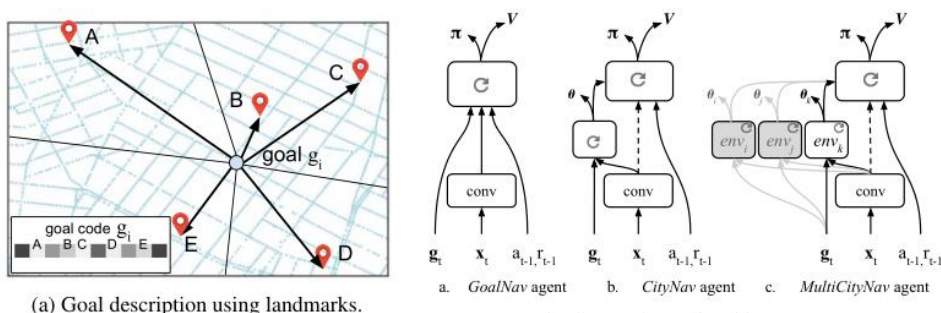
(6) Learning to Navigate in Cities Without a Map

DeepMind 在这篇论文中展示了一种室外的交互式的导航环境，该环境使用来自谷歌街景的第一人称视角图像，并游戏化该环境来训练 agent。尽管谷歌街景图像已经非常标准了，但是人脸和汽车的牌照还是比较模糊、无法辨认。这个 agent 可以使用视觉信息（谷歌街景图像的像素）学会在多个城市中导航。注意此处的导航是通常意义上的导航，并非驾驶，DeepMind 未使用过交通信息，也没有尝试对车辆控制进行建模。

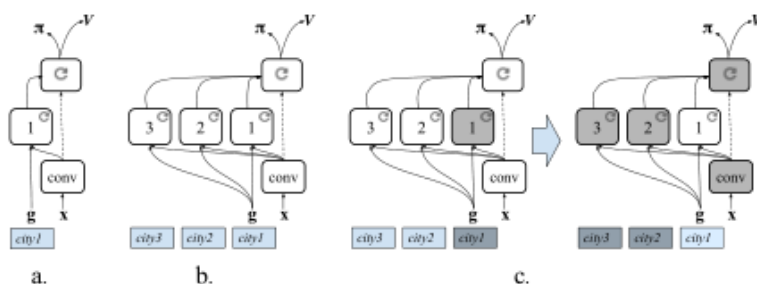
当 agent 到达目标地点时会得到奖励（目标地点是指定的，如经纬度坐标），我们可以将 agent 看作一个没有携带地图、却需要运输货物的快递员。随着时间的推移，它将学会如何穿越整个城市并最终到达目的地：



在下面这个对目标描述的图示中，我们显示了一组附近的有 5 个较近地标和 4 个遥远的地标的例子。Agent 需要到达的目的地 g 是一个向量，是目的地与每个地标之间距离归一化之后 softmax 得到的结果。



DeepMind 还展示了其可以在多个城市中学习执行导航任务，然后稳定地泛化至新的城市。当 agent 访问新城市的时候，人类希望它可以学习一系列新的地标，但是却没必要重新学习视觉表征或行为。因此，DeepMind 使用了 Multi-City 的架构——首先在多个城市中进行预训练，然后冻结策略网络、视觉卷积网络和多个新城的市特定路径，只训练当前城市的特定路径，从而达到不遗忘之前所学知识的前提下获取新知识的效果：



(a) Diagram of transfer learning experiments.

(7) Navigating New York City through Grounded Dialogue

与 Deepmind 那篇单纯的运用视觉信息来导航不同的是，这篇论文将对话（NLP）的内容加入，是真正意义上的城市级别的跨媒体导航。

Facebook 的人工智能研究室（FAIR）在这篇论文中提出了一个新的研究课题——Talk the Walk，其目的是让一对 agent，即一个导游和一个游客通过自然语言相互交流，共同实现一个共同的目标：让游客到达正确的位置。

导游可以接触到地图（论文中使用了纽约城的数据），知道目标的位置，但他不知道游客具体在哪里，我们可以理解为这是一个“云导游”，他不跟在游客的身边；游客对真实的环境有 360 度的观测，但他没有地图，因此也就不知道目标位置，更不知道去地图的路。因此这两个 agent 之间需要通过沟通一起工作，才能成功解决任务。下图该任务的一个示例：



我们可以看到，Talk the Walk 是首个将所有三个要素结合在一起的任务：感知（游客 agent 观察环境）、行为（游客 agent 在环境中导航），以及语言交互达成目标（导游 agent 为游客 agent 提供引导帮助其实现目标）。

论文中使用的纽约城的街景地图数据，是 MTurk 众包手动收集的五个纽约市街区的 360 度视图。这些街景环境被整合到 ParlAI 中，ParlAI 是 Facebook 的一个用于训练 AI 的框架，支持很多任务，包含的数据集包括 SQuAD, bAbI tasks, MS MARCO, MCTest, WikiQA, WebQuestions 等等。

实验中使用的自然语言数据，也是 MTurk 用几周时间收集出来的真人对话，其中包含大约 10k 个成功的导航对话。平均来说，人类需要超过 62 次行动（对话和走路）才能顺利到达目标地点。

虽然我们的目标是让 agent 通过自然语言来处理收集到的信息，但论文中显示：当 agent 使用“合成语言”或者是“连续的向量”进行通信的时候，任务的完成效果远远好于使用“自然语言”，因为前两者更依赖于使用简单的符号来传达信息和位置（即使失去了自然语言的结构，但是关键词被生成了出来）。这种不那么自然的数据通信方式不仅优于人类的聊天，还能让 agent 比人在自然语言聊天中更简单快速找到前进的道路。

Method	Decoding	Utterance
Observations		(Bar)
Actions		-
Human		a field of some type
Supervised	greedy sampling beam search	at a bar sec just hard to tell which is a restaurant ? im at a bar
Policy Grad.	greedy sampling	bar from bar from bar and righth righth bulding bulding which bar from bar from bar and bar righth bulding bulding..

FAIR 还提出了一个名为 MASC (Masked Attention for Spatial Convolution) 的机制, 这个机制可以让 agent 快速从语言模型中解析对方回应的关键字内容, 利用该机制可以使测试结果的准确性翻倍:

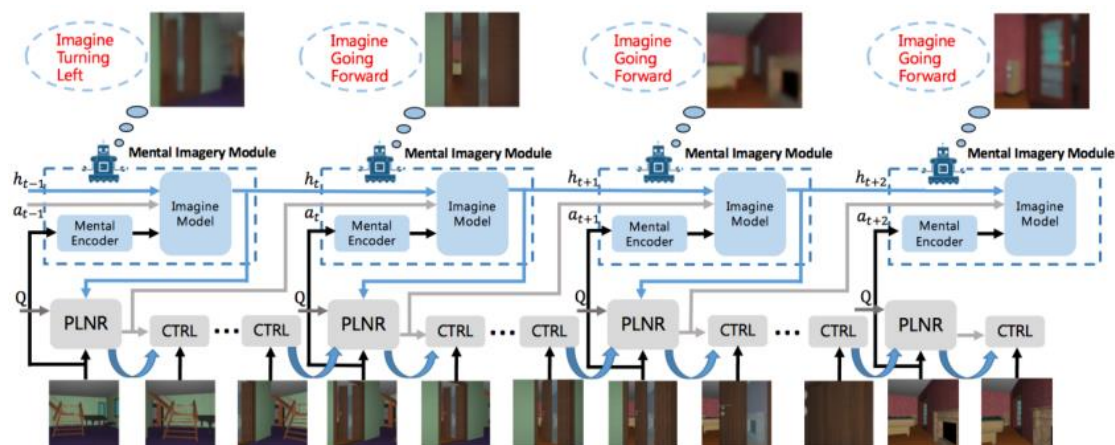
Model	Decoding	Train	Valid	Test		Train	Valid	Test	#steps
Random		6.25	6.25	6.25	Random	18.75	18.75	18.75	-
Human utterances		23.46	15.56	16.17	Human	76.74	76.74	76.74	15.05
Supervised	sampling	17.19	12.23	12.43	Best Cont.	89.44	86.35	88.33	34.47
	greedy	34.14	29.90	29.05	Best Disc.	86.23	82.81	87.08	34.83
	beam (size: 4)	26.21	22.53	25.02	Best NL	39.65	39.68	50.00	39.14
Policy Grad.	sampling	29.67	26.93	27.05					
	greedy	29.23	27.62	27.30					

结果显示, 最好的沟通模型 (emergent communication model) 的准确率达到 了 70%, 优于从人类话语沟通中的结果——大约 20%, 这表明人类很不善于定位, 可能是由于人类并不总能很好地传达自己的观察和行动。而从沟通模型中找到位置的最好的定位模型 (continuous communication, with MASC), 在整个导航任务测试中也有着 88.33% 准确率的表现, 同样超过了人类 76.74% 的准确率。

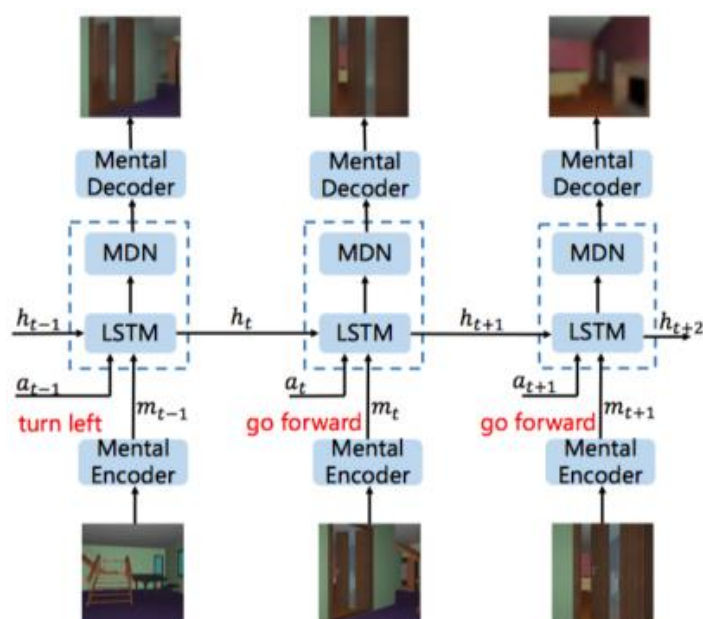
个人的思考:

(1) 针对世界模型的想法

从上面的综述中我们可以看到: 最近几年的无地图跨媒体导航任务中基于模型的强化学习方法逐渐成为主流, 其核心思想就是对世界建模, 但完全对周边环境建模是不现实的, 因为要涉及到一些物理的约束, 比较难以实现。因此我们可以换一个思路——只对人脑进行建模, 建立一个“心像”模块以模拟人类在做某件工作前需要静下心来思考怎样进行这一脑部活动, 这个模块跟 UCSB 的《Look Before You Leap》的思想类似, 但不局限于那篇文章中一个简简单单的函数。我们可以这样做:



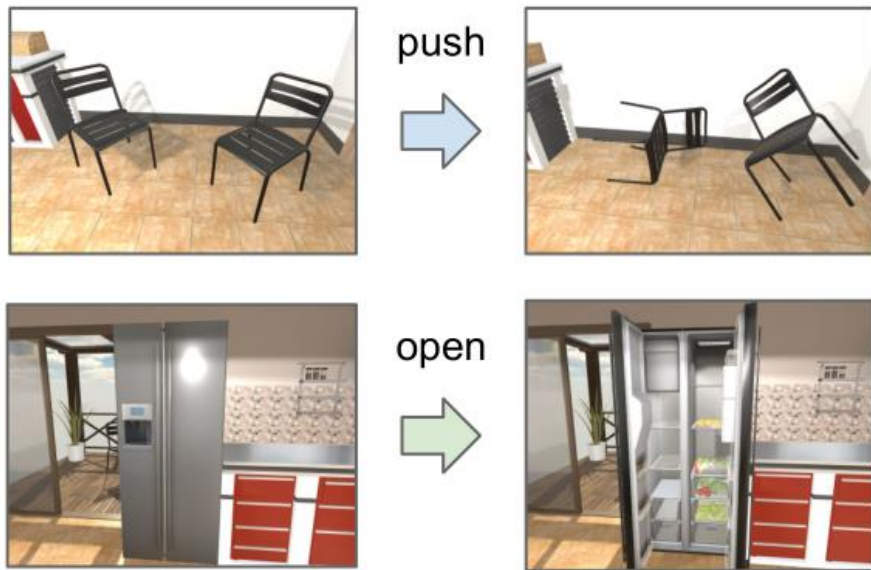
下面是 Mental Imagery Module 的结构：



解码后可可视化的图像就是当前状态下做出某种动作后转换到下一帧的状态。

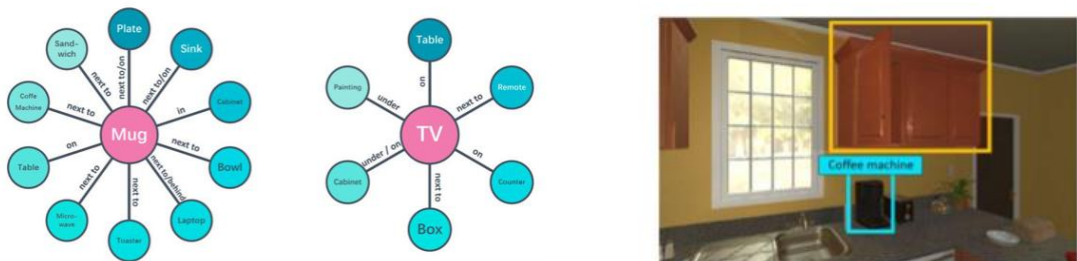
(2) 针对可交互的周围环境与知识图谱和推理的想法

在之前的综述中，无论是室内的导航任务还是室外的导航任务，其场景模型都是不可以变动的，只允许类比于人的 agent 在里面自由活动、变换视角。但是这一切在一个仿真器——AI2-THOR 的诞生后发生了变化。AI2-THOR 由论文《Target-driven Visual Navigation in Indoor Scenes using Deep Reinforcement Learning》首次提出，论文《IQA: Visual Question Answering in Interactive Environments》、《Visual Semantic Navigation using Scene Priors》相继使用。其是由艾伦人工智能研究所（AI2）、斯坦福大学、卡耐基梅隆大学、华盛顿大学、南加州大学合作完成的，它为人工智能 agent 提供了一个高度仿真的室内装修效果图画风的世界，在这个世界中 agent 可以和里面的各种家具家电进行交互——比如说打开冰箱、推倒椅子、把电脑放在桌子上以及关闭大厅吊灯等等：



为了让 agent 与场景的交互尽可能接近真实，AI2-THOR 除了包含表面上能看到的高质量 3D 场景之外，背后还有 Unity 3D 引擎，能让其中的物体遵循现实世界的物理规则来运动，也就是让交互动作尽可能真实。

论文《IQA》在这个场景建立一个类似于 VQA 的数据集，扩展了以往 VQA 只能根据不变的 image 或是 video 进行回答的限制，而论文《Visual Semantic Navigation using Scene Priors》则是纯粹的视觉语义导航，其任务的具体目标是，以一个场景的随机位置作为起点，仅依靠代理获得的 RGB 图像来找到指定类别物体的位置。当目标物体类别的一个实例出现在视野范围内，并且代理与物体实例的距离低于设定的阈值，便认为任务成功完成：



上图就是一个例子，当我们知道咖啡机的位置的时候，需要估计马克杯的位置。右图中黄色的方框标出了代理估计的马克杯位置，在咖啡机上方的壁橱，此处用到的场景先验是从相应的知识图谱中的到的，即马克杯是配合咖啡机使用的，所以马克杯一般放置于咖啡机附近。

根据上面三篇论文的思想，再结合跨媒体导航任务，我觉得可以对知识图谱进行修改。比如添加指令：“我要喝咖啡”，知识图谱中该事件包括咖啡机、咖啡豆以及马克杯，或者也可以在现有的知识图谱上加一些对话的内容进去，比如代理需要询问指令“我要喝咖啡”需要什么物品，当人类告知后再次进行导航任务，与环境进行交互、对三件物品逐一进行“Pick”操作，从而完美实现指令的要求，服务人类。

当然，针对家庭服务型机器人，我们也可以考虑多智能体协同工作的情况，这个思路我还未想到好的应用场景，就不介绍了。

参考文献:

- [1] Hermann K M , Hill F , Green S , et al. Grounded Language Learning in a Simulated 3D World[J]. 2017.
- [2] Wang X , Xiong W , Wang H , et al. Look Before You Leap: Bridging Model-Free and Model-Based Reinforcement Learning for Planned-Ahead Vision-and-Language Navigation[J]. 2018.
- [3] Fried D , Hu R , Cirik V , et al. Speaker-Follower Models for Vision-and-Language Navigation[J]. 2018.
- [4] Wang X , Huang Q , Celikyilmaz A , et al. Reinforced Cross-Modal Matching and Self-Supervised Imitation Learning for Vision-Language Navigation[J]. 2018.
- [5] Das A , Datta S , Gkioxari G , et al. Embodied Question Answering[J]. 2017.
- [6] Mirowski P , Grimes M K , Malinowski M , et al. Learning to Navigate in Cities Without a Map[J]. 2018.
- [7] De Vries H , Shuster K , Batra D , et al. Talk the Walk: Navigating New York City through Grounded Dialogue[J]. 2018.
- [8] Gordon D , Kembhavi A , Rastegari M , et al. IQA: Visual Question Answering in Interactive Environments[J]. 2017.
- [9] Zhu Y , Mottaghi R , Kolve E , et al. Target-driven Visual Navigation in Indoor Scenes using Deep Reinforcement Learning[J]. 2016.
- [10] Yang W , Wang X , Farhadi A , et al. Visual Semantic Navigation using Scene Priors[J]. 2018.