AV 对多层抽象的需求

M Minsky 在他的 The Emotion Machine 一书里认为[1]，人之所以有智慧，是因为人类具备对事物进行多层抽象的能力。当在某层具象思维里找不到答案的时候，人脑就会提高抽象级别，在另外一个概括层面上换一个角度去思考问题，他认为这种变通能力（resourcefulness）是人类智慧所独有的。Minsky 总结的三个抽象层见图 1.

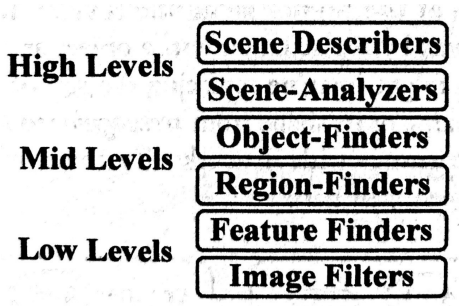

图 1

自动驾驶的表现与人类的行为之间目前存在很大的差异，其中原因之一就是二者工作在不同的认知抽象层面。智能机器一般在图 1 中的 Low Levels 和 Mid Levels，而人类则是从 Hight Levels 开始寻求答案，因为高级抽象思维速度最快，当找不到答案是才下探一层分析细节。

以驾驶为例，人脑优先关注的是图 1 中所示的 Scene Describers 和 Scene- Analyzers，比如，我们只关心一个地标的大概轮廓，只关心一辆大卡车是不是跟我们太紧，而不会关心这个地标大楼有多少层，窗户有多少，不会关心这辆大卡车的尺寸和精确的速度。也就是说，高层抽象更关心对象的性质，以及事物之间的关系。如果说低层抽象类似于解析几何，那么高级抽象更像拓扑几何。

机器思维的方向是与人类相反的，起点是像素，然后逐步抽象出物体、区域，要想得到高层抽象结论就必须经历大规模运算，目前无法直接开始高层抽象运算，和人类的直觉思维不是同一种方法，见图 2.

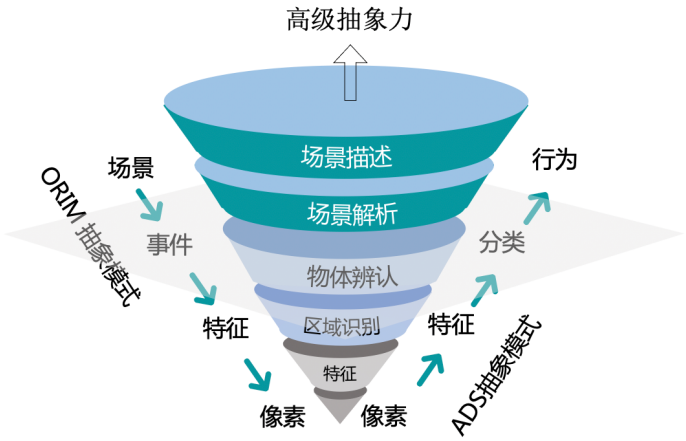

图 2 两种不同的抽象方向

目前 AV 思维与人脑相比，前者更注重战术层，按照 DIKW 智能等级分类，处于 Class 0 和 Class I。如果想进化到 Class II 和 Class III 级智能，则必须建立 Social Relation Model 和 Inference Future Model [2]。

Table 1 - DIKW wisdom hierarchy, the four-class consciousness model and AV intelligence frameworks

DIKW Wisdom Hierarchy by Information science	Four-class Consciousness Model by [3]	AD Planning Policy Hierarchy
Data：Observed symbols that represent properties of objects, events and their environments.	**Class 0_ Plants**: Physical sensing	**Feedback Model**: Streaming content generated by sensors such as cameras, LiDAR, radar, ultrasonics, GPS and dynamic HD maps.
Information：Revealing relationships in the data to find an answer to "who", "what", "when", "where", or "how many" questions.	**Class I_ Reptile**: Space consciousness	**Time-space Model**: Identifying measures in terms of object classification, object positioning, object kinematics, and infrastructure property detection; collision prediction; etc.
Knowledge: Actionable information being able to answer "how" questions.	**Class II_ Mammal**: Social relations consciousness	**Social Relation Model**: Object behavior prediction; maneuver planning.
Wisdom: Principles and values are determined and "why" question is answered.	**Class III_ Human**: Future consciousness	**Inference Future Model**: Solution creation; alternatives evaluation; driving environment envision on a scenario level.

图 3 [2]

自动驾驶的安全保障就像人类的战争或者一个公司的管理，需要有战术层和战略层的配合才能高效完成任务。但是目前 AV 更注重战术应对，把安全当作一种确定性过程，以交通参与个体为分析对象，而没有以场景为对象的分析手段，缺少系统性的战略判断与决策。

场景判断的观察和分析对象是场景，是场景要素之间的关联性质，而不是场景要素的物理特性，也就是说抽象层固定在“关系”上。所以，场景的描述、表达、定义、观察、传感方法与实体感知有所不同，需要重新开发。

场景分析有两种方法。第一种是根据经验教训对场景进行风险识别，比如，后面的大型货车跟随太近会有危险、行驶在其他车辆的盲区里会有危险、靠站的公交车后面会突然跑出行人、皮球后面会有小孩追逐，还有一些交互行为的礼仪要求、舒适性要求、社会接受度要求，等等，也就是所谓的“防御驾驶”规则，必须要找到技术方法将这些规则写进 AV 并且能够执行。应当注意，如果采用端到端技术方案，要想让其通过学习去掌握一本防御驾驶手册是非常困难的，因为端到端学习有一个“跷跷板”效应，也就是不会保证一个永恒的 ground truth 内容不变。

在 SOTIF 的四个象限中，第一种方法解决的是第 II 象限“显示规则”的问题（图 4）。“防御驾驶”利用“举一反三”的正向推理方法推知在学习训练里未出现过的、但是具有威胁性的新场景，预见“known”类危险的各种发生方式，避免以大海捞针方式显露“长尾”问题。

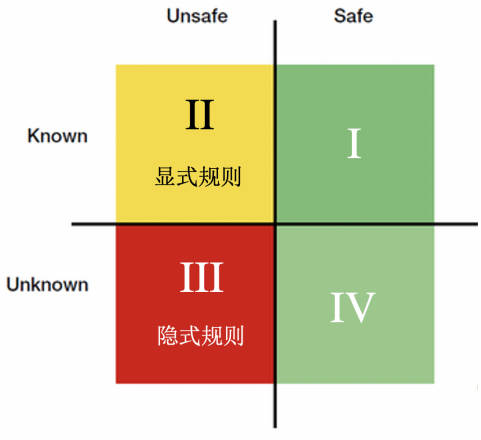

图 4

第二种方法是根据实际发生的事故对风险进行提炼总结。这类风险的背后原因人们并不一定很清楚，但是根据经验记录，某些场景元素的组合确实会导致危险发生，这就好比“此处事故多发”的警示牌，虽然路过的司机并不关心以前具体发生过什么样的事故，但是此时一定会额外谨慎驾驶。如果纵观百万、千万次真实事故，总结事发时的场景要素组合，那么就相当于把动态的“此景事故多发”（注意：不是“此处”）警示牌随时给司机或者 AV 系统发出预警，进而避免类似的历史悲剧重演。

在 SOTIF 的四个象限中，第二种方法解决的是第 III 象限“隐式规则”的问题（图 4）。：既然不知道有危险（Unknown），就说明这是一种人类认知不可理解的现象，无法通过推理分析正向总结其规律性，也无法描述和定义，而只能通过对事实结果利用归纳方法进行“举三反一”性质的逆向推测。“风险快照”事故模型[3]正是可以通过观察大量实发事故揭示不可理喻危险成因的工具，使 unknown 转化成 known。

参考文献

[1] Marvin Minsky, The Emotion Machine- Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind, Simon & Schuster, 2006

[2] S Qiu et al, Planning Automated Driving with Accident Experience Referencing and Common-sense Inferencing, https://doi.org/10.48550/arXiv.2301.10892

[3] S Qiu, 根据 NHTSA 数据训练的事故预测模型