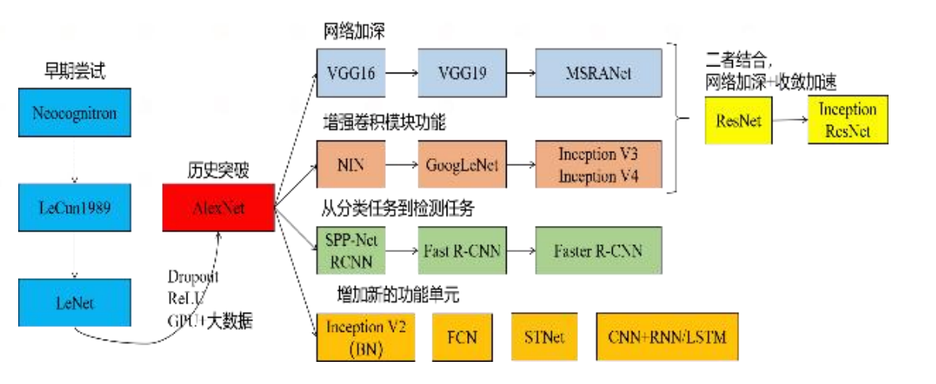
基于关键场景的智能算法安全边界评估方法

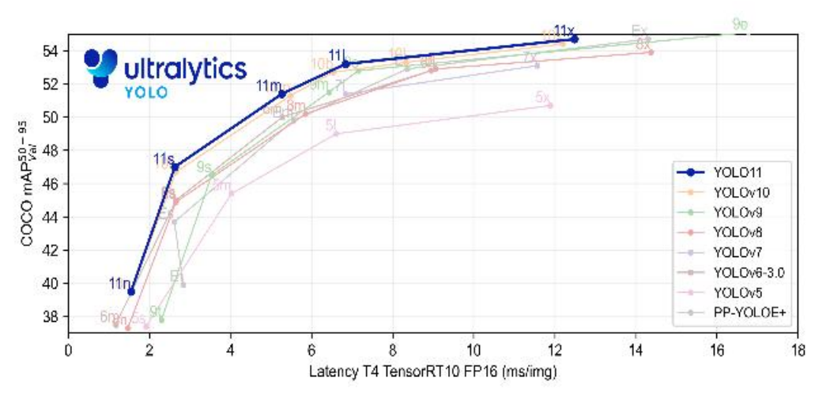
董乾，薛云志，孟令中，陈贺，杨光，王鹏淇，任红萍，龚优迪，李瑞
联系方式：董乾，dongqian@iscas.ac.cn

研究背景

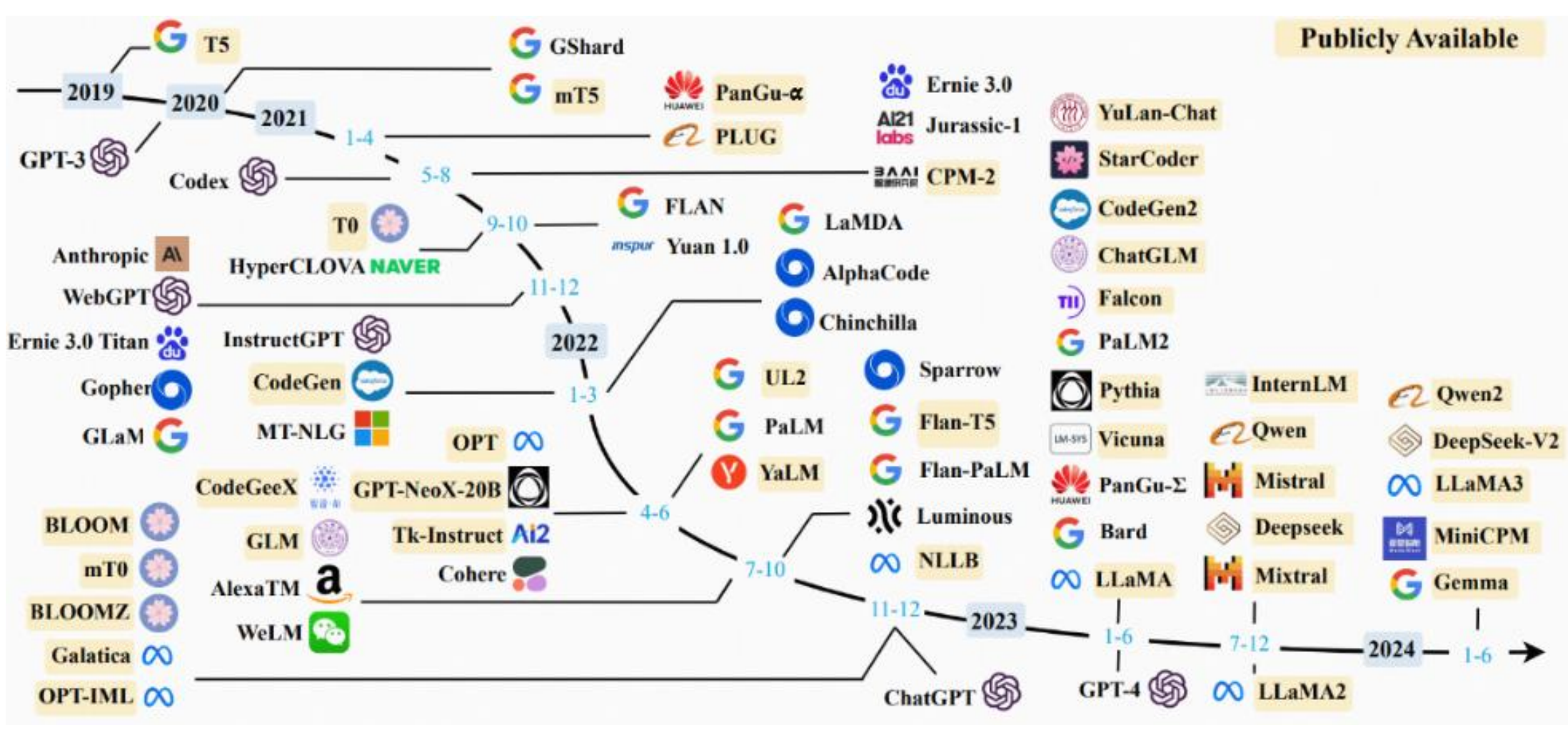
智能算法处于新一波的飞速发展阶段

- 智能算法种类显著增加，性能快速优化


神经网络种类显著增加



算法性能快速优化



大模型爆发

- 应用落地前景广阔，自动驾驶、机器人、大模型等

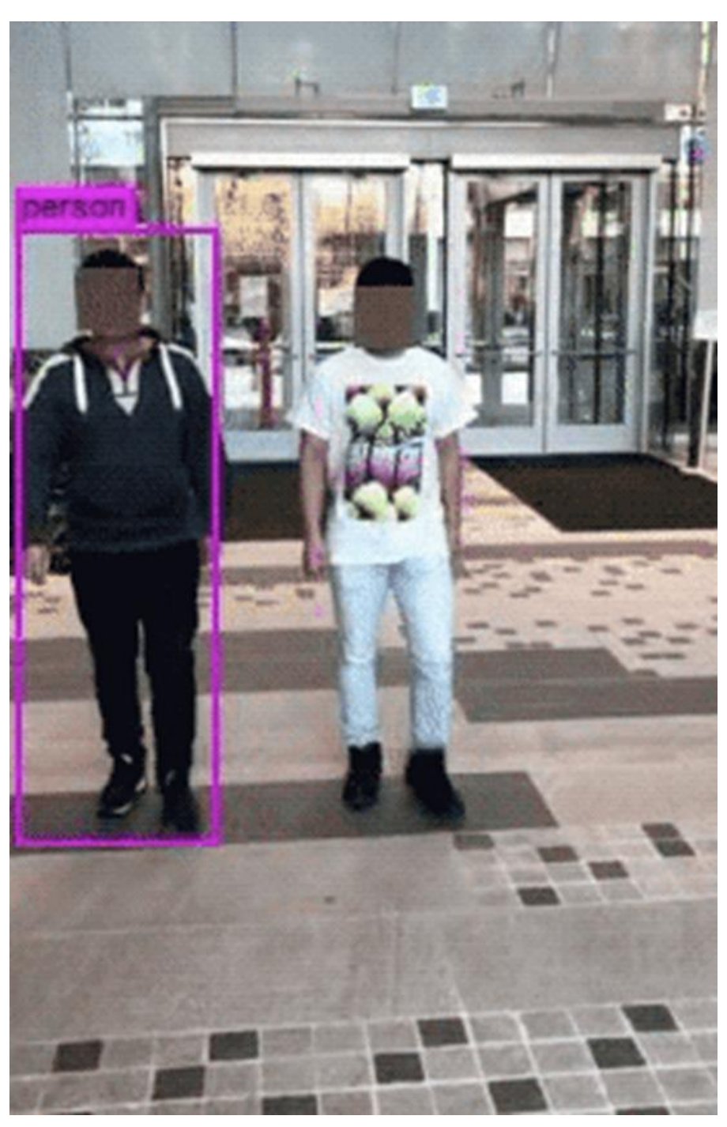

自动驾驶


机器人

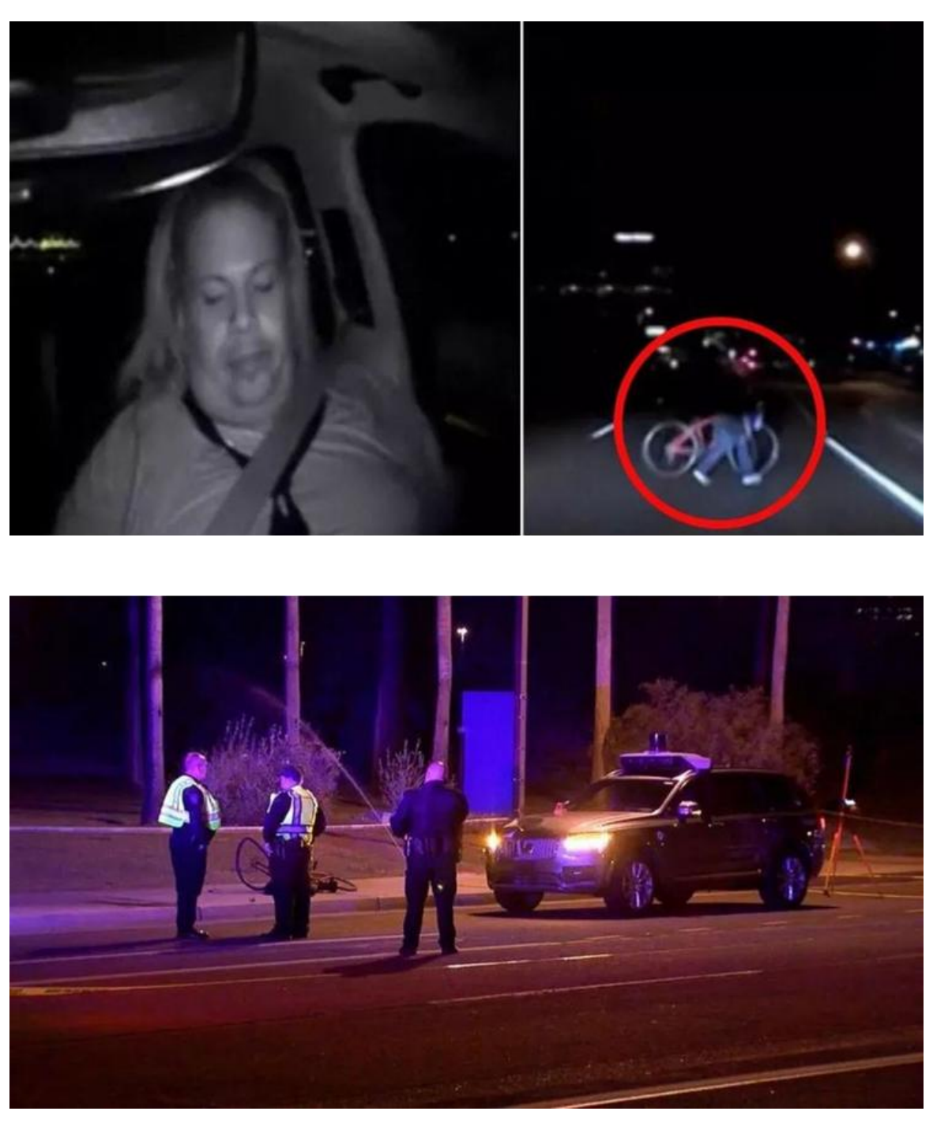

智能制造

大模型

目前智能算法存在较大安全风险和隐患，需要重点关注安全边界评估


目标检测漏检失效

- 对抗攻击方法欺骗AI。白色T恤上不同颜色的块干扰智能识别算法，识别错误。


Uber撞人事件

- 自动驾驶汽车撞上推着自行车行人。车辆自动驾驶系统由于设计缺陷错误分类行人，延迟了反应时间，导致发生碰撞。


理想L9识别失效

- 理想L9辅助驾驶系统，将广告牌误识别为汽车，紧急制动功能突然启动。

总体流程

智能算法应用场景构建 → 关键场景识别 → 评估指标体系构建与指标计算 → 安全边界划分与波动性分析

关键技术

场景形式化描述技术

- 场景中实体描述，实体名称、属性、类别等。

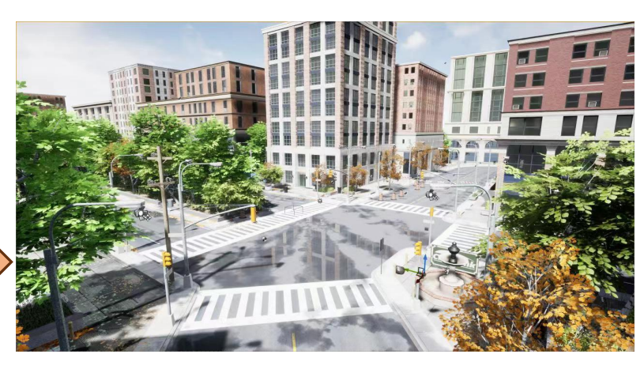
```
<xsd:complexType name="Vehicle">
    <xsd:all>
        <xsd:element name="ParameterDeclarations" type="ParameterDeclarations" minOccurs="0"/>
        <xsd:element name="BoundingBox" type="BoundingBox"/>
        <xsd:element name="Performance" type="Performance"/>
        <xsd:element name="Axles" type="Axles"/>
        <xsd:element name="Properties" type="Properties"/>
        <xsd:element name="Sensors" type="Sensors"/>
        <xsd:element name="Loads" type="Loads"/>
    </xsd:all>
    <xsd:attribute name="name" type="String" use="required"/>
    <xsd:attribute name="vehicleCategory" type="VehicleCategory" use="required"/>
</xsd:complexType>
```

车辆实体

- 场景中环境描述，天气、时间、道路等。

```
<xsd:complexType name="Environment">
    <xsd:all>
        <xsd:element name="ParameterDeclarations" type="ParameterDeclarations" minOccurs="0"/>
        <xsd:element name="TimeOfDay" type="TimeOfDay" minOccurs="0"/>
        <xsd:element name="Weather" type="Weather" minOccurs="0"/>
        <xsd:element name="RoadCondition" type="RoadCondition" minOccurs="0"/>
    </xsd:all>
    <xsd:attribute name="name" type="String" use="required"/>
</xsd:complexType>
```

城市环境



- 场景中动态行为描述，行为、触发条件、时序逻辑等。

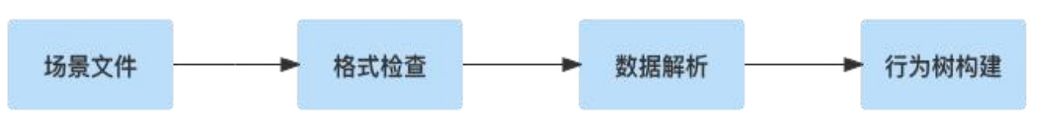
```
<ManeuverGroup  name     = "c1ManeuverGroup"
                maximumExecutionCount    = "1">
    <Actors  selectTriggeringEntities     = "false">
        <EntityRef  entityRef    =    "c1"/>
    </Actors>
    <CatalogReference   catalogName      = "Overtake"
                        entryName    = "OvertakeEgoVehicle">
        <ParameterAssignments>
            <ParameterAssignment parameterRef   = "OvertakingVehicle"
                                 value = "c1"/>
        </ParameterAssignments>
    </CatalogReference>
</ManeuverGroup>
```

超车行为

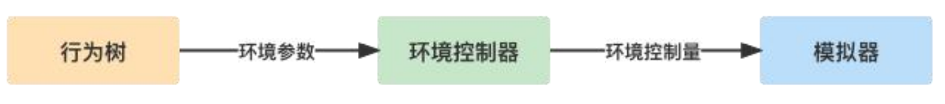
基于行为树机制的场景动态控制技术

场景控制行为树构建与状态控制

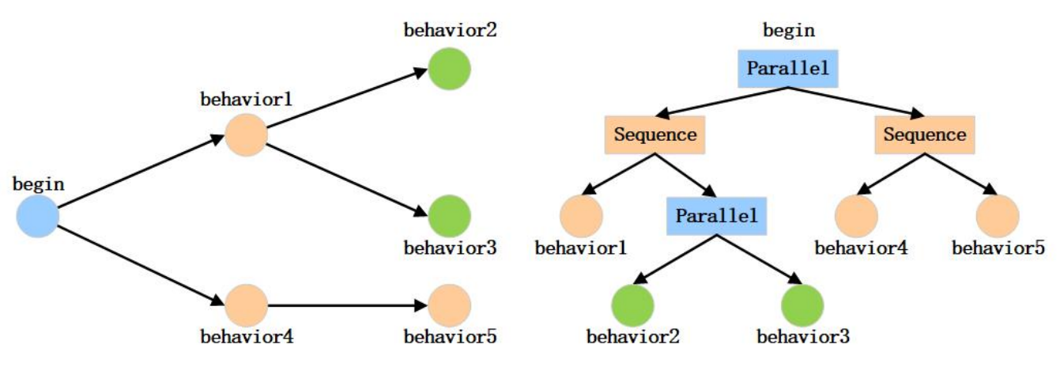
- 节点状态：成功、失败、正在运行；
- 节点执行时序逻辑：并行、串行、循环。



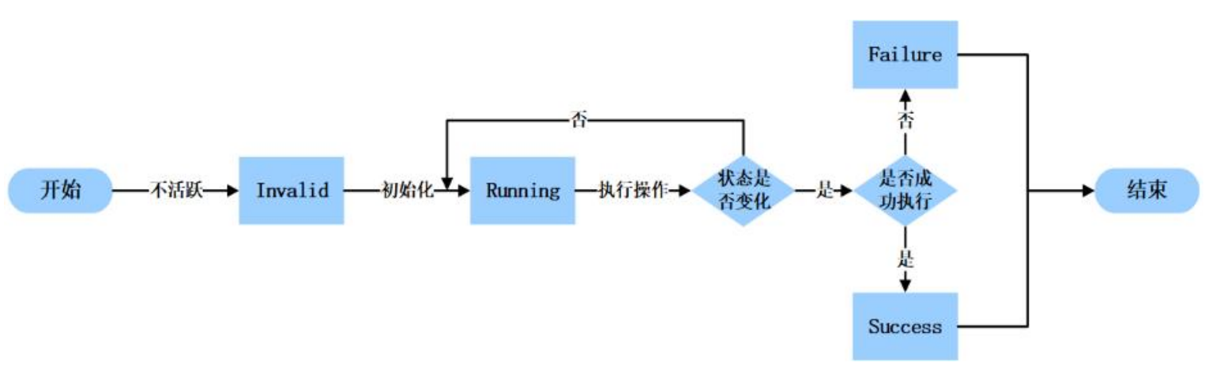
行为树构建流程



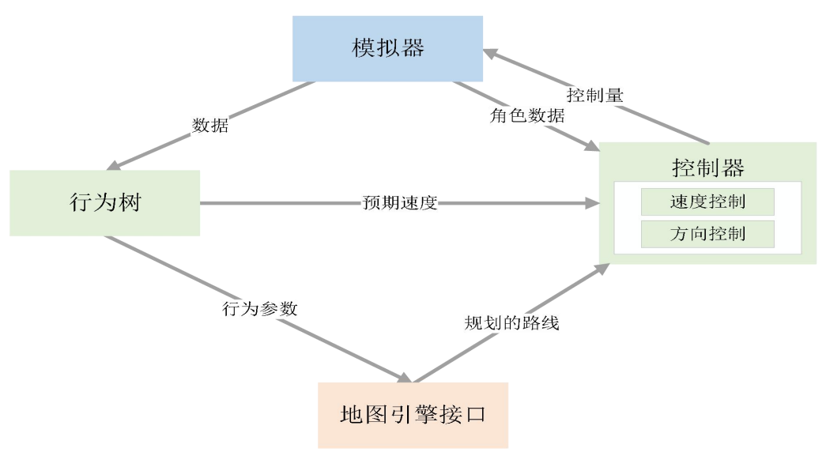
环境状态相关原子行为机制



将故事板中的时序关系转为行为树示例



行为的状态控制流程

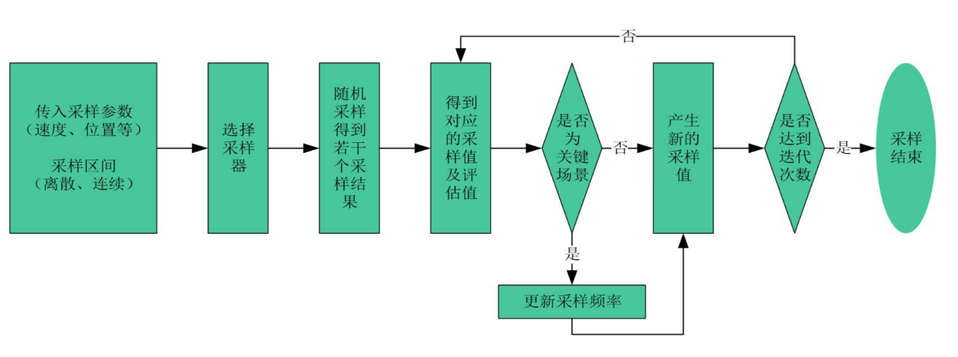


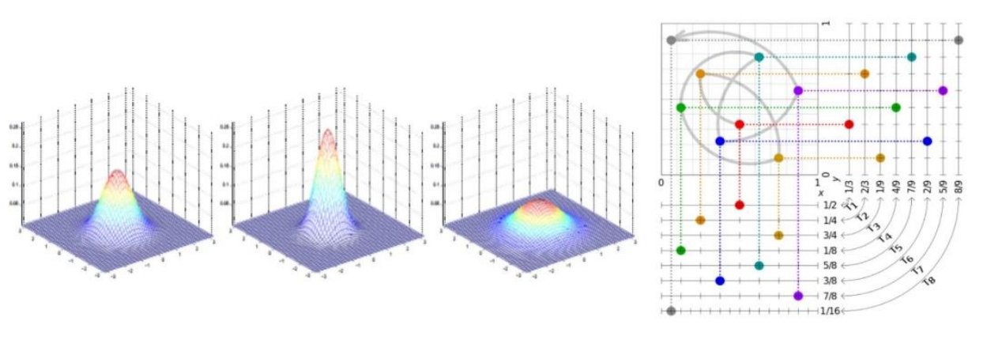
物体运动相关的原子行为机制

关键场景识别技术

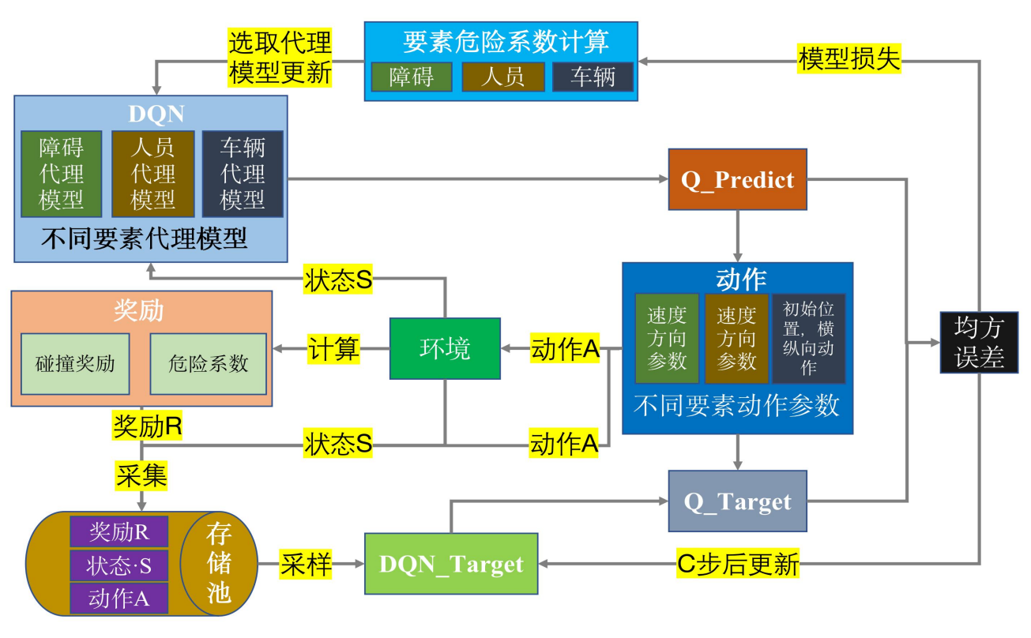
基于采样的关键场景识别

- 基于Gaussion、Halton等采样器对场景参数空间进行搜索，快速搜索到关键场景。




Gaussion采样　　Halton采样器

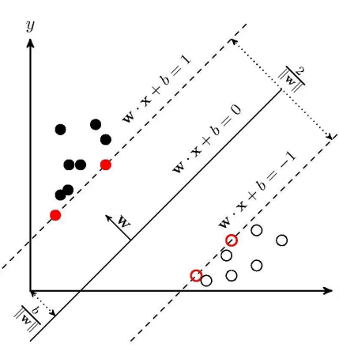
基于强化学习的的关键场景识别



识别方法	准确性	效率	多样性	可控性
采样算法	良好	一般	良好	良好
强化学习	优秀	优秀	优秀	良好

基于机器学习的智能算法安全边界划分技术

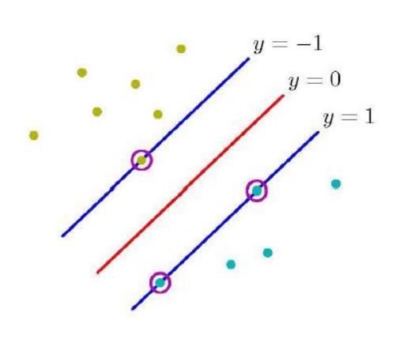
核函数



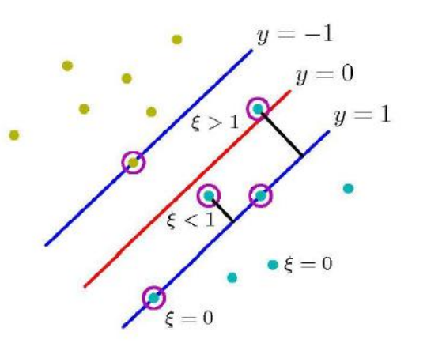
支持向量机示意图

常见核函数

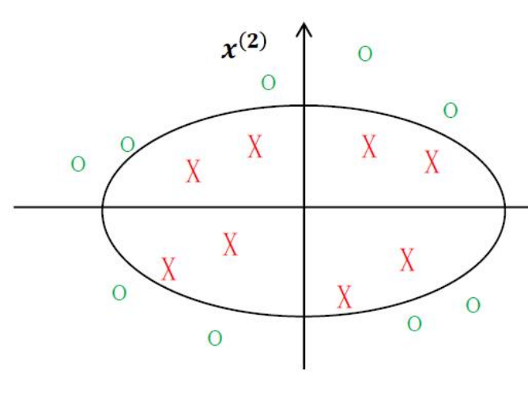
离群点处理



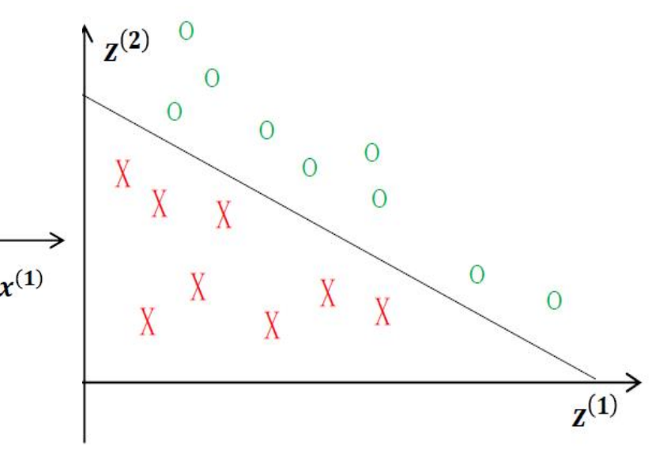
硬间隔



软间隔



线性不可分支持向量机



线性可分支持向量机

特点

- **高维数据处理**：映射到高维空间，有效处理复杂数据。
- **泛化能力强**：通过最大化间隔减少过拟合风险。
- **小样本适用**：仅依赖关键的支持向量，适合少量数据。
- **非线性问题**：使用核函数将非线性转为线性问题。

受中国科学院青年创新促进会资助