Vehicle Energy Optimization Systems (VEOS) Phase One 车辆能源优化系统
==========

- Goal
- State Definition
- Action Definition

总览

1. 将优化表述为可解或近似可解的问题（**SARP**γ模型）
    - 通用的可以学习的问题
    - 在线的优化过程
    - 尽可能简化的模型：通用性好，问题可解性好
        - 实际自动驾驶/能源优化系统是复杂问题
        - 取得简化问题描述和不遗漏重要因素之间的平衡
2. 理解问题：
    - 行动价值函数属性：
        - 形状，连续性，奇异性
        - 行动分辨率，行动周期，观测周期
    - 定义奖励，把**目标传达给智能体**：
        - 长期奖励，即时奖励之间的关系
        - 奖励的传播情况
    - 系统状态的轨迹和奖励之间的关系
    - 奖励稀疏的情况下是否定义多目标奖励加速学习
        - 引导系统学习
        - 从失败中学习
        - Off Policy学习
    - 利用专家知识引导智能体
    - 专家知识与认知偏差的关系
        - 专家知识利用特征和线性组合系数
        - 使用预定义的超参数，导致认知偏差(归纳Inductive Bias)
        - 归纳偏差提高训练效率,减少样本需求量
    - 神经网络架构的选择
        - 价值函数的网络可能不同于目标检测的网络

            - 基于Actor-Critic方法共享骨干网
            - 满足价值网与策略网兼容性定理
            - 计算量差异
3. 建立仿真
    - 通过仿真建立基本智能体模型
    - 产生学习用的数据
    - 通过Self-Play改进性能（节能，省时）
        - 将单智能体的性能优化表述为竞争环境下的多智能体演化问题
        - 两车或多车的竞速问题
        - 纳什均衡
    - 多车协作的策略
    - 分级策略
        - 构造子策略
        - 能源管理系统:调节SOC特性/油门特性
        - 车道保持系统:调节车道跟踪状态/道路边沿跟踪
    - 安全策略
    - 探索策略
        - 智能体不依赖专家知识引导，自主发现最优策略
4. 迁移学习

电动车运行模式

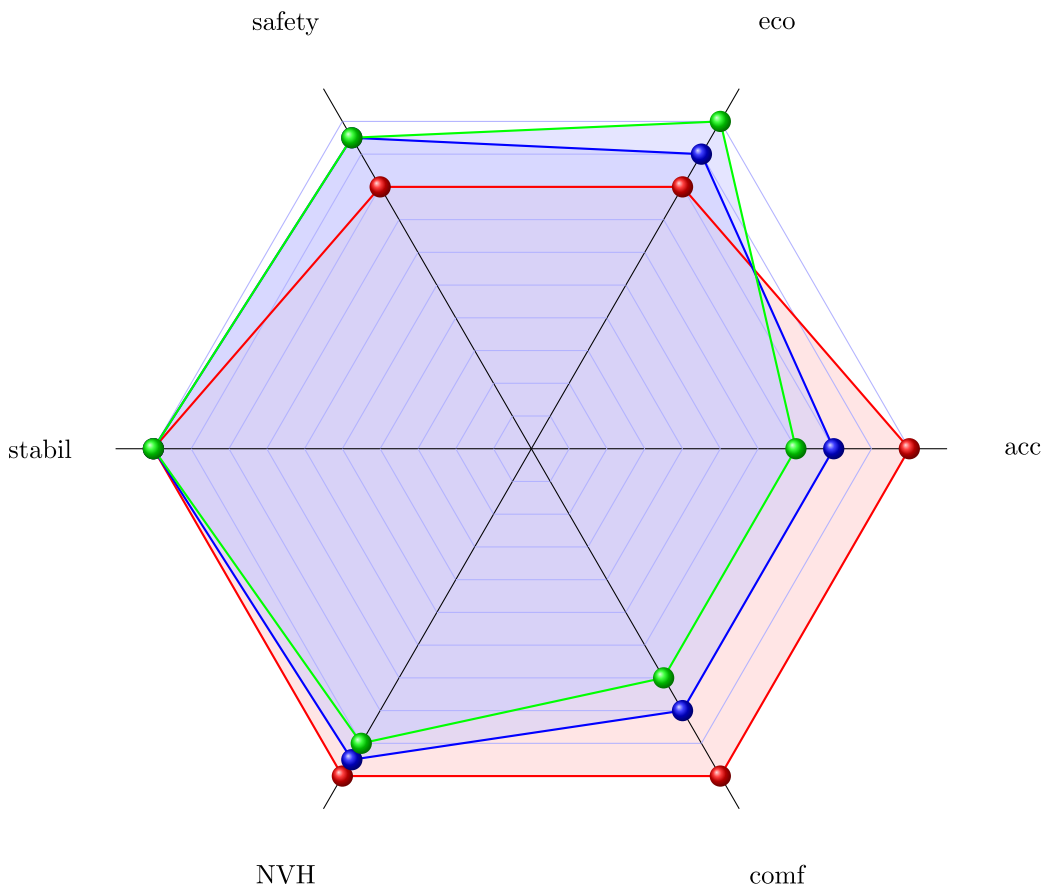
可调节参数:
限制最大扭矩,最大功率,最高转速,热管理,VCU标定,电池SOC标定

```
\documentclass[margin=10pt]{standalone}
\usepackage{tkz-kiviat}
\begin{document}

\begin{tikzpicture}[label distance=.15cm]
\tkzKiviatDiagram[radial style/.style ={-},
lattice style/.style ={blue!30}]%
{acc,eco,safety,stabil,NVH,comf}
\tkzKiviatLine[thick,color=red,
mark=ball,
ball color=red,
mark size=4pt,
fill=red!20](10,8,8,10,10,10)
\tkzKiviatLine[thick,color=blue,mark=ball,
ball color=blue,
mark size=4pt,
fill=blue!20,
opacity=.5](8,9,9.5,10,9.5,8)
\tkzKiviatLine[thick,color=green,mark=ball,
ball color=green,
mark size=4pt,
fill=blue!20,
opacity=.5](7,10,9.5,10,9,7)


\end{tikzpicture}

\end{document}
```



第一阶段目标

第一阶段目标是通过动态调整VCU标定参数改变电机工况，以达到降低能耗的目标。

通过形式化描述为强化学习，把优化过程转化成实时在线学习模式。

完整SARPγ模型

- **S**tate: 系统状态（车辆状态+道路状况）对观测的最小充分描述（马尔可夫决策过程,当前的状态描述），环境状态（真实环境的全知模型）
- **A**ction: 行动（加速度，刹车，转向，导航决策...)，**策略参数化**（初始化可以是随机选定）
- **R**eward：即时奖励（到达目的地+1，用时t，能耗e；可正可负，必须是一个标量，

可通过加权转换成标量）**通过采样得到即时奖励**

- **P**robability：（Transition Probabiltiy）状态迁移概率，描述系统动态 **通过采样得到下一个状态**
- γ：即时奖励的时间折扣系数

> 先做简化假设，只依赖车端状态，不依赖道路状态，先固定大部分标定参数，路线，路况，司机驾驶模式，只改变关键参数以验证学习效果，学习出一个**经济模式**和人工调试的进行比对。

最小化**SARP**γ 模型

- State 系统状态： **油门踏板开度（请求力矩），车速**; 电池SOC，能量回收模式，转向状态，
- Action 行动： **标定量TQD_TrqLeadSport_MAP**（初始化可以是随机选定）
- Reward 即时奖励: **瞬时能耗e/电机瞬时功率**
- Transition Probabiltiy 状态迁移概率:
- γ：即时奖励的时间折扣系数 **0.9/0.99/... (时间步长100ms, 1s, 1 min)**

状态空间

离散化

- 油门开度 $[throttle_{min}, throttle_{max}]$, [0:0.1:1]
- 车速 $[V_{min}, V_{max}]$:[0kmh:5kmh:100kmh]

行动空间

离散化

表格Tab[throttle,V]=T; $T \in [T_{min}, T_{max}]$: [0:5 NM:1000 NM]

即时奖励

瞬时功率/力矩/电机功耗+（行驶里程+最低加速度）

折扣系数

0.9，0.99, ...