

面向学习型自动驾驶决策的可解释人工智能方法

挑战3

分层式决策模型及解释

基于奖励机强化学习的
自动驾驶决策及解释
(第4章)

形式化交通规则

奖励机设计

Shapley值、反事实
解释

挑战1

挑战2

通用解释方法

改进的Shapley值归因解释方法
(第2章)

Shapley值基值
公理化描述

正定Shapley值
基值设计

正定Shapley值

Shapley值引导的反事实解释方法
(第3章)

反事实解释
问题构建

Shapley值引导
目标优化

挑战4

端到端决策模型及解释

基于密集卷积网络的端到
端自动驾驶决策及解释
(第5章)

端到端决策建模

语义Shapley值
归因解释

语义Shapley值引导
反事实解释

可解释自动驾驶决策实车实验 (第6章)

奖励机强化学习可解释自动驾驶决策验证

端到端可解释自动驾驶决策验证

平台支撑

自动驾驶平台搭建

人机系统设计