Artificial Intelligence 316 (2023) 103829

Contents lists available at ScienceDirect

Artificial Intelligence

journal homepage: www.elsevier.com/locate/artint

Reward (Mis)design for autonomous driving ☆

W. Bradley Knox [a,b,*], Alessandro Allievi [a,b], Holger Banzhaf [c], Felix Schmitt [d], Peter Stone [b,e]

[a] Robert Bosch LLC, United States of America
[b] The University of Texas at Austin, United States of America
[c] Robert Bosch GmbH, Germany
[d] Bosch Center for Artificial Intelligence, Germany
[e] Sony AI, United States of America

ARTICLE INFO

Article history:
Received 10 January 2022
Received in revised form 30 September 2022
Accepted 24 November 2022
Available online 13 December 2022

Keywords:
Reinforcement learning
Reward design
Utility
Cost
Safety
Risk
Autonomous driving

ABSTRACT

This article considers the problem of diagnosing certain common errors in reward design. Its insights are also applicable to the design of cost functions and performance metrics more generally. To diagnose common errors, we develop 8 simple sanity checks for identifying flaws in reward functions. We survey research that is published in top-tier venues and focuses on reinforcement learning (RL) for autonomous driving (AD). Specifically, we closely examine the reported reward function in each publication and present these reward functions in a complete and standardized format in the appendix. Wherever we have sufficient information, we apply the 8 sanity checks to each surveyed reward function, revealing near-universal flaws in reward design for AD that might also exist pervasively across reward design for other tasks. Lastly, we explore promising directions that may aid the design of reward functions for AD in subsequent research, following a process of inquiry that can be adapted to other domains.

1. Introduction

Treatments of reinforcement learning often assume the reward function is given and fixed. However, in practice, the correct reward function for a sequential decision-making problem is rarely clear. Unfortunately, the process for designing a reward function (i.e., *reward design*)—despite its criticality in specifying the problem to be solved—is given scant attention in introductory texts.[1] For example, Sutton and Barto's standard text on reinforcement learning [45, pp. 53–54, 469] devotes merely 4 paragraphs to reward design in the absence of a known performance metric. Anecdotally, reward design is widely acknowledged as a difficult task, especially for people without considerable experience doing so. Further, Dulac-Arnold et al. [14] recently highlighted learning from "multi-objective or poorly specified reward functions" as a critical obstacle hampering the application of reinforcement learning to real-world problems. Additionally, the problem of reward design is highly related to the more general problem of designing performance metrics for optimization—whether manual or automated optimization—and is equivalent to designing cost functions for planning and control (Section 2), making a discussion

☆ This paper is part of the Special Issue: "Risk-aware Autonomous Systems: Theory and Practice".

* Corresponding author.
E-mail address: bradknox@cs.utexas.edu (W.B. Knox).

[1] Unless otherwise noted, any discussion herein of reward design focuses on the specification of the environmental reward, *before any shaping rewards are added*. We also focus by default on manual reward specification, which differs from inverse reinforcement learning and other methods for learning reward functions. However, we discuss the application of this work to such methods in Section 5.4.

https://doi.org/10.1016/j.artint.2022.103829
0004-3702/