Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 112 (2017) 61–70

Procedia Computer Science

www.elsevier.com/locate/procedia

21st International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

Optimal program for autonomous driving under Bentham- and Nash-type social welfare functions

Keita Kinjo[a,i], Takeshi Ebina[b]

[a]Faculty of Economics, Okinawa International University, 2-6-1 Ginowan, Ginowan City, Okinawa, 901-2701,Japan
[b]Institute of Social Sciences, Shinshu University,3-1-1, Asahi, Matsumoto City, Nagano, 390-8621,Japan

Abstract

The purpose of this paper is to formally define and solve ethical problems of how an artificial vehicle (AV) determines its driving behavior when there are some passengers in the AV and some pedestrians on a street. We construct a mathematical model introducing mainly two Bentham- and Nash-types social welfare functions, and derive optimal solutions. We show the optimal solutions are completely different depending on the functions and their parameters. Our contribution is that policymakers or managers of AVs can discuss the problem and determine an algorithm for autonomous driving by formalizing the situation and offering the optimal solutions.

© 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of KES International

Keywords: Autonomous vehicle, Social welfare function, Ethics of artificial intelligence, Utilitarianism, Economics

1. Introduction

Artificial intelligence (AI) and robots are playing a more significant role in many situations today, from nursing to autonomous driving to military uses. Amidst this trend, decision-making and behaviors of AI and robots sometimes face social dilemmas from an ethical perspective, which is becoming a problem.[1,2] In particular, autonomous driving is known to have ethical issues.[3] It is a significant area of concern, and discussions are taking place on specific issues

[i] Corresponding author. Tel.: (+81)-98-892-1111; fax: (+81)-98-893-3273
E-mail address: keita.kinjo@okiu.ac.jp

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of KES International
10.1016/j.procs.2017.08.024

and algorithms in this context.[4ii]

An example of a past study on autonomous driving is the decision-making problem investigated by Bonnefon et al. [5,6]. Based on an online survey, Bonnefon et al. analyzed an assessment of autonomous driving by people under three scenarios:

A) Consider a scenario in which a person crosses the road in front of an *Autonomous Vehicle* (*AV*). If the car continues to move forward, it will hit the person. On the other hand, if the car turns to avoid the person, it will crash into a wall, killing the passenger. How should the AV behave in this scenario? This is a problem in which the life of either the pedestrian or the passenger is prioritized over the other.

B) Next, consider a scenario in which many people are crossing the road, the automobile collides into a wall to avoid hitting the crowd, and the passenger dies. This is a problem in which the lives of either the crowd or the passenger is prioritized over the other.

C). Finally, consider a scenario in which many people are crossing the road in front of the car while it is driving autonomously. Turning the car to avoid the crowd results in running over another person. Such a scenario asks the question of whether the AV should drive straight or make a turn. This is a problem in which the lives of either the crowd or an individual are prioritized over the other upon comparing the lives of many against the life of an individual. Similar problems have been known for many years in the study of ethics, such as with the trolley problem.

These scenarios demonstrate the ethical problems that exist when automating cars using AI.[iii] Many studies have examined what choices people make under various ethical scenarios. However, few studies have discussed how the approaches to ethical problems should be defined formally, and how the optimal values vary in such formal settings. In particular, as the cases presented by Bonnefon et al. show, there is a strong need to solve problems under formal settings when ethical problems related to priorities in human damage are implemented in AI decision-making and in actual programs.

To formally define the ethical problems presented by Bonnefon et al., this study introduced social welfare functions used in economics as objective functions of AVs, and solved these functions. Finally, using the results, the impact of objective functions and changes in the parameters of control in AVs was discussed.

The problem raised by Bonnefon et al. is related to the problem of how social welfare should be defined. Several social welfare functions have been proposed in the field of economics. A social welfare function represents the sum of the welfare of individuals such as consumers, producers, and the government. Policymakers and social planners make decisions on policies to maximize this sum. The objective function for the policymakers is called a social welfare function. Some social welfare functions consider efficiency, while others consider equality. One way where economists might determine how best to balance competing objectives of efficiency and equality is to specify a social welfare function.[7,8] The details are discussed in the next section.

However, this study did not consider which function is preferred. This needs to be defined through surveys of the people, as studies by Bonnefon et al. have done. This is a matter of whether the passenger should choose upon purchasing the system, and whether it should be decided democratically upon designing the legal system.

This study found that optimal control varies by social welfare function in AVs. A major contribution of this study was the identification of an AV control for specific social welfare functions to enable the implementation of the program and a comparison of the results. Specifically, the study found that AV control under the Bentham welfare function (see the next section) becomes very extreme (boundary solutions), while the damages are halved by the control under the Nash welfare function (see the next section). The approach in this study proposed a roadmap to introducing the knowledge in welfare economics to autonomous driving and AI.

This paper is structured as follows. Section 2 introduces relevant studies, and model settings are defined in section 3. Optimal values are derived analytically per target function in section 4. Calculations are performed in section 5.

[ii] For example, Anderson [9,10] proposed a more specific algorithm to implement ethics in AI in general. There are systems such as "Jeremy" (based on utilitarianism) and W.D. (based on Ross's Moral Theory).

[iii] Similar problems arise in automation in marketing, in which utility for the people in business management is maximized. In marketing, segmentation is performed to maximize the utility for the target customers, and more effective actions are taken for specific customers. An example of this is a recommendation or visualization system. [11] For typical products, the problem of providing recommendations to specific customers for the purpose of efficiency does not occur. However, in cases where social assets are used (known as social marketing, for example, to promote medical exams), inequality arises when policies are specialized for certain active people. This is a problem that concerns the effective allocation of resources.

متن کامل مقاله ← دریافت فوری

ISIArticles
مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات