Computational Modelling of Goals

Junqing Chen
Yuanpei College
Peking University
2000017795@stu.pku.edu.cn

Abstract

Current artificial intelligence models severely suffer from the problems of poor generality. Deep neural networks trained on quantities of data can only handle a specific task, sometimes even under many constraints. The reason for this phenomenon lies in the fact that current models can not model expressive concepts like functionality, rationality goals and etc. In this article, we will mainly focus on how to model goals in a computational framework and we hope this will contribute to a more generalized artificial intelligence model.

1 Introduction

Goal, or intentionality is ubiquitous and important in our daily life and behavioral decisions. Almost all of our movements and behaviors are motivated by one or many goals, which can be either explicit or implicit, and either long-term or instant. For example, when we are driving, the destination is simply our goal and it guides us to plan driving route by distance and road conditions.

However, representing goals on computer is not as direct and natural as in physical world. Goals can be various, from positive ones to malicious ones, from physical ones to mental ones, from real ones to imaginary ones and more. It is hard to represent them all in a unified paradigm so that it can then be represented on computer.

In the following sections, we will analyze the concept of goal more detailedly and illustrate three possible ways to represent goals in computational ways, rule-based, Bayesian inference and inverse reinforcement learning.[1]

2 Rule-based modelling of goal

As mentioned in Introduction section, the main difficulty of goal representation is how to utilize a unified paradigm to describe arbitrary goals. In this section, we simply bypass this problem by modelling goals under explicit rules and constraints.

To better explain our ideas, consider the driving example first. In the driving scenario, we will have a destination goal g_1. This means that we want to be as close to the destination x_g as possible at any time $x_t = f(t)$, then we can turn this goal to an optimization problem as

$$\min_f \int \|x_g - f(t)\|_2 \mathrm{dt}$$

Possibly, we still have another goal, that we do not want our speed at any time $v_t = g(t)$ to exceed the speed limitv_l, we can be also represented as a naive optimization problem as

$$\min_g \|v_l - g(t)\|_2$$

More generally, we can represent this idea with following equation

$$\min_{f_1,\ldots,f_n} \sum_{i=1}^{n} \omega_i O_i(f(x_i), c_i(x_i))$$

where x_i means explicit metrics like speed, position and etc on which rules and constraints c_i like speed limit apply. O_i refers to a specific operator for the i-th metric. In above examples, the operators are integral and 2-norm respectively. w_i are weights for different metrics. For example, we may want to assign the position constraint a higher weight than speed limit.

3 Bayesian inference modelling of goal

The disadvantage of rule-based modelling is apparent. We must design specific operators O_i for different goals. And the constraints c_i have to be given explicitly, which are often implicitly integrated into agents' behaviors. For example, we can infer the speed limit by observing sequences of driving cars' speeds.

Therefore, we put forward a more general way to model goal with Bayesian inference. Specifically, we can represent our goal as

$$G = \arg\max_G P(A|G, O)$$

where A means action, O means observation, and G means our goal. It also seems to represent it as

$$G = \arg\max_G P(G|O, A)$$

The first mathmatical interpretation is based on the idea that agents' action is based on its goal and observation. The latter one is based on the idea that we can infer the agents' goal by acquiring its action and observations.

References

[1] Brian D Ziebart, Andrew L Maas, J Andrew Bagnell, Anind K Dey, et al. Maximum entropy inverse reinforcement learning. In *Aaai*, volume 8, pages 1433–1438. Chicago, IL, USA, 2008. 1