Discussion about How to Represent Goals

Chenhao Zhou
Yuanpei College
Peking University
zhouch@stu.pku.edu.cn

Abstract

The construction of intention is increasingly becoming a central position within human cognitive functioning. The judgments of intentions can provide the rationality of actions, helping with the understanding and memorizing of the previous action, and prediction about the future action. The representation of goals have become an essential component in current computational framework. This essay would involve the psychological researches on intentions, trying to give legitimacy to various method of goals perception. The main part contains the brief analysis about three typical methods and the their ways to represent goals.

1 Introduction

Intentions could be seen as a hidden target in mind, drives a series of goals, decisions, plans and so on. When we observe others in motion, we usually care little about the surface behaviors they exhibit. What matters are their underlying intentions. Abundant evidence in psychology research has shown that our humans are goal-directed agents, having ability to perceive, plan, make decisions, and achieve goals. The interpretation about actions via intentions is natural and common for humans. Six-month-old infants are capable of perceiving people's goals, and would become surprised when the further actions mismatch the underlying intentions. In addition, the understanding and memorizing of actions can be efficient by goal-driven processing, along with the simplification of complex action details: one latent intention can represent several highly dissimilar motion patterns.

Crucially, obtain a unified representation of goals is intuitively difficult for its variety and various degrees of abstraction. In addition, the representation of goals should consider some features, such as the short-term stability and long-term generalizability, enabling the achievement of the intended goal state with different actions across contexts; and the rationality of relevant actions, goal-state, and situation constraints [3]. Building blocks for intent in a computational agent would be discussed in the following sections, along with some psychological theories about intentions.

2 Psychological theories about intentions

In this section we take three theories proposed by several psychologists, which analyzes the underlying mechanism about intentions and actions.

Simulation theory Blakemore and Decety [2] reviews the psychophysical and functional neuroimaging evidence, then propose the simulation theory, which claims that the interpretation of actions via intentions might rely on simulating the observed action and mapping it onto our own experiences and intention representations.

Rationality principle Another possible mechanism has been proposed by several psychologists is the "rationality principle", which states that humans achieve their intentions rationally by maximizing their utility while minimizing their costs.

Teleological stance Humans tend to interpret and draw inferences about other's goal-directed actions. Based on evidence from psychological experiment, Gergely and Csibra [4] propose that one-year-olds may apply a non-mentalistic interpretational system called "teleological stance". They summarized the "teleological stance" as a naive theory of rational action that allows them to interpret actions in a variety of different contexts.

3 Typical ways to perceive goals

Since the construct of intention is being granted a central position within human cognition, researchers have proposed various approaches to embody intentions representation on AI. In this section we would discuss three typical ways in intentions inference to seek the appropriate way to represent intentions given different contexts.

3.1 Inverse Planning

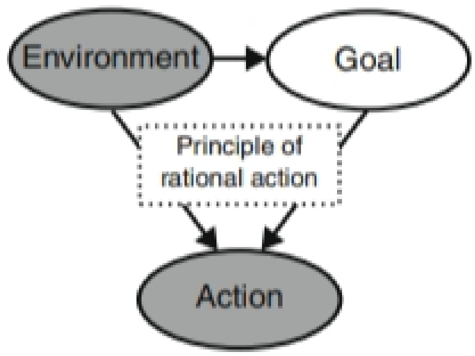

Figure 1: Diagram with causal graph notation, modelling infants' early-developing action understanding competency.

Incorporating planning into expressions of intentions is a natural idea, as plans always consist of implicit goals to achieve. Cognitive studies [4] have shown that humans are strongly inclined to interpret events as a series of goals driven by the intentions of agents. Based on this teleological stance theory which explains infants' early-developing action understanding competency, see Fig. 2, Baker et al. [1] formalize this intuitive theory as a Bayesian inverse planning framework.

Specifically, the sequence of rational actions given current environment should be inverted to infer the goal from an action sequence, thus generating an action sequence from the goal. The functional form of the causal relation between Environment, Goal and Action is given by rational probabilistic planning in Markov decision problems, and goal inference is performed by Bayesian inversion of this model of planning.

The inverse planning method demonstrates rational human action process, involving a conditional probabilistic inference of intentions, which allow us to assess how well and in what ways people's mental representations of the world correspond to reality.

3.2 Inverse reinforcement learning

Another intuitive view is modelling intention as a rational policy mapping several contexts to action. The underlying reward or cost function of person can be seen as a rational degree, which can be infer along with the policy from the demonstrated behaviors. Considering the future goal states that are far in terms of both space and time, Rhinehart and Kitani [5] proposed an algorithm to learn a decision-theoretic human activity model via an online Inverse Reinforcement Learning (IRL) technique.

In the design of markov decision process, a special type of state called goal state was involved to denote states where the person has achieved a goal. The goal can be seen as the place which is meaningful (i.e., kitchen, office) or simply the location where people stops. And the prediction of the final goal of a person's action sequence can be posed as solving for the MAP estimate of the posterior over goals. Thus in the algorithm, the jointly discover states, transitions, goals, and the reward function of the underlying Markov Decision Process model was discovered. This method successfully overcome changeable over the course of the action sequence, adjusting an online policy to predict the goals.

3.3 Grammar parser

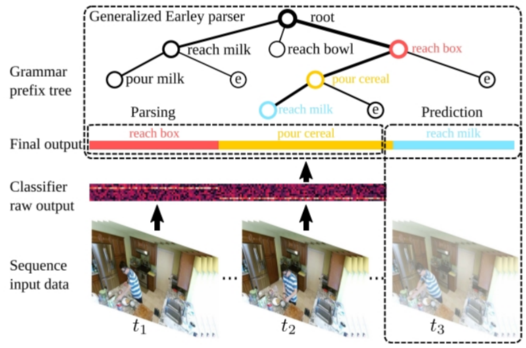

Figure 2: An overview of the parsing algorithm

In addition, a grammar parser could be used to acquire a graphical representation from a sequential motion input. And the intentions are captured from the plan which parse graph indicates. For example, a traditional grammar parser, the Earler parser was used in [6] to parse the sequence data, further making top-down future predictions.

Specifically, the generalized Earley parser segments and labels the sequence data into a label sentence in the language of a given grammar. The parsing process is expanding a grammar prefix tree and searching heuristically in this tree according to the input prefix probabilities. Then future predictions can be made based on the grammar prefix tree. This grammar parser involves the non-Markovian property of grammar to implicitly model the intentions' variety.

4 Conclusion

This essay gives a brief overview of three typical methods in intention detection and prediction. All of those methods should be chosen according to specific contexts and supposed to be accompanied by corresponding adjustments. A more general representation way of intention is urgently needed in research of cognitive AI.

References

[1] Chris L Baker, Rebecca Saxe, and Joshua B Tenenbaum. Action understanding as inverse planning. *Cognition*, 113(3):329–349, 2009. 2

[2] Sarah-Jayne Blakemore and Jean Decety. From the perception of action to the understanding of intention. *Nature reviews neuroscience*, 2(8), 2001. 1

[3] Gyorgy Gergely and Gergely Csibra. Teleological reasoning in infancy: the naive theory of rational action. *Trends in cognitive sciences*, 7(7):287–292, 2003. 1

[4] Gyorgy Gergely and Gergely Csibra. 'obsessed with goals': Functions and mechanisms of teleological interpretation of actions in humans. *Acta psychologica*, 124(1):60–78, 2007. 1, 2

[5] Nicholas Rhinehart and Kris M. Kitani. First-person activity forecasting with online inverse reinforcement learning. *International Conference on Computer Vision (ICCV)*, 2017. 2

[6] Qi Siyuan, Jia Baoxiong, Huang Siyuan, Wei Ping, and Zhu Song-Chun. A generalized earley parser for human activity parsing and prediction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(8):2538–2554, 2020. 3