The Framing Problem for Visual Representation Modules in Neuro-Symbolism Paradigm for Artificial Intelligence

Dingzhou Fei

Department of Psychology, Wuhan University, 430072, China

ABSTRACT

Research on the visual computation has received a great deal of attention The demands on human-based cognitive computing are very high. for example, the implementation of autonomous driving functions depends on the effective cooperation of three components: perception, decision making and execution. The perception layer identifies people, objects, and signs on the road by simulating the human eye through multi-dimensional sensors; the decision layer evaluates and makes decisions through pre-processing such as algorithm fusion and feature extraction; the data is fused and output to each control unit in the execution layer; and finally, the hardware mechanism in the execution layer makes feedback actions to realize the full set of autonomous driving operations. In this work, we try to show that this kind of symbolism framework will encounter the so-called "the framing problem" as earlier attempts to formalize the changes of knowledge in event flows. The framing problem is the problem of finding a sufficient set of axioms for a feasible description for environments of the robot or automatic driving. Even the visual-spatial hybrid computation with additional default inference has shown its potentiality in the automation, there still be the frame problem. This means that the symbolic methodology has its own logical flaw.

Keywords: automatic driving, visual computation, frame problem, logical pro-gramming

INTRODUCTION

Autonomous driving, also known as driver-less driving, computerized driving or wheeled mobile robotics, is a cutting-edge technology that relies on computer and artificial intelligence technology to complete, safe and efficient driving without human intervention. There are six levels of autonomous driving: from level in which a driver-operated car to the level in which cars will be no longer controlled by the driver. While deep learning methods are used in autonomous driving research, hybrid visual perception solutions based on a combination of vision and semantics have their own value. These values relate to interpretability, human-centered artificial intelligence and industrial standardization to suit ethical and legal requirements This work aim to investigate the potentiality of the paradigm (Aditya, S et al, 2015)

Research on the visual and spatial-temporal computation has received a great deal of attention. The demands on human-based cognitive computing are very high. for example, the implementation of autonomous driving functions depends on the effective cooperation of three components: perception, decision making and execution. The perception layer identifies people, objects, and signs on the road by simulating the human eye through multi-dimensional sensors; the decision layer evaluates and makes decisions through pre-processing such as algorithm fusion and feature extraction; the data is fused and output to each control unit in the execution layer; and finally, the hardware mechanism in the execution layer makes feedback actions to realize the full set of autonomous driving operations.

In latest paper "Commonsense visual sensemaking for autonomous driving on generalized neuromyotonic online abduction integrating vision and semantics", which appeared in journal "Artificial intelligence'' (*Vol.299,2021*) (Suchan, Jakob et al,2021), the authors claimed that it incorporates the latest technologies in visual computing and has been developed as a modular framework that can be generally used in hybrid architectures for real-time perception and control. It has been developed as a modular framework that can be generally used in hybrid architectures for real-time perception and control. This advancement represents a useful side of symbolism.

FRAME PROBLEM

The framing problem is naturally not a problem for connectionism, because the mathematical basis behind it is not formal logic. But among old-school AI, formal logic is still dominant, and implicit in the framing problem is also the flaw in old-school AI's use of symbolic computation as a model of the mind and perception. While end-to-end vision and control based on deep learning has (arguably) been successful for self-driving cars, integrating hybrid vision and semantic solutions at each step, there are clear needs in fulfilling the fundamental legal and ethical responsibilities involving interpretable, human-centered AI and industrialization.

The emergence of non-monotonic reasoning (NMR) in the 1980s benefited from the close correlation of temporal reasoning research in the field of action. The framework problem (McCarthy J. and Hayes, P, 1969) clearly stipulates all the influencing conditions of an action. It is logically infeasible to ignore it. It played a central role in the research of the NMR form representing the default mechanism, especially in the field of reasoning about actions and changes. Focusing on correctly capturing the law of inertia, this is a dynamic default reasoning that can be expressed as "unless there is evidence to the contrary, the value of fluency will not change." NMR is also helpful for dealing with other typical characterization problems in action theory, such as ramification and qualification problems (Moore R. C,1985). Even the visual-spatial hybrid computation with additional default inference has shown its potentiality in the automation, there still be the frame problem. This means that the symbolic methodology has its own logical flaw.

In fact, the framing problem describes the problem of expressing facts about the world using first-order logic (FOL). Representing the state of a robot or an automatic driving in traditional FOL requires the use of several axioms that simply imply that things in the environment do not change arbitrarily. For example, we describe a 'block world' in terms of rules about stacking blocks on top of each other. In the FOL system, other axioms need to be used to infer the environment (for example, blocks cannot change position unless they are moved).

EXPRESSING VISUAL SEMANTICS FOR AUTOMATIC DRIVING USING LOGICAL PROGRAMMING

In this section, we use logical programming to express high level actions and knowledge without formal details due to page limitation for this paper. The language for the expressions will be intuitive and semantic as possible as we can.

Visual perception is generic and modular, at the high-level computational level of 'common sense and semantics, while being combined with low-level deep learning methods capable of computing raw features in visual data. The paper [2] demonstrates the importance of semantics-driven approaches rooted in knowledge representation and reasoning (KR), and the implementation of such a combination, suggesting that this is a promising research approach.

The combination of logical methods with neural models is also a development of Marr's visual models in addressing research questions related to interpretable and human-centered artificial intelligence, especially from the perspective of the (perceptual) sensing of dynamic visual images (Marr, 2010).

This is a common scenario in driving: a car (c) is driving ahead and is implied to turn right; during this time, the person (p) is riding a bicycle (b), which is located overall to the right of c, and is moving forward. The car c turns to the right, during which time the person on the bicycle <p, b> is not visible. Subsequently after a while the person on the bicycle <p, b> appears again. This episode is lustrated as below **Fig.1**

Fig 1: An illustration example for the episode as described by above passage. A car and a cyclist went in different tracks and the car witnessed the cyclist going away from him and disappeared for a while, eventually and reappeared. The time stamps t represents distinct temporary points. **Source** see reference (Suchan, Jakob,2021)

This example in fact is program ***P*** and can be used to illustrate how to obtain the model about actions and cognition and changes for them. For example, the followings can be considered as necessary:

- For projection and interpolation of missing information, for example, when a cyclist <p, b> is occluded, what assumptions can be made about it? What assumptions can be made about the cyclist<p, b> in the case of being blocked; how can this assumption support the plan for the next action?
- Maintain the identity of objects on a semantic level, for example, in the presence of occlusion, missing, and noisy quantitative data.
  Errors in detection and tracking: the ability to make default assumptions, for ex ample, assumptions about persistent objects and/or object properties
- Maintain consistent beliefs and respect (domain-neutral) common-sense standards, such as constitutive and indirect influences, continuity of time and space, and position changes caused by movement
- Inferring/calculating counterfactuals, in a manner similar to human cognitive ability, mental simulation for introspection, human cognitive ability for the purpose of introspection, performing "hypothetical" reasoning tasks, etc.

The paper (Suchan, Jakob,2021) provides a representation framework based on logical programming, including a visually meaningful ontological framework and formal representation foundations, while at the same time focusing on common sense representation aspects related to spatial, spatial-temporal, and motion modelling, events, and other aspects related to spatial-temporal dynamic modelling and reasoning are also considered

Table 1 Formal presentation for the ontologies in **Fig 1,** is written by the logical programming (for example, situational calculi for expression actions and state changes). **Source**: see reference (Suchan, Jakob,2021).

As for the other high-level objects in motion at spatial -temporal scenes, like arbitrary rectangles, circles, polygons, and mereotopology, incidence, orientation, moving style, etc., can be defined using logical programming language. Here the formal specifications are not listed because of space restriction for this paper, too**.**

THE ASP FRAMEWORK CANNOT AVOID THE FRAME PROBLEM

Set Programming (ASP) is a declarative problem-solving paradigm, rooted in Logic Programming and Nonmonotonic Reasoning, which has been gaining increasing attention during the past years. But in recent time, it became a emerging paradigm in artificial intelligence, like automatic driving because it has interpretable control strategies in the deep learning paradigm (Lifschitz, V,2017).
A normal logic program has the following form:

```
Fact: A0
Rule: A0 :- L1, …, Ln
Integrity Constraint: :- L1, …, Ln
```

The proposition contained in `Fact` is unconditionally true. `fact` or the head A_0 of a rule is atom, which can be a constant or a function. the body L_j of a rule or integrity constraint is a literal with the form `A` or `not A`, representing positive and negative literal respectively, where `A` is also atom.

Interpretation of a rule: if all positive literals in the body of a rule are true and all negative literals are satisfied, then it follows that the head of the rule is true.

Interpretation of an integrity constraint: all literals in the body of an integrity constraint cannot be satisfied at the same time.

When all the atoms in a normal logic program satisfy all the facts, rules and integrity constraints, then the set of these atoms is called a model. in ASP, when each atom in a model can be derived acyclically, then the model is an answer set. If `A` is an atom, then `-A` denotes the complement of `A`, i.e., classical negation. Unlike `not A`, which is true when and only when `A` is false, it is only required that `A` and `-A` are not both true. Depending on the logic program, neither `-A` nor `A` may be present in the answer set, thus expressing a state of affairs where A is not known to be true or false. `-A` may appear in the header. Note: in ASP, the double negation i.e. `not not A`, `not not A` satisfies when and only when `A` satisfies. This can be used to solve a number of problems with cycles.

We will provide two answer sets or model for a program **P** (see an example in **Fig. 1**)

In order to show this **P** has two opposite semantics.

```
% Model 1
% a = cycle, p = person, c = car
% det = be detected by the car driver, trk= a cycle or person in the track
% observed(a,p,c) = in all time the driver in the car observed a person and a cyclist
a (t + 1) :-a(t), time (t)
p (t + 1) :-p(t), time(t)
a(t+1), a(t), p(t+1),p(t): -occluded(a,t),occluded(p,t) % default assumption for unseen objects
a(t), p(t); - observed(a,p,c)
observed(a,p,c): -det(c,a,t), det (c, p,t), trk, time(t)
```

then we have a model{`a(t), p(t), observed(a,p,c), occluded(a,t),occluded(p,t)`}, which shows that a person and a cyclist all time appear. This is normal result that we desire for, and standard model.

```
% Model2, unusual case
% model2 is the same as rules in the model1, but except below
not a(t+1), not a(t), not p(t+1), not p(t): -occluded(a,t),occluded(p,t)
% default assumption for unseen objects is no longer valid.
:-occluded(a,t),occluded(p,t)for some t
```

Then we have another model {not a(t), not(p), -occluded(a,t),occluded(p,t); for some t }. This model2 indicate that the person and cyclist may not the same ones as they were before, and the driver in the car will not detect and witness the changes in objects in history of driving in the tracks in the example scene. If the drivers are not able to or unsure to changes in traffic roads, then they face risk because of mistakes in judgement and decisions during driving Obviously, default rules are critical hypothesis for understanding the frame problem in changed world, like the traffics.

CONCLUSIONS

Default rules in ASP are associated with the frame problem, depending on the hypothesis of world inertia, which is warrant for the successful self-driving. How to tackle the frame problem is an issue for logical programming-based control strategy.

REFERENCES

Aditya, S., Yang, Y., Baral, C., Fermuller,C., Aloimonos,Y., (2015), "Visual commonsense for scene understanding using perception, semantic parsing and reasoning", in *2015 AAAI Spring Symposium Series*

Marr, D., (2010), "*Vision: A Computational Investigation Introduction the Human Representation and Processing of Visual Information*". The MIT Press

Lifschitz, V., "Achievements in answer set programming", (2017), *Theory Practice of Logical Program*. 17 (5–6), 961–973

McCarthy J. Hayes, P, (1969), "Some philosophical problems from the standpoint of artificial intelligence". In Meltzer, B. and Michie, D. and Swann, M. (eds.) *Machine intelligence* 4, 463-502, Edinburgh University Press

Moore R. C., "Semantical considerations on nonmonotonic logic". (1985), *Artificial intelligence*, 25(1), 75-94,

Suchan, Jakob, Bhatt, Mehul, Varadarajan, Srikrishna, (2021), "Commonsense: Visual sense-making for autonomous driving: on generalised neurosymbolic online abduction integrating vision and semantics", *Artificial Intelligence* 299, October

.