Artificial Intelligence Reading Club

Chapter 02 & Chapter 03

Ruoding Wang

411624184@qq.com

2020.10

Table of contents

Chapter 02

1. The intelligence problem and *Intelligence Science*
2. Existing methods and standards
3. Some guiding principles

1. The intelligence problem

About the main problem

(1) What is the intelligence problem?

The *intelligence problem*: the problem of understanding how unintelligent components can combine to generate human-level intelligence;

understanding human intelligence: the endeavor to understand how the human brain embodies a solution to this problem;

human-level artificial intelligence: the project of making computers with human-level intelligence.

(2) Why is the intelligence problem important?

a. human intelligence poses a problem for a naturalistic worldview;

b. Surmounting the human-level intelligence problem also has enormous technological benefits which are obvious enough.

1. *Intelligence Science*

What is Intelligence Science?

A new subfield within cognitive science towards understanding human intelligence.
(P.S. for lack of a better name)

What must be answered by Intelligence Science?

Qualification problem
How does the mind retrieve or infer in so short a time the exceptions to its knowledge?

Relevance problem
Of the enormous amount of knowledge people have, how do they manage to retrieve the relevant aspects of it to sort from many of the possible interpretations of a verbal utterance or perceived set of events?

Integration problem
How does the mind solve problems that require different computational methods?

2.Existing Methods and Standards are not Sufficient

Formal Linguistics

It is possible therefore to do great linguistics without addressing the computational problems (e.g. the relevance problem from the last section) involved in human-level language use.

Neuroscience

1. Cognitive neuroscience
the field is more concerned with what parts of the brain embody a solution to the intelligence problem, not how they actually solve the problem.

2. Computational neuroscience
concerned with explaining complex computation in terms of the interaction of less complex parts.
(cognitive modeling below applies to computational neuroscience)

2.Existing Methods and Standards are not Sufficient

Artificial Intelligence

Tow goals:
1)A formal or empirical demonstration that an algorithm is consistent with, approximates, or converges on some normative standard.

2) Demonstrating with respect to some metric that an algorithm or system is faster, consumes fewer resources and/or is more accurate than some alternative(s).

Experimental Psychology

There is nothing specifically in this methodology that focuses the field on solving the intelligence problem.

The field’s standards mainly regard the accuracy and precision of theories, not the level of intelligence they help explain.

2.Existing Methods and Standards are not Sufficient

Cognitive Modeling: The Model Fit Imperative

The author thinks:
Some of the practices and standards of the cognitive modeling community, while being well-suited for understanding many aspects of cognition, are not sufficient to, and sometimes even impede progress towards, understanding human-level intelligence.

Why?

2.Existing Methods and Standards are not Sufficient

Cognitive Modeling: The Model Fit Imperative

What is the model fit imperative？

The main approach to modeling today is to create a model of human cognition in a task that fits existing data regarding their behavior in that task and, ideally, predicts behavior in other versions of the task or other tasks altogether.

Analogy: Aviary Science

Problem:

It is possible to collect data about human cognition, build fine models that fit the data and accurately predict new observations – it is possible to do all this without actually helping to understand human intelligence.

2.Existing Methods and Standards are not Sufficient

Why are these methods and standards not sufficient?

1. each field's standards make it possible to reward work that is not highly relevant to understanding human intelligence;

2. there is nothing in these standards to encourage researchers to discover each field's gaps in its explanation of human intelligence;

3. these standards can actually make it difficult for significant advances towards understanding human-intelligence to gain support and recognition.

3. Some guiding principles

Understanding human-intelligence should be its own subfield

It needs its own scientific standards and funding mechanisms.

Model the right data

The most important datum for intelligence scientists to model is that humans are intelligent.

Take AI seriously

On a technical level, search, neural networks, Bayesian networks, production rules, etc. were all in part ideas developed by AI researchers but which play an important role in cognitive modeling today.

3. Some guiding principles

Have a success

The field needs a success to show that real progress is capable soon.
We thus need metrics that push the state of the art but are at the same time realistic.

Develop realistic metrics

Developing realistic methods for measuring a system's intelligence would
1)make it possible to confirm that the ideas underlying it are an important part of solving the intelligence problem.
2)increase confidence in the prospects of intelligence science enabling quicker demonstrations of progress.

Difficulty

Challenge problems should be difficult enough so that a solution to them requires a significant advance in the level of intelligence it is possible to model.

3. Some guiding principles

Ease

Challenge problem should be as simple as possible so that real progress is made while avoiding extraneous issues and tasks.

Incremental

It should be possible to demonstrate advances towards the goal short of actually achieving it.

General

The extent to which a challenge problem involves issues that underlie cognition in many domains makes progress towards solving that problem more important.

Discussion

Chapter 03

1.Introduction

What is Psychometric AGI? (Brief Definition)

Psychometric AGI (PAGI; pronounced “pay guy”), in a nutshell, is the brand of AGI that anchors AGI science and engineering to explicit tests, by insisting that for an information-processing (i-p) artifact to be rationally judged generally intelligent, creative, wise, and so on, the artifact must be capable of passing a suitable, well-defined test of such mental power(s), even when it hasn’t seen the test before.

What is the Piaget-MacGyver Room (PMR)?

It is such a room that an i-p artifact can credibly be classified as general-intelligent if and only if it can succeed on *any* test constructed from the ingredients in this room.

1.Introduction

What is hypothetico-deduction?

In hypothetico-deduction one creates hypotheses $h1, h2, \ldots, hn$, conditionals of the form $hi \rightarrow ri$, and then tests to see whether the results ri do indeed obtain, following upon an instantiation of hi. If ri doesn't obtain, *modus tollens* immediately implies that hi is to be rejected.

What is analogico-deduction?

When analogical reasoning undergirds either the generation of the hypotheses or the conditionals, or the negation of a result ri, the overall process falls under analogico-deduction.

2. Theory Foundation of Psychometric AGI

What is general intelligence?

Some i-p artifact is intelligent if and only if it can excel at *all* established, validated tests of neurobiologically normal cognition, even when these tests are new for the artifact.

Psychometric AGI

Psychometric AGI is the field devoted to building i-p artifacts capable of at least solid performance on all established, validated tests of intelligence and mental ability, without having seen these tests beforehand at all; the class of tests in play here includes not just the rather restrictive IQ tests, but also tests of the many different forms of intelligence seen in the human sphere.

Newell’s three routes towards General Machine Intelligence

- **“You Can’t Play 20 Questions with Nature and Win.” Newell (1973)**

  He says “man is an information processor” and offers three possibilities for addressing the fragmentary nature of the study of mind as computer:

  - **“Complete Processing Models”**

  Of course today’s cognitive architectures can be traced back to this first possibility.

  - **“Analyze a Complex Task”**

  To accept a single complex task and do all of it, for example, playing chess.

  - **“One Program for Many Tasks”**

  To stay with the diverse collection of small experimental tasks and to construct a single system to perform them all.

Descartes's two tests

- Descartes suggested two tests to use in order to separate mere machines from human persons.

- The first anticipates the so-called "**Turing Test**."

The first is, that they could never use speech or other signs as we do when placing our thoughts on record for the benefit of others.

- The second anticipates the **Piaget-MacGyver Room**.

Although machines can perform certain things as well as or perhaps better than any of us can do, ……we may discover that they did not act from knowledge, but only for the disposition of their organs.

Descartes's two tests

- PMR is designed specifically to test for the level of proficiency in using what Descartes here refers to as a "universal instrument."

- For while reason is a universal instrument which can serve for all contingencies, these organs have need of some special adaptation for every particular action.

3. Piaget's Magnet Challenge

Magnet Challenge

See in Fig. 3.3

Piaget's cognitive stage IV: formal operation

Neurobiologically normal humans can reason accurately and quickly over formulas expressed in the logical system known as first-order logic.

Hypothetico-deductive method

To provide an explanation is to rule out hypotheses until one arrives deductively at ϕ.

Methods: modus tollens/disjunctive syllogism.

4. Modeling human cognition with Slate+LISA

What is LISA?

Learning and Inference with Schemas and Analogies

To create a neurally-plausible model of analogical reasoning by using a hybrid connectionist and symbolic architecture.

What is Slate?

?

4. Modeling human cognition with Slate+LISA

- Analogico-deductive reasoning

Analogical reasoning can often be useful in generation of hypotheses and theories to explain unfamiliar phenomena. (for example, wave theory of sound→wave theory of light)

Gou's reasoning process

1. Knows that More_P(x.y):a pair of boxes *x* is more likely to stop the rotating bar than another pair of boxes *y*.
2. Knows that Heavier(x,y) :some boxes are heavier than others.
3. Has some knowledge of the transitivity of weight.
4. Knows Causes(Heavier(x,y),More_P(x,y)) : weight is causally responsible for the bar's stopping.

4. Modeling human cognition with Slate+LISA

See Fig. 3.5

D is then subjected to reasoning by analogical inference. To do this, **D** must first be divided into two subsets: **Dsource** and **Dtarget**.

.

⇩

See Fig. 3.6 (produced by LISA)

⇩

D'(See Fig. 3.5)

⇩

D' is returned to Slate, subjected to deductive reasoning

4. Modeling human cognition with Slate+LISA

From D∪D', one of following things would be derived:

1. A testable hypothesis
2. A contradiction

See Fig. 3.5

D' contains *More_P*(*circle,square*)
But Gou's experiment shows that ¬*More_P*(*circle,square*)

Conclusion: Weight is not the cause.

Thank you for your time!