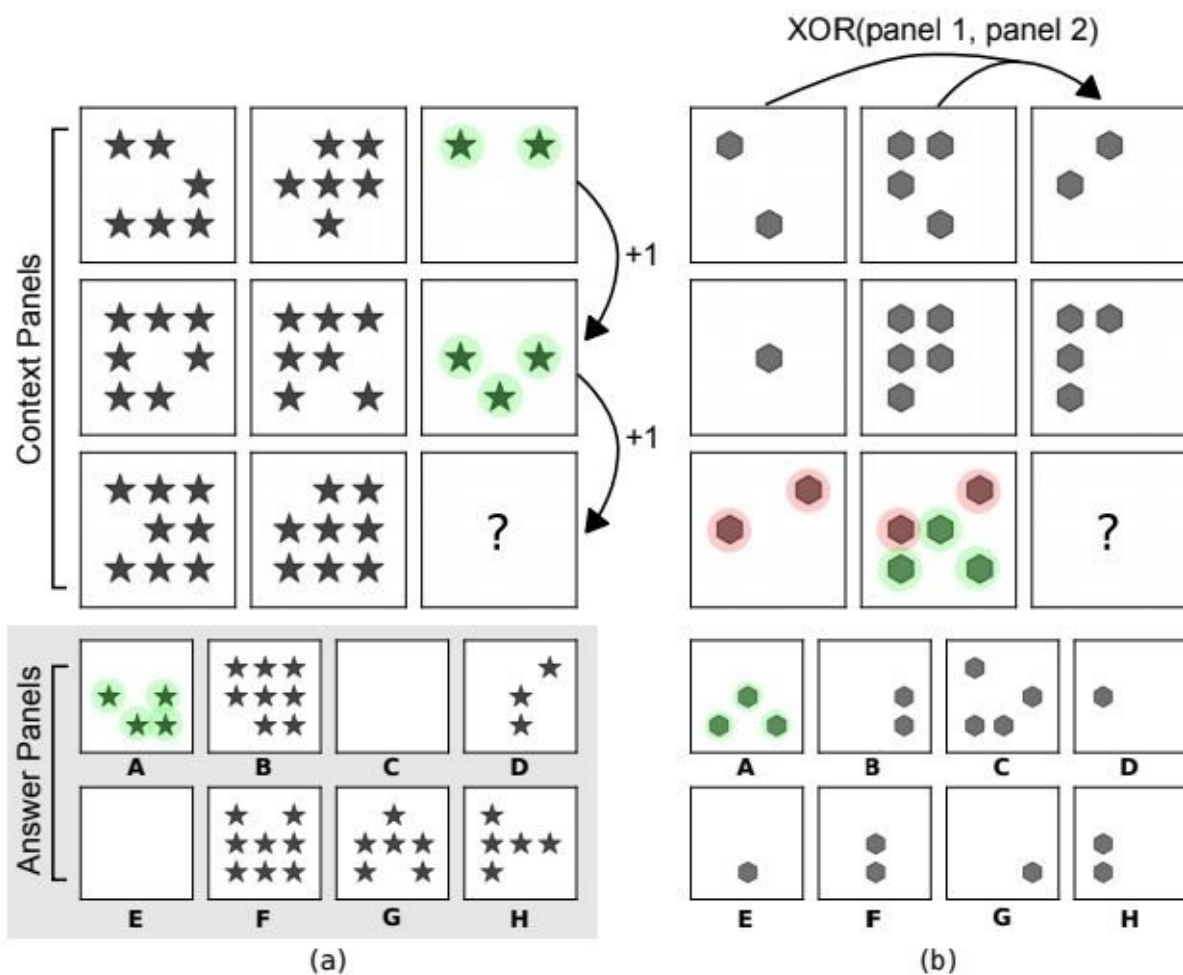


Are you eating your relish with dogs?

Testing, testing AI

July 17 2018, by Nancy Owano



Raven-style Progressive Matrices. In (a) the underlying abstract rule is an arithmetic progression on the number of shapes along the columns. In (b) there is an XOR relation on the shape positions along the rows (panel 3 = XOR(panel 1, panel 2)). Other features such as shape type do not factor in. A is the correct

choice for both. Credit: arXiv:1807.04225 [cs.LG]

Testing, testing: DeepMind sits AI down for an IQ test. While the AI performance results are not staggering in trumping or matching human reasoning, it is a start. AI scientists recognize that establishing their capacity to reason about abstract concepts has proven difficult. DeepMind wanted to see how AI could perform and the team proposed a dataset and challenge to probe abstract reasoning.

Can AI match our abilities for [abstract reasoning](#)? Will deep neural networks be better able to solve abstract visual [reasoning](#) problems in the future? The DeepMind researchers have certainly been on the case.

Their paper, "Measuring abstract reasoning in neural networks," is on arXiv. Authors are David Barrett, Felix Hill, Adam Santoro, Ari Morcos, Timothy Lillicrap, from DeepMind. You can check out what they were looking for and how they tested. The paper basically focuses on an approach for measuring abstract reasoning in learning machines. In their discussion, the team said, yes, there has been progress in reasoning and abstract representation learning in neural nets—but the extent to which these models exhibit anything like general abstract reasoning "is the subject of much debate."

The models to succeed had to cope with generalization regimes in which the training and test data differed. They said they presented an architecture with a structure designed to encourage reasoning. Results: Mixed bag. They said their [model](#) was proficient at certain forms of generalization, but weak at others.

Nonetheless, it is noteworthy that they explored ways to measure and elicit stronger abstract reasoning in neural networks.

"Standard human IQ tests often require test-takers to interpret perceptually simple visual scenes by applying principles that they have learned through everyday [experience](#)," said a DeepMind blog. "We do not yet have the means to expose machine learning agents to a similar stream of 'everyday experiences', meaning we cannot easily measure their ability to transfer knowledge from the real world to visual reasoning tests. Nonetheless, we can create an experimental set-up that still puts human visual reasoning tests to good use."

They proceeded to build a generator for matrix problems with a set of abstract factors. The team is encouraging more research in abstract reasoning, and they made their [dataset](#) publicly available.

Big-picture question is if scientists can achieve humanlike analytical reasoning capabilities.

While their IQ test-giving results might have been a mixed bag, the researchers do not see this as a game of winning or giving up. They will keep up their work to explore strategies for improving generalization and explore future models. As *CIO Dive* remarked, "Intelligent assistants have been fed mountains of data to help consumers in almost every conceivable area, yet when presented with unknown problems can still fall short."

The authors wrote, in their abstract, "we propose a dataset and challenge designed to probe abstract reasoning, inspired by a well-known human IQ test. To succeed at this challenge, models must cope with various generalisation 'regimes' in which the training and test data differ in clearly-defined ways. We show that popular models such as ResNets perform poorly, even when the training and test sets differ only minimally, and we present a novel architecture, with a structure designed to encourage reasoning, that does significantly better."

CIO Dive described their tests as [visual](#) IQ tests. In the process, the authors were interested to see performance in abilities for generalizing when [test](#) data were different.

Matching AI with human abilities for abstraction continues to be an uphill battle.

As *CIO Dive*'s Alex Hickey wrote, AI would need to distinguish different meanings between "eating spaghetti with cheese" and "eating spaghetti with dogs."

The paper commented that testing the capabilities of neural nets can be tricky and [neural networks](#) have their pitfalls, given their capacity for memorization and ability to exploit superficial statistical cues.

More information: Measuring abstract reasoning in neural networks, arXiv:1807.04225 [cs.LG] arxiv.org/abs/1807.04225

Abstract

Whether neural networks can learn abstract reasoning or whether they merely rely on superficial statistics is a topic of recent debate. Here, we propose a dataset and challenge designed to probe abstract reasoning, inspired by a well-known human IQ test. To succeed at this challenge, models must cope with various generalisation 'regimes' in which the training and test data differ in clearly-defined ways. We show that popular models such as ResNets perform poorly, even when the training and test sets differ only minimally, and we present a novel architecture, with a structure designed to encourage reasoning, that does significantly better. When we vary the way in which the test questions and training data differ, we find that our model is notably proficient at certain forms of generalisation, but notably weak at others. We further show that the model's ability to generalise improves markedly if it is trained to predict symbolic explanations for its answers. Altogether, we introduce and

explore ways to both measure and induce stronger abstract reasoning in neural networks. Our freely-available dataset should motivate further progress in this direction.

© 2018 Tech Xplore

Citation: Are you eating your relish with dogs? Testing, testing AI (2018, July 17) retrieved 10 April 2024 from <https://techxplore.com/news/2018-07-relish-dogs-ai.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.