BLOG POSTS ABOUT COMMON SENSE AI BY RONAN DALE

"COLOURLESS GREEN IDEAS SLEEP FURIOUSLY": APPROACHES TO FORMALISING SEMANTICS IN COMPUTATIONAL LINGUISTICS

The principal objective of this module for me is to establish a foundation in research into formalising text into a data-structure to capture its meaning so that others can use what I've surveyed and deduced myself, then build upon it to ascend up this Mount Everest of an end goal which is to create a formalisation of semantics similar to how 'The Sims' formalises life in the real world.

The phrase 'formalising text into a data-structure' seems quite straightforward initially; shapes have been formalised like that thanks to research into Computer Vision so surely this can be repeated for text. However, Computer Vision and Graphics have been researched extensively; it is a mature field, whereas my research indicates that considerably less scholars have invested time into this aspect of Natural Language Understanding.

This blog describes my approaches, both successful and unsuccessful, to gathering information on a problem that Professor Teufel (University of Cambridge) describes as far from being solved; "The second wave of AI, right now, is soon going to fail because too much trickery and even self-trickery is used."

APPROACHES TO RESEARCH

From the perspective of a first-year student such as myself, research papers are a nightmare. Between presumed understanding of advanced concepts in mathematics and descriptions of ideas that span several pages, distinguishing between valuable information and material of no relevance to me was always going to be challenging. In order to face that problem however, you need to acquire relevant research papers. This was the painfully arduous at times.

I started with keyword searching which could be considered a 'lottery-based approach' because while you are likely to find some relevant papers which include the terminology that you use, it's possible that there are other insightful papers which are omitted from the search results because they address the same problem differently.

I would actively discourage other researches from doing this for a long time because you're unlikely to gather anything meaningful apart from the names of important people from the field of research. This in itself constitutes as something of value because it leads us directly to our next research approach: searching by people who have made a contribution to the field.

Keyword searching for phrases like 'formalising text into data structure' and 'rule-based approach to formalising text' yielded a limited supply of papers to work with. 'Formalising natural language' was surprisingly successful because I read a paper which lead me on to researching tools from Computational Linguistics. It described 4 tools for formalising grammars; XFST, GPSG, LFG and HPSG. Initially, I was hesitant to accept that researching linguistics papers would be of any value. However, I reminded myself that formalising text into a data structure is a problem which applies Machine Learning to a linguistics field of research.

I missed countless results that are related representing text by using the keywords selected above; they did not account for graph-based representations of the meaning of text which from my perspective, is a more semantically oriented approach to formalising text.

For my second and third blog entries, my research approach changed slightly as I became aware of existing simulations to represent common sense. I found a YouTube channel called

'Two minute papers' which showcases the latest highly cited papers. In addition, I found DARPA's YouTube channel which actually has published their vision for an approach to reaching the general-purpose robot that has common sense and their approach places great importance on applying Psychology studies to simulations. This led me to then look at how Behavioural Psychologists experiment with children to find out what they know about the world at for example, 4 months old. If someone mentions a technology/module in their blog, it's no harm in clicking on the link and reading about it yourself because you may find synergy between it and some other module like I've described in my 3rd Blog entry.

My objective in my mind for this entire research project has been "If I had to give a TED talk on approaching the Representation of Common Sense Knowledge for AI, have I enough general knowledge about the field that people would enjoy it and learn something?". This project is no different to any other which involves research; the more sources, the better. It amazes me that I've been all over the internet to gather this information, between BBC, PwC, Google AI, MIT, TED talks, Cycorp, Harvard Psychology Dept and DARPA.

FIGURES OF INTEREST: NOAM CHOMSKY

The first important name I came across while researching Linguistics is Avram Noam Chomsky (1928 – Present). In 1957, he published a monograph called 'Syntactic Structures' which argues that natural language semantics and syntax should be considered independent of each other; sentences which are syntactically coherent can have no meaning. His famous example is *'Colourless green ideas sleep furiously'*. The syntactic structure [ADJ, ADJ, NOUN, VERB, ADV] is acceptable yet the individual words when put together are nonsense.

One valuable point that I deduced from reading about "Syntactic Structures" is that **a metric for success** for our working machine is for it to be able to **recognise nonsense**. The ability to do that would indicate to me that the machine is clever; it uses reasoning and understands each of the words in that sentence as well as why they make no sense in that order. A true measure of this intelligence would be if the machine could explain **why** that combination of words is invalid. In our example case, 'Colourless green ideas sleep furiously', a well-designed machine would argue ['Something cannot be colourless and coloured at the same time.', 'Sleep is a passive state, so adverbs related to aggression don't make sense'] just as a human would.

Furthermore, Chomsky distinguishes between a 'grammatical' and 'ungrammatical' sentences where 'grammatical' sentences are intuitively acceptable to a native speaker. His famous example is grammatical, but not statistically probable. He then relates phonology to semantics, arguing that how a word is said means more than the word itself.

FIGURES OF INTEREST: CHRISTOPHER D. MANNING

Professor Manning lectures and researches Natural Language Processing at the Stanford University. His focus is on a computational linguistics approach to parsing, Natural Language Inference and Multilingual Language Processing. In fact, he was a key developer in the SNLI corpus.

The paper describing the process of creation this corpus https://nlp.stanford.edu/pubs/snli_paper.pdf

What is Natural Language Inference? The task of determining if a hypothesis is true, undetermined or false given a premise. A metric for success for this is if a machine can determine the nature of the hypothesis having been shown only the premise and in addition, demonstrate 'common sense' by explaining why it has reached the decision. The reason that we are concerned with NLI is because our final product will likely be trained on the rules of the world so that it can take input like a movie synopsis, then reproduce an accurate simulation of what was described in it. Another student on my course is working on Automated Text Generation, so one of my visions would be for their tool to produce a novel synopsis, then for mine to understand the situation it describes and generate a simulation to represent it. If you are interested in this vision, I recommend that you explore Generative Adversarial Networks.

FIGURES OF INTEREST: KRISHNA P. GUMMADI

Professor Gummadi researches at the Max Planck institute and is very interested in the ethics of AI which is used to make decisions which affect people. He deals with AI that make decisions about 'grey-areas' like predictive policing. In the Turing Talk 2019, he highlighted startling biases within machine learning systems like facial recognition software and a recidivation predictor called COMPAS. Some facial recognition tools have a success rate as low as 66% for black females despite being 99% accurate when identifying white males. Our end goal also relies on **unbiased decision making** not only founded in **rule-based logic** but also in common sense given how ambiguous the English language can be. If our machine is trained on biased data, its common sense will be warped and the simulations we produce may not be an objective representation of what was actually described.

He also highlights bias within an existing semantic representation: Word2Vec which shows the translation between 'Man' and 'Computer' as of the same magnitude of that between 'Woman' and 'Housekeeper'.

Gummadi is also interested in multilinguistic semantics in the form of machine translation. This field could be bolstered by an investigation into representing semantics because of how many phrases do not semantically translate from source to target language; "Tengo un hambre que da calambre." literally translates to "I've a hunger that gives cramps" but is semantically equivalent to "I'm starving." Making a machine understand the concepts that are represented by words will inevitably lead to improved translation quality and would mean that our end-goal could reproduce the same simulation regardless of the language of the input text.

FIGURES OF INTEREST: SEBASTIAN PADÓ

I found this researcher because he is a keynote speaker for RANLP which took place in Bulgaria during Summer 2019. His expertise lies in Computational Linguistics; having carried out postdoctoral research in Stanford. His research interests specifically address our 'Everest': developing graphical representations for the meaning of natural language words, phrases and documents obtained from a corpora.

WHY IS REPRESENTING THE MEANING OF WORDS IMPORTANT?

The value of representing the meaning of words could be described anecdotally in terms of a band playing a song; your interpretation of a song is highest when all instruments are playing together. If you only listen to one instrument, your understanding of the overall song is greatly reduced. The same is true in NLU; if you isolate a word from its context, it is semantically ambiguous. Although context is key, syntactic ambiguity is another challenge that an intelligent machine would have to overcome. Take the Winograd Schema; a list of 150 sentences containing a referent (it) and accompanied by a couple of words, each changing what the referent refers to. If the machine can score highly on Winograd Schema examples, it's a sign that it has common sense and contextual awareness.

REQUIREMENTS FOR FORMALISING TEXT AS A DATASTRUCTURE 1: COMPUTATIONAL POWER

In the paper 'Language Modelling with Gated Convolutional Networks', Dauphin, Fan, Auli and Grangier propose that a metric of success for a model. The throughput is the number of tokens being processed per unit time. Parallel processing accelerates this. Responsiveness is described as processing tokens in sequence. Typically, responsiveness and throughput are inversely proportional however batching can make them directly proportional to each other.

TEACHING AI COMMON SENSE

The term ‘Artificial Intelligence’ can be woefully misleading because from my perspective, it is naïve to consider a machine which is excellent at recognising patterns and following rules to be intelligent in the same sense that a human is. In 1975, Noam Chomsky debated against Jean Piaget on the motion of how linguistic knowledge is acquired. Piaget proposed that when a baby is born, its brain is like a blank slate with no aptitude for learning. However, Chomsky refuted this proposal and suggested that a new-born’s brain has innate abilities to learn. This is known as Chomskian Nativism versus Piagetian Constructivism. Common sense is relatively easy to identify, yet surprisingly difficult to define. The sentence “The film received no applause from the audience because it was too boring.” What was too boring? We as humans understand instinctively that the subject of the referent ‘it’ is the film because ‘boring’ describes what the audience thought of the film and to describe an audience as ‘boring’ in this context is somewhat nonsensical.

Oren Etzioni is a key figure of interest in the quest for transcending the boundary between ML/Deep Learning and genuine machine understanding. At the Allen Institute for Artificial Intelligence, Etzioni is pioneering a project which has the objective of endowing computers with common sense: Project Mosaic. He argues that most AI systems in the present day are highly capable when it comes to performing the specific task they have been trained to do, yet woeful at adapting to changes. Before describing Project Mosaic, it’s useful to note two examples of this lack of adaptability.

In March 2016, Googles Deep Learning acquisition DeepMind, trained an AI to play Go, a game with a number of possible game-states that exceeds the number of atoms in the universe. Lee Sedol held 18 World Championship titles and is a 9-Dan level player (the equivalent of Djokovic or Nadal in tennis). He was invited to participate in five matches against AlphaGo. He resigned in the first three matches and was visibly alarmed at the capability of this machine. In the second match, move 37 by AlphaGo would have been executed by 1 in 10,000 human players; even the commentators could not initially justify why that position was chosen until it contributed to the winning strategy. Despite being this ‘clever’ and having been trained on millions of games, Lee Sedol made it resign in the fourth match because he had created a game-state sufficiently complicated that AlphaGo hadn’t enough prior knowledge of it to make strategic plays. The first move it made in error was one that no human would make because it yielded no potential advantage.

It’s also worth touching on research done with Atari’s Breakout. When an AI learned to play this game using Deep Learning, it realised that the most optimal technique is to break through a line of blocks and then rebound off the top wall and the edges. However, when another group of researchers changed fixed aspects of the game such as moving the paddle up a few pixels or by making unbreakable blocks, the AI’s performance deteriorated severely because it had not been trained to deal with these modifications. The conclusion of the research was the AI had a different style of intelligence to humans; it was able to perform one task with fixed variables really well but cannot adapt to a simple change in conditions like a human would.

Formalising common sense has been approached in both Computer Vision and Natural Language Understanding through Mosaic. I began by reading about their 'SWAG' (Situations with Adversarial Generations) dataset and its testing process. This was published by Rowan Zellers, Yonatan Bisk, Roy Schwartz and Yejin Choi and was published at the 2018 Conference on Empirical Methods in NLP. This dataset was made possible by contracting workers from Mechanical Turk to filter the 4 possible endings to each sentence in the dataset. All possible measures were taken to mitigate racial bias in the dataset. However, given its nature of being ActivityNet and LSMDC video captions, this is an impossible task to perfect. According to PwC, "Machine learning (especially deep learning) models can be duped by malicious inputs known as 'adversarial attacks'. It is possible to find input data combinations that can trigger perverse outputs from machine learning models, in effect 'hacking' them."

Each record in the dataset consists of the start of a caption and four possible endings. One of them is the correct caption. One of them is an alternative future which didn't happen, but it is acceptable to reason that it could have. Another is a future which is unlikely to happen based on common sense; in the example below, the striker would be going against the principles of football by openly giving the ball to an opponent. The last option is grammatically and contextually nonsense. It would make no sense for this to be the end of the caption.

For example:

"The forward received the football that was passed to him. He"

A) "kicked it in the direction of the net to try to score." (**Most likely**)
B) "was tackled by an opposing defender." (**Second-most likely**)
C) "passed it to an opposing defender." (**Unlikely**)
D) ", did her makeup!" (**Nonsense**)

When creating and testing the model, other NLP tools were called upon, including SpaCy's dependency parser and GloVe (Stanford's 300-dimension vector representation of semantics). The dependency parser was used to extract the head verb from each possible ending, as well as its dependent object which GloVe was the first word representation style to be tested. The others were Numberbatch (300-dimension) and ELMo (1024-dimension).

REACHING BEYOND THE VENEER OF INTELLIGENCE

A Literature overview and evaluation of figures of interest, existing tools and datasets in the field of common sense ontologies.

Common sense knowledge is a focus of the Paul Allen Institute for AI research.

The first paper to address our area of interest is 'Programs with Common Sense' by John McCarthy published in 1958. It proposed a hypothetical program called 'Advice Taker' which converted logic into representable information. He advocated for computers to manipulate common instrumental statements which have been written in a suitable formal language (possibly predicate logic). It will then draw conclusions from a list of premises which will be declarative or imperative sentences. ChatBots today often include the ability to take in (listen) to statements from the user and those statements will then be stored in its knowledge base. However, it's **how** it handles those statements to make reasoned decisions that matters.

Link to paper and peer review:

http://www-formal.stanford.edu/jmc/mcc59.pdf

Why is Advice Taker important to us?

A functional 'Advice Taker' represents the final goal of the project – to make a computer's understanding of the world improvable by simply feeding it statements. 'Programs with Common Sense' also defines a metric for success: "A program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows."

Cyc Dataset

OpenCyc's Knowledge Base (KB) is comprised of the following entities: 42,000 relational predicates which connect terms, 500,000 collections, 1,500,000 general concepts and 24,500,000 rules. It is represented in full 1st order logic and understands concepts such as time and context.
Cyc's goal is to be the universally accepted ontology for any application which requires knowledge-based reasoning and does NOT NEED to use Machine Learning; it's a ready to use package.
Cyc uses 3 types of reasoning: abductive, deductive and inductive:
Abductive: Start with at least one observation, then find the simplest and most likely reason why it happened. Abductive conclusions have a remnant of uncertainty or doubt, usually being expressed as "most likely".
Deductive: Reasoning from one or more general factual statements to reach a logically certain conclusion; Example: (statements: "John is older than Kate. Kate is older than Conan" conclusion: "John is older than Conan")
Inductive: Reasoning from observation of specific premises in the form of evidence to a general conclusion; if the premises are true, then the conclusion is probably true.
A priority within the dataset is **reusability** through making **generalisations**; instead of asserting "Doctors need medical training.", "Paramedics require medical training." and "Nurses need medical training", Cyc generalises that "All medical professionals need medical training" and then generalises that "All doctors are medical professionals" etc.
Case Study 1: Cleveland Clinic
When building a Cyc application for Cleveland Clinic, only around 120,000 new entities needed to be added because the data set's general knowledge encompassed much of what it needs to know for clinical trial cohort selection processes; in fact, it already possessed 99.5% of the knowledge required to be fit to do this! It filtered through 5 million patient records to find the optimal 82 candidates for a new cardiovascular therapy which can take humans 10 weeks. It took Cyc 10 minutes.
Case Study 2: Multinational Oil Company
Cyc was used to increase oil yield by 10% and also improved the safety conditions of workers on an oil rig.
This is a visualisation of the taxonomic relationships inside Cyc
https://www.youtube.com/watch?v=8NRvhGm9EOg

The key finding from this video is the 'Wikipedia-like' format that the data structure adopts; from each individual node (representing an object), the statically represented

properties/characteristics of that object accompany it like they would in a table of a SQL database.
Cyc acquires this background knowledge through a process called ‘Structured Knowledge Source Integration’ or SKSI. This enables Cyc to access, query, assimilate and merge external structured data sources like databases or spreadsheets and can draw upon information obtained from several of these external sources when making a conclusion or mediate between semantically similar knowledge sources. Having mentioned SQL, Cyc has its own
query language: CycL and the statements are in the following format:

(thereExists ?CITY
(and
(isa ?CITY USCity)
(geographicallySubsumes
NewEngland-USRegion ?CITY)
(weather ?CITY RainyArea)))
“Is it raining somewhere in New England?”
In 2006, a paper was published by 3 employees of Cycorp which describes the process of integrating semi-structured data sources such as websites into Cyc to further increase querying power.

For every noun and verb in its vocabulary, this dataset has the word and all its ‘senses’ (depending on the context, the same word can take on a different meaning).
There was a TedX talk in Austin, Texas by Doug Lenat, their CEO which was useful because it presents his vision for the future of AI. I found his remarks about current AI assistants like Siri, referring to them as ‘the dog that brings you your newspaper without actually being able to understand the news’ and I went on to demonstrate this with Mitsuku by teaching it knowledge about Pokemon and then asking it questions which would require one layer of reasoning and it completely failed to relate the pieces of information I had provided.

Link to talk: https://www.youtube.com/watch?v=2w_ekB08ohU

Short blog on Cyc: https://mc.ai/common-sense-knowledge-crucial-for-the-success-of-ai-systems/

Specifically, I find Cyc offers advantages in the following ways:

- Its inference engine has been supporting one of the most expressive knowledge representation languages (CycL)
- Its HL Module framework allows the ability to add additional special purpose reasoners optimized for specific types of problems
- Its SKSI technology enables a mature means of interfacing semantically with databases and triple stores allowing inference and database/triple store queries to be intermingled.
- Its ontology with linkages to English and other languages to a lesser extent stands up well to related systems and has often provided me with useful insight when ontologising new knowledge domains via OpenCyc or ResearchCyc.

- It comes with a variety of ontology development environments that support ontological development across multiple asynchronous developers using a variety of text- and GUI-based authoring tools.

MathCraft

This is Cyc's version of a game-based representation of knowledge. At a high level, its purpose is to develop the learning of not just the child who plays the game, but also the AI behind it. The approach to learning used puts the child in the teaching role. They accompany the AI in a spaceship environment which has a series of challenges which can be solved by walking the AI through their line of thinking. The AI will attempt some puzzles and approach them incorrectly. Then, the child must point out any flaws in its logic and explain the correct approach.

What value do I see in this?

MathCraft's premise is applicable to not just teaching mathematics, but also teaching language and simulating a conversation between an AI character and a human.

Stanford Blog – Harry Collins

https://web.stanford.edu/group/SHR/4-2/text/collins.html

Looking at an existing formal representation of Common Sense: IsisWorld

In short, it's an open source commonsense simulator for AI research which is based on a 3D virtual environment in which human commonsense tasks are mapped to real world problems. Write commands in Python to set up the environment and then you expect an output which is similar to what I had described a while ago in my diary: a simulation of exactly what was specified in text. In fact, according to the docs, "It's easy to write *scenario templates* and create custom evaluation scenarios." and "Environment descriptions are procedurally generated: the sizes and orientation of the kitchen are determined each time the environment is generated as an effort to prevent AI systems from "overfitting" the task.

```
def environment():
    k = kitchen()
    put_in_world(k)

    f = fridge()
    put_in(f, k)

    ralph = IsisAgent("Ralph)
```

MIT urges researchers to focus on metareasoning (Marvin Minsky paper) "You can't think about thinking about thinking without thinking about thinking about something." A formalised problem domain would be beneficial.

A set of concrete problem domains

A reasoner that solves problems in a set of concrete problem domains

A metareasoner that solves problems in a reasoner that solves problems in a set of concrete domain problems

Examples of Canonical Problem Domains

Go (AlphaGo Zero mastered this domain)

Atari Breakout

Chess (IBM Deep Blue)

These canonical domains aren't ideal for common sense reasoning about typical scenarios in the real world because solutions to canonical problems will always "overfit" the problem domain, and not generalise to other classes of problems. Even though Chess and Heist Negotiations share lots of parallels as shown by the Professor in La Casa de Papel representing the state of the hostage situation outside the Spanish Mint, an AI specifically trained to master the domain of chess would need knowledge transfer capabilities and an awareness of the domain of events that can take place during a heist in order to apply their thought processes to successfully negotiating against the police.

Examples of Generalised Problem Domains

Turing Test and its successors/alternative benchmarks from Allen AI institute

Examples of Physical Problem Domains

RoboCup (football for robots)

Criteria for a good problem domain according to MIT Mind

Easily understandable: i.e. a researcher studying language acquisition can have gross understanding of how his model affects planning and learning.

Tests only relevant components of intelligence, abstracting away "irrelevant" parts. –

Continual/Scalable benchmarks that extend from easy to difficult.

Modular whenever possible, recognizing the merits of specialization (and the limitations of humans).

CLEVERER: An AI which represents what happens in a video as a summary of text

https://www.youtube.com/watch?v=bVXPnP8k6yo

This is a newly released paper and one which summarises videos. Image classification has been around in computer vision for around 5 years and has been able to describe an image in relatively specific detail, being able to infer based on a human's pose, the task that they are carrying out and even more, infer hidden objects that are in their hand like a ball. Video summaries are created by the AI observing the actors in the video and events carried out by them. In addition to the summary, the AI demonstrates one of the characteristics I feel is required in a general purpose Common Sense AI: In addition to providing a formal description of the events in the video, it is able to perform question answering about unseen events/future events which is very impressive. In the example, the question is about 3D shapes made out of different materials colliding. It is actually able to quote the material of the shape that will be hit next by the main actor which means that it has a grasp of intuitive physics, realising that objects move with a continuous trajectory and spatial awareness, realising that the ball will collide with the object nearest to it in the direction it's moving.

According to the video, YouTube will probably be interested in this AI because it would revolutionise its search engine because then the video's tags could include a summary of what it shows, meaning that you could find a video without knowing its title or any of the tags but instead by describing what happens in it and then your normal semantics similarity NLP tools could be used to compare your description of the video you want with the AI's descriptions of videos and only show ones that match!

I can see a lot of applications for this in the reverse, just like with what has been accomplished with novel image generation in Computer Vision. What if instead of the input being the video and the output being the summary, that the input was the summary and the output was a novel video which represents accurately what has been described in the summary. A dataset of common sense rules would increase the quality of the video because the AI could make logical inferences about what the human didn't mention but expects to be present in the video.

Based on this paper, I feel that although Computer Vision is a much more mature field, Natural Language Understanding synergises really effectively with it in this paper. I haven't got around to looking at whether or not this is vulnerable to 'one-pixel' attacks but if it is, then that's an example of where the veneer of intelligence just isn't enough because the human eye would never make that mistake.

SO… WHAT'S THE PLAN?

https://arxiv.org/ftp/arxiv/papers/1810/1810.07528.pdf

Overview of Current Approaches

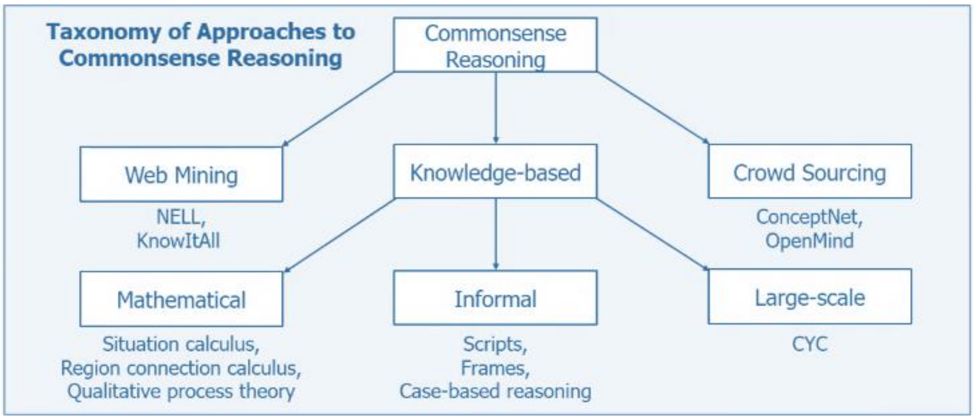


Figure 1: DARPA Taxonomy of Current Approaches into CS Reasoning

DARPA describes Common-Sense as one of the biggest barriers stopping Narrow AI from evolving into General AI. Today, systems are programmed for a specific problem or are trained with big datasets and labelled examples so there is a function to compute any expected task. It acknowledges that Cyc has been the most successful first-generation attempt at formalising commonsense knowledge into a network of information. As stated in my learning log, although it contains 42,000 predicates, 500,000 collection types, 1,500,000 general concepts and 24,500,000 general assertions, it simply isn't enough because symbolic logic representation is brittle.

Objective: A Commonsense Service

What will it enable AI to do?

- Understand natural language questions which AREN'T in a specific formatted ontology and provide answers which can be justified/explained
- Understand new situations and adapt to them
- Monitor the reasonableness of their actions
- Transfer/apply learning to new domains

What should be encoded into a successful Commonsense service?

- Anything an average adult human knows without debate.
- Intuitive Physics
- Intuitive Psychology
- Common Facts about the world
- Core domains of human cognition: objects, agents, places, numbers, forms and social beings

"There is an agent called Any Fool. Any Fool knows exactly what everyone knows, so if he knows something, that's common sense" – John McCarthy

Why Intuitive Physics?

Young children have a rich knowledge of the physical world without needing any kind of qualification in Physics; they just understand concepts like motion.

Why Intuitive Psychology?

A basic understanding of why people do a certain action to meet a certain goal; general abductions, deductions and inductions. For example, "Person goes to bed which means that Person is tired".

Why Common Facts?

These are simply basic pieces of useful information about the world that all its inhabitants have.

Developmental Psychologists will be very important in overseeing the success of a Commonsense service because that field is deeply mature when it comes to analysis of infant common sense. For example, Harvard University has funded research into the Core Knowledge Systems of Childhood Cognition. Intuitive Physics, External Actors and Spatial Reasoning are the three core knowledge systems that when demonstrated, would show that the Commonsense service is progressing towards the goal outlined with the $70 Million U.S funding* which has been invested in solving this problem.

What approaches should we use in creating a Commonsense Service?

- Learning Grounded Representations
- Learning Predictive Models from experience
- Learning Commonsense Knowledge from the internet (NELL and NEIL)
- Understanding and Modelling Childhood Cognition

There MUST be a test environment with which the AI interacts. From this environment, it must learn the foundations of commonsense from videos or simulated experiences.

Another possible approach is improving how we currently use machine learning and crowdsourcing to create a benchmark which defines whether or not a system has commonsense. This benchmark could be in the form of an assessment (AiQ Test) which combines a set of NLU and Image based questions which test commonsense areas of Common Facts, Intuitive Physics and Intuitive Psychology.

Allen AI is currently developing a set of benchmark tests for AI. However, from a developmental psychologist's perspective, there are some techniques which could potentially be applied to an AI to investigate childhood cognition qualities like Intuitive Physics: *"Your* *baby watches pairs of short videos of physical events. On one side, something normal happens: e.g., a ball rolls off a table and falls to the ground. On the other side, something surprising happens e.g., the ball rolls off a table and falls UP!"*

https://lookit.mit.edu/studies/cfddb63f-12e9-4e62-abd1-47534d6c4dd2/

This is a link to a crowdsourced MIT Cognitive Development study in which babies are recorded reacting to possible and impossible physical situations. *"**Where your baby chooses to look can tell us about his or her expectations of how the physical world works**. Although your baby isn't ready to study physics, he or she is already learning the basics: Should things fall up or down or not at all? Should they keep going once they start moving?"*

From my point of view, I consider tests like these to be more measurable because of their format; they're not traditional QA intelligence tests where you're trying to see if an AI is clever or not.

Another crowdsourced study in Cognitive Development from MIT: "How does the physical world in which we live shape the abstract world in which we think? This study will help us find out what kinds of **basic shape information draw infants' attention**. Such attentional preferences might guide infants' learning about more abstract geometric phenomena, like identifying categories of shapes and learning the names of those shape categories (e.g., 'square,' 'rectangle,' or 'triangle')."

https://lookit.mit.edu/studies/3053a70c-53d2-4016-8cd5-55ebefb50814/

MIT is trying to understand how these sorts of beliefs relate to each other, whether babies pass through discrete "stages" of understanding, and how much individual babies' moods and "looking personalities" affect their responses."

This is the timeline that DARPA is using which shows when a cognitive function in 3 Core Domains begins developing within a child and also when it is expected to have fully developed. Note that expression of empathy is one of the last skills to develop in a newborn which is likely also going to be the case with Strong AI. Ideally, a milestone for Strong AI would be that it can duplicate the cognitive skills of an 18 month old child.

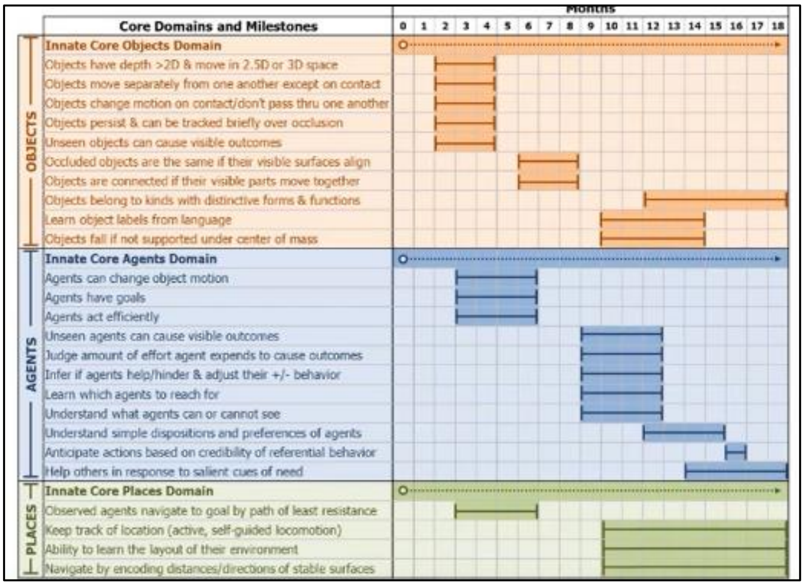


Figure 2: DARPA Chart showing Development within Core Domains of a newborn over time

Another key observation is that a young child will gauge the importance of an actor's action based on how much physical energy they exert.

Consider malicious behaviour at a young age; Child A builds a linear tower of blocks. Child B realises that building that tower takes effort and by knocking it down, Child A will be upset.

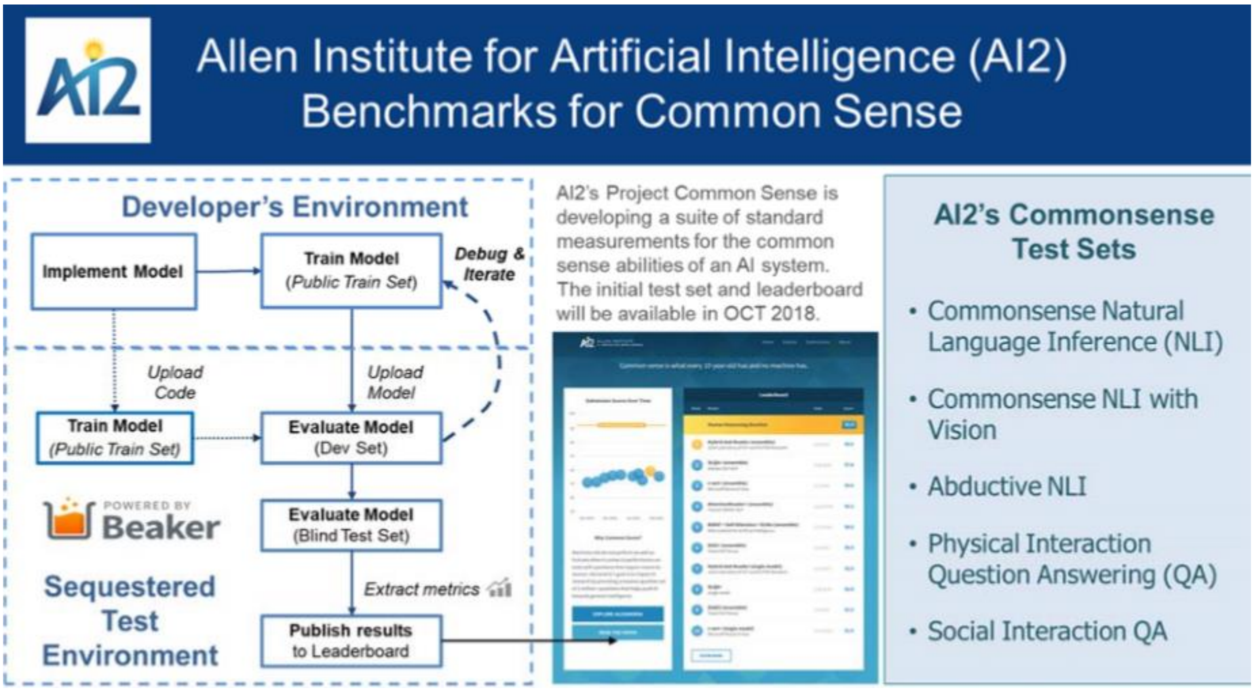
By age 4 months, a child should understand that an object's trajectory is continuous; it cannot instantaneously leap through time. (Spatio-temporal Continuity)

By age 5 months, a child should realise that objects do not instantaneously pop in and out of existence. (Object Permanence)

By age 10 months, a child should realise that objects keep their shapes (Shape Constancy)

Furthermore, there are other behavioural and cognitive psychological experiments which could be used to assess the commonsense of an AI like an awareness of the source of a stimulus. Take the Little Albert experiment which in summary, placed a toddler in a pen with a rat and observed no emotional response. Then, the rat was placed in the pen again but this time, an external source of loud noise was also a stimulus. This upset the child and he was subsequently removed from the pen. Finally, he was returned to the pen and this time the only stimulus was the rat. Despite there being no noise this time, the child was visibly upset as if he attributed the source of the loud noise to be the rat.

At the moment, there are two benchmark tests; Allen AI2's Mosaic dataset of 113k multiple choice questions.



In summary, I feel that a plan to move forward with finding a formal representation for common sense knowledge to be manipulated by an AI will involve experts from Psychology, Mathematics, Linguistics and Computer Science coming together with a lot of funding from several leading research institutions like Paul Allen, Max Planck, Stanford and MIT. I used the example of the heist from La Casa de Papel in my diary and its messages actually transfer over to AI research quite well. Firstly, the team behind the heist/plan is doing something that nobody has attempted but somebody has planned thoroughly. Secondly, all participants are financially incentivised for the plan to succeed and the outcome of the mission revolutionises a country or in the case of a general purpose AI, the entire world. Thirdly, there are certainly unknown variables and setbacks ahead; it's unlikely that in DARPA's case, their benchmarks for cognitive development within an AI will be realised in just 4 years given that

A massive part of the plan is to make use of existing approaches in constructing the final representation, just as the thieves from the heist co-ordinate themselves based on their skills shown in previous smaller heists. It's possible that it may just be a matter of embedding a knowledge base of rules inside a simulation of the world in order to start making measurable progress; in this case it would be by m

In videos about Neural Networks, I've heard the presenter say that 'AI learns in the same way a child does; by showing it examples' but I don't think the two learning styles are the same by any means. Babies and toddlers are highly receptive to learning language because they do nothing but listen for the first 18 – 24 months of their life. This is a considerably longer 'training period' than Neural Networks. Referring back to Meena, it was fully trained in just over a month and all the training data is text-based as opposed to speech which contains the emotions of the speaker.

In addition, a powerful aspect of the human mind is our capability to make reasonable generalisations. Similar to how natural language generation can create new narratives which are ideally linked to the context of the current text body, if a machine could take a set of rules which was missing one and then infer it based on the others in the way bidirectional encoders are given the start and end of a sentence and fill in the missing word, this would be a way of assessing how capable the machine is of coming up with new rules based on old ones and this learning process could be supervised so that it if it comes up with something that is completely irrational and unlike the actual answer, it gets shown the actual answer to complete that rule set OR gets told that it's incorrect, given a clue by perhaps adding a weight to one or more of the rules that are used in generalising the next rule and then coming up with the rule again.

At a Big Data talk during Digital DNA, I remember hearing a quote that was 'If Big Data hasn't solved your problem, you just don't have enough of it yet.' In a way, I suppose Meena is testament to that because of seq2seq's 2.6 Billion parameters and Google's 3.14GB of social media conversations. Imagine being trained on 3.14GB of rules about how the world works as well; that would add so much more context to the conversations and would unlock more of the rationale behind why someone said something.

USEFUL LINKS

Project Mosaic's homepage

https://mosaic.allenai.org/

Mosaic's Computer Vision common sense project

https://visualcommonsense.com/

R2C Common sense paper submitted by Rowan Zellers

https://arxiv.org/abs/1811.10830

Overview of the SWAG dataset

https://rowanzellers.com/swag/#about

Empirical Methods in NLP 2018 Conference Slide – Rowan Zellers

https://drive.google.com/file/d/1vHH9kqufVdTWzFC734VQYd2Gz2DRpIsN/view

Problems with teaching AI Common Sense blog on ML Mastery

https://machinelearningmastery.com/statistical-language-modeling-and-neural-language-models/

How to teach AI Common Sense blog by Wired

https://www.wired.com/story/how-to-teach-artificial-intelligence-common-sense/

Debate between Noam Chomsky and Jean Piaget on learning language being an innate or acquired trait of a baby's brain

https://anthrosource.onlinelibrary.wiley.com/doi/pdf/10.1525/aa.1982.84.1.02a00330

Paper on the development of SWAG

https://www.semanticscholar.org/paper/SWAG%3A-A-Large-Scale-Adversarial-Dataset-for-Zellers-Bisk/af5c4b80fbf847f69a202ba5a780a3dd18c1a027

The Alexandria Project

http://alexandria-project.eu/

Overview of the Chomsky/Piaget debate

https://boingboing.net/2018/11/13/naive-learning.html

NYU

https://cacm.acm.org/magazines/2015/9/191169-commonsense-reasoning-and-commonsense-knowledge-in-artificial-intelligence/abstract