Teaching AI Common Sense (and examples of where it shows absolutely none…)

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”

1. “kicked it in the direction of the net to try to score.” (**Most likely**)
2. “was tackled by an opposing defender.” (**Second-most likely**)
3. “passed it to an opposing defender.” (**Unlikely**)
4. “, 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). (More on these shortly).

Here are links to the

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>