Self Learning AI - AlphaGo

Rushika Ghadge, Pooja More, Sachee Nene

Abstract— An artificial intelligence conquered humanity with its most complex and smartly developed games.But,it is not smart enough to live with its alternative, a self learning artificial intelligence which required no human input and aims to progress in areas such as imagination, memory and learning. This paper talks about Google's attempt to bring intelligence to another level with its development of AlphaGo which uses reinforcement learning for teaching complex movements.It also works as an algorithm to solve many real world problems with a goal to solve intelligence.This is achieved by combining the best techniques from machine learning and system's neuroscience to build powerful general purpose learning algorithms.

Index Terms— Artificial Intelligence(AI), Central Processing Unit(CPU), Convolution Neural Network(CNN), Graphical Processing Unit(GPU),Machine Learning(ML), Neural Network(NN), Tensor Processing Unit(TPU)

1 Introduction

After chess, Go game had been considered as a challenge of Artificial Intelligence. Go game's search space is quite complex and it was thought that it will take humans long time to win this game. But ,Google's DeepMind made an attempt to design and develop a system with altogether new set of algorithms that can challenge a human Go player.

It turned into a reality when this AI programme, AlphaGo beat the professional human Go players,leaving the world amazed [1].

AlphaGo project was started in 2014 as a testing bed to check how Google's DeepMind algorithm of neural network with deep learning could win over Go. Its algorithm is a mixture of search tree, machine learning techniques with reinforcement training of humans and virtual players. Since its win over the premier players, it has boosted the research in the field of AI. It has even proved that NN algorithm and deep learning can be used for other purposes because they were not traditionally meant for playing Go but instead taught on how to play Go. This effectiveness of AlphaGo is its potential to predict its own chances of winning and adjusting the game plan accordingly.

2 How it evolved

AI has made a significant change in a variety of domains from image classification to speech recognition.With this, its algorithms have also evolved year by year.It started off with AlphaGo which got transformed into more better versions and reached upto its fifth version which is AlphaGo Zero. Every version has its own pros and cons & proved to be superior over its predecessor.The evolution of AlphaGo occurred as follows :

2.1 AlphaGo Fan

It was introduced in October 2015.With its development,it became the first computer Go program which defeated the European Go champion Fan Hui on a full-sized board by winning all the five matches.The version used 176 GPUs with Elo rating 3144 on distributed machines instead of CPUs,so the processing improved drastically.

2.2 AlphaGo Lee

This version came into play in March 2016.It made a significant mark when it won over second-ranked Go champion Lee Seedol by winning four matches out of five.It made development over its previous version by using 48 TPUs with Elo rating 3739 on distributed machines than GPU,which was even more faster than CPU and GPU.

2.3 AlphaGo Master

It was introduced in December 2016.This used 4 TPUs with Elo rating 4858 on a single machine and it was 3 times stronger than AlphaGo Lee.It was also used in a teaching tool developed by DeepMind to analyze rates of winning of various Go openings,which were unable to play real games.

2.4 AlphaGo Zero

This version skips to train human amateur on how to play Go,and rather learns playing by competing against itself.It makes use of reinforcement learning,by becoming its own teacher.It uses 4 TPUs on a single machine with Elo rating 5185.This technique is more stronger because it is not limited to human knowledge.

3 How it Works

AlphaGo uses supervised learning algorithm where input to it is a training set along with their labels.A training set here is the current game state of Go board, and training labels could be win or loss for current player depending on game state [2].

It uses NN which has multiple layers.Each layer contains a number of neurons.Each neuron receives an input which is the output of neurons from the previous layers. The inputs are added together and are further processed by a nonlinear activation function.

AlphaGO uses CNN which is a subtype of neural network.It is used to process images.

Image is given as an input to convolution network, in this case AlphaGo uses Go games board as an image. At every layer the image goes through a series of filters. The network restricts itself to filtering operations and thus are efficient on processing image data. It is a feedforward network. After a series of filtering operations the output of convolution

network is obtained. This network is used for the applications like face recognition etc. where pattern recognition is needed.No search operation or backtracking is performed.

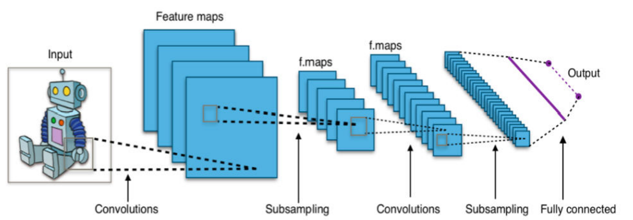

Fig. 1. Working of Convolution Network which has filters in each layer.

AlphaGo uses a tree search procedure and a convolution network.Convolution network works along with the tree search procedure.Tree search procedure uses a brute force approach and CNN are designed to learn on their own.
Three of the convolution networks are trained which includes two policy network and one value network. Input to these networks is current state of game.

The value network returns the probability of a win for a particular player depending on the current state of game.It provides answer to the question "Whether the current player can win the game depending on current game state?". Input given to value network is 'current game-state' and output is single value which returns probability of a player to win the game.

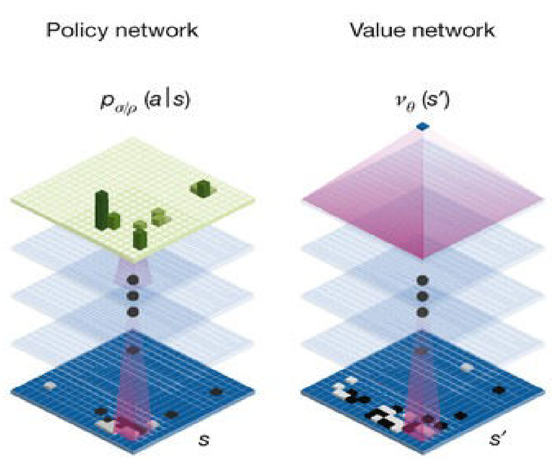

Fig. 2. Representation of policy network and value network.

The policy network returns the most promising move which has high chances to win the game.The output of policy network is larger in size which is directly proportional to boards size. It gives probability value for moves depending on current board state.Moves with higher probability is selected.Both networks are trained.The policy network was trained by providing data of 30 million moves which was played by human experts. For this it used data at the KGS Go Server. It gave an accuracy of 57% to predict the human moves. The policy network played against itself and its result which has 30 million game move was used to train the value network.Value network is like an evaluation function but it is learned rather than designed.

4 Applications

Apart from gaming, DeepMind also plans to venture in other fields,such as[3]:

4.1 Health Service

DeepMind has made a collaboration with UK National Health Service so that the algorithms can be applied to clinical data to improve diagnosis or treatment plans.

4.2 Smartphones

Using DeepMind's algorithm in smartphone's will give the user an altogether different and personalized experience.

4.3 E-commerce

To improve the user's experience by recommending products to them as the interaction of the site visitor can be seen as a sequence of turns with the action to click on the link or not.

5 Challenges in Applying AlphaGo's Technology

As it is said a coin has two sides, advancement of technology is the need of the hour but along with it comes the real world problems and the need to tackle them so that the machine can work in an unsupervised way.Identifying what makes up the state of the world,the actions to be carried out and then representing the states and actions in machine's memory are considered as the problems faced in modelling the world into format the machine can work[4].These states here can be equivalent to Go board's move.In Go they can be very low and poorly structured.
So,here Deep Learning can be used to extract useful representations,but there are some prerequisites :

- The type of network needs to be chosen,the interconnection pattern between different layers of neurons and size of the neuron.
- We also need examples of previous states of the world,the actions and the decisions taken by the human i.e the training data.
- However we are not often sure which state will arrive after which action and so we need to look at the uncertainty as well.

6 Future Scope

AlphaGo's algorithm makes the task of searching complex

data simpler.With this,the team aims to develop more advanced algorithms that might help scientists to overcome issues like reducing energy consumption,inventing raw materials or finding cures for diseases.A teaching tool is also aimed to be developed to show the analysis of Go's positions,how the program thinks and proving with an insight to see the game from Go's perspective.

7 Conclusion

AI has this potential to help our society cross paths through new knowledge which is beneficial and AlphaGo can make this possible. AlphaGo as a competitor has inspired Go players to try new strategies in this 3000 year-old game.
AlphaGo Zero its next version is a generalised algorithm which can be applied to solve real time problems.

References

[1] "AlphaGo",https://www.techopedia.com/definition/31955/alphago, 2016.

[2] "DeepMind",https://deepmind.com/research/alphago",2016.

[3] Elizabeth Gibney, "What Google's winning Go algorithm would do next",http://www.nature.com/news/what-google-s-winning-go lgorithm-will-do-next-1.19573 ,2016.

[4] James Vincent. "DeepMind's Go-playing AI doesnt need human help to beat us anymore",www.theverge.com/2017/10/18/16495548/deepmind-ai-go-alphago-zero-self-taught,2017.