

Survey on Video game AI based on Deep learning

Wangzhihui Mei 2019124044

CCNU-UOW JI

Abstract

1 Introduction

Great progress has been made in artificial intelligence technology in recent years. In 2016, Alphabet AlphaGo, a computer Go program developed by DeepMind, successfully defeated Li Shishi, who has been considered the world's top player for nearly 15 years. AlphaGo's success is extremely. This is mainly due to its use of deep learning algorithms. This article introduces the application of deep learning in games from a broader perspective.

2 Deep Learning

Deep learning is a machine learning model that has shined in recent years. Its main method is to achieve better learning results by training multiple layers of neural networks. Common multilayer network structures include multilayer perceptrons and convolutional neural networks. Networks and recurrent neural networks, etc. A method of hierarchical training using unsupervised learning is applicable to some deep network structures [78]. Later, researchers at the University of Montreal Analyze the help of unsupervised learning for deep structures [9] and the reasons for the failure of original training methods [10], and propose parameter initialization methods [10] and activation functions [11] suitable for deep structures. Breaking through the bottleneck of training algorithms and computing power, deep learning is widely used in artificial intelligence-related Field, and has made great progress on a number of research issues. Typical application scenarios include Image classification [12], Object detection [1314], Video classification [15], Scene parsing [16] and Shadow detection [17], Speech recognition in Speech understanding [18], Prosody prediction and Text-to-speech syn-thesis [2021], Parsing in natural language processing [22], Machine translation [2324], and Contextual entity linking [25] and Sentiment analysis [26] and Information retrieval [2728] in data mining.

In general, in the presence of a large number of training samples, deep learning often performs well on prediction problems. Why is deep learning so effective compared to traditional supervised learning models? We think that the reason can be roughly attributed to three points: the idea of end-to-end learning, the ability of non-linear learning, and scalability in the face of large-scale data

2.1 End-to-end Learning

The traditional machine learning process is often divided into multiple independent modules. Error propagation is not transmitted between these modules during the training process, so the previous module will not be adjusted according to the training results of the subsequent modules. For example, in a typical natural language processing problem, traditional methods will take separate steps such as tokenization, POS tagging, parsing, and semantic analysis. In such a process, Each step is an independent learning task, which requires a large number of labeled training samples. It is expensive to label a large number of training samples for different learning tasks, and the wrong predictions made by each step (regardless of the cause) will also be Affects subsequent tasks. Even in a single learning task (such as syntactic analysis), feature extraction is often a preprocessing step independent of training. In deep learning processes, trainable hierarchical representations (Trainable hierarchical representation) replaces predefined features Representation, and the representation of each layer will be adjusted during training according to the error information transmitted by the subsequent layers, which is conducive to the optimization of the objective function [31]. On the other hand, in each layer of training in deep learning Does not depend on additional learning targets and training samples (Of course, deep learning can also use additional labels to provide a secondary objective function for the training of the middle layer). This not only allows limited resources to be preferentially used to label training samples for the final learning target , Also the flexibility to use ready-made labeled data sets

2.2 Non-linear Learning

Deep learning builds a mapping function from the original data input to the final prediction target by stacking multiple layers of neural networks. Traditional machine learning models are good at finding linear transformations of data to output targets, but in reality, from data input to The mapping of target output is often complex and non-linear. In deep neural networks, each layer is a non-linear mapping from input to output. These multi-layer non-linear transformations form a hierarchical representation of the data, As the number of layers increases, the representation is also more abstract and more invariant. For example, for image data, the deep learning structure first extracts the representation of the object boundary, and the subsequent network layers extract the relevant information from the representation of the boundary. The feature of the object part, and the feature of the object part is then used as the input of the next layer of the neural network to obtain the feature vector of the object [32]; for text data, the deep learning structure first extracts the word feature representation (Word representation), and then The network layer extracts the

features of the sentence from the features of the word through non-linear transformation, and the features of the sentence are The feature is used as the input of the next layer of the neural network to obtain the feature vector of the article. Although theoretically, a neural network structure with more than two layers can express any function, but the deeper network structure is more efficient for a specific set of functions Expressive ability, which can express complex functions with fewer parameters (such as a few network nodes) [3336]. This efficient expressiveness is necessary for complex artificial intelligence tasks. The expression of deep networks Advantages have obvious help for neural network learning: variables at the bottom and high levels can share statistical features, thereby improving learning efficiency. Because there is no such hierarchical feature representation in the two-layer neural network, we generally don't consider the two-layer The network is a deep learning structure. Similarly, non-linear models built directly on the original input space, such as decision trees and support vector machines (non-linear Kernel), are not considered deep learning models [36].

2.3 Scalability

Learning complex, non-linear functions, of course, requires more training samples. Too few training samples can lead to serious over-fitting problems. This not only makes how to obtain a large amount of labeled data a key, it is also important for machine learning. Faced with the scalability of a large number of training samples, a test was proposed. Although traditional machine learning algorithms (such as SVM (Support Vector Machine) and Kernel method) can also be used to learn non-linear transformations, such computational complexity is very High: At least $O(N^2)$, N is the number of training data samples. When the amount of data is too large, this algorithm is obviously difficult. The training methods of deep learning are mostly based on stochastic gradient descent, not only need not be calculated The pairwise relationship of the training samples often does not even need to traverse all the training samples, so that a larger data set can be flexibly used. At the same time, the backpropagation algorithm can quickly calculate the gradient of the entire network and allow training errors Effectively propagate to the underlying feature space [37]. With the increase of training samples, the performance of traditional supervised learning algorithms The phenomenon of diminishing return appears. On the contrary, the effect of deep learning algorithms can be significantly improved with the increase of trainable data sets. This advantage comes from the flexible expression ability of deep network structure, so that deep learning can Express complex set of functions with simple structure [5, 38].

3 Deep Learning in Game-playing

Game-playing is a crucial problem of artificial intelligence. A game that can objectively determine the winner not only provides a perfect test platform for artificial intelligence algorithms, but also allows

computer players and human players to compare against each other. According to different classification methods, games can be divided into single People games and two-player (or multiplayer) games, board games and video games, cooperative games and battle games, etc. Two-player games or more are often called games, and games can be divided into perfect information games. games (such as Go and Chess) and incomplete information games (such as poker and military chess). Early artificial intelligence algorithms often relied on search algorithms, which is very common in simple, limited-search games (such as Tic-Tac-Toe) Effective. As search difficulty increases, more complex games often use algorithms that combine search with machine learning, especially algorithms for reinforcement learning. Reinforcement learning itself is not deep learning. It is an important branch of machine learning , Which focuses on solving sequential decision problems [3940]. The most advanced algorithms in many games are currently Based on reinforcement learning. For example, the most effective Go algorithms before AlphaGo (such as CrazyStone and Zen) are a combination of reinforcement learning and Monte-Carlo tree search (MCTS). Compared to unsupervised learning And supervised learning, reinforcement learning focuses on two unique issues: Exploration vs. exploitation and Temporal credit assignment. A reinforcement learning algorithm needs to answer two basic questions:

1. How Evaluate a policy;
2. How to find the optimal strategy for a problem.

The application of deep learning in games is often to better solve these two problems by assisting reinforcement learning.

3.1 Deep Reinforcement Learning

Reinforcement learning traditional research focuses on learning tabular representation or linear function approximation [39]. For large-scale and complex sequential decision-making processes in reality, simple tabular representation and linear approximation It is not enough. Deep learning can provide end-to-end, non-linear function approximation for reinforcement learning, so that reinforcement learning can solve more realistic and complex problems such as how to express the state of the chess board in Go. On the other hand, in order to solve Partially observable Markov decision problems (POMDP) are commonly observed in games.Reinforcement learning algorithms need to effectively process sequences of actions and observations (such as effectively summarizing historical actions and observations). Is a state representation) to find the optimal strategy. In the fields of audio, video, and text, deep learning has been proven to successfully learn sequence representations, so it is also used to learn state representations in POMDP [41 43]. Reinforcement learning algorithms optimized by deep learning, commonly known as deep learning Degree reinforcement learning. A review of deep reinforcement learning can be found in [30].

3.2 Deep Learning applied in Video Game

In recent years, deep reinforcement learning has been increasingly used in games, including multiple single-player and multiplayer games. As mentioned earlier, the advantage of deep learning is that it can learn non-linear from a large amount of training data. Representation, so as to achieve better prediction results. When deep learning is applied in a new field, we need to pay attention to three questions: What is the goal of the prediction? Where does the training data come from? What does the learning result represent?

4 Methodologies

We take a class of games commonly used in the literature as an example to sort out the development of deep reinforcement learning in games. We found that when deep learning is introduced to games, its prediction target is often every possible under a certain game state. Values or probabilities (e.g., valuations and strategies) corresponding to the actions of the player, the training data is often derived from the process of the computer playing the game (to be precise, the state-action-reward record in sequence), and the learning results Often a non-linear representation of the state and strategy of the game.

4.1 Atari Games

Arcade learning environment (ALE) is a common evaluation platform for artificial intelligence researchers. Its role is similar to the ImageNet challenge in the field of image classification [44]. ALE provides an Atari 2600 (ATARI 2600) simulator And about 50 games (mostly single-player games) [44]. These games are built on the same screen of 160 pixels wide and 210 pixels high, each pixel has 128 colors. Each game has different actions Space, including up to 18 possible actions. To play these games well, a successful algorithm needs to simultaneously solve the two challenges of game state representation and strategy selection, which is almost tailored for deep reinforcement learning: Deep reinforcement learning algorithms can Learn the representation of the current game state from a high-dimensional, but also partially observable screen, and learn the optimal strategy from sparse, highly delayed reward information (Reward).

Bellemare et al. Of the University of Alberta first used the linear function approximation SARS algorithm and the following four new general artificially designed feature sets on ALE: 1) First divide the screen into blocks of the set, using one for each block The vector indicates whether each color appears, and the set of these vectors is used as the basic (BASIC) feature set of the current game screen; 2) The paired combination of basic features is added to the basic feature set, and these combinations constitute the BASS feature set; 3) First extract the objects on the screen, and classify the objects on the screen by clustering, and use the information between multiple frames to infer the position and speed of these objects. The category, position and speed of all objects on the screen constitute the current game The DISCO feature set of the picture; 4) Use locality-sensitive hashing for the game picture, and use the resulting low-dimensional

representation as the LSH feature set [44]. In their subsequent work, they proposed a method of Contingency awareness, Based on the original feature set, additional features are added to indicate which elements in the screen are directly input by the player [46]. Belle-mare et al. Subsequently proposed a method to further expand the feature set by using tug-of-war sketch [47]. But in general, the reinforcement learning algorithm based on linear value function approximation is far away Far weaker than human players. In addition, manually designing feature functions is not an easy task

Prior to 2013, the successful application of neural networks in strong chemistry practice was quite limited. The one exception was that Tesauro of the IBM Watson Research Center successfully trained neural networks with reinforcement learning to solve backgammon [48].

DeepMind's Wang et al. Proposed a new "dueling" neural network architecture (Dueling network). This architecture separates the value function of the game state and the dominance function of state-related actions, which makes the valuation of each action no longer independent. Different Actions sharing more generalized state-value functions are useful to reduce the estimation bias for different actions [57].

Although not a familiar game, Atari Games provided an excellent soil and evaluation platform for the development of deep learning in games. Many of these ideas and techniques were then applied to other game problems.

4.2 Go

Go is one of the most sophisticated complete information games in existence. Compared to other board games (such as chess and checkers), the search space for Go is huge, and the description and evaluation of the situation are extremely difficult. With the convolutional neural network in The Atari game has achieved great success, and people expect the same method to be used to solve computer Go. However, it turns out that it is difficult to directly draw a gourd to approximate the strategy or value function of computer Go. Investigate its The reason is that the strategy and situation judgment of Go is too complicated. Because the true value of each chessboard state is unknown, if there is not a good enough initial strategy as the basis of reinforcement learning, it is trained using inaccurate movements and reward information. Neural networks may not be able to make reliable predictions. To solve this problem, researchers have thought of using human chessboards as training sets for deep neural networks, because in the human chessboard, the choice of each move and the outcome of each game are Known. Researchers expect that strategies learned from human chess records are more suitable as initial strategies for reinforcement learning.

Clark et al. From the University of Edinburgh used the historical chess records of human players to train convolutional neural networks, so that they can predict the strategy of human players (that is, the next move in the current situation). Such neural networks are also called strategic networks. 41.1% and 44.4% prediction accuracy rates were obtained on each data set. These strategic networks use feature sets designed specifically for Go. For example, symmetric information is hard-coded into the input of a deep neural network, and

unreasonable positions on the board Blocked. Even if the prediction accuracy is less than 50%, the trained convolutional neural network can defeat the well-known Go program GnuGo (but lose to the more advanced Go program Fuego) [61].

Researchers at the University of Toronto such as Maddison et al. And DeepMind trained a 12-layer deep convolutional neural network with a historical chessboard of a network battle platform (KGS). Its prediction strategy for human chess players can reach 55% accuracy. Train well Convolutional God of Meridian battle against GnuGo can achieve a 97% win rate, and can match the most advanced Monte Carlo tree search algorithm that simulates 2 million times per step [62]. Similar methods have also been used by Tian and other Facebook developers Computer Go program in DarkForest [63].

It is worth mentioning that there is a bias in the strategic network trained from human expert chess records. This is because the training goal (predicting the next hand) and the actual goal (winning the game) are inconsistent. To solve this problem, AlphaGo will The strategy network trained with chess records is then adjusted through self-play. The idea is the same as that of the Atari game using a self-combat sequence to train the strategy network. [64]

The strategy network trained from the human chessboard makes the basic strategy good enough, which makes the state-action-reward information observed by the reinforcement learning algorithm in the self-matches closer to the real situation of the optimal strategy (Ground-truth) AlphaGo then used the state of the game and the outcome of the game to train a value network to predict the value of any board state. This value network further improves AlphaGo's chess power. In order to make the training data as independent as possible, Al-phaGo Only one position is selected for training in Pan Zi war [64]. This idea is similar to the Atari game where the selection of experience playback is as uniform and random as possible.

On the other hand, Monte Carlo tree search is still the most advanced algorithm for processing large and complex sequential decisions, including computer Go. As the number of random simulations increases, the search tree becomes larger, and the value estimation also changes. More accurate. However, its computational overhead is too large for the actual Go game. Therefore, researchers have also tried to use deep learning (or simpler linear learning) to help Monte Carlo tree search to improve the The accuracy of the value prediction or the computational cost of the search is reduced. AlphaGo introduces a new Monte Carlo tree search algorithm. This algorithm combines Monte Carlo simulation and the value and strategy network to reduce the computational cost. The value network is used to evaluate the current chessboard. State (that is, predicting the value of a given position), which effectively reduces the depth of the search plan. The strategic network is used to select the next step (that is, predicting the probability distribution of the next step of a given position), thereby effectively Reduced the search width. The training of these networks is the precision of supervised learning of human expert actions and reinforcement learning of self-play Co. AlphaGo take this beating European champion and the world's top players go Shishi[64].

4.3 Poker

The challenge of poker games comes from incomplete information, that is, players can only partially observe historical events, but can't see the information of their opponents. Bowling et al. From the University of Alberta proposed an approximate Nash equilibrium solution for Texas Hold'em. The basic idea of this scheme is to repeatedly iterate the self-match between two Regret minimization algorithms [65]. Recently, researchers from Yakovenko et al. And Columbia University trained a convolutional neural network for poker in a similar way, and used it in Three common poker games can compete with human experts [66]. Heinrich et al. And Silver et al. Of DeepMind have proposed deep reinforcement learning algorithms that are more suitable for incomplete information games, which they call "neural virtual self-play" (Neural fictitious self-play (NFSP)). The main idea of NFSP is to approximate the classic "fictitious play" model in game theory, which does not rely on prior knowledge and can be used in a two-person zero-sum game (Zero-sum games) or multiplayer potential games (Potential games) by converging to Nash equilibrium through self-matching [67].

4.4 Other Games

Atari, Go, and poker are just the tip of the iceberg in computer games. Interestingly, deep learning has not been widely used in other classic games. This may be because computers have long been defeated in most classic games (such as chess and checkers). Human masters. This has to some extent lost researchers' interest in further improving algorithms. Even so, the idea of deep learning or broader supervised learning is scattered in other games. In addition to the backgammon mentioned earlier, Draughts program Chinook first extracts and builds a database of opening strategies from the chess records of human experts, and then selects the optimal action through the alpha-beta search algorithm and the strategy evaluation function for leaf nodes [68]. Recent research hotspots Gradually shifted from board games to video games, such as Microsoft using the game "Minecraft" as a research platform for testing artificial intelligence, and DeepMind claims that their next challenge is StarCraft. Compared to board games, these The status of real-time strategy games is more complex, and the information is more incomplete. The choice is also greater.

5 Conclusion

Gaming has always been an important branch of artificial intelligence. The success of deep learning in other fields has brought unprecedented inspiration to game artificial intelligence. For example, computer Go experienced the earliest rule-based algorithms, heuristic-based evaluation algorithms, and Monte Carlo tree search algorithms, reinforcement learning-based algorithms, and ultimately qualitative leaps from deep reinforcement learning. In the near future, we will see deep learning algorithms and ideas being used in more and more games. The future research direction should gradually shift from classic games to more

complex and more incomplete multiplayer games, especially video games. There are still many problems to be solved in deep learning in games, such as how to put the knowledge that has been learned Applied to new problems, how to use expert knowledge more effectively, how to learn new knowledge in self-matches, how to change game strategies based on opponents, etc. We also look forward to seeing successful game algorithms being applied in various industries (such as Medical and education), and really improve the lives of ordinary people.

References