

AI for System and System for AI

Conless Pan

2023 年 7 月 24 日

1 Introduction

2 AI for System

3 System for AI

The Rise of Machine Learning

Midjourney: Art in the Age of Artificial Intelligence



The Rise of Machine Learning

The screenshot shows the ChatGPT interface. On the left is a sidebar with a search bar at the top, followed by a "Today" section containing a list of interactions. Below that is a "Previous 7 Days" section with a similar list. At the bottom of the sidebar are buttons for "Upgrade to Plus" and "New". The main area has a header "ChatGPT" with two tabs: "GPT-3.5" and "GPT-4". Below the header are three sections: "Examples", "Capabilities", and "Limitations", each with a list of bullet points. A message input field at the bottom says "Send a message" and includes a "NEW" badge. A footer at the bottom of the page contains a copyright notice and navigation icons.

Today

- ACM Growth: Change & Efficiency

Previous 7 Days

- Upbeat Study Buddies
- 提供HTML版本信息
- Shell Usage: \$0 and ""
- 提取网址至输入文件
- Shell Task: Summarize Request
- 正则表达式概述
- New chat
- APP观察脚本编写 (4 words)

Previous 30 Days

- Data Structures for CS
- New chat
- Financial Aid Application Requirements
- A*在自动驾驶中的重要性
- CPU温度解决方案
- 环法自行车赛结果未知
- 提取微信文章至Word

Upgrade to Plus NEW

conlesspan@outlook.com ...

ChatGPT

Examples

- "Explain quantum computing in simple terms" →
- "Got any creative ideas for a 10 year old's birthday?" →
- "How do I make an HTTP request in Javascript?" →

Capabilities

- Remembers what user said earlier in the conversation
- Allows user to provide follow-up corrections
- Trained to decline inappropriate requests

Limitations

- May occasionally generate incorrect information
- May occasionally produce harmful instructions or biased content
- Limited knowledge of world and events after 2021

Send a message

Free Research Preview. ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT July 20 Version

The Rise of Machine Learning

When we talk about the rise of machine learning, people usually raise these questions:

The Rise of Machine Learning

When we talk about the rise of machine learning, people usually raise these questions:

- What is machine learning?

The Rise of Machine Learning

When we talk about the rise of machine learning, people usually raise these questions:

- What is machine learning?
- Do you know its history?

The Rise of Machine Learning

When we talk about the rise of machine learning, people usually raise these questions:

- What is machine learning?
- Do you know its history?
- Why is it so important today?

The Rise of Machine Learning

When we talk about the rise of machine learning, people usually raise these questions:

- What is machine learning?
- Do you know its history?
- Why is it so important today?

But I don't want to talk about them, cause I'm not interested about AI.

Applications of Machine Learning

Anyway, AI is a useful tool.

Applications of Machine Learning

Anyway, AI is a useful tool.

- Generative AI

Applications of Machine Learning

Anyway, AI is a useful tool.

- Generative AI
- AI for science

Applications of Machine Learning

Anyway, AI is a useful tool.

- Generative AI
- AI for science
- Others

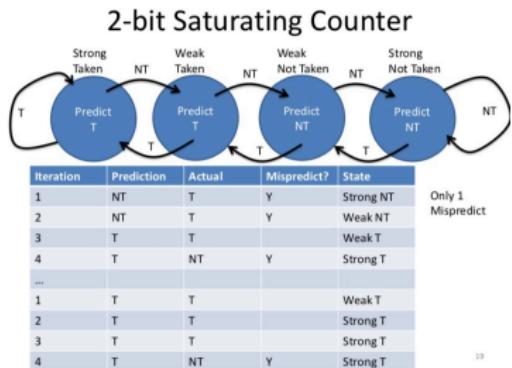
How can AI promote our research of computer system?

Let's compare these two games:¹

¹ smith1998study.

How can AI promote our research of computer system?

Let's compare these two games:¹

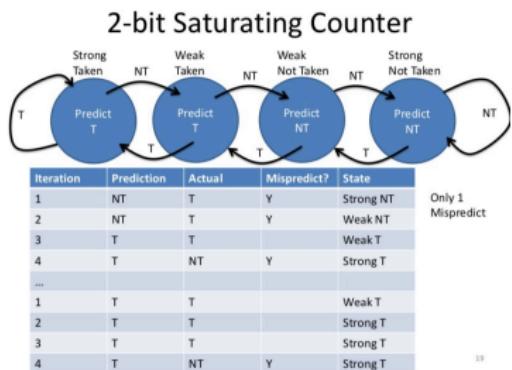


13

¹ smith1998study.

How can AI promote our research of computer system?

Let's compare these two games:¹



¹ smith1998study.

How can AI promote our research of computer system?

And it turns out that...²³

²**zouzias2021branch.**

³**silver2017mastering.**

How can AI promote our research of computer system?

And it turns out that...²³

The screenshot shows a red header with the arXiv logo and navigation links for 'Search...', 'Help | Advanced'. Below is a grey header with 'Computer Science > Machine Learning' and a submission date 'Submitted on 25 Jun 2021'. The main title is 'Branch Prediction as a Reinforcement Learning Problem: Why, How and Case Studies' by 'Anastasios Zouzias, Kleovoulos Kalaitzidis, Boris Grot'. The abstract discusses stagnating improvements in branch predictor (BP) efficacy and a lack of fresh ideas in design, calling for RL. It argues that RL facilitates systematic reasoning about BP designs. The text is in a monospace font.

Comments: 6 pages, appeared in ML workshop for Computer Architecture and Systems 2021

Subjects: Machine Learning (cs.LG); Artificial Intelligence (cs.AI)

Cite as: arXiv:2106.13429 [cs.LG]

or arXiv:2106.13429v1 [cs.LG] for this version

<https://doi.org/10.48550/arXiv.2106.13429>

²[zouzias2021branch](#).

³[silver2017mastering](#).

How can AI promote our research of computer system?

And it turns out that...²³

arXiv > cs > arXiv:2106.13429

Computer Science > Machine Learning

[Submitted on 25 Jun 2021]

Branch Prediction as a Reinforcement Learning Problem: Why, How and Case Studies

Anastasios Zouzias, Kleovoulos Kalaitzidis, Boris Grot

Recent years have seen stagnating improvements to branch predictor (BP) efficacy and a dearth of fresh ideas in branch predictor design, calling for fresh thinking in this area. This paper argues that looking at BP from the viewpoint of Reinforcement Learning (RL) facilitates systematic reasoning about, and exploration of, BP designs. We describe how to apply the RL formulation to branch predictors, show that existing predictors can be succinctly expressed in this formulation, and study two RL-based variants of conventional BPs.

Comments: 6 pages, appeared in ML workshop for Computer Architecture and Systems 2021

Subjects: Machine Learning (cs.LG); Artificial Intelligence (cs.AI)

Cite as: arXiv:2106.13429 [cs.LG]

(or arXiv:2106.13429v1 [cs.LG] for this version)

<https://doi.org/10.48550/arXiv.2106.13429>

nature

Explore content ▾ About the journal ▾ Publish with us ▾ Subscribe

nature > articles > article

Published: 19 October 2017

Mastering the game of Go without human knowledge

David Silver✉, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel & Demis Hassabis

Nature 550, 354–359 (2017) | [Cite this article](#)

345k Accesses | 4205 Citations | 2528 Altmetric | [Metrics](#)

²[zouzias2021branch](#).

³[silver2017mastering](#).

How can AI promote our research of computer system?

Their similarities:

How can AI promote our research of computer system?

Their similarities:

- ① Decision-making problems

How can AI promote our research of computer system?

Their similarities:

- ① Decision-making problems
- ② A finite state given as input

How can AI promote our research of computer system?

Their similarities:

- ① Decision-making problems
- ② A finite state given as input
- ③ A simple decision required as output

How can AI promote our research of computer system?

Their similarities:

- ① Decision-making problems
- ② A finite state given as input
- ③ A simple decision required as output

The role that AI plays: looking for a fitting function $y = f(x)$, which calculates the correct/best y with given x .

How does AI perform its job?

The perceptron is introduced at first:⁴

⁴jimenez2001dynamic.

How does AI perform its job?

The perceptron is introduced at first:⁴

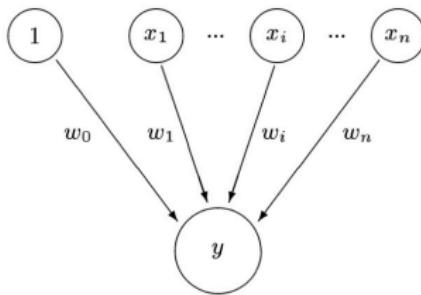


Figure 1: Perceptron Model. The input values x_1, \dots, x_n , are propagated through the weighted connections by taking their respective products with the weights w_1, \dots, w_n . These products are summed, along with the bias weight w_0 , to produce the output value y .

⁴jimenez2001dynamic.

How does AI perform its job?

The perceptron is introduced at first:⁴

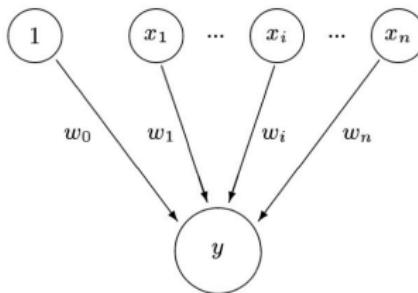
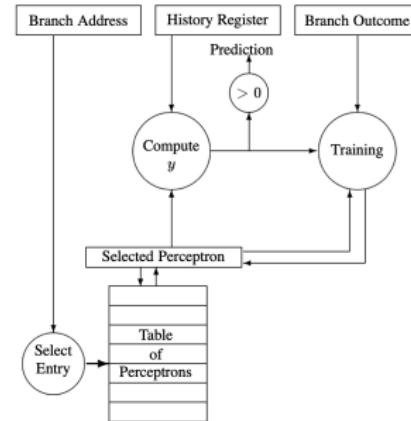


Figure 1: Perceptron Model. The input values x_1, \dots, x_n , are propagated through the weighted connections by taking their respective products with the weights w_1, \dots, w_n . These products are summed, along with the bias weight w_0 , to produce the output value y .



⁴jimenez2001dynamic.

Applications

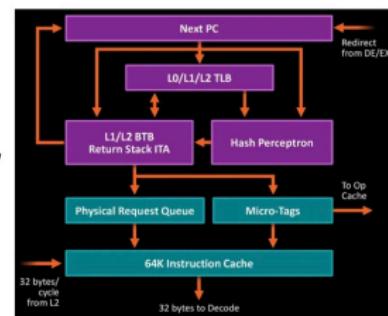
AMD Zen microarchitecture, the start of "AMD YES".

Front End [edit]

The Front End of the Zen core deals with the **in-order** operations such as **instruction fetch** and **instruction decode**. The instruction fetch is composed of two paths: a traditional decode path where instructions come from the **instruction cache** and a **μ OPs cache** that are determined by the **branch prediction** (BP) unit. The instruction stream and the branch prediction unit track instructions in 64B windows. Zen is AMD's first design to feature a **μ OPs cache**, a unit that not only improves performance, but also saves power (the **μ OPs cache** was first introduced by **Intel** in their **Sandy Bridge** microarchitecture).

The **branch prediction** unit is decoupled and can start working as soon as it receives a desired operation such as a redirect, ahead of traditional instruction fetches. AMD still uses a **hashed perceptron system** similar to the one used in **Jaguar** and **Bobcat**, albeit likely much more finely tuned. AMD stated it's also larger than previous architectures but did not disclose actual sizes. Once the BP detects an indirect target operation, the branch is moved to the Indirect Target Array (ITA) which is 512 entry deep. The BP includes a 32-entry return stack.

In Zen, AMD moved the instruction TLB to BP (to much earlier in the pipeline than in previous architectures). This was done to allow for more-aggressive prefetching by allowing the physical address to be retrieved at an earlier stage. The BP is capable of storing 2 branches per BTB (Branch Target Buffer) entry, reducing the number of BTB reads necessary. ITLB is composed of:



Advanced version

Advanced neural network is also being introduced to many subfields of computer system...

Reinforcement is what you need

Let's go over the basic concept of reinforcement learning.

When agent receives a reward or a feedback, it updates its estimation or probability

$$\mathbb{E}(r_{a_i|s_i}), \mathbb{P}(r_{a_i|s_i} > 0)$$

Reinforcement is what you need

Let's go over the basic concept of reinforcement learning.

- A virtual agent who makes decisions

When agent receives a reward or a feedback, it updates its estimation or probability

$$\mathbb{E}(r_{a_i|s_i}), \mathbb{P}(r_{a_i|s_i} > 0)$$

Reinforcement is what you need

Let's go over the basic concept of reinforcement learning.

- A virtual agent who makes decisions
- A state space $S = \{s_i\}$

When agent receives a reward or a feedback, it updates its estimation or probability

$$\mathbb{E}(r_{a_i|s_i}), \mathbb{P}(r_{a_i|s_i} > 0)$$

Reinforcement is what you need

Let's go over the basic concept of reinforcement learning.

- A virtual agent who makes decisions
- A state space $S = \{s_i\}$
- An action space $A = \{a_i\}$

When agent receives a reward or a feedback, it updates its estimation or probability

$$\mathbb{E}(r_{a_i|s_i}), \mathbb{P}(r_{a_i|s_i} > 0)$$

Reinforcement is what you need

Let's go over the basic concept of reinforcement learning.

- A virtual agent who makes decisions
- A state space $S = \{s_i\}$
- An action space $A = \{a_i\}$
- Rewards $r_{a_i|s_i}$

When agent receives a reward or a feedback, it updates its estimation or probability

$$\mathbb{E}(r_{a_i|s_i}), \mathbb{P}(r_{a_i|s_i} > 0)$$

How can our works on computer system promote research of AI?

Key idea: build systems to adapt, simplify and accelerate machine learning.