

Modelos Neurais mais Transparentes e Compactos usando Esparsidade

Gonalo M. Correia

Prémio Vencer o Adamastor, Fevereiro 2023

Introdução

Adaptively Sparse Transformers

Efficient Marg. of Discrete Latent Variables via Sparsity

Impacto e Conclusões

O sucesso das redes neuronais

O sucesso das redes neuronais

- Modelos poderosos de aprendizagem automática

O sucesso das redes neuronais

- Modelos poderosos de aprendizagem automática
- Capacidade de aprender representações complexas

O sucesso das redes neuronais

- Modelos poderosos de aprendizagem automática
- Capacidade de aprender representações complexas
- Aplicações em áreas como visão computacional, processamento de linguagem natural, etc.

O sucesso das redes neuronais

- Modelos poderosos de aprendizagem automática
- Capacidade de aprender representações complexas
- Aplicações em áreas como visão computacional, processamento de linguagem natural, etc.
- Resultados de ponta

Artificial intelligence beats eight world champions at bridge

**Victory marks milestone for AI as
bridge requires more human skills
than other strategy games**

des neuronaux

Artificial intelligence beats eight world champions at bridge

**Victory marks milestone for AI
bridge requires more human skill
than other strategy games**

des neuronaux

AI 'outperforms' doctors diagnosing breast cancer



Fergus Walsh

Medical correspondent

@BBCFergusWalsh

Artificial intelligence

beats
chan

Victory
bridge
than ot

des neuronais

Expresso50

Conversar com uma máquina
nunca foi tão real: chatGPT dá
conselhos, escreve poemas, lê o
Expresso e gosta de futebol



Fergus Walsh

Medical correspondent

@BBCFergusWalsh

Limitações

Limitações

- Quantidade de dados necessários

Limitações

- Quantidade de dados necessários
- Dificuldade de interpretar os resultados

Limitações

- Quantidade de dados necessários
- Dificuldade de interpretar os resultados
- Quantidade de recursos computacionais necessários

Limitações



**Harvard
Business
Review**

AI Can Outperform Doctors. So Why Don't Patients Trust It?

by Chiara Longoni and Carey K. Morewedge

Limitações

≡ WIRED

SUBSCRIBE

AI Can Do Great Things—if It Doesn't Burn the Planet

The computing power required for AI landmarks, such as recognizing images and defeating humans at Go, increased 300,000-fold from 2012 to 2018.

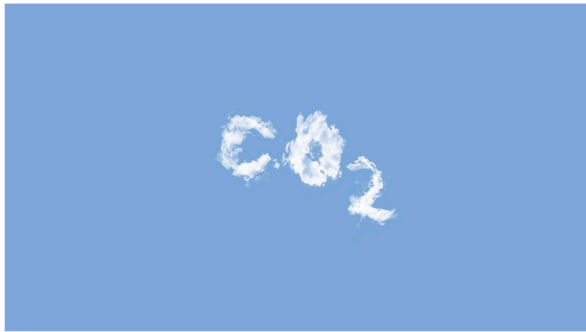


AI Can Outperform Doctors. So Why Don't Patients Trust It?

by Chiara Longoni and Carey K. Morewedge

Limitações

The carbon footprint of ChatGPT



≡ WIRED

SUBSCRIBE

Do Great Things—if It Burn the Planet

power required for AI
such as recognizing images and
games at Go, increased 300,000-
fold to 2018.

perform Why Don't Just It?

by Chiara Longoni and Carey K. Morewedge

Objetivos do trabalho

**Modelos Neurais
mais Transparentes e Compactos
usando Esparsidade**

Objetivos do trabalho

**Modelos Neurais
mais **Transparentes** e Compactos
usando Esparsidade**

Objetivos do trabalho

Modelos Neurais
mais Transparentes e Compactos
usando Esparsidade

Objetivos do trabalho

Modelos Neurais
mais Transparentes e Compactos
usando Esparsidade

Introdução

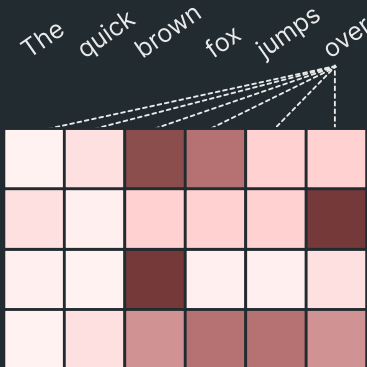
Adaptively Sparse Transformers

Efficient Marg. of Discrete Latent Variables via Sparsity

Impacto e Conclusões

Tornar a rede esparsa

LMs usam representações **densas**.

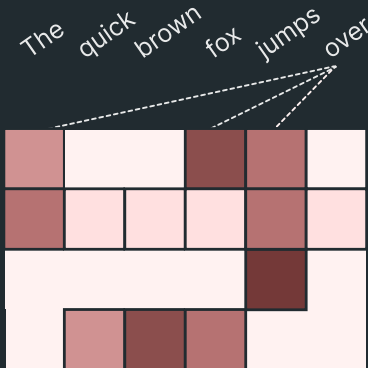


Tornar a rede esparsa

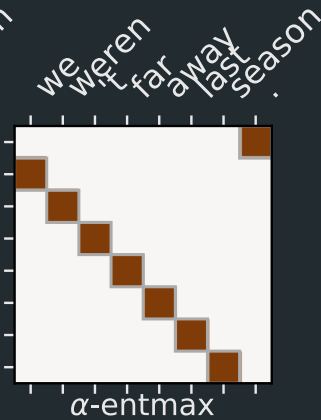
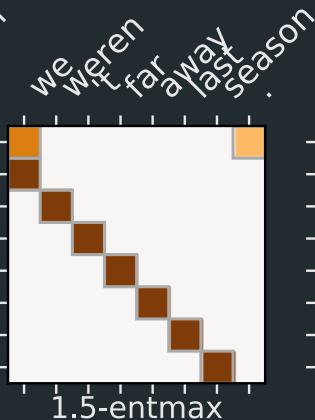
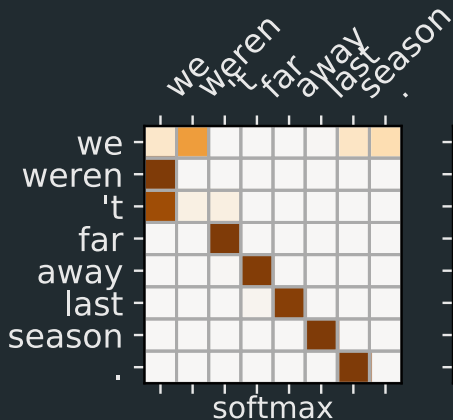
LMs usam representações **densas**.

A nossa solução foi **esparsificar**:

- interpretabilidade
- descobrir estruturas linguísticas



Maior interpretabilidade



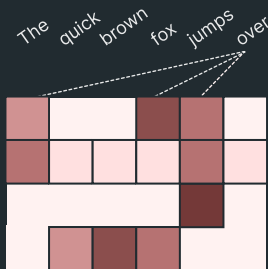
Ideias-chave

Introduzimos uma forma de usar **esparsidade adaptativa** em redes neuronais, melhorando a **transparência** das mesmas.

Ideias-chave

Introduzimos uma forma de usar **esparsidade adaptativa** em redes neurais, melhorando a **transparência** das mesmas.

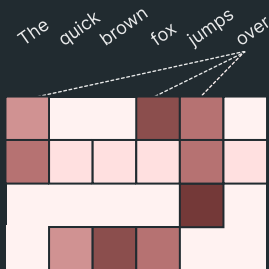
esparsidade



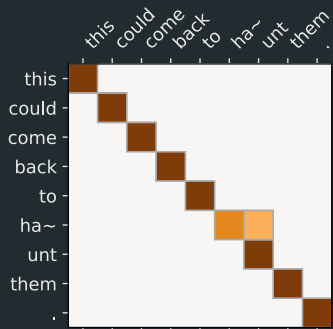
Ideias-chave

Introduzimos uma forma de usar **esparsidade adaptativa** em redes neurais, melhorando a **transparência** das mesmas.

esparsidade



mais interpretável



Introdução

Adaptively Sparse Transformers

Efficient Marg. of Discrete Latent Variables via Sparsity

Impacto e Conclusões

Soluções anteriores

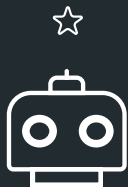


Método	sucesso (%)	# mensagens
--------	-------------	-------------

<i>Monte Carlo</i>		
--------------------	--	--

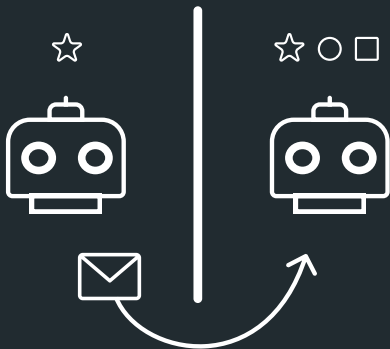
<i>Marginalização</i>		
-----------------------	--	--

Soluções anteriores



Método	sucesso (%)	# mensagens
<i>Monte Carlo</i>		
<i>Marginalização</i>		

Soluções anteriores



Método	sucesso (%)	# mensagens
--------	-------------	-------------

<i>Monte Carlo</i>		
--------------------	--	--

<i>Marginalização</i>		
-----------------------	--	--

Soluções anteriores



Método	sucesso (%)	# mensagens
--------	-------------	-------------

Monte Carlo

Marginalização

Soluções anteriores



Método	sucesso (%)	# mensagens
<i>Monte Carlo</i>		
<i>Marginalização</i>		
Denso	93.37 ± 0.42	256

Soluções anteriores



Método	sucesso (%)	# mensagens
<i>Monte Carlo</i> SFE	33.05 ± 2.84	1
<i>Marginalização</i> Denso	93.37 ± 0.42	256

A nossa solução



Método	sucesso (%)	# mensagens
<i>Monte Carlo</i>		
SFE	33.05 ± 2.84	1
<i>Marginalização</i>		
Denso	93.37 ± 0.42	256
Esparsa		

A nossa solução



Método	sucesso (%)	# mensagens
<i>Monte Carlo</i>		
SFE	33.05 ± 2.84	1
<i>Marginalização</i>		
Denso	93.37 ± 0.42	256
Esparsa		

A nossa solução



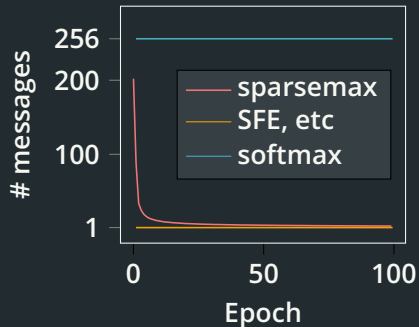
Método	sucesso (%)	# mensagens
<i>Monte Carlo</i>		
SFE	33.05 ± 2.84	1
<i>Marginalização</i>		
Denso	93.37 ± 0.42	256
Esperso	93.35 ± 0.50	

A nossa solução



Método	sucesso (%)	# mensagens
<i>Monte Carlo</i>		
SFE	33.05 ± 2.84	1
<i>Marginalização</i>		
Denso	93.37 ± 0.42	256
Esparsa	93.35 ± 0.50	3.13 ± 0.48

A nossa solução



Ideias-chave

Introduzimos um novo método para treinar, de forma mais eficiente, redes **compactas**, usando **esparsidade**.

Ideias-chave

Introduzimos um novo método para treinar, de forma mais eficiente, redes **compactas**, usando **esparsidade**.

eficiente



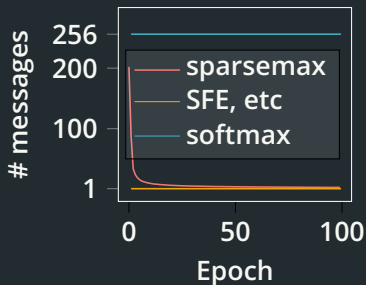
Ideias-chave

Introduzimos um novo método para treinar, de forma mais eficiente, redes **compactas**, usando **esparsidade**.

eficiente



esparso, se necessário



Introdução

Adaptively Sparse Transformers

Efficient Marg. of Discrete Latent Variables via Sparsity

Impacto e Conclusões

Impacto

Impacto

- Artigos que desenvolveram mais métodos com **esparsidade**, focados na transparência e eficiência que isso traz

Impacto

- Artigos que desenvolveram mais métodos com **esparsidade**, focados na transparência e eficiência que isso traz
- Esperamos que este trabalho tenha contribuído para IA mais responsável

Impacto

- Artigos que desenvolveram mais métodos com **esparsidade**, focados na transparência e eficiência que isso traz
- Esperamos que este trabalho tenha contribuído para IA mais responsável



Center for
Responsible AI

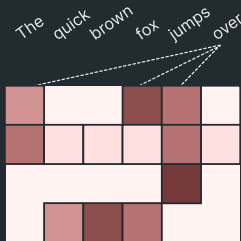
Conclusões

Com **esparsidade**, demos passos para levar modelos neuronais mais perto da versão 2.0

Conclusões

Com **esparsidade**, demos passos para levar modelos neuronais mais perto da versão 2.0

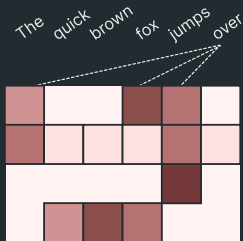
transparência



Conclusões

Com **esparsidade**, demos passos para levar modelos neuronais mais perto da versão 2.0

transparência



mais compacto



Agradecimentos