Rio 3.5, model merging e a geometria dos pesos

Rio 3.5 foi lançado e já chegou com uma polêmica: É original ou é um kage bushin do naruto?

2026-06-17



Figure 1: Capa do artigo

Rio 3.5 foi lançado e já chegou com uma polêmica: É original ou é um kage bushin do naruto?

Mas como sempre bora olhar com mais carinho pro que importa Como que combinamos modelos.

Oi sou a professora Artemísia Weyl, vem comigo entender e refletir sobre o assunto.



Figure 2: Imagem da publicação sobre o Rio 3.5

Segundo a análise publicada pela Nex, o Rio 3.5 seria aproximadamente uma combinação entre Nex-N2-Pro e Qwen, em uma proporção próxima de 60% e 40%.

E essa discussão é super legítima, transparência na área do open source, principalmente em open Models é extremamente importante.

© Artemisia Weyl 2024–2026 — Todos os direitos reservados.

Mas pro público comum que ler uma notícia dessas fica “como assim uniram os modelos?

“. isso dá certo?

Imagine que você tem dois mapas da mesma cidade.

Um tem informações somente sobre pet shops, outro tem infos sobre restaurantes.

Quando sobrepomos os dois, não destruímos a informação anterior, correto?

Apenas criamos uma representação mais rica daquele território.

Durante muito tempo se acreditou que redes neurais não funcionariam dessa forma, inclusive.

Em 2018, Garipov et al, começaram a observar que modelos treinados separadamente nem sempre terminavam em regiões isoladas.

Em muitos casos existiam caminhos conectando esses modelos sem que a qualidade das respostas piorasse significativamente.

Em linguagem simples: Os modelos pareciam morar no mesmo bairro matemático.

E Claro que na época isso mudou a forma como pesquisadores enxergam o espaço de parâmetros.

Um modelo com centenas de bilhões de parâmetros pode ser visto como um ponto em um espaço geométrico gigantesco, treinar um modelo em suma, é procurar regiões onde o erro seja pequeno.

O surpreendente é que essas regiões parecem muito mais conectadas do que se esperava.

E segura o raciocínio ai

Em 2022, Wortsman et al, lançam um trabalho chamado Model Soups.

Super simples, pegam-se vários modelos derivados da mesma base e calcular uma média entre os seus pesos.

Algo conceitualmente dessa forma:

$$\text{Modelo Novo} = \frac{\text{Modelo A} + \text{Modelo B} + \text{Modelo C}}{3}$$

O resultado surpreendeu muita gente e esse resultado surpreendeu. vários casos a combinação preservava o desempenho e às vezes até melhorava.

Quando olhamos para isso pela ótica da geometria, começa a fazer sentido.

Não estamos misturando conhecimento de forma aleatória.

Estamos navegando dentro de uma paisagem matemática extremamente complexa, procurando regiões onde diferentes soluções convivem próximas umas das outras.

É mais parecido com traçar uma rota entre cidades vizinhas.

Isso ajuda a entender por que a corrida da IA deixou de ser apenas uma corrida por mais parâmetros.

O avanço passou a depender também de como organizamos conhecimento já existente.

Por isso eu acho curioso quando alguém descreve model merging como uma gambiarra Na prática estamos vendo uma consequência direta de propriedades geométricas profundas das redes neurais modernas.

A polêmica acabou apresentando para muita gente uma ideia que não tava sendo discutida pela grande massa das pessoas: Modelos de inteligência artificial podem ser combinados.

E quando entendemos essa geometria, percebemos que os LLMs modernos têm tanto a ver com álgebra linear e otimização, quanto com linguagem.

© Artemisia Weyl 2024–2026 — Todos os direitos reservados.

Gostou?

Comenta.

© Artemisia Weyl 2024–2026 — Todos os direitos reservados.