

Bom, falando sobre essa etapa. Qual é a nossa estratégia aqui para entender. Dentro de um modelo de linguagem grandes, inserções, incorporações, qual que é a nossa estratégia?

Bom, o que a gente vem falando? Primeiro eu tenho um volume de dados muito grande, eu preciso trabalhar essas informações de uma forma rápida, que traga numa linguagem o mais próximo do mundo possível, literalmente tentando adivinhar o que as pessoas estão escrevendo, para busca de texto, tradução, entre outros.

A gente olhou a parte sobre, literalmente esmiuçando aquilo que é escrito, para que a gente consiga meio que reaproveitar isso, e entendendo qual é o próximo passo, literalmente com completa frase. Ok? Tá, e o que está faltando aqui? Está faltando fazer sentido. Então, não é apenas a questão de probabilidades, o quão provável é que a palavra A vai aparecer depois da palavra B.

Então, aqui a gente está olhando, até mesmo por isso que tem essa questão de vetores esse comparativo. Mas sim em tornar isso de uma forma que faça sentido. Faça sentido na questão de conjugação de verbos ou até mesmo se encaixe um texto, não fique aquela coisa, nossa, por que foi colocar essa palavra aqui se isso não tem nada a ver com o contexto? É mais ou menos essa estratégia.

Então, quando a gente fala sobre a parte de tokens olhando para inserções, a gente está pensando aqui como a gente vai conseguir encaixar isso dentro do nosso vetor. Não esqueça que agora a gente está trabalhando com o vetor. Então, colocando isso, olhando para um lado...semântico podemos dizer assim, né? Que vai encaixar essa estratégia.

Então, aqui a gente vai ter os nossos valores, ou seja, cada uma das nossas palavras, elas receberam um valor atribuído e agora o que é mais útil para nós é pensar num modelo, né? Aqui acredito que seria um modelo multidimensional né? Aonde eu vou colocar cada uma dessas coisas. Quem vai ficar mais próximo do que, em questão de probabilidade. Então as palavras relacionadas elas devem ficar mais perto. A gente estava falando sobre o cachorro está latindo alto para o gato e tudo mais e percebeu que tem um "skate" aqui perdido na jogada. Então o token dele ficou como 10. Ele está bem mais longe das outras palavras. "Cachorro" e "Gato" está aqui meio que próximo do outro. O "latido" e o "miado" também estão relativamente próximos. Porque está relacionado ao som, obviamente que, esses animais fazem.

Então, quando a gente fala sobre a inserção, a gente está também colocando o quão próximo um token, digamos assim, vai ficar perto do outro, com base na questão da probabilidade.

Então, a gente tem aqui um cenário onde a gente tem uma, duas, três dimensões, mas a gente pode trazer aqui muitas outras. E, obviamente, mais estratégias de como a gente pode trazer essas probabilidades.

Então, percebam que.. para que tudo isso faça sentido, para que a gente tenha uma inteligência artificial que nos traga as respostas, consiga criar frases, textos inteiros, que a gente olhe e, gente do céu, é sério que não foi uma pessoa que escreveu isso? Até mesmo pela questão da semântica e tudo mais. Com base não apenas em encaixar aqui ou encaixar ali, mas sim de olhar para grandes bases de dados, conseguir desmiuçar cada uma dessas palavras,

entender probabilidades e, também, essa questão de estratégias. Olhando aí para as inserções, por exemplo, de como conseguir encaixar as palavras de uma forma que elas façam sentido.

Então, agora eu acredito que fica muito mais fácil da gente entender Uau, como que conseguiu-se fazer tudo isso? É claro que as aplicações de inteligência artificial usam muito mais coisas, um conjunto de regras, mas olhando assim passo a passo fica bem mais fácil a gente entender a lógica por trás de tudo, né?

Agora vamos para o nosso próximo modelo, que é o de atenção