

Então, aqui quando a gente fala desse modelo, a gente segue falando sobre a questão do token e as próprias inserções.

Então, nesse sentido, o que a gente vai ter? Aqui a gente vai usar a estratégia de token, mas esse token na verdade, ele vai nos ajudar a meio que particionar as informações que já existem, e a partir daí a gente vai conseguir direcionar, porque o que acontece? Se a gente está olhando para frases por exemplo, textos e tudo mais. Como que eu sei a próxima palavra para colocar? Para completar uma frase, às vezes mais de uma palavra. Então vamos de novo, acessa a informação e muito treinamento. Então, quanto mais eu treino a minha ferramenta para ela avaliar as probabilidades, então é mais fácil da gente conseguir ser assertivo.

Então, o primeiro passo é a gente conseguir fazer esse modelo de inserções, como a gente está colocando aqui, através de vários vetores de token. E entender como que a gente vai conseguir fazer essas separações. A partir daí, tudo vai fazer um pouco mais de sentido. Então, a gente vai ter tanto a questão das camadas de atenção que a gente coloca aqui, que são os valores que a gente vai trazer aqui para os tokens. Então, estou determinando o que faz mais sentido e qual é a probabilidade de após tal palavra, ser essa daqui que vai ser colocada escrita depois. Tem alguma outra coisa antes que vai determinar isso sim ou não. Então, aqui a gente precisa determinar se isso está refletindo ou não. E, claro, daí a gente vai ter o nosso decodificador.

Então, aqui essas relações são usadas para prever a sequência mais provável. Então, digamos, eu estou redigindo um e-mail. Estou iniciando ali pelo horário muito provável dizer se eu digo bom a próxima palavra é dia. Que outra palavra vou escrever agora? Você escreveu um boa e passou do meio-dia bem provável vou dizer boa tarde, ne? Quando eu estou encerrando ali o e-mail: att.

ou se não fico à: disposição. Para nós isso é uma coisa nossa mas né que novidade todo mundo fala isso. Mas não é uma questão de todo mundo eu falo falo você fala. A forma como cada um escreve, mas sim a probabilidade disso acontecer.

Então, aqui a gente tem o nosso modelo explicado aqui do método transformador, aonde eu tenho as informações que a gente vai ter, então meu texto de entrada, a minha geração de tokens, ou seja, eu estou aqui particionando as minhas informações, estou validando a questão de inserção, a codificação. Então, aqui eu vou passar por cada uma das etapas e eu tenho os meus recursos. Então, aqui no outro ponto, meu decodificador, aqui sim eu tenho um valor de entrada, uma codificação e aqui eu vou encaixando cada uma das etapas até ter, de fato, o modelo de saída esperado. Então, quando a gente está trabalhando com isso, a gente precisa entender o seguinte, que nós vamos ter um modelo de processamento de linguagem natural, mas ele tem evoluído muito ao longo dos anos.

E, claro, ainda que a gente consiga hoje ter um resultado aparentemente excelente, a ideia é que cada probabilidade dessa aqui, que é testada, quanto mais ele vai avaliando e reavaliando. Ele tende a se tornar, sim, mais assertivo. Então, aqui dentro dos modelos que a gente tem, o que acontece? A gente vai fazer essa validação, ou seja, eu estou trabalhando com um modelo de token dividindo cada uma dessas etapas, para entender como, primeiro, o meu bloco codificador vai conseguir criar essas representações do meu vocabulário, do meu treinamento e como esse modelo decodificador vai gerar de fato, as sequências de linguagem dentro de um cenário que faça sentido aí para nós.

Então, a gente tem esse modelo, né, pensando em uma quantidade muito grande de dados, várias probabilidades, várias situações diferentes, e como

que nós vamos testar isso, meio que por fatias, para garantir aí o melhor retorno o melhor resultado.