



**Data Science
Academy**

www.datascienceacademy.com.br

Processamento de Linguagem Natural

Avaliação e Perplexidade



A perplexidade é frequentemente usada para medir a utilidade de um modelo de linguagem (basicamente, uma distribuição de probabilidade sobre sentença, frases, sequência de palavras, etc.).

Ao avaliar um modelo de linguagem, um bom modelo é aquele que tende a atribuir maiores probabilidades aos dados de teste (isto é, é capaz de prever as frases nos dados de teste com alta precisão). Para um conjunto de teste com palavras $W = w_1, w_2, \dots, w_N$, a perplexidade do modelo no conjunto de teste é:

$$PP(W) = P(w_1, w_2, w_3)^{-1/N}$$

Observe que a perplexidade é normalizada pelo número de palavras. Perplexidade é a medida da probabilidade de um determinado modelo de linguagem prever os dados de teste.

No entanto, a perplexidade não é uma maneira definitiva de determinar a utilidade de um modelo de linguagem. Um modelo com pouca perplexidade em um conjunto de teste pode não funcionar igualmente bem em um aplicativo do mundo real, cujos dados podem não ser desenhados a partir da mesma distribuição que o conjunto de teste. No entanto, na falta de meios eficientes para avaliar o modelo de linguagem, a perplexidade é uma métrica útil para comparar os modelos de linguagem.

Um modelo ruim diz que cada palavra é igualmente provável e prediz uma probabilidade uniforme para cada parâmetro, $1/N$. A perplexidade de um modelo ruim é o tamanho do vocabulário, N . Isso não é bom porque as palavras não são igualmente prováveis; certas palavras e frases tendem a juntar-se. **Se melhorar o modelo de linguagem, a perplexidade diminui.**

Se o tamanho do vocabulário for 10.000 e nossa perplexidade é 100, então, em média, para cada palavra na sequência dos dados de teste, reduzimos nossas opções de 10.000 palavras para 100.