



Processamento de Linguagem Natural com Transformers

Processamento de Linguagem Natural com Transformers

O Que é Modelagem de Tópicos?

Modelagem de Tópicos é um tipo de técnica para descobrir os tópicos abstratos "ocultos" que ocorrem em uma coleção de documentos. É frequentemente usada em Processamento de Linguagem Natural (PLN) e mineração de texto para descobrir a estrutura semântica latente dos documentos.

A modelagem de tópicos é uma forma de aprendizado não supervisionado que busca descobrir "tópicos" ou temas que estão presentes em uma coleção de documentos (ou outros tipos de dados textuais).

O objetivo da modelagem de tópicos é identificar os tópicos, ou seja, as distribuições de palavras que ocorrem frequentemente juntas e a proporção desses tópicos em cada documento.

Existem várias técnicas diferentes para modelagem de tópicos, sendo uma das mais populares a Alocação Latente de Dirichlet (LDA - Latent Dirichlet Allocation). A LDA assume que cada documento é uma mistura de um número pequeno de tópicos e que cada palavra no documento é atribuída a um dos tópicos do documento. A LDA é um método não supervisionado, o que significa que não necessita de dados de treinamento rotulados. E uma alternativa baseada em Inteligência Artificial e na arquitetura BERT, é o BERTopic.

A modelagem de tópicos tem uma variedade de usos, incluindo:

- Agrupamento de documentos: documentos com tópicos similares podem ser agrupados juntos.
- Resumo de textos: os tópicos podem ser usados para resumir o conteúdo de um documento.
- Recomendação de conteúdo: se você sabe quais tópicos interessam a um usuário, pode recomendar novos documentos que abordem esses tópicos.
- Análise de tendências: em uma coleção de documentos ao longo do tempo, você pode ver como os tópicos (ou a importância deles) mudam.

A modelagem de tópicos é uma ferramenta poderosa para entender grandes coleções de textos.