Nome: Fabio Grassiotto
RA: 890441
IA024 - Aula 2_3
Leitura do Artigo “Attention Is All You Need - Vaswani et al”

Resumo do Artigo:

Este é um artigo seminal, importantíssimo para a pesquisa em processamento de linguagem desde sua publicação em 2017. No artigo, os autores apresentam uma nova arquitetura de rede neural, chamada de Transformers.
O objetivo inicial dos autores foi propor um modelo alternativo às redes neurais sequenciais recorrentes para o modelamento de linguagem e sistemas de tradução simultâneos.
No artigo, os autores apresentam a arquitetura do modelo que consiste em duas seções distintas, um codificador e um decodificador. A seção do codificador é responsável por processar entradas e gerar representações intermediárias, normalmente na forma de embeddings de texto, enquanto que o decodificador utiliza as representações para gerar a saída final.
O modelo emprega como técnica principal o mecanismo de auto-atenção, que é explorado no artigo. Os autores descrevem a implementação deste mecanismo, apresentado como um sistema de mapeamento de entradas e saídas. A auto-atenção é descrita como um mecanismo que permite que o modelo capture dependências entre diferentes palavras dentro de uma sequência de entrada. Ao fazer isso, ele aprende a se concentrar nas palavras mais relevantes para cada posição, o que melhora significativamente sua capacidade de entender o contexto e executar tarefas como tradução ou outras operações baseadas em sequência.
Os autores também descrevem a necessidade de se criar uma codificação para a posição das entradas dentro de uma sequência, o que é conseguido através da adição de *positional encodings* aos dados de embeddings de texto.
Após a descrição dos mecanismos utilizados para a proposta da rede neural, os autores descrevem o processo de treinamento do modelo, detalhando o dataset utilizado, o hardware necessário e as técnicas utilizadas, como o tipo de otimizador e a regularização do treinamento.
Finalmente, o artigo apresenta os resultados obtidos pelo modelo, na tarefa de tradução de inglês para alemão e de inglês para francês do dataset WMT2014. Na época, notou-se que o modelo proposto apresentou um nível de performance medido através da métrica BLEU superior aos modelos sequenciais, a um custo de 25% de treinamento.
Na minha opinião, este é um artigo bem organizado, que conseguiu apresentar um novo modelo que trouxe ganhos substanciais para as tarefas de tradução e outras. No artigo, os autores explicam de forma clara os mecanismos utilizados, discorrem sobre os processos de treinamento e apresentam o ganho de performance atingido com modelos utilizando a tecnologia de Transformers.