

Tradução Automática Usando Atenção

Código

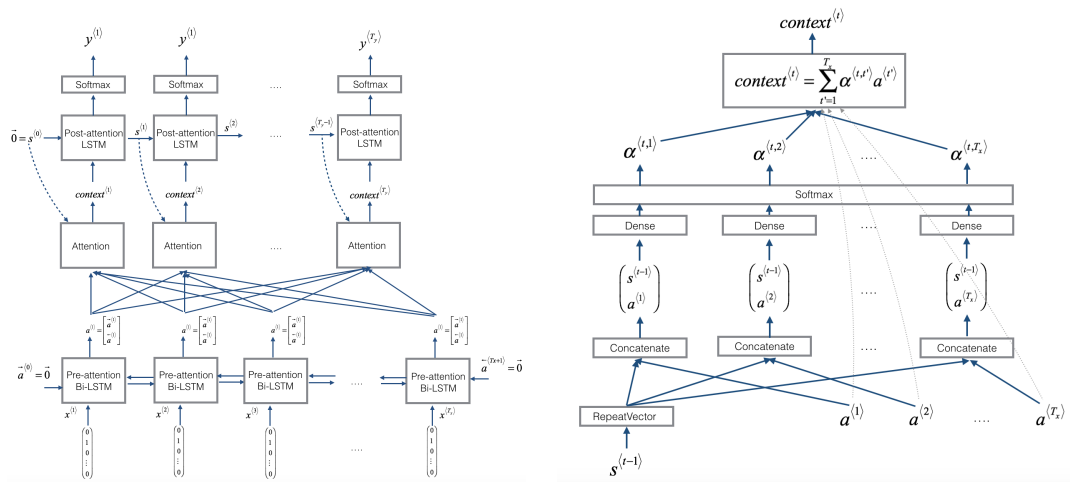
Uma das tarefas em que algoritmos de *deep learning* mais apresentaram evolução nos últimos anos foi a tradução automática de textos (vide Google Tradutor). Porém, a tradução entre idiomas propriamente dita é uma tarefa avançada, que envolve elementos além do escopo deste curso, além de exigir um alto poder computacional.

Neste contexto, este projeto propõe uma tarefa de tradução simplificada em que o objetivo é converter datas de um formato que seres humanos costumam usar para escrever (em inglês), por exemplo, “17th of October, 2023” ou “tuesday october 17 2023” para um formato padronizado e mais amigável ao computador, usando o padrão YYYY-MM-DD. No caso da data mencionada isso seria “2023-10-17”.

A pasta “files” contém dados que serão usados no treinamento do seu modelo, funções auxiliares que poderão ser usadas se desejarem (opcional) além do modelo pré-treinado:

- dataset: lista de tuplas com pares (data em formato humano, data em formato de máquina).
- human_vocab: dicionário mapeando os caracteres nas datas em formato humano para índices inteiros.
- machine_vocab: dicionário mapeando os caracteres nas datas em formato de máquina para índices inteiros.
- inv_machine_vocab: dicionário inverso de machine_vocab, mapeando de volta dos índices para os caracteres.
- string_to_int.py: função que converte uma string para uma lista de inteiros usando human_vocab ou machine_vocab (possivelmente com *padding*).
- plot_attention_map.py: função que desenha o mapa de atenção.
- model.h5: modelo já pré-treinado e pronto para ser usado na etapa de inferência.

1. Implemente um modelo de atenção associado a LSTM para tradução similar ao visto em aula e cujo diagrama esquemático é apresentado abaixo.



2. Compare o desempenho do seu modelo (testando para alguns exemplos) fazendo o treinamento do zero por 1 época e por 50 épocas com o desempenho usando o modelo pré-treinado.
3. Teste diferentes valores para o número de unidades na pré-ativação e na pós-ativação e verifique a influência no resultado final testando para alguns exemplos de datas.
4. Mostre o mapa de atenção para alguns exemplos de datas.

Relatório

1. Introduza o problema de tradução automática, sua importância e principais métodos usados na atualidade.
2. Descreva os passos e a arquitetura de tradução usada em seu projeto bem como a intuição/motivação por trás desses passos. Faça um resumo de aspectos teóricos e práticos do algoritmo.
3. Mostre os resultados do seu código testando para datas com formatos de sua escolha.
4. Discuta os seus resultados, formule hipóteses que expliquem os resultados bons ou ruins que obteve, comente sobre o papel do número de unidades de pré e pós-ativação e o número de épocas de treinamento, discuta os mapas de atenção que gerou, comente sobre possíveis melhorias que poderiam ser adicionadas ao modelo, etc.

5. Liste as referências usadas na escrita do relatório e na produção do seu código.