

Materiais:

Q: Onde encontro os materiais de apoio e os arquivos notebooks para execução dos exercícios?

R: O material de apoio consta no ícone de folha a4 ao lado do título da disciplina.

Ferramentas:

Pytorch

Essas bibliotecas mais novas passam por constantes atualizações e por tal motivo mudam constantemente suas estruturas e as estruturas das funções utilizadas durante as demonstrações dos professores, causando incompatibilidade. Assim, se você encontrou um erro indicando depreciação de código ou algo assim, por favor nos comunique para podermos atualizar os notebooks afetados.

Conceitos e Exercícios:

Q: Estava tentando entender o cálculo do positional encoding e apresentado no slide 76 da aula 4. No slide são apretnadas as equações para os índices pares e ímpares. Não consegui reproduzir os valores apresentados no exemplo para o texto "Je suis étudiant", não consegui chegar ao mesmo resultado do slide.

R: Os valores dos slides são meramente ilustrativos. Na equação do positional encoding existem 3 variáveis diferentes.

"pos": que é a posição da palavra/caracter "i": é a posição da palavra no word embedding treinado (no caso se "je" for a 24 palavra treinada no word embedding "i" terá valor igual a 24). "dmodel": o tamanho das embeddings (podendo variar para cada modelo, no paper original sendo 512).

Dessa forma, na imagem existem duas variáveis que não temos como ter certeza do valor utilizado "i" e "dmodel". Como sugestão do professor, caso tenha interesse em validar manualmente o cálculo do positional embedding, o estudante pode usar o Notebook resolvido dessa aula.

Q: Sobre dados não-estruturados, semi-estruturados.

R: O e-mail é encarado como não estruturado quando não temos nenhuma ideia de sua estrutura, a imagem de um celular possui um certo nível de informação estruturada, como por exemplo as imagens do celular que possuem tags com informação de resolução, geolocalização, etc.

Q: Conceito de cluster na tela (inserido no vídeo) está errado no tempo 39:00

R: Foi descrito o cluster de computação em nuvem ao invés do agrupamento de documentos (cluster de documento).

Q: Estou com dificuldade em realizar a aplicação da ferramenta word2vec para o que eu preciso. Já entendi o conceito por trás do algoritmo, mas ainda não consigo entender como aplicar isso para classificar os documentos de patentes que tenho.

R: Na verdade o Word2Vec só irá transpor tuas sentenças todas para forma de representações numéricas de acordo com um critério como a frequência de palavras e etc. O Word2Vec vai transformar a sentença para INPUTs do seu modelo classificador. Nesse caso é possível usar uma árvore de decisão, regressão linear ou qualquer outro classificador.

** FAQ gerado com base em comentários até o dia 30/11/2022.*