

Trabalho de Redes Neurais Artificiais

Componente Curricular: GEX618 – Inteligência Artificial

Créditos: 4

Ano/semestre: 2022/2

Curso/fase de oferta:

Ciência da Computação (noturno) / 7ª fase

Professor(es): Felipe Grando

1. Descrição

Trabalho individual ou em duplas de implementação em Python, análise e apresentação de um sistema de reconhecimento de voz usando o modelo *whisper* (OpenAI). O material de apoio sobre o modelo pode ser encontrado no Moodle do CCR.

O grupo deve coletar um conjunto de dados (arquivos de áudio) com características variadas (músicas, diferentes sotaques, gravação simples de voz, conversas simultâneas, ruído de fundo etc.). O conjunto de dados deve conter no mínimo 10 amostras. Cada amostra deve estar acompanhada de uma transcrição oficial ou esta deverá ser realizada pelos próprios integrantes do grupo de forma manual.

O programa implementado deverá aplicar o modelo *whisper* em cada amostra para transcrever os dados automaticamente além de computar como métricas de avaliação do modelo a quantidade de palavras erradas (que não são idênticas a transcrição oficial/manual) em cada amostra, o percentual de palavras erradas em cada amostra (relação entre a quantidade de erros em relação ao tamanho do corpo textual da amostra) e o tempo total de execução do modelo em cada amostra.

O programa deverá salvar em um arquivo de resultados único as características de cada amostra (nome, tamanho, idioma, tipo e outros relevantes) acompanhadas de sua transcrição oficial/manual, da transcrição realizada pelo modelo, do resultado das métricas de avaliação computadas e do tamanho do modelo utilizado.

Selecionar uma das seguintes questões de pesquisa para explorar no trabalho:

1. Qual é a diferença na eficiência entre os diferentes modelos disponíveis na biblioteca *whisper* (tiny, base, small, médium, large)?
2. Qual é a diferença na eficiência dos modelos disponíveis na biblioteca *whisper* considerando diferentes idiomas (escolher ao menos dois idiomas)?
3. Qual é a diferença na eficiência dos modelos específicos para o inglês e os multilíngues (entre tiny.en e tiny, base.en e base, small.en e small, médium.en e médium)?
4. Qual é a diferença na eficiência do modelo considerando diferentes sotaques/regionalismos dentro de uma mesma linguagem? Escolher um idioma para explorar como os exemplos abaixo:
 - a. português do sul, nordeste, norte ou do interior versus capital;
 - b. inglês dos EUA, do Reino Unido, da Austrália;
 - c. francês do Haiti, França, Guiana.

5. Qual é a diferença na eficiência do modelo em diferentes situações ambientais? Explorar um elemento como os exemplos abaixo:
 - a. Níveis de ruído de ambiente no fundo;
 - b. Volume/intensidade do áudio;
 - c. Cadência da fala (rápida, lenta, cantada);
 - d. Múltiplas vozes versus vozes únicas;
 - e. Timbres diferenciados de voz.
6. Qual é a diferença na eficiência do modelo considerando amostras de um mesmo áudio em qualidades (taxas de bits) ou formatos de arquivo distintos?
7. Qual é a diferença na eficiência do modelo considerando regionalismos/gírias e/ou frases desconexas e sem sentido real na linguagem?
8. Outras questões de pesquisa podem ser utilizadas pelo grupo desde que verifiquem com o professor se estão adequadas considerando o escopo do trabalho proposto.

Os resultados obtidos pelo grupo devem ser comparados e analisados pelos mesmos. As conclusões obtidas, assim como a descrição do que e como foi realizado devem ser apresentados para a turma. A apresentação acontecerá de forma assíncrona (deverá ser gravada) e o modelo *whysper* deve ser utilizado para realizar a transcrição do áudio da apresentação do grupo.

2. Entrega e prazos

A entrega do trabalho deverá ser realizada no Moodle até o dia indicado na plataforma (não serão aceitas entregas atrasadas). Deverão ser entregues o código em Python utilizado, um arquivo com os resultados (conforme consta na descrição), um arquivo com a apresentação e um arquivo com a transcrição da apresentação. As amostras utilizadas, arquivos de áudio, devem ser anexadas/entregues no Moodle se as mesmas couberem dentro do limite de 10MB impostas pela plataforma, caso contrário, poderão ser omitidas.

3. Método de Avaliação

A avaliação do trabalho será dividida em duas etapas: avaliação do trabalho entregue no Moodle e a avaliação da apresentação. Será atribuída uma nota de 0 a 10 para cada etapa avaliativa, sendo que cada uma das partes comporá 50% da nota final do trabalho.

A avaliação do trabalho de implementação entregue será a mesma para todos os integrantes do grupo e considerará os seguintes itens:

- criatividade e complexidade da implementação e dos dados/amostras utilizadas/coletadas, considerando a coerência dos mesmos com a questão de pesquisa escolhida para ser explorada;
- organização e documentação do código (separação adequada do código em funções, nomes significativos para variáveis, estruturas e funções, comentários de código em partes importantes);
- organização, clareza e completude dos resultados considerando também o escopo e aderência à questão de pesquisa escolhida.

A avaliação da apresentação do trabalho será individual por integrante do grupo e considerará os seguintes itens:

- criatividade, organização e preparo;
- objetividade e clareza na fala e nos elementos auxiliares utilizados (slides, gráficos, imagens, outros);
- adequação ao tempo definido, de no máximo 15min;

Curso de Ciência da Computação

- deve ficar claro na apresentação a questão de pesquisa explorada, quais foram os dados/amostras utilizados, os resultados obtidos e as conclusões obtidas nas análises dos mesmos.

Este trabalho deverá ser entregue no Moodle até a data prevista (cadastrada no Moodle) e valerá uma nota de 0 a 10 que comporá 50% da nota final do CCR.

Será dado nota 0 (zero) para todos os estudantes que não entregarem o trabalho dentro do prazo estipulado ou que plagiarem o trabalho em sua totalidade ou em considerável parte de um colega ou de fonte externa.