ID da Contribuição: **7** Tipos: **Trabalho consolidado ou em conclusão**

De Embeddings a Modelos de Linguagem: Uma Análise Comparativa de Extratores de Caraterística para Geração de Gestos com Texto-Only e Multimodal

quinta-feira, 4 de dezembro de 2025 11:10 (12 minutos)

A geração de gestos co-verbais expressivos e contextualmente apropriados é crucial para a naturalidade na interação humano-agente. Embora os Modelos de Linguagem Grandes (LLMs) tenham demonstrado grande potencial para essa tarefa, persistem questões sobre a integração ideal de características multimodais e as capacidades de modelos menores e mais acessíveis. Este estudo apresenta uma avaliação sistemática e comparativa de sete pipelines de geração de gestos, utilizando uma robusta arquitetura baseada em difusão. Investigamos o impacto de extratores de características de áudio (WavLM, Whisper) e texto (Word2Vec, Llama-3.2-3B-Instruct) para avaliar a contribuição relativa de cada modalidade. Demonstramos que é possível alcançar desempenho de ponta utilizando um LLM significativamente menor (3B parâmetros) do que benchmarks anteriores, sem sacrificar a qualidade. Nossos resultados, baseados em métricas objetivas e uma avaliação perceptual abrangente, revelam que os pipelines que incorporam o Llama-3.2-3B-Instruct não apenas superam as referências em adequação semântica e semelhança humana, mas também são percebidos como mais apropriados por avaliadores humanos.

Autor: GOMEZ SANCHEZ, Johsac Isbac (Estudante)

Co-autor: COSTA, Paula (Unicamp)

Apresentador: GOMEZ SANCHEZ, Johsac Isbac (Estudante)

Classificação da Sessão: Sessões orais