Briefing do Desafio - Detecção de Desinformação e Anomalias em Dados Multimídia

Instituto Apoema
tecnologia e inovação

Tribunal Regional Eleitoral
de Goiás

INF
INSTITUTO DE INFORMÁTICA

UFG
UNIVERSIDADE FEDERAL DE GOIÁS

Título:
Busca Multimodal em Vídeos: Localização de Conteúdos Baseada em Texto Descritivo

Desafio:
Como podemos criar um sistema inovador que, a partir de uma descrição textual (por exemplo, "Homem andando a cavalo"), localize de forma precisa e rápida os vídeos que melhor correspondam a esse conteúdo, mesmo em grandes repositórios de dados multimídia?

Resumo do Contexto:
Com o crescimento exponencial de vídeos em plataformas de mídia, encontrar conteúdos relevantes se tornou um grande desafio. Métodos baseados exclusivamente em metadados, como títulos e tags, frequentemente falham em capturar a riqueza semântica do conteúdo visual. Este desafio propõe o desenvolvimento de um sistema que utilize aprendizado de máquina multimodal para correlacionar descrições textuais e vídeos. Isso envolve a extração de embeddings tanto visuais quanto textuais, possibilitando uma busca mais intuitiva e precisa. O sistema deve identificar automaticamente cenas ou segmentos relevantes em vídeos, correlacionando-os com a semântica da consulta textual.

Resultados Esperados:
- Desenvolvimento de um sistema de busca que retorne os vídeos mais relevantes para a descrição fornecida, com uma precisão Top-5 de pelo menos 85%.
- Uso de modelos de aprendizado multimodal, como CLIP (Contrastive Language–Image Pretraining), para geração de embeddings compartilhados entre texto e vídeo.
- Implementação de uma interface de busca que permita ao usuário fornecer descrições textuais e visualizar os resultados organizados por relevância.
- Avaliação detalhada do sistema utilizando métricas como Recall@K, Mean Reciprocal Rank (MRR) e precisão baseada em feedback de usuários.
- Teste prático com vídeos de diferentes categorias e complexidades semânticas, destacando cenas correspondentes às descrições fornecidas.

Características do Desafio:
- **Visão Computacional e Processamento de Vídeos:**
    - Identificação de cenas relevantes por meio de extração de frames e características visuais em vídeos.

- **Processamento de Linguagem Natural (NLP):**
    - Geração de representações vetoriais de descrições textuais para análise semântica.
- **Aprendizado Multimodal:**
    - Integração de diferentes modalidades (texto e vídeo) em um espaço compartilhado para análise de similaridade.
- **Sistemas de Recuperação de Informação:**
    - Implementação de algoritmos eficientes de busca, indexação e ranqueamento de conteúdos audiovisuais.
- **Modelagem de Similaridade Semântica:**
    - Desenvolvimento de métricas que quantifiquem a correlação entre representações textuais e visuais.