

Título: Frontiers of Large Language Models: Empowering Decision Optimization, Scene Understanding, and Summarization Through Advanced Computational Approaches

Resumen:

El advenimiento de los Large Language Models (LLMs) marca una fase transformadora en el campo de la Inteligencia Artificial (IA), significando el cambio hacia sistemas inteligentes y autónomos capaces de una comprensión y toma de decisiones complejas. Esta tesis profundiza en las capacidades multifacéticas de los LLMs, explorando sus posibles aplicaciones en la optimización de decisiones, la comprensión de escenas y tareas avanzadas de resumen de video en diversos contextos.

En el primer segmento de la tesis, el foco está en la comprensión semántica de escenas de Vehículos Aéreos No Tripulados (UAVs). La capacidad de proporcionar instantáneamente datos de alto nivel y señales visuales sitúa a los UAVs como plataformas ideales para realizar tareas complejas. El trabajo combina el potencial de los LLMs, los Visual Language Models (VLMs), y los sistemas de detección objetos de última generación para ofrecer descripciones de escenas matizadas y contextualmente precisas. Se presenta una implementación práctica eficiente y bien controlada usando microdrones en entornos complejos, complementando el estudio con métricas de legibilidad estandarizadas propuestas para medir la calidad de las descripciones mejoradas por los LLMs. Estos avances podrían impactar significativamente en sectores como el cine, la publicidad y los parques temáticos, mejorando las experiencias de los usuarios de manera exponencial.

El segundo segmento arroja luz sobre el problema cada vez más crucial de la toma de decisiones bajo incertidumbre. Utilizando el problema de Multi-Armed Bandits (MAB) como base, el estudio explora el uso de los LLMs para informar y guiar estrategias en entornos dinámicos. Se postula que el poder predictivo de los LLMs puede ayudar a elegir el equilibrio correcto entre exploración y explotación basado en el estado actual del sistema. A través de pruebas rigurosas, la estrategia informada por los LLMs propuesta demuestra su adaptabilidad y su rendimiento competitivo frente a las estrategias convencionales.

A continuación, la investigación se centra en el estudio de las evaluaciones de bondad de ajuste de las Generative Adversarial Networks (GANs) utilizando la Signature Transform. Al proporcionar una medida eficiente de similitud entre las distribuciones de imágenes, el estudio arroja luz sobre la estructura intrínseca de las muestras generadas por los GANs. Un análisis exhaustivo utilizando medidas estadísticas como las pruebas de Kruskal-Wallis proporciona una comprensión más amplia de la convergencia de los GANs y la bondad de ajuste.

En la sección final, la tesis introduce un nuevo benchmark para la síntesis automática de videos, enfatizando la integración armoniosa de los LLMs y la Signature Transform. Se propone un enfoque innovador basado en los componentes armónicos capturados por la

Signature Transform. Las medidas son evaluadas extensivamente, demostrando ofrecer una precisión convincente que se correlaciona bien con el concepto humano de un buen resumen.

Este trabajo de investigación establece a los LLMs como herramientas poderosas para abordar tareas complejas en diversos dominios, redefiniendo la optimización de decisiones, la comprensión de escenas y las tareas de resumen de video. No solo establece nuevos postulados en las aplicaciones de los LLMs, sino que también establece la dirección para futuros trabajos en este emocionante y rápidamente evolucionante campo.