Instituto Tecnológico y de Estudios Superiores de Monterrey
Campus Monterrey

Posgrados

Avance 2. Ingeniería de características

Integrantes del equipo de trabajo

Carlos Daniel Villena Santiago A01795127
Edwin David Hernandez Alejandre A01794692
Gustavo Andres Garcia Anguiano A01795493

05 de octubre del 2025

1. Contexto del Proyecto: Construct AI

Construct AI es un sistema innovador de **Inteligencia Artificial generativa multimodal** que se apoya en un enfoque **RAG (Retrieval-Augmented Generation)**. Su principal objetivo es la comprensión y asistencia avanzada sobre **planos arquitectónicos**.

El sistema se distingue por integrar:

- **Imágenes segmentadas** de planos.
- Un **conjunto de conocimiento textual** estructurado en formato *Markdown*, generado previamente con **Gemini AI**.

Este diseño permite una recuperación de información visual y textual de forma semántica, simplificando la interpretación de planos a través del lenguaje natural.-----**2. Ingeniería de Características en Construct AI**

La ingeniería de características en Construct AI se centra en la **creación, transformación y gestión de embeddings multimodales**. Estos embeddings representan la información visual y textual en espacios vectoriales comparables, siendo clave para alcanzar los objetivos 2.3 (crear nuevas características para mejorar el rendimiento) y 2.4 (mitigar sesgos y acelerar la convergencia).

A. Construcción de Características1. Modelo de Imágenes: CLIP

- **Modelo:** openai/clip-vit-base-patch16
- **Ubicación:** src/ingest/image_ingest_pipeline.py:228–246
- **Dimensión:** 512
- **Uso:** Vectoriza los segmentos de planos arquitectónicos tras las etapas de segmentación y OCR.
- **Proceso:**
    - Se aplica un **OCR previo** sobre las imágenes para extraer el texto incrustado.
    - Los resultados del OCR se asocian a cada fragmento visual segmentado.
    - CLIP procesa la imagen junto con su texto contextual, generando un embedding semántico que captura tanto relaciones espaciales como textuales.

Justificación Técnica: CLIP permite mapear imágenes y texto en un espacio vectorial común, creando representaciones abstractas y robustas. Estas características de alto nivel reemplazan a las técnicas tradicionales (como *one-hot encodings* o *feature scaling* manual), expresando la similitud conceptual entre diversas secciones del plano

B. Modelo de Texto: BGE-M3

- **Modelo:** BAAI/bge-m3
- **Ubicación:** src/utils/model_manager.py:62–97
- **Dimensión:** 1024 (definida en config_retrieval.yaml:14)
- **Uso:** Vectoriza las consultas textuales y las descripciones en Markdown del *base-knowledge*.
- **Técnicas Utilizadas:**
    - *Mean pooling* para consolidar representaciones de tokens.
    - *Normalización L2* para asegurar coherencia en magnitud y dirección.

Justificación Técnica: El modelo BGE-M3 transforma el texto narrativo técnico (descripciones arquitectónicas, materiales, dimensiones) en un espacio vectorial uniforme, alineado con CLIP para facilitar la comparación semántica entre modalidades. Esto constituye una forma moderna de **generación y codificación de nuevas características**, análoga a los procesos clásicos de *feature engineering* en datasets tabulares.**B. Normalización y Escalamiento**

El sistema aplica **normalización L2** sobre todos los embeddings (tanto de CLIP como de BGE), asegurando que cada vector tenga una magnitud unitaria. Esta práctica garantiza que la **distancia coseno** refleje únicamente la similitud direccional entre vectores, eliminando el efecto de la magnitud y asegurando equidad entre características.

Justificación Metodológica:

- Asegura que las comparaciones entre modalidades (imagen ↔ texto) se realicen en condiciones equitativas.
- Mejora la **estabilidad numérica** y la **convergencia** del proceso de recuperación semántica.
- Equivale funcionalmente al *standard scaling* de variables tradicionales, pero en espacios de alta dimensión.

C. Selección y Extracción de Características

En *Construct AI*, la selección de características no se basa en métodos estadísticos clásicos (como ANOVA o PCA), sino en una **curación y filtrado semántico** implícito en las propiedades de los embeddings y su gestión en MongoDB.

Técnicas Aplicadas:

1. **Filtrado Semántico:** Solo se conservan los embeddings que superan un umbral de similitud coseno con respecto a conceptos relevantes del dominio arquitectónico. Este umbral actúa como un criterio de selección automática, reduciendo la redundancia y el ruido en la base vectorial.
2. **Extracción Implícita:** Los modelos CLIP y BGE realizan una *reducción de dimensionalidad* interna a través de sus capas de proyección, similar al efecto de un **PCA entrenado**. Esto optimiza la densidad informativa sin comprometer el contexto semántico.
3. **Gestión en MongoDB:** Los embeddings normalizados se almacenan directamente en MongoDB junto con sus metadatos, lo que permite realizar búsquedas vectoriales eficientes y contextualizadas.

Resultado: Este esquema reduce la complejidad del modelo, mejora los tiempos de recuperación y facilita la interpretación de resultados al mantener solo las características más informativas.

D. Validación y Métricas

Para evaluar la calidad de los embeddings y la efectividad del *retrieval*, se utiliza la **distancia coseno** como métrica principal. Esta medida permite verificar que los embeddings más cercanos correspondan a imágenes y textos conceptualmente similares, garantizando la consistencia semántica.

Justificación: La distancia coseno se alinea con el principio de normalización L2, facilitando la evaluación sin requerir métricas supervisadas (como *accuracy* o *recall@k*), dado que el sistema se encuentra en etapa de desarrollo exploratorio.

E. Conclusiones

En el marco metodológico **CRISP-ML(Q)**, esta fase se enmarca dentro del bloque de **Data Preparation**, donde los datos son transformados en representaciones útiles para el modelado. Las principales conclusiones son:

1. La ingeniería de características en *Construct AI* se implementa a través de **modelos de embeddings multimodales**, que generan representaciones abstractas capaces de capturar información espacial, visual y lingüística.
2. El uso combinado de **OCR, CLIP y BGE-M3** extiende el concepto de ingeniería de características hacia un dominio multimodal, donde las características no son creadas manualmente, sino aprendidas y optimizadas por modelos preentrenados.
3. La **normalización L2** y la **selección semántica** garantizan consistencia, reducen redundancia y mitigan sesgos, mejorando la capacidad de generalización.
4. Este enfoque acelera la convergencia del sistema de recuperación y genera una base sólida para la fase de modelado generativo con **Gemini AI**, que actúa como modelo de síntesis en la etapa posterior.
5. En resumen, la aplicación de estos procesos cumple plenamente con los objetivos del avance 2: crear nuevas características significativas y mitigar riesgos de sesgo en el modelado.

Referencias

- Visengeryeva, L. et al. (2023). *CRISP-ML(Q): The ML Lifecycle Process.* MLOps INNOQ.
- Galli, S. (2022). *Python Feature Engineering Cookbook.* Packt Publishing.
- OpenAI. (2021). *CLIP: Connecting Text and Images.*
- BAAI. (2023). *BGE-M3: Multilingual Embedding Model for Semantic Retrieval.*