**CSE508 Information Retrieval**

**Winter 2024**

**MT23017- Ankur Tiwari**

**Multimodal Retrieval System**

**Problem Statement:**

The objective is to construct a Multimodal Retrieval System capable of suggesting relevant images and review text based on user-provided textual or image inputs.

**Approach:**

**Initial Data Processing**:

The system initiates by parsing a CSV dataset, extracting image URLs and review text while managing exceptions for invalid URLs or image retrieval errors.

Rows with multiple image URLs for a single Product ID are segregated into distinct rows.

Image Feature Extraction from URL:

Image preprocessing transformations are defined, and a pre-trained ResNet-18 model is loaded.

Features are extracted from image URLs using the ResNet-18 model after applying the defined transformations.

**Text Preprocessing:**

Text preprocessing is executed on the 'Review Text' column of the dataset.

Processes encompass HTML tag removal, text conversion to lowercase, tokenization, punctuation removal, stop word removal, and lemmatization.

Calculating TF-IDF for Processed Review Text:

Tokenized texts are processed to compute Term Frequency (TF), Inverse Document Frequency (IDF), and TF-IDF scores.

TF-IDF scores are transformed into a DataFrame for subsequent analysis.

**Computing Cosine Similarity:**

Cosine similarity is computed between features extracted from input images and features in the dataset.

Cosine similarity based on TF-IDF scores is determined between input review text and review texts in the dataset.

Composite similarity scores are derived by averaging cosine similarities from both text and image domains.

**Model and Formulas:**

**ResNet-18**:

ResNet-18 architecture comprises 18 layers, including convolutional, pooling, and fully connected layers.

It incorporates residual connections to address the vanishing gradient problem, facilitating effective training of deeper networks.

**TF-IDF Calculation**:

TF is calculated by dividing the count of each word in tokenized text by the total number of words in the document.

IDF is computed based on how many documents contain each word, followed by a logarithmic transformation.

TF-IDF scores are obtained by multiplying TF scores with IDF scores for each word in the document.

**Results:**

**Top Comparable Pictures:**

The system successfully identifies visually similar photos to the input image, exhibiting cosine similarity scores alongside review descriptions.

**Top Comparable Reviews**:

Utilizing TF-IDF ratings, the system suggests reviews with text akin to the input review, providing insights into comparable textual evaluations and corresponding photos.

**Composite Similarity Rankings:**

By computing composite similarity scores accounting for both text and image domains, the system delivers a holistic view of similarity, enabling comprehensive recommendations.

**Conclusion:**

The developed system furnishes an efficient approach for recommending related photos and reviews grounded on input image URLs and review content. Integration of computer vision and natural language processing techniques ensures a comprehensive recommendation strategy, with scope for further optimization and exploration of advanced similarity criteria to enrich suggestions.









