CSE508_Winter2024_A2_Report_2021080

Report on Building a Retrieval System for Image and Text Data

1. Introduction:
In today's digital age, the abundance of multimedia data, including images and textual content, necessitates effective retrieval systems to efficiently search and retrieve relevant information. In this report, we detail the development of a retrieval system that integrates both image and textual data to enable users to find relevant content based on their queries.

2. Objective:
The objective of this project is to develop a retrieval system that leverages both image and text data for similarity computation and retrieval. By combining features extracted from images and textual reviews, the system aims to provide more accurate and comprehensive search results to users.

3. Methodology:

Data Loading and Preprocessing:
The project begins with loading a dataset containing product information, including image URLs and textual reviews. Pandas is used to read the dataset, and preprocessing steps are applied to extract relevant information from the data.

Image Preprocessing:
Images are preprocessed using techniques such as resizing, contrast enhancement, rotation, flipping, brightness adjustment, and exposure adjustment. These preprocessing steps ensure that images are in a standardized format and enhance their visual quality for feature extraction.

Feature Extraction from Images using MobileNet:
Feature extraction is performed on images using the MobileNet model pre-trained on the ImageNet dataset. This model enables the extraction of high-level features from images, which can then be used for similarity computation and retrieval.

Text Preprocessing using NLTK:
Textual reviews undergo preprocessing using NLTK (Natural Language Toolkit). This involves lower-casing, tokenization, removal of punctuation and stopwords, stemming, and lemmatization to standardize the textual data and prepare it for further analysis.

TF-IDF Calculation:
TF-IDF (Term Frequency-Inverse Document Frequency) scores are calculated for textual reviews using scikit-learn's TfidfVectorizer. This step computes the importance of each word in the reviews, allowing for the representation of textual data in a numerical format.

Similarity Computation for Images:
Cosine similarity is calculated between images based on their extracted features. This enables the identification of similar images in the dataset, facilitating image-based retrieval.

Similarity Computation for Textual Reviews:
Similarly, cosine similarity is computed between textual reviews based on their TF-IDF scores. This enables the identification of similar textual content, facilitating text-based retrieval.

Composite Similarity Computation:
To provide comprehensive search results, composite similarity scores are computed by combining both

image and text similarity scores. This ensures that the retrieval system considers both modalities when ranking search results.

Challenges and Potential Improvements:

The project acknowledges various challenges in building a retrieval system, including the semantic gap between image features and textual content, data heterogeneity, scalability concerns, domain-specific processing, evaluation metrics selection, user interaction, and ethical considerations. Potential improvements are suggested to address these challenges, such as improving feature representations, integrating multiple modalities, efficient indexing and retrieval, domain-specific processing, and considering ethical and privacy concerns.

4. Results and Discussion:

The retrieval system successfully computes similarity scores for both images and textual reviews, enabling users to search for relevant content based on their queries. By combining image and text data, the system provides more comprehensive search results compared to systems that consider only one modality. The computed composite similarity scores offer a holistic view of content relevance, considering both image and text similarity.

5. Conclusion:

In conclusion, the developed retrieval system demonstrates the effectiveness of integrating image and text data for content-based retrieval. By leveraging feature extraction techniques, similarity computation, and composite score computation, the system offers users a powerful tool for finding relevant information across different modalities. Further research and development could focus on addressing the identified challenges and implementing potential improvements to enhance the system's performance and usability.

6. References:

- Scikit-learn Documentation: https://scikit-learn.org/

- NLTK Documentation: https://www.nltk.org/

- Keras Documentation: https://keras.io/

- OpenAI: https://openai.com/

This report outlines the methodology, challenges, results, and potential improvements in building a retrieval system for image and text data, providing insights into the development process and future directions for enhancement.