

MULTIMEDIA IMAGE RETRIEVAL SYSTEM USING MACHINE LEARNING ON DISTRIBUTED FRAMEWORK

ABSTRACT

The rapid growth of multimedia data has necessitated the development of efficient image retrieval systems. In this study, we propose a novel multimedia image retrieval system that leverages the power of machine learning within a distributed framework, specifically utilizing the Hadoop ecosystem. Our system aims to enhance the accuracy, efficiency, and scalability of image retrieval by integrating advanced machine learning techniques and the distributed computing capabilities offered by Hadoop.

To improve retrieval accuracy, we employ state-of-the-art machine learning algorithms, including deep learning and convolutional neural networks (CNNs), to extract high-level features from images. These extracted features capture the semantic content and visual characteristics of the images, enabling more accurate similarity matching during the retrieval process. By leveraging the power of deep learning, our system can effectively represent and understand the visual information present in images, leading to improved retrieval performance.

In order to handle the vast amount of multimedia data, we adopt a distributed architecture using Hadoop, a widely-used open-source framework for distributed computing. Hadoop provides a scalable and fault-tolerant environment for processing and analyzing large-scale datasets. Within this distributed framework, we distribute the retrieval tasks across multiple computing nodes, enabling parallel processing and improved efficiency. Each node in the Hadoop cluster is equipped with computational resources and trained machine learning models, allowing for simultaneous retrieval of images from different parts of the dataset.

Furthermore, our system incorporates sophisticated indexing and organization techniques to optimize the storage and retrieval of images within the Hadoop Distributed File System (HDFS). The extracted high-level features are utilized to create compact and informative representations of images, enabling efficient indexing and faster retrieval times. This indexing scheme, combined with the distributed computing capabilities of Hadoop, allows for more precise and efficient searching of images based on their content and visual similarity.

We evaluate the performance of our proposed system using diverse and challenging multimedia datasets within the Hadoop environment. Extensive experiments are conducted to compare the retrieval performance of our system against traditional image retrieval methods. The experimental results demonstrate the superior accuracy and efficiency of our system, with significant improvements in retrieval precision and recall.

In conclusion, our multimedia image retrieval system presents a novel approach that integrates advanced machine learning techniques, such as deep learning and CNNs, within a

distributed computing framework utilizing Hadoop. The combination of machine learning and Hadoop offers a powerful solution for accurate and efficient image retrieval, capable of handling large-scale multimedia datasets. Our research contributes to the field by demonstrating the effectiveness of machine learning and Hadoop in capturing and utilizing high-level image features, and the potential for distributed computing in multimedia retrieval tasks.