Personal Home Media Server Using Machine Learning-Based NLP Processing

Author: Manay Lodha

I am developing a personal home media server designed to efficiently organize and stream movies I have legally extracted from my Blu-ray collection. This project utilizes natural language processing (NLP) techniques, specifically word embeddings and recurrent neural networks, to improve metadata management and content categorization within my media server.

Project Overview

The core of this project leverages **Word2Vec embeddings** and **LSTM-based neural networks** to analyze and process metadata associated with my movie collection. By training on a dataset of user reviews, descriptions, and related metadata, this system enhances the categorization and retrieval of movies in my personal library. This allows for **smarter search functionality**, personalized recommendations, and an intuitive media browsing experience.

Technical Approach

- **Dataset Processing:** The system utilizes a dataset containing movie reviews, which undergoes **preprocessing** by removing HTML tags, special characters, and case normalization.
- **Word Embeddings:** Word2Vec embeddings are trained on the cleaned text corpus, providing a meaningful vector representation of movie-related terms.
- **Neural Network Architecture:** A **bi-layer LSTM model** with dropout layers is implemented to learn patterns in the movie metadata, enabling sentiment-aware tagging and enhanced categorization.
- **Integration with Media Server:** The processed metadata will be used to **dynamically tag and organize** movies, allowing efficient search and streaming based on genre, user sentiment, or custom-defined categories.

Planned Features for Home Media Server

- **AI-driven movie recommendations** based on extracted metadata.
- **Automated tagging** of movies using NLP-based feature extraction.
- **Custom streaming interface** with intelligent search capabilities.
- **Scalability for larger collections**, allowing integration with additional datasets.

This project showcases how machine learning techniques can enhance personal media experiences by transforming how we store, search, and stream digital content. The system will be deployed locally, ensuring complete privacy and efficiency in organizing my personal Blu-ray movie library.