

Chapter 10

Enhancing Big Data Analytics and Recommendation Systems With Python: Hierarchical Clustering and Filtering Using Slicing Technique

P. Thamilselvan
Bishop Heber College, India

ABSTRACT

The advancement of technology has led to an exponential increase in the volume, velocity, and variety of data generated, necessitating the development of effective methods for analyzing and extracting valuable insights from large datasets. This research focuses on enhancing big data analytics and recommendation systems using Python, specifically employing hierarchical clustering and a filtering approach with the slicing technique. This study proposes a novel approach to leverage Python's capabilities in processing and analyzing big data. Hierarchical clustering algorithms organize and structure data into hierarchical groups, enabling efficient exploration and extraction of relevant information. Additionally, a filtering mechanism is integrated with the slicing technique, allowing for identifying and extracting specific subsets of data based on predefined criteria. Experiments are conducted using real-world datasets in the context of recommendation systems to evaluate the approach's effectiveness.

DOI: 10.4018/978-1-6684-7100-5.ch010

INTRODUCTION

Big Data consists of vast data from diverse and independent sources, making traditional software difficult to manage (Beyer & Laney, 2012). Analyzing and extracting helpful information for future operations is a crucial challenge in Big Data applications. Recommendation Systems (RSs) are sophisticated instruments that aid consumers in making product or service selections. Recommendation systems confront two formidable challenges in applications containing enormous amounts of data. The first challenge is making decisions promptly. It is crucial that recommendation systems efficiently process and provide user-satisfying suggestions in light of the immense quantities of available data. The second challenge is generating optimal service recommendations from various alternatives. As the number of services and options increases, it becomes more challenging to identify the most relevant and customized offers for users. These obstacles must be surmounted for recommendation systems to provide timely and high-quality recommendations within the context of big data applications.

Clustering to integrate similar services has proven highly effective at reducing aggregate data size (Rajaraman & Ullman, 2012). Clustering and collaborative filtering are the two primary phases of this procedure (Zheng et al., 2013). Each subset corresponds to a discrete cluster consisting of services with similar characteristics. It is possible to considerably reduce computational time by limiting the number of services subject to the collaborative filtering phase to a subset within a particular cluster. Notably, this approach leverages the inherent relevance of ratings ascribed to similar services within a given cluster, significantly improving the accuracy of recommendations derived from user ratings (Bellogín et al., 2013).

CLUSTERING

Clustering has the potential to significantly reduce data volume by clustering together similar services (Rajaraman & Ullman, 2012). Clustering and collaborative filtration are the two fundamental phases of this procedure. Clustering is an essential preparatory step that facilitates the subdivision of enormous datasets into smaller, more manageable subsets (Zheng et al., 2013). Each cluster consists of services with comparable characteristics, enabling a more precise analysis. It is possible to significantly reduce computational time during the collaborative filtering phase by limiting the number of services within a cluster relative to the total benefits. Moreover, given the inherent relevance of ratings about services within a cluster as opposed to ratings about disparate services, the accuracy of recommendations based on user ratings can be substantially improved (Bellogín et al., 2013).

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/enhancing-big-data-analytics-and-recommendation-systems-with-python/326085

Related Content

Orchestrating Multi-Agent Applications With Langflow and Large Language Models (LLMs)

Soly Mathew Biju, Sara Majeedand Pradnya Bhagwat (2026). *Advanced Interdisciplinary Applications of Large Language Models and AI Agents* (pp. 1-30). www.irma-international.org/chapter/orchestrating-multi-agent-applications-with-langflow-and-large-language-models-llms/391838

Geospatial Data Visualization With Folium

K. G. Suma, Gurram Sunitha, J. Avanija, Mohammad Gouse Galetyand Chinthapatla Pranay Varna (2024). *Geospatial Application Development Using Python Programming* (pp. 187-208). www.irma-international.org/chapter/geospatial-data-visualization-with-folium/347438

Vert.X for Kotlin Restful API

(2023). *Principles, Policies, and Applications of Kotlin Programming* (pp. 394-451). www.irma-international.org/chapter/vertx-for-kotlin-restful-api/323941

Exploring Python's Powerful Data Collections

Mastan Vali Shaikand Rajasrkaran Selvaraju (2023). *Advanced Applications of Python Data Structures and Algorithms* (pp. 162-181). www.irma-international.org/chapter/exploring-pythons-powerful-data-collections/326083

Machine Learning for Geospatial Analysis: Enhancing Spatial Understanding and Decision-Making

Robin Rohit Vincent, E. Sakthivel, Meena Kumari, F. Nishaand A. Rohini (2024). *Ethics, Machine Learning, and Python in Geospatial Analysis* (pp. 168-195). www.irma-international.org/chapter/machine-learning-for-geospatial-analysis/345909