

## Chapter 7

# Cognitive Mining for Exploratory Data Analytics Using Clustering Based on Particle Swarm Optimization: Cognitive Mining for Exploratory Data Analytics

**Suriya Murugan**

*Bannari Amman Institute of Technology, India*

**Sumithra M. G.**

*Bannari Amman Institute of Technology, India*

**Logeswari Shanmugam**

*Bannari Amman Institute of Technology, India*

### **ABSTRACT**

*This chapter examines the exploratory data analytics that require statistical techniques on data sets which are in the form of object-attribute-time format and referred to as three-dimensional data sets. It is very difficult to cluster and hence a subspace clustering method is used. Existing algorithms like CATSeeker are not actionable and its 3D structure complicates the clustering process, hence they are inadequate to solve this clustering problem. To cluster these three-dimensional data sets, a new centroid-based concept is introduced in the proposed system called clustering using particle swarm optimization (CPSO). This CPSO framework can be applied to financial and stock domain datasets through the unique combination of (1) singular value decomposition (SVD), (2) particle swarm optimization (PSO), and (3) 3D frequent item set mining which results in efficient performance. CPSO framework prunes the entire search space to identify the significant subspaces and clusters the datasets based on optimal centroid value.*

DOI: 10.4018/978-1-5225-7522-1.ch007

## INTRODUCTION

### Data Mining

Data mining is the process of discovering new patterns from large data sources. Knowledge discovery from databases (KDD) is an interdisciplinary subfield of computer science which is used for discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

The analysis step involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization and online updating. Data mining is not only for the analysis of large-scale data or information processing but is also generalized to any kind of computer decision support system, including artificial intelligence, machine learning, and business intelligence.

Data mining uses information from past data to analyze the outcome of a particular problem or situation that may arise. Data mining works to analyze data stored in data warehouses that are used to store that data that is being analyzed.

Managers also use data mining to decide upon marketing strategies for their product. They can use data to compare and contrast among competitors.

Data mining interprets its data into real time analysis that can be used to increase sales, promote new product, or delete product that is not value-added to the company. Data mining interprets its data into real time analysis that can be used to increase sales and promote new product. Data mining mostly is used in decision making process which is also called business intelligence. Business-related decision-making is made using data mining techniques. Data mining is the entire process of applying computer methodology for knowledge discovery.

Steps in data mining:

- **Data Cleaning:** It is a phase in which noise and irrelevant data are removed from the collection.
- **Data Integration:** In this stage, multiple data sources, often heterogeneous may be combined in a common source.
- **Data Selection:** At this stage, the data relevant to the analysis is decided on and retrieved from the data collection.
- **Data Transformation:** It is also known as data consolidation. It is a phase in which the selected data is transformed into forms appropriate for mining procedure.
- **Data Mining:** It is the crucial step in which clever techniques are applied to extract patterns potentially useful.
- **Pattern Evaluation:** In this step, strictly interesting patterns representing knowledge are identified based on given measures.
- **Knowledge Representation:** It is the final phase in which the discovered knowledge is visually represented to the user. This essential step uses visualization techniques to help users understand and interpret the data mining results. Finally, the output will be represented in some human readable format which will be easy to understand.

18 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/cognitive-mining-for-exploratory-data-analytics-using-clustering-based-on-particle-swarm-optimization/218395](http://www.igi-global.com/chapter/cognitive-mining-for-exploratory-data-analytics-using-clustering-based-on-particle-swarm-optimization/218395)

## Related Content

---

### Bayesian Kernel Methods: Applications in Medical Diagnosis Decision-Making Processes (A Case Study)

Arti Saxena and Vijay Kumar (2021). *International Journal of Big Data and Analytics in Healthcare* (pp. 26-39).

[www.irma-international.org/article/bayesian-kernel-methods/268416](http://www.irma-international.org/article/bayesian-kernel-methods/268416)

### Digital Sovereignty and Cross-Border Data Flows: Legal Fragmentation and Governance Challenges

Hemendra Singh (2026). *Digital Sovereignty in the Global Data Economy: Rights, Regulation, and Resistance* (pp. 291-318).

[www.irma-international.org/chapter/digital-sovereignty-and-cross-border-data-flows/412421](http://www.irma-international.org/chapter/digital-sovereignty-and-cross-border-data-flows/412421)

### Data-Driven Decision Making for Achieving Sustainability in Apparel Business Value Chain

Kamalendu Pal (2024). *Data-Driven Decision Making for Long-Term Business Success* (pp. 395-414).

[www.irma-international.org/chapter/data-driven-decision-making-for-achieving-sustainability-in-apparel-business-value-chain/335585](http://www.irma-international.org/chapter/data-driven-decision-making-for-achieving-sustainability-in-apparel-business-value-chain/335585)

### Effective E-Healthcare System: Cache Invalidation Mechanisms for Wireless Data Access in Mobile Cloud Computing

Harshit Sinha, Gaurav Raj, Tanupriya Choudhury and Praveen Kumar (2018). *International Journal of Big Data and Analytics in Healthcare* (pp. 10-27).

[www.irma-international.org/article/effective-e-healthcare-system/223164](http://www.irma-international.org/article/effective-e-healthcare-system/223164)

### A Study on Mathematical Modeling of HIV/AIDS, Malaria, and Dengue Fever

Preeti Kalra, Shoket Ali and Ather Aziz Raina (2024). *Exploring Medical Statistics: Biostatistics, Clinical Trials, and Epidemiology* (pp. 279-295).

[www.irma-international.org/chapter/a-study-on-mathematical-modeling-of-hiv-aids-malaria-and-dengue-fever/351179](http://www.irma-international.org/chapter/a-study-on-mathematical-modeling-of-hiv-aids-malaria-and-dengue-fever/351179)