Chapter 10
Exploratory Cluster Analysis Using Self-Organizing Maps:
Algorithms, Methodologies, and Framework

Nuno C. Marques
https://orcid.org/0000-0002-3019-3304
NOVA-LINCS, SST, Universidade NOVA de Lisboa, Portugal

Bruno Silva
https://orcid.org/0000-0001-7873-2111
EST, Polytechnic Institute of Setúbal, Portugal

ABSTRACT

As the volume and complexity of data streams continue to increase, exploratory cluster analysis is becoming increasingly important. In this chapter, the authors explore the use of artificial neural networks (ANNs), particularly self-organizing maps (SOMs), for this purpose. They propose additional methodologies, including concept drift detection, as well as distributed and collaborative learning strategies and introduce a new open-source Java ANN library, designed to support practical applications of SOMs across various domains. By following our tutorial, users will gain practical insights into visualizing and analyzing these challenging datasets, enabling them to harness the full potential of our approach in their own projects. Overall, this chapter aims to provide readers with a comprehensive understanding of SOMs and their place within the broader context of artificial neural networks. Furthermore, we offer practical guidance on the effective development and utilization of these models in real-world applications.

DOI: 10.4018/978-1-6684-9591-9.ch010

Copyright © 2023, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

INTRODUCTION

Self-organizing maps (SOMs), also known as Kohonen maps (Kohonen, 1982), represent a significant advancement in artificial intelligence, providing a robust framework for data visualization and analysis. Their ability to transform complex, high-dimensional data into simple, comprehensible visual representations has made them an indispensable tool in a variety of fields, from computer science to bioinformatics, finance, and beyond. Since their inception in the early 1980s by Teuvo Kohonen, SOMs have significantly contributed to our understanding of complex systems, pattern recognition, and data representation (Kohonen, 2001). SOMs provide a robust method for visualizing and comprehending intricate data structures. By organizing high-dimensional input data into a low-dimensional grid, they facilitate the identification of hidden patterns, clusters, and relationships within the data, which might otherwise be challenging to discern (Vesanto, 1999).

In many scientific and engineering domains, researchers often encounter datasets with numerous variables and complex relationships. SOMs offer a unique solution to this challenge, enabling visualization and interpretation. The ability of self-organizing maps to uncover latent patterns within datasets has played a pivotal role in data mining and knowledge discovery. By identifying clusters and similarities in data, SOMs support the exploration of large datasets to extract valuable information. SOMs have proven instrumental in discovering new relationships, trends, and correlations, enabling researchers and scientists to gain valuable insights into data distributions and spatial relationships (Oja, Kaski, & Kohonen, 2003). Today, SOMs continue to empower informed decision-making and drive innovation across diverse domains, including genomics, finance, environmental sciences, marketing, healthcare, and social sciences.

SOMs have served as a crucial milestone in the development of artificial neural networks. By demonstrating how a simplified model of the brain's self-organization could be applied to data analysis, SOMs laid the foundation for subsequent advancements in neural network research. They have contributed to the evolution of deep learning, reinforcement learning, and other neural network architectures, leading to breakthroughs in fields such as computer vision, natural language understanding, and robotics (Goodfellow, Bengio, & Courville, 2016).

This chapter aims to guide readers through the role of SOMs in machine learning and introduce a publicly available neural network framework for exploratory cluster analysis. We will explore the practical application of these techniques using the *Wine Dataset*, a popular choice for machine learning and data mining tasks. By following along, readers will gain practical insights into the power of SOMs in handling complex, multi-dimensional data.

The chapter commences with an introduction to the capabilities of SOMs. This is followed by a detailed examination of the *Wine Dataset* using SOMs, where we elucidate complex relationships through the Unified Distance Matrix (U-Matrix) and Component Planes. The *Wine Dataset* serves as a comprehensive case study for SOMs, and we further explore the component planes, a visualization technique that shows the distribution of different feature values in the map, and a *feature clustering* technique based on hierarchical clustering. The focus then shifts to the application of SOMs in data stream contexts, specifically through the Ubiquitous Self-Organizing Map (UbiSOM). We present theoretical results and discuss the philosophical implications of this approach, as well as practical results from real-time exploratory analysis. We introduce the UbiSOM Library and discuss the application of *multiSOM* visualization with UbiSOM. This technique enhances interpretability by enabling simultaneous visualization of multiple dimensions and multiple SOMs. Following this, we illustrate the application of multiple SOMs to the

25 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/exploratory-cluster-analysis-using-self-organizing-maps/332606

Related Content

Designing Interactive Architecture: Lessons Learned from a Multi-Professional Approach to the Design of an Ambient Computing Environment
Mikael Wiberg (2009). *International Journal of Ambient Computing and Intelligence (pp. 1-18).*
www.irma-international.org/article/designing-interactive-architecture/34032

Understanding AI: A Primer for Leaders
(2024). *Holistic Approach to AI and Leadership (pp. 25-49).*
www.irma-international.org/chapter/understanding-ai/349172

An Empirical Performance Measurement of Microsoft's Search Engine and its Comparison with Other Major Search Engines
Xiannong Meng, Song Xingand Ty Clark (2007). *International Journal of Intelligent Information Technologies (pp. 65-81).*
www.irma-international.org/article/empirical-performance-measurement-microsoft-search/2419

Optimal Strategy for the Smartphone Industry in Taiwan: HTC Case Study
Yi-Fen Chen, Bi-Chu Chen, Wen-Yu Chen, Chia-Wen Tsaiand Wei-Hung Lin (2012). *International Journal of Intelligent Information Technologies (pp. 62-79).*
www.irma-international.org/article/optimal-strategy-smartphone-industry-taiwan/74830

Agent-Based Modeling for Simulation of Complex Business Systems: Research Design and Validation Strategies
Christopher Langdon (2008). *Intelligent Information Technologies: Concepts, Methodologies, Tools, and Applications (pp. 1852-1861).*
www.irma-international.org/chapter/agent-based-modeling-simulation-complex/24377