

# Chapter 10

## Deep Clustering

**M. Parimala Boobalan**  
VIT University, India

### ABSTRACT

*Clustering is an unsupervised technique used in various application, namely machine learning, image segmentation, social network analysis, health analytics, and financial analysis. It is a task of grouping similar objects together and dissimilar objects in different group. The quality of the cluster relies on two factors: distance metrics and data representation. Deep learning is a new field of machine learning research that has been introduced to move machine learning closer to artificial intelligence. Learning using deep network provides multiple layers of representation that helps to understand images, sound, and text. In this chapter, the need for deep network in clustering, various architecture, and algorithms for unsupervised learning is discussed.*

### INTRODUCTION

Growth in digital data and storage methodologies has resulted in collection of a huge database. The process of extracting or finding relevant and hidden information from large databases in the field of Data Mining is a powerful technology with great potential for analysis of meaningful information in data warehouse. Knowledge discovery task involves various steps, selection of target data, pre-processing of data, transformation, extracting meaningful pattern and interpreting discovered pattern. It also involves the process of extracting hidden information from large databases, which is a powerful tool to help companies to focus on predicting behavior of future trends and improve performance of the organization by analyzing future market trends.

The process of clustering involves grouping similar objects into one cluster and dissimilar objects into different cluster. The quality of clusters is evaluated based on two measures namely, intra-cluster similarity and inter-cluster similarity. The similarity among objects within the cluster is defined as intra-cluster similarity whereas the similarity of objects in different clusters is called as inter-cluster similarity. The cluster with low inter-cluster similarity and high intra-cluster similarity is considered to be a quality cluster. Clustering is classified into different types namely Partition based clustering, Hierarchical clustering, Model-based Clustering, Density based clustering and Graph clustering.

DOI: 10.4018/978-1-5225-7862-8.ch010

## Deep Clustering

The important task in partition based clustering is to group set of objects into  $k$  disjoint points such that points within group are similar. Given a set of  $n$  data points  $\{x_1, x_2 \dots x_n\}$ , the algorithm partition into  $k$  clusters  $\{C_1, \dots, C_k\}$  such that each data point  $x_i$  is assigned to a unique cluster  $C_i$ . There are two types of Hierarchical clustering namely Agglomerative (merge) and Divisive (divide). In the Agglomerative method, clusters are measured on the basis of distance. Initially, each object is formed as a cluster and then cluster centroids that are close to each other are merged. The Divisive method starts with a single cluster that consists of all data points and iteratively splits into various subgroups. This process continues until each cluster contains single object.

Density based clustering is a method that groups objects on the basis of density. DBSCAN algorithm is a widely used density based algorithm that detects clusters with arbitrary shape and varied density. The main idea of this algorithm is that the number of objects within neighborhood must be greater than or equal to the threshold points.

Clustering technique finds extensive application in the field of business and financial data, bioinformatics, telecommunication and health applications. Credit card holders can be grouped on the basis of usage of card, purchase pattern, money spent on card, frequency of use and location of a card used. This information can be very useful for market analysis to find a group to which promotional activities can be targeted that might be of interest. This analysis can be of mutual benefit to both card holders and sellers. Market analysis is based on lifestyle, past purchase behavior or their demographic characteristics.

Clustering can also be applied in a chain store which wants to examine profit of outlets similarly placed, on the basis of variables like social neighborhood, purchase pattern, vicinity to other shops and so on. Cluster analysis has also been used widely in areas of medicine such as psychiatry, disease modelling, gene modelling and disease diagnostics. The other applications of clustering are grouping policy holders with average claim cost, creating thematic maps by grouping feature space, document classification and cluster weblog data to discover groups of similar patterns.

Search engines initially try to group similar objects in one cluster and dissimilar objects in different cluster. The result of searching for an object is based on the nearest similar object of key data. Efficient searching depends on quality of cluster derived. Clustering algorithm can also be applied in Wireless Sensor Network by detecting the cluster head which collects the complete information from the nodes in a cluster. It can also be effectively used in academic community to monitor the progress of students in academics where each group denotes different levels of performance.

Clustering process is widely utilized in Health applications to group cancerous data set and to locate people with similar symptoms. Clustering plays an essential role behind the search engines that initially try to group similar objects in one cluster and dissimilar objects in other cluster. The result of search for an object is based on nearest similar object of key data and depends on quality of cluster derived.

Over the last few years Deep Learning was applied to hundreds of problems, ranging from computer vision to natural language processing. In many cases Deep Learning outperformed previous work. Deep Learning is heavily used in both academia to study intelligence and in the industry in building intelligent systems to assist humans in various tasks. Both machine and deep learning are subsets of artificial intelligence, but deep learning represents the next evolution of machine learning. In machine learning, algorithms created by human programmers are responsible for parsing and learning from the data. They make decisions based on what they learn from the data. Deep learning learns through an artificial neural network that acts very much like a human brain and allows the machine to analyze data in a structure very much as humans do. Deep learning machines don't require a human programmer to tell them what to do

14 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/deep-clustering/227851](http://www.igi-global.com/chapter/deep-clustering/227851)

## Related Content

---

### Deep Learning in Early Detection of Alzheimer's: A Study

Anitha S. Pillai and Bindu Menon (2019). *Handbook of Research on Deep Learning Innovations and Trends* (pp. 153-162).

[www.irma-international.org/chapter/deep-learning-in-early-detection-of-alzheimers/227849](http://www.irma-international.org/chapter/deep-learning-in-early-detection-of-alzheimers/227849)

### Complex-Valued Boltzmann Manifold

Masaki Kobayashi (2009). *Complex-Valued Neural Networks: Utilizing High-Dimensional Parameters* (pp. 1-26).

[www.irma-international.org/chapter/complex-valued-boltzmann-manifold/6762](http://www.irma-international.org/chapter/complex-valued-boltzmann-manifold/6762)

### A Theoretical Framework for Parallel Implementation of Deep Higher Order Neural Networks

Shuxiang Xu and Yunling Liu (2016). *Applied Artificial Higher Order Neural Networks for Control and Recognition* (pp. 351-361).

[www.irma-international.org/chapter/a-theoretical-framework-for-parallel-implementation-of-deep-higher-order-neural-networks/152110](http://www.irma-international.org/chapter/a-theoretical-framework-for-parallel-implementation-of-deep-higher-order-neural-networks/152110)

### Fall Behavior Recognition Based on Deep Learning and Image Processing

He Xu, Leixian Shen, Qingyun Zhang and Guoxu Cao (2020). *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications* (pp. 1394-1409).

[www.irma-international.org/chapter/fall-behavior-recognition-based-on-deep-learning-and-image-processing/237941](http://www.irma-international.org/chapter/fall-behavior-recognition-based-on-deep-learning-and-image-processing/237941)

### Efficient Large-Scale Stance Detection in Tweets

Yilin Yan, Jonathan Chen and Mei-Ling Shyu (2020). *Deep Learning and Neural Networks: Concepts, Methodologies, Tools, and Applications* (pp. 667-683).

[www.irma-international.org/chapter/efficient-large-scale-stance-detection-in-tweets/237898](http://www.irma-international.org/chapter/efficient-large-scale-stance-detection-in-tweets/237898)