



Types Of Data Used In Cluster Analysis - Data Mining
January 28, 2020



[Home](#) [Data Mining](#) [Data Warehousing](#) [About Us](#) [Privacy Policy](#)

[STING, E](#)

April 06, 2020

Home > clustering > Grid-Based Clustering - STING, WaveCluster & CLIQUE

Grid-Based Clustering - STING, WaveCluster & CLIQUE

April 06, 2020



Attribute Oriented Induction In Data Mining - Data Characterization
February 05, 2020



Grid-Based Clustering

Grid-Based Clustering method uses a multi-resolution grid data structure.

(Partitional Clustering Methods - [read here](#))

(Hierarchical Clustering Methods - [read here](#))

(Density-Based Clustering Methods - [read here](#))

Several interesting methods

- › **STING** (a **S**Tatistical **I**Nformation **G**rid approach) by Wang, Yang, and Muntz (1997)
- › **WaveCluster** by Sheikholeslami, Chatterjee, and Zhang (VLDB'98) - A multi-resolution clustering approach using wavelet method
- › **CLIQUE** - Agrawal, et al. (SIGMOD'98)

[CONTACT US](#)

Name

Email *

Message *

STING - A Statistical Information Grid Approach

STING was proposed by Wang, Yang, and Muntz (VLDB'97).

In this method, the spatial area is divided into rectangular cells.

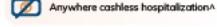
SOCIAL PLUGIN



MaxProtect Insurance

Coverage Starting from 1 crore!*

Enjoy benefits like



IL Take Care app



Get Covered Now!

Disclaimer:
*This policy provides coverage of 1 crore and other sum insured options as per requirement. For more details please refer to the policy T&C.
^Anywhere Cashless - As per policy T&C. This facility will be activated based on confirmation from the respective hospital. Cashless facility will be provided only if the hospital is a member of the network. ICICI Lombard GIC Ltd. is not liable for any claims arising out of non-membership of the hospital by the network.
ICICI Lombard GIC Ltd. under Kwality and Lombard logo belongs to ICICI Lombard GIC Ltd. The advertisement does not contain any endorsement or recommendation of the products or services mentioned. Please read the policy & its terms & conditions in detail on our website, terms & conditions, please read the terms & conditions of the policy & its addendum available on our website. ICICI Lombard GIC Ltd., Mumbai, Maharashtra - 400026, India. Reg. No. 195, Tel. No. 022 19191234, Fax No. 022 19191235, Email ID - info@icicilombard.com, Website - www.icicilombard.com, CIN (L12020MH40000L12020), MacPower, CIAL, PSONIC, FV1320, ADW18



Types Of Data Used In Cluster Analysis - Data Mining

January 28, 2020



Grid-Based Clustering - STING, WaveCluster & CLIQUE

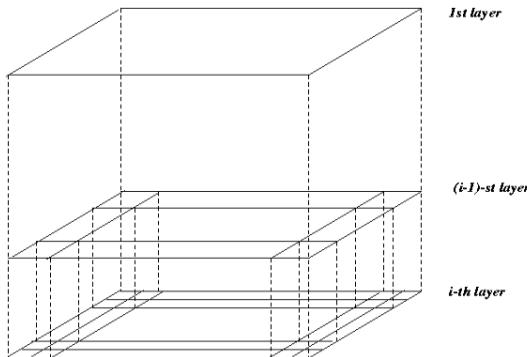
April 06, 2020



Attribute Oriented Induction In Data Mining - Data Characterization

February 05, 2020

There are several levels of cells corresponding to different levels of resolution.



For each cell, the high level is partitioned into several smaller cells in the next lower level.

The statistical info of each cell is calculated and stored beforehand and is used to answer queries.

MaxProtect Insurance

Coverage Starting from 1 crore!*

Enjoy benefits like



IL Take Care app



Get Covered Now!

Disclaimer:
*This policy provides coverage of 1 crore and other sum insured options as per requirement. For more details please refer to the policy T&C.
^Anywhere Cashless - As per policy T&C. This facility will be activated based on confirmation from the respective hospital. Cashless facility will be provided only if the hospital is a member of the network. ICICI Lombard GIC Ltd. is not liable for any claims arising out of non-membership of the hospital by the network.
ICICI Lombard GIC Ltd. under Kwality and Lombard logo belongs to ICICI Lombard GIC Ltd. The advertisement does not contain any endorsement or recommendation of the products or services mentioned. Please read the policy & its terms & conditions in detail on our website, terms & conditions, please read the terms & conditions of the policy & its addendum available on our website. ICICI Lombard GIC Ltd., Mumbai, Maharashtra - 400026, India. Reg. No. 195, Tel. No. 022 19191234, Fax No. 022 19191235, Email ID - info@icicilombard.com, Website - www.icicilombard.com, CIN (L12020MH40000L12020), MacPower, CIAL, PSONIC, FV1320, ADW18



The parameters of higher-level cells can be easily calculated from parameters of lower-level cell

- › Count, mean, s, min, max
- › Type of distribution—normal, uniform, etc.

Then using a top-down approach we need to answer spatial data queries.

Then start from a pre-selected layer—typically with a small number of cells.

For each cell in the current level compute the confidence interval.

CONTACT US

Name

Email *

Message *

MaxProtect Insurance
Coverage Starting from 1 crore!*

Enjoy benefits like

- Anywhere cashless hospitalization^
- IL Take Care app
- 24x7 customer support

Get Covered Now!

Disclaimer: *This policy provides coverage of 1 crore and other sum insured options as per requirement. For more details please refer to the policy brochure. ^Anywhere Cashless - As per policy T&C. This facility will be activated based on claim from the respective ICLIC Lombard General Insurance Company Limited. ICLIC Lombard General Insurance Company Limited is a registered trademark by ICLIC Lombard GIC Ltd under Indian and Lombard logo belongs to ICLIC Lombard GIC Ltd. The advertisement above is for ICLIC Lombard General Insurance Company Limited. ICLIC Lombard House, 4th, Veer Savarkar Marg, New Mumbai, Maharashtra - 400020. Tel: 022 61910123, Fax: 022 61910124, Email: info@icliclombard.com, Website: www.icliclombard.com CIN (L45200MH2006PLC29464, Regd No. 100198)



Types Of Data Used In Cluster Analysis - Data Mining

January 28, 2020



Grid-Based Clustering - STING, WaveCluster & CLIQUE

April 06, 2020



Attribute Oriented Induction In Data Mining - Data Characterization

February 05, 2020

Now remove the irrelevant cells from further consideration.

When finishing examining the current layer, proceed to the next lower level.

Repeat this process until the bottom layer is reached.

Advantages:

It is Query-independent, easy to parallelize, incremental update.

O(K), where K is the number of grid cells at the lowest level.

Disadvantages:

All the cluster boundaries are either horizontal or vertical, and no diagonal boundary is detected.

WaveCluster

It was proposed by Sheikholeslami, Chatterjee, and Zhang (VLDB'98).

It is a multi-resolution clustering approach which applies wavelet transform to the feature space

Save up to 75% now!
Don't miss out.
Shop now

- › A wavelet transform is a signal processing technique that decomposes a signal into different frequency sub-band.

It can be both grid-based and density-based method.

CONTACT US

Input parameters:

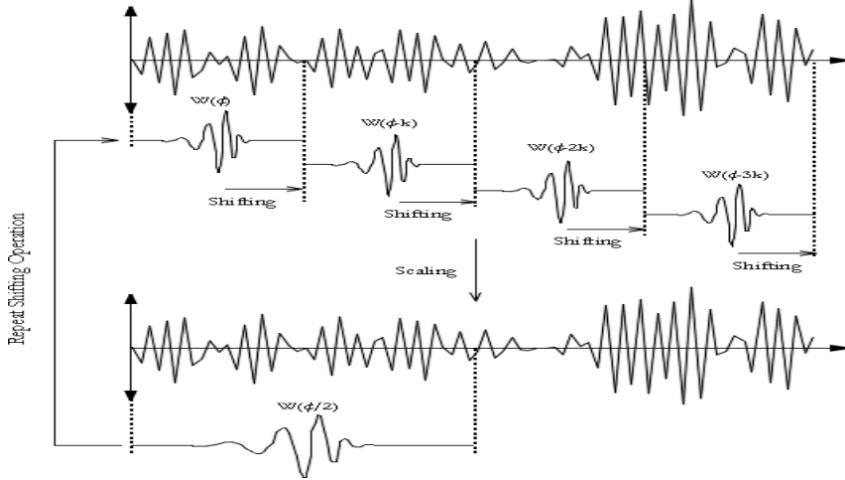
- › No of grid cells for each dimension
- › The wavelet, and the no of applications of wavelet transform.

Name

Email *

Message *

SOCIAL PLUGIN



How to apply the wavelet transform to find clusters

- It summarizes the data by imposing a multidimensional grid structure onto data space.

Types Of Data Used In Cluster Analysis - Data Mining
January 28, 2020

Grid-Based Clustering - STING, WaveCluster & CLIQUE
April 06, 2020

Attribute Oriented Induction In Data Mining - Data Characterization
February 05, 2020



Save up to 75% now!
Don't miss out.
[Shop now](#)

- These multidimensional spatial data objects are represented in an n-dimensional feature space.
- Now apply wavelet transform on feature space to find the dense regions in the feature space.
- Then apply wavelet transform multiple times which results in clusters at different scales from fine to coarse.

Why is wavelet transformation useful for clustering

- It uses hat-shape filters to emphasize region where points cluster, but simultaneously to suppress weaker information in their boundary.
- It is an effective removal method for outliers.
- It is of Multi-resolution method.
- It is cost-efficiency.

Major features:

- The time complexity of this method is O(N).
- It detects arbitrary shaped clusters at different scales.
- It is not sensitive to noise, not sensitive to input order.
- It only applicable to low dimensional data.

CLIQUE - Clustering In QUEst

It was proposed by Agrawal, Gehrke, Gunopulos, Raghavan (SIGMOD'98).

POPULAR POSTS



Types Of Data Used In Cluster Analysis - Data Mining

January 28, 2020



Grid-Based Clustering - STING, WaveCluster & CLIQUE

April 06, 2020



Attribute Oriented Induction In Data Mining - Data Characterization

February 05, 2020

Udemy

Save up to 75% now!
Don't miss out.

[Shop now](#)

CONTACT US

Name

Email *

Message *

SOCIAL PLUGIN



It is based on automatically identifying the subspaces of high dimensional data space that allow better clustering than original space.

POPULAR POSTS



Types Of Data Used In Cluster Analysis - Data Mining
January 28, 2020



Grid-Based Clustering - STING, WaveCluster & CLIQUE
April 06, 2020



Attribute Oriented Induction In Data Mining - Data Characterization
February 05, 2020



CLIQUE can be considered as both density-based and grid-based:

- › It partitions each dimension into the same number of equal-length intervals.
- › It partitions an m-dimensional data space into non-overlapping rectangular units.
- › A unit is dense if the fraction of the total data points contained in the unit exceeds the input model parameter.
- › A cluster is a maximal set of connected dense units within a subspace.

Partition the data space and find the number of points that lie inside each cell of the partition.

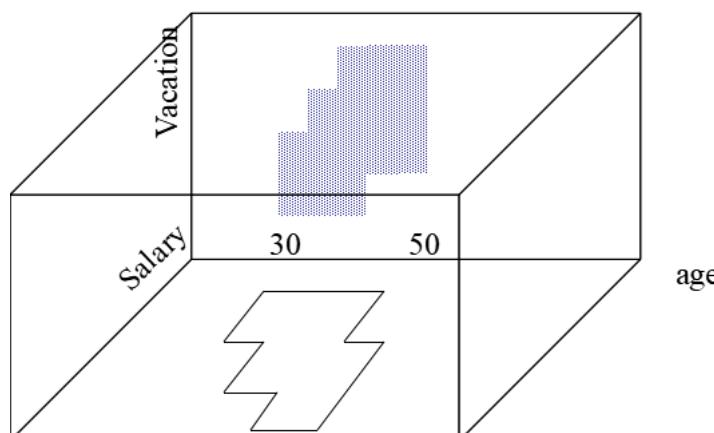
Identify the subspaces that contain clusters using the [Apriori principle](#).

Identify clusters:

- › Determine dense units in all subspaces of interests.
- › Determine connected dense units in all subspaces of interests.

Generate minimal description for the clusters:

- › Determine maximal regions that cover a cluster of connected dense units for each cluster.
- › Determination of minimal cover for each cluster.



CONTACT US

Name

Email *

Message *

Advantages

It automatically finds subspaces of the highest dimensionality such that high-density clusters exist in those subspaces.

SOCIAL PLUGIN



It is insensitive to the order of records in input and does not presume some canonical data distribution.

POPULAR POSTS



Types Of Data Used In Cluster Analysis - Data Mining
January 28, 2020



Grid-Based Clustering - STING, WaveCluster & CLIQUE
April 06, 2020



Attribute Oriented Induction In Data Mining - Data Characterization
February 05, 2020

It scales linearly with the size of input and has good scalability as the number of dimensions in the data increases.

Disadvantages

The accuracy of the clustering result may be degraded at the expense of the simplicity of the method.

Summary

Grid-Based Clustering -> It is one of the methods of cluster analysis which uses a multi-resolution grid data structure.

Subscribe us for more content on **Data**.



Tags clustering

[f Facebook](#) [Twitter](#) [p Pinterest](#) [in LinkedIn](#) [Email](#)

YOU MAY LIKE THESE POSTS



Model-Based Clustering - COBWEB

April 10, 2020



Grid-Based Clustering - STING, WaveCluster & CLIQUE

April 06, 2020



Density-Based Clustering - DBSCAN, OPTICS & DENCLUE

April 01, 2020

POST A COMMENT

0 Comments

To leave a comment, click the button below to sign in with Google.

SIGN IN WITH GOOGLE

CONTACT US

Name

Email *

Message *

[Home](#) [About Us](#)

Copyright 2021 @ Data Mining 365

SOCIAL PLUGIN





Types Of Data Used In Cluster Analysis - Data Mining

January 28, 2020



Grid-Based Clustering - STING, WaveCluster & CLIQUE

April 06, 2020



Attribute Oriented Induction In Data Mining - Data Characterization

February 05, 2020



The banner features the Udemy logo at the top right. To the left, there's promotional text: "Save up to 75% now! Don't miss out." Below this is a red button with the text "Shop now". On the right side of the banner is a portrait of a smiling man with a beard.