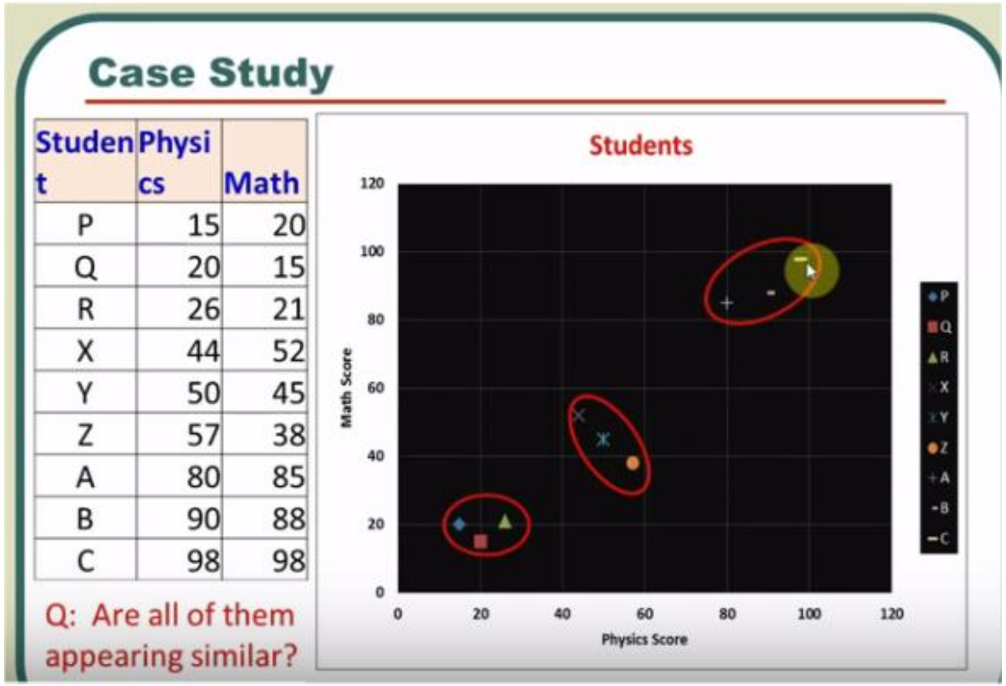
Cluster Analysis:
Global Clustering Methods

Outline

- Introduction
- Clustering Methods
- Clustering Categorical and Mixed Data
- Multiple Clusterings
- Temporal clustering



Student	Physics	Math
P	15	20
Q	20	15
R	26	21
X	44	52
Y	50	45
Z	57	38
A	80	85
B	90	88
C	98	98

(From: [Video])

Clustering

- is part of an exploratory data analysis where users have little a priori information.
- serves to *define groups automatically* (unsupervised learning).

Introduction

- Members of a cluster are closer to each other than to members of other clusters.
  - The definition of proximity is based on a distance measure that may incorporate different weights for the attributes.
- A cluster *summarizes homogeneus regions.*
- A cluster may contain subclusters – the understanding of this hierarchical relation is essential.

Introduction

Relation between *clustering* and *visual analytics*?

- „Since data organization and presentation are among the objectives of cluster analysis, graphic tools to display the results of clustering are important. They should let one „see" whether the grouping of the data has brought out any latent structure …" (Ling, 1973)
- „Graphic display is not only a useful aid to cluster analysis but essential for the assessment of real clusters … whether certain clusters can be seen as 'real'." (Ling, 1973)

Introduction

What is clustering good for?

- Learn the structure of data, e.g. define subgroups for customer segmentation (business analytics)
- A preprocess for selective visualization (show only certain clusters) or for focus-context visualization (show cluster representatives as overview and all instances as detailed view)
- A preprocess for classification, e.g. as input for a decision tree search

We have two tasks:

- Clustering to efficiently partition a dataset
- Displaying and exploring clustering results to interpret partioning and support reasoning

Introduction

- Clustering algorithms maximizes *intercluster coherence*
- Clustering: part of an explorative knowledge discovery process.
- When the distribution is largely unknown, clustering is performed iteratively *with different methods* and *parameters* until plausible results arise.
- When selecting a clustering method, researchers make assumptions about the data and employ a *clustering model*
- The cluster model relates to
  - the distribution of data,
  - the expected shapes of clusters and
  - their relations.

Introduction

According to the model, clustering is performed

- In a hierarchic or non-hierarchic manner
- In a fuzzy or binary manner (hard)
- In a deterministic or non-deterministic manner
- Using various *distance measures*

Clustering results in a decomposition of the data elements.

Outliers (not belonging to any cluster) are possible with some approaches

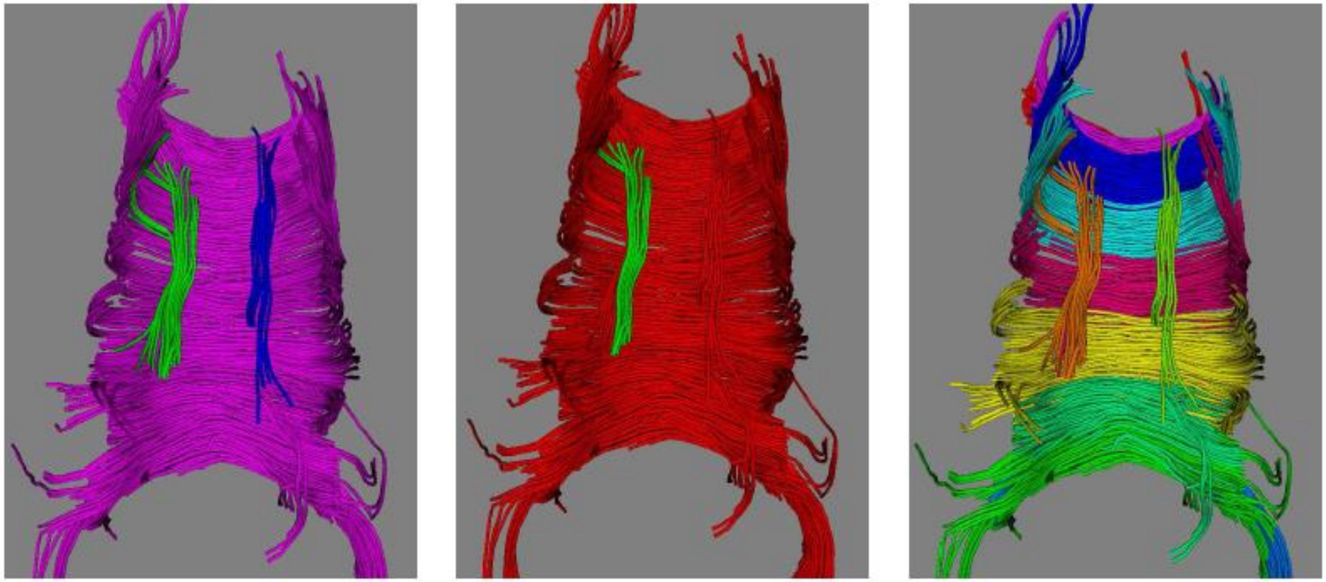
For further analysis and visualization it is often useful to determine a *cluster representative*

Introduction

Clustering is applied

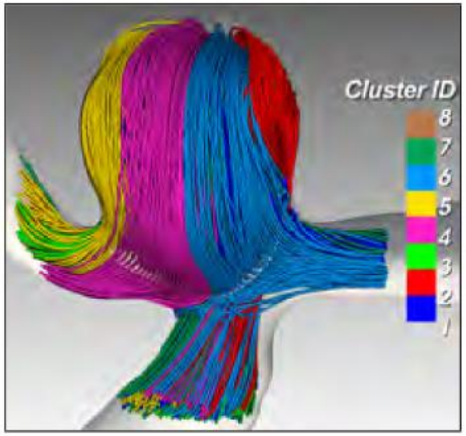
- to low and high-dimensional data,
- to documents (based on words),
- to scalar and categorical data or a mixture of both,
- to image data, such as CT and MRI,
- to streamlines (for supporting flow visualization),
- to fiber tracts (from brain MRI data) and
- even to time-dependent data where elements are clustered according to temporal changes.

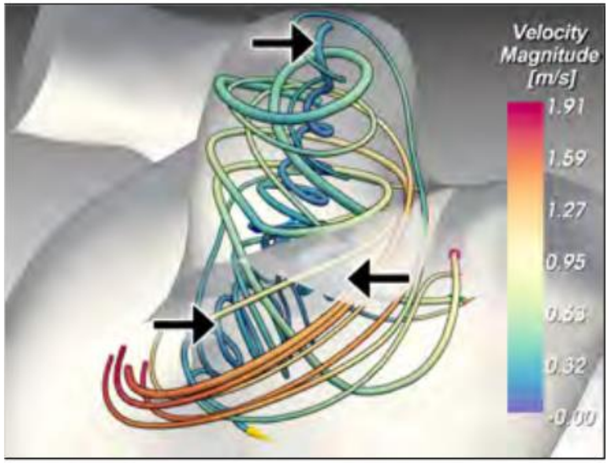
If applied to image data, clustering is similar to segmentation.



Different results of clustering of fiber tract data to identify crucial fiber tracts (Moberts, 2005). Left: correct, middle: incorrect, right: incomplete.

It is easier to interactively merge clusters (right image) than to separate an erroneously connected cluster.





Clustering of streamlines to support the visual analysis of blood flow in aneurysms.

Left: full set of streamlines color-coded according to cluster-id.
Right: (different aneurysm): Per cluster one cluster representative was determined (Oeltze, 2014) (Video)

Clustering Methods

- Choice of clustering methods strongly influences results
  – K-Means clustering favors Voronoi-cell-shaped clusters
  – Density-based clusters enables arbitrary shapes
- Even experienced users try different clustering methods and parameters

  → Quality assessment is essential to support this explorative process.

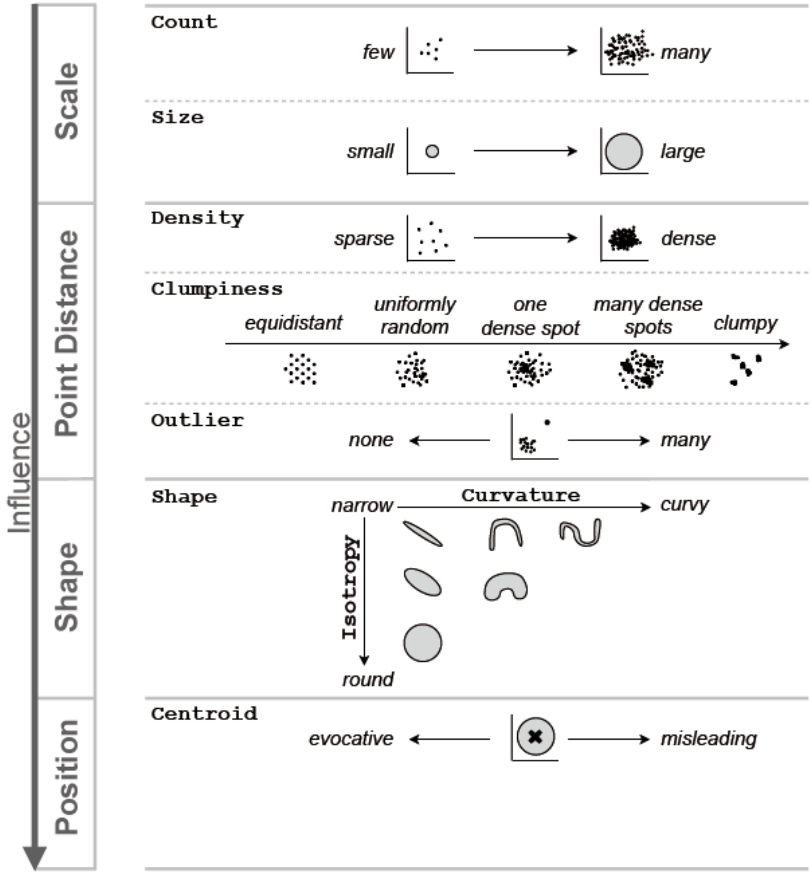
Clustering Methods

- Classification of Clustering problems
- Requirements
- Methods:
  - K-Means
  - Optics
  - Density-Based
  - Agglomerative Hierarchical
  - Clustering Ensembles

Clustering Methods: Classification

Why we need so many clustering methods?

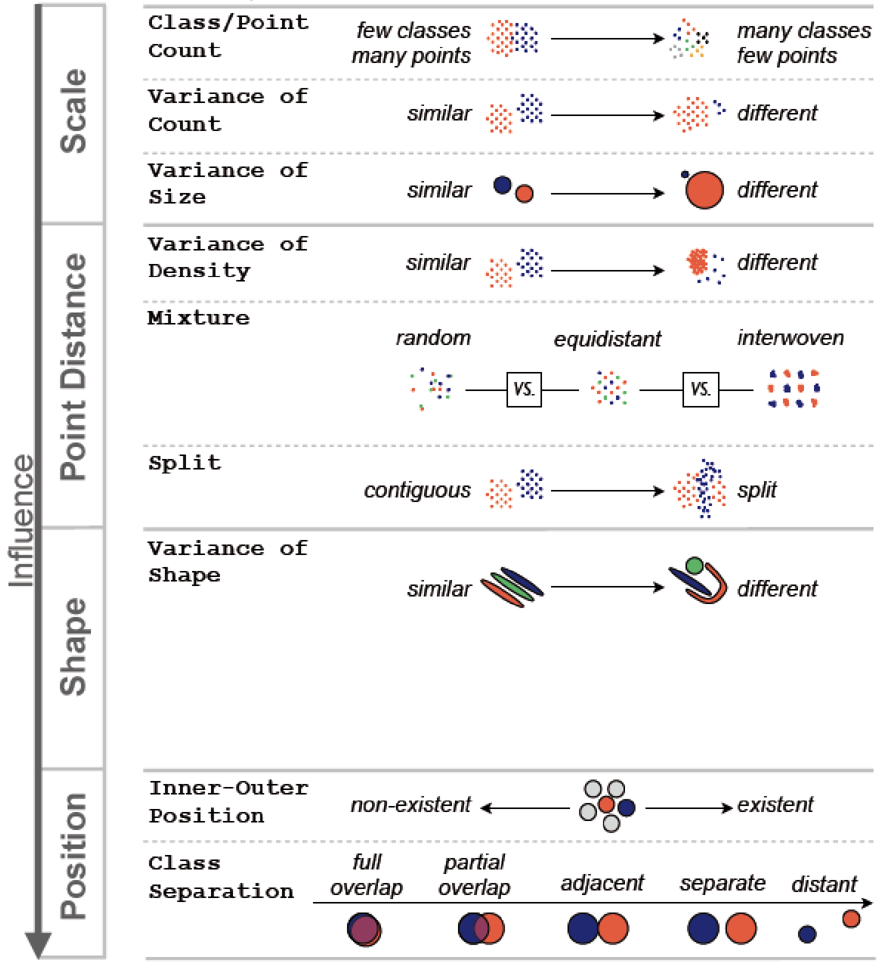
Because clusters are different.



Clusters are discriminated according to four categories of *within class* features. These features are not completely orthogonal, e.g. cluster shape and centroid mutually influence each other (From: Sedlmaier, 2012)

Why we need so many clustering methods (II)?

Because the relation between clusters are different.



Clusters are discriminated according to four categories of *between class* features. An essential feature is „Split" – whether or not clusters are spatially coherent (From: Sedlmaier, 2012)

Clustering Methods

Results:

- Hard clustering: Each object is assigned to a cluster (or considered as outlier)
- Fuzzy Clustering: assignment of degree of membership
- Per cluster, a representative may be generated
- Prototype generating clustering: Generates, hull surfaces, spherical hulls, regression lines

Clustering Methods: Requirements

Ideally clustering methods

- Are highly scaleable to many objects and many dimensions
- May create clusters of arbitrary shapes
- Are robust against noise
- Create plausible results, even for regions with different amounts of density

Clustering Methods: Models

Clustering may be based on

- A distance model (objects belong to a *cluster*$_i$ if they are closer to a *cluster center*$_i$ than to any other *cluster center*$_j$) – k-means and variants
- A density model (objects belong to a cluster if the density in the surrounding is high compared to the average density) – DB-SCAN, Optics
- A hierarchy model where clusters are assumed to exist at different levels

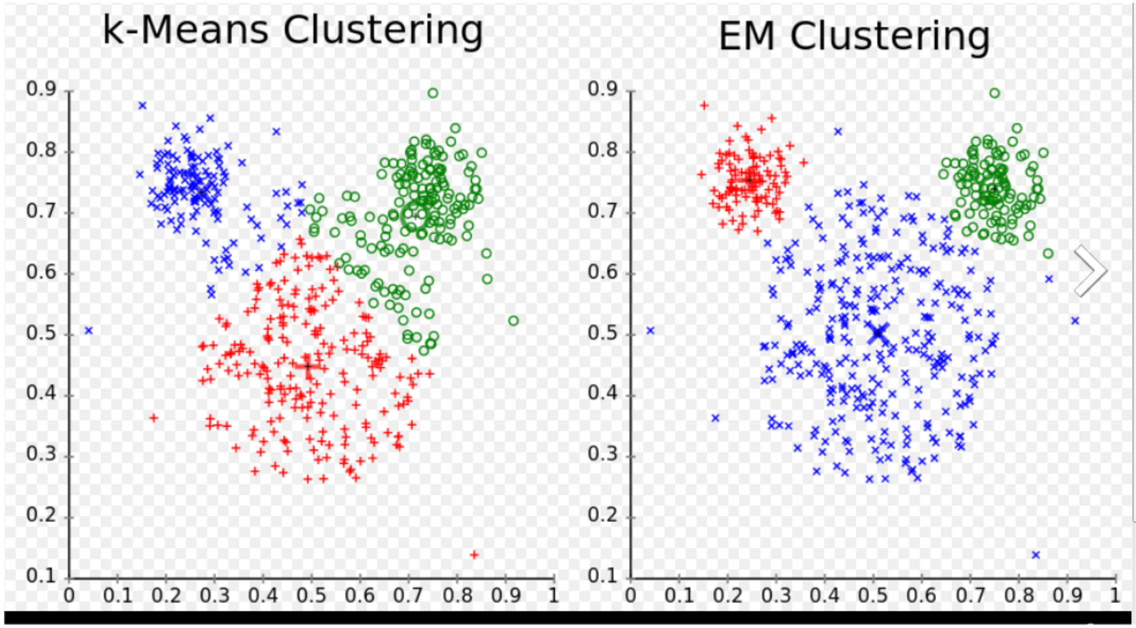
Clustering Methods: k-means

- k-means: Partioning of a dataset in *k* groups (MacQueen, 1967), follows a centroid model (a cluster is defined by its center)
- **Strategy**: Define *k* clusters such that the sum of squared differences from cluster elements to cluster centroids is minimal. → optimization problem, variance minimization

$$J = \sum_{i=1}^{k} \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2$$

- Since the problem is NP-hard, algorithms approximate the solution
- Algorithm starts with random initialization of *k* centers, e.g. with the first *k* elements, and then iteratively improves the selection until the influence on the result gets small.

Clustering Methods: k-means



Source: https://commons.wikimedia.org/wiki/File:ClusterAnalysis_Mouse.svg

Comparison of two clustering methods. The tendency to produce equally-sized clusters leads to worse results for k-means (compared to expectation maximization).

Ideally, k-means clustering produces a voronoi decomposition.

The boundaries of clusters with k-Means corresponds to a Voronoi decomposition based on the cluster centers.

Clustering Methods: k-means

Properties:

- The Euclidean distance metric restricts the application to numerical data.
- Clusters are restricted to be convex.
- Not robust against outliers, that often move the cluster centers considerably
- k must be known in advance (With cluster validity measures, k may also be estimated, but that is time-consuming)
  - The a priori knowledge of k is not typical for exploratory data analysis.
  - If k is chosen too small, clustering groups together wrong elements.
  - If k is choosen too high, clustering separates elements that belong together.
- The clusters should have approximately similar size (boundaries are in the middle between cluster centers).
- The result strongly depends on the initialization – it is not deterministic.

Part of many libraries, e.g. Weka,OpenCV

Clustering Methods: k-means

Improvements:

- Aiming at better performance, e.g., with hierarchical data structures such as the k-d tree or with better start points to improve convergence (k-Means++, k-Medoids)
  - K-Medoids choose the point nearest to the computed center point as new center
- Extend the results to fuzzy membership (*fuzzy c-means*). Membership of an element to a cluster is maximum for the cluster center and minimum at the cluster borders (Bezdek, 1981)
  - Often used in image segmentation, e.g. brain data with gray matter, white matter, cerebrospinal fluid
  - Similar properties like k-means

Clustering Methods: Fuzzy C-Means

With fuzzy clustering objects on the boundaries between cluster centers are not forced to fully belong to one of them, but are assigned membership degrees between 0 and 1 indicating partial memberships μ_{ik}.

The sum of the partial memberships is always 1 (Bezdek, 1981).

$$\mu_{ik} \in [0,1],\ 1 \leq i \leq N,\ 1 \leq k \leq c,$$

$$\sum_{k=1}^{c} \mu_{ik} = 1,\ 1 \leq i \leq N,$$

$$0 < \sum_{i=1}^{N} \mu_{ik} < N,\ 1 \leq k \leq c.$$

Clustering Methods: Fuzzy C-Means

The functional to be minimized reflects the partial memberships:

$$J(\mathbf{X};\mathbf{U},\mathbf{V}) = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^m \|\mathbf{x}_k - \mathbf{v}_i\|_A^2$$

The solution of this functional is a nonlinear optimization problem, solved e.g. with genetic algorithms or simulated annealing.

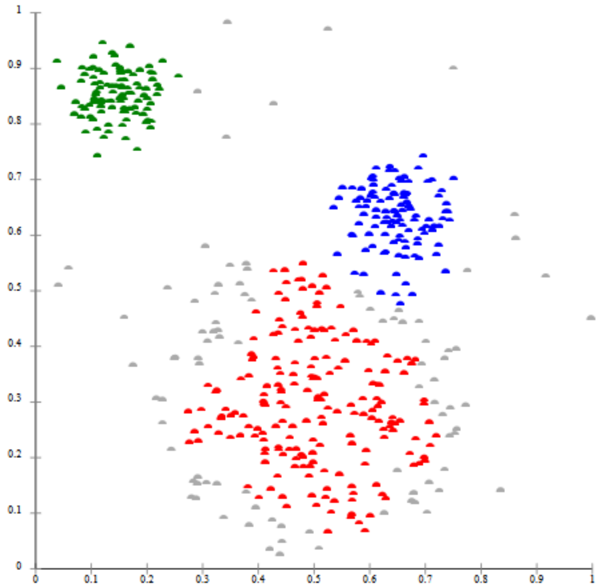
Like k-Means, fuzzy C-means generates hyperspherical convex clusters.

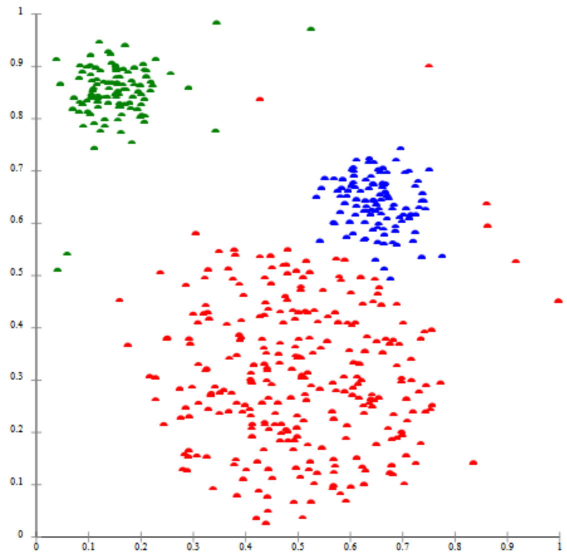
Fuzzy clustering with arbitrary shapes are based on modified distance functions, e.g. the Gustafsson Kessel algorithm.

Clustering Methods: Optics

- Several clustering methods are *density-based*, searching for clusters that are characterized by an above-average local density.
- Density-based methods do not require an a priori number of expected clusters and generate clusters of *arbitrary shapes*.
- Optics and DB-SCAN are essential density-based clustering techniques.

Clustering Methods: Optics

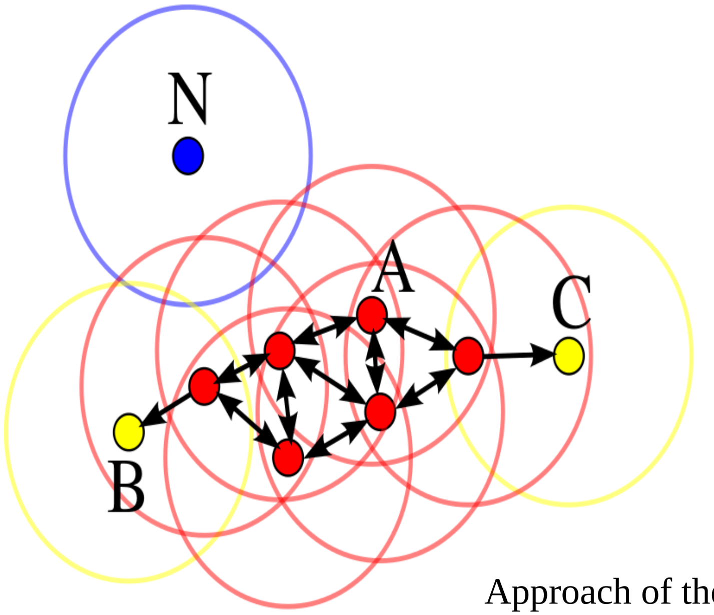




Clusters with different density are challenging to determine. DB-SCAN (left) generates many outliers, in particular close to the cluster with the lowest density (left), whereas OPTICS handles such situations well (right) (From: Wikipedia)

Clustering Methods: Density-Based

- DBSCAN (Density-Based Spatial Clustering of Applications with Noise, [Ester, 1996])
- Basic concept: “density-connectivity“ → two objects are *density connected* , if there is a chain of dense objects that connect the objects



Approach of the DBSCAN-method [12].

Clustering Methods: Hierarchical

- Hierarchical clustering generates a hierarchy of clustering results, typically shown in a dendrogramm
- Elements are connected and assigned to a cluster, if their distance is below a threshold.
- If the threshold is increased, low level clusters get connected to higher level clusters.
- Thus, a hierarchy arises bottom up (agglomerative)
- Essential parameters of hierarchical clustering are:
  - Distance function, e.g. Euclidean, Manhattan, Mahalanobis distance
  - Linkage criterion

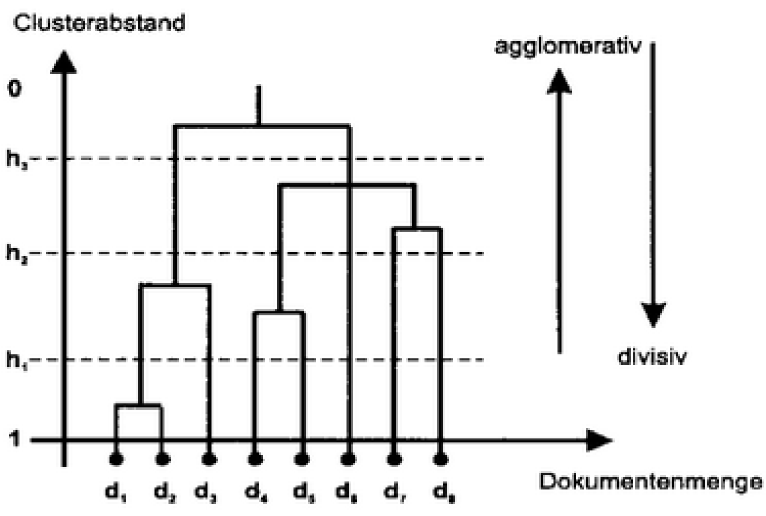
Linkage criteria define which clusters contain the closest point of elements that are merged. Popular criteria are:

- Single-linkage, considers the pair of points from different clusters with the smallest distance
  - produces long thin clusters
- Complete linkage
  - All element pairs are considered to determine a maximum distance. A pair of clusters is merged if this maximum distance is minimum.
- Ward's method
  - The error sum of squares for each pair of elements of two clusters is determined. Clusters are merged when this sum is minimal.

Clustering Methods: Hierarchical

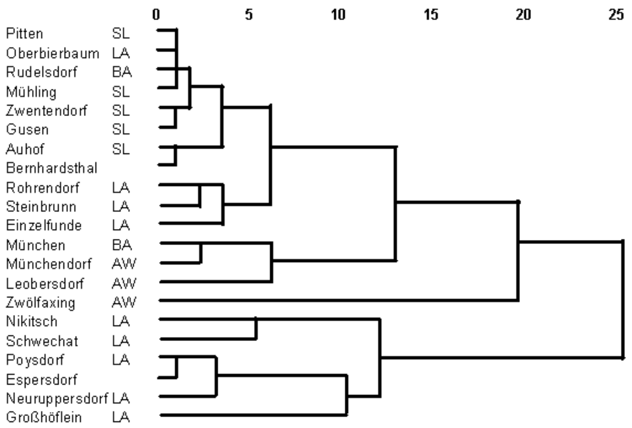
Dendrogramm as result of a hierarchical clustering.

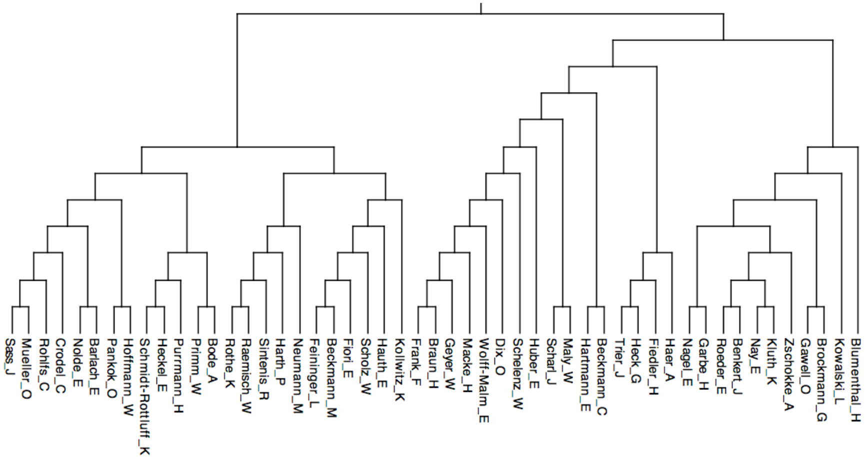
- Relative positions of subtrees are determined and integrated in a larger tree by placing subtrees as close as possible (Reingold, 1981)
- The method is not suitable for large element numbers
- For navigation, the user needs a facility to adjust the hierarchy level (feedback: number of sel. clusters)



Dendrogramm (of 8 documents).
(From: Dehner, 2007).

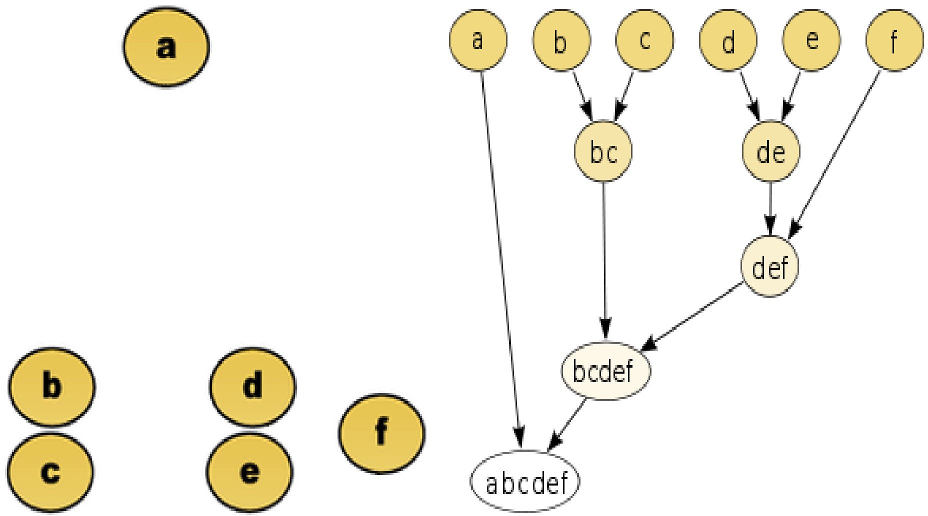
Clustering Methods: Hierarchical



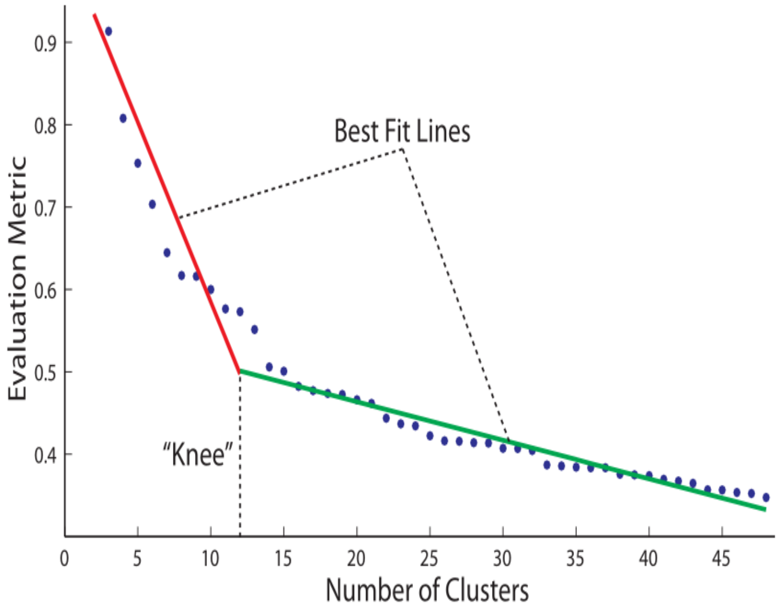


- Dendrogramm for larger datasets. Dendrograms may be horizontally or vertically aligned.
- **Left**: Dendrograms indicate very old skeletons, the location where they were found and their similarity (from: Link)
- **Right**: Artists are analyzed w.r.t. similarity, assessed by participation in exhibitions (from: Link).

Choosing the number of clusters/the threshold



Approach of the AHC-method (From: Wikipedia).



L-Method to determine an appropriate cluster number [From: Salvador, 2004].

Clustering Methods: Discussion

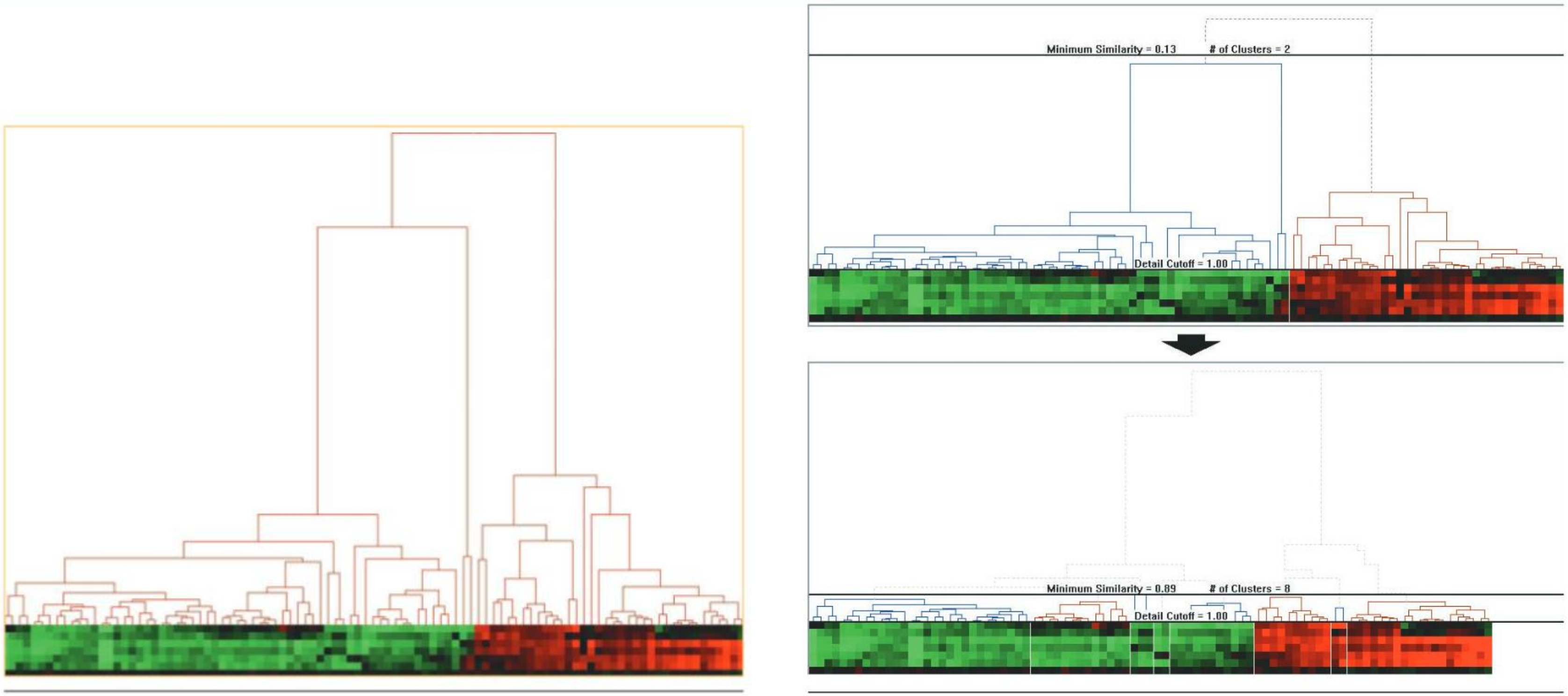
- Agglomerative hierarchical clustering is slow ($O(n^3)$, scalability req. not fulfilled)
- Refined methods employ hierarchical data structures (well-balanced trees) to represent cluster features (see [van Long, 2009] for an overview)
- Hierarchical clustering is available in Matlab, Mathematica, SPSS, SAS, …

Clustering Methods: Hierarchical

Interactive exploration of dendrograms (Seon, 2002)

- Select subtrees to collapse/display
- Adjust a global similarity threshold (range slider)
- Combine with a detailed view to display features of the selected cluster
- Enable comparison of two clustering results

Clustering Methods: Hierarchical



Larger dendrogramms representing clusters of genes according to expression profiles (color-coded). Genes with similar expression profiles may be responsible for different functions. The minimum similarity is reduced (right) to restrict the visualization (From: Seo, 2002)

Clustering Methods: Hierarchical

Clusters with the given similarity level are shown with alternating red and blue lines.

For the selected cluster (yellow) the gene names are highlighted and the expression profile data is embedded with a yellow line (From: Seo, 2002)

Clustering Categorical and Mixed Data

- High-dimensional data, e.g. census data often comprise ordinal and categorical data (low ranges and in case of categorical data no ordering)
- Most clustering methods are tailored to scalar data.
- For categorical data: The distance between two points is zero, if the attributes of both points are equal, otherwise 1 (*known as overlap metric*).
- For mixed data, an integration of binary and categorical data is necessary as part of an overall distance measure, e.g. the Heterogeneous Euclidean Overlap metric [Wilson, 1997]
- Alternative: discretize the scalar data to categorical data. Search for ideal cutoff points that minimize the entropy in the resulting classes [Fayyad, 1993]

Clustering Categorical and Mixed Data

- K-Prototypes (Huang, 1998)
  - example for a method appropriate for categorical data
  - requires an a priori estimation of the number of clusters

Clustering Categorical and Mixed Data

K-Prototypes (Huang, 1998)

- The overall distance measure is a sum of Euclidean distances of numerical values and the sum of categorical distances

$$d(X_1, Q_l) = \sum_{j=1}^{m_r} (x_{ij}^r - q_{lj}^r)^2 + \gamma_l \sum_{j=1}^{m_c} \delta(x_{ij}^c, q_{lj}^c)$$

- When applied to numeric data, K-Prototypes equals K-means
- γ_1 is derived from the variance of the numerical data.
- After initial clusterings, γ_1 is adapted

Discussion: later publications refined distance metrics and initialization, including K-modes by the same author.

Clustering with Constraints

- The results of clustering may be improved by further (in addition to cluster models) with a priori knowledge:
  - *Must-link constraints*: The user specifies a set of objects and „asks“ the algorithm to group these objects in the same cluster.
  - *Cannot-link constraints*: The selected objects should belong to different clusters.
- A clustering algorithm usually cannot fulfil all constraints within one clustering result.
- The degree to which the constraints are met may be an evaluation criterion.
- Clustering in general is an unsupervised method; clustering with constraints is supervised.

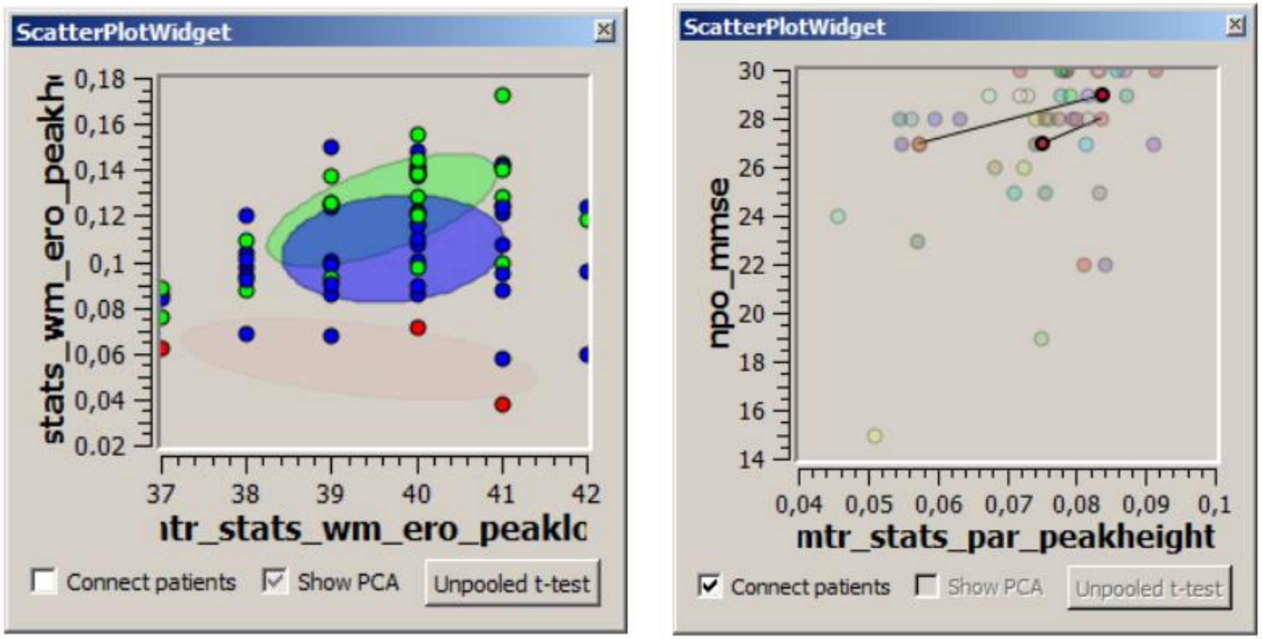
Clustering with Constraints

Examples:

- A clustering method should group debtors (Schuldner) w.r.t. the likelihood of non-payment.
- From the past, many data are available with income, marital status, children, rent, … labeled with either successful payment or non-payment.
- Clustering should find groups with payment and with non-payment, but should avoid clusters with both payment and non-payment.
- Better prediction who can and will pay
- Similar: Patients with symptoms of diseases that will or will not need hospitalization

Combining Cluster Analysis and Regression

- Statistical analysis is focussed on computing regression (linear, quadratic, logarithmic, Poisson, …)
- Regression is computed for the whole dataset thus „averaging“ over all regions
- Low correlation according to a regression type may be the result of different distinct regions where the correlation is high in at least one region
- Cluster analysis may reveal these distinct regions
- Regression analysis is then performed per cluster (later we discuss explorative regression in a separate lecture)



Cluster analysis in a small epidemiological dataset combined with linear regression (From: Stenwijk, 2010).

Temporal Clustering

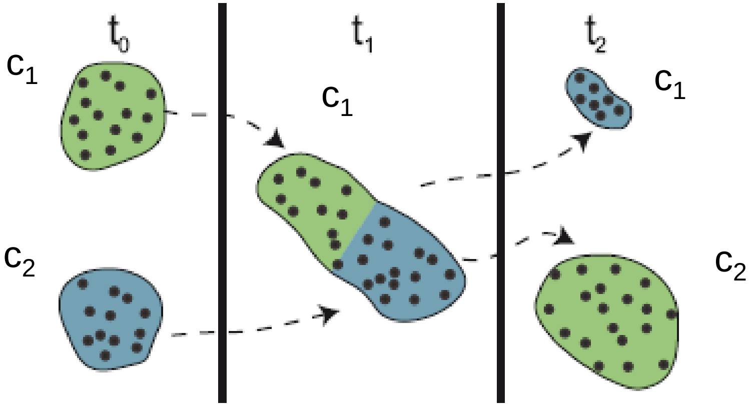
- Many HD datasets are available that reflect changes over time, e.g. data on student cohorts, cohorts in epidemiology, marketing-relevant data.
- Distance metric may consider each point in time indepently or cluster the time series w.r.t. similar trends (slightly increasing, steady, …)
- Clustering in these settings is useful but requires to reflect how clusters emerge, merge, split or disappear (temporal clustering, Turkay, 2011)
  - Temporal cluster view to assess the structural quality of clusters and type of structural changes. The cluster view displays the cluster quality and structural changes by encoding each cluster as axis in a parallel coordinate view.

Temporal Clustering

Frequent questions of analysts (see also Turkay, 2011):

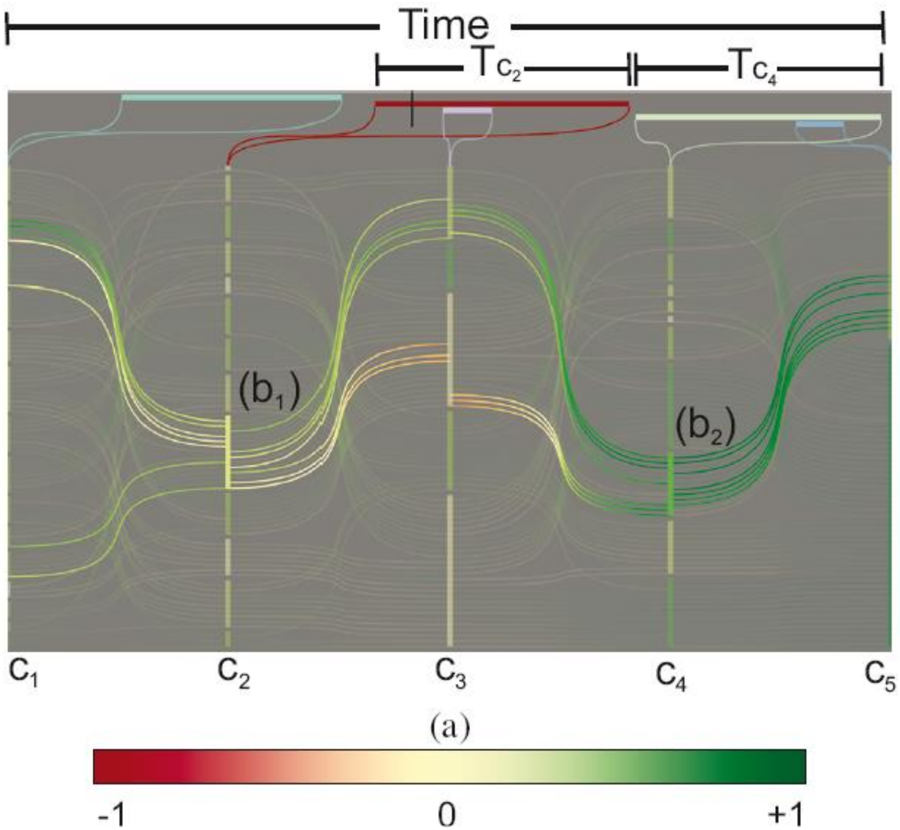
- How does the quality of clusters change over time?
- What type of structural changes do clusters exhibit?
- What are the characteristics of items that „move“ between clusters, e.g. move from slim to obese person or from healthy to pathological?
- What are the (relative) sizes of clusters over time?

Temporal Clustering



Clustering is performed indepently at time points t0, t1 and t2. Membership of items to clusters indicates that the cluster $c_1\_t_0$ and $c_2\_t_0$ merge to $c_1\_t_1$ and then split to $c_1\_t_2$ and $c_2\_t_2$. However, the trajectory of individual items is not recognizeable (From Turkay, 2011)

Temporal Clustering

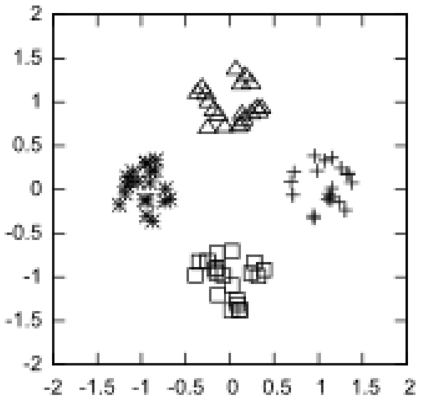


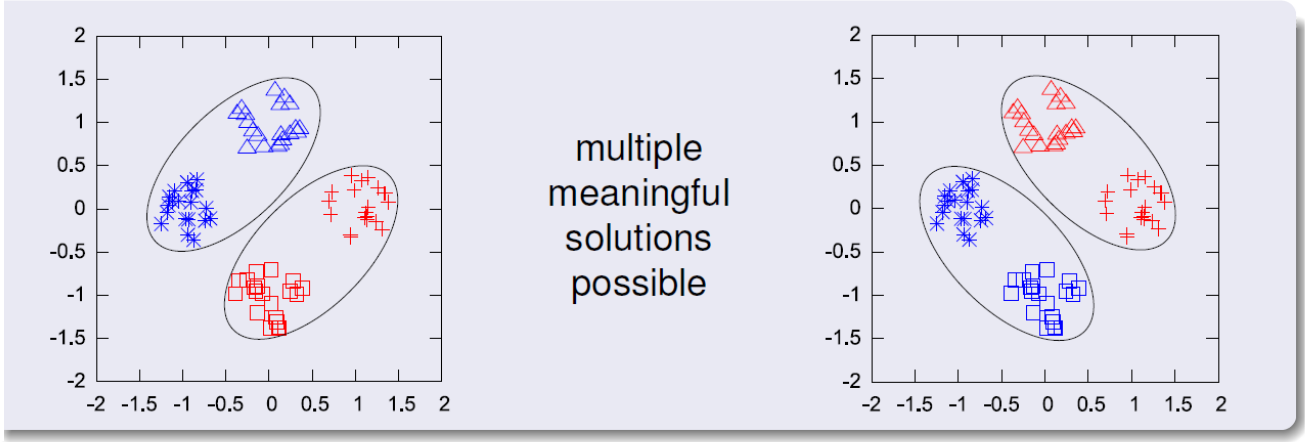
Cluster view – an adapted parallel coordinate plot. Temporal clusters are characterized by a temporal interval and by their members. Color indicates the silhouette coefficient of each cluster as a quality measure (next lecture) (From: Turkay, 2011).

Multiple Clustering

- Clustering results depend on
  - the algorithm used,
  - on the parameters and in case of stochastic algorithms
  - on the initialization.
- The display of a single clustering result is not ideal given the uncertainty about the right solution.
- Often, there is not even one right solution!
- More appropriate is the computation and visualization of a set of different clustering results (alternative clusterings).
- Naive approach: sample parameter space and stochastic behavior.
  - While this supports a cluster stability analysis, it leads to many similar clusters and does not support an analyst well.

Multiple Clustering



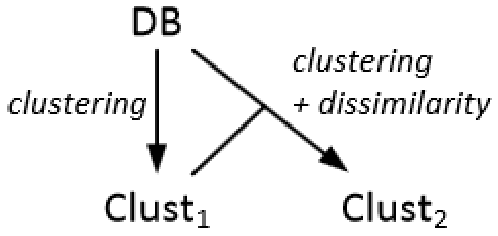


Alternative clustering results (From: Müller, 2013)

In an epidemiological application, Hielscher et al. 2014 performed 35 DB-Scan clustering with variations of the *minPoint* and *epsilon* parameters.

Strategies:

- Compute a first clustering C_1 and compute a second clustering C_2 that fulfils two properties:
  - The clustering quality is high.
  - The difference between C_2 and C_1 is high.



- Example: COALA (Bae, 2006)
- Extension: compute further clusterings C_{i+1} with a high quality and a high distance to all C_i.
- Compute multiple clusters simultaneously (aiming at high distances)

Applications

- There are many applications in different areas, e.g. in finance data, security analysis, and medicine.
- Document clustering is growing (Steinbach, 2000).
  - K-means and hierarchical clustering are employed.
  - Textmining and ontologies are integrated (Hotho, 2003).

Summary

- Clustering supports explorative data analysis
- A variety of clustering methods exists fitting to different *cluster models*.
- Multiple clusterings may provide alternative (and valid) perspectives on the data.
  - Visual analysis to understand the result and the influence of parameters
- Clustering is evaluated by means of quality measures (silhouette coefficient, …) and expert feedback based on appropriate visualizations
  - Cluster quality has many aspects – too many to be sufficiently represented by one scalar quality value

Summary and Outlook

- Global clustering restricted to 10-15 dimensions (Hund, 2016)
- For higher-dimensional data: search for clusterable subspaces and cluster within them (next lecture)
- Temporal clustering as a current research topic
  - Clusters arise, disappear, merge, change

References

Bae, E., & Bailey, J. 2006. COALA: A Novel Approach for the Extraction of an Alternate Clustering of High Quality and High Dissimilarity.In: IEEE Int. Conference on Data Mining (ICDM), pp. 53-62.

James C. Bezdek (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*.

Nan Cao, David Gotz, Jimeng Sun, Huamin Qu, "DICON: Interactive Visual Analysis of Multidimensional Clusters," *IEEE Trans. Vis. Comput. Graph.* Vol. 17(12):2581-2590, 2011

Matthias Dehmer (2007). *Strukturelle Analyse Web-basierter Dokumente*, DUV Verlag

M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in Proc. of Knowledge Discovery and Data Mining, KDD, pp. 226–231, 1996.

U. M. Fayyad and K. B. Irani (1993). Multi-interval discretization of continuous-valued attributes for classification learning, *Proc. of Internat. Joint Conf. on Artificial Intelligence*, pp. 1022-1029

S. Glaßer, K. Lawonn, B. Preim. Visualization of 3D Cluster Results for Medical Tomographic Image Data. In *Proc. of Conference on Computer Graphics Theory and Applications (VISIGRAPP/GRAPP)*, pp. 169-176, 2014

S. Glaßer, S. Roscher, B. Preim. Adapted Spectral Clustering for Evaluation and Classification of DCE-MRI Breast Tumors, *Bildverarbeitung für die Medizin (BVM),* pp. 198-203, 2014

RJ Hathaway, JC Bezdek. "Visual cluster validity for prototype generator clustering models", Pattern Recognition Letters 24 (9), 1563-1569

Hotho, Andreas, Steffen Staab, and Gerd Stumme. "Ontologies improve text document clustering." Proc. of IEEE International Conference on Data Mining, 2003.

Huang Z. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery*. 1998;2(3):283-304

References (II)

M Hund, I Färber, M Behrisch, A Tatu, T Schreck, DA Keim, T Seidl (2016). „Visual Quality Assessment of Subspace Clusterings", *Proc. of KDD Workshop on Interactive Data Exploration and Analytics (IDEA'16)*

Johnson, T., Kwok, I., and Ng, R.T. 1998. Fast computation of 2-dimensional depth contours. In Proc. Int. Conf. on Knowledge Discovery and Data Mining (KDD)

P. Klemm, L. Frauenstein, D. Perlich, K. Hegenscheid, H. Völzke, B. Preim. Clustering Socio-demographic and Medical Attribute Data in Cohort Studies, *Bildverarbeitung für die Medizin (BVM),* pp. 180-185, 2014

Kriegel H.-P., Kröger P., Zimek A.: Outlier Detection Techniques. *Tutorial. In: 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining* (PAKDD 2009).

R. L. Ling. 1973. A computer generated aid for cluster analysis. *Commun. ACM* 16, 6 (1973), 355-361.

S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," *IEEE Tools with Artificial Intelligence*, pp. 576–584, 2004.

T. C. Sprenger, R. Brunella, and Markus H. Gross. H-blob: a hierarchical visual clustering method using implicit surfaces. *Proc. of IEEE Visualization*, pp. 61–68, 2000

Bart Moberts, Anna Vilanova, Jarke J. van Wijk: Evaluation of Fiber Clustering Methods for Diffusion Tensor Imaging. *IEEE Visualization* 2005: 9-16

J. B. MacQueen: Some Methods for classification and Analysis of Multivariate Observations. In: Proc. of Berkeley Symposium on Mathematical Statistics and Probability, Univ. of California Press, 1967, pp. 281–297

E. Müller, S. Günnemann, I. Färber, and T. Seidl. "Discovering Multiple Clustering Solutions: Grouping Objects in Different Views of the Data", Tutorial at the International Conference on Machine Learning, Atlanta, USA (2013)

S. Oeltze, D. J. Lehmann, A. Kuhn, G. Janiga, H. Theisel, B. Preim: Blood Flow Clustering and Applications in Virtual Stenting of Intracranial Aneurysms. *IEEE Trans. Vis. Comput. Graph.* 20(5): 686-701 (2014)

References (III)

Oliynyk, Andriy; Bonifazzi, C.; Montani, F.; Fadiga, L., "Automatic online spike sorting with singular value decomposition and fuzzy C-mean clustering", *BMC Neuroscience*", 2012", Vol. 13(1): 1--19

Preparata, F. and Shamos, M. 1988. Computational Geometry: an Introduction. Springer

F. J. Rohlf. “Adaptive hierarchical clustering schemes”, Systematic Zoology, 19:58-82, 1970

E. M. Reingold and J. S. Tilford. Tidier drawings of trees. *IEEE Transactions on Software Engineering*, 7(2):223–238, 1981

Michael Sedlmair, A. Tatu, Tamara Munzner, Melanie Tory: A Taxonomy of Visual Cluster Separation Factors. *Comput. Graph. Forum* 31(3): 1335-1344 (2012)

J. Seo, B. Shneiderman: Interactively Exploring Hierarchical Clustering Results. *IEEE Computer* 35(7): 80-86 (2002)

Steinbach, M., G. Karypis V. Kumar. "A comparison of document clustering techniques." *Proc. of KDD workshop on text mining*. Vol. 400. No. 1. 2000.

C. Turkay, J. Parulek, N. Reuter, and H. Hauser. Interactive visual analysis of temporal cluster structures. Comput. Graph. Forum, 30(3):711–720, 2011.

T. Van Long (2009). Visualizing high-density clusters in multidimensional data, PhD thesis, School of Engineering and Science, Jacobs University, Bremen

Willett, Peter. "Recent trends in hierarchic document clustering: a critical review." *Information Processing & Management* 24.5 (1988): 577-597

D. R. Wilson and T. R. Martinez (1997). Improved Heterogeneous Distance Functions, *Journal of Artificial Intelligence Research*, Vol. 6: 1-34

H. Zhou, X. Yuan, H. Qu, W. Cui, and B. Chen. Visual clustering in parallel coordinates. *Computer Graphics Forum*, Vol. 27(3):1047–1054, 2008