Ludwig
Maximilians
Universität
München
LMU

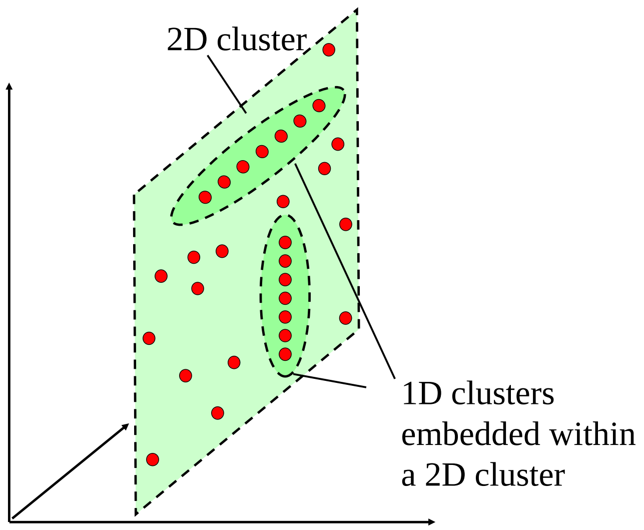
Institut für Informatik

HiSC

Finding Hierarchies of Subspace Clusters

Objective and Definitions

Hierarchically Nested Subspace Clusters



Subspace Dimensionality of a Point *p*

- Variance along attribute A_i

$$VAR_{A_i}(NN_k(p)) = \frac{\sum_{q \in NN_k(p)} (\pi_{A_i}(q) - \pi_{A_i}(p))^2}{|NN_k(p)|}$$

$\pi_S(o)$: projection of point *o* into subspace *S*

- Subspace preference vector

$$w_i(p) = \begin{cases} 0, if \; VAR_{A_i}(NN_k(p)) > \alpha \\ 1, if \; VAR_{A_i}(NN_k(p)) \le \alpha \end{cases}$$

- Subspace dimensionality

$$\lambda(p) = \sum_{i=1}^{d} \begin{cases} 1, if \; w_i(p) = 0 \\ 0, if \; w_i(p) = 1 \end{cases}$$

Subspace Preference Vector / Dimensionality of a Pair of Points (*p, q*)

- Subspace preference vector

$$w(p,q) = w_p \wedge w_q$$

- Subspace dimensionality

$$\lambda(p,q) = \sum_{i=1}^{d} \begin{cases} 1, \text{ if } \; w_i(p,q) = 0 \\ 0, \text{ if } \; w_i(p,q) = 1 \end{cases}$$

Subspace Distance between two Points *p* and *q*

$$SDIST(p,q) = (d_1, d_2), \text{ where}$$
$$d_1 = \lambda(p,q) + \begin{cases} 1, \text{ if } \; \max\{d_{w_p}(p,q), d_{w_q}(p,q)\} > \alpha \\ 0, \text{ else} \end{cases}$$
$$d_2 = d_{\overline{w}(p,q)}(p,q)$$

$$d_{w_p}(p,q) = \sum_{i=1}^{d} w_i(p)(\pi_{A_i}(p) - \pi_{A_i}(q))^2 \;:$$

weighted Euclidean distance w.r.t. *w* between two points *p* and *q*

Database Systems LMU

E. Achtert, C. Böhm, H.-P. Kriegel,
P. Kröger, I. Müller-Gorman, A. Zimek

www.dbs.ifi.lmu.de

Ludwig
Maximilians
Universität
München

HiSC

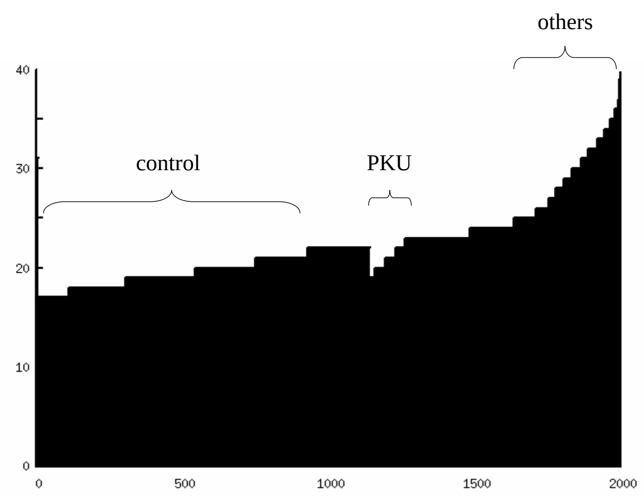
Finding Hierarchies of Subspace Clusters

Algorithm and Results

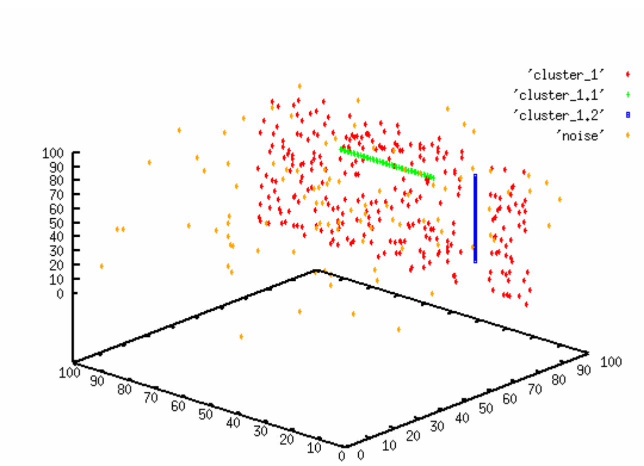
Algorithm HiSC

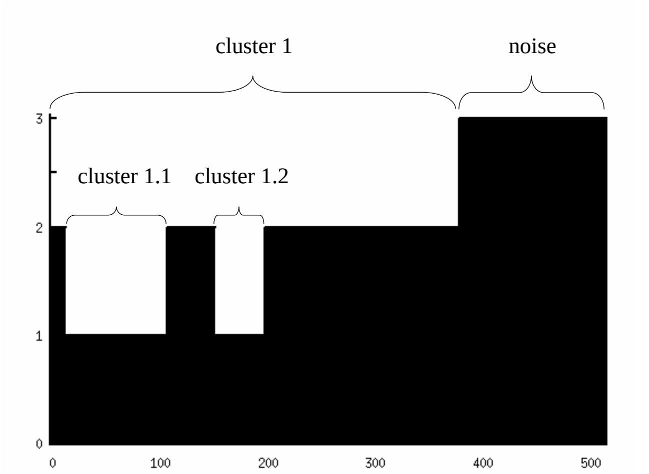
```
algorithm HiSC(DB, k, α)
  pq := empty priority queue;
  for each p ∈ DB do
    compute w_p;
    p.SDIST = ∞;
    insert p into pq;
  while pq ≠ ∅ do
    o := pq.next();
    for each p ∈ pq do
      pq.update(p, SDIST(o,p));
    append o to cluster order;
return cluster order;
```

Result on Metabolome Dataset



Result on Synthetic Dataset





Database Systems LMU

E. Achtert, C. Böhm, H.-P. Kriegel,
P. Kröger, I. Müller-Gorman, A. Zimek

www.dbs.ifi.lmu.de