LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
INSTITUTE FOR INFORMATICS
DATABASE SYSTEMS GROUP

Subspace Clustering, Ensemble Clustering, Alternative Clustering, Multiview Clustering: What Can We Learn From Each Other?

MultiClust@KDD 2010

Hans-Peter Kriegel, Arthur Zimek

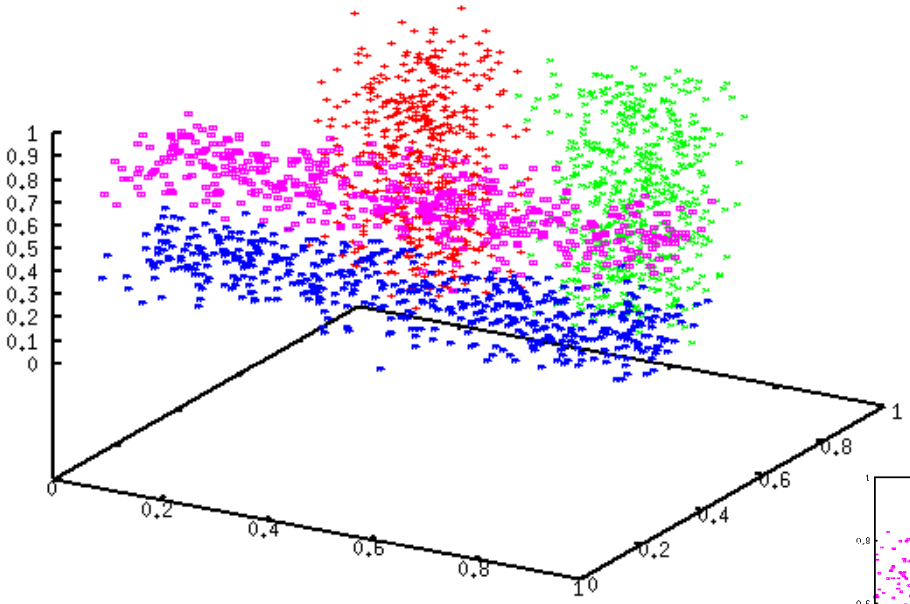
Ludwig-Maximilians-Universität München
Munich, Germany
http://www.dbs.ifi.lmu.de
{kriegel, zimek}@dbs.ifi.lmu.de

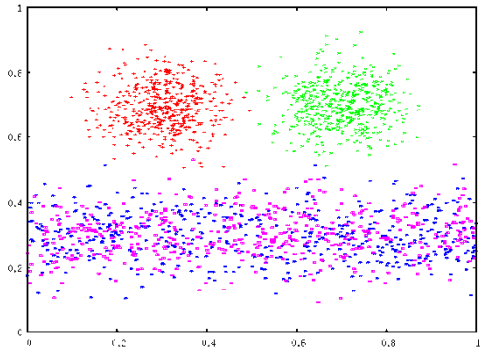
Outline

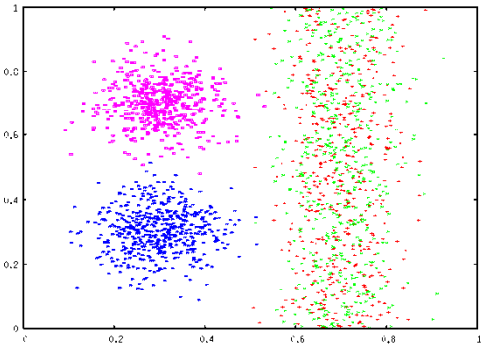
1. Subspace Clustering
2. Ensemble Clustering
3. Alternative Clustering
4. Multiview Clustering
5. Discussion

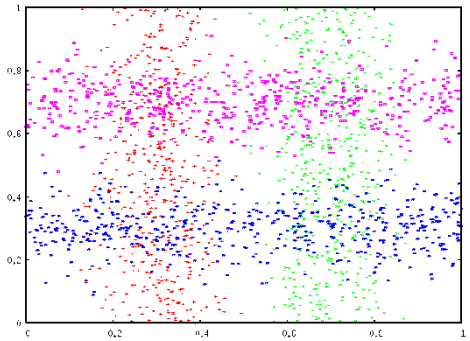
Subspace Clustering

- Task: identify clusters of similar objects
- similarity defined w.r.t. a certain subspace of the data space
- different subspaces for different clusters





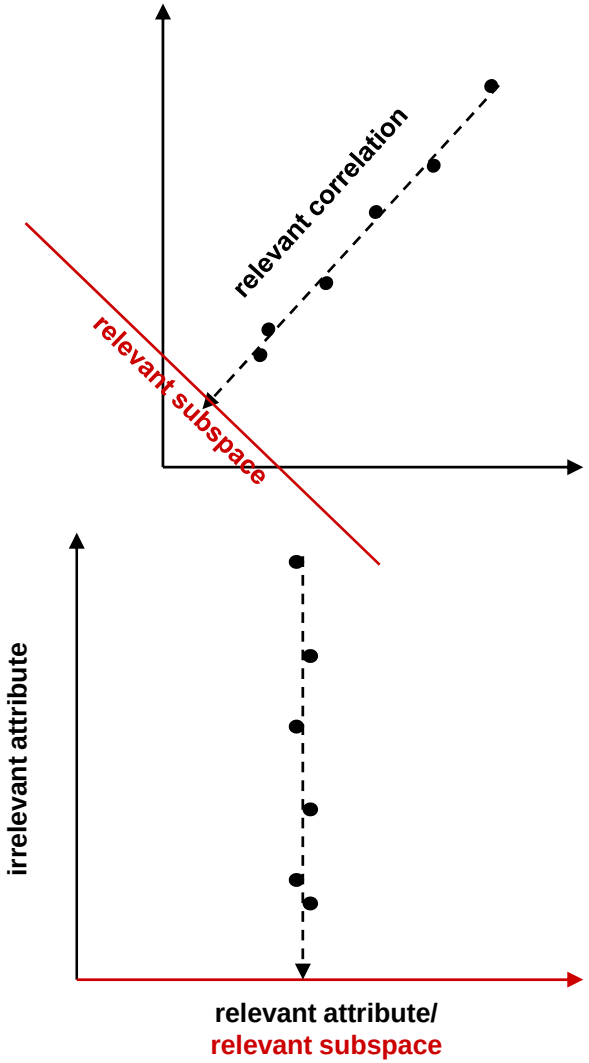




Subspace Clustering

- Subspaces: different
  - selection
  - weighting
  - combination

  of attributes
- learn subspace and clustering simultaneously (interdepency)
- strategies:
  - top-down (learn spatial characteristics of initially built sets of objects)
  - bottom-up (learn 1-d clusters, combine them to 2-d clusters, etc. (APRIORI)) => many irrelevant clusters



Ensemble Clustering

- basic idea: combine different clusterings to obtain one single, more reliable clustering
- tasks:
  - how to create *diverse* clusterings
  - how to *combine* different clusterings
- induce diversity of clusterings
  - use different feature-subsets
  - use different database subsets
  - use different clustering algorithms
- correspondence between clusterings
  - useful for judging on redundancy of clusters?
  - a lot of different answers – but: could it not be that different clusterings are just different, yet both meaningful?

Alternative Clustering

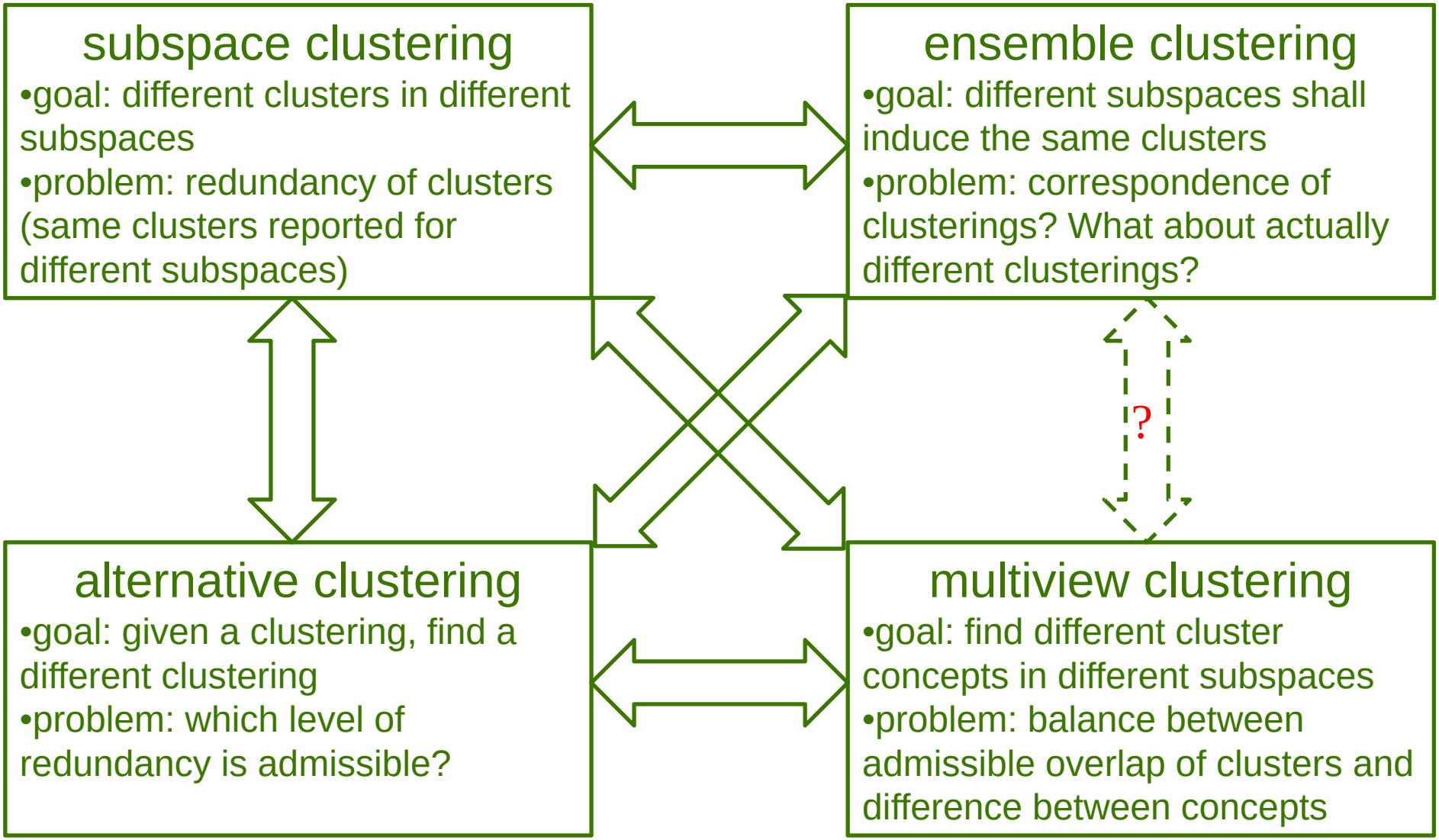
- given a clustering, use diversity or non-redundancy as a constraint to find a different clustering
- techniques:
  - ensemble techniques
  - use different subspaces
- relationship to subspace clustering:
  - subspace clustering can learn from the treatment of non-redundancy
  - alternative clustering can learn to allow for a certain level of redundancy

Multiview Clustering

- seek different clusterings in different subspaces
- special case of alternative clustering?
  - constraint: orthogonality of subspaces
- special case of subspace clustering?
  - allowing maximal overlap of clusters
  - seeking minimally redundant clusters by accommodating different concepts
- emphasizes the observation known from subspace clustering:

  highly overlapping clusters in different subspaces need not be redundant nor meaningless

Discussion


subspace clustering
•goal: different clusters in different subspaces
•problem: redundancy of clusters (same clusters reported for different subspaces)
ensemble clustering
•goal: different subspaces shall induce the same clusters
•problem: correspondence of clusterings? What about actually different clusterings?
?
alternative clustering
•goal: given a clustering, find a different clustering
•problem: which level of redundancy is admissible?
multiview clustering
•goal: find different cluster concepts in different subspaces
•problem: balance between admissible overlap of clusters and difference between concepts

Discussion

- how should we treat diversity of clustering solutions?
  - should diverse clusterings always be unified (ensemble)?
  - under which conditions is a unification of diverse clusterings meaningful?
- can we learn from diversity itself?
  - again ensemble: exceptional clustering in one subspace will be outnumbered and lost – could it not be especially interesting?
- how to treat redundancy (esp. overlap)?
  - when does a cluster qualify as redundant w.r.t. another cluster, when does it represent a different concept (despite a certain overlap)?

- how to assess similarity between clustering solutions?
  - possible overlap between clusters makes this problem really difficult
  - no simple mapping