The Blind Men and the Elephant

Arthur Zimek

Ludwig-Maximilians-Universität München

Talk at NII, Tokyo
23.03.2015

The Blind Men and the Elephant

Outline

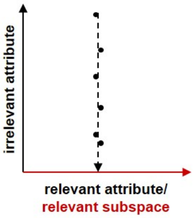
Clustering Paradigms

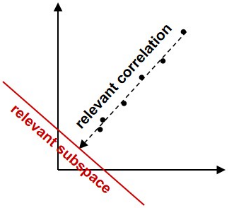
- Subspace Clustering:
  Find clusters in subspaces of the data. The subspaces where clusters reside are previously unknown.
- Ensemble Clustering:
  Derive various clustering results, unify different results to a single, supposedly more reliable result.
- Alternative Clustering:
  Given some clustering result, find a different clustering (can also be seen as a way of constraint clustering).
- Multi-View Clustering:
  Discover different clusterings in different subspaces.

Outline

Relevant and Irrelevant Attributes

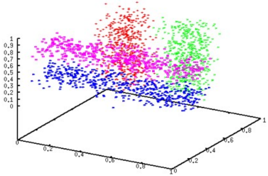
- feature relevance
  - subset of features relevant for clustering
  - cluster can be identified in this subspace only
- feature correlation
  - a subset of features can be correlated
  - relevant subspace is arbitrarily oriented



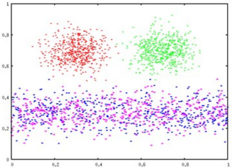


Local Feature Relevance

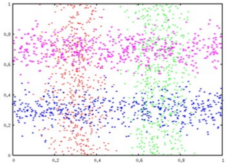
different feature subsets/correlations are relevant for different clusters/outliers



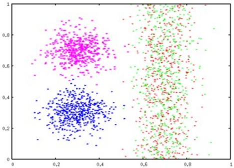
3D data set with four clusters



projection on x/z (relevant for red/green cluster)



projection on x/y



projection on y/z (relevant for blue/purple cluster)

Problem Setting

Clustering in High-Dimensional Data

Search for clusters in (in general arbitrarily oriented) subspaces of the original feature space

Challenges:

- Find the correct subspace of each cluster
  - Search space:
    - all possible arbitrarily oriented subspaces of a feature space
    - infinite
- Find the correct cluster in each relevant subspace
  - Search space:
    - "Best" partitioning of points
    - NP-complete

Circular Dependency

- circular dependency:
  - in order to determine the correct subspace of a cluster, we need to know (at least) some cluster members
  - in order to determine the correct cluster membership, we need to know the subspaces of all clusters
- solution strategy
  - integrate subspace search into clustering process
  - requires heuristics to solve
    - the clustering problem and
    - the subspace search problem

    simultaneously

Assumptions and Categories of Approaches

- one common assumption to restrict the search space: we look for clusters in axis-parallel subspaces only
- subspace search traversal can
  - start from one-dimensional spaces and combine them (bottom-up)
  - start in the full-dimensional space and deselect attributes (top-down)
- result set of clusters can
  - partition the data into disjoint clusters ("projected clustering")
  - find all clusters in all subspaces (possibly with a huge overlap/redundancy) ("subspace clustering")
- both taxonomies are not strict
- other assumptions lead to other approaches (correlation clustering, pattern-based/ co-/ bi-clustering [Kriegel et al., 2009, Madeira and Oliveira, 2004])

Summary

- Rich literature – some surveys by Kriegel et al. [2009, 2012], Sim et al. [2013], Zimek [2013], Zimek et al. [2014].
- Essential point here: many approaches are interested in finding potentially *different* clustering results in *different* subspaces.
- Problem: very similar clusters might be found over and over again in different subspaces (redundancy).

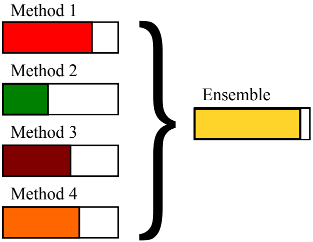
Outline

Ensembles for Classification

Assume a binary classification problem
(e.g., “does some item belong to class ‘A’ or to class ‘B’?”)

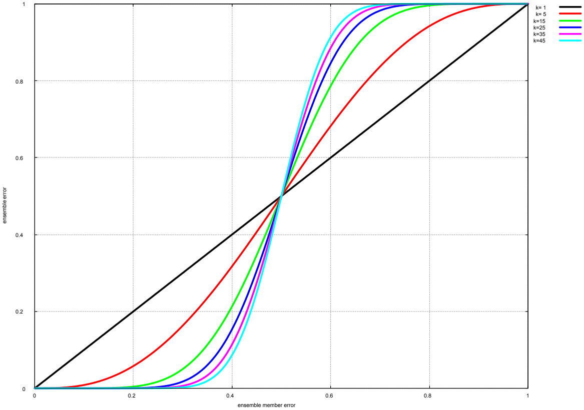
- in a “supervised learning” scenario (i.e., we have examples for ‘A’ and examples for ‘B’), we can learn a model (i.e., train a classifier)
- some classifier (model) decides with a certain accuracy
- error rate of the classifier: in how many cases (percentage) is the decision wrong?
- “ensemble”: ask several classifiers, combine their decision (e.g., majority vote)

Ensembles for Classification



- the ensemble will be much more accurate than its components, *if*
  - the components decide independently,
  - and each component decides more accurate than a coin.
- In supervised learning, the literature provides a well developed theory for ensembles.

Error-Rate of Ensembles for Classification



$$\bar{p}\,(k,p) = \sum_{i=\lceil k/2 \rceil}^{k} \binom{k}{i} p^i (1-p)^{k-i}$$

(relates to Condorcet's Jury theorem [Marquis de Condorcet, 1785])

Weak Theory in Clustering

Error-rate and similar concepts are not clearly applicable in clustering, yet the basic intuition is transferred from classification to clustering:

Motivation:

All the ensemble members are committing errors, hopefully not too many, but on different cases, *if* the members are independent, i.e., diverse.
By combining such diverse base methods into an ensemble, the different errors should level out.

Therefore the basic steps are:

- construct diverse enough (but “accurate”) clusterings for the same data set
- combine the diverse decisions into an ensemble

Weak Theory in Clustering

Error-rate and similar concepts are not clearly applicable in clustering, yet the basic intuition is transferred from classification to clustering:

Motivation:

All the ensemble members are committing errors, hopefully not too many, but on different cases, *if* the members are independent, i.e., diverse.
By combining such diverse base methods into an ensemble, the different errors should level out.

Therefore the basic steps are:

- construct diverse enough (but "accurate") clusterings for the same data set
- combine the diverse decisions into an ensemble

Weak Theory in Clustering

Error-rate and similar concepts are not clearly applicable in clustering, yet the basic intuition is transferred from classification to clustering:

Motivation:

All the ensemble members are committing errors, hopefully not too many, but on different cases, *if* the members are independent, i.e., diverse.
By combining such diverse base methods into an ensemble, the different errors should level out.

Therefore the basic steps are:

- construct diverse enough (but "accurate") clusterings for the same data set
- combine the diverse decisions into an ensemble

Weak Theory in Clustering

Error-rate and similar concepts are not clearly applicable in clustering, yet the basic intuition is transferred from classification to clustering:

Motivation:

All the ensemble members are committing errors, hopefully not too many, but on different cases, *if* the members are independent, i.e., diverse.
By combining such diverse base methods into an ensemble, the different errors should level out.

Therefore the basic steps are:

- construct diverse enough (but "accurate") clusterings for the same data set
- combine the diverse decisions into an ensemble

Theory on Combination and Diversity

- Combination procedures for cluster ensembles are relatively well explored [Topchy et al., 2004, 2005, Long et al., 2005, Fred and Jain, 2005, Caruana et al., 2006, Domeniconi and Al-Razgan, 2009, Gullo et al., 2009, Hahmann et al., 2009, Singh et al., 2010].
- There are only a few studies on the impact of diversity in cluster ensembles, such as the work by Kuncheva and Hadjitodorov [2004], Brown et al. [2005], Hadjitodorov et al. [2006], Hadjitodorov and Kuncheva [2007].

Heuristics for Diversity

- Possible heuristics to obtain diverse results (as discussed by Strehl and Ghosh [2002]):
  - non-identical sets of features
  - non-identical sets of objects
  - different clustering algorithms
- first strategy clearly related to subspace clustering
- has been pursued in many approaches, e.g., by Fern and Brodley [2003], Topchy et al. [2005], Bertoni and Valentini [2005], but:
  - typically, projections are random
  - not different solutions are sought in different subspaces, but true clusters are supposed to be more or less equally apparent in different randomized projections

Ensemble Clustering vs. Subspace Clustering

- It is probably interesting to account for the possibility of different, yet meaningful, clusters in different subspaces, so ensembles should unify only similar (?) clusters.
- Possibly, subspace clustering can benefit from advanced clustering diversity measures in ensemble clustering [Strehl and Ghosh, 2002, Fern and Brodley, 2003, Hadjitodorov et al., 2006, Gionis et al., 2007, Fern and Lin, 2008], in order to avoid redundancy.

Outline

Constraint Clustering

- classification (supervised learning): a model is learned based on complete information about the class structure of a (training) data set
- clustering (unsupervised learning): a model is fit to a data set without using prior information
- semi-supervised clustering is using *some* information, e.g., some objects may be labeled
- this partial class information is used to derive must-link constraints or cannot-link constraints [Klein et al., 2002, Bade and Nürnberger, 2008, Basu et al., 2008, Davidson and Ravi, 2009, Lelis and Sander, 2009, Zheng and Li, 2011, Campello et al., 2013, Pourrajabi et al., 2014]

Motivation by Gondek and Hofmann [2005]:

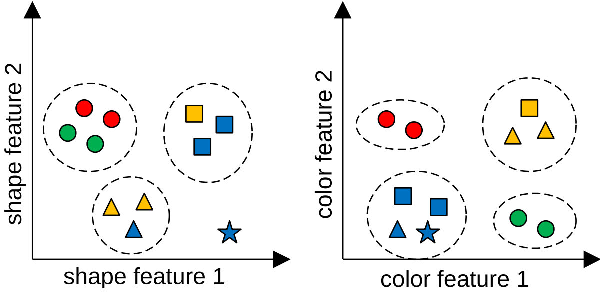
"users are often unable to *positively* describe what they are looking for, yet may be perfectly capable of expressing what is *not* of interest to them"

- constraint of non-redundancy, given some clustering
- objects clustered together in the given clustering *should not* be clustered together in a new clustering [Gondek and Hofmann, 2004, Bae and Bailey, 2006, Jain et al., 2008, Davidson et al., 2010, Niu et al., 2010, Dang and Bailey, 2010, Dang et al., 2012]
- a common heuristic to get a different clustering: use different subspaces [Qi and Davidson, 2009, Günnemann et al., 2009] – allowing *some* redundancy?
- survey: **?**

Outline

Semantically Different Subspaces

- some methods treat different subsets of attributes or different data representations separately based on a different semantic of these subsets
  - example: color features vs. texture features



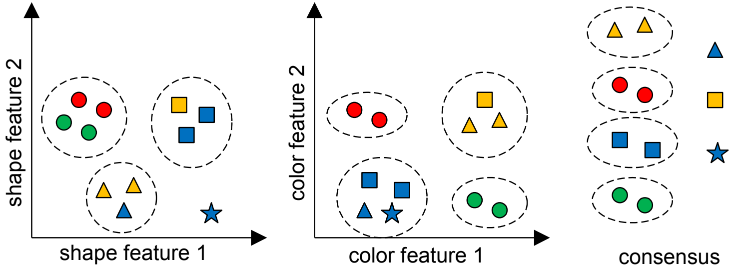
  - distances in the combined color-texture space are meaningless

Different Goals of Multi-View Clustering

- find the *same* concepts realized in different feature subsets (the 'ensemble idea' – related to co-learning/co-training) [Blum and Mitchell, 1998, Bickel and Scheffer, 2004, Sridharan and Kakade, 2008, Chaudhuri et al., 2009, Kumar and Daumé, 2011, Kriegel and Schubert, 2012]
- find *different* concepts realized in different feature subsets (the 'subspace clustering' idea – but subspaces are semantically meaningful) [Cui et al., 2007, Jain et al., 2008]
  - special case of alternative clustering (constraint: orthogonality of the subspaces) [Dang et al., 2012]
  - or special case of subspace clustering allowing maximal overlap yet seeking minimally redundant clusters by accommodating different concepts [Günnemann et al., 2009]

Outline

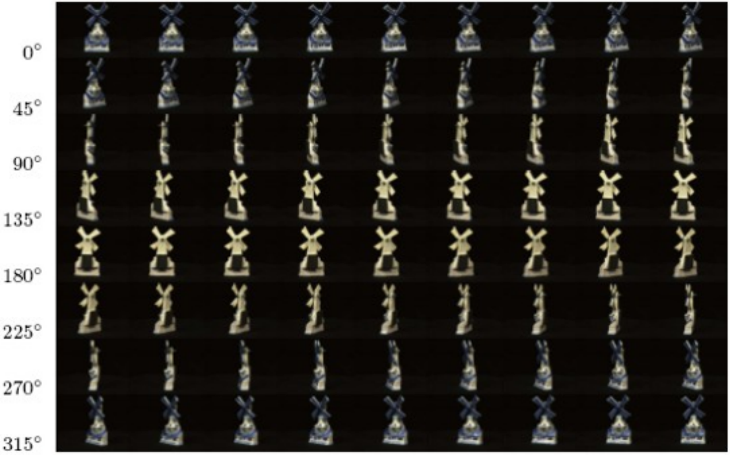
Different Meaningful Clusterings?



Different Meaningful Clusterings?

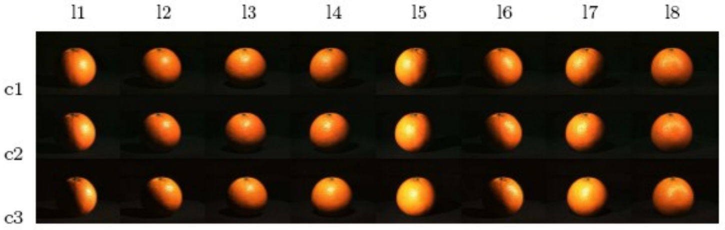
Example: ALOI [Geusebroek et al., 2005]

Example: ALOI [Geusebroek et al., 2005]



Different Meaningful Clusterings?

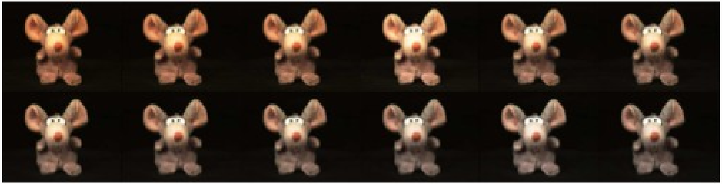
Example: ALOI [Geusebroek et al., 2005]



Different Meaningful Clusterings?

Example: ALOI [Geusebroek et al., 2005]

2175 K	2250 K	2325 K	2400 K	2475 K	2550 K
110	120	130	140	150	160



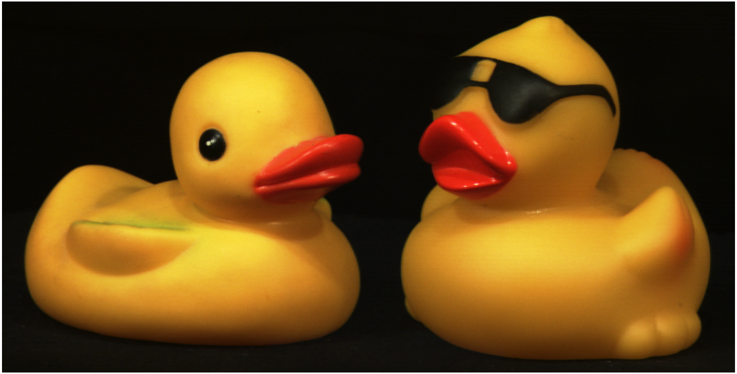
170	180	190	210	230	250
2625 K	2675 K	2750 K	2850 K	2975 K	3075 K

Different Meaningful Clusterings?

Example: ALOI [Geusebroek et al., 2005]

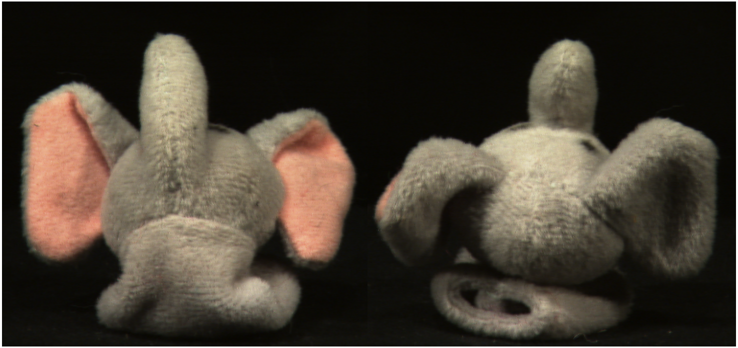
Different Meaningful Clusterings?

Example: ALOI [Geusebroek et al., 2005]



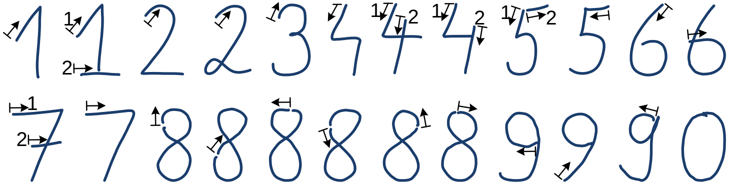
Different Meaningful Clusterings?

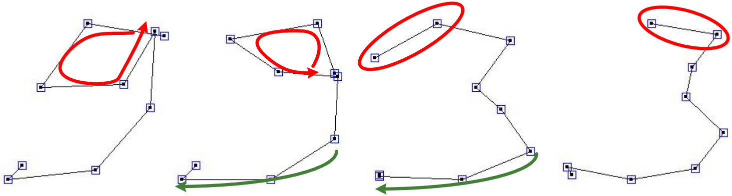
Example: ALOI [Geusebroek et al., 2005]



Different Meaningful Clusterings?

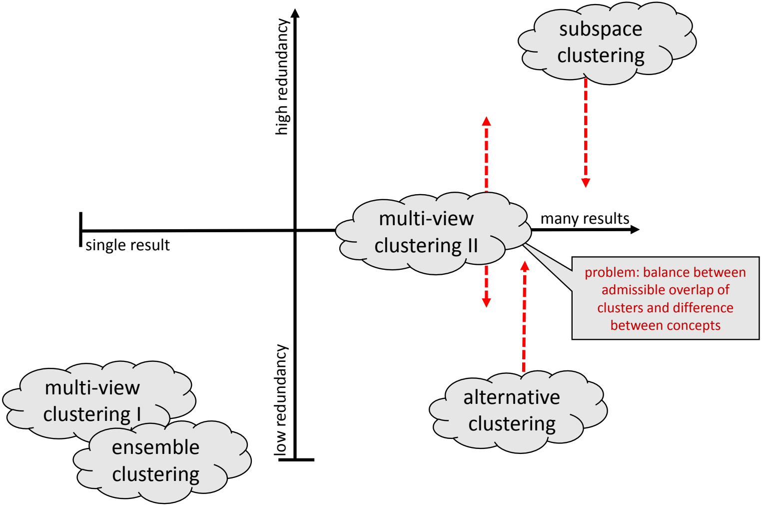
Example: Pendigits [Bache and Lichman, 2013]





observation by Färber et al. [2010]

Clustering Paradigms



Outline

Clustering Paradigms

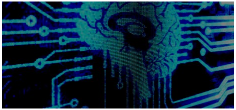
Clustering Uncertain Data

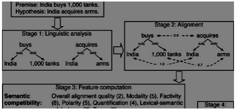
Conclusions

Uncertain Data

uncertain data can occur in very different scenarios, such as

- sensor readings
- recognition and parsing
- predictions and extrapolations
- machine learning tasks
- etc. . . .





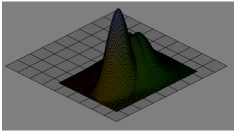
Example: Geo-Spatial Data

geo-spatial data may be uncertain due to

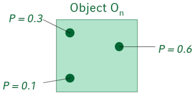
- erroneous/inexact GPS readings
- triangulation errors
- human error
- etc. ...

uncertain spatial data may be represented

- continuously

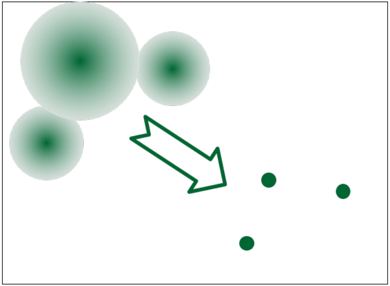


- discretely



Dealing with Uncertainty

approach 1: clean
(i.e., remove uncertainty)



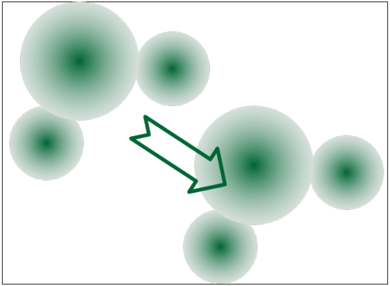
pro:

- can use traditional DBMS (and clustering. . .)

con:

- cleaning non-trivial
- can results be trusted?

approach 2: manage
(i.e., keep uncertainty)



pro:

- preserves information
- can provide confidence

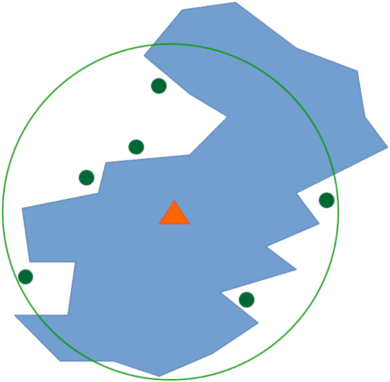
con:

- specialized DBMS (and clustering. . .)

Possible Worlds Semantics

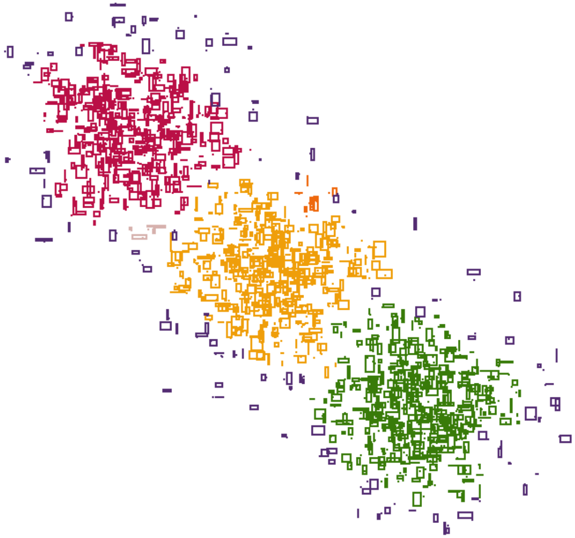
Cleaning non-trivial:
e.g., constraints on data may not be fulfilled when using aggregates.

- GPS readings around a lake
- the mean of all readings is not a valid position

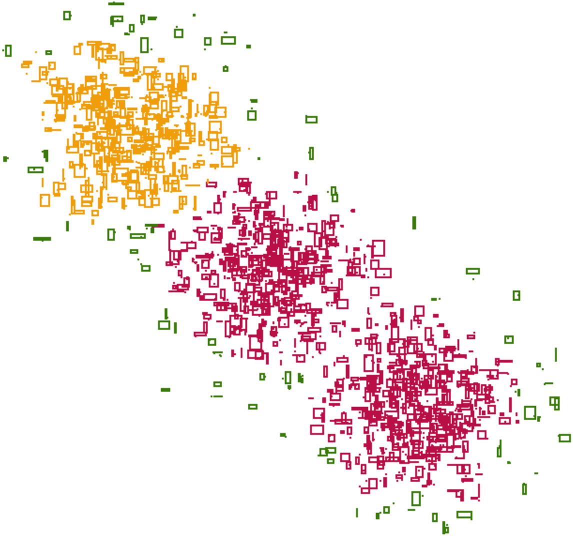


solution: sampling possible worlds

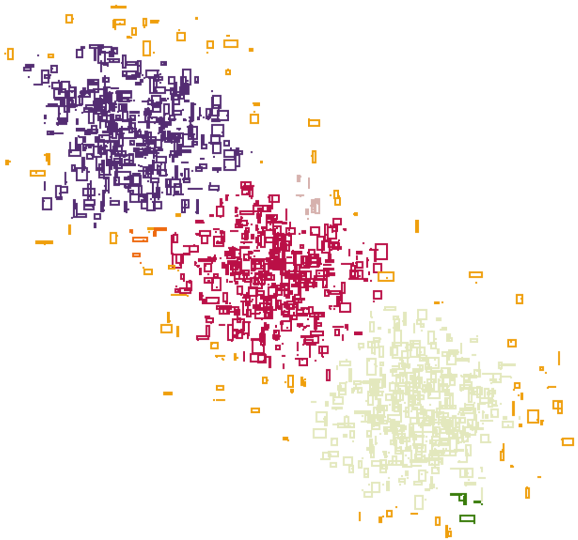
Sampling of Possible Worlds



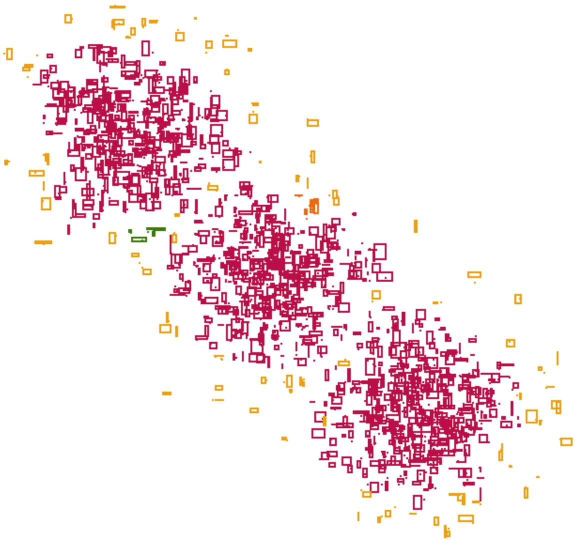
Sampling of Possible Worlds



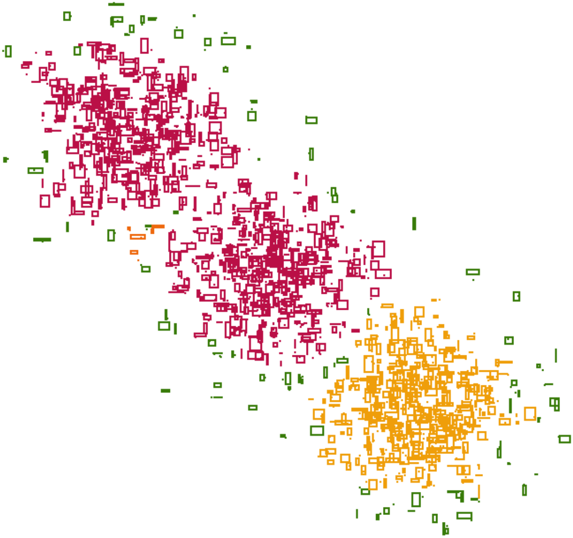
Sampling of Possible Worlds



Sampling of Possible Worlds

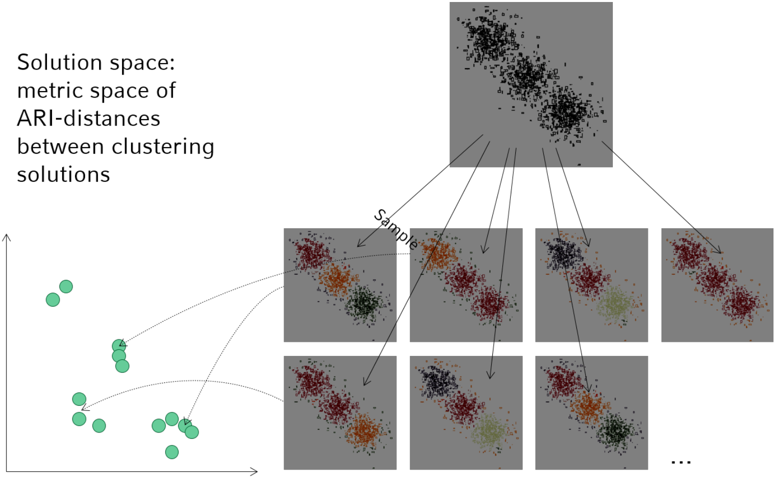


Sampling of Possible Worlds



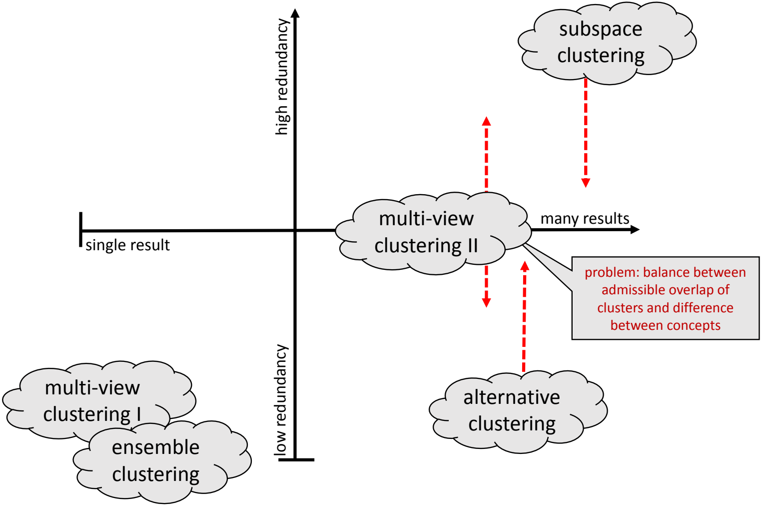
Representative Solutions

Solution space: metric space of ARI-distances between clustering solutions



Outline

Conclusions



Sources

Zimek and Vreeken: *The blind men and the elephant: On meeting the problem of multiple truths in data from clustering and pattern mining perspectives. Machine Learning, 98(1–2):121–155, 2015.*

http://doi.org/10.1007/s10994-013-5334-y

Färber, Günnemann, Kriegel, Kröger, Müller, Schubert, Seidl, and Zimek: *On using class-labels in evaluation of clusterings. In MultiClust: 1st International Workshop on Discovering, Summarizing and Using Multiple Clusterings Held in Conjunction with KDD 2010, Washington, DC, 2010.*

Züfle, Emrich, Schmid, Mamoulis, Zimek, and Renz: *Representative clustering of uncertain data. In Proceedings of the 20th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), New York, NY, pages 243–252, 2014.*

http://doi.org/10.1145/2623330.2623725

Thank you for your attention!

References I

K. Bache and M. Lichman. UCI machine learning repository, 2013. URL http://archive.ics.uci.edu/ml.

K. Bade and A. Nürnberger. Creating a cluster hierarchy under constraints of a partially known hierarchy. In *Proceedings of the 8th SIAM International Conference on Data Mining (SDM), Atlanta, GA*, pages 13–23, 2008.

E. Bae and J. Bailey. COALA: a novel approach for the extraction of an alternate clustering of high quality and high dissimilarity. In *Proceedings of the 6th IEEE International Conference on Data Mining (ICDM), Hong Kong, China*, pages 53–62, 2006. doi: 10.1109/ICDM.2006.37.

S. Basu, I. Davidson, and K. Wagstaff, editors. *Constraint Clustering: Advances in Algorithms, Applications and Theory*. CRC Press, Boca Raton, London, New York, 2008.

A. Bertoni and G. Valentini. Ensembles based on random projections to improve the accuracy of clustering algorithms. In *16th Italian Workshop on Neural Nets (WIRN), and International Workshop on Natural and Artificial Immune Systems (NAIS), Vietri sul Mare, Italy*, pages 31–37, 2005. doi: 10.1007/11731177_5.

S. Bickel and T. Scheffer. Multi-view clustering. In *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM), Brighton, UK*, pages 19–26, 2004. doi: 10.1109/ICDM.2004.10095.

LMU

References II

A. Blum and T. Mitchell. Combining labeled and unlabeled data with Co-training. In *Proceedings of the 11th Annual Conference on Computational Learning Theory (COLT), Madison, WI*, pages 92–100, 1998. doi: 10.1145/279943.279962.

G. Brown, J. Wyatt, R. Harris, and X. Yao. Diversity creation methods: a survey and categorisation. *Information Fusion*, 6:5–20, 2005. doi: 10.1016/j.inffus.2004.04.004.

R. J. G. B. Campello, D. Moulavi, A. Zimek, and J. Sander. A framework for semi-supervised and unsupervised optimal extraction of clusters from hierarchies. *Data Mining and Knowledge Discovery*, 27(3):344–371, 2013. doi: 10.1007/s10618-013-0311-4.

R. Caruana, M. Elhawary, N. Nguyen, and C. Smith. Meta clustering. In *Proceedings of the 6th IEEE International Conference on Data Mining (ICDM), Hong Kong, China*, pages 107–118, 2006. doi: 10.1109/ICDM.2006.103.

K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan. Multi-view clustering via canonical correlation analysis. In *Proceedings of the 26th International Conference on Machine Learning (ICML), Montreal, QC, Canada*, pages 129–136, 2009. doi: 10.1145/1553374.1553391.

References III

Y. Cui, X. Z. Fern, and J. G. Dy. Non-redundant multi-view clustering via orthogonalization. In *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM), Omaha, NE*, pages 133–142, 2007. doi: 10.1109/ICDM.2007.94.

X. H. Dang and J. Bailey. Generation of alternative clusterings using the CAMI approach. In *Proceedings of the 10th SIAM International Conference on Data Mining (SDM), Columbus, OH*, pages 118–129, 2010.

X. H. Dang, I. Assent, and J. Bailey. Multiple clustering views via constrained projections. In *3rd MultiClust Workshop: Discovering, Summarizing and Using Multiple Clusterings Held in Conjunction with SIAM Data Mining 2012, Anaheim, CA*, 2012.

I. Davidson and S. Ravi. Using instance-level constraints in agglomerative hierarchical clustering: theoretical and empirical results. *Data Mining and Knowledge Discovery*, 18:257–282, 2009. doi: 10.1007/s10618-008-0103-4.

I. Davidson, S. S. Ravi, and L. Shamis. A SAT-based framework for efficient constrained clustering. In *Proceedings of the 10th SIAM International Conference on Data Mining (SDM), Columbus, OH*, pages 94–105, 2010.

C. Domeniconi and M. Al-Razgan. Weighted cluster ensembles: Methods and analysis. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 2(4): 17:1–40, 2009. doi: 10.1145/1460797.1460800.

References IV

X. Z. Fern and C. E. Brodley. Random projection for high dimensional data clustering: A cluster ensemble approach. In *Proceedings of the 20th International Conference on Machine Learning (ICML), Washington, DC*, pages 186–193, 2003.

X. Z. Fern and W. Lin. Cluster ensemble selection. *Statistical Analysis and Data Mining*, 1(3):128–141, 2008. doi: 10.1002/sam.10008.

A. L. N. Fred and A. K. Jain. Combining multiple clusterings using evidence accumulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6):835–850, 2005. doi: 10.1109/TPAMI.2005.113.

I. Färber, S. Günnemann, H.-P. Kriegel, P. Kröger, E. Müller, E. Schubert, T. Seidl, and A. Zimek. On using class-labels in evaluation of clusterings. In *MultiClust: 1st International Workshop on Discovering, Summarizing and Using Multiple Clusterings Held in Conjunction with KDD 2010, Washington, DC*, 2010.

J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders. The Amsterdam Library of Object Images. *International Journal of Computer Vision*, 61(1): 103–112, 2005. doi: 10.1023/B:VISI.0000042993.50813.60.

A. Gionis, H. Mannila, and P. Tsaparas. Clustering aggregation. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), 2007. doi: 10.1145/1217299.1217303.

LMU

References V

The Blind Men and the Elephant

Arthur Zimek

D. Gondek and T. Hofmann. Non-redundant data clustering. In *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM), Brighton, UK*, pages 75–82, 2004. doi: 10.1109/ICDM.2004.10104.

D. Gondek and T. Hofmann. Non-redundant clustering with conditional ensembles. In *Proceedings of the 11th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Chicago, IL*, pages 70–77, 2005. doi: 10.1145/1081870.1081882.

F. Gullo, A. Tagarelli, and S. Greco. Diversity-based weighting schemes for clustering ensembles. In *Proceedings of the 9th SIAM International Conference on Data Mining (SDM), Sparks, NV*, pages 437–448, 2009.

S. Günnemann, E. Müller, I. Färber, and T. Seidl. Detection of orthogonal concepts in subspaces of high dimensional data. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM), Hong Kong, China*, pages 1317–1326, 2009. doi: 10.1145/1645953.1646120.

S. T. Hadjitodorov and L. I. Kuncheva. Selecting diversifying heuristics for cluster ensembles. In *7th International Workshop on Multiple Classifier Systems (MCS), Prague, Czech Republic*, pages 200–209, 2007. doi: 10.1007/978-3-540-72523-7_21.

References VI

S. T. Hadjitodorov, L. I. Kuncheva, and L. P. Todorova. Moderate diversity for better cluster ensembles. *Information Fusion*, 7(3):264–275, 2006. doi: 10.1016/j.inffus.2005.01.008.

M. Hahmann, P. B. Volk, F. Rosenthal, D. Habich, and W. Lehner. How to control clustering results? Flexible clustering aggregation. In *Proceedings of the 8th International Symposium on Intelligent Data Analysis (IDA), Lyon, France*, pages 59–70, 2009. doi: 10.1007/978-3-642-03915-7_6.

P. Jain, R. Meka, and I. S. Dhillon. Simultaneous unsupervised learning of disparate clusterings. *Statistical Analysis and Data Mining*, 1(3):195–210, 2008. doi: 10.1002/sam.10007.

D. Klein, S. D. Kamvar, and C. D. Manning. From instance-level constraints to space-level constraints: Making the most of prior knowledge in data clustering. In *Proceedings of the 19th International Conference on Machine Learning (ICML), Sydney, Australia*, pages 307–314, 2002.

H.-P. Kriegel and M. Schubert. Co-RCA: Unsupervised distance-learning for multi-view clustering. In *3rd MultiClust Workshop: Discovering, Summarizing and Using Multiple Clusterings Held in Conjunction with SIAM Data Mining 2012, Anaheim, CA*, pages 11–18, 2012.

References VII

H.-P. Kriegel, P. Kröger, and A. Zimek. Clustering high dimensional data: A survey on subspace clustering, pattern-based clustering, and correlation clustering. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 3(1):1–58, 2009. doi: 10.1145/1497577.1497578.

H.-P. Kriegel, P. Kröger, and A. Zimek. Subspace clustering. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(4): 351–364, 2012. doi: 10.1002/widm.1057.

A. Kumar and H. Daumé. A co-training approach for multi-view spectral clustering. In *Proceedings of the 28th International Conference on Machine Learning (ICML), Bellevue, Washington, DC, USA*, pages 393–400, 2011.

L. I. Kuncheva and S. T. Hadjitodorov. Using diversity in cluster ensembles. In *Proceedings of the 2004 IEEE International Conference on Systems, Man, and Cybernetics (ICSMC), The Hague, Netherlands*, pages 1214–1219, 2004. doi: 10.1109/ICSMC.2004.1399790.

L. Lelis and J. Sander. Semi-supervised density-based clustering. In *Proceedings of the 9th IEEE International Conference on Data Mining (ICDM), Miami, FL*, pages 842–847, 2009. doi: 10.1109/ICDM.2009.143.

References VIII

B. Long, Z. Zhang, and P. S. Yu. Combining multiple clustering by soft correspondence. In *Proceedings of the 5th IEEE International Conference on Data Mining (ICDM), Houston, TX*, pages 282–289, 2005. doi: 10.1109/ICDM.2005.45.

S. C. Madeira and A. L. Oliveira. Biclustering algorithms for biological data analysis: A survey. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 1(1):24–45, 2004. doi: 10.1109/TCBB.2004.2.

M. J. A. N. C. Marquis de Condorcet. *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. L'Imprimerie Royale, Paris, 1785.

D. Niu, J. G. Dy, and M. I. Jordan. Multiple non-redundant spectral clustering views. In *Proceedings of the 27th International Conference on Machine Learning (ICML), Haifa, Israel*, pages 831–838, 2010.

M. Pourrajabi, D. Moulavi, R. J. G. B. Campello, A. Zimek, J. Sander, and R. Goebel. Model selection for semi-supervised clustering. In *Proceedings of the 17th International Conference on Extending Database Technology (EDBT), Athens, Greece*, pages 331–342, 2014. doi: 10.5441/002/edbt.2014.31.

References IX

Z. J. Qi and I. Davidson. A principled and flexible framework for finding alternative clusterings. In *Proceedings of the 15th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Paris, France*, pages 717–726, 2009. doi: 10.1145/1557019.1557099.

K. Sim, V. Gopalkrishnan, A. Zimek, and G. Cong. A survey on enhanced subspace clustering. *Data Mining and Knowledge Discovery*, 26(2):332–397, 2013. doi: 10.1007/s10618-012-0258-x.

V. Singh, L. Mukherjee, J. Peng, and J. Xu. Ensemble clustering using semidefinite programming with applications. *Machine Learning*, 79(1-2):177–200, 2010.

K. Sridharan and S. M. Kakade. An information theoretic framework for multi-view learning. In *Proceedings of the 21st Annual Conference on Learning Theory (COLT), Helsinki, Finland*, pages 403–414, 2008.

A. Strehl and J. Ghosh. Cluster ensembles – a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3: 583–617, 2002.

A. Topchy, A. Jain, and W. Punch. Clustering ensembles: Models of concensus and weak partitions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1866–1881, 2005. doi: 10.1109/TPAMI.2005.237.

References X

A. P. Topchy, M. H. C. Law, A. K. Jain, and A. L. Fred. Analysis of consensus partition in cluster ensemble. In *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM), Brighton, UK*, pages 225–232, 2004. doi: 10.1109/ICDM.2004.10100.

L. Zheng and T. Li. Semi-supervised hierarchical clustering. In *Proceedings of the 11th IEEE International Conference on Data Mining (ICDM), Vancouver, BC*, pages 982–991, 2011. doi: 10.1109/ICDM.2011.130.

A. Zimek. Clustering high-dimensional data. In C. C. Aggarwal and C. K. Reddy, editors, *Data Clustering: Algorithms and Applications*, chapter 9, pages 201–230. CRC Press, 2013.

A. Zimek and J. Vreeken. The blind men and the elephant: On meeting the problem of multiple truths in data from clustering and pattern mining perspectives. *Machine Learning*, 98(1–2):121–155, 2015. doi: 10.1007/s10994-013-5334-y.

A. Zimek, I. Assent, and J. Vreeken. Frequent pattern mining algorithms for data clustering. In C. C. Aggarwal and J. Han, editors, *Frequent Pattern Mining*, chapter 16, pages 403–423. Springer, 2014. doi: 10.1007/978-3-319-07821-2_16.

References XI

A. Züfle, T. Emrich, K. A. Schmid, N. Mamoulis, A. Zimek, and M. Renz. Representative clustering of uncertain data. In *Proceedings of the 20th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), New York, NY*, pages 243–252, 2014. doi: 10.1145/2623330.2623725.