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

Hans-Peter Kriegel Arthur Zimek
Ludwig-Maximilians-Universität München
Oettingenstrasse 67, 80538 München, Germany
http://www.dbs.ifi.lmu.de
{kriegel,zimek}@dbs.ifi.lmu.de

ABSTRACT
Though subspace clustering, ensemble clustering, alternative clustering, and multiview clustering are different approaches motivated by different problems and aiming at different goals, there are similar problems in these fields. Here we shortly survey these areas from the point of view of subspace clustering. Based on this survey, we try to identify problems where the different research areas could probably learn from each other.

1. INTRODUCTION
Clustering is the problem of finding a set of groups of similar objects within a data set while keeping dissimilar objects separated in different groups or the group of noise. Two major issues can be distinguished in the clustering problem in general, namely (i) the adopted paradigm and algorithmic approach to clustering and (ii) the definition and assessment of similarity vs. dissimilarity. Although clustering in general is a rather dignified data mining task, different specialized techniques to enhance clustering solutions have been recently brought forward in the literature. Here we shortly survey the techniques named ensemble clustering, alternative clustering, multiview clustering, and subspace clustering. Each of these techniques has been developed with different motivations yet there are striking similarities. However, there are also — plain as well as subtle — differences. In this contribution, we want to shed only some light on these differences and similarities by highlighting the problems occurring in several of these areas. It is our hope that we can trigger discussion among researchers active in these different areas. Although these areas aim at different goals, researchers may learn from each other how to treat similar problems.

As we are active in research on subspace clustering [32] yet only observing so far research on ensemble clustering, alternative clustering, or multiview clustering, we try to survey these fields of research mainly from the point of view of subspace clustering. In the following, we first give an overview of the problems and basic solutions of subspace clustering (Section 2). Afterwards we discuss ensemble clustering approaches (Section 3) and touch alternative clustering (Section 4) and multiview clustering (Section 5). Finally (Section 6), we collect questions in these areas where answers from other areas may be helpful in improving the field. Let us emphasize, however, that we do not aim at solving all the open questions or at unifying different research areas but just hope to inspire discussion between different research areas.

Proc. 1st International Workshop on Discovering, Summarizing and Using Multiple Clusterings (MultiClust 2010) in conjunction with 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2010), Washington, DC

2. SUBSPACE CLUSTERING
Subspace clustering refers to the task of identifying clusters of similar vectors where the similarity is defined w.r.t. a subspace of the data space. The subspace is not necessarily (and actually is usually not) the same for different clusters within one clustering solution. The key-issue in subspace clustering is the definition of similarity taking into account only certain subspaces. Different weighting, different selections, or different combinations of attributes of a data set are equivalent to defining or deriving different subspaces where the desired properties of a model of similarity suitable to a given application domain may be expressed appropriately. Which subspaces are important for the similarity measure is to be learned during the clustering process, since for different clusters within one and the same clustering solution usually different subspaces are relevant. Hence subspace clustering algorithms cannot be thought of as usual clustering algorithms using just a different definition of similarity but the similarity measure and the clustering solution are derived simultaneously and depend on each other.

In [33], a short theoretical overview is provided on this topic. In [32], the problem is discussed in depth, also differentiating several subproblems and surveying a number of example algorithms. Although there is a wide variety of task definitions for clustering in subspaces, the term "subspace clustering" in a narrower sense does also relate to a special category of clustering algorithms in axis-parallel subspaces. Another family of clustering in axis-parallel subspaces is called "projected clustering". Recent experimental evaluation studies covered some selections of these specific subtypes of subspace clustering [38, 40] and these subtypes may be the most interesting ones for considering relationships with the other research areas discussed in this survey.

Clustering in axis-parallel subspaces is based on the distinction between relevant and irrelevant attributes. This distinction generally assumes that the variance of attribute values for a relevant attribute over all points of the corresponding cluster is rather small compared to the overall range of attribute values whereas the variance for irrelevant attributes within a given cluster is high or indistinguishable from the values of the same attribute for other clusters. For example, one could assume a relevant attribute for a given cluster being normally distributed with a small standard deviation whereas irrelevant attribute values are uniformly distributed over the complete data space. The geometrical intuition of these assumptions relates to the points of a cluster being widely scattered in the direction of irrelevant axes while being densely clustered along relevant attributes. When selecting the relevant attributes only, the cluster would appear as a full-dimensional cluster, while in the full dimensional space (including also the irrelevant attributes) the cluster points form a hyperplane parallel to the irrelevant axes. Due to this geometrical appearance, this type of cluster is addressed as "axis-parallel subspace cluster".

Since the number of possible axis-parallel subspaces where clusters could reside is exponential in the dimensionality of the data space, the main task of research in the field was the development of appropriate subspace search heuristics. Starting from the pioneering approaches to axis-parallel subspace clustering, there have been pursued two opposite basic techniques for searching subspaces, namely (a) *top-down search* [1] and (b) *bottom-up search* [2].

(a) The rational of top-down approaches is to determine the subspace of a cluster starting from the full dimensional space. This is usually done by determining a subset of attributes for a given set of points — potential cluster members — such that the points meet the given cluster criterion when projected onto the corresponding subspace. Obviously the dilemma is that for the determination of the subspace of a cluster at least some cluster members must be identified. On the other hand, in order to determine cluster memberships, the subspace of each cluster must be known. To escape from this circular dependency, most of the top-down approaches rely on a rather strict assumption, which has been called the *locality assumption* [32]. It is assumed that the subspace of a cluster can be derived from the local neighborhood (in the full dimensional data space) of the cluster center or the cluster members. In other words, it is assumed that even in the full dimensional space, the subspace of each cluster can be learned from the local neighborhood of cluster representatives or cluster members. Other top-down approaches that do not rely on the locality assumption use random sampling as a heuristic in order to generate a set of potential cluster members.

(b) For a bottom-up search, the exponential search space of all possible subspaces of a data space that needs to be traversed is seen as being equivalent to the search space of the frequent item set problem in analysis of market baskets in transaction databases [3]. Each attribute represents an item and each subspace cluster is a transaction of the items representing the attributes that span the corresponding subspace. Finding item sets with frequency 1 then relates to finding all combinations of attributes that constitute a subspace containing at least one cluster. This observation is the rational of most bottom-up subspace clustering approaches. The subspaces that contain clusters are determined starting from all 1-dimensional subspaces that accommodate at least one cluster by employing a search strategy similar to frequent itemset mining algorithms. To apply any efficient frequent itemset mining algorithm, the cluster criterion must implement a downward closure property (also called monotonicity property): *If subspace S contains a cluster, then any subspace $T \subseteq S$ must also contain a cluster.* The anti-monotonic reverse implication, *if a subspace T does not contain a cluster, then any superspace $S \supseteq T$ also cannot contain a cluster*, can be used for pruning, i.e. excluding specific subspaces from consideration. Let us note that there are bottom-up algorithms that do not use an APRIORI-like subspace search but instead apply other search heuristics.

Subspace clustering in a narrower sense pursues the goal to find all clusters in all subspaces of the entire feature space. This goal obviously is defined to correspond to the bottom-up technique used by these approaches, based on some anti-monotonic property of clusters allowing the application of the APRIORI search heuristic. The pioneer approach for finding all clusters in all subspaces coining the term "subspace clustering" for this specific task has been CLIQUE [2]. Variants include [11, 41, 31, 4, 37, 5, 39, 35]. Since the initial problem formulation of finding "all clusters in all subspaces" is rather questionable since the information gained by retrieving such a huge set of clusters with high redundancy is not very useful, subsequent methods often concentrated on possibilities of concisely restricting the result set of clusters by somehow assessing and reducing the redundancy of clusters, for example to keep only clusters of highest dimensionality. It also should be noted that the statistical significance of subspace clusters (as defined in [37]), is not an anti-monotonic property and hence does in general not allow for APRIORI-like bottom-up approaches finding only *meaningful* clusters.

3. ENSEMBLE CLUSTERING

In the area of supervised learning, combining several self-contained predicting algorithms to an ensemble to yield a better performance in terms of accuracy than any of the base predictors, is backed by a sound theoretical background[16, 17, 47]. In short, a predictive algorithm can suffer from several limitations such as statistical variance, computational variance, and a strong bias. *Statistical variance* describes the phenomenon that different prediction models result in equally good performance on training data. Choosing arbitrarily one of the models can then result in deteriorated performance on new data. Voting among equally good classifiers can reduce this risk. *Computational variance* refers to the fact, that computing the truly optimal model is usually intractable and hence any classifier tries to overcome computational restrictions by some heuristics. These heuristics, in turn, can lead to local optima in the training phase. Obviously, trying several times reduces the risk of choosing the wrong local optimum. A restriction of the space of hypotheses a predictive algorithm may create is referred to as *bias* of the algorithm. Usually, the bias allows for learning an abstraction and is, thus, a necessary condition of learning a hypothesis instead of learning by heart the examples of the training data (the latter resulting in random performance on new data). However, a strong bias may also hinder the representation of a good model of the true laws of nature one would like to learn. A weighted sum of hypotheses may then expand the space of possible models.

To improve over several self-contained classifiers by building an ensemble of those classifiers requires the base algorithms being accurate (i.e., at least better than random) and diverse (i.e., making different errors on new instances). It is easy to understand why these two conditions are necessary and also sufficient. If several individual classifiers are not diverse, then all of them will be wrong whenever one of them is wrong. Thus nothing is gained by voting over wrong predictions. On the other hand, if the errors made by the classifiers were uncorrelated, more individual classifiers may be correct while some individual classifiers are wrong. Therefore, a majority vote by an ensemble of these classifiers may be also correct. More formally, suppose an ensemble consisting of k hypotheses, and the error rate of each hypothesis is equal to a certain $p < 0.5$ (assuming a dichotomous problem), though errors occur independently in different hypotheses. The ensemble will be wrong, if more than $k/2$ of the ensemble members are wrong. Thus the overall error rate $\bar{p}$ of the ensemble is given by the area under the binomial distribution, where $k \geq \lceil k/2 \rceil$, that is for at least $\lceil k/2 \rceil$ hypotheses being wrong:

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

The overall error-rate is rapidly decreasing for an increasing number of ensemble members.

In the unsupervised task of clustering, the theory for building ensembles is less clear yet. Improvement by application of ensemble techniques have been demonstrated empirically, though. Approaches have concentrated on creating diverse base clusterings and then combining them somehow to a unified single clustering. The approaches differ in (a) how to create diversity and (b) how to combine different clusterings.

(a) As sources of diversity, [44] discuss (i) non-identical sets of features, (ii) non-identical sets of objects, and (iii) different clustering algorithms. Obviously, the first of these strategies is somehow related to subspace clustering and has been pursued in different ensemble clustering approaches [20, 45, 8]. Usually, however, the projections used here are random projections and not different clusters are sought in different subspaces but true clusters are supposed to be more or less equally apparent in different random projections. It is probably interesting to account for the possibility of different yet meaningful clusters in different subspaces. For example, the authors of [20] are aware of possibly different clusters existing in different subspaces. Nevertheless, their approach aims at a single unified clustering solution, based on the ensemble framework of [44].

(b) How to derive the correspondence between different clustering solutions in order to combine them is the other dominant question in research on clustering ensembles, see e.g. [46, 45, 36, 10, 18, 26, 29, 43]. The correspondence between different clusterings is a problem not encountered in classification ensembles. From the point of view of subspace clustering, the correspondence problem in ensemble clustering is also an interesting topic as there are no suitable automatic evaluation procedures for the possibly highly complex and overlapping clusters obtained in different subspaces.

It is our impression that these two topics in research on ensemble clustering directly relate to certain questions discussed in subspace clustering:

(a) A lesson research in ensemble clustering may want to learn from subspace clustering could be that diversity of clusterings could be a worthwhile goal in itself. We should differentiate here between significantly differing clusterings and just varying yet similar (i.e., correlated) clusterings [34, 6]. We believe, however, that it can be meaningful to unify the latter by some sort of consensus while it is in general not meaningful to try to unify substantially different clusterings.

(b) As we have seen above (Section 2), redundancy is a problem in traditional subspace clustering. Possibly, subspace clustering can benefit from advanced clustering diversity measures in ensemble clustering [44, 20, 28, 23, 21]. These measures, however, are usually based on some variant of pairwise mutual information where the overlap of clusters (i.e., the simultaneous membership of some subsets of objects in different clusters) is a problem.

4. ALTERNATIVE CLUSTERING

One direction in clustering based on constraints has been the constraint of diversity or non-redundancy, resulting in the discovery of different clustering solutions, sequentially or simultaneously [24, 25, 7, 14, 15, 13]. The motivation behind these approaches is that some results may already be known for a specific data set yet other results should be obtained that are new and interesting (cf. the classical definition of knowledge discovery in data as *"the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data"* [19]). Reproducing known or redundant patterns by clustering algorithms does not qualify as identifying *novel* patterns. A problem could be that the already known facts are dominant in a data set. The idea is, thus, to constrain some clustering algorithm by a set of patterns or links between objects that are *not* to identify. As they put in in [25]: *"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"*.

Some of these approaches again use ensemble techniques. Here, however, we are more interested in the relationship between these approaches and the area of subspace clustering. Hence, for our context, the interesting idea in this area is to use different subspaces as one possibility to find different clustering solutions [42]. As these approaches seek diversity usually constrained by non-redundancy, clearly subspace clustering research tackling the high redundancy level of subspace clusters can learn from these approaches. However, to allow a certain degree of redundancy could be meaningful as in turn can be learned from subspace clustering, allowing the overlap of clusters. In different subspaces, one subset of objects could belong to two different but meaningful clusters and hence increase the redundancy level of these clusters without rendering the report of both overlapping clusters meaningless. Indeed, considerations in this direction can actually be found in the research area of subspace clustering [27]. Also on part of alternative clustering research it has been conceded that it may be desirable to enable the conservation of certain already known properties of known concepts while seeking different clusters [42].

5. MULTIVIEW CLUSTERING

Multiview clustering [9, 12, 30] seeks clusterings in different subspaces of a data space. This kind of approach can be seen as a special case of seeking alternative clusterings (the constraint being the orthogonality of the subspaces) or as a special case of subspace clustering allowing maximal overlap yet seeking minimally redundant clusters by accommodating different concepts (as proposed e.g. in [27]). Because of these close relationships, we do not go into detail here. Let us note, though, that these approaches shed light on the observation learned in subspace clustering that highly overlapping clusters in different subspaces (i.e., certain subsets of objects may belong to several clusters simultaneously) need not be redundant nor meaningless.

6. CONCLUSION

In this paper, we shortly surveyed the research areas of subspace clustering, ensemble clustering, alternative clustering, and multiview clustering. We restricted our survey to the essential parts in order to highlight where similar problems are touched in different areas. As possible topics for the discussion between different areas we have identified:

1. How to treat diversity of clustering solutions? Should diverse clusterings always be unified? Allegedly, they should not — but under which conditions is a unification of divergent clusterings meaningful and when is it not?

2. Contrariwise, can we learn also from diversity itself? If in an ensemble of several clusterings in several arbitrary random subspaces, one clustering is exceptionally different from the others, it will be outnumbered in most voting procedures and lost. Could it not be especially interesting to report?

3. How to treat redundancy of clusters, especially in the presence of overlap between clusters? When does a cluster qualify as redundant w.r.t. another cluster, and when does it represent a different concept although many objects are part of both concepts? We have seen research on subspace clustering more and more trying to get rid of too redundant clusters while research on alternative clustering recently tends to allow some degree of redundancy. May there be a point where both research directions meet?

4. How to assess similarity between clustering solutions or single clusterings? Again, the presence of overlap between clusters increases the complexity of a mapping of clusters where the label correspondence problem makes it already non-trivial.

A problem unresolved so far and relevant for Multiview Clustering, Alternative Clustering and Subspace Clustering is how to evaluate clusterings that can overlap and relate to different concept levels. This problem is discussed more detailed (though without a concrete solution yet) in [22].

Although our perspective may be biased from the point of view of subspace clustering, we hope to stimulate discussion among researchers from different fields and encourage to learn from each other.

7. REFERENCES

[1] C. C. Aggarwal, C. M. Procopiuc, J. L. Wolf, P. S. Yu, and J. S. Park. Fast algorithms for projected clustering. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD), Philadelphia, PA*, 1999.

[2] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan. Automatic subspace clustering of high dimensional data for data mining applications. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD), Seattle, WA*, 1998.

[3] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD), Minneapolis, MN*, 1994.

[4] I. Assent, R. Krieger, E. Müller, and T. Seidl. DUSC: dimensionality unbiased subspace clustering. In *Proceedings of the 7th IEEE International Conference on Data Mining (ICDM), Omaha, NE*, 2007.

[5] I. Assent, R. Krieger, E. Müller, and T. Seidl. INSCY: indexing subspace clusters with in-process-removal of redundancy. In *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM), Pisa, Italy*, 2008.

[6] J. Azimi and X. Fern. Adaptive cluster ensemble selection. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI), Pasadena, CA*, 2009.

[7] 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*, 2006.

[8] 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*, 2005.

[9] S. Bickel and T. Scheffer. Multi-view clustering. In *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM), Brighton, UK*, 2004.

[10] 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*, 2006.

[11] C. H. Cheng, A. W.-C. Fu, and Y. Zhang. Entropy-based subspace clustering for mining numerical data. In *Proceedings of the 5th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), San Diego, CA*, pages 84–93, 1999.

[12] 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*, 2007.

[13] 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*, 2010.

[14] I. Davidson and Z. Qi. Finding alternative clusterings using constraints. In *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM), Pisa, Italy*, 2008.

[15] 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*, 2010.

[16] T. G. Dietterich. Ensemble methods in machine learning. In *First International Workshop on Multiple Classifier Systems (MCS), Cagliari, Italy*, 2000.

[17] T. G. Dietterich. Ensemble learning. In M. A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 405–408. MIT Press, 2nd edition, 2003.

[18] 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.

[19] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. Knowledge discovery and data mining: Towards a unifying framework. In *Proceedings of the 2nd ACM International Conference on Knowledge Discovery and Data Mining (KDD), Portland, OR*, 1996.

[20] 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*, 2003.

[21] X. Z. Fern and W. Lin. Cluster ensemble selection. *Statistical Analysis and Data Mining*, 1(3):128–141, 2008.

[22] 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.

[23] A. Gionis, H. Mannila, and P. Tsaparas. Clustering aggregation. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1), 2007.

[24] D. Gondek and T. Hofmann. Non-redundant data clustering. In *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM), Brighton, UK*, 2004.

[25] 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*, 2005.

[26] F. Gullo, A. Tagarelli, and S. Greco. Diversity-based weighting schemes for clsutering ensembles. In *Proceedings of the 9th SIAM International Conference on Data Mining (SDM), Sparks, NV*, 2009.

[27] 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*, 2009.

[28] S. T. Hadjitodorov, L. I. Kuncheva, and L. P. Todorova. Moderate diversity for better cluster ensembles. *Information Fusion*, 7(3):264–275, 2006.

[29] 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*, 2009.

[30] P. Jain, R. Meka, and I. S. Dhillon. Simultaneous unsupervised learning of disparate clusterings. *Statistical Analysis and Data Mining*, 1(3):195–210, 2008.

[31] K. Kailing, H.-P. Kriegel, and P. Kröger. Density-connected subspace clustering for high-dimensional data. In *Proceedings of the 4th SIAM International Conference on Data Mining (SDM), Lake Buena Vista, FL*, 2004.

[32] 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.

[33] P. Kröger and A. Zimek. Subspace clustering techniques. In L. Liu and M. T. Özsu, editors, *Encyclopedia of Database Systems*, pages 2873–2875. Springer, 2009.

[34] T. Li and C. Ding. Weighted consensus clustering. In *Proceedings of the 8th SIAM International Conference on Data Mining (SDM), Atlanta, GA*, 2008.

[35] G. Liu, K. Sim, J. Li, and L. Wong. Efficient mining of distance-based subspace clusters. *Statistical Analysis and Data Mining*, 2(5-6):427–444, 2009.

[36] B. Long, Z. Zhang, and P. S. Yu. Combining multiple clustering by sift correspondence. In *Proceedings of the 5th IEEE International Conference on Data Mining (ICDM), Houston, TX*, 2005.

[37] G. Moise and J. Sander. Finding non-redundant, statistically significant regions in high dimensional data: a novel approach to projected and subspace clustering. In *Proceedings of the 14th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Las Vegas, NV*, 2008.

[38] G. Moise, A. Zimek, P. Kröger, H.-P. Kriegel, and J. Sander. Subspace and projected clustering: Experimental evaluation and analysis. *Knowledge and Information Systems (KAIS)*, 21(3):299–326, 2009.

[39] E. Müller, I. Assent, S. Günnemann, R. Krieger, and T. Seidl. Relevant subspace clustering: Mining the most interesting non-redundant concepts in high dimensional data. In *Proceedings of the 9th IEEE International Conference on Data Mining (ICDM), Miami, FL*, 2009.

[40] E. Müller, S. Günnemann, I. Assent, and T. Seidl. Evaluating clustering in subspace projections of high dimensional data. In *Proceedings of the 35nd International Conference on Very Large Data Bases (VLDB), Lyon, France*, 2009.

[41] H. Nagesh, S. Goil, and A. Choudhary. Adaptive grids for clustering massive data sets. In *Proceedings of the 1st SIAM International Conference on Data Mining (SDM), Chicago, IL*, 2001.

[42] 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*, 2009.

[43] V. Singh, L. Mukherjee, J. Peng, and J. Xu. Ensemble clustering using semidefinite programming with applications. *Machine Learning*, 79(1-2), 2010.

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

[45] 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), 2005.

[46] 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*, 2004.

[47] G. Valentini and F. Masulli. Ensembles of learning machines. In *Proceedings of the 13th Italian Workshop on Neural Nets, Vietri, Italy*, 2002.