for the best subspace may come at prohibitive computational costs. This work is in line with the few filter approaches that exist in the literature (e.g. [1,3]) which limit themselves to the efficient identification of promising subspaces only, leaving the further cluster analysis to subsequent steps.

When searching for potentially small clusters in a noisy environment, we face various problems: (1) If the clusters are relatively small, global correlation measures may respond to them only marginally such that the chosen thresholds are not passed. (2) Density variations in single variables alone may cause high-dimensional spots look dense (but do not establish an worthwhile high-dim. cluster). (3) Any kind of density estimation involves some kind of threshold selection (e.g. the sampling area) and the impact of the selection may be easily underestimated. (4) Many weapons to reduce runtime (e.g. subsampling) do not apply successfully if a clusters size is only a small fraction of the noise.

The new ROSMULD algorithm (**r**anking **o**f **s**ubspaces by the **m**ost **u**nlikely high **l**ocal **d**ensity) overcomes these difficulties. By means of a rank-order transformation, all attributes become uniformly distributed, which eliminates density variations in single attributes. For each data point the subspace with the most surprisingly high data density is identified. Only if this density exceeds the expected density significantly, the data object *votes* for the respective subspace. Thresholds are automatically derived from the desired sensitivity (e.g. a cluster should have at least a density f times higher than the background noise). An exhaustive search for the most suprising subspace is avoided by employing new bounds on the used interestingness measure (without loosing completeness of the search).

ROSMULD successfully identifies subspaces with very small clusters and does not report any interesting subspace if the attributes are mutually independent. It performs also well on data sets with prominent and well-separated clusters. Compared to subspace clustering algorithms (cf. comparison in [5]) ROSMULD performs very competetive. For further details we refer to [2].

References

1. C. Baumgartner, K. Kailing, H.-P. Kriegel, P. Krüger, and C. Plant. Subspace Selection for Clustering High-Dimensional Data. In *ICDM*, 2004.
2. F. Höppner. A subspace filter supporting the discovery of small clusters in very noisy datasets. *Proc. 26th Int. Conf. on Scientific and Statistical Database Management - SSDBM '14*, 2014.
3. K. Kailing, H. Kriegel, P. Kröger, and S. Wanka. Ranking interesting subspaces for clustering high dimensional data. In *PKDD*, volume 2838, pages 241–252, 2003.
4. H.-P. Kriegel, P. Kröger, and A. Zimek. Clustering high-dimensional data. *ACM Transactions on Knowledge Discovery from Data*, 3(1):1–58, Mar. 2009.
5. E. Müller, S. Günnemann, I. Assent, and T. Seidl. Evaluating clustering in subspace projections of high dimensional data. *Proceedings of the VLDB*, 2(1):1270–1281, 2009.
6. K. Sim, V. Gopalkrishnan, A. Zimek, and G. Cong. A survey on enhanced subspace clustering. *Data Mining and Knowledge Discovery*, 26(2):332–397, Feb. 2012.