Recent Trends in Computational Intelligence in Life Sciences

Udo Seiffert[1,2], Frank-Michael Schleif[3], and Dietlind Zühlke[4,5]

1- Fraunhofer IFF Magdeburg, Dept. Biosystems Engineering
Sandtorstrasse 22, 39106 Magdeburg, Germany

2- University of Magdeburg, Institute of
Electronics, Signal Processing and Communications,
P.O. Box 4120, 39016 Magdeburg, Germany

3- University of Bielefeld, AG Theoretical Computer Science,
Universitätsstrasse 21-23, 33615 Bielefeld, Germany

4- RWTH Aachen University, Dept. Information Systems,
Ahornstrasse 55, 52056 Aachen, Germany

5- Fraunhofer FIT Sankt Augustin, Dept. Life Science Informatics,
Schloss Birlinghoven, 53754 Sankt Augustin, Germany

Abstract. Computational intelligence generally comprises a rather large set of – in a wider sense – adaptive and human-like data analysis and modelling methods. Due to some superior features – such as generalisation, trainability, coping with incomplete and inconsistent data, etc. – computational intelligence has found its way into numerous applications in almost all scientific disciplines. A very prominent field among them are life sciences that are characterised by some unique requirements in terms of data structure and analysis.

1 Introduction

Life sciences cover a large research field with challenging questions in domains such as (bio-)chemistry, biology, environmental research, or medicine. Not only recent technological developments allow the generation of large and very complex data sets in these fields. Often such data is no longer manageable by humans. Challenges are e.g. in the context of high-throughput processing to handle the high frequency of incoming data and its high-dimensionality by means of a large number of measured features. Also the structure of the measured data representing an object of interest is often challenging because the data may be very heterogeneous – combining different measurement sources – or can be of high-content type. Another specific challenge in life science supporting systems is the interface to the domain expert. Only those experts can provide suitable validity evaluation and gain semantic insight using the models built from the data.

Computational intelligence (CI) [1] methods have been successfully applied to biological and biomedical problems for several years, refer to [2, 3] for some early work as well as to [4, 5, 6]. Computational intelligence has a strong potential to be used to pre-process, model, and analyse such data focused on specific questions arising from the domain. Thus new strategies are needed to cope with the complexity of life science applications. A very effective way is to employ

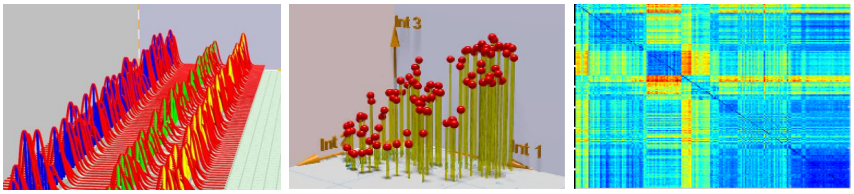

Fig. 1: Large spectra sets are common in life science with more and more sample by e.g. new imaging technologies (plot a). The spectra are typically encoded in less, but often still high-dimensional feature sets (plot b). Finally (dis-)similarities are frequently uses to represent the whole data set for clustering or classification models (plot c).

explicit or implicit domain expert knowledge about the data or the analysis task. As such knowledge may be available in very different forms – by means of appropriate (bio-)physical models, data specific distance measures, or analysis specific data encoding techniques – the strategies of integration are different. One can learn the parameters of a problem-specific model from available test data. In other cases the knowledge is used in the design of adaptive analysis algorithms to generate the desired meta-information out of the available data.

Examples where computational intelligence methods are used for data analysis are e.g. the analysis of spectroscopic data (see Figure 1) with a large number of measurements to reveal different types of biochemical information in a sample cohort; the automatic monitoring of cell cultures using microscopic images of the cells [7]; enabling systems in cancer research combining information from immune-histological images, gene-expression analysis, and clinical data, to mention just a few.

In medical research novel high-throughput measurement methods have led to an increasing amount of data providing potentially relevant information for the identification of disease markers or to support the development of new medical treatments and models. Additionally the results of different sample analyzers and multiple sources of expert information like specialized databases and ontology sources have to be integrated and are accessible. Computational intelligence methods play an important role and have been excessively utilized [8, 9, 10, 11, 12, 13, 14].

Clinical researchers need efficient methods to handle these large amount of data and effective tools to transfer this data into meta-data and information. This includes methods to explore measured data by means of clustering under different criteria, to evaluate data and models with respect to accuracy and reliability and approaches to provide maximum interpretability of models. Based on cognitive concepts, such as learning and prototypes, some machine learning tools approach already these needs but the complexity of the data, by means of varying data density, high dimensionality, and model reliability, are still very challenging.

For example in pathology or biomedical applications the labeling of data ac-

cording to different recognition tasks may already be challenging for the domain experts in two different ways: the vast amount of data to be labeled and the vagueness of the underlying concepts in biology. One concrete example, where experts struggle in labeling the data, is the identification of tumor tissue in immune-histological slice images of tumor probes. One possibility to handle this problem is to first cluster the tissues on the slices in an unsupervised manner, using tissue representations that – according to the domain expert – are relevant like graininess or intensity of different immune-histological markers. The result of this clustering can then be validated, annotated and corrected by the human domain expert for further supervised analysis.

In plant biology the *-omics* (*transcriptomics*, *proteomics*, *metabolomics*) era generated – along with typically huge data sets – the demand for complex and particularly integrative data analysis [15]. Computational intelligence paradigms in combination with state-of-the-art statistical methods have been extensively utilised to cope with large-scale data that is often incomplete and/or inconsistent. Particularly their excellent way to incorporate non-explicitly available expert knowledge made computational intelligence often the method of choice [16, 17, 18].

As a next step, the routine utilisation of *functional genomics* and genetical approaches along with automated and completely controlled growth chambers and greenhouses has led to an increasing demand for high-throughput phenotyping (*phenomics*). Not surprisingly, this is reflected in accordingly required data analysis [19] and spatiotemporal modelling [20]. Here, the generalisation ability of machine learning approaches has turned out to be an advantageous feature to cope with inherent biological variation, both between individuals [21] and along the time line of plant development [22].

Within the next years it is generally anticipated to find an increasing importance of

- data analysis on structured data with application specific semantics,
- non-standard metrics and dissimilarity measures to cope with application specific data properties,
- unbalanced and sparsely represented data that reflects the typical data structure in life sciences,
- approximation, visualization, and data reduction techniques to cope with large scale problems,
- reliability improvement and confidence estimates for model interpretation.

All of these topics do not just lead to a nice playground for computational intelligence paradigms. They can often only be addressed successfully by computational intelligence. In this context the present paper reflects a number of recent developments without claiming to provide a complete view.

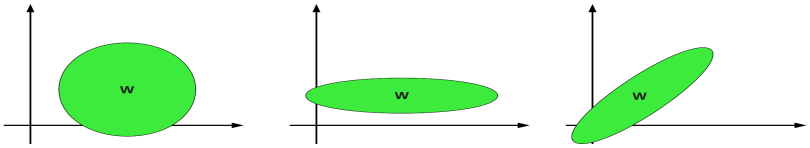

Fig. 2: Effect of different simple metric adaptations with respect to a cluster center (prototype) $\mathbf{w}$ in a 2D space. Left standard euclidean-, middle weighted euclidean-, right matrix-weighted euclidean distance. Details are given e.g. in [26]. More advanced data and problem specific metric adaptations have been found to be very effective in different computational intelligence approaches [28].

2 Development of novel methodologies

2.1 Non-standard metrics and dissimilarity measures

In general it is a promising approach to reflect characteristic data properties in the utilised data processing pipeline. This typically leads to an increased performance in tasks, such as clustering, classification, and non-linear regression, that are commonly addressed by machine learning methods.

One possible way to achieve this is to adapt the used metric according to the underlying data properties and application, respectively [23, 24, 25, 26], see Figure 2.

Spectral data has moved into the focus of data analysis recently. Particularly in case of prototype-based neural networks, that are essentially based on similarity measures, a very interesting approach is to consider the individual spectral signatures as *densities*, that are positive functions (patterns), not necessarily normalised but finite measures [27]. In general, a pairwise directed distance D between these densities is called a metric, if the following three conditions hold:

1. $D(\mathbf{X}\|\mathbf{W}) \geq 0$ (positive definiteness),
2. $D(\mathbf{X}\|\mathbf{W}) = D(\mathbf{W}\|\mathbf{X})$ (symmetry),
3. $D(\mathbf{X}\|\mathbf{Z}) \leq D(\mathbf{X}\|\mathbf{W}) + D(\mathbf{W}\|\mathbf{Z})$ (triangle inequality).

However, if only condition 1 is satisfied by a particular distance measure, it is not a metric but it is referred to as a divergence. Assuming that spectra are positive functions, which is typically the case, divergences could be applied as similarity, or more precisely dissimilarity measures to compare different spectral signatures. Common divergences are for example Kullback-Leibler, Hellinger, or Jensen-Shannon.

The concept of divergences is on no account restricted to the Euclidean space. A more general approach is to consider an abstract space where common divergences, such as Kullback-Leibler-, Csiszár-Morimoto-, or Bregman-divergence can be generalised towards Alpha-, Beta-, and Gamma-divergences. The fundamental properties of the underlying divergences remain present. Particularly the

Gamma-divergence seems to be very robust in terms of outliers [29]. Moreover, novel divergences offering tailored properties can be developed [30].

The properties of divergences are typically controlled by parameters. This tuning can be done in an elegant way by integration of divergences as dissimilarity measures or cost functions into machine learning approaches. In return, these machine learning approaches benefit from an extended choice of available dissimilarity measures and cost functions. As mentioned before, prototype- and vector-quantisation-based paradigms seem particularly suitable [28, 31].

In [32] a general method for the assessment of data attribute variability is suggested. This method provides a mathematically thorough characterisation of a certain attribute sensitivity using several general similarity measures. The properties of this method are shown by means of multi-spectral image analysis and segmentation.

The representation of biomedical objects is often composed from data coming from different sources (e.g. spectra as well as clinical data and image based information). For every type of data (data from one source) suitable similarity measures are (or become) available and thus appropriate integration of those measures into one similarity measure for the whole object representation is needed. Approaches towards this integration are matter of ongoing research (see e.g. [33] and [34]).

2.2 Efficient data processing for large scale problems

The novel data generation approaches applied in life sciences generate large amounts of data which may still be sparse in the data space due to the common high dimensionality of the measurements. Computational intelligence methods are used in multiple ways to reduce the dimensionality of the data [35, 36, 37] to summarize large data sets [16, 8] or to describe them in a more compact manner [38], providing maximum information.

While the basics of these steps are well known some recent trends can be identified leading to significantly improved models and approximation schemes. Recently the concept of random projection [39], or the more advanced method of compressed sensing [40], got significant attention [41, 42]. Employing the Johnson-Lindenstrauss Lemma (see e.g. [42]) already a random projection of the data to a lower (but not to low) data space preserves enough structure in the data to keep most of the information accessible. More recent approaches, based on additional mathematical results and proofs allow also the reconstruction of the original signal based on a rather small number of projected measurements, such that, theoretically already the recording of the data could be simplified. These findings are going to be used in many computational intelligence methods to simplify learning and data modelling tasks [42].

Another recent approach to simplify learning models in computational intelligence deals with relational data. This is very common also in life sciences, e.g. if the similarity between proteins is given by scores where the original data are not accessible or not available as regular feature vectors. Relational learning methods (see e.g. [43]) employ (dis-)similarity matrices, which can get huge for

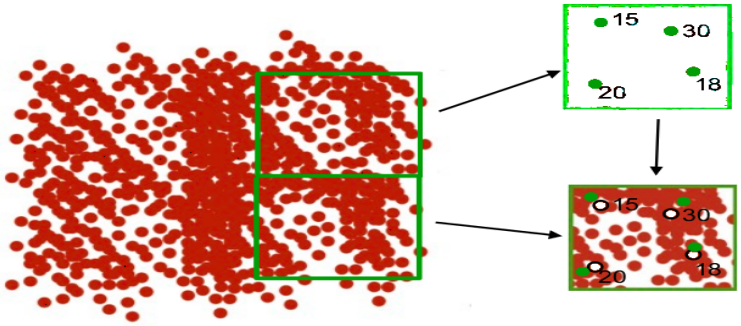

Fig. 3: Schema of patch clustering. The data set (left) is to large to fit into memory. Only a patch is processed and represented by a small cluster model (right top). The cluster model is used subsequently together with multiplicities (data statistics). The new patch and the former cluster model are reclustered to obtain a refined model (right bottom). Details in [43].

high-throughput data like in sequencing experiments. There are at least two approaches for large (dis-)similarity matrices, namely patch processing [43, 44] (see Figure 3), processing the similarities in patches while approximating the remaining data statistics in a smooth manner, and matrix approximation e.g. by the Nystroem approximation [45]. Both approaches rely implicitly on the fact that the information encoded in the data is still accessible also if a rather large amount of data points is ignored or only summarized. A further approach to approximate large data sets is accessible by the core-set theory [46] that becomes also more and more prominent in computational intelligence for life science.

2.3 Making models more trustful

Computational intelligence methods in the life science are often used for the prediction of some properties of a new item with respect to a model. To make such models applicable under true life settings the confidence, reliability and interpretability of the models gets most relevant. Many computational intelligence approaches like prototype based models show good interpretation properties but do not or only minor allow to estimate the confidence of a prediction or to judge the reliability of the models by itself. Initial steps for corresponding extensions have been suggested in [47, 48, 12, 49], see Figure 4. A further aspect is the incorporation of additional knowledge to make the models more realistic and therefore trustful as shown e.g. in [50] but also to combine different sources of information to support the models in learning *appropriate* solutions, like in [33].

Visualization is another key topic towards the interpretability of models built from the available data. Challenges in this field comprise: the handling of non-linearity of dimension reductions for visualization, compression and exhaustive evaluation on large data sets, and domain adequate interfaces for the human expert. In [51] discriminative dimension reduction for labeled data is realised

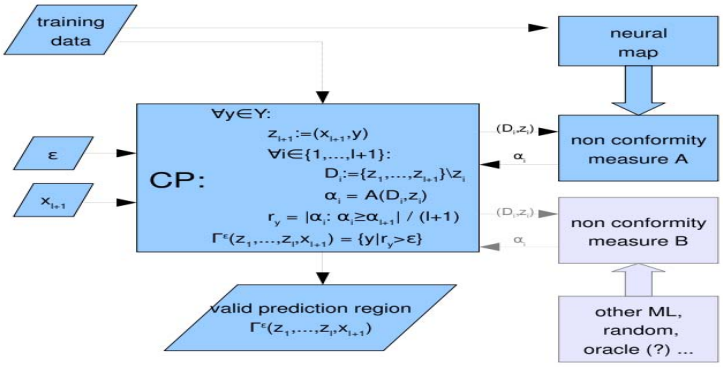

Fig. 4: Schema of conformal prediction [48]. The approach is suitable to judge the confidence and credibility of predictions made by a classifier method.

using an adaptive local dissimilarity measure.

3 Conclusions

The existence of big project consortia, a large number of primary publications, edited book volumes, and the second special session at ESANN within five years on this topic clearly demonstrates the importance of computational intelligence for life science applications. Although concrete applications on the one hand can be found in an extremely wide range and are typically rather specific then again, there are many challenging applications that share certain life science characteristics. Currently these seem to be, besides and beyond sheer data set size and the need of high-throughput processing, particular data properties such as sparseness and non-standard (non-Euclidean, non-correlation, etc.) and mixed data spaces. This is accompanied by properties that are also prominent outside the life sciences, such as spectrality and implicitly given expert knowledge. All this together benefits from and virtually requires computational intelligence based approaches.

References

[1] Andries P. Engelbrecht. *Computational Intelligence: An Introduction*. John Wiley and Sons, Hoboken, NJ, USA, 2003.

[2] Udo Seiffert, Lakhmi C. Jain, and Patrick Schweizer, editors. *Bioinformatics using Computational Intelligence Paradigms*, volume 176 of *Studies in Fuzziness and Soft Computing*. Springer-Verlag, Heidelberg, 2005.

[3] Udo Seiffert, Barbara Hammer, Samuel Kaski, and Thomas Villmann. Neural networks and machine learning in bioinformatics – theory and applications. In Michel Verleysen, editor, *Proceedings of the 14. European Symposium on Artificial Neural Networks ESANN 2006*, pages 521–532, Evere, Belgium, 2006. D-Side Publications.

[4] Gary B. Fogel, David W. Corne, and Yi Pan, editors. *Computational Intelligence in Bioinformatics*. IEEE Press Series on Computational Intelligence. John Wiley and Sons, Hoboken, NJ, USA, 2008.

[5] Tomasz G. Smolinski, Mariofanna G. Milanova, and Aboul-Ella Hassanien, editors. *Computational Intelligence in Biomedicine and Bioinformatics*, volume 151 of *Studies in Computational Intelligence*. Springer-Verlag, Berlin, 2008.

[6] Arpad Kelemen, Ajith Abraham, and Yuehui Chen, editors. *Computational Intelligence in Bioinformatics*, volume 94 of *Studies in Computational Intelligence*. Springer-Verlag, Berlin, 2008.

[7] Dirk Malthan, Roland Huchler, Albrecht Brandenburg, Hagen Thielecke, Cornelia Hildebrandt, and Dietlind Zühlke. Live cell monitoring – development of an automated cell cultivation and monitoring system. In *LabAutomation 2008 : Where Science, Technology and Industry Come Together. Final Program & Abstracts.*, page 74 ff, 2008.

[8] Sotiris K. Tasoulis, Dimitris K. Tasoulis, and Vassilis P. Plagianakos. Enhancing principal direction divisive clustering. *Pattern Recognition*, 43(10):3391–3411, 2010.

[9] Kenneth Lo, Florian Hahne, Ryan R. Brinkman, and Raphael Gottardo. flowClust: A bioconductor package for automated gating of flow cytometry data. *BMC Bioinformatics*, 10, 2009.

[10] Sounak Chakraborty and Ruixin Guo. A Bayesian hybrid Huberized support vector machine and its applications in high-dimensional medical data. *Computational Statistics and Data Analysis*, 55(3):1342–1356, 2011.

[11] Ingrid Branne, Kai Labusch, and Amir Madany Mamlouk. Sparse coding for feature selection on genome-wide association data. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6352 LNCS(PART 1):337–346, 2010.

[12] Frank-Michael Schleif, Thomas Villmann, Markus Kostrzewa, Barbara Hammer, and Alex Gammerman. Cancer informatics by prototype networks in mass spectrometry. *Artificial Intelligence in Medicine*, 45(2-3):215–228, 2009.

[13] Julien Boccarda, Alexandros Kalousisb, Melanie Hilariob, Pierre Lantéric, Mohamed Hanafid, Gérard Mazerollese, Jean-Luc Wolfendera, Pierre-Alain Carrupta, and Serge Rudaz. Standard machine learning algorithms applied to UPLC-TOF/MS metabolic fingerprinting for the discovery of wound biomarkers in Arabidopsis thaliana. *Chemometrics and Intelligent Laboratory Systems*, 104(1):20–27, 2010.

[14] Zsuzsanna Dosztányi, Bálint Mészáros, and Bálint Mészáros. Bioinformatical approaches to characterize intrinsically disordered/unstructured proteins. *Briefings in Bioinformatics*, 11(2):225–243, 2010.

[15] Nese Srinivasulu, Ljudmilla Borisjuk, Björn Junker, Hans-Peter Mock, Hardy Rolletschek, Udo Seiffert, Winfriede Weschke, and Ulrich Wobus. Barley grain development: Toward an integrative view. In Kwang W. Jeon, editor, *International Review of Cell and Molecular Biology*, volume 281, pages 49–89. Elsevier Academic Press, Burlington, 2010.

[16] Barbara Hammer, Alexander Hasenfuss, Frank-Michael Schleif, Thomas Villmann, Marc Strickert, and Udo Seiffert. Intuitive clustering of biological data. In *Proceedings of the IEEE International Joint Conference on Neural Networks IJCNN 2007*, pages 1877–1882, Piscataway, NJ, USA, 2007. IEEE Press.

[17] Thomas Villmann, Erzsébet Merényi, and Udo Seiffert. Machine learning approaches and pattern recognition for spectral data. In Michel Verleysen, editor, *Proceedings of the 16. European Symposium on Artificial Neural Networks ESANN 2008*, pages 433–444, Evere, Belgium, 2008. D-Side Publications.

[18] Felix Bollenbeck, Rainer Pielot, Diana Weier, Winfriede Weschke, and Udo Seiffert. Intermodality registration of NMRi and histological section images using neural networks regression in Gabor feature space. In *Proceedings of the IEEE Symposium on Computational Intelligence for Image Processing CIIP 2009*, pages 27–34. IEEE Press, 2009.

[19] Udo Seiffert, Felix Bollenbeck, Hans-Peter Mock, and Andrea Matros. Clustering of crop phenotypes by means of hyperspectral signatures using artificial neural networks. In *Proceedings of the 2nd IEEE Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing WHISPERS 2010*, pages 31–34. IEEE Press, 2010.

[20] Tobias Baum, Aura Navarro-Quezada, Wolfgang Knogge, Dimitar Duchkov, Patrick Schweizer, and Udo Seiffert. HyphArea – Automated analysis of spatiotemporal fungal patterns. *Journal of Plant Physiology*, 168:72–78, 2011.

[21] Felix Bollenbeck, Diana Weier, Wolfram Schoor, and Udo Seiffert. From individual intensity voxel data to inter-individual probabilistic atlases of biological objects by an interleaved registration-segmentation approach. In *Proceedings of the 4. International Conference on Computer Vision and Applications VISAPP 2009*, volume 1, pages 125–129, 2009.

[22] Rainer Pielot, Udo Seiffert, Bertram Manz, Diana Weier, Frank Volke, and Winfriede Weschke. 4D warping for analysing morphological changes in seed development of barley grains. In *Proceedings of the 3. International Conference on Computer Vision and Applications VISAPP 2008*, volume 1, pages 335–340. INSTICC and University of Madeira, 2008.

[23] Hong Chang and Dit-Yan Yeung. Locally linear metric adaptation for semi-supervised clustering. In *Proceedings of the 21. International Conference on Machine Learning*, ICML '04, pages 153–160, New York, NY, USA, 2004. ACM.

[24] Barbara Hammer, Marc Strickert, and Thomas Villmann. Supervised Neural Gas with general similarity measure. *Neural Processing Letters*, 21:21–44, 2005.

[25] Marika Kästner, Barbara Hammer, Michael Biehl, and Thomas Villmann. Generalized functional relevance learning vector quantization. In Michel Verleysen, editor, *Proceedings of the 19. European Symposium on Artificial Neural Networks ESANN 2011*, page numbers to be obtained from ToC of this proceedings book, Evere, Belgium, 2011. D-Side Publications.

[26] Banchar Arnonkijpanich, Alexander Hasenfuss, and Barbara Hammer. Local matrix adaptation in topographic neural maps. *Neurocomputing*, 74(4):522–539, 2011.

[27] Thomas Villmann. Utilization of divergences to compare density functions. Personal communication on the 2nd Mittweida Workshop on Computational Intelligence MiWoCi 2010, June 2010.

[28] Thomas Villmann, Sven Haase, Frank-Michael Schleif, and Barbara Hammer. Divergence based online learning in vector quantization. In *Artificial Intelligence and Soft Computing*, volume 6113 of *Lecture Notes in Computer Science*, pages 479–486. Springer-Verlag, Berlin, 2010.

[29] Hironori Fujisawa and Shinto Eguchi. Robust parameter estimation with a small bias against heavy contamination. *Journal of Multivariate Analysis*, 99(9):2053 – 2081, 2008.

[30] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and Shun-Ichi Amari. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. John Wiley and Sons, Ltd, 2009.

[31] Dong-Chul Park, Chung Nguyen, and Yunsik Lee. Content-based classification of images using centroid neural network with divergence measure. In *AI 2006: Advances in Artificial Intelligence*, volume 4304 of *Lecture Notes in Computer Science*, pages 729–738. Springer-Verlag, Berlin, 2006.

[32] Mark Strickert, Björn Labitzke, Andreas Kolb, and Thomas Villmann. Multispectral image characterization by partial generalized covariance. In Michel Verleysen, editor, *Proceedings of the 19. European Symposium on Artificial Neural Networks ESANN 2011*, page numbers to be obtained from ToC of this proceedings book, Evere, Belgium, 2011. D-Side Publications.

[33] Dietlind Zühlke, Frank-Michael Schleif, Tina Geweniger, and Thomas Villmann. Learning vector quantization for heterogeneous structured data. In Michel Verleysen, editor, *Proceedings of the 18. European Symposium on Artificial Neural Networks ESANN 2010*, pages 271–276, Evere, Belgium, 2010. D-Side Publications.

[34] Milton García-Borroto and José Ruiz-Shulcloper. Selecting prototypes in mixed incomplete data. In Alberto Sanfeliu and Manuel Cortes, editors, *Progress in Pattern Recognition, Image Analysis and Applications*, volume 3773 of *Lecture Notes in Computer Science*, pages 450–459. Springer Berlin, 2005.

[35] Luis F. Giraldo, Fernando Lozando, and Nicanor Quijano. Foraging theory for dimensionality reduction of clustered data. *Machine Learning*, 82(1):71–90, 2011.

[36] Longcun Jin, Wanggen Wan, Yongliang Wu, Bin Cui, and Xiaoqing Yu. A general framework for high-dimensional data reduction using unsupervised Bayesian model. *Communications in Computer and Information Science*, 98 CCIS(PART 2):96–101, 2010.

[37] Nir Ailon and Bernard Chazelle. Faster dimension reduction. *Communications of the ACM*, 53(2):97–104, 2010.

[38] Martijn Dijkstra and Ritsert C. Jansen. Optimal analysis of complex protein mass spectra. *Proteomics*, 9(15):3869–3876, 2009.

[39] Angshul Majumdar and Rabab K. Ward. Robust classifiers for data reduced via random projections. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 40(5):1359–1371, 2010.

[40] David Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52:1289 – 1306, 2006.

[41] David Donoho. High-dimensional data analysis: The curses and blessings of dimensionality, 2010. http://www-stat.stanford.edu/donoho/Lectures/AMS2000/Curses.pdf (30.09.2010).

[42] Richard G. Baraniuk and Michael B. Wakin. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 9(1):51–77, 2009.

[43] Barbara Hammer and Alexander Hasenfuss. Clustering very large dissimilarity data sets. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5998 LNAI:259–273, 2010.

[44] Zhu Xibin and Barbara Hammer. Patch affinity propagation. In Michel Verleysen, editor, *Proceedings of the 19. European Symposium on Artificial Neural Networks ESANN 2011*, page numbers to be obtained from ToC of this proceedings book, Evere, Belgium, 2011. D-Side Publications.

[45] Mu Li, James T. Kwok, and Bao-Liang Lu. Making large-scale Nyström approximation possible. In *Proceedings of the International Conference on Machine Learning (ICML)'2010*, 2009.

[46] Liang Chang, Xiao-Ming Deng, Sui-Wu Zheng, and Yong-Qing Wang. Scaling up kernel grower clustering method for large data sets via core-sets. *Acta Automatica Sinica*, 34(3):376–382, 2008.

[47] David Baehrens, Timon Schroeter, Stefan Harmeling, Motoaki Kawanabe, Katja Hansen, and Klaus-Robert Müller. How to explain individual classification decisions. *Journal of Machine Learning Research*, 11:1803–1831, 2010.

[48] Alex Gammerman and V. Vovk. Hedging predictions in machine learning. *The Computer Journal*, 50(2):151–163, 2007.

[49] Bin Tong, ZhiGuang Qin, and Einoshin Suzuki. Topology preserving SOM with transductive confidence machine. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6332 LNAI:27–41, 2010. Conference of 13th International Conference on Discovery Science, DS 2010; Conference Date: 6 October 2010 through 8 October 2010.

[50] Volkan Vural, Glenn Fung, Balaji Krishnapuram, Jennifer G. Dy, and R. Bharat Rao. Using local dependencies within batches to improve large margin classifiers. *Journal of Machine Learning Research*, 10:183–206, 2009.

[51] Kerstin Bunte, Barbara Hammer, Axel Wismüller, and Michael Biehl. Adaptive local dissimilarity measures for discriminative dimension reduction of labeled data. *Neurocomput.*, 73:1074–1092, March 2010.