Supervised Neural Gas for Classification of Functional Data and Its Application to the Analysis of Clinical Proteom Spectra

Frank-Michael Schleif[1], Thomas Villmann[1], and Barbara Hammer[2]

[1] University Leipzig, Dept. of Medicine, 04107 Leipzig, Germany
Tel.: +49(0)3419718896
{schleif,villmann}@informatik.uni-leipzig.de
[2] TU-Clausthal, Dept. of Math. & C.S., 38678 Clausthal-Zellerfeld, Germany
Tel.: +49(0)53237271[86][39]
hammer@in.tu-clausthal.de

Abstract. The analysis of functional data, is a common task in bioinformatics. Spectral data as obtained from mass spectrometric measurements in clinical proteomics are such functional data leading to new challenges for an appropriate analysis. Here we focus on the determination of classification models for such data. In general the available approaches for this task initially transform the spectra into a vector space followed by training a classifier. Hereby the functional nature of the data is typically lost, which may lead to suboptimal classifier models. Taking this into account a wavelet encoding is applied onto the spectral data leading to a compact *functional* representation. Further the Supervised Neural Gas classifier is extended by a functional metric. This allows the classifier to utilize the functional nature of the data in the modeling process. The presented method is applied to clinical proteom data showing good results.

Keywords: supervised neural gas, functional, data analysis, clinical proteomics, wavelet analysis, spectra preprocessing.

1 Introduction

Applications of mass spectrometry (ms) in clinical proteomics have gained tremendous visibility in the scientific and clinical community [10,4]. One major objective is the search for potential classification models for cancer studies. For this purpose, efficient analysis and visualization of large high-dimensional data sets derived from patient cohorts is crucial. Additionally, it is necessary to apply statistical analysis and pattern matching algorithms to attain validated signal patterns. Here we focus on the determination of classification models discriminating between multiple classes. A powerful tool to achieve such models with high generalization abilities is available with the prototype based Supervised Neural Gas algorithm (SNG) [11]. Like all nearest prototype classifier algorithms, SNG heavily relies on the metric d, usually the standard euclidean metric. For high-dimensional data as they occur in proteomic patterns, this choice is not adequate due to two reasons, first the functional nature of the data should be kept as far as

possible, second the noise present in the data set accumulates and likely disrupts the classification taking a standard euclidean approach. Thus, a functional representation of the data with respect to the used metric and a weighting or pruning of especially (priory not known) irrelevant function parts of the inputs, would be desirable. Therefore we focus on a functional distance measure as recently proposed in [6] which will be referred as functional metric. Further feature selection is applied based on a statistical pre-analysis of the data. Hereby a discriminative data representation is necessary. The extraction of such discriminant features is critical for spectral data and typically done by a parametric peak picking procedure. This peak picking is often focus of criticism because peaks may be insufficiently detected and the functional nature of the data is partially lost. To avoid this difficulties we focus on a wavelet encoding of the spectral data to get discriminative features. Thereby the obtained wavelet coefficients are sufficient to reconstruct the signal, still containing all relevant information of the spectra. However this better discriminating set of features is typically more complex and hence a robust approach to determine the desired classification model is needed. The paper is organized as follows: first the bioinformatic methods, namely the wavelet encoding of the spectral data and the SNG algorithm with the functional distance measure is presented. In the second part the clinical data are described and the introduced methods are applied in the analysis of the proteom spectra followed by some experiments, the paper is closed by the conclusions.

2   Bioinformatic Methods

The classification of mass spectra involves in general the two steps peak picking to locate and quantify positions of peaks within the spectrum and feature extraction from the obtained peak list. In the first step a number of procedures as baseline correction, optional denoising, noise estimation and normalization must be applied. Upon these prepared spectra the peaks have to be identified by scanning all local maxima and the associated peak endpoints followed by a S/N thresholding such that one obtains the desired peak list.

The procedure of baseline correction and recalibration (alignment) of multiple spectra is standard, and has been done using ClinProTools in this paper (details in [4])[1]. Here we propose an alternative feature extraction procedure preserving all (potentially small) peaks containing relevant information by use of the discrete wavelet transformation (DWT). The feature extraction has been done by Wavelet analysis using the Matlab Wavelet-Toolbox[2], due to the local analysis property of wavelet analysis the features can still be related back to original mass position in the spectral data which is essential for further biomarker analysis. In a first step a feature selection procedure using the Kolmogorov-Smirnoff test (KS-test) was applied. Thereby the test was used to identify features which show a significant ($p < 0.01$) discrimination between the two groups (cancer,control).

---

[1] Biomarker software available at http://www.bdal.de
[2] The Matlab Wavelet-Toolbox can be obtained from www.mathworks.com

This is done in accordance to [12] were also a generation to a multiclass experiment is given. The roughly reduced set has been further processed by SNG to obtain a classification model with a *small*, ranked set of features, crossvalidated by a 10-fold cross validation procedure.

2.1   Feature Extraction and Denoising by Bi-orthogonal Discrete Wavelet Transform

Wavelets have been developed as powerful tools [1,7] used for noise removal and data compression. The discrete version of the continuous wavelet transform leads to the concept of a multiresolution analysis (MRA). This allows a fast and stable wavelet analysis and synthesis. The analysis becomes more precise if the wavelet shape is adapted to the signal to be analyzed. For this reason one can apply the so called bi-orthogonal wavelet transform[2] which uses two pairs of scaling and wavelet functions. One is for the decomposition/analysis and the other one for reconstruction/synthesis. The advantage of the bi-orthogonal wavelet transform is the higher degree of freedom for the shape of the scaling and wavelet function. In our analysis such a smooth synthesis pair was chosen to avoid artifacts. It can be expected that a signal in the time domain can be represented by a small number of a relatively large set of coefficients from the wavelet domain. The spectra are reconstructed in dependence of a certain approximation level L of the MRA which can be considered as a hard-thresholding. The denoised spectrum looks similar to the reconstruction as depicted in Figure 1. The starting point for an argumentation is the simplest example of a MRA which can be defined by the characteristic function $\chi_{[0,1)}$. The corresponding wavelet is the so-called *Haar* wavelet. Assume that the denoised spectrum $f \in L_2(\mathbb{R})$ has a peak with endpoints $2^j k$ and $2^j(k+1)$, the integral of the peak can be written as

$$\int_{2^j k}^{2^j(k+1)} f(t)dt = \int_{\mathbb{R}} f(t)\chi_{[2^j k, 2^j(k+1))}(t)dt$$

Obviously the right hand side is the Haar DWT scaling coefficient $c_{j,k} = \langle f, \psi_{j,k} \rangle$ at scale $a = 2^j$ and translation $b = 2^j k$. One obtains approximation- and detail-coefficients [2]. The approximation coefficients describe a generalized peak list of the denoised spectrum encoding primal spectral information and depending on the level L which is determined with respect to the measurement procedure. For linear MALDI-TOF spectra a device resolution of $500 - 800Da$ can be expected. This implies limits to the minimal peak width in the spectrum and hence, the reconstruction level of the Wavelet-Analysis should be able to model corresponding peaks. A level $L = 4$ is appropriate for our problem (see Figure 1). Applying this procedure including the KS-test on the spectra with an initial number of 22306 measurement points per spectrum one obtains 602 wavelet coefficients used as representative features per spectrum, still allowing a reliable functional representation of the data. The coefficients were used to reconstruct the spectra and the final functional representation was based on the maximal intensity values at the mass positions encoded by the wavelet coefficients. Subsequently, the data were processed by SNG in a 10-fold cross validation procedure.
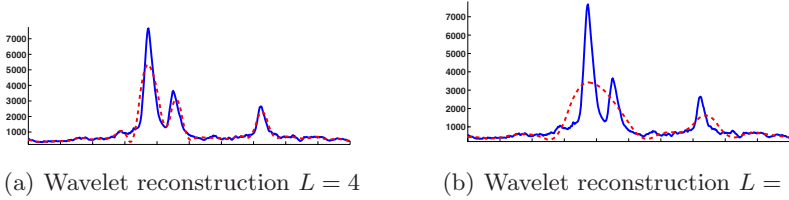
(a) Wavelet reconstruction $L = 4$ (b) Wavelet reconstruction $L = 5$

Fig. 1. Wavelet reconstruction of the spectra with $L = 4, 5$, x-mass positions, y-arbitrary unit. The original signal is plotted with the solid line. One observes that a wavelet analysis with $L = 5$ is to rough to approximate the sharp peaks.

2.2   Supervised Neural Gas for Functional Data

Supervised Neural Gas (SNG) is considered as a representative for prototype based classification approaches as introduced by KOHONEN. Different prototype classifiers have been proposed so far [5,9,3,11] as improvements of the original approach. The SNG has been introduced in [11] and combines ideas from the Neural Gas algorithm (NG)introduced in [8] with the Generalized learning vector quantizer (GLVQ) as given in [9]. Subsequently we give the basic notations and some remarks to the integration of alternative metrices into Supervised Neural Gas (SNG). Details on SNG including convergence proofs can be found in [11].

Let us first clarify some notations: Let $c_{\mathbf{v}} \in \mathcal{L}$ be the label of input $\mathbf{v}$, $\mathcal{L}$ a set of labels (classes) with $\#\mathcal{L} = N_{\mathcal{L}}$. Let $V \subseteq \mathbb{R}^{D_V}$ be a finite set of inputs $\mathbf{v}$. LVQ uses a fixed number of prototypes (weight vectors, codebook vectors) for each class. Let $\mathbf{W} = \{\mathbf{w_r}\}$ be the set of all codebook vectors and $c_{\mathbf{r}}$ be the class label of $\mathbf{w_r}$. Furthermore, let $\mathbf{W}_c = \{\mathbf{w_r}|c_{\mathbf{r}} = c\}$ be the subset of prototypes assigned to class $c \in \mathcal{L}$.

The task of vector quantization is realized by the map Ψ as a winner-take-all rule, i.e. a stimulus vector $\mathbf{v} \in V$ is mapped onto that neuron $\mathbf{s} \in A$ the pointer $\mathbf{w}_s$ of which is closest to the presented stimulus vector $\mathbf{v}$,

$$\Psi_{V \to \mathcal{A}}^{\lambda} : \mathbf{v} \mapsto \mathbf{s}(\mathbf{v}) = \operatorname{argmin}_{\mathbf{r} \in A} d^{\lambda}(\mathbf{v}, \mathbf{w_r}) \tag{1}$$

with $d^{\lambda}(\mathbf{v}, \mathbf{w})$ being an arbitrary differentiable similarity[3] measure which may depend on a parameter vector λ. For the moment we take λ as fixed. The neuron $\mathbf{s}(\mathbf{v})$ is called winner or best matching unit. The subset of the input space

$$\Omega_{\mathbf{r}}^{\lambda} = \{\mathbf{v} \in V : \mathbf{r} = \Psi_{V \to A}(\mathbf{v})\} \tag{2}$$

which is mapped to a particular neuron $\mathbf{r}$ according to (1), forms the (masked) receptive field of that neuron forming a Voronoi tesselation. If the class information of the weight vector is used, the boundaries $\partial\Omega_{\mathbf{r}}^{\lambda}$ generate the decision boundaries for classes. A training algorithm should adapt the prototypes such that for each class $c \in \mathcal{L}$, the corresponding codebook vectors $\mathbf{W}_c$ represent the

---

[3] A similarity measure is a non-negative real-valued function, which, in contrast to a distance measure does not necessarily fulfill the triangle inequality and the symmetry property.

class as accurately as possible. This means that the set of points in any given class $V_c = \{\mathbf{v} \in V | c_\mathbf{v} = c\}$, and the union $\mathcal{U}_c = \bigcup_{\mathbf{r}|_{\mathbf{w}_\mathbf{r} \in \mathbf{W}_c}} \Omega_\mathbf{r}$ of receptive fields of the corresponding prototypes should differ as little as possible.

Supervised Neural Gas (SNG) constitutes a combination of GLVQ and NG. Again, let $\mathbf{W}_c = \{\mathbf{w}_\mathbf{r} | c_\mathbf{r} = c\}$ be the subset of prototypes assigned to class $c \in \mathcal{L}$ and K_c its cardinality. Further we assume to have m data vectors $\mathbf{v}_i$. As pointed out in [11], the neighborhood learning for a given input $\mathbf{v}_i$ with label c is applied to the subset $\mathbf{W}_c$. The respective cost function is

$$Cost_{SNG}(\gamma) = \sum_{i=1}^{m} \sum_{\mathbf{r}|\mathbf{w}_\mathbf{r} \in \mathbf{W}_{c_i}} \frac{h_\gamma(\mathbf{r}, \mathbf{v}_i, \mathbf{W}_{c_i}) \cdot f(\mu_\lambda(\mathbf{r}, \mathbf{v}))}{C(\gamma, K_{c_i})} \tag{3}$$

with $f(x) = (1 + \exp(-x))^{-1}$, $h_\gamma(\mathbf{r}, \mathbf{v}, \mathbf{W}) = \exp\left(-\frac{k_\mathbf{r}(\mathbf{v}, \mathbf{W})}{\gamma}\right)$ and $\mu_\lambda(\mathbf{r}, \mathbf{v}) = \frac{d_\mathbf{r}^\lambda - d_{\mathbf{r}_-}^\lambda}{d_\mathbf{r}^\lambda + d_{\mathbf{r}_-}^\lambda}$ whereby $d_{\mathbf{r}_-}^\lambda$ is defined as the squared distance to the best matching prototype but labeled with $c_{\mathbf{r}_-} \neq c_\mathbf{v}$, say $\mathbf{w}_{\mathbf{r}_-}$ and $d_\mathbf{r}^\lambda = d^\lambda(\mathbf{v}, \mathbf{w}_\mathbf{r})$. The neighborhood cooperativeness makes sure that prototypes are spread faithfully among data of their respective classes. Note that $\lim_{\gamma \to 0} Cost_{SNG}(\gamma) = Cost_{GLVQ}$ holds [11]. Hence, for vanishing neighborhood the SNG also becomes optimal in the sense of margin analysis, as detailed below. However, if the neighborhood range γ is large, typically at the beginning of the training, the prototypes of one class share their responsibilities for a given input. Hence, neighborhood cooperation is involved such that initialization of the prototypes is not longer crucial. Given a training example $(\mathbf{v}_i, c_i)$ all prototypes $\mathbf{w}_\mathbf{r} \in \mathbf{W}_{c_i}$ and the closest wrong prototype $\mathbf{w}_{\mathbf{r}_-}$ are adapted. Taking now

$$\xi_\mathbf{r}^+ = \frac{2 \cdot d_{\mathbf{r}_-}^\lambda}{(d_\mathbf{r}^\lambda + d_{\mathbf{r}_-}^\lambda)^2} \text{ and } \xi_\mathbf{r}^- = \frac{2 \cdot d_\mathbf{r}^\lambda}{(d_\mathbf{r}^\lambda + d_{\mathbf{r}_-}^\lambda)^2} \tag{4}$$

we get the update rules

$$\triangle \mathbf{w}_\mathbf{r} = \epsilon^+ \cdot \xi_\mathbf{r}^+ \cdot \frac{f'|_{\mu_\lambda(\mathbf{r}, \mathbf{v})} \cdot h_\gamma(\mathbf{r}, \mathbf{v}_i, \mathbf{W}_{c_i})}{C(\gamma, K_{c_i})} \cdot \frac{\partial d_\mathbf{r}^\lambda}{\partial \mathbf{w}_\mathbf{r}} \tag{5}$$

$$\triangle \mathbf{w}_{\mathbf{r}_-} = -\epsilon^- \cdot \sum_{\mathbf{r}|\mathbf{w}_\mathbf{r} \in \mathbf{W}_c} \xi_\mathbf{r}^- \cdot \frac{f'|_{\mu_\lambda(\mathbf{r}, \mathbf{v})} \cdot h_\gamma(\mathbf{r}, \mathbf{v}_i, \mathbf{W}_{c_i})}{C(\gamma, K_{c_i})} \cdot \frac{\partial d_{\mathbf{r}_-}^\lambda}{\partial \mathbf{w}_{\mathbf{r}_-}} \tag{6}$$

We remark that only for correct prototypes the neighborhood cooperativeness is applied. Yet, one could also include neighborhood cooperation for wrong prototypes. However, as shown in [11] this yields instabilities of learning.

Incorporation of a Functional Metric to SNG. As pointed out before, the similarity measure $d^\lambda(\mathbf{v}, \mathbf{w})$ is only required to be differentiable with respect to λ and $\mathbf{w}$. The triangle inequality has not to be fulfilled necessarily. This leads to a great freedom in the choice of suitable measures and allows the usage of non-standard metrics in a natural way. We now review the functional metric as given

in [6], the obtained derivations can be plugged into the above equations leading to SNG with a functional metric, whereby the data are functions represented by vectors and, hence, the vector dimensions are spatially correlated.

Common vector processing does not take the spatial order of the coordinates into account. As a consequence, the functional aspect of spectral data is lost. For proteom spectra the order of signal features (peaks) is due to the nature of the underlying biological samples and the measurement procedure. The masses of measured chemical compounds are given ascending and peaks encoding chemical structures with a higher mass follows chemical structures with lower masses. In addition multiple peaks with different masses may encode parts of the same chemical structure and hence are correlated.

LEE proposed a distance measure taking the functional structure into account by involving the previous and next values of x_i in te i-th term of the sum, instead of x_i alone. Assuming a constant sampling period τ, the proposed norm is:

$$\mathcal{L}_p^{fc}(\mathbf{v}) = \left(\sum_{k=1}^{D} (A_{k-1}(\mathbf{v}) + A_{k+1}(\mathbf{v}))^p \right)^{\frac{1}{p}} \tag{7}$$

with

$$A_k(\mathbf{v}) = \begin{cases} \frac{\tau}{2}|v_k| & \text{if } 0 \le v_k v_{k-1} \\ \frac{\tau}{2}\frac{v_k^2}{|v_k|+|v_{k-1}|} & \text{if } 0 > v_k v_{k-1} \end{cases} \quad B_k(\mathbf{v}) = \begin{cases} \frac{\tau}{2}|v_k| & \text{if } 0 \le v_k v_{k+1} \\ \frac{\tau}{2}\frac{v_k^2}{|v_k|+|v_{k+1}|} & \text{if } 0 > v_k v_{k+1} \end{cases} \tag{8}$$

are respectively of the triangles on the left and right sides of x_i. Just as for L_p, the value of p is assumed to be a positive integer. At the left and right ends of the sequence, x_0 and x_D are assumed to be equal to zero. The derivatives for the functional metric taking $p = 2$ are given in [6]. Now we consider the scaled functional norm where each dimension v_i is scaled by a parameter $\lambda_i > 0$ $\lambda_i \in (0, 1]$ and $\sum_i \lambda_i = 1$. Then the scaled functional norm is:

$$\mathcal{L}_p^{fc}(\lambda\mathbf{v}) = \left(\sum_{k=1}^{D} (A_{k-1}(\lambda\mathbf{v}) + A_{k+1}(\lambda\mathbf{v}))^p \right)^{\frac{1}{p}} \tag{9}$$

with

$$A_k(\lambda\mathbf{v}) = \begin{cases} \frac{\tau}{2}\lambda_k|v_k| & \text{if } 0 \le v_k v_{k-1} \\ \frac{\tau}{2}\frac{\lambda_k^2 v_k^2}{\lambda_k|v_k|+\lambda_{k-1}|v_{k-1}|} & \text{else} \end{cases} \quad B_k(\lambda\mathbf{v}) = \begin{cases} \frac{\tau}{2}\lambda_k|v_k| & \text{if } 0 \le v_k v_{k+1} \\ \frac{\tau}{2}\frac{\lambda_k^2 v_k^2}{\lambda_k|v_k|+\lambda_{k+1}|v_{k+1}|} & \text{else} \end{cases} \tag{10}$$

The prototype update changes to:

$$\frac{\partial \delta_2^2(\mathbf{x}, \mathbf{y}, \lambda)}{\partial x_k} = \frac{\tau^2}{2}(2 - U_{k-1} - U_{k+1})(V_{k-1} + V_{k+1})\triangle_k \tag{11}$$

with

$$U_{k-1} = \begin{cases} 0 \text{ if } 0 \leq \triangle_k\triangle_{k-1} \\ \left(\frac{\lambda_{k-1}\triangle_{k-1}}{\lambda_k|\triangle_k|+\lambda_{k-1}|\triangle_{k-1}|}\right)^2 \text{ else} \end{cases} , U_{k+1} = \begin{cases} 0 \text{ if } 0 \leq \triangle_k\triangle_{k+1} \\ \left(\frac{\lambda_{k+1}\triangle_{k+1}}{\lambda_k|\triangle_k|+\lambda_{k+1}|\triangle_{k+1}|}\right)^2 \text{ else} \end{cases}$$

$$V_{k-1} = \begin{cases} 1\lambda_k \text{ if } 0 \leq \triangle_k\triangle_{k-1} \\ \frac{\lambda_k|\triangle_k|}{\lambda_k|\triangle_k|+\lambda_{k-1}|\triangle_{k-1}|} \text{ else} \end{cases} , V_{k+1} = \begin{cases} 1\lambda_k \text{ if } 0 \leq \triangle_k\triangle_{k+1} \\ \frac{\lambda_k|\triangle_k|}{\lambda_k|\triangle_k|+\lambda_{k+1}|\triangle_{k+1}|} \text{ else} \end{cases}$$

and $\triangle_k = x_k - y_k$ Using this parametrization one can emphasize/neglect different parts of the function for classification. This distance measure can be put into SNG as shown above and has been applied subsequently in the analysis of clinical proteom spectra.

3   Analysis of Proteomic Data

Subsequently the proposed data processing scheme is applied to clinical ms spectra taken from a cancer study. Thereby we focus on a reliable encoding and analysis of the spectral data as well as on the generation of a classification model indicating discriminative features.

3.1   Clinical Sample Preparation and MS Data Acquisition

Sample preparation and profile spectra analysis were carried out using the CLINPROT system (Bruker Daltonik, Bremen, Germany). Plasma samples from 45 cancer patients and 50 controls were prepared using the MB-WCX Kit (Bruker Daltonik, Bremen, Germany). Purifications were performed according to the product description. Sample preparation onto the Anchor Chip target was done using HCCA matrix. Profiling spectra were generated on a linear autoflex MALDI-TOF MS (Bruker Daltonik, Bremen, Germany).

3.2   Analysis with Supervised Neural Gas

The preprocessed set of spectra and the corresponding wavelet coefficients are now analyzed using the Supervised Neural Gas Algorithm (SNG) extended by a functional metric. We reconstructed the spectra based upon the discriminative wavelet coefficients determined by the Kolmogorov-Smirnoff test as explained above. In a second step maximal intensities for the ranges encoded by the wavelet coefficients has been determined. The spectra are subsequently considered only upon these maximal intensities leading to a significant reduced representation of the spectra as depicted in Figure 2 but still sufficient to keep all necessary information as evaluated by experts. This leads to a complexity reduction of the signals as well as to a smoothing of the spectral signal. Here we used the full spectra for parametrized functional norm i.e. all $\lambda_i = 1$. The original signal with approx. 22000 sampling points had been processed with only 600 remaining points still incoding the significant parts of the signal relevant for discrimination
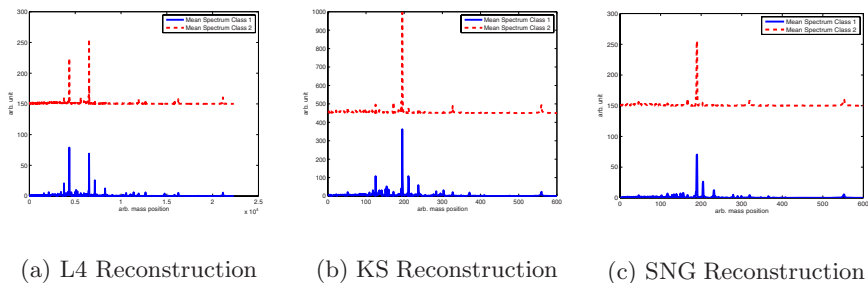
(a) L4 Reconstruction   (b) KS Reconstruction   (c) SNG Reconstruction

Fig. 2. Visualization of the different representations of the spectral data with respect to wavelet encoding and feature selection. The first figure shows a part of the original spectra for the two classes reconstructed with $L = 4$. The second figure shows the same part after encoding by wavelet analysis and KS-Test application as explained before one clearly observed that dominating common peaks are removed. The third figure shows the reduced representation of the spectra by linear interpolation between maximal intensities at encoding regions, as finally used for the modeling.

between the classes. The SNG classifier with functional metric obtains a cross-validation accuracy of 84% using functional metric and 82% by use of standard Euclidean metric, in addition the standard deviation of the different crossvalidation runs is smaller using the functional metric. Thereby the results with the wavelet processed spectra are slightly better than using standard peak lists, with 81% crossvalidation accuracy.

4   Conclusions

The presented initial data interpretation of proteom data demonstrate that the functional analysis and model generation using SNG with functional metric in combination with a wavelet based data preprocessing provides an easy and efficient detection of classification models and a good analysis of the high-dimensional data. The usage of wavelet encoded spectra features is especially helpful in detection of small differences which maybe easily ignored by standard approaches as well as to significantly reduce the number of points needed in further processing steps. Thereby the signal must not be reduced to peak lists but could be preserved in its functional representation. SNG was able to process high-dimensional functional data and observed good regularization. By use of the Kolmogorv-Smirnoff test we found a ranking of the features related to mass positions in the original spectrum which allows for identification of most relevant feature dimensions which was used for the reduction of the large number of wavelet coefficients and to prune irrelevant regions of the spectrum. In future analysis it could be interesting to reduce the feature set by incorporating knowledge about fractionation of the chemical compounds leadings to multiple related peaks of the same source. Alternatively one could optimize the scaling parameters of the functional norm directly during classification learning by so

called relevance learning as shown in [3] for scaled Euclidean metric. Conclusively, wavelet based spectra encoding in combination with SNG equipped by a functional metric is an interesting alternative to standard approaches allowing more flexibility in problem modeling as well as the control of the data processing task. It combines efficient model generation with automated data pretreatment and intuitive analysis.

References

1. Rieder, A., Louis, A.K., Maaß, P.: Wavelets: Theory and Applications. Wiley, Chichester (1998)
2. Cohen, A., Daubechies, I., Feauveau, J.-C.: Biorthogonal bases of compactly supported wavelets. Comm. Pure Appl. Math. 45(5), 485–560 (1992)
3. Hammer, B., Strickert, M., Villmann, T.: Supervised neural gas with general similarity measure. Neural Processing Letters 21(1), 21–44 (2005)
4. Ketterlinus, R., Hsieh, S-Y., Teng, S-H., Lee, H., Pusch, W.: Fishing for biomarkers: analyzing mass spectrometry data with the new clinprotools software. Bio. techniques 38(6), 37–40 (2005)
5. Kohonen, T.: Self-Organizing Maps. Springer Series in Information Sciences, vol. 30. Springer, Heidelberg (1997) (2nd Ext. Ed. 1997)
6. Lee, J., Verleysen, M.: Generalizations of the lp norm for time series and its application to self-organizing maps. In: Cottrell, M., (ed) 5th Workshop on Self-Organizing Maps, vol. 1, pp. 733–740 (2005)
7. Leung, A., Chau, F., Gao, J.: A review on applications of wavelet transform techniques in chemical analysis: 1989-1997. Chem. and Int. Lab. Sys. 43(1), 165–184 (1998)
8. Martinetz, T.M., Berkovich, S.G., Schulten, K.J.: 'Neural-gas' network for vector quantization and its application to time-series prediction. IEEE Trans. on Neural Networks 4(4), 558–569 (1993)
9. Sato, A., Yamada, K.: Generalized learning vector quantization. In: Touretzky, D.S., Mozer, M.C., Hasselmo, M.E. (eds.) Advances in Neural Information Processing Systems 8. Proceedings of the 1995 Conference, pp. 423–429. MIT Press, Cambridge, MA, USA (1996)
10. Villanueva, J., Philip, J., Entenberg, D., Chaparro, C.A., et al.: Serum peptide profiling by magnetic particle-assisted, automated sample processing and maldi-tof mass spectrometry. Anal. Chem. 76, 1560–1570 (2004)
11. Villmann, T., Hammer, B.: Supervised neural gas for learning vector quantization. In: Polani, D., Kim, J., Martinetz, T. (eds.) Proc. of the 5th German Workshop on Artificial Life (GWAL-5). Akademische Verlagsgesellschaft - infix, pp. 9–16. IOS Press, Berlin (2002)
12. Waagen, D.E., Cassabaum, M.L., Scott, C., Schmitt, H.A.: Exploring alternative wavelet base selection techniques with application to high resolution radar classification. In: Proc. of the 6th Int. Conf. on Inf. Fusion (ISIF'03), pp. 1078–1085. IEEE Press, New York (2003)