SOM-based Peptide Prototyping for Mass Spectrometry Peak Intensity Prediction

Alexandra Scherbart[1], Wiebke Timm[1,2], Sebastian Böcker[3] and Tim W. Nattkemper[1]
[1]Applied Neuroinformatics Group, Faculty of Technology, Bielefeld University
[2]Intl. NRW Graduate School of Bioinformatics and Genome Research, Bielefeld University
[3]Bioinformatics Group, Jena University
email: ascherba@techfak.uni-bielefeld.de

Keywords: Peak Intensity Prediction, Self-Organizing Map, Local Linear Map, Maldi-MS

***Abstract*— In todays bioinformatics, Mass spectrometry (MS) is the key technique for the identification of proteins. A prediction of spectrum peak intensities from pre computed molecular features would pave the way to better understanding of spectrometry data and improved spectrum evaluation. We propose a neural network architecture of Local Linear Map (LLM)-type based on Self-Organizing Maps (SOMs) for peptide prototyping and learning locally tuned regression functions for peak intensity prediction in MALDI-TOF mass spectra. We obtain results comparable to those obtained by ν-Support Vector Regression and show how the SOM learning architecture provides a basis for peptide feature profiling and visualisation.

1 Introduction

In todays bioinformatics, Mass spectrometry (MS) is the key technique for the identification of proteins. Matrix-assisted laser desorption ionization (MALDI) is one of the most often used technique for the analysis of whole cell proteomes in high-throughput experiments. There are different applications of MALDI-MS where the prediction of peak heights (referred to as intensities) in the spectrum are needed for further improvements: Protein identification is commonly done by comparing the peak's masses from a spectrum – the so called protein mass finger print (PMF) – to theoretical PMFs in a data base, generating a score for each comparison. Different tools are available for this purpose. For an overview see [SCB05]. These tools rarely use peak intensities, because there is no model to calculate the theoretical PMFs directly. The use of peak intensities could improve the reliability of protein identification without lowering the error rate, as was shown by Elias *et al.* for tandem MS[EGK$^+$04].
Another application of MALDI where peak intensities are important is quantitative proteomics, where proteins in a complex sample are quantified or protein abundances across different samples are compared.

For the prediction of MALDI PMF there has been one study so far by Gay *et al.* who applied different regression and classification algorithms[GBHA02]. Tang *et al.* used multi-layer neural networks to predict peptide detectabilities (i.e. the frequency with which peaks occur in spectra) in LC/MS ion trap spectra[TAA$^+$06] which is a related problem.

An algorithmic approach for peak intensity prediction is a non-trivial task because of several obstacles: The extraction of PMF from spectra is a signal processing task which can not be done perfectly. Data from this domain is always very noisy and contains errors introduced by preprocessing steps in the wet lab as well as in signal processing. Misidentifications may even lead to wrong sequences. Intensity values can be distorted due to the unknown scale of spectra. It is nearly impossible to come by a large enough data set from real proteins where the content is known, i.e. there is no perfect gold standard, because of the not reproducible and non-unique peptide/intensity relation.

To overcome these obstacles to predict peak intensities in MALDI-TOF spectra based on a pre-selected training set of peptide/peak intensity pairs, a method is needed, that is able to (a) determine peptide profiles and (b) learn locally tuned regression functions for peak intensity prediction. For this purpose we consider a artificial neural net architecture of Local Linear Map(LLM)-type, since it combines unsupervised (a) and supervised (b) learning principles based on (Multi-) Self-Organizing Maps (Multi-SOMs) [GS06].

The LLM-architecture is well suited for this task due to its transparency. It is simple to implement, can cope with large data sets, is easy adaptable to new data by a slight deviation of the parameters without loss of information. Other than for example support vector regression (SVR) it can be used for data mining once adapted in a straight forward manner, as demonstrated in this work. We propose a combination of unsupervised and supervised learning architecture with comparable results in predicting the peaks' intensities to ν-Support Vector Regression (SVR).The mixture of linear experts derives implicit models for characterizing peptides and feature analysis as an unsupervised learning task. The second step consists of supervised adaptation of the neural network and prediction of peaks' intensities.

Proceedings of the 6th International Workshop on Self-Organizing Maps (WSOM 2007)
Published by the Neuroinformatics Group, Bielefeld University, Germany, ISBN 978-3-00-022473-7
All contributions to WSOM 2007 are available online at: http://biecoll.ub.uni-bielefeld.de

2 Materials and Methods

2.1 Data

In this study we use two datasets $\mathbf{A}$ and $\mathbf{B}$ of peptides of MALDI mass spectra. The first one,$\mathbf{A}$, consists of 66 spectra of 29 different proteins, with 16 of these proteins being present in multiple spectra, whereas dataset $\mathbf{B}$ consists of 200 spectra of 137 different proteins with 39 of these proteins occurring multiple times.
Peak extraction steps include soft filtering, baseline correction, peak picking and isotopic deconvolution in the corresponding raw spectra. The resulting list of peaks is matched against masses derived from a theoretical tryptic digestion. These steps for $\mathbf{A}$ (and $\mathbf{B}$ respectively) result in 857 (1631) matched peaks corresponding to 415 (1135) different peptides.
For preprocessing, normalization of the intensities is necessary, because spectra do not have the same scale. For a MALDI spectrum the exact amount of protein sample that leads to it is not known, nor is it possible to scale spectra belonging to the same protein by the same amount. There is no peptide that connects the scales of spectra.
The normalized intensities are therefore computed following two different heuristics, in order to use values from different spectra together. In the remainder, the intensities refer to scaling the original matched peaks intensity I_{orig} after denoising and baseline correction by the sum of all $i = 1, \ldots, N$ values in the whole spectrum yielding I_S:

$$I_S = \frac{1000 \cdot I_i^{\text{orig}}}{\sum_{i=1}^{N} I_i^{\text{orig}}} \quad \text{with} \quad I_i^{\text{orig}} = I_i - B_i - N_i.$$

Subsequently, the natural logarithm of the intensities is applied to compute the final intensity output values.

2.2 Feature Sets

One of the most important questions in conjunction with finding a model for predicting peak intensities is the representation of the peptides. A suitable feature space is the precondition for success of any machine learning method.

We combine different properties of peptides to represent features of a peptide. Amino acid frequencies, typically used in bioinformatics, in conjunction with chemical features of the peptides are used to create the heuristically selected 18-dimensional feature space is built by different types of characterization we assume to be relevant for MALDI ionization and additional features that are chosen in an ad hoc feature forward selection.
Most of the peptides in the data set occur multiple times in different spectra with different intensity values. To eliminate outliers (potential noisy peptides) and to map each peptide to one unique value, the α-trimmed mean of all intensities per distinct peptide with $\alpha = 50\%$ is computed. It is defined as the mean of the center 50% of an ordered list. In the case of less than 4 peptides in the list a simple mean is taken. In the remainder, we refer to data points from $\mathbf{A}$ and $\mathbf{B}$ in the heuristically selected feature space with the α-trimmed mean calculated target values as datasets $\mathbf{A_{TM}}$ and $\mathbf{B_{TM}}$.

2.3 Local Linear Map

The task of mass spectrometry prediction and peptide prototyping corresponds to the task of unsupervised clustering as well as classification and supervised prediction. The problem can be stated as follows:
Given a training set $\Gamma = \{(\mathbf{x}, \mathbf{y})_i, i = 1, \ldots, N\}$, consisting of input-output pairs: peptide patterns $\mathbf{x}_i$ which are elements of feature space $\mathcal{X} = \mathbb{R}^t$, and real-valued outputs, i.e. intensities, $\mathbf{y}_i \in \mathbb{R}$. One promising approach would be to find a set of clusters and prototypes representing the data points best according to the statistical properties of the data provided. After assigning every input point to a prototype, a prediction of a real-valued output Y has to be done.

For determining peptide prototypes and learning into the mapping the output, i.e. intensity space, we propose to use a SOM variant of the Local Linear Map (LLM)-architecture. The LLM combines unsupervised vector quantisation algorithm for computing a voronoi tessellation of the input space $\mathcal{X}$ with supervised techniques for feature classification.

The artificial neural net (ANN) of Local Linear Map-type [Rit91] was indeed originally motivated by the Self-Organizing Map by Kohonen [Koh82] and has been shown to be a valuable tool for the fast learning of non-linear mappings $\mathcal{C} : \mathbb{R}^{d_{in}} \rightarrow \mathbb{R}^{d_{out}}$.
A LLM consists of n_l nodes $\mathbf{v}_i, i = 1, \ldots, n_l$. Each node consists of a triple $\mathbf{v}_i = (\mathbf{w}_i^{in}, \mathbf{w}_i^{out}, \mathbf{A}_i)$. The vectors $\mathbf{w}_i^{in} \in \mathbb{R}^{d_{in}}$ are used to build prototype vectors adapting to the statistical properties of the input data $\mathbf{x}_\xi \in \mathbb{R}^{d_{in}}$ provided. The vectors $\mathbf{w}_i^{out} \in \mathbb{R}^{d_{out}}$ approximate the distribution of the target values $\mathbf{y}_\xi \in \mathbb{R}^{d_{out}}$. The matrices $\mathbf{A}_i \in \mathbb{R}^{d_{in} \times d_{out}}$ are locally trained mappings from the input to the output space.

Vector quantisation by Self-Organizing Maps For the task of unsupervised training in the input space we apply SOMs and combine it with the LLM into the output space. In addition, we explored an extension of recently proposed Multi-Self-Organizing Maps (Multi-SOMs) as a set of several neural networks, each of them accounting for certain input space data structures.
The partner SOMs in a Multi-SOM are not necessarily identical, but can differ in size, dimension and topology. For our purpose we use K identical 2-dimensional partner SOMs, denoted as $K - n \times m$ SOMs.
There are two extreme configurations of Multi-SOMs demonstrating the range of paradigms that can be realized with an M-SOM: With $K = 1$, the M-SOM consists of just a single, classical SOM. The other extreme situation, with K partner SOMs, having minimal size $n = m = 1$, the M-SOM performs a K-Means-algorithm.

Proceedings of the 6th International Workshop on Self-Organizing Maps (WSOM 2007)
Published by the Neuroinformatics Group, Bielefeld University, Germany, ISBN 978-3-00-022473-7
All contributions to WSOM 2007 are available online at: http://biecoll.ub.uni-bielefeld.de

In the unsupervised training phase the $n \cdot m$ prototype vectors of the winning SOM are adapted following SOM learning rule: The neurons are adapted according to their distance between the input pattern $\mathbf{x}_\xi$ and winning prototype $\mathbf{v}_\kappa$. The learning procedure changes the weights according to the gaussian neighborhood function h_σ. After adapting the prototypes, each of the input vectors $\mathbf{x}$ can be associated with its closest prototype as a winner-takes-all (WTA) rule: $\mathbf{w}_\kappa = \arg\min_{\mathbf{w}} \left\{ \|\mathbf{x} - \mathbf{w}_i^{in}\| \right\}$. We also applied Neural Gas clustering [MBS93] instead of SOM leaving the input space for comparison.

Training of local mappings from input to the output space is performed in the second step. Subsequently to unsupervised adaptation and tessellation of the input space $\mathcal{X}$, a local expert is assigned to each of the $K \cdot n \cdot m$ voronoi cells. The mapping of an arbitrary input vector $\mathbf{x}$ to an output $\mathcal{C}(X)$ is computed by

$$\mathcal{C}(\mathbf{x}) = \mathbf{w}_\kappa^{out} + \mathbf{A}_\kappa \left(\mathbf{x} - \mathbf{w}_\kappa^{in} \right)$$

by the corresponding local expert $\mathbf{w}_\kappa$. The weights $\mathbf{w}_i^{out}$ and the linear map $\mathbf{A}_i$ are changed iteratively applying the learning rules[1]:

$$\begin{aligned} \Delta \mathbf{w}_i^{out} &= \epsilon^{out} \cdot h_\sigma \cdot (\mathbf{y}_\xi - \mathcal{C}(\mathbf{x}_\xi)) \,, \\ \Delta \mathbf{A}_i &= \epsilon^{\mathbf{A}} \cdot h_\sigma \cdot (\mathbf{y}_\xi - \mathcal{C}(\mathbf{x}_\xi)) \cdot \frac{(\mathbf{x}_\xi - \mathbf{w}_i^{in})^T}{\left\| \mathbf{x}_\xi - \mathbf{w}_i^{in} \right\|^2}. \end{aligned}$$

2.4 Evaluation

About 10% of the centered and normalized data are used for validation and put aside. The remaining dataset is used to train the LLM and to find the best parameter set using 10-fold Cross-Validation (CV). So, the remaining dataset is split into 10 portions and one set is used for testing performance of the selected model. It was ensured that peptides from one spectrum as well as peptides occurring in more than one spectrum are found in only one of the portions.

Model selection: Grid search over the parameter space $P = \left(K, n, m, \epsilon^A\right)$ is performed to determine optimal parameters for learning. The remaining learning parameters for the LLM were set to initial values $\epsilon^{out} = 0.3, \epsilon^{in} = 0.5$ and $\sigma = 2$ decreasing exponentially over time to final values $\epsilon^{out} = 0.01, \epsilon^{in} = 0.01$ and $\sigma = 0.4$. A 10-fold-CV is done for each parameter set. For every point in the parameter space the prediction accuracy for every training/test set is determined by squared pearson-correlation coefficient r^2 and root mean square error RMSE of the test set. The choice of the best parameter set is made by the best mean r^2 over all 10 test sets while training the learning algorithm.

Model assessment: The final model with the optimal parameters is chosen. To validate its prediction (generalization) error on new data, the validation set is used, which has not taken part in training.

[1] It has to be stated that for Multi-SOMs with $K > 1$, the local experts are adapted following Neural-Gas (NG) learning rule replacing $h_\sigma = h_\sigma\left(r_\kappa(\mathbf{x}, l)\right)$, because the grid structure can no longer be hold up.

3 Results

We compare the prediction performance for the two datasets $\mathbf{A_{TM}}$ and $\mathbf{B_{TM}}$ containing peptides mapped by α-trimmed mean (2.2). The following results are evaluated with respect to the squared pearson correlation r^2 and $RMSE$ for the 10 test sets and the validation set. The validation set of $\mathbf{A_{TM}}$ consists of 44 items and for $\mathbf{B_{TM}}$ of 112 items.

3.1 Peptide Prototyping

A display of the prototype vectors resulting from the Self-Organizing Map training allows a profiling of peptides. In the following Fig. 1 the resulting parallel coordinates plot for six prototypes in case of $\mathbf{B_{TM}}$ is shown.

The correlation of input space reflects in the prototype distribution. We can see that some of the features ('GB500','Y') show a correlation to the mass, while no such tendency can be observed for other features. The six prototypes take three to five levels for each feature, two or more prototypes sharing the same region. 'OOBM850104' (measure of non-bonded energy), 'ROBB760107' (information measure) and 'M' (no. of methionine) show the least similarity to any other feature.

If we look closer at the prototypes, it can be seen that there are existing pairs covering outmost areas in data space in almost all features (prototypes 1 and 6 except for 'H') as well as prototypes covering contrary regions in data space (e.g. prototypes 1 and 6, 3 and 5,2 and 4). Another thing to be noted is that prototypes 1 and 3 are near to each other for almost all features, except for 'ARGP820102' and 'F' , where they split up to almost the extremes. Similar behavior can be observed for the prototypes 5 and 6.

'ROBB760107'(information measure), 'FINA770101' and 'KHAG800101' (kerr constant increment) show the most even distribution of prototypes. Thus the prototypes share ranges in certain features and split up for others, achieving a nice spread in data space: For a data point that is similar to two prototypes sharing their space for a set of features, other features decide which prototype it is assigned to.

3.2 Predicting Peak Intensities

We compare the prediction capabilities of the LLM for the two data sets $\mathbf{A_{TM}}$ and $\mathbf{B_{TM}}$. The evaluation is done as described in 2.4. We perform a grid search over all parameters and the parameter set is chosen yielding the best mean r^2 of training/test sets. For the exact results of r^2 and RMSE see Tab. 1. A scatter plot with target vs. predicted values for dataset $\mathbf{B_{TM}}$ predicted by a $N = 2 \times 3$-SOM-LLM is shown in Fig. 2.

Proceedings of the 6th International Workshop on Self-Organizing Maps (WSOM 2007)
Published by the Neuroinformatics Group, Bielefeld University, Germany, ISBN 978-3-00-022473-7
All contributions to WSOM 2007 are available online at: http://biecoll.ub.uni-bielefeld.de

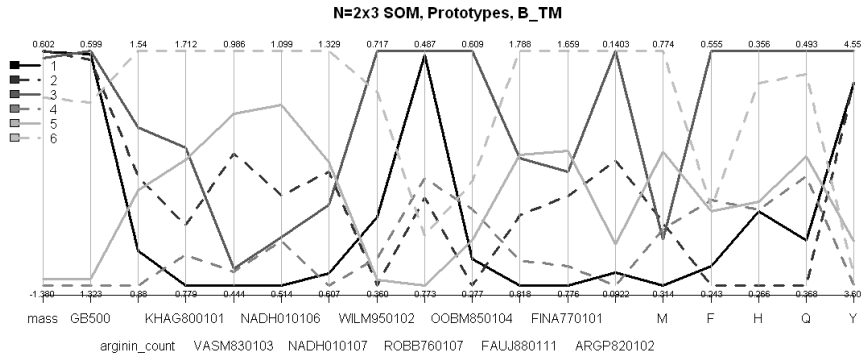

Figure 1: Parallel coordinates plot for six prototypes in case of $\mathbf{B_{TM}}$. For every feature the range of values covered by the prototypes is shown.

3.3 Comparison prediction performance of SOM to Neural Gas

For a comparison of the prediction and generalization performance of SOM and Neural Gas, the evaluation of the studied datasets $\mathbf{A_{TM}}$ and $\mathbf{B_{TM}}$ are shown. Due to the the difference of topology of these learning methods, the SOMs topology was kept fixed as $N = n \times 2$, $(K = 1)$, and its performance compared to the one of NG with N neurons.
In Fig. 3 an iteration over $N = 1, \ldots, 18$ is done. The figure (a) shows the corresponding correlations (r^2) for the 10 test sets as well for the validation set.

4 Discussion

Our results show that the SOM-LLM-approach combining data mining and supervised learning yields similar results in prediction accuracy to our first approach utilizing ν-SVR [TBTN06] and Neural Gas.

From our results it is clear that peak intensities can be characterized and predicted by the use of the heuristically selected feature set with high prediction accuracy. The visual inspection of the prototypes reveals that the peptides can be grouped around a set of approximately 6 profiles. Those seem to have individual mappings to peak intensity which can be discussed with biochemical experts.

In Fig. 3 the results of the prediction accuracy determined by r^2 and RMSE with SOMs and Neural Gas for data set $\mathbf{A_{TM}}$ are summed up. Both SOM and NG yield a similar behavior with respect to prediction performance.

Table 1: Comparison of capabilities of the LLM in predicting intensities for the studied datasets $\mathbf{A_{TM}}$ and $\mathbf{B_{TM}}$. The evaluation was done for N prototypes in case of SOM and NG.

	Test		*Valid*	
Dataset	r^2	RMSE	r^2	RMSE
$\mathbf{A_{TM}, N = 2 \times 2}$	**0.438**	1.126	0.260	1.356
$\mathbf{A_{TM}, N = 4}$	0.416	1.158	0.117	1.765
$\mathbf{A_{TM}, \nu - SVR}$	0.399	1.03	0.311	1.140
$\mathbf{B_{TM}, N = 2 \times 3}$	0.252	1.119	**0.403**	0.954
$\mathbf{B_{TM}, N = 6}$	0.251	1.114	0.355	0.993
$\mathbf{B_{TM}, \nu - SVR}$	**0.292**	1.082	0.381	0.973

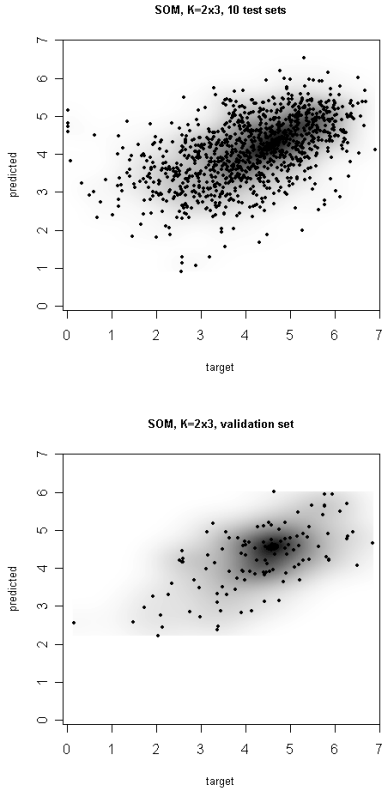

Figure 2: Scatterplot with target vs. predicted values for dataset $\mathbf{B_{TM}}$ predicted by a $N = 2 \times 3$-SOM-LLM. The upper plot shows the resulting predicted values of the 10 test sets, the lower shows the predicted values of the validation set.

Proceedings of the 6th International Workshop on Self-Organizing Maps (WSOM 2007)
Published by the Neuroinformatics Group, Bielefeld University, Germany, ISBN 978-3-00-022473-7
All contributions to WSOM 2007 are available online at: http://biecoll.ub.uni-bielefeld.de

As the number of neurons increases ($N > 4$, where the optimum is reached), the clustering error for both learning paradigms decreases, while the prediction error and correlation for the 10 test sets also decrease and increase for the validation set. Furthermore, it can be observed that though a worse clustering error of SOM-learning, the prediction error for the 10 test sets as well as for validation is smaller than that of NG-learning and yields better prediction performance. The number of neurons is a critical size due to overfitting of training data. In Fig. 4 the results of the prediction accuracy for different number of prototypes are summed up. The mean performance of the 10 test sets is compared to the performance of the validation set. The fact that some test sets performance is worse (especially for dataset $\mathbf{A_{TM}}$) than the performance of the chosen validation set, can be explained by the static choice of the set. The different portions of $\mathbf{A_{TM}}$ set yield a wide spread of correlation, resulting in high standard deviation of r^2 over all the portions. There are test sets that seem significantly worse in prediction performance over all training sets. There exists a positive correlation to the number of test set examples. The studied datasets differ in their prediction performance and can be separated well in Fig. 4(a) and (b). For dataset $\mathbf{A_{TM}}$ a wide spread in test and validation performance can be observed for NG-LLM as well as SOM-LLM, whereas for using SOM architecture in both cases better results in correlation and prediction error are achieved. Future investigation will investigate how to immprove the NG based results. From the evaluation and experiments of the underlying heuristically selected feature space can be assumed to very compact with very sparse located outliers off the center. A strong hint for this assertion is the prediction performance in case of only 1 neuron (see.Fig. 3).

There are two conflicting targets that have to be reached: First is minimization of the clustering error, and second the minimization of prediction error yielding maximization of prediction performance. The trade-off in generalization performance can be observed between adaptation to the data and training of local mappings from input to output depending on size and topology for both LLM-learning algorithms.

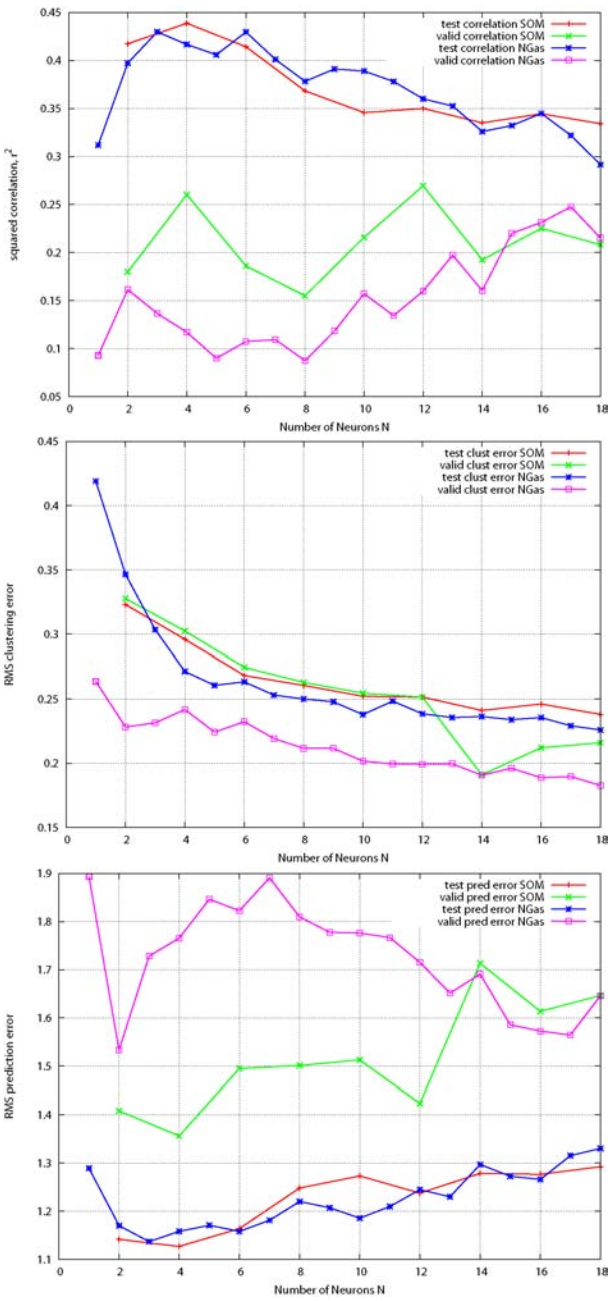

Figure 3: Results of the prediction accuracy for $\mathbf{A_{TM}}$. Iteration is done over $N = 1, \ldots, 18$ number of neurons, with $N = 2 \times n$ (SOM),$K = 1$, and N (NGas) neurons respectively. (a) For every evaluation the results of the mean performance of the test sets is plotted as well as the performance of the validation set. The best mean test correlation is yielded by a 2×2-SOM. While the prediction accuracy for the 10 test sets decreases, the prediction accuracy for validation set increases proportional to the number of neurons. (b) It can be stated that the clustering error for SOM tends to be worse than for NGas. (c) The prediction error increases for SOM and NG in case of test sets to the same degree, whereas the prediction error of the validation set for the SOM is much smaller than that for NG.

5 Conclusions

We propose an algorithmic approach for peak intensity prediction in MALDI-TOF spectra. The proposed model for peptide prototyping and prediction of peak intensities with the architecture of Local Linear Map-type has been shown to be a valuable neural network tool for these tasks combining unsupervised and supervised learning architecture. The LLM includes determining peptide profiles in the data set and the mixture of linear expert are able to learn locally tuned regression functions for peak intensity prediction. The heuristically selected feature space is a good choice as the characteristics of peptides are reflected. Some fea-

Proceedings of the 6th International Workshop on Self-Organizing Maps (WSOM 2007)
Published by the Neuroinformatics Group, Bielefeld University, Germany, ISBN 978-3-00-022473-7
All contributions to WSOM 2007 are available online at: http://biecoll.ub.uni-bielefeld.de

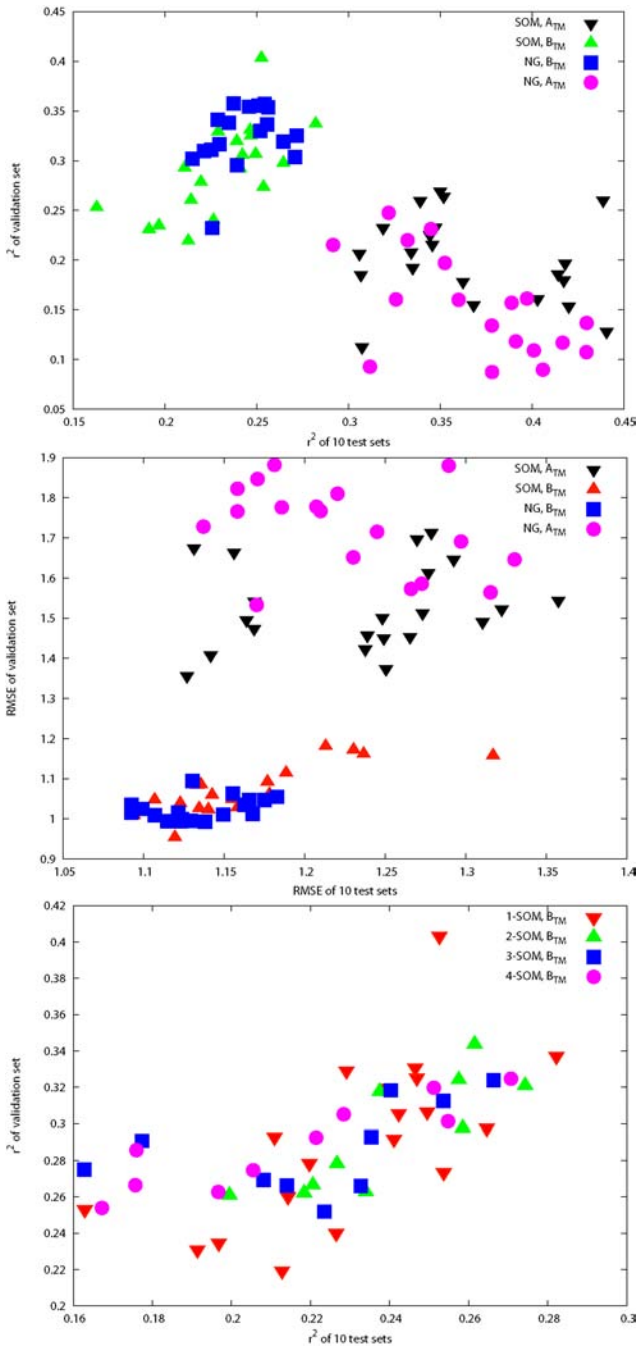

Figure 4: Results of the prediction accuracy measured by (a) r^2 and (b)RMSE for the evaluated data sets $\mathbf{A_{TM}}$ and $\mathbf{B_{TM}}$ with SOM as well as NGas. For every evaluation the results of the mean performance of the test sets is plotted against the performance of the validation set. The prediction accuracy of the two datasets can be separated well, where $\mathbf{B_{TM}}$ yields (a) a lower test correlation, but higher validation correlation, whereas for $\mathbf{A_{TM}}$ is the opposite. (b) $\mathbf{A_{TM}}$ yielding higher error compared to $\mathbf{B_{TM}}$ and a wide spread of error. (c) For this plot a discrimination according to the number of M-SOMs used is done. The best results are found for $K = 1$-SOM with $N = 2 \times 3$ neurons.

tures do not contribute to the assignment of data points to one of the prototypes. If this is due to the number of other features slightly correlated to each other or if they really carry no information with respect to the target values has to be the subject of further studies. The experiments with the considered data set have demonstrated the capabilities of the SOM-LLM approach in direct comparison to NG-LLM as well as ν-SVR.

References

[EGK+04] J. E Elias, F. D Gibbons, O. D King, F. P Roth, and S. P Gygi. Intensity-based protein identification by machine learning from a library of tandem mass spectra. *Nat Biotechnol*, 22(2):214–219, Feb 2004.

[GBHA02] S. Gay, P.-A. Binz, D. F Hochstrasser, and R. D Appel. Peptide mass fingerprinting peak intensity prediction: extracting knowledge from spectra. *Proteomics*, 2(10):1374–1391, Oct 2002.

[GS06] N. Goerke and A. Scherbart. Classification using multi-soms and multi-neural gas. In *Proc. of IEEE World Congress on Computational Intelligence*, 2006.

[Koh82] T. Kohonen. Self-organized formation of topologically correct feature maps. In *Biological Cybernetics*, volume 43, pages 59–69, 1982.

[MBS93] T. M. Martinetz, S. G. Berkovich, and K. J. Schulten. 'neural gas' network for vector quantization and its application to time-series prediction. In *IEEE Trans. Neural Networks*, volume 4, pages 558–569, 1993.

[Rit91] Helge Ritter. Learning with the self-organizing map. In T. Kohonen et al., editor, *Artificial Neural Networks*, pages 379–384, Amsterdam, 1991. Elsevier Science Publishers.

[SCB05] I. Shadforth, D. Crowther, and C. Bessant. Protein and peptide identification algorithms using MS for use in high-throughput, automated pipelines. *Proteomics*, 5(16):4082–4095, Nov 2005.

[TAA+06] H. Tang, R. J Arnold, P. Alves, Z. Xun, D. E Clemmer, M. V Novotny, J. P Reilly, and P. Radivojac. A computational approach toward label-free protein quantification using predicted peptide detectability. *Bioinformatics*, 22(14):e481–e488, Jul 2006.

[TBTN06] W. Timm, S. Böcker, T. Twellmann, and T. W. Nattkemper. Peak intensity prediction for pmf mass spectra using support vector regression. In *Proc. of the 7th International FLINS Conference on Applied Artificial Intelligence*, 2006.