s c c h
software competence center
hagenberg

FUZZY.LOGIC.LAB.LINZ

Advances in Knowledge-Based Technologies

Proceedings of the Master and PhD Seminar Winter term 2008/09, part 1

Softwarepark Hagenberg
SCCH, Room 0/2
November 26th, 2008

Software Competence Center Hagenberg
Softwarepark 21
A-4232 Hagenberg
Tel. +43 7236 3343 800
Fax +43 7236 3343 888
www.scch.at

Fuzzy Logic Laboratorium Linz
Softwarepark 21
A-4232 Hagenberg
Tel. +43 7236 3343 431
Fax +43 7236 3343 434
www.flll.jku.at

Program

9:00–10:15 Session 1 ***(Chair: Bernhard Moser)***

9:00 Holger Schöner:
Transfer Learning – An Overview of "Lifelong Learning"

9:45 Frank Bauer:
Parallel MRI – Quality and Speed, a contradiction?

10:15 Coffee Break

10:30–11:30 Session 2 ***(Chair: Roland Richter)***

10:30 Henrike Stephani:
Hierarchical Clustering of Filtered and Unfiltered Terahertz Spectra

11:00 Bettina Heise:
Differential Phase Contrast OCT: Measurements and Object Reconstruction

Transfer Learning:
An Overview of "Lifelong Learning"

Holger Schöner
Software Competence Center Hagenberg
email *holger.schoener@scch.at*

Joint PhD-Seminar by FLLL and SCCH on November 26, 2008

Abstract: *Transfer Learning as a research area in the field of Machine Learning developed mainly during the middle of the 90s [4, 8] and gained momentum again a few years ago [1–3, 5–7]. It is based on the idea, that it is not advantageous to start learning each new problem from scratch, but that it should help to use knowledge about similar tasks, which one has already learned. As is the case with human learning, obviously. The goals pursued are usually better generalization, need for less data, a better understanding of connections between different learning problems, and a more time and resource efficient learning. In my presentation I will provide an introduction to the topic, and present some selected methods for performing transfer learning.*

Keywords: *transfer learning, inductive bias*

References

[1] Rie Kubota Ando and Tong Zhang. A framework for learning predictive structures from multiple tasks and unlabeled data. *Journal of Machine Learning Research*, 6: 1817–1853, 2005. URL `http://portal.acm.org/citation.cfm?id=1194905`.

[2] Jonathan Baxter. A model of inductive bias learning. *Journal of Artificial Intelligence Research*, 12:149–198, 2000. doi: 10.1.1.37.8738. URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.37.8738`.

[3] Steffen Bickel, Jasmina Bogojeska, Thomas Lengauer, and Tobias Scheffer. Multi-task learning for hiv therapy screening. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2008.

[4] Rich Caruana. Multitask learning. *Machine Learning*, 28(1):41–75, 1997. URL `http://portal.acm.org/citation.cfm?id=262872\&dl=`.

[5] J. Madrid-Sanchez, E. Parrado-Hernandez, and A. Figueiral-Vidal. Selective multitask learning by coupling common and private representations. Learning from Multiple Sources Workshop, December 2008.

[6] David Pardoe and Peter Stone. Boosting approaches for regression transfer. In *Unpublished (NIPS 2008)*, Learning from Multiple Sources Workshop, December 2008.

[7] Jaakko Peltonen, Yusuf Yaslan, and Samuel Kaski. Variational bayes learning from relevant tasks only. Learning from Multiple Sources Workshop, December 2008.

[8] Sebastian Thrun. Is learning the n-th thing any easier than learning the first. *In Advances in Neural Information Processing Systems (NIPS)*, 8:640–646, 1996. doi: 10.1.1.44.2898. URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.44.2898`.

Parallel MRI – Quality and Speed, a contradiction? –

Frank Bauer
Fuzzy Logic Laboratorium Linz-Hagenberg
Johannes Kepler University Linz
Softwarepark 21
4232 Hagenberg
AUSTRIA
e-mail: frank.bauer@jku.at

Magnetic resonance imaging is a medical imaging technique which is very flexible and can be used to measure a high number of different properties of human tissue. In contrast to other imaging techniques it does not use ionizing and hence potentially dangerous radiation. Applications are:

- MRI: The classical imaging
- Diffusion MRI: The anisotropy of tissue (e.g. in the brain) due to varying diffusion coefficients in different directions is measured
- Functional MRI: Measuring the brain activity
- Voxelwise Molecule Spectra
- MREIT: The electrical properties of tissue can be measured
- Elastography: The elastic properties can be measured

All of these require the fast acquisition of MRI data. One possibility to bypass the physical boundaries is the usage of parallel MRI, which yields an instable bi-linear inverse problem. In this talk we will present different possibilities to solve this problem as well as the inherent problems the method poses itself:

- Regularized Gauss-Newton Regularization
- Landweber-Kaczmarz with various image enhancing steps
- Landweber Regularization with Energy correction.

Corresponding reconstructions from real data will be shown. Depending on the time just the last reconstruction idea will be presented.

Hierarchical Clustering of Filtered and Unfiltered Terahertz Spectra

Wavelet Shrinkage and Savitzky-Golay Filtering in Raw Data Euclidean Distance Clustering of Spectra with Broad Bandwidth Features

Henrike Stephani
Fraunhofer Institute for Industrial Mathematics (ITWM)
Fraunhofer Platz 1
67663 Kaiserslautern
Henrike.Stephani@itwm.fraunhofer.de

Abstract

With the goal to perform Clustering on spectral data, two standard Chemometrics filtering techniques are performed. The test data consists of Terahertz measurements of different chemical compounds. The filtering techniques are Savitzky-Golay filters and Wavelet shrinkage. They are applied as a preprocessing step to raw data based hierarchical clustering. The shrinkage as well as the Savitzky-Golay filter are performed with different parameters depending on the frequency interval of the spectrum. The intervals are found by a variance analysis. The distance measure is Euclidean distance, and the clustering algorithm is agglomerative hierarchical average link clustering. The performance of the clustering is significantly improved by filtering the data. The cluster dendrogram shows clearer separability and the cluster distances are more stable. Both filters yield similar results though Wavelet shrinkage seems to have a slightly better ability to deal with noise.

1 Introduction

With the advancement of Terahertz detectors and emitters, Terahertz Spectroscopy is becoming a valuable method for different commercial applications. One of the most interesting one is non-invasive testing. Pharmaceutical compounds have characteristically shaped absorption spectra as well as explosives do. At the same time usual packaging materials such as ceramics, paper, and carton as well as most cloth are non absorbent. That is why it is possible to detect or examine materials within [12]. Therefore, Terahertz measurements are especially valuable in pharmaceutical quality control as well as drug and explosives detection. The analysis of Terahertz spectra from the above describe angle belongs to the category of Chemometrics, an interdisciplinary research area that is increasing constantly [8], [9]. Our special interest here shall lie in testing different filtering methods on the spectra with the final goal to perform raw data clustering. Although the amount of different Terahertz-Spectra still is limited it increases continually. Therefore, the challenge of categorizing those spectra has to be faced.

In raw data clustering as well as in feature based clustering preprocessing in form of filtering is necessary due to noise occurring during the detection. As can be seen in figure 1, in comparison with infrared spectroscopy the characteristic features in Terahertz spectra are quite different. While in the former the peaks are sharp and isolated and are directly related to the ingredients of the measured component in the later they are broad and overlapping. This is caused by the different molecular excitements in Terahertz and infrared spectroscopy [4]. Filtering, therefore, has to consider different necessities. One wants to preserve features that differentiate compounds and suppress noise. There is a variety of possibilities to determine fingerprint regions in spectroscopy but most of these methods aim at fingerprinting in mass spectroscopy or infrared spectroscopy. We shall now

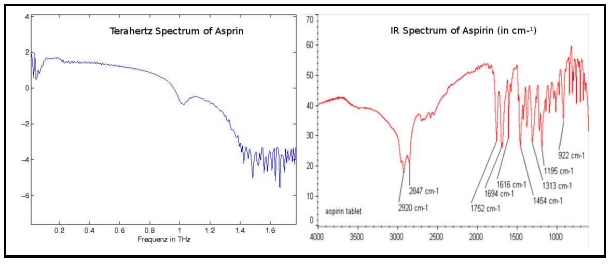

Figure 1: Terahertz and IR Spectrum. While in the IR spectrum the peaks stand for different components in the sample the shape of the THz spectrum is broad and related to the overall structure of the sample.

test two methods, namely Wavelet shrinkage and Savitzky-Golay filters. In addition to that a variance analysis of each spectrum is performed to find the respective interval containing its relevant characteristics.

2 Methods

2.1 Filtering

In Chemometrics one of the most popular filters is the Savitzky-Golay filter. Given a window of points $x_{-m}, ..., x_0, ..., x_m$ one wants to calculate the optimal fit of a polynomial f of degree k to these points. Optimality is defined by minimizing the mean squared distance.
Instead of interpolating a new polynomial in each new point, Savitzky and Golay proved that it is possible to determine the values of the respective coefficients beforehand. These coefficients are independent from the measured data. They depend only on the size of the filtering window, and the degree of the polynomial. This is due to the fact, that one isn't interested in the actual polynomial but rather only wants to know its value in the center point x_0 [3]. The advantage of this approach is that peaks are generally well preserved, where normal mean filters lead to a broadening and flattening. But at the same time it is as easily and quickly computed as a normal mean filter is.

Another approach is using wavelet shrinkage[1]. The main idea of Wavelet transformation is a loss free hierarchical decomposition of a signal based on a basis of so called Wavelets which can be considered band pass filters with certain desirable properties. The main advantage over the Fourier transform is the possibility to express characteristics with respect to as well frequency as time. In that way it is possible to control the degree of influence certain frequencies have on the signal depending on the interval they occur in. Although the amount of information one gets is quite sophisticated the discrete Wavelet transformation is computationally very efficient thanks to a down-sampling mechanism that is applied during the procedure. The computational complexity of calculating the wavelet coefficients of n data points is only $O(n)$ (in contrast to the Fast Fourier Transform with $O((n)log(n))$) [11].

2.2 Clustering

There is a broad variety of clustering algorithms applied in many different areas. Usually they are divided in at least two groups: partitional and hierarchical algorithms. While in partitional clustering the data is distributed to classes on one level, in hierarchical approaches the clustering is carried out iteratively. Both methods hold advantages and disadvantages. Partitional algorithms are often faster than hierarchical ones and can revise decision once made. Hierarchical algorithms, on the other hand, provide better interpretability. The optimal number of clusters doesn't have to be known beforehand and neither have initial cluster centers or starting values. Another important advantage is that the algorithms do not only produce clusters but also information about the distances between the classes (for more thorough information on clustering algorithms see [6]). We shall focus on hierarchical clustering because of the above mentioned advantages.

The procedure of classical agglomerative clustering operates in the following way: Beginning with one sample per cluster in each iteration the closest clusters are being united. Assuming a given distance between the samples the closeness of the cluster is defined by the link function. This function was generalized by Lance and Williams [7]. When merging the clusters C_j and C_k the distance of the resulting cluster C_{jk} to another cluster C_l will be:

$$D(C_{jk}, C_l) = \alpha_j D(C_j, C_l) + \alpha_k D(C_k, C_l) +$$
$$\beta D(C_j, C_k) + \gamma |D(C_j, C_l) - D(C_k, C_l)|.$$

In case of $\{\alpha_j, \alpha_k, \beta, \gamma\} = \{1/2, 1/2, 0, -1/2\}$ this is the minimum of the individual distances. The resulting clustering is called single link clustering and is one of the most frequently used methods. A disadvantage of this approach is, that in case of a relatively high variance within one cluster and a relatively low distance between clusters the probability of misclassification is high. Especially in case of the assumed existence of a prototype for the cluster using the average link distance yields better results. Here the coefficients are set to $\{\frac{\overline{C_j}}{\overline{C_j + C_k}}, \frac{\overline{C_k}}{\overline{C_j + C_k}}, 0, 0\}$. In this way, outliers don't have a dominant effect on the clustering.

3 Application

The test data is given in form of Terahertz measurements of five different compounds: PABA, tartaric acid, acetylsalicylic acid (ASS), salicylic acid and lactose. We have nine different measurements per compound and want the clustering to automatically divide the set into these five groups.
The absorption of a femtosecond pulse by these compounds is measured, i.e. a time resolved pulse spectrum is given. For better interpretability only the Fourier - transformed spectra are considered. For comparability of measurements taken at different times and places, it is important to take a reference measurement into account [4]. Transformed into Fourier space, one calculates the transmission T by dividing the sample I by the reference I_0, i.e. $T = \frac{I}{I_0}$. If we compare the Fourier transformed spectrum of the compound and

the logarithmic spectrum of the transmission we can see that there are frequencies where high transmission values are reached although in Fourier Space there is no noteworthy amplitude. This is mostly due to the decrease of I_0 and the resulting near zero divisions. To avoid this effect a variance analysis of each spectrum is proposed.
Each spectrum is divided into intervals of 16 data points. In each interval X the variance of the values is calculated:

$$var = \frac{1}{\#\{X\}} * \sqrt{\sum_{x_i \in X} (x_i - \overline{X})^2}.$$

If the variance stays under a certain threshold for more than three intervals, the first of these three is considered the break off interval. After this no more relevant information is going to be retrieved. Figure 2 shows the break off points of three different example spectra with variance threshold 0.2. As one can see each spectrum has its own. If all spectra are given beforehand, information beyond the maximum of these points is discarded. The other break off points are used later in the filtering procedure.

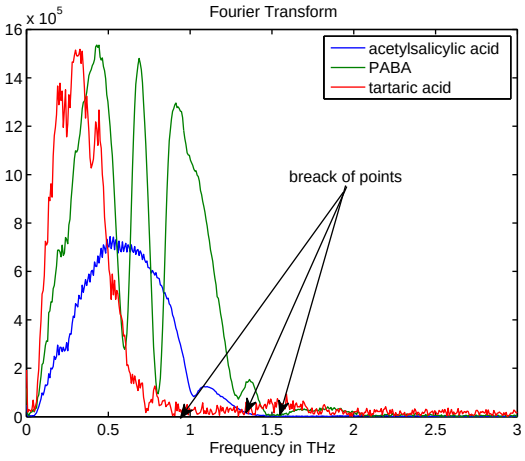

Figure 2: Example of Fourier transformed spectra, showing the cutting off points after variance calculation.

3.1 Filtering

After Finding the relevant regions the logarithmic transmission of each spectrum is calculated and the filters are applied. For the Wavelet transformation we use a Symmlet8 Wavelet and the coarsest level of coefficients being four. The last two levels of coefficients are shrunk to zero the first two are shrunk to zero beyond the threshold gained by the above described variance analysis. In that way the coarse overall appearance is maintained even after the individual break off points. Figure 3, left side, shows some example spectra, unfiltered and wavelet filtered by this method. One can see that although the smoothing is quite extensive the peaks are preserved well.
The Savitzky-Golay filter was applied in a similar way: using finer parameters before the individual cut off indexes and coarser for the intervals beyond. The parameters were altogether chosen quite generously to suit the broad bandwidth of these spectra. They are before: degree $= 3$, window size $= 9$, and after: degree $= 2$, window size $= 15$. Generally, by visual appearance, it seems that the Wavelet filter is doing better in coping with noisy areas.

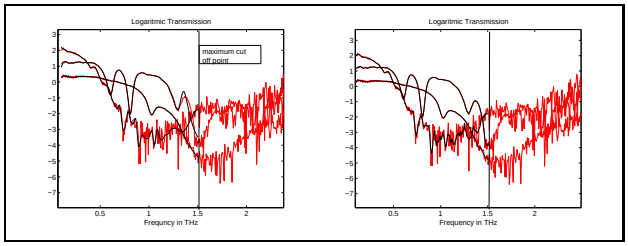

Figure 3: Unfiltered (red) and filtered (black) spectra of ASS, tartaric acid, and PABA. Left: Wavelet, Right: Savitzky-Golay.

3.2 Clustering

We now apply classical agglomerative clustering on the data. We use Euclidean distance and the average link function. As one can see in figure 4 the main difference between

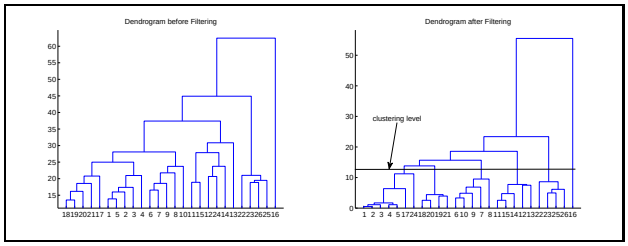

Figure 4: Dendrogram after clustering the unfiltered (left) and filtered (right) spectra.

the filtered and the unfiltered dendrograms is the total distance between the clusters. In the unfiltered case they are much higher. Although one can see the different groups in this dendrogram as well, it is difficult to find a common clustering level. The difficulty of automatically doing so is even higher. In contrast to that in the filtered dendrogram

one can chose levels around 10 and yield a good clustering result. This can be seen in figure 5. Both filtering methods performed quite similar in the clustering. Wavelet shrinkage showed clearer separation properties on the data set.

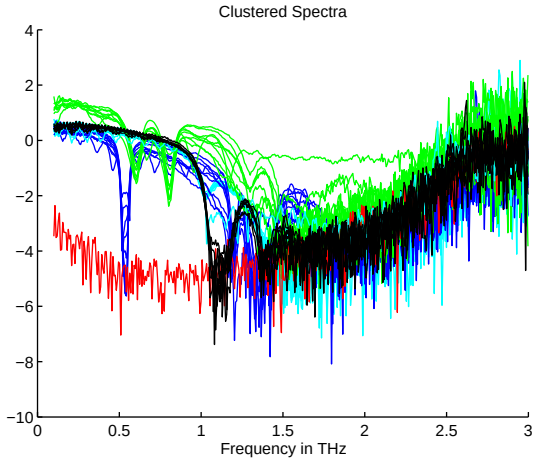

Figure 5: Clustered spectra after filtering and coarseness level 11.

4 Conclusion and Further Work

The goal of this paper was to test preprocessing methods, especially filtering, on Terahertz spectra with the final goal of clustering the spectra. Two popular methods: Wavelet shrinkage and Savitzky-Golay filters were applied. In clustering both tested methods gained good results and brought a notable improvement in comparison to the unfiltered transmission spectra. Both methods performed quite similarly. Wavelet filtering shows a slightly better capability of handling noise. This should be further investigated. In this paper it was tested to cluster with Euclidean distance on the filtered raw data. There are possibilities to classify in the wavelet domain that seem to yield good results [2]. Clustering with Euclidean distance holds problems in such high dimensional spaces. Therefore, the clustering of time series is often done not on raw data but on features that are extracted beforehand [10]. Hence, the possibility of performing a feature selection on the filtered data should be explored. A comparison of the filtering methods when doing the feature extraction should be done. In hierarchical clustering there are different attempts to perform such a feature extraction in a hierarchical way while performing the actual clustering [13] [5].

References

[1] F. Ehrentreich, S. Nikolov, M. Wolkenstein, and H. Hutter. The wavelet transform: A new preprocessing method for peak recognition of infrared spectra. *Microchimica Acta*, 128(3):241–250, 1998.

[2] R. Galvão, S. Hadjiloucas, J. Bowen, and C. Coelho. Optimal discrimination and classification of THz spectra in the wavelet domain. *Optics Express*, 11(12):1462–1473, 2003.

[3] M. Golay and A. Savitzky. Smoothing and differentiation of data by simplified least square procedures. *Analytical Chemistry*, 36:1627–1639, 1964.

[4] S. Gorenflo. *A Comprehensive Study of Macromolecules in Composites Using Broadband Terahertz Spectroscopy*. PhD thesis, Freiburg (Breisgau), 2006.

[5] S. Guha, R. Rastogi, and K. Shim. Cure: an efficient clustering algorithm for large databases. *Information Systems*, 26(1):35–58, 2001.

[6] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Comput. Surv.*, 31(3):264–323, September 1999.

[7] G. Lance and W. Williams. A General Theory of Classificatory Sorting Strategies: 1. Hierarchical Systems. *The Computer Journal*, 9(4):373, 1967.

[8] B. Lavine and J. Workman. Chemometrics. *Analytical Chemistry*, 78:4137–4145, 2006.

[9] B. Lavine and J. Workman. Chemometrics. *Analytical Chemistry*, 80:4519–4531, 2008.

[10] T. W. Liao. Clustering of time series data - a survey. *Pattern Recognition*, 38:1857 – 1874, 2005.

[11] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999.

[12] C. Schmuttenmaer. Exploring dynamics in the far-infrared with terahertz spectroscopy. *Chemical Reviews*, 104(4):1759–1780, 2004.

[13] T. Zhang, R. Ramakrishnan, and M. Livny. Birch: an efficient data clustering method for very large databases. *SIGMOD Rec.*, 25(2):103–114, 1996.

Differential Phase Contrast-Optical Coherence Tomography: Imaging and Object Reconstruction

Bettina Heise
Institute for Knowledge-based Mathematical Systems

November 25, 2008

1 Introduction

Arising from white light interferometry Optical Coherence Tomography (OCT) was invented in 1991 starting in medical application fields. It became a well-established technique for visualization of retina and skin diseases where cross-sectional imaging could provide valuable information about structure and modifications in tissues. Recently the capacity of OCT for material investigation in technical applications was more and more recognized. The possibility to explore embedded microstructure, buried interfaces or changes of refractive index in sub-surface regions of transparent or turbid samples shows the performance of this technique. OCT as a low coherence interferometry technique is based on the interferometric measurement of the intensity and time delay of the backscattered or backreflected light depending on the sample penetration depth. Internal material interfaces and inhomogeneities can be detected and visualized as depth resolved reflectivity signals. Utilizing a modified OCT configuration with respect to polarization sensitivity or a sheared probing beam, additionally phase information can be gained. The phase retardation or phase shift allows to draw conclusion about birefringence and anisotropies in the sample in case of Polarization Sensitive-OCT (PS-OCT) , or about optical path length (OPL) changes due to thickness or refractive index variations in case of Differential Phase Contrast-OCT (DPC-OCT).