ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

Classifying Scotch Whisky from near-infrared Raman spectra with a Radial Basis Function Network with Relevance Learning

Andreas Backhaus[1], Praveen C. Ashok[2], Bavishna B. Praveen[2], Kishan Dholakia[2], Udo Seiffert[1]

1- Biosystems Engineering, Fraunhofer Institute for Factory Operation and Automation (IFF), Magdeburg, Germany

2- SUPA, School of Physics and Astronomy, University of St. Andrews North Haugh, Fife, Scotland, UK

Abstract. The instantaneous assessment of high-priced liquor products with minimal sample volume and no special preparation is an important task for quality monitoring and fraud detection. In this contribution the automated classification of Raman spectra acquired with a special optofluidic chip is performed with the use of a number of Artificial Neural Networks. A standard Radial Basis Function Network is adopted to incorporate relevance learning and showed robust classification performance across classification tasks. The acquired relevance weighting per feature dimension can be used to reduce the number of features while retaining a high level of accuracy.

1 Introduction

The automated, on-line assessment of high-priced liquor products is essential for the standardization and quality monitoring in liquor production as well as potential fraud detection. An ideal sensor should be compact for mobile applications and require no special sample preparation while measure quality instantaneously. In [1] an optofluidic chip was presented that uses Raman spectroscopy to acquire a Raman spectrum of the fluid sample. This spectra proofed to be characteristic for whisky brands, age or type of maturing cask. A 1-Nearest Neighbourhood (1NN) classifier produced high accuracies in classifying Whisky age [1].

In this contribution we examine a number of Artificial Neural Network classifier for automated classification of age, distillery, cask and product variety from whisky spectra. It is shown that a small Artificial Neural Network model is capable of high classification accuracy. In order to reduce processing time and sensor costs, a strategy of feature selection, introduced as 'relevance learning' by [2], was used. We show that relevance learning can be integrated easily into a standard RBF Network which shows robust classification performance across datasets. The found weighting of feature dimensions was used to reduce the number of channels in the spectral data.

2 Data Acquisition

The procedure to acquire the Raman spectra from whiskey samples is shown in detail in [1]. In Raman spectroscopy a sample is illuminated with a laser beam. The laser light interacts with molecular vibrations, phonons or other excitations in the system, resulting in the energy of the laser photons being shifted up or

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

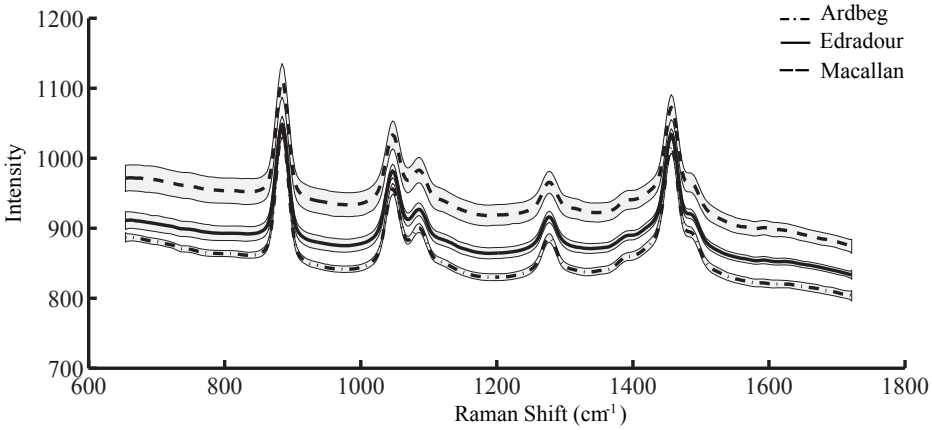

Fig. 1: Raman Spectra: Depicted are the mean Raman spectra for three different whiskey varieties together with the standard deviation. The whiskey varieties produce a distinct peak pattern in the spectra and a varying offset.

down. The shift in energy gives information about the vibrational modes in the system. Raman spectroscopy is commonly used in chemistry, since vibrational information is specific to the chemical bonds and symmetry of molecules. Therefore, it provides a fingerprint by which molecules can be identified.

Whisky samples of $20\mu l$ were directly loaded into the microfluidic chip without any preparation. After Raman acquisition, any remaining liquid at the sample inlet was wiped off and 40 μl of deionized water rinsed the system. Raman excitation was performed with 200 mW of laser power at a wavelength of 785 nm.

Six commercially available Scotch whisky brands and their variants were used to build the dataset. The 'Age' dataset consists of Glenfiddich at the age of 12, 14 and 18. The dataset 'Cask' is taken from a Glenmorangie matured in different casks. The 'Distillery' dataset groups all available data according to the brand. Finally the 'Product' dataset consists of different varieties of Bruichladdich Whisky. For each class, 400 Raman spectra were taken. Each dataset was scaled so the maximum across spectral bands was one. Figure 1 shows average spectra for three whisky classes with standard deviation. In Figure 2, for each dataset, the two most discriminant dimensions from an LDA projection are shown.

Relevance Radial Basis Function (rRBF) Network

The original Radial Basis Function (RBF) network was introduced in [3]. A RBF network is a two layer neural network with a first layer of prototypical spectra $\mathbf{w}_r$. A second layer calculates the network output

$$y_k(\mathbf{v}) = \sum_r u_{rk}\phi\left(d\left(\mathbf{v}, \mathbf{w}_r, \boldsymbol{\lambda}\right)\right)$$

with $\phi(x) = \exp\left(-\frac{x}{2\sigma^2}\right)$. The number of outputs is the number of classes.

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

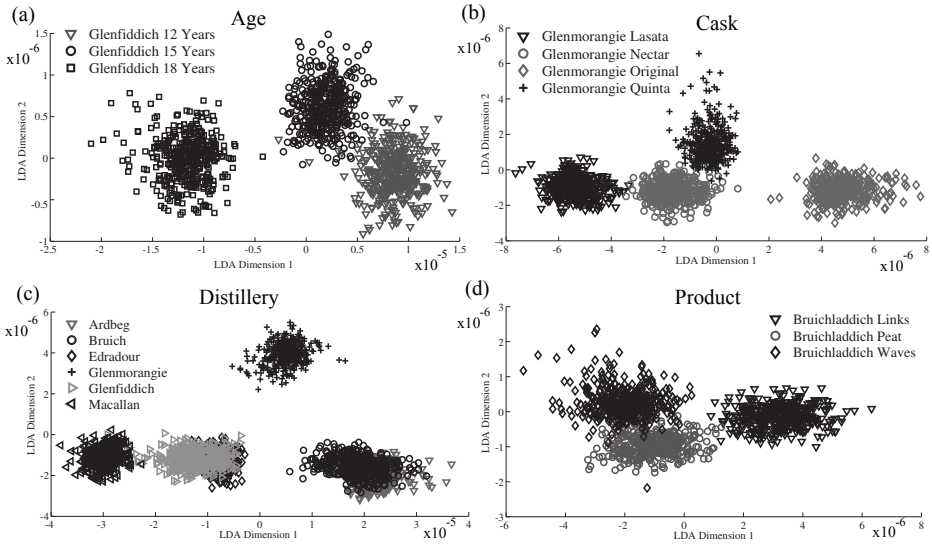

Fig. 2: Data sets used for classification: (a) Glenfiddich of different ages; (b) Glenmorangie matured in different casks; (c) Whiskey from different distilleries; (d) Different varieties of Bruichladdich; The two highest discriminative LDA components are shown.

The target vector $\mathbf{t}^j$ of the j-th sample is set to a 1-of-n coding with n being the number of classes. The (dis)similarity of input spectrum vs. prototype is calculated by weighted Euclidean distance,

$$d\left(\mathbf{v}, \mathbf{w}_r, \boldsymbol{\lambda}\right) \quad = \quad \sum_i \lambda_i \left(v_i - w_{ir}\right)^2$$

where λ_i is termed the relevance factor [2, 4] and weights each spectral band and will be adapted by the model training process.

The output weight u_{rk} is yielded by direct update [5].

$$\mathbf{U}^T \quad = \quad \Phi^{\dagger}\mathbf{T}$$

where $\mathbf{U} = (u_{rk})$, $(\mathbf{T})_{jk} = t_k^j$ and $(\Phi)_{jr} = \phi\left(d\left(\mathbf{v}^j, \mathbf{w}_r, \boldsymbol{\lambda}\right)\right)$ and † denotes the pseudo inverse. The parameters $\mathbf{w}_r$, σ_r, and $\boldsymbol{\lambda}$ are updated through minimizing

$$E\left(\mathbf{V}, \mathbf{W}, \boldsymbol{\lambda}\right) = \frac{1}{2}\sum_j \sum_k \left\{y_k\left(\mathbf{v}^j\right) - \mathbf{t}_k^j\right\}^2$$

which is achieved by gradient descent with momentum term. In the first iteration $k = 0$ relevance parameters are updated according to

$$\Delta\lambda_i\left(k\right) \quad = \quad -\epsilon^{\lambda}\frac{\partial E}{\partial \lambda_i}$$

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

Table 1: Test accuracy of classification; averaged across 5-fold cross-validation; standard deviation in brackets.

Method	Age	Cask	Distillery	Product
RBF	0.974 (0.010)	0.991 (0.006)	0.942 (0.011)	0.993 (0.004)
rRBF	0.981 (0.012)	0.994 (0.004)	0.949 (0.010)	0.973 (0.014)
GRLVQ	0.825 (0.073)	0.899 (0.043)	0.735 (0.037)	0.938 (0.015)
SRNG	0.874 (0.017)	0.978 (0.014)	0.740 (0.006)	0.968 (0.012)
SVM	0.885 (0.020)	0.962 (0.017)	0.841 (0.018)	0.941 (0.019)
1NN	0.980 (0.007)	0.994 (0.001)	0.953 (0.007)	0.980 (0.007)

while for any other iteration $k \geq 1$ updates are

$$\Delta\lambda_i(k) = -(1-\alpha)\,\epsilon^{\lambda}\frac{\partial E}{\partial \lambda_i} + \alpha\Delta w_r(k-1)$$

with α being the 'momentum term'. Updates for w_{ir} and σ_r are performed accordingly. The partial derivative for the relevance factor is as follows

$$\frac{\partial E}{\partial \lambda_i} = -\sum_j \sum_k \left\{ y_k\left(\mathbf{v}^j\right) - \mathbf{t}_k^j \right\} \sum_r u_{rk}\phi\left(d\left(\mathbf{v}^j, \mathbf{w}_r, \boldsymbol{\lambda}\right)\right) \frac{\left(v_i^j - w_{ir}\right)^2}{2\sigma_r^2}$$

The rRBF has been shown to have a robust classification performance in comparison to other prototype based approaches with relevance learning like Generalized Relevance Learning Vector Quantization (GRLVQ) [2] and Supervised Relevance Neural Gas (SRNG) [4] in classifying nutritional states in plants presented in [6].

3 Machine Learning

For the classification a number of different methods were implemented or used from 3rd party libraries and tested for their classification accuracy. A 1-Nearest Neighbourhood (1NN) classifier is used where all the training data is kept and the label of a presented test vector is the label of the best matching (Euclidean distance) training vector. A Standard GRLVQ and SRNG implementation are used with three prototypes per class. Weight adaptation is performed by stochastic gradient descent. The complete training data is presented to the network for 50 times. The rRBF network is setup with three times the number of classes in each dataset in the hidden layer to match network complexity. Weight adaptation is performed with conjugate gradient descent with momentum term. Training is performed till the step size reached a lower threshold. The output of the network is a 1ofN coding of the class label. As comparison, a Support Vector Machine (C-SVM) [7] with linear kernel was trained using the freely available libSVM package [1]. Classification is performed in a 5-fold cross validation setup. Accuracy is averaged and standard deviation calculated.

[1] *www.csie.ntu.edu.tw/~cjlin/libsvm/*

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

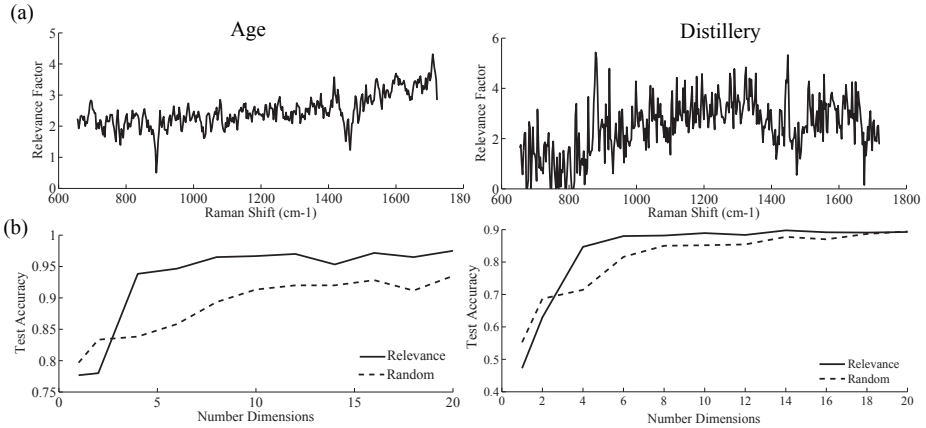

Fig. 3: Relevance Learning: (a) Depicts the weighting per Raman shift for the classification of age and distillery; (b) Compares a reduction of dimensions starting with just one Raman Shift value for random selection or selection guided by relevance weighting.

The described rRBF Network models optimize its energy function by adaptation of per-spectral band weighting. This weight vector is used to order the spectral bands due to their relevance. In order to check model performance on reduced spectral information, a rRBF network model was trained on the largest weighted bands as input and continuously added bands with the highest input dimensionality of 20. The rRBF networks were trained with the same setup and parameters as above. Test accuracy on unseen data is evaluated for each n-dimensional input space. As control, a matching number of spectral bands are chosen at random.

4 Results

In Table 1 test accuracies for all four dataset are depicted. The 1NN classifier showed robust performances of very high accuracy, comparable to the similar approach in [1]. However, while a 1NN classfier has to retain and match a data vector against a large dataset, the RBF as well as the rRBF managed to produce similar results in accuracy with a much smaller model and offer a much more efficient classification approach. The SVM performed similarly well with lower performance in the age and distillery classification. The performance of the GRLVQ and SRNG was as well dependent on the dataset with accuracy results ranging from mid 70% to over 90%.

In Figure 3 the results of the feature selection process from relevance learning are shown for the classification of distillery and age. Feature dimensions in Fig. 3a are weighted differently and by building a rRBF model based on the highest weighted dimensions, high accuracy can be achieved with less dimensions then selected by a random selection (see Fig. 3b).

ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 25-27 April 2012, i6doc.com publ., ISBN 978-2-87419-049-0.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100967420.

5 Conclusion

In this paper we presented an extension to the Radial Basis Function Network in order to weight features according to their relevance for a given task. The methodology is similar to relevance learning used GRLVQ and SRNG. The rRBF showed robust performances across a number of classification tasks. Near-infrared Raman spectra from Scotch whisky were classified for whiskey age, maturing cask, distillery and product variety with very high accuracy. The rRBF produces a relevance pattern that can be used to select the most important features for a given task and reduce sensor cost as well as processing time.

The feature selection technique and therefore dimension reduction using a relevance vector is an integral part of the classification model and is optimized due to the model objective function. This sets it into contrast to common bottom-up dimension reduction techniques like PCA and feature selection strategies for example by mutual information [8], which are independent of the actual classification or regression model. The selection strategy by relevance learning however does not deal with the likely co-linearity of the data e.g. its high correlation of neighbouring spectral bands. The spectra is seen as functional pattern [9]. For future work a regular sub-sampling of the spectra should be considered compared to the random selection strategy taken as control.

The optofluidic chip used for recording the Raman spectra is an ideal sensing device for detecting the quality of alcoholic beverages from a small sample volumes and with low acquisition time while being portable and easy to use. The combination of pattern producing sensing devices and advanced pattern recognition algorithms offers a powerful tool with applications in food inspection, environmental monitoring and medical diagnostics.

References

[1] Praveen C. Ashok, Bavishna B. Praveen, and K. Dholakia. Near infrared spectroscopic analysis of single malt scotch whisky on an optofluidic chip. *Opt Express*, 19(23):22982–22992, Nov 2011.

[2] Barbara Hammer and Thomas Villmann. Generalized relevance learning vector quantization. *Neural Networks*, 15:1059–1068, 2002.

[3] John Moody and Christian J. Darken. Fast learning in networks of locally tuned processing units. *Neural Computation*, 1:281–294, 1989.

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

[5] Cristopher M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press., 1995.

[6] Andreas Backhaus, Felix Bollenbeck, and Udo Seiffert. Robust classification of the nutrition state in crop plants by hyperspectral imaging and artificial neural networks. In *In Proc. 3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing*, Lissabon, Portugal, 2011.

[7] Vladimir Vapnik and Alexey Chervonenkis. *Theory of Pattern Recognition*. Nauka, 1974.

[8] Fabrice Rossi, Amaury Lendasse, Damien François, Vincent Wertz, and Michel Verleysen. Mutual information for the selection of relevant variables in spectrometric nonlinear modelling. *Chemometrics and Intelligent Laboratory Systems*, 80(2):215 – 226, 2006.

[9] Thomas Villmann, Erzsébet Merenyi, and Udo Seiffert. Machine learning approaches and pattern recognition for spectral data. In *Proc. of the 16. European Symposium on Artificial Neural Networks ESANN 2008, Bruges, Belgium*, D-Side Publications, pages 433–444, Bruges, Belgium, 2008.