s c c h
software competence center
hagenberg

FUZZY.LOGIC.LAB.LINZ

Advances in Knowledge-Based Technologies

Proceedings of the Master and PhD Seminar Summer term 2012, part 2

Softwarepark Hagenberg
SCCH, Room 0/2
5 July 2012

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

Session 1. Chair: Bernhard Moser

13:00 Mariya Zhariy:

Cumulative Wavefront Reconstructor for the Shack-Hartmann Sensor

13:30 Jean-Luc Bouchot:

The discrepancy norm of continuous functions: approximation and optimisation

14:00 Gernot Stübl:

Fusion of patchmatch and discrepancy norm

Session 2. Chair: Edwin Lughofer

14:30 Francisco Serdio:

Fault Detection By Residual Analysis

15:00 Carlos Cernuda:

Waveband Selection in NIR Spectra using Enhanced Genetic Operators

15:30 Ciprian Zavoianu:

On the Performance of Master-Slave Parallelization Methods for Multi-Objective Evolutionary Algorithms

Session 3. Chair: Roland Richter

16:00 Lucie Urbanova:

Derivation Digraphs for Similarity-Based Functional Dependencies

16:30 Wolfgang Heidl:

How do you decide? Connecting personal traits with machine-learned behavior

Inverse Problems and Imaging
Volume 5, No. 4, 2011, 893–913

Web site: http://www.aimSciences.org

CUMULATIVE WAVEFRONT RECONSTRUCTOR FOR THE SHACK-HARTMANN SENSOR

Mariya Zhariy, Andreas Neubauer,[1] Matthias Rosensteiner,[2] and Ronny Ramlau[1]

[1]Industrial Mathematics Institute, Johannes Kepler University Linz
A-4040 Linz, Austria

[2]MathConsult GmbH, Altenbergerstrae 69, A-4040 Linz, Austria

Abstract. We present a new direct algorithm aiming at the reconstruction of the optical wavefront from the Shack-Hartmann sensor measurements in Single Conjugate Adaptive Optics (SCAO) systems. The objective of an adaptive optics system designed for a large telescope can be only achieved if the wavefront reconstruction is sufficiently fast. Our scheme does not contain any explicit regularization for the reconstruction process but is still able to provide a good quality of reconstruction. The analysis of quality is given for three varying parameters: the diameter of the telescope, the number of subapertures and the level of photon noise. It has been shown both analytically and numerically that the quality of the reconstruciton, measured by the Strehl ratio, is reasonable for the small photon noise level and increases with the increasing number of subapertures for the same telescope size. The impact of the photon noise on the reconstruction gets higher with the increasing telescope diameter. The computational complexity of the method is linear in the number of unkowns. Counting all summation and multiplication steps the scaling factor is 14. Moreover, due to its simple structure, the cumulative reconstructor is pipelinable and parallelizable, which makes the effective computation even faster.

2000 *Mathematics Subject Classification:* 78A10,78M25.

Key words and phrases: Adaptive optics, inverse problems, Cumulative wavefront reconstructor.

Technical Report
FLLL–TR–1201

FUZZY.LOGIC.LAB.LINZ

The discrepancy norm of continuous functions: approximation and optimisation

July 4, 2012

Jean-Luc Bouchot
email: jean-luc.bouchot@jku.at

Abstract — In this presentation we reintroduce an approximation formula for the discrepancy of discrete signals and extend it for general continuous function on σ finite measurable spaces.

Together with this approximation, estimation of convergence and sensitivity of the approximation are given.

Finally, a formula for approximating the derivative of the discrepancy correlation is also given. First results show a good approximation power, together with low computations.

Key words — *optimisation, discrepancy norm, p-norm approximation*

Institut für Wissensbasierte Mathematische Systeme — Fuzzy Logic Laboratorium Linz

Johannes Kepler Universität Linz
Altenberger Str. 69
A-4040 Linz
Austria

softwarepark hagenberg

Softwarepark 21
A-4232 Hagenberg
Austria

1 Introduction

In this talk we start by recalling a first result obtained for signals defined on spaces with discrete uniform measures, as introduced in [1]. In this work we had introduced an approximation formula for the discrepancy norm [6] based on p-norms. Then as a result, we had shown that this approximation can be efficiently used in an alignment problem due to its implementation by two convolutions.

Here we investigate such approximations in a more general context of signals defined on (eventually) continous non-uniform measure spaces.

Moreover we compute formulas for the estimation of the derivative of the discrepancy correlation function as we would need it for practical applications of gradient descent like algorithms.

2 Approximation of one dimensional signals' discrepancy

2.1 Motivation

This subsection is intended as an introduction for our next results and is based on the work of Bouchot *et al.* [1].

In this case, we consider the counting measure μ_c and function of bounded support. Without loss of generality, we have $f : D := \{1, \cdots, N\} \to \mathbb{R}$ and the discrepancy norm reads

$$\|f\|_D = \max_{i,j \in D} |\sum_{k=i}^{j} f_k| = \max_{j \in D}\{0, \sum_{k=1}^{j} f_k\} - \max_{i \in D}\{0, \sum_{k=1}^{i} f_k\} \tag{1}$$

which can be implemented in a linear time by means of integral images [2, 7].

Note that the 0 in the $\max, \min$ operations is essential as the sums should be taken starting at $-\infty$. But the function can be extended with 0 outside the domain D. We can also equivalently write $\|f\|_D = \max_{j \in D_0} \sum_{k=0}^{j} f_k - \min_{i \in D_0} \sum_{k=0}^{i} f_k$ where we define $D_0 = D \cup \{0\}$ and $f_0 = 0$.

Before we recall the idea of the approximation and its property, we give some definitions

Definition 1 (Generalized mean)**.** *Let ϕ be a continuous continous invertible monotone function, let $f = \{f_k\}_{k=1}^{N}$, we can define [4] a mean as*

$$M_\phi(f) = \phi^{-1}\left(\frac{\sum_{k=1}^{N} \phi(f_k)}{N}\right) \tag{2}$$

Moreover, as we would expect from a mean, we have $\min_k f_k \leq M_\phi(f) \leq \max_k f_k$. (see [3] Ch II for the proof)

Now if we consider taking L^p norms as ϕ in the previous definition, this yields the following p means:

$$M_p(f) = \left(\frac{\sum_{k=1}^{N} |f_k|^p}{N}\right)^{1/p} \tag{3}$$

In our previous work we have introduced an approximation of the discrepancy norm for discrete one dimensional signals as described in the following theorem.

Theorem 1. *Let $f \in \ell^1 = L(\mathbb{Z}, \mu_c)$, we define*

$$\begin{aligned} \gamma_p : \ell^1 &\to \mathbb{R}^+ \\ f &\mapsto \gamma_p(f) := \ln\left(\frac{M_p(\chi(f))}{M_{-p}(\chi(f))}\right) \end{aligned} \tag{4}$$

where $\chi(f)(k) := \exp(\sum_{l=0}^{k} f_l)$ is the exponential of the cumulative function.

Then the followings holds:

$$\gamma_p(f) \leq ||f||_D < \gamma_p(f) + \frac{2}{p}\ln(N+1) \tag{5}$$

and γ_p is positive definite.

The proofs and more details on this theorem can be found in [1].

Corollary 1 (Choice of p)**.**

$$\forall \epsilon > 0, \forall f \in \ell^1, \forall p \in \mathbb{R}, p \geq p^* := \frac{2}{\epsilon}\ln(N+1) \Rightarrow |\gamma_p(f) - \|f\|_D| \leq \epsilon \tag{6}$$

This last part allows to compute an optimal approximation in a practical way. Indeed, due to the power p coming in the approximation process, the danger of getting overflow is getting higher and higher with finest approximation.

2.2 Approximation in the continuous case

Now that we have seen how practical and close the approximation can be, we want to extend this to continuous measurable functions.

From now on we go back to the setting of measure space on $\mathbb{R}$: $(\mathbb{R}, \Sigma, \mu)$ with μ being a finite measure.

In the same way, we can introduce the continuous counter-part of the generalized means

$$M_\phi(f) = \phi^{-1}\left(\frac{\int_{\mathbb{R}} \phi(f(x))d\mu(x)}{\int_{\mathbb{R}} d\mu}\right) \tag{7}$$

And the following bounding property still holds $\inf_{x\in\mathbb{R}} f(x) \leq M_\phi(f) \leq \max_{x\in\mathbb{R}} f_(x)$ (where $\inf$ and $\sup$ are understood as their essential definitions: $\mu\{x : f(x) > \sup f\} = 0$). We also re-define the p-means M_p in the same way.

The proposed approximation is reintroduced in the same way:

$$\begin{aligned} \Gamma_p : L^1(\mathbb{R}, \mu) &\to \mathbb{R}^+ \\ f &\mapsto \Gamma_p(f) := \ln\left(\frac{M_p(\chi(f))}{M_{-p}(\chi(f))}\right) \end{aligned} \tag{8}$$

where χ is defined as the exponential of the cumulative function.

The following theorem holds:

Theorem 2 (Approximation of the discrepancy norm for 1D continuous functions)**.** Γ_p *defines an approximation of the discrepancy norm in the sense that*

i) $\forall f \in L(\mathbb{R},\mu), \Gamma_p(f) \underset{p\to\infty}{\longrightarrow} \|f\|_D$

ii) $\Gamma_p(f) \leq \|f\|_D$

iii) $\forall \epsilon > 0, \exists p^* : \forall p \in \mathbb{R}, p \geq p^* \Rightarrow |\Gamma_p(f) - \|f\|_D| \leq \epsilon$

Proof. Direct calculus and application of the following lemmas and inequalities give the result.

Lemma 1 ($p-$norm approximation of the ∞ norm)**.** *Let* f *be in* $L^q \cap L^\infty$ *for a certain* $q \in \mathbb{R}$*. It holds:*

$$\|f\|_p \underset{p\to\infty}{\longrightarrow} \|f\|_\infty \tag{9}$$

A proof of this lemma can be found in Ch.III, Theorem 14F. of [5].

$$\forall f \in L^p(\mathbb{R},\mu) \cap L^q(\mathbb{R},\mu), 1 \leq p \leq q \leq \infty, \|f\|_p \leq \mu(\mathbb{R})^{1/p-1/q}\|f\|_q \tag{10}$$

Lemma 2 (Eventual convergence of L^p means towards ∞ norm)**.** *Let* $\epsilon\prime > 0$ *and* $g \in L^q \cap L^\infty$ *for a certain* $q \in \mathbb{R}$*. Let* $p_0 = p_0(t,\epsilon\prime,g) = \frac{\ln\left(\frac{\mu(E_g(t))}{\mu(\mathbb{R})}\right)}{\ln\left(\frac{1-\epsilon\prime}{1-t}\right)}$*, with* $E_g(t) = \{x : |g(x)| > \|g\|_\infty(1-t)\}$*. It holds*

$$\forall p > \max\{q, p_0\}, \frac{M_p(g)}{\|g\|_\infty} \geq 1 - \epsilon\prime \tag{11}$$

□

Remarks This lower bound depends strongly on the function f through the set $E_f(t)$; this is due to the non-equivalence of norms in the continuous case (in the discrete case, the bounds given for the theorem were crisp and uniform).

In the proof we have made use of an upper level set $E_f(t)$ and have proven that we get an upper bound for $t = \epsilon/2$. It would work for $t = \epsilon/4$ too and more generally for all $0 < t < \epsilon$. We have no idea at the moment what would be the influence of this on the estimated p^*.

Finding a nice p^* is a crucial problem for pratical applications. Indeed because we are taking the p^{th} power of an exponential, the risk to encounter overflow is high. Therefore we need to ensure a good approximation but keeping the power p as small as possible.

Example of lower bounds For illustration purposes let us see this effect on some toy examples.

We consider three functions:

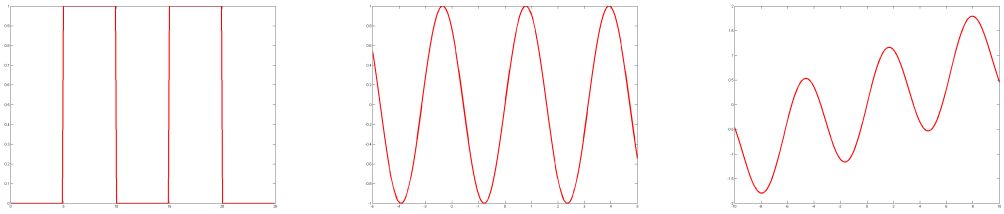
$$\begin{aligned} f_1 : [0,5] &\to \mathbb{R} \\ x &\mapsto 1 \text{ if } x \in [1,2[\cap[3,4],\ 0 \text{ elsewhere} \end{aligned} \tag{12}$$

$$\begin{aligned} f_2 : [-10,10] &\to \mathbb{R} \\ x &\mapsto \sin(x) \end{aligned} \tag{13}$$

$$\begin{aligned} f_3 : [-10,10] &\to \mathbb{R} \\ x &\mapsto a \cdot x + \sin(x) \end{aligned} \tag{14}$$

where we have chosen a to be around $1/10$. In our case, we consider a measure as a combination of uniformly distributed diracs (every 0.02 for f_1 and every 0.01 for the two others.)

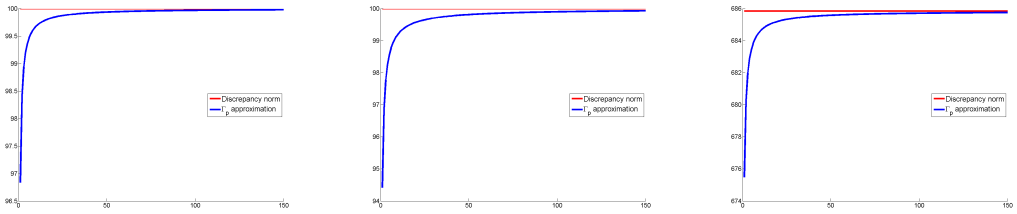
f_1, f_2, f_3 are illustrated in Fig. 1.



(a) 2 steps function: 250 data (b) Horizontal wave: 2001 data (c) Oblique wave: 2001 data

Figure 1: Our dataset of 1 dimensional toy functions used in our paper.

We start by analyzing the convergence of the $p-$norms approximations towards. The two first statements of Theorem 2 tell that Γ_p gets to $\|\cdot\|_D$ from below with p increasing. Figs. 2 depict this behaviour. One sees that the values of Γ_p functions increase drastically with p close to 0 and tends to stabilize to $\|\cdot\|_D$ early.



(a) 2 Steps function (b) Horizontal sine wave (c) Oblique sine wave

Figure 2: Convergence of the Γ_p functions towards the discrepancy norm. The horizontal axis corresponds to changing values of p while the vertical one is the output of the Γ_p and discrepancy norm functions.

The next point we need to have a look at is the reliability of our estimator for the value of p^*. While this value is clearly not optimal (we have left potentially a lot of the function away when deriving the estimation), we would like it to be relatively small anyway. As this is the value we are going to use afterwards for our tests, we need it to be small enough, in order not to get into machine overflow.

We have gahtered results based on our three toy functions. For all of them, we have ntoed the p^* defined in Theorem 2, and the smallest integer value, denoted by $\overline{p^*}$, for which the approximation up to an ϵ_0 holds. Moreover, as it can be seen in the proof of the theorem, one can tune the threshold used for the decomposition of the function. This threshold has also been analyzed for the horizontal sine wave function.

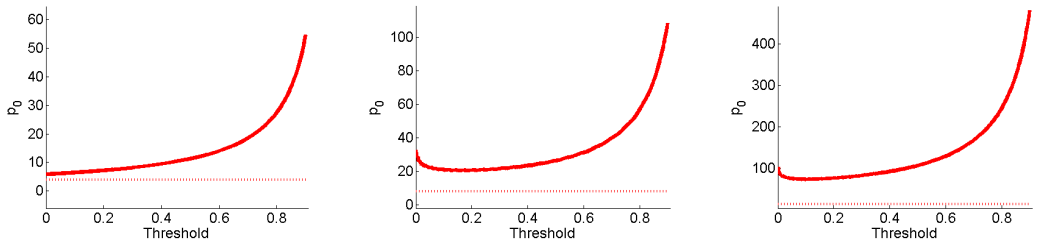
The results are summarized in Table. 1.

It comes out that we indeed get close approximate of the discrepancy norm of a function by means of $p-$norms approximation. The estimator of a good p^* is however strongly dependent on

Function	ϵ_0	α	p^*	$\Gamma_{p^*}(f)$	$\|f\|_D$	$\overline{p^*}$
2 steps	0.1	1/2	100.77	99.97	100	32
	5	0.2	2.68	98.81	100	1
Horizontal sines	0.1	1/4	257.73	99.96	100.00	101
	0.1	1/2	348.72	99.97	100.00	101
	0.1	3/4	666.56	99.98	100.00	101
	0.05	1/2	767.06	99.98	100.00	212
	2	1/2	13.45	99.39	100.00	4
Oblique sines	1	1/2	105.99	685.72	685.85	13
	2	1/4	44.64	685.55	685.85	7

Table 1: Examples of approximation and lower bound estimations on some toy functions.

the choice of the threshold. However first tests tend to show that in when given a threshold in the order of 0.1 or 0.2 yield stable and reliable results.



(a) 2 Steps function (b) Horizontal sine wave (c) Oblique sine wave

Figure 3: Computed values of the p^* estimator. The continuous curvs show the estimated p^* while the the dotted straight line corresponds to the $\overline{p^*}$ as described in the Table 1

Figs. 3 show the behaviour of the estimation of p^* for different thresholding values. For the rest of the paper we will use in our applications a thresholding of 0.2 which seems to lie in a rather stable area and which achieves low values of p^*.

3 Differentiation of the approximation

In this subsection we want to introduce a way of optimizing the discrepancy norm based correlation function. Given a function f and a reference g, we want to find the optimal translation parameter t^* such that $\|f_{t^*} - g\|_D$ is minimal, where $\forall t \in \mathbb{R}, f_t(x) = f(x-t)$. We consider the following optimization problem:

$$t^* = \underset{t\in\mathbb{R}}{\operatorname{argmin}} \|f_t - g\|_D \tag{15}$$

We would like to optimize the objective function $J(t) = \|f_t - g\|_D$ making use of the monotonicity property of the discrepancy norm when facing misaligned functions. However, due to its definition with $\max$ and $\min$, the discrepancy norm is not everywhere differentiable (it is however

almost everywhere differentiable du to Rademacher's theorem). Moreover, as it can be seen from Fig. 4, the objective function shows some kind of plateau on some clinical cases which yield any gradient based method to fail in finding a correct solution. Therefore, we want to make use of the previous approximation to compute an approximated gradient to the objective function. As we will see, we can derive formula based on a scalar product which allows us to compute the derivative very efficiently.

Theorem 3 (Derivation of the discrepancy correlation)**.** *The discrepancy correlation function's derivative $J(t)$ can be approximated in the following way:*

$$\frac{\partial J}{\partial t}(t) \approx \left\langle \int_{x=-\infty}^{\cdot} f'(x-t)\,\mathrm{d}\mu(x), \left(\frac{\chi(-f_t+g)}{\|\chi(-f_t+g)\|_p}\right)^p - \left(\frac{\chi(f_t-g)}{\|\chi(f_t-g)\|_p}\right)^p \right\rangle \tag{16}$$

Proof. We need to compute the derivative of the discrepancy norm of a difference. As it has been seen in the previous section, it can be approximated by means of L^p norms and we get

$$\frac{\partial J}{\partial t}(t) \approx \frac{\partial \Gamma_p}{\partial t}(f_t - g) = \frac{\partial}{\partial t}\left(\frac{1}{p}\ln\left(\|\chi(f_t-g)\|_p^p\right) + \frac{1}{p}\ln\left(\|\chi(-f_t+g)\|_p^p\right) + C\right) \tag{17}$$

Now application of the chain rule and the theorem of differentiation under the integral gives the result. □

While this formula seems complicated, we can see that on real discrete signals applications, it can be simplified a little

Corollary 2 (Derivation for discrete uniform measure)**.** *Assume μ is a uniform discrete measure on a bounded domain. Then the derivation of the discrepancy autocorrelation function can be approximated by*

$$\frac{\partial J}{\partial t}(t) \approx \left\langle f_t, \left(\frac{\chi(-f_t+g)}{\|\chi(-f_t+g)\|_p}\right)^p - \left(\frac{\chi(f_t-g)}{\|\chi(f_t-g)\|_p}\right)^p \right\rangle \tag{18}$$

or equivalently:

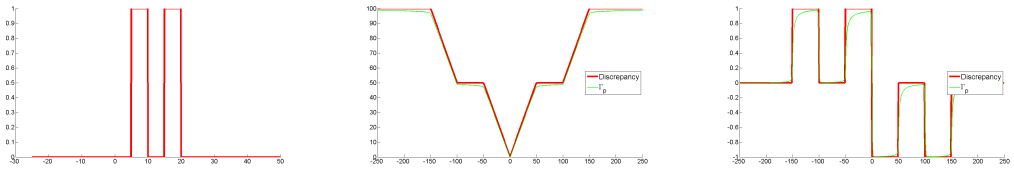
$$\frac{\partial J}{\partial t}(t) \approx \sum_{i=1}^{N} f_t(i)\left(\left(\frac{\chi(-f_t+g)(i)}{\|\chi(-f_t+g)\|_p}\right)^p - \left(\frac{\chi(f_t-g)(i)}{\|\chi(f_t-g)\|_p}\right)^p\right) \tag{19}$$

where N denotes the size of the discrete finite signal (i.e. vector) we consider.

4 Conclusion

We have seen how the discrepancy correlation function can be approximated and differentiated by means of p norms for 1 dimensional eventually continuous signals. Besides analytic formalae, we have propose some way to select beforehand the values of the power sufficient to get a good approximation. As it comes out on practical examples, these power might not be optimal they are still however small enough to get reliable results without overflow troubles.

Future work will include extension to higher dimensional signals and practical applications of such results for image alignment.



(a) Extended two steps function (b) Autocorrelation function: discrepancy and Γ_p (c) Approximate derivatives

Figure 4: This figure shows how a gradient descent based optimization algorithm might fail when trying to locally optimize the discrepancy correlation function. Indeed the correlation function (second figure) shows some plateau where the derivative (illustrated on the third figure) is 0. On the other hand, using an appropriate $p-$norm approximation allows to overcome this effect while keeping a really close objective function.

References

[1] BOUCHOT, J.-L., HIMMELBAUER, J., AND MOSER, B. On autocorrelation based on hermann weyl's discrepancy norm for time series analysis. In *IJCNN* (2010), IEEE, pp. 1–7.

[2] CROW, F. Summed-area tables for texture mapping. *ACM SIGGRAPH Computer Graphics 18*, 3 (1984), 207–212.

[3] HARDY, G., LITTLEWOOD, J., AND PÓLYA, G. *Inequalities*. Cambridge Univ Pr, 1952.

[4] KOLMOGOROV, A., AND CASTELNUOVO, G. *Sur la notion de la moyenne*. G. Bardi, tip. della R. Accad. dei Lincei, 1930.

[5] LOOMIS, L. *An introduction to abstract harmonic analysis*. Van Nostrand Reinhold, 1953.

[6] MOSER, B. A similarity measure for image and volumetric data based on hermann weyl's discrepancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 33*, 11 (2011), 2321–2329.

[7] VIOLA, P., AND JONES, M. Robust real-time object detection. *International Journal of Computer Vision 57*, 2 (2001), 137–154.

Fusion of Patchmatch and Discrepancy Norm

Gernot Stübl and Bernhard Moser

July 5, 2012

Abstract

This talk presents the current work in progress of applying discrepancy norm to a patchmatch based algorithm for generating approximate nearest neighbour fields. The goal is to improve the behaviour of the original patchmatch algorithm with patches of high frequency and accelerate the algorithm through recent theoretic insights.

Two potential improvements have been identified: (a) enhance the random phase by integrating a RANSAC like global optimization procedure and utilize the Lipschitz property of the discrepancy norm as well as (b) utilize the discrepancy norm as similarity function in the propagation phase. The main focus of the talk regards (b). Experiments on a standardized database have shown that discrepancy norm and L_1 have different strengths and weaknesses in different regions of an image. Therefore a weighted combination is of both similarity measures is envisioned.

A first explorative version of such a weighting function is done by analysing the settling sequence in the self reference case. Experiments on the dataset showed that a simplistic learning algorithm can classify if a settling sequence is shorter with discrepancy norm or L_1 norm with a F-measure of 69% . This motivates to invest further research into this direction. However the current results show over the whole dataset only minor improvements, so further effort has to be taken into refining the learning algorithm and weighting function.

Fault Detection by Residual Analysis

Francisco Serdio[a], Ewdin Lughofer[a]

[a]*Department of Knowledge-Based Mathematical Systems, Johannes Kepler University Linz, Austria*

Abstract

Principal Component Analysis (PCA) has been widely used in the field of fault detection, with a good degree of success. But this approach becomes not applicable when the process at hand presents time-varying behaviors, mainly due to the assumption of linear correlation between the variables made by PCA. We propose an alternative to PCA in the field of fault detection, describing the framework to use when developing the approach and describing also its modeling capabilities. Our approach is residual-based, and focus its attention in the track of the residuals over time. Our framework establish data cleaning, system identification, model training and model testing stages. The system identification extracts the variables that identify each channel in the process. The model training stage trains a model per channel according to the identified system. The approach allows to train several models per channel and select the best one, but this is still on the outlook. Once we have a model explaining each channel in the process, the best ones are selected according to its quality, $R2-adjusted$. This opens the door to exclude low-quality and/or instable models, bringing into play the concept of cascadability. Cascadability gives the opportunity to include new models when new channels appear, replace models when there are changes on the process and remove models when the quality is either degraded or not good enough according to established parameters. The testing stage get the residuals of the identified systems and calculate both global and local error bars as an uncertainty measure. Residuals are normalized according to the uncertainty and subsequently tracked online with a dynamic tolerance band. The residuals outside the tolerance band are them classified as faults, so the system triggers a fault alarm. We test our approach with three different models -linear model, transformed linear model and fuzzy model-, using both global and local error bars when expressing the uncertainty. We show results where its clear that our approach outperforms PCA most of the times either regardless the type of the model used or the way used to articulate the uncertainty. Our results are encouraged by an statistical test, manifesting the pairwise preference of the methods. From the results we also reinforce the existing literature reporting the (sometimes) non-applicability of PCA in time-varying processes.

Keywords: Fault detection, data-driven regression modeling, non-linear transformations, Box-Cox, fuzzy systems extraction, dynamic residual analysis, Principal Component Analysis.

Email addresses: francisco.serdio@jku.at (Francisco Serdio), ewdin.lughofer@jku.at (Ewdin Lughofer)

Waveband Selection in NIR Spectra Using Enhanced Genetic Operators

Carlos Cernuda*[a], Edwin Lughofer[a], Wolfgang Märzinger[b], Wolfram Summerer[c]

[a]*Department of Knowledge-Based Mathematical Systems, Johannes Kepler University Linz, Austria*
[b]*i-RED Infrarot Systeme GmbH, Linz, Austria*
[c]*Recendt GmbH, Linz, Austria*

Abstract

Nowadays the techniques employed in data acquisition provide huge amounts of data. Some parts of the information are related to the others, making desirable a way to reduce the number of variables, i.e. dimensionality reduction, loosing as less information as possible, in order to decrease computational times and complexity when applying any Data Mining technique, e.g. for classification or regression purposes. Genetic Algorithms(GA) offer the possibility of selecting which variables contain the most relevant information in order to represent all the original ones. The traditional genetic operators seem to be too general, leading to results which could be improved by means of designed genetic operators that employ some available problem specific information.

Especially, when dealing with calibration by means of NIR spectral data, which use to contain thousands of variables, it is known that not isolated wavelengths but wavebands allow a more robust model design. This aspect should be taken into account when crossing individuals. We propose three crossover operators specifically designed for calibration with NIR spectral data, based on a pseudo-random 2-points crossover where the first point is randomly chosen and the selection of the second point is guided by problem specific information, and we compare their performance against state of the art operators. The chosen fitness function is partial least squares regression (PLSR), because it is fast and widely used in chemometrics. Our benchmark consists of two real world high dimensional data sets, corresponding to polyetheracrylat (PEA), where hydroxyl number, viscosity and acidity are on-line monitored; and melamine resin production, where the chilling point is considered in order to regulate the condensation. The former contains a very low number of samples, each of them measured several times. Once the variables are selected, we check their performance on PLSR (linear) and FLEXFIS (non linear), in order to check that the chosen algorithm for the fitness function is not relevant in the results achieved by the selected variables. We show that designed operators promote wavebands selection, lead to in general better good quality solutions, and converge faster and smoother to them than S-o-A operators.

Keywords: Variable Selection, Dimensionality Reduction, Genetic Algorithm, Forward Selection, Variable Bands, Design Genetic Operators.

Email address: carlos.cernuda@jku.at (Carlos Cernuda*)

On the Performance of Master-Slave Parallelization Methods for Multi-Objective Evolutionary Algorithms

Alexandru-Ciprian Zăvoianu [a,c] Edwin Lughofer [a]
Wolfgang Amrhein [a,b] Erich Peter Klement [a,c]

[a] *Department of Knowledge-based Mathematical Systems / Fuzzy Logic Laboratorium Linz-Hagenberg, Johannes Kepler University of Linz, Austria*

[b] *Institute for Electrical Drives and Power Electronics, Johannes Kepler University of Linz, Austria*

[c] *ACCM, Austrian Center of Competence in Mechatronics, Linz, Austria*

Abstract

This paper is focused on a comparative analysis of the performance of two master-slave paralellization methods: the standard *generational* scheme (GEN-MSPS) and a *steady-state asynchronous* scheme (SSA-MSPS). Both parallelization methods can be used to drastically improve the physical optimization time of multi-objective evolutionary algorithms (MOEAs) that rely on time-intensive fitness evaluation functions. We assesd both the quantitative and the qualitative impact of opting for a certain parallelization scheme and the obtained empirical results indicate that, when dealing with a time-wise fitness evaluation distribution that has a significant amount of variance, the SSA-MSPS seems to be the better parallelization option.

Key words: evolutionary computation, multi-objective optimization, performance comparison, master-slave parallelization, steeady-state evolution.

Derivation Digraphs for Similarity-Based Functional Dependencies

(Abstract for seminar: Advances in Knowledge-Based Technologies Summer 2012) *

Lucie Urbanova and Vilem Vychodil

DAMOL (Data Analysis and Modeling Laboratory)
Dept. Computer Science, Palacky University, Olomouc
17. listopadu 12, CZ–77146 Olomouc, Czech Republic
lucie.urbanova01@upol.cz, vychodil@acm.org

Abstract. We present a graph-based method of reasoning with similarity-based functional dependencies in an extension of the Codd's model of data, which generalize the ordinary functional dependencies by considering similarity of attribute values. The main results show that a degree to which an similarity-based functional dependency is semantically entailed from a set (or a graded set) of other similarity-based functional dependencies can be characterized by existence of particular directed acyclic graph

1 Introduction

This paper is a continuation of our previous work [3, 4] where we have introduced the concept of a ranked table over domains with similarities, which is a counterpart to the concept of a table (relation) of the ordinary Codd's model. The motivation for our extension is the fact that in many situations is desirable to consider similarities on domains instead of equalities. Assume that a buyer (user of database) want to buy a family car and is searching for a car sold for 7 000 €, which is three years old. Then the user is interested not only in cars sold for the exact price 7 000 € and being exactly three years old, but also in cars which are sufficiently "close" to his requirements (expressed as query in database). The "closeness" can be formalize using similarity relations (reflexive and symmetric) on domains values. Then each tuple in the answering data table will have a *rank* value to denote how much a tuple matches the user's query. Ranks have mainly comparative meaning: the higher the rank the better match, and will come from complete residuated lattice, see section 2 for details. The data table equipped with similarity relations on domains and ranks assigned to tuples will be called a ranked data table (shortly an RDT). An example of such a table can be seen in Table 1. The table consists of the following attributes:

* Supported by grant no. P103/11/1456 of the Czech Science Foundation and internal grant of Palacky University no. PrF_2012_029. DAMOL is supported by project reg. no. CZ.1.07/2.3.00/20.0059 of the European Social Fund in the Czech Republic.

Table 1. Cars for sale at 7 000 € from year 2009

	Brand	Model	Price	Year	Fuel	Km
0.97	Renault	Scénic	6 940	2009	Diesel	115 556
0.8	Renault	Scénic	7 200	2010	Diesel	101 478
0.75	Opel	Zafira	7 500	2008	Diesel	130 656
0.54	Citroen	C3 Picasso	7 925	2009	Diesel	109 015
0.1	Volkswagen	Caddy	8 600	2008	Diesel	122 855

Brand, *Model*, *Price* (in euros), *Year* (year of production), *Fuel*, *Km* (state of the odometer in kilometers).

Another situations when similarities naturally appears is when searching for functional dependencies in a relational database. The interpretation of an ordinary functional dependence $A \Rightarrow B$ in a given data table $\mathcal{D}$ is the following: if any two tuples have the same values on all attributes from A, then they have the same values on all attributes from B. As one can see, there are no non-trivial functional dependencies in Table 1, however, one can intuitively see that there are "similarity" dependencies, for example: Similar year of production imply similar state of tachometer. Thus it seems natural to interpret functional dependence $A \Rightarrow B$ as follows: if any two tuples have similar values on attributes from A, then they have similar values on attributes from B. In our model we consider sets A, B as fuzzy sets over some finite set of attributes R and we allow the functional dependence to be satisfied in degrees, not only 0 or 1, see section 2 for more details.

In the original Codd's model semantic entailment is used to determine whether a functional dependency follows from a set of other functional dependencies and plays very important role in eliminating redundancy. The best known inference system has been proposed by Armstrong [1] and can be seen as a system consisting of two rules [13]: (Ax): infer $A \cup B \Rightarrow B$, and (Cut) from $A \Rightarrow B$ and $B \cup C \Rightarrow D$ infer $A \cup C \Rightarrow D$. An interesting alternative graph-based approach that is also aimed at possible automated proving has been proposed by Maier in [14], see also [15] for an extensive description and its application for theorem proving. The goal of this paper is to show a graph-based inference for similarity-based functional dependencies (shortly SBFDs) in RDTs, where the attributes are graded.

2 Preliminaries

We recall basic notions of directed graphs, residuated lattices, and our extension of Codd's relational model. Details can be found in [2, 5, 10, 11, 8, 6].

A directed graph (a digraph) is a pair $\mathbf{D} = \langle V, A \rangle$, where V is a nonempty finite set of elements called vertices and A is a binary relation $A \subseteq V \times V$, each $\langle v, w \rangle \in A$ is called an arc. V and A are called the vertex set and the arc set of $\mathbf{D}$, respectively. If $\langle v, w \rangle \in A$, we say that the arc $\langle v, w \rangle$ leaves v and enters w. A digraph $\mathbf{D} = \langle V, A \rangle$ is acyclic (in short, $\mathbf{D}$ is a DAG) if there is no finite sequence

$v_1, \ldots, v_k$ ($k \geq 2$) of vertices from V such that $v_1 = v_k$ and $\langle v_i, v_{i+1} \rangle \in A$ for all $i = 1, \ldots, k-1$, i.e., if $\mathbf{D}$ does not contain a cycle in the usual sense.

A complete residuated lattice with a truth-stressing hedge (shortly, a hedge) [11, 12], which serves as a structure of truth degrees, is an algebra $\mathbf{L} = \langle L, \wedge, \vee, \otimes, \rightarrow, {}^*, 0, 1 \rangle$ such that $\langle L, \wedge, \vee, 0, 1 \rangle$ is a complete lattice with 0 and 1 being the least and the greatest element of L, respectively; $\langle L, \otimes, 1 \rangle$ is a commutative monoid (i.e. $\otimes$ is commutative, associative, and $a \otimes 1 = 1 \otimes a = a$ for each $a \in L$); $\otimes$ and $\rightarrow$ satisfy so-called adjointness property: $a \otimes b \leq c$ iff $a \leq b \rightarrow c$ for each $a, b, c \in L$; hedge * satisfies $1^* = 1$, $a^* \leq a$, $(a \rightarrow b)^* \leq a^* \rightarrow b^*$,$a^{**} = a^*$, for each $a, b \in L$, $a_i \in L$ ($i \in I$). Elements a of L are interpreted as truth degrees. The operations $\otimes$ and $\rightarrow$ are (truth functions of) "fuzzy conjunction" and "fuzzy implication" and are called a multiplication and residuum, respectively. Hedge * is a (truth function of) logical connective "very true", see [11, 12].

A special case of a complete residuated lattice with hedge is the two-element Boolean algebra $\langle \{0, 1\}, \wedge, \vee, \otimes, \rightarrow, {}^*, 0, 1 \rangle$, denoted by $\mathbf{2}$, which is the structure of truth degrees of the classical logic.

2.1 Ranked Data Tables over Domains with Similarities

We consider a set of *attributes* Y. Any subset $R \subseteq Y$ is called a *relation scheme.* For each attribute $y \in Y$ we consider its *domain* D_y, i.e. a set of all possible values of the attribute. In addition, we equip each domain D_y with a binary $\mathbf{L}$-relation $\approx_y$ on D_y which satisfies the following conditions: (i) for each $u \in D_y$: $u \approx_y u = 1$, and (ii) for each $u, v \in D_y$: $u \approx_y v = v \approx_y u$. Each binary $\mathbf{L}$-relation $\approx_y$ on D_y satisfying (i) and (ii) shall be called a *similarity*; pair $\langle D_y, \approx_y \rangle$ is called a *domain with similarity.* Let $R \subseteq Y$ be a relation scheme and denote by $\mathrm{Tupl}(R)$ the Cartesian product of domains D_y ($y \in R$). Each $r \in \mathrm{Tupl}(R)$ is called a *tuple over* R. Obviously, $r(y) \in D_y$ for each $y \in R$; $r(y)$ is called the *y-value of r*.

Definition 1. Let $R \subseteq Y$ be a relation scheme. A ranked data table on R over $\{\langle D_y, \approx_y \rangle \mid y \in R\}$ (shortly, an RDT) is any finite $\mathbf{L}$-set $\mathcal{D}$ in $\mathrm{Tupl}(R)$. The degree $\mathcal{D}(r)$ to which r belongs to $\mathcal{D}$ is called a rank of tuple r in $\mathcal{D}$.

According to Definition 1, if $\mathcal{D}$ is an RDT on R over $\{\langle D_y, \approx_y \rangle \mid y \in R\}$ then $\mathcal{D}$ is a map $\mathcal{D} \colon \mathrm{Tupl}(R) \rightarrow L$ assigning to each tuple r over R a degree $\mathcal{D}(r) \in L$ which is called a rank. Note that $\mathcal{D}$ is, in fact, an n-ary $\mathbf{L}$-relation between domains D_y ($y \in Y$) since $\mathcal{D}$ is a map from $\prod_{y \in R} D_y$ to L. For an RDT $\mathcal{D}$ on R, tuples t_1, $t_2 \in \mathrm{Tupl}(R)$ and $\mathbf{L}$-set $C \in L^R$ of attributes, we introduce a degree to which t_1 and t_2 have similar values on attributes from C

$$t_1(C) \approx_{\mathcal{D}} t_2(C) = (\mathcal{D}(t_1) \otimes \mathcal{D}(t_2)) \rightarrow \bigwedge_{y \in R} (C(y) \rightarrow t_1(y) \approx_y t_2(y)) \qquad (1)$$

Now the validity of similarity-based functional dependency (shortly SBFD) is captured by the following definition. A degree $||A \Rightarrow B||_{\mathcal{D}}$ to which $A \Rightarrow B$ is

true in a RDT $\mathcal{D}$ is defined by

$$||A \Rightarrow B||_{\mathcal{D}} = \bigwedge_{t_1,t_2 \in \mathrm{Tupl}(R)} \Big((t_1(A) \approx_{\mathcal{D}} t_2(A))^* \rightarrow (t_1(B) \approx_D t_2(B)) \Big) \qquad (2)$$

Remark 1. By basic rules of semantics of predicate fuzzy logic, $t_1(C) \approx_{\mathcal{D}} t_2(C)$ is a truth degree of formula "if t_1, t_2 are from $\mathcal{D}$ then for each attribute y from C, t_1 and t_2 have similar values on y", and $||A \Rightarrow B||_{\mathcal{D}}$ is a truth degree of formula "for any tuples t_1, t_2: if t_1 and t_2 have very similar values on attributes from A then they have similar values on attributes from B".

We first define a semantic entailment from a collection T of SBFD's. We will say that a datatable $\mathcal{D}$ is a model of theory T if for each SBFD from T, a degree to which $A \Rightarrow B$ is truth in $\mathcal{D}$ is higher or equal to a degree prescribed by T,

$$\mathrm{Mod}(T) = \{\mathcal{D} | \text{ for each } A, B \in \mathbf{L}^R : T(A \Rightarrow B) \leq ||A \Rightarrow B||_D\} \qquad (3)$$

Then a degree to which $A \Rightarrow B$ semantically follows from a theory T of SBFD's is given by: $||A \Rightarrow B||_T = \bigwedge_{\mathcal{D} \in \mathrm{Mod}(T)} ||A \Rightarrow B||_{\mathcal{D}}$.

In our previous work we have shown that for finite $\mathbf{L}$ there is an axiomatization which uses (Ax) and (Cut) as in Section 1 (only with sets replaced by $\mathbf{L}$-sets and $\cup$ being the union of $\mathbf{L}$-sets defined componentwise using $\vee$) with an additional deduction rule of multiplication: (Mul): from $A \Rightarrow B$ infer $c^* \otimes A \Rightarrow c^* \otimes B$, where $c \in L$ and $x \otimes C$ is an $\mathbf{L}$-set in Y defined by $(x \otimes C)(y) = x \otimes C(y)$ for all $y \in Y$. The inference system consisting of (Ax), (Cut), and (Mul) is complete in the following sense: $||A \Rightarrow B||_T = 1$ iff $T \vdash A \Rightarrow B$, i.e. iff there is a proof of $A \Rightarrow B$ from T. This result (ordinary-style completeness [6]) characterizes SBFDs which follow semantically from T to degree 1. There is also a result on graded-style (Pavelka-style [16]) completeness saying that

$$||A \Rightarrow B||_T = \bigvee \{c \in L \,|\, T \vdash A \Rightarrow c \otimes B\}, \qquad (4)$$

i.e. that the degree to which $A \Rightarrow B$ semantically follows from T is a supremum of degrees $c \in L$ for which $A \Rightarrow c \otimes B$ are provable from T in the ordinary sense. Details can be found in [6].

3 Derivation Acyclic Digraphs for SBFDs

We now introduce derivation diagrams as particular acyclic digraphs where vertices are labeled by attributes from Y and degrees from $\mathbf{L}$. The arcs of the diagrams will correspond to SBFDs from an input theory and indicate which formulas from the theory are used in the process of inference. In what follows, $\mathbf{L}^*$ is a complete residuated lattice with hedge.

Definition 2 (T-based $\mathbf{L}^*$-derivation DAG). Let T be a set of SBFDs over Y.

1. Any $\mathbf{D} = \langle V, \emptyset \rangle$ such that $\emptyset \neq V \subseteq Y \times L$ is a T-based $\mathbf{L}^*$-derivation DAG;

2. If $\mathbf{D} = \langle V, A\rangle$ is a T-based $\mathbf{L}^*$-derivation DAG and there are $E \Rightarrow F \in T$, attribute $y \in Y$, and vertices $\langle y_1, a_1\rangle \in V, \ldots, \langle y_k, a_k\rangle \in V$ such that for

$$s_0 = \bigwedge\{E(y) \to 0 \,|\, y \in Y \text{ and } y \notin \{y_1, \ldots, y_k\}\}, \tag{5}$$
$$s_1 = \bigwedge\{E(y_i) \to a_i \,|\, i = 1, \ldots, k\}, \tag{6}$$
$$m = \bigvee\{a \in L \,|\, \langle y, a\rangle \in V\}, \tag{7}$$
$$d = \big((s_0 \wedge s_1)^* \otimes F(y)\big) \vee m, \tag{8}$$

we have $d > m > 0$, then $\mathbf{D}' = \langle V', A'\rangle$, where

$$V' = V \cup \{\langle y, d\rangle\}, \tag{9}$$
$$A' = A \cup \{\langle\langle y_i, a_i\rangle, \langle y, d\rangle\rangle \,|\, i = 1, \ldots, k\}, \tag{10}$$

is a T-based $\mathbf{L}^*$-derivation DAG.

Remark 2. (a) As one can see, the definition of T-based $\mathbf{L}^*$-derivation DAGs is recursive. The base step says that any set of unconnected vertices is a T-based $\mathbf{L}^*$-derivation DAG. The meaning of the vertices is the following: if $\langle y, a\rangle \in V$, we can interpret the fact that the attribute y is assumed valid at least to degree a.

(b) In the second step, the definition postulates that more complex T-based $\mathbf{L}^*$-derivation DAGs result from simpler ones by adding a vertex and arcs leading from vertices related to antecedents of SBFDs from T.

If $\mathbf{D}$ is a T-based $\mathbf{L}^*$-derivation DAG, we let $\mathbf{D}(y) = \bigvee\{a \in L \,|\, \langle y, a\rangle \in V\}$ and call $\mathbf{D}(y)$ the *yield of* $\mathbf{D}$ *on* y. Clearly, the yield of $\mathbf{D}$ corresponds to (7), i.e., we can interpret it as the degree to which y is known to be valid according to $\mathbf{D}$. Moreover, $\langle y, a\rangle \in V$ is called an *initial vertex* of $\mathbf{D}$ is $\langle y, a\rangle$ has no incoming arcs (i.e., no arc in $\mathbf{D}$ enters $\langle y, a\rangle$).

The following notions introduces derivation digraphs related to SBFDs:

Definition 3 (T-based $\mathbf{L}^*$-derivation DAG for $E \Rightarrow F$). Let $\mathbf{D} = \langle V, A\rangle$ be a T-based $\mathbf{L}^*$-derivation DAG. Then $\mathbf{D}$ is called a T-based $\mathbf{L}^*$-derivation DAG for $E \Rightarrow F$ if $\{\langle y, E(y)\rangle \,|\, y \in Y \text{ and } E(y) > 0\}$ is the set of initial vertices of $\mathbf{D}$ and $\mathbf{D}(y) \geq F(y)$ for all $y \in Y$.

4 Completeness

We now turn our attention to the completeness by which we mean a characterization of the semantic entailment by the existence of $\mathbf{L}^*$-derivation DAGs.

Theorem 1. *Let T be a theory. If $T \vdash A \Rightarrow B$, then there is a T-based $\mathbf{L}^*$-derivation DAG for $A \Rightarrow B$.*

Theorem 2. *Let T be a theory. If there is a T-based $\mathbf{L}^*$-derivation DAG for $A \Rightarrow B$, then $T \vdash A \Rightarrow B$.*

The next theorem follows from Theorem 1 and Theorem 2 and completeness of fuzzy attribute logic, see [7].

Theorem 3. *If* $\mathbf{L}$ *is finite, then* $||A \Rightarrow B||_T = 1$ *iff there is a* T*-based* $\mathbf{L}^*$*-derivation DAG for* $A \Rightarrow B$*.*

Furthermore, we can express the graded-style completeness as follows:

Theorem 4. *If* $\mathbf{L}$ *is finite, then* $||A \Rightarrow B||_T$ *is the greatest degree* $a \in L$ *such that there is a* T*-based* $\mathbf{L}^*$*-derivation DAG for* $A \Rightarrow a \otimes B$*.*

Remark 3. From the point of view of the syntax, functional dependencies are the same formulas as attribute implications in formal concept analysis [9] but their interpretation is different. An interesting property is that both the different interpretations of the if-then rules yield the same notion of semantic entailment. As a result, one can use a single inference system for reasoning with both attribute implications and functional dependencies. The same hold in fuzzy setting for SBFD's and graded attribute implication in fuzzy formal concepts analysis [7]. Thus the developed theory can be used for graded attributes implication as well [17].

5 Illustrative Example

In following example, we use a particular finite linearly ordered Łukasiewicz algebra with $L = \{0, 0.1, \ldots, 0.9, 1\} \subseteq [0, 1]$ as a structure of truth degrees. We consider the genuine ordering of real numbers, i.e., $\wedge$ and $\vee$ in (5)–(8) are the usual minima and maxima, respectively. Moreover, $\otimes, \rightarrow$ are Łukasiewicz operations (see [11]) and * is identity.

We demonstrate here how to obtain a T-based $\mathbf{L}^*$-derivation DAG. Assume we have a theory T:

$$T = \{\{^{0.7}/y_3\} \Rightarrow \{^{1}/y_4\}, \{^{0.7}/y_1, {}^{0.8}/y_2\} \Rightarrow \{^{0.6}/y_3\},$$
$$\{^{0.6}/y_1, {}^{0.9}/y_3\} \Rightarrow \{^{0.9}/y_5, {}^{0.1}/y_6\}, \{^{0.5}/y_2, {}^{0.6}/y_4\} \Rightarrow \{^{0.7}/y_6\}\}$$

and consider $A = \{^{0.7}/y_1, {}^{0.7}/y_2\}$, $B = \{^{0.6}/y_4, {}^{0.6}/y_6\}$. Suppose we are interested in whether $||A \Rightarrow B||_T = 1$ or not. According to Theorem 3, $||A \Rightarrow B||_T = 1$ iff there is T-based $\mathbf{L}^*$-derivation DAG for $A \Rightarrow B$. The desired T-based $\mathbf{L}^*$-derivation DAG for $A \Rightarrow B$ can be seen in Figure 1, notice that the set of initial vertices is equal to A and $B \subseteq B_{\mathbf{D}}$. From Theorem 2, we know that $A \Rightarrow B$ is provable from T.

References

1. Armstrong, W.W.: Dependency structures of data base relationships. In: Rosenfeld, J.L., Freeman, H. (eds.) Information Processing 74: Proceedings of IFIP Congress. pp. 580–583. North Holland, Amsterdam (1974)

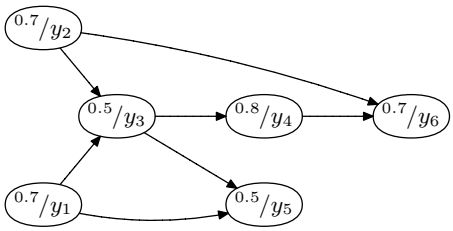


Fig. 1. T-based $\mathbf{L}^*$-derivation DAG for $A \Rightarrow B$.

2. Bang-Jensen, J., Gutin, G.: Digraphs: Theory, Algorithms and Applications. Springer Monographs in Mathematics, Springer (2010)
3. Belohlavek, R., Vychodil, V.: Logical foundations for similarity-based databases. In: LNCS 5667: Database Systems for Advanced Applications: Dasfaa 2009 International Workshop. pp. 137–151. Springer (2009)
4. Belohlavek, R., Vychodil, V.: Query systems in similarity-based databases: logical foundations, expressive power, and completeness. In: Rosenfeld, J.L., Freeman, H. (eds.) Proc. ACM SAC. pp. 1648–1655 (2010)
5. Belohlavek, R.: Fuzzy Relational Systems: Foundations and Principles. Kluwer Academic Publishers, Norwell, MA, USA (2002)
6. Belohlavek, R., Vychodil, V.: Data tables with similarity relations: functional dependencies, complete rules and non-redundant bases. In: Database Systems for Advanced Applications: 11th International Conference, DASFAA 2006, Proceedings
7. Belohlavek, R., Vychodil, V.: Attribute implications in a fuzzy setting. In: Missaoui, R., Schmidt, J. (eds.) Formal Concept Analysis, Lecture Notes in Computer Science, vol. 3874, pp. 45–60. Springer Berlin / Heidelberg (2006)
8. Belohlavek, R., Vychodil, V.: Codd's relational model from the point of view of fuzzy logic. Journal of Logic and Computation 21, 851–862 (Oct 2011)
9. Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1st edn. (1997)
10. Goguen, J.A.: The logic of inexact concepts. Synthese 19, 325–373 (1979)
11. Hájek, P.: Metamathematics of Fuzzy Logic. Kluwer Academic Publishers, Dordrecht, The Netherlands (1998)
12. Hájek, P.: On very true. Fuzzy Sets and Systems 124(3), 329–333 (2001)
13. Holzer, R.: Knowledge acquisition under incomplete knowledge using methods from formal concept analysis: Part I. Fundam. Inform 63(1), 17–39 (2004)
14. Maier, D.: Minimum covers in relational database model. Journal of the ACM 27(4), 664–674 (Oct 1980)
15. Maier, D.: Theory of Relational Databases. Computer Science Pr, Rockville, MD, USA (1983)
16. Pavelka, J.: On fuzzy logic I, II, III. Zeitschrift für Mathematische Logik und Grundlagen der Mathematik 25, 45–52, 119–134, 447–464 (1979)
17. Urbanova, L., Vychodil, V.: Derivation digraphs for graded if-then rules (submitted)

How do you decide? Connecting personal traits with machine-learned behavior.

Wolfgang Heidl WOLFGANG.HEIDL@PROFACTOR.AT
Profactor Research, 4407 Steyr-Gleink, AUSTRIA

In this paper we propose a statistical model that links personal traits to properties of machine-learned behavior models. If successfully validated, the model would be a useful tool in the exploratory phase of behavior research. The model has originally been developed for investigating gender differences in visual inspection (Heidl, 2012). It is organized in two nested levels. At the lower level we model decision behavior of each subject by machine-learned decision trees. Properties of the decision trees are linked to gender at the upper level.

The proposed model is not restricted to the given setting but is applicable to a broad set of tasks for which responses to stimuli can be measured in a controlled experiment. Correlations between task behavior as captured by machine learned models an personal traits different to gender, such as age, education and attitudes can readily be computed. To investigate the wider utility of the proposed model, we have conducted a regression analysis between visual inspection behavior of 100 subjects and a set of 88 personal traits retrieved from a post-experiment questionnaire.

Visual inspection is carried out for quality control of products at the end of a production line. This task is quite often done by women. Their job is to make a quick accept/reject decision and to sort out the bad products. Manufacturing companies often argue that women have more endurance in performing this task and also make decisions with better reproducibility. Motivated from this casually observed gender difference we have carried out a carefully designed experiment to investigate these differences in detail (Heidl et al., 2010; Heidl et al., 2011). The subject's decision behavior has then been modeled by CART decision trees (Breiman et al., 1993). From the trees 15 meta-features (Heidl et al., 2012) have been extracted that give various measures of tree complexity as well as importance of input features for the subject's decisions.

For each of the extracted meta-features we have fit a linear regression model using the personal traits as explanatory variables. Fitting of the regression models has been done with the Lasso algorithm (Tibshirani, 1996). By using the L1-norm of the coefficient vector as regularization penalty, the Lasso combines coefficient shrinkage and subset selection to control the effective number of parameters. The significance of the correlations represented by the regression models is assessed by computing their cross validation (CV) error in a permutation test. The associated p-value is obtained as the fraction of permutations for which the CV-error is lower than on the observed data.

Results show significant correlations to subsets of the 88 personal traits exist for six of the 15 decision model meta-features. Among those meta-features are four that already showed significant gender differences. For those meta-features gender is again contained in the subset of explanatory variables chosen by the Lasso. Most interesting are the significant links identified for two of the meta-features describing input feature importance for the subject's decision for which gender differences are insignificant. In those links gender is indeed not present in the subset of explanatory variables.

The results show that the proposed model allows for generalization from a single, pre-selected trait (gender) to a set of explanatory variables, among which relevant subsets are chosen in CV-guided procedure. It has to be noted that this approach corresponds to a much wider and more exploratory research question than asking for gender differences. As such, for a given sample size lower observed significance levels are to be expected. Intuitively, the more possibly noisy explanatory variables are considered, the more likely it is to observe spurious results. This intuition is formally reflected in the dependence of test statistics on the number of model parameters (Cohen, 1988). Having a tool for exploratory research does not free the practitioner from the necessity of domain knowledge to interpret results. However, given such knowledge, the identified links can act as valuable hints for building hypothesis for later confirmatory research.

References

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1993). *Classification and regression trees*. Boca Raton, FL: Chapman and Hall/CRC.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates. 2 edition.

Heidl, W. (2012). Population decision modeling. *Advances in Knowledge-Based Technologies, Master and PhD Seminar, Summer term 2012, part 1* (pp. 21–28). Hagenberg, Austria.

Heidl, W., Thumfart, S., Lughofer, E., Eitzinger, C., & Klement, E. P. (2010). Classifier-based analysis of visual inspection: Gender differences in decision-making. *Proceedings of SMC2010, IEEE Conference on Systems, Man and Cybernetics* (pp. 113–120). Istanbul, Turkey.

Heidl, W., Thumfart, S., Lughofer, E., Eitzinger, C., & Klement, E. P. (2011). Decision tree-based analysis suggests structural gender differences in visual inspection. *Proceedings of AIA2011, IASTED International Conference on Artificial Intelligence and Applications* (pp. 142–149). Innsbruck, Austria.

Heidl, W., Thumfart, S., Lughofer, E., Eitzinger, C., & Klement, E. P. (2012). Machine learning based analysis of gender differences in visual inspection decision making. *Information Sciences*. In 2nd round of review.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, *58*, 267–288.