ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

Multispectral image characterization by partial generalized covariance

Marc Strickert[a*], Björn Labitzke[a], Andreas Kolb[a], and Thomas Villmann[b]

[a] Institute for Vision and Graphics (IVG), University of Siegen, Germany
[b] Department of Mathematics, University of Applied Sciences Mittweida, Germany
[*] Corresponding author: strickert@informatik.uni-siegen.de

Abstract. A general method is presented for the assessment of data attribute variability, which plays an important role in initial screening of multi- and high-dimensional data sets. Instead of the commonly used second centralized moment, known as variance, the proposed method allows a mathematically rigorous characterization of attribute sensitivity given not only Euclidean distances but partial data comparisons by general similarity measures. Depending on the choice of measure different spectral features get highlighted by attribute assessment, this way creating new image segmentation aspects, as shown in a comparison of Euclidean distance, Pearson correlation and γ-divergence applied to multi-spectral images.

Keywords. Unsupervised attribute assessment, partial generalized correlation, distance pursuit.

Multi- and hyper-spectral image acquisition systems going beyond the RGB color space, such as acquired by liquid crystal tunable filters or imaging mass spectrometry, provide a wealth of data that impose challenges on the analysis, because of the size, the complexity, and the typically non-Euclidean nature of spectral data sets [1]. Recent work allows to effectively address the size problem by employing multi-CPU or GPU-accelerated computing [5]. Non-Euclidean nature is addressed by general data similarity measures like Minkowski distances [2], matrix metrics [3], Pearson correlation [4], or, very recently, divergence measures [6]. In the following metrics, divergence measures, and dissimilarities will be addressed as distances in the non-mathematical intuitive way.

Initial data exploration of yet unknown or unlabeled multi-dimensional data sets often starts with the calculation of basic statistics of the data set attributes, such as histograms or statistical moments such as mean, variance, skewness and kurtosis. Relationships between attributes are usually characterized by the covariance matrix which is also the main ingredient for principal component analysis (PCA), one of the most widely utilized dimension reduction methods. Along with variance and covariance quantities come normalization methods like z-score transformation, i.e. mean-centered attribute divided by attribute variance, and whitening, i.e. linearly de-correlated attributes with identity covariance matrix. All these concepts are based on the implicit assumption of underlying Euclidean data spaces. Particularly the growing popularity of alternative similarity measures requires a reconsideration of variance quantification. The main reason is consistency of the analysis pipeline. For example, if clustering based on Pearson correlation similarity is desired, it might be adverse to choose standard PCA for data preprocessing, or z-score normalization of spectral data might

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

be prohibited if the main processing part utilizes a divergence measure. Divergences are designed for the comparison of non-negative data, and out of a variety we concentrate on γ-divergence-measures, because they include the well-known Kullback-Leibler and Cauchy-Schwarz divergences as special cases.

Here, a very general variational approach is presented for unsupervised assessment of attribute (co-)variability for differentiable metrics or similarity measures taking real-valued vectors as arguments. Application examples with Pearson correlation measure and the flexible γ-divergence-measure are provided in the fruitful domain of multi-spectral image characterization.

1 Assessment of attribute variation

Generally, if two vectors $\mathbf{x}, \boldsymbol{w} \in \mathbb{R}^d$ from a data set are given the key idea pursued here is to measure the minimum efforts needed to transform the variable vector $\boldsymbol{w}$ into a representation of maximum similarity with the target vector $\mathbf{x}$. During that optimum transformation, changes of attributes in $\boldsymbol{w}$ are integrated over and recorded. This procedure, called distance pursuit (DP), can be formalized and solved for each data pair of interest in a mathematically rigorous optimization framework. As a result, the attribute variability explaining a directed relationship of data vectors is quantified for a given data measure. In a summarization step, all recorded attribute changes are turned into a common notion of variability analogous to standard covariance for Euclidean distance.

1.1 Distance pursuit

Formally, even for plain vector pairs it is impossible to find closed form analytical solutions to optimum vector transformation in the general case. Yet, arbitrary good approximations can be obtained in an iterative way outlined in Algorithm 1.

First, a monotonic distance sequence $\boldsymbol{s} = \{s_0 = \mathrm{d}(\mathbf{x}, \boldsymbol{w}), \ldots, s_t = \mathrm{d}_{target}\}$ from the initial vector distance to the maximum possible degree of similarity d_{target} is created with $\mathrm{d}_{target} = 0$ for metrics and divergence measures and

Algorithm 1 distance pursuit $\mathsf{DP}_{\{\boldsymbol{b},g\}}(\mathbf{x}, \boldsymbol{w}, t)$

{input $\mathbf{x}$ target vector ; $\boldsymbol{w}$ source vector ; t number of steps}
$g \leftarrow 0$; $\boldsymbol{b} \leftarrow \mathbf{0}_d$ {initialize target variables}
for $s \leftarrow s_1$ to s_t {visit distance sequence elements, excluding s_0} **do**
  $\boldsymbol{v} \leftarrow \arg\min_{\boldsymbol{w}^*} |s - \mathrm{d}(\mathbf{x}, \boldsymbol{w}^*)|$ {optimum vector at s, starting at $\boldsymbol{w}^* = \boldsymbol{w}$}
  $\boldsymbol{\Delta} \leftarrow \boldsymbol{v} - \boldsymbol{w}$ {differential displacement}
  $\boldsymbol{w} \leftarrow \boldsymbol{w} + \boldsymbol{\Delta}$ {move on to identified location $\boldsymbol{w} = \boldsymbol{v}$}
  $\boldsymbol{b} \leftarrow \boldsymbol{b} + \boldsymbol{\Delta}$ {integrate changes per attribute}
  $g \leftarrow g + \sqrt{\langle \boldsymbol{\Delta} \rangle^2}$ {line integral}
end for
return g, $\boldsymbol{b}$

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

$\mathrm{d}_{target} = 1$ for Pearson correlation. An equidistant sampling of $t = 10$ sub-intervals, such as done in this work, is pragmatic but not mandatory. The vector $\boldsymbol{w}$ moves along the minimum path following increasing steps of similarity towards the target vector $\mathbf{x}$. During the transformation of $\boldsymbol{w}$, two quantities are collected: g, the overall line integral summing up path fragment lengths and $\boldsymbol{b}$, the integrated differential attribute vector. Thus, a piecewise linear approximation in d-dimensional Euclidean space is conducted to effectively assess attribute properties of the data under the distance measure of interest.

Line 4 of the Algorithm 1 provides the important identification of the next position of vector $\boldsymbol{w}$ for getting a bit more similar to vector $\mathbf{x}$. Gradient-based optimization methods can be used for minimizing the distance discrepancies between the current position and the next step of the target distance imposed by line 3. For distances $\mathrm{d}(\mathbf{x}, \boldsymbol{w}^*)$, such as the ones listed in Table 1 being utilized in this work, the gradient of the arg min operation, required for optimization, is

$$\delta = -\mathsf{sign}(s - \mathrm{d}(\mathbf{x}, \boldsymbol{w}^*)) \cdot \partial \mathrm{d}(\mathbf{x}, \boldsymbol{w}^*)/\partial \boldsymbol{w}^* \,. \tag{1}$$

Alternatively, a least squares expression can be employed in line 4, but it tends to generate numerical underflows during convergence.

Common gradient-based methods find zero discrepancy solutions desired in line 4. Yet, since optima for reaching a given similarity s are not unique, only minimum norm results for Δ are valid to get minimum path lengths. Else, for example with Minkowski metrics usually two points along the search line would yield valid optima at distance s to the target vector $\mathbf{x}$. Possible oscillations would be integrated out in line 7 for attribute variability, but the line integral in line 8, based on repeated calculations of $\langle \boldsymbol{\Delta} \rangle^2 := \langle \boldsymbol{\Delta}, \boldsymbol{\Delta} \rangle$, would be over-estimated.

Different gradient-based optimizers were tried using a reference implementation of Minkowski metrics with line integrals being standard Euclidean distance and attribute variability being standard variance, irrespective of the choice of the metric order p. Memory-limited Broyden-Fletcher-Goldfarb-Shanno turned out to provide the best mix of speed, memory requirement, and accuracy, in comparison to full BFGS, conjugate gradients, and steepest gradient descent.

2 Generalized partial covariance

The result of the distance pursuit algorithm is used as building block in a general formula for measuring attribute variability of data set $\mathbf{X} = (x_{lk})_{l=1\ldots n, k=1\ldots d}$ with n data vectors (in rows) containing d attributes. The reformulated text book term of standard variance

$$\sigma_k^2 = \frac{1}{n-1} \cdot \sum_{l=1}^{n} (x_{lk} - \mu_k)^2 = \frac{1}{2 \cdot n \cdot (n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (x_{ik} - x_{jk})^2 \tag{2}$$

can be transformed into the generalized partial covariance expression

$$V_{kl} = \frac{1}{G} \cdot \sum_{i=1}^{n} \sum_{j=1}^{|\mathcal{I}^i|} \mathsf{DP}_{b_k}(\mathbf{x}_i, \mathbf{x}_{\mathcal{I}_j^i}, t) \cdot \mathsf{DP}_{b_l}(\mathbf{x}_i, \mathbf{x}_{\mathcal{I}_j^i}, t) \,, \quad k, l = 1 \ldots d \,. \tag{3}$$

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5. Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

These values V_{kl} quantify the overall attribute contribution of all desired pairs of data vectors $\mathbf{x}_i$ connected to vectors indexed by $\mathcal{I}^i$. For each pair, the contribution of attributes l and k is calculated using the DP algorithm. Thus, Equation 3 quantifies attribute variability given data connectivity and distance measure.

The normalization constant $G = 2 \cdot (-z + \sum_{i=1}^{n} |\mathcal{I}^i|)$ is twice the number of pairwise comparisons, excluding the number z of non-contributing self-comparisons $i = \mathcal{I}^i_j$. For all pairs of data $\mathcal{I}^i = \{1 \dots n\}$ for $i = 1 \dots n$ this leads to $G = 2 \cdot (-n + n^2)$ which yields an unbiased estimation of the variance $\sigma_k^2 = V_{kk}$ of the k-th attribute for the Euclidean distance. This case allows a connection to Equation 2, because the DP algorithm yields the simple difference $\boldsymbol{b} = \mathbf{x}_i - \mathbf{x}_j$.

The partial generalized covariance matrix $\boldsymbol{V}$ is necessarily symmetric due to the commutative calls to DP in Equation 3. This matrix can be used to calculate the partial generalized correlation matrix $\boldsymbol{R}$ according to the formula

$$R_{kl} = V_{kl} / \sqrt{V_{kk} \cdot V_{ll}} \,. \tag{4}$$

For all data pairs and Euclidean distance $\boldsymbol{R}$ contains exactly what could be calculated much more efficiently with pairwise Pearson correlation from Table 1. Else, the runtime complexity of DP is $\mathcal{O}(t \cdot G \cdot d^2)$ for covariance and $\mathcal{O}(t \cdot G)$ for variance which can easily become a bottleneck. Yet, the proposed method allows custom data connectivities and similarity measures. Euclidean assumptions like mean values $\mu_{\mathbf{x}}$ of data vectors are not needed, because distance-specific centers of gravity are implicitly computed by the double sum in Equation 3.

3 Application to multi-spectral image data

Multi-spectral images were acquired by using the Nuance EX camera utilizing its liquid crystal tunable filter. The filter operates in a visual and near infrared wavelength range from $450nm$ to $950nm$ and has a full width at half maximum characteristic of $10nm$, leading to 51 channels per pixel. Food analysis is important for quality control and nutraceutics, thus, a representative multi-spectral image of a sliced cucumber fruit at a resolution of 416x408 pixels is taken as analysis target. The two targeted complementary perspectives for partial generalized attribute variance assessment are: frequency channel correlations involving pairs of pixels and per pixel variance involving pairs of channels.

Pixel variance is assessed for adjacent channels using the index set $\mathcal{I}^i = i \pm 1$ clipped to valid indices. This requires DP for 100 pairs of channels represented by one of the 51 frequency-specific monochrome image layers, resulting in an image of partial generalized variance per pixel. Involving adjacent channels uses only local differences for detecting more subtle relative changes than by global comparisons of distant and different frequency channels. This is shown in the first two images in the top row of Figure 1 changing from standard (i.e. global) to partial (i.e. local) variance. For Pearson correlation and γ-divergence the shadow cast rather than the cucumber peel gets emphasized. This is, because both measures are β-scaling invariant to vectors $\beta \cdot \mathbf{x}$ and thus optimally aligned w.r.t. scaling, i.e. being able to highlight variability in non-alignable structures like noise-containing shadows, which is useful for filtering different aspects.

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5.
Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

Measure	Formula $\mathrm{d}(\mathbf{x}, \boldsymbol{w}) =$
Minkowski distance	$(\sum_{i=1}^{d} \lvert x_i - w_i \rvert^p)^{1/p}$
Pearson correlation	$\frac{\langle \mathbf{x}-\mu_{\mathbf{x}},\, \boldsymbol{w}-\mu_{\boldsymbol{w}} \rangle}{\sqrt{\langle \mathbf{x}-\mu_{\mathbf{x}} \rangle^2 \cdot \langle \boldsymbol{w}-\mu_{\boldsymbol{w}} \rangle^2}}$
γ-divergence	$\log \frac{\langle \mathbf{x}^{\gamma+1}, \mathbf{1} \rangle^{1/(\gamma\cdot(\gamma+1))} \cdot \langle \boldsymbol{w}^{\gamma+1}, \mathbf{1} \rangle^{1/(\gamma+1)}}{\langle \mathbf{x}, \boldsymbol{w}^{\gamma} \rangle^{1/\gamma}}$

Measure	Derivative $\partial \mathrm{d}(\mathbf{x}, \boldsymbol{w}) / \partial \boldsymbol{w} =$
Minkowski distance	$-\mathsf{sign}(\mathbf{x} - \boldsymbol{w}) \circ \lvert \mathbf{x} - \boldsymbol{w} \rvert^{p-1} \cdot (\sum_{i=1}^{d} \lvert x_i - w_i \rvert^p)^{\frac{1}{p}-1}$
Pearson correlation	$\left(\frac{\mathbf{x}-\mu_{\mathbf{x}}}{\langle \mathbf{x}-\mu_{\mathbf{x}},\, \boldsymbol{w}-\mu_{\boldsymbol{w}} \rangle} - \frac{\boldsymbol{w}-\mu_{\boldsymbol{w}}}{\langle \boldsymbol{w}-\mu_{\boldsymbol{w}} \rangle^2} \right) \cdot \frac{\langle \mathbf{x}-\mu_{\mathbf{x}},\, \boldsymbol{w}-\mu_{\boldsymbol{w}} \rangle}{\sqrt{\langle \mathbf{x}-\mu_{\mathbf{x}} \rangle^2 \cdot \langle \boldsymbol{w}-\mu_{\boldsymbol{w}} \rangle^2}}$
γ-divergence	$\frac{\boldsymbol{w}^{\gamma}}{\langle \boldsymbol{w}^{\gamma+1}, \mathbf{1} \rangle} - \frac{\mathbf{x} \circ \boldsymbol{w}^{\gamma-1}}{\langle \mathbf{x}, \boldsymbol{w}^{\gamma} \rangle}$

Table 1: Distance measures and their derivatives. $\mu_{\mathbf{x}}$ and $\mu_{\boldsymbol{w}}$ denote the mean values of vectors $\mathbf{x}$ and $\boldsymbol{w}$, respectively; sign refers to the signs of vector components; $\circ$ is the component-wise Hadamard product; vector powers operate as powers on each vector component; $\mathbf{1}$ is the d-dimensional vector of ones; $\langle \cdot, \cdot \rangle$ is the cross product of two vectors; $\langle \cdot \rangle^2$ is the sum of squares of vector components. For **Minkowski distance**, real values $p \geq 1$ are allowed. For **Pearson correlation** and its derivative see [4]. For **γ-divergence** the expression $\mathrm{d}(\mathbf{x}, \boldsymbol{w})$ with vector arguments refers to the discrete version of $\mathsf{D}(\mathbf{x} || \boldsymbol{w})$ with density distributions; the measure and its derivative are discussed in [6], and a choice of $\gamma \in [0; 1]$ is recommended.

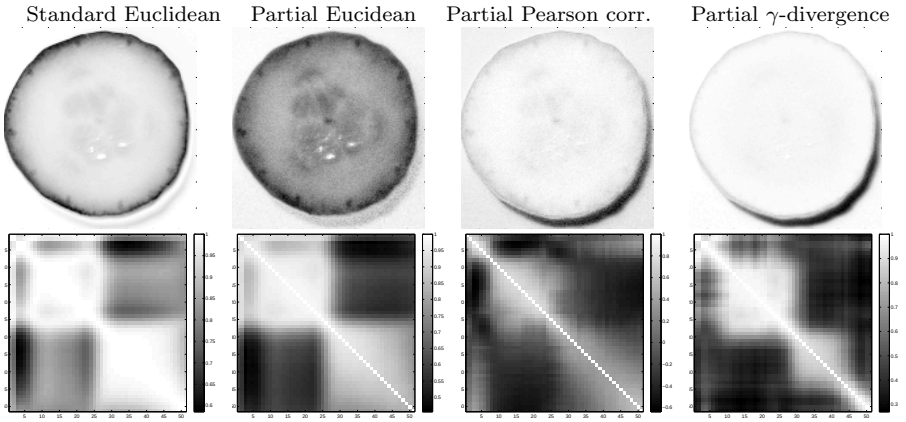

Fig. 1: Cucumber pixel variance (top row, darker means higher) and frequency channel correlation (bottom row, brighter means higher) for, f.l.t.r., standard (all pairs) Euclidean distance and partial (adjacent data pairs) Euclidean distance, Pearson correlation, and γ-divergence for $\gamma = 0.1$.

ESANN 2011 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 27-29 April 2011, i6doc.com publ., ISBN 978-2-87419-044-5. Available from http://www.i6doc.com/en/livre/?GCOI=28001100817300.

Channel correlations are assessed in the 8-neighborhood of each pixel. The resulting partial generalized correlation matrices of all pairs of 51 channels are shown in the bottom row of Figure 1. Generally, there are two major clusters split in the middle (at about channel 25), and the first two channels are a bit outstanding. The range for partial Euclidean correlation (0.45–1) is a bit larger than for global correlation (0.59–1). Even richer are the partial correlation patterns for Pearson correlation (range: -0.6–1) and for γ-divergence (range: 0.29–1) which reveals four clusters along the diagonal line (two large, two small).

4 Conclusions

A generalized way has been presented for calculating partial covariance and correlation of attributes in multidimensional data for custom data similarity measures and connectivity structures. It contains standard covariance and correlation of Euclidean data as special cases. For large data sets, the current implementation in R, available for download at `http://mi.informatik.uni-siegen.de/projects_data/dp.zip`, requires extraordinary runtime, because multiple nonlinear optimizations are required for each compared vector pair. A parallel version is targeted in near future.

The application to multi-spectral data has shown the ability to provide richer partial covariance structures by utilizing local comparisons. Furthermore, data invariance properties of Pearson correlation and γ-divergence allowed to identify shadow areas rather than interior fruit segments being highlighted by Euclidean distance. Unsupervised and supervised data analysis tasks along data processing pipelines with other than Minkowski distances will benefit from proposed method, because the measure-specific assessment of attribute variability is a very fundamental requirement. Several tightly connected methods like generalized whitening operation and PCA are yet to be exploited in future research.

We thank three anonymous reviewers for their constructive comments. Research is founded by DFG Graduiertenkolleg 1564 'Imaging new Modalities'.

References

[1] C.-I. Chang. *Hyperspectral Data Exploitation: Theory and Applications*. Wiley, 2007.

[2] J. Kivinen, M. K. Warmuth, and B. Hassibi. The p-norm Generalization of the LMS Algorithm for Adaptive Filtering. *IEEE Trans. on Signal Processing*, 54:1782–1793, 2006.

[3] P. Schneider, K. Bunte, H. Stiekema, B. Hammer, T. Villmann, and M. Biehl. Regularization in matrix relevance learning. *IEEE Trans. on Neural Netw.*, 21(5):831–840, 2010.

[4] M. Strickert, F.-M. Schleif, T. Villmann, and U. Seiffert. *Similarity-Based Clustering - Recent Developments and Biomedical Applications*, volume 5400 of *Lecture Notes in Computer Science*, chapter Unleashing Pearson Correlation for Faithful Analysis of Biomedical Data, pages 70–91. Springer, 2009.

[5] R. Tsuchiyama, T. Nakamura, T. Iizuka, A. Asahara, and S. Miki. *The OpenCL Programming Book*. Fixstars Corporation, 2010.

[6] T. Villmann, S. Haase, F.-M. Schleif, B. Hammer, and M. Biehl. The mathematics of divergence based online learning in vector quantization. In F. Schwenker and N. El Gayar, editors, *Artificial Neural Networks in Pattern Recognition*, volume 5998 of *LNCS*, pages 108–119. Springer Berlin / Heidelberg, 2010. 10.1007/978-3-642-12159-3_10.