Defect Detection using Nonparametric Regression

Siana Halim
Industrial Engineering Department-Petra Christian University
Siwalankerto 121-131 Surabaya- Indonesia
halim@petra.ac.id

Abstract: To compare signals, we first model them as nonparametric regression setup, we then wish to test either those signals are significantly the same against they are significantly different. To perform a test, first we need to measure the distance between two nonparametric regression and use this distance as test statistic for testing the null hypothesis. Typically, the distribution of test statistic under the hypothesis null is not known. This problem can be handled by deriving the asymptotic approximation for unknown distribution that holds for sample size infinitely. However, this approach practically cannot be applied in signal, since the structure of the data is frequently too complicated. We then used bootstrap tests, we move from our original data to the bootstrap world of pseudo data vector or resample. We apply this method to image processing for detecting defect on the texture. We model the images as 2D Gasser-Mueller Kernel Density with rotational-ellipsoidal support function, to estimate the regression function. Moreover, we let the errors correlated in their neighborhoods. We use standardized the modification of the Mallows distance between these two estimates, to test the hypothesis and construct spatial bootstrap to get the distribution of the test statistic. The spatial bootstrap is needed to preserve the bound of a pixel to its neighborhood.

Keywords: 2D nonparametric regression, testing hypothesis for signals, bootstrap.

AMS subject: 62P30

Introduction

Comparing signals have been studied for a long time and applied in many disciplined of science. In this study we applied the comparing signal for modeling defect on the texture, especially the pattern defect on the texture. The idea of detecting defect on the texture`s pattern is the same as we compare the not defected “signal” or series of the texture to the defected one.

We first model those signals as nonparametric regression setup, we then wish to test either those signals are significantly the same against they are significantly different. To perform a test, first we need to measure the distance between two nonparametric regression and use this distance as test statistic for testing the null hypothesis.

Methods

Hypotheses Analysis

To compare those signals, we first model them as the following nonparametric regression setup, for simplicity we assume that the size of the image is n by n.

$$Y_{ij} = m^{I}\ x_{ij}\ + \varepsilon_{ij},$$
$$Y_{ij} = m^{II}\ x_{ij}\ + \varepsilon_{ij},\ i,j = 1,\dots,n \tag{1}$$

where the $\varepsilon_{11},\dots,\varepsilon_{nn},\varepsilon_{11},\dots,\varepsilon_{nn}$ are independent with mean zero and finite variance, $Var\ \varepsilon_{ij}\ = Var\ \varepsilon_{ij}\ = \sigma^2(x_{ij})$ and uniformly bounded fourth moments $E\varepsilon_{ij},\ E\varepsilon_{ij} \leq C < \infty, i,j = 1,\dots,n$.
For sake of simplicity, we only consider the case of equidistant x_i on a compact set, say [0,1]. (Detail of nonparametric regression can be seen in [4]). The hypotheses analysis in this work is the extended version of the hypothesis analysis for comparing signal and images, developed by Franke and Halim [3,4]. Instead of doing in one dimensional setting, we now work for two dimensional one.

We wish to test either those signals are significantly the same against they are significantly different, i.e.,
$H_0: m^{I}\ x_i\ = m^{II}\ x_i\ = m\ x_i\ , i = 1,\dots,n$ against
$H_1: m^{I}\ x_i\ \neq m^{II}\ x_i$ for some i

Kernel Smoothing

To model an image as a regression, first, we consider an equidistant grid of pixels

$$x_{ij} = \ \frac{i}{n} - \frac{1}{2n}, \frac{j}{n} - \frac{1}{2n}\ = \frac{1}{n}\ i,j\ - \frac{1}{2n}; i,j = 1,\dots,n \tag{2}$$

in the unit square $\mathrm{A} = [0,1]^2$ and a function $m: [0,1]^2 \to \mathbb{R}$ to be estimated from data, i.e.. the gray levels of the image as follows:

$$Y_{ij} = m\ x_{ij}\ + \varepsilon_{ij}; i,j = 1,\dots,n \tag{3}$$

Where the noise is part of a stationary random field $\varepsilon_{ij}, -\infty < i,j < \infty$, with zero-mean and finite variance.

We use the Gasser-Muerller-type kernel to estimate$m(x)$. For that purpose we decompose A into squares $A_{ij} =\ x \in A; \frac{i-1}{n} \leq u_1 \leq \frac{i}{n}; \frac{j-1}{n} \leq u_2 \leq \frac{j}{n}\ , 1 \leq i,j \leq n$
such that x is the mid point of A_{ij}, then estimate m using:

$$m\ x,h\ =\ {}_{i,j=1}^{n}\ {}_{A_{ij}} K_h\ x - u\ du\, Y_{ij} \tag{4}$$

where $K: \mathbb{R}^2 \to \mathbb{R}$ is a given kernel function and for the bandwidth vector $h = (h_1, h_2)$, the rescaled kernel is$K_h = \frac{1}{h_1 h_2} K\ \frac{u_1}{h_1}, \frac{u_2}{h_2}$

Statistical Properties of Smoothers

We assume that the design points are generated by a positive, Lipschitz continuous density function f, and the following conditions on the kernel K

(K1) K has support (-1,1)

(K2) K is Lipschitz continuous

(K3) $_{-1}^{1} K \ u \ du = 1$

(K4) $_{-1}^{1} u K \ u \ du = 0$

To simplify notation, we write the index in the following way, $z = (i, j)$ such that, (4) can be written as

$$Y_z = m \ x_z \ + \varepsilon_z, z \in I_n = \{1, \dots, n\}^2$$

Letε_z, z $\in \mathbb{Z}^2$, is strictly stationary random field on the integer lattice with $\mathbb{E}\varepsilon_z = 0$, $Var \ \varepsilon_z = r \ 0 \ < \infty$ and autocovariances $r \ z \ = cov \ \varepsilon_{z'+z} \ , z, z' \in \mathbb{Z}^2$

<u>Mean Square Error of Gasser-Mueller Estimators</u>

Suppose$H = \begin{matrix} h_1 & 0 \\ 0 & h_2 \end{matrix}$, i.e. the bandwidth matrix is diagonal then, the bias is

$$\mathbb{E}m \ x, H \ - m \ x \ = \tfrac{1}{2} V_K \ \ h_1^2 m^{\ 2,0} \ \ x \ + h_2^2 m^{\ 0,2} \ \ x \ ^2 + o \ \ h \ ^2 + O \ \tfrac{1}{n} \qquad (5)$$

and the variance will be

$$\mathrm{Var}\, m \ x, H \ = \frac{1}{n^2 h_1 h_2} f \ 0,0 \ Q_K + O \ \frac{h_1 + h_2}{n^3 h_1^2 h_2^2} \qquad (6)$$

where the constant $V_K = \ u_1^2 K \ u \ du = \ u_2^2 K \ u \ du$ and $Q_K = \ K^2 \ u \ du$ depend only on the kernel K.m is twice continuously differentiable, $m^{\ \alpha,\beta} \ \ x \ = \frac{\partial^\alpha}{\partial x_1^\beta} \frac{\partial^\alpha}{\partial x_2^\beta} m \ x \ , \alpha, \beta \geq 0$. Let $f(\omega, \omega')$ denote the spectral density of the random field ε_{ij}, i.e., the Fourier transform of the autocovariances, we have $f \ 0,0 \ = \ _{i,j=-\infty}^{\infty} r \ i, j \ = \ _z r(z)$. (Franke *et al.*[5])

Selecting the smoothing parameter

The performance of the estimate depends crucially on the bandwidth $h = (h_1, h_2)$of $m(x, h)$. We consider the problem of selecting an optimal global bandwidth h_{opt} such that mean integrated square error

$$\mathrm{mise}\, m \ ., h \ = \ \mathrm{E} \ m \ x, h \ - m \ x \ ^2 w_h \ x \ dx$$

is asymptotically minimized. $w_h$is a weight function with support $x; h_1 \leq x_1 \leq 1 - h_1, h_2 \leq x_2 \leq 1 - h_2$ and $w_h \ x \ dx = 1$, which we employ, for sake of simplicity, to avoid boundary effects.

Asymptotic mean square error (amse) $m(., h)$ is minimized for $h_{ASY} = \ h_{1ASY}, h_{2ASY}$ given by

$$h_{1ASY}^{12} = \ \frac{f(0,0) Q_k}{2n^2 V_k^2}^{\ 2} \ \frac{m^{\ 0,2} \ (x)}{m^{\ 2,0} \ (x)^{\ 5}}$$

$$h_{2ASY}^{12} = \ \frac{f(0,0) Q_k}{2n^2 V_k^2}^{\ 2} \ \frac{m^{\ 2,0} \ (x)}{m^{\ 0,2} \ (x)^{\ 5}} \qquad (7)$$

The asymptotic mean integrated square error (mise) is minimized for $h_{IASY} = (h_{1IASY}, h_{2IASY})$ given by

$$h_{1IASY}^6 = \frac{f(0,0) Q_K}{n^2 V_K^2} \ \frac{I_{02}}{I_{20}}^{\frac{3}{4}} \ \frac{1}{\overline{I_{20} I_{02}} + I_{11}}$$

$$h_{2IASY}^{6} = \frac{f(0,0)Q_K}{n^2 V_K^2}\left(\frac{I_{20}}{I_{02}}\right)^{\frac{3}{4}} \frac{1}{\sqrt{I_{20}I_{02}} + I_{11}} = \left(\frac{I_{20}}{I_{02}}\right)^{\frac{3}{2}} h_{1IASY} \tag{8}$$

Where $I_{kl} = \int \left(m^{(2,0)}(x)\right)^{k} \left(m^{(0,2)}(x)\right)^{l} w_h(x)\, dx, 0 \leq k, l \leq 2$

Estimating the second derivative of m

$$m^{(2,0)}(x,h) = \sum_{i,j=1}^{n} \int_{A_{ij}} K_h^{(2,0)}(x-u)\, du\, Y_{ij}$$

$$m^{(0,2)}(x,h) = \sum_{i,j=1}^{n} \int_{A_{ij}} K_h^{(2,0)}(x-u)\, du\, Y_{ij} \tag{9}$$

where $K_h^{(2,0)}(u) = \frac{1}{h_1^3 h_2}\frac{\partial^2}{\partial u_1^2}K\left(\frac{u_1}{h_1},\frac{u_2}{h_2}\right)$; $K_h^{(0,2)}(u) = \frac{1}{h_1 h_2^3}\frac{\partial^2}{\partial u_2^2}K\left(\frac{u_1}{h_1},\frac{u_2}{h_2}\right)$

Plugging the derivative estimate into the integrands, we get the estimate of I_{kl}.

Performing Test

To perform a test, first we need to measure the distance between $\mu_h^I(x)$ and $\mu_h^{II}(x)$ and use this distance as test statistic for testing the null hypothesis.Following, Haerdle and Mammen [7], we use standardized L_2-distance between these two estimates, i.e.

$$T_n = n\sqrt{\overline{h}} \int \left(\mu_h^I(x) - \mu_h^{II}(x)\right)^2 dx \tag{10}$$

*The modifi*cation of the Mallows distance (Haerdle and Mammen[9])

$$d(\mu,\nu) = Inf_{X,Y}\left\{\mathbb{E}\|X-Y\|^2 \wedge 1 : \mathcal{L}(X) = \mu, \mathcal{L}(Y) = \nu\right\} \tag{11}$$

Convergence in this distance is equivalent to weak convergence.

Testing withBootstrap

We have to decide either those signals are significantly the same (i.e., there is no defect present on a surface) against they are significantly different (i.e., the defect presents on a surface). Typically, a test is performed by calculating some function $T(\mathrm{Y})$ of the data and comparing it with some bound C_α, chosen as the $(1-\alpha)$ quantile of the distribution of $T(\mathrm{Y})$ under the hypothesis H_0. If $T(\mathrm{Y}) \leq C_\alpha$, we accept c as compatible with the data, otherwise we reject it in favor of H_1. αis the prescribed probability of an error of the first kind, i.e., under the H_0, we have $pr(T(Y) > C_\alpha) = \alpha$. Now, constructing the test becomes a problem of determining C_α. However, the distribution of test statistic $T(Y)$ under $H_0$is not known. The classical approach to handle this problem is by deriving the asymptotic approximation for unknown distribution that holds for sample size $N \to \infty$. However, this approach practically cannot be applied in signal and image analysis, since the structure of the data is frequently too complicated.

We then used bootstrap tests, we move from our original data $Y$to the bootstrap world of pseudo data vector or resample Y^*. The resample Y^* may be artificially generated from the original data and has a similar random structure as Yitself. Then, we consider the test statistic $T(Y^*)$ calculated from the bootstrap data Y^* and determine the $(1-\alpha)$-quantile $C_\alpha^*$of its distribution: $pr^*\{T(Y^*) > C_\alpha^*\} = \alpha$, where pr^* denotes the conditional probability given the data Y. (Franke and Halim [4])

The $(1-\alpha)$-quantile C_α^* can be computed numerically using Monte Carlo simulation as follows

1. generate a realization $Y^*(b)$ of the bootstrap data and then calculate $T_b^* = T(Y^*(b))$ repeat for $b = 1, \dots, B$
2. order $T_1^*, \dots, T_B^*$ such that $T_{(1)}^* \leq \cdots \leq T_{(B)}^*$
3. set $C_{\alpha,B}^* = T^*_{([(1-\alpha)B])}$, where $[x]$ denotes the largest integer $\leq x$.

The applicability of the bootstrap data Y^* depends on the way the bootstrap data Y^* are generated as well as the test statistic $T(Y)$ considered. To construct the Y^* for the image, first, we estimated the residual as follow

$$\varepsilon_{ij} = Y_{ij} - \mu^I(x_{ij}) \;\; ; \quad \varepsilon_{ij} = Y_{ij} - \mu^{II}(x_{ij}) \tag{12}$$

centering the residual by their sample mean, we achieve

$$\varepsilon_{ij}^0 = \varepsilon_{ij} - \frac{1}{n}\sum_{i,j=1}^{n} \varepsilon_{ij}$$
$$\varepsilon_{ij}^0 = \varepsilon_{ij} - \frac{1}{n}\sum_{i,j=1}^{n} \varepsilon_{ij} \tag{13}$$

We, then construct out bootstrap samples by

$$Y_{ij}^* = \mu_g^I(x_{ij}) + \varepsilon_{ij}^{*0}; \;\; Y_{ij}^* = \mu_g^{II}(x_{ij}) + \varepsilon_{ij}^{*0} \tag{14}$$

where $\varepsilon_{ij}^{*0}, \varepsilon_{ij}^{*0}$ are the centering residual.

For the construction of $\varepsilon_{ij}^{*0}, \varepsilon_{ij}^{*0}$; we constructed using spatial bootstrap to preserved the bound of a pixel to its neighborhood. First, we compute the spatial covariance matrix of ε_{ij} and ε_{ij} and generated both bootstrap residual of them based on that bound.

The spatial covariance of ε_{ij} and ε_{ij}, is computed between a pair of ε_{ij} and ε_{ij} respectively located at points separated by the distance h. The covariance function can be written as a product of a variance parameter, σ^2 times a positive definite correlation function $\rho(h)$, i.e., $Cov(h) = \sigma^2\rho(h)$.

Denote φ the basic parameter of the correlation function and name it the *range parameter*. Some of the correlation functions will have an extra parameterκ, the *smoothness parameter*. $K_\kappa(x)$ denotes the modified Bessel function of the third kind of order *kappa*. In the equations below the functions are valid for $\varphi > 0$and $\kappa > 0$, unless stated otherwise (Diggle and Ribeiro [1]).

Cauchy

$$\rho(h) = \left[1 + \left(\frac{h}{\varphi}\right)^2\right]^{-\kappa}$$

Generalized Cauchy

$$\rho(h) = \left[1 + \left(\frac{h}{\varphi}\right)^{\kappa 2}\right]^{-\left(\frac{\kappa 1}{\kappa 2}\right)}, \kappa 1 > 0, 0 < \kappa 2 \leq 1$$

Circular

Let $\theta = \min\left(\frac{h}{\varphi}, 1\right)$, $\gamma(h) = \frac{2}{\pi}\left(\theta\sqrt{1-\theta^2} + \sin^{-1}\sqrt{\theta}\right)$

Then, the circular model is given by

$$\rho(h) = \begin{cases} 1 - \gamma(h) & \text{if } h < \varphi \\ 0 & \text{otherwise} \end{cases}$$

Cubic $\rho(h) = 1 - 7\left(\frac{h}{\varphi}\right)^2 - 8.75\left(\frac{h}{\varphi}\right)^3 + 3.5\left(\frac{h}{\varphi}\right)^5 - 0.75\left(\frac{h}{\varphi}\right)^7$ if $h < \varphi$, 0 otherwise

Gaussian $$\rho(h) = \exp\left(-\left(\frac{h}{\varphi}\right)^2\right)$$

Exponential $$\rho(h) = \exp(-h/\varphi)$$

Matern $$\rho(h) = \frac{1}{2^{\kappa-1}\Gamma(\kappa)}\left(\frac{h}{\varphi}\right)^{\kappa} K_{\kappa}\left(\frac{h}{\varphi}\right)$$

Spherical $\rho(h) = 1 - 1.5\left(\frac{h}{\varphi}\right) + 0.5\left(\frac{h}{\varphi}\right)^3$ if $h < \varphi$, 0 otherwise

In this work we chose the correlation $\rho(h)$ as Gaussian model.

Now, the bootstrap test statistics can be constructed as follows (Franke and Halim [3,4]).

$$T_n = nh^{1/2}\int\left(\mu_h^I(x) - \mu_h^{II}(x)\right)^2 dx \quad (15a)$$

$$\sim h^{1/2}\sum_{i,j=1}^{n}\left(\mu_h^I(x_{ij}) - \mu_h^{II}(x_{ij})\right)^2 \quad (15b)$$

$$= \frac{1}{nh^{1/2}}\sum_{i,j=1}^{n}\left(\sqrt{nh}\left(\mu_h^I(x_{ij}) - \mu_h^{II}(x_{ij})\right)\right)^2 \quad (15c)$$

$$= \frac{1}{nh^{1/2}}\sum_{i,j=1}^{n}\left(\sqrt{nh}\left(\mu_h^I(x_{ij}) - m(x_{ij}) + m(x_{ij}) - \mu_h^{II}(x_{ij})\right)\right)^2 \quad (15d)$$

$$\sim h^{1/2}\sum_{i,j=1}^{n}\left(\mu_h^I(x_{ij}) - \mu_g^I(x_{ij}) + \mu_g^{II}(x_{ij}) - \mu_h^{II}(x_{ij})\right)^2 \quad (15e)$$

Under the hypothesis H_0, we use two forms of the test statistics based on (15b) or (15d)with the bootstrap samples. From now on, we call them as $T1_n$ and $T2_n$ respectively, and we set

$$t1_n^* = h^{1/2}\sum_{i,j=1}^{n}\left(\mu_h^{*I}(x_{ij}) - \mu_h^{*II}(x_{ij})\right)^2 \quad (16)$$

$$t2_n^* = h^{1/2}\sum_{i,j=1}^{n}\left(\mu_h^{*I}(x_{ij}) - \mu_g^{*I}(x_{ij}) + \mu_g^{*II}(x_{ij}) - \mu_h^{*II}(x_{ij})\right)^2 \quad (17)$$

Using one of these two functions then we can set the $C_{\alpha,B}^* = t1_{(\lfloor(1-\alpha)B\rfloor)}^*$ or $C_{\alpha,B}^* = t2_{(\lfloor(1-\alpha)B\rfloor)}^*$ and deduce either the hypothesis is rejected (the defect presents in the image) or failed to reject (no defect presents in the image).

Result and Discussion

Some examples of pattern`s defect detection are given in Figure 1.

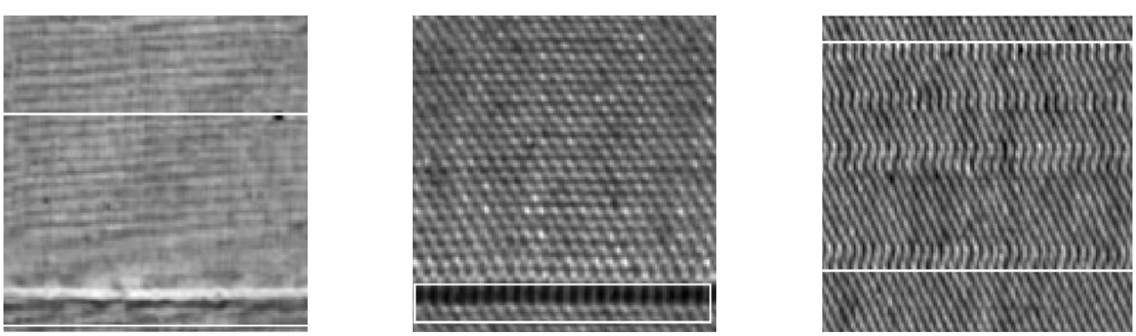

Figure 1. Some examples of pattern`s defect detection

Conclusion and Remark

So far, the methods presented here can handle pattern defect detection. However, there are some limitations that these methods cannot overcome, and therefore should be handled for the future work, i.e., capturing many types of defect on a single texture; capturing many defects on several location of large texture.

References

1. Diggle, P. J. and Ribeiro, P. J., *Model-Based Geostatistics*, Springer, New York, 2007
2. Efron, B., and Tibshirani, R. J., *An Introduction to the Bootstrap*. Chapman & Hall, 1994.
3. Franke, J., and Halim, S., Wild Bootstrap Tests: Regression Models for Comparing Signal and Images, *IEEE Signal Processing Magazine,* July 2007, pp. 31-37.
4. Franke, J., and Halim, S., A Bootstrap Test for Comparing Images in Surface Inspection, presented at DFG-Priority Program 1114 Technical Report 150. [Online] Available: www.math.uni-bremen.de/zetem/DFG-Schwerpunkt/
5. Franke, J., Halim, S. and numerous co-authors, Structural Adaptive Smoothing Procedure, in *Understanding Complex System*, Springer: Heidelberg, 2008.
6. Haerdle, W.; *Applied Nonparametric Regression*; Springer, [Online] Available: http://fedc.wiwi.hu-berlin.de/xplore/ebooks/html/anr/
7. Haerdle, W. and Mammen, E.; Comparing Nonparametric versus Parametric Regression Fits. *The Annals of Statistics,* 12(4), 1993, pp. 1926-1947.
8. Tang, L., Schucany, W. R., Woodward, W. A., and Gunst, R. F., *A Spatial Bootstrap*, Technical Report, Department of Statistical Science, University of Southern Methodist, [Online] Available: *http://smu.edu/statistics/techreports/tr337.pdf*