AIP Conference Proceedings

Sensors closeness test based on an improved [0, 1] bounded Mahalanobis distance Δ2

Maz Jamilah Masnan, Nor Idayu Mahat, Ali Yeon Md Shakaff, and Abu Hassan Abdullah

Citation: AIP Conference Proceedings **1691**, 050017 (2015); doi: 10.1063/1.4937099
View online: http://dx.doi.org/10.1063/1.4937099
View Table of Contents: http://scitation.aip.org/content/aip/proceeding/aipcp/1691?ver=pdfcov
Published by the AIP Publishing

Sensors Closeness Test Based on an Improved $[0,1]$ Bounded Mahalanobis Distance Δ^2

Maz Jamilah Masnan[1, a)], Nor Idayu Mahat[2, b)], Ali Yeon Md Shakaff[3, c)] and Abu Hassan Abdullah[3, d)]

[1]*Institute of Engineering Mathematics, Universiti Malaysia Perlis, Aras Bawah, Pusat Kejuruteraan, Kampus Tetap Pauh Putra, 02600 Arau, Perlis*
[2]*College of Arts and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah*
[3]*Center of Excellence for Advanced Sensor Technology (CEASTech), Lot 16-21, Pusat Pengajian Jejawi 2, Jalan Jejawi Permatang, 02600 Jejawi, Arau, Perlis*

[a)]Corresponding author: mazjamilah@unimap.edu.my
[b)]noridayu@uum.edu.my
[c)]aliyeon@unimap.edu.my
[d)]abu.hassan@unimap.edu.my

Abstract. Mahalanobis distance Δ^2 values are commonly in the range of 0 to $+\infty$ where higher values represent greater distance between class means or points. The increase in Mahalanobis distance is unbounded as the distance multiply. To certain extend, the unbounded distance values pose difficulties in the evaluation and decision for instance in the sensors closeness test. This paper proposes an approach to [0, 1] bounded Mahalanobis distance Δ^2 that enable researcher to easily perform sensors closeness test. The experimental data of four different types of rice based on three different electronic nose sensors namely InSniff, PEN3, and Cyranose320 were analyzed and sensor closeness test seems successfully performed within the [0, 1] bound.

INTRODUCTION

This paper discusses the $[0,1]$ bounded Mahalanobis distance Δ^2 applied in the sensors closeness test. Sensors closeness test examines the closeness or similarity of data from different sensors, which implies whether the sensors are close or similar. Such test is commonly applied in the area of multi sensor data fusion where input data from different sensors are used for further classification analysis. In order to fuse these sensors data, it is good to know whether data to be fused are similar for the correct low data fusion level is applied. Otherwise, intermediate or high level data fusion may be appropriate. Raju and Wang (1994) employed Mahalanobis distance Δ^2 for testing sensors data closeness by comparing two sensors at a time with the aim to identify reliability of sensor data. However, in their analyses the distances obtained were unbounded. Commonly, Mahalanobis distance Δ^2 can takes values from 0 to $+\infty$ which is impractical in decision making especially when the magnitude of Mahalanobis distance Δ^2 is very large, leaving the question of how large is large unanswered and most of the time it is unknown.

In order to overcome these flaws, we adopted a concept of bounded Mahalanobis distance Δ^2 which devoted by Ray and Turner (1992). Their paper mainly discussed the application of Mahalanobis distance-based criteria to evaluate new feature for classification. Due to the failure of using common Mahalanobis distance Δ^2 in representing

Innovation and Analytics Conference and Exhibition (IACE 2015)
AIP Conf. Proc. 1691, 050017-1–050017-6; doi: 10.1063/1.4937099

the average separability of g classes, they transformed the original range of Mahalanobis distance Δ^2 values, $\Delta^2 \in [0,+\infty)$, to lies within a finite range $[0,1]$.

Further discussions of this concept are discussed in the following order; concept of Mahalanobis distance Δ^2 in sensors closeness test, proposed $[0,1]$ bounded Mahalanobis distance Δ^2, application, numerical analysis and results, and finally conclusion.

CONCEPT OF MAHALANOBIS DISTANCE Δ^2 IN SENSORS CLOSENESS TEST

Assuming two sensors A and B with population mean vectors $\boldsymbol{\mu}_{\mathbf{A}}$ and $\boldsymbol{\mu}_{\mathbf{B}}$, and the invertible common covariance matrix $\boldsymbol{\Sigma}$ for sensors A and B, then the Mahalanobis distance Δ^2 between the two sensing devices is given by equation (1)

$$\Delta^2 = \left(\boldsymbol{\mu}_{\mathbf{A}} - \boldsymbol{\mu}_{\mathbf{B}}\right)^{\mathbf{T}} \boldsymbol{\Sigma}^{-1} \left(\boldsymbol{\mu}_{\mathbf{A}} - \boldsymbol{\mu}_{\mathbf{B}}\right). \tag{1}$$

Usually the population parameters are unknown and are estimated using the corresponding sample values. Suppose a sensor device has p array of sensors with n observations, then we have the matrix in (2) with the sample mean vector given by (3), and the mean $\bar{X}_j$ is defined by (4) where m is the sample size

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix} \tag{2}$$

$$\bar{\mathbf{X}} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_p \end{bmatrix} \tag{3}$$

$$\bar{X}_j = \frac{\sum_{i=1}^{m} x_{ji}}{m} \tag{4}$$

Let the sample size n_1 and n_2 of sensors A and B, respectively, such that $n = n_1 + n_2$. Let $\bar{\mathbf{X}}_{\mathbf{A}}$ and $\bar{\mathbf{X}}_{\mathbf{B}}$ be the corresponding sample mean vectors of sensors A and B. Then, the estimated Mahalanobis distance , D^2, is defined by equation (5), where $\mathbf{S}$ in equation (6) is the pooled variance-covariance matrix of sensing devices A and B, respectively.

$$D^2 = \left(\bar{\mathbf{X}}_{\mathbf{A}} - \bar{\mathbf{X}}_{\mathbf{B}}\right)^T \mathbf{S}^{-1} \left(\bar{\mathbf{X}}_{\mathbf{A}} - \bar{\mathbf{X}}_{\mathbf{B}}\right) \tag{5}$$

$$\mathbf{S} = \frac{(n_1 - 1)\mathbf{S}_{\mathbf{A}} + (n_2 - 1)\mathbf{S}_{\mathbf{B}}}{n} \tag{6}$$

Based on (5), the two sensors are considered similar or close if D^2 gives small values. Otherwise, the two sensors are dissimilar.

Proposed [0, 1] Bounded Mahalanobis Distance Δ^2

Raju and Wang (1994), elaborated three types of distance, i.e. straight line distance, statistical distance and Mahalanobis distance, and claimed that Mahalanobis distance Δ^2 is more general and applicable to a wide variety of cases. On this basis, we work further on sensors closeness test and took that as the fundamental concept with the adoption of improved $[0, 1]$ bounded Mahalanobis distance Δ^2. Earlier, Ray and Turner (1992) forwarded in detail the Mahalanobis distance Δ^2 and its relationships with the Bayesian probability of error, both within the distribution free case and in the case when feature vector follows a Gaussian distribution. They proposed Mahalanobis distance $\Delta^2 \in [0, 1]$ to reduce the drawback of getting high value of average distance which fails to represent the average separability of *g* classes. In order to transform Mahalanobis distance $\Delta^2 \in [0, +\infty)$ to bounded Mahalanobis distance $\Delta^2 \in [0, 1]$, Ray and Turner (1992) derived the transformed Δ^2 (denoted by Δ_1^2) given by equation (7), where π_1 and π_2 are the *a priori* probabilities of the respective classes.

$$\Delta_1^2 = \frac{\pi_1 \pi_2 \Delta^2}{1 + \pi_1 \pi_2 \Delta^2} . \tag{7}$$

Further details of the transformations, properties as well as some related proofs are referred to Ray and Turner (1992). In sensors closeness test, the transformation $\Delta_1^2 \in [0, 1]$ is achieved by applying the concept in equation (7). The estimated transformed Δ_1^2 is defined by equation (8)

$$D_1^2 = \frac{\pi_1 \pi_2 D^2}{1 + \pi_1 \pi_2 D^2} . \tag{8}$$

For the sensor closeness test, the *a priori* probability of π_1 and π_2 are considered equal where $\pi_1 = \pi_2 = 0.5$, due to the assumption of equally importance of both sensors to the analysis. Thus, equation (8) can be further simplified to equation (9) after some algebraic processes. Hence, the estimated transformed Mahalanobis distance $D^2{}_1$ of sensor devices A and B is obtained using equation (9)

$$D_1^2 = \frac{D^2{}_{AB}}{4 + D^2{}_{AB}} . \tag{9}$$

Interestingly, Ray and Turner (1992) have also illustrated some properties of the Mahalanobis distance $\Delta_1^2 \in [0, 1]$ in terms of the boundedness, monotonicity, symmetry with respect to π_1 and π_2, property of $\Delta_1^2(\pi_1 = p) < \Delta_1^2(\pi_1 = q)$ iff $\min(p, 1-p) < \min(q, 1-q)$, and finally the relationships of the transformed Mahalanobis distance $\Delta_1^2 \in [0, 1]$ with the Bayesian error probability.

APPLICATIONS

We tested the transformed Mahalanobis distance $\Delta_1^2 \in [0, 1]$ in the sensors closeness test that involved three different electronic nose sensors; InSniff, PEN3, and Cyranose320. Basically these sensors detect and classify substance based on its gases or volatile compounds. InSniff is a portable electronic nose sensor developed by UniMAP (Universiti Malaysia Perlis), and it consists of twelve different metal oxide sensors (MOS) used to detect specific gases or volatile compounds. PEN3 is a portable electronic nose from Win Muster Airsense (WMA) Analytics Inc., Germany, comprises of ten different MOS sensor arrays. While Cyranose320 is a portable system from Smith Detection™ (Pasadena, CA, USA) consisting of 32 individual polymer sensors made up of various conducting polymers to sense variety of volatile compounds.

Four different samples of rice such as basmathi (Moghul), local glutinous rice, ordinary rice (Super Import) and aromatic rice (Jasmine) were selected for the experiment. All rice was stored in air-tight stainless steel container to ensure no bias towards the storage effect. The experiments were implemented in a closed laboratory. The laboratory temperature (22^{o}C) and humidity (70%) were measured using hydrometer 506-HI from Testo UK. The controlled environment is to minimize humidity and temperature variation during the experiment to ensure data repeatability.

About 50 grams of each rice sample was weighed using an electronic balance and poured into a tin canister which was wrapped tightly with a paraffine film wrapper. The sample was left idle for ten minutes at room temperature for the sample to reach the equilibrium state before the experiment was performed. The instruments applied static headspace sampling technique. The headspace gas was pumped into the sensor chamber at a constant rate via a Teflon-tubing connected to the instrument front-end during data acquisition process. Each sample measurement was conducted in two different batches of samples and data acquisition measurements were replicated for at least 10 times.

Each sensor produced different data dimensions mainly due to the number of array sensors attached in every sensor (or known as the number of features referred as p). For example, the number of features for all the sensors are $p(Insniff) = 12$, $p(PEN3) = 10$, and $p(Cyranose320) = 32$. We presumed that all the sensors are similar due to its common function of detecting different rice types based on its volatile compounds.

NUMERICAL ANALYSES AND RESULTS

The aim of this section is to evaluate the transformed Mahalanobis distance $\Delta_1^2 \in [0,1]$ in the sensors closeness test. Recorded data from InSniff, PEN3, and Cyranose320 for four different samples of rice (basmathi, local glutinous rice, ordinary rice and aromatic rice) were applied. Due to unequal number of features from the three sensors, we decided to include only 10 features from each sensor to be included for further analysis, as well as to ensure the variance-covariance matrix and it inverse is obtainable. Thus, only 10 features with the highest means were included for evaluation. Once data were ready, we employed equations (5) and (6) to find the estimated unbounded Mahalanobis distance D^2 for every pair of sensors applied.

TABLE 1. The unbounded Mahalanobis distance for sensors closeness test based on basmathi (in the upper diagonal), and glutinous (in the lower diagonal).

	InSniff	PEN3	Cyranose320
InSniff	0	6,830,208.51 [3]	8,982,099.22 [2]
PEN3	2,603,550.08 [3]	0	52,240,106.41 [1]
Cyranose320	16,756,572.78 [2]	112,434,005.47 [1]	0

TABLE 2. The unbounded Mahalanobis distance for sensors closeness test based on ordinary rice (in the upper diagonal), and aromatic rice (in the lower diagonal).

	InSniff	PEN3	Cyranose320
InSniff	0	4,230,141.03 [3]	5,445,185.70 [2]
PEN3	9,944,068.92 [3]	0	50,429,303.18 [1]
Cyranose320	15,070,734.49 [2]	39,687,137.09 [1]	0

Tables 1 and 2 show the estimated Mahalanobis distance D^2 for each pair of InSniff-PEN3, InSniff-Cyranose320, and PEN3-Cyranose320 for four different types of rice. Obviously, the magnitudes of the estimated Mahalanobis distance D^2 for each pair of sensor are huge even though these sensors detected volatile compounds emitted by the common rice samples. For instance, sensor closeness test for basmathi (upper diagonal) in Table 1 shows that PEN3 and Cyranose320 is the most dissimilar sensors with the highest estimated Mahalanobis distance D^2 value (52,240,106.41), while InSniff and PEN3 is still considered dissimilar but with the lowest estimated Mahalanobis distance D^2 value (6,830,208.51). For the purpose of confirmation, sensor closeness test towards the same sensor were also carried out. The results were clearly proven along the diagonal of Tables 1 and 2, where the estimated Mahalanobis distance D^2 values for InSniff-InSniff, PEN3-PEN3 and Cyranose320- Cyranose320 are all equal to zero.

TABLE 3. The bounded Mahalanobis distance for sensors closeness test based on basmathi (in the upper diagonal), and glutinous (in the lower diagonal).

	InSniff	**PEN3**	**Cyranose320**
InSniff	0	0.99999766 [3]	0.99999822 [2]
PEN3	0.99999385 [3]	0	0.99999969 [1]
Cyranose320	0.99999905 [2]	0.99999986 [1]	0

TABLE 4. The bounded Mahalanobis distance for sensors closeness test based on ordinary rice (in the upper diagonal), and aromatic rice (in the lower diagonal).

	InSniff	**PEN3**	**Cyranose320**
InSniff	0	0.99999622 [3]	0.99999706 [2]
PEN3	0.99999839 [3]	0	0.99999968 [1]
Cyranose320	0.99999894 [2]	0.99999960 [1]	0

On the contrary, when equation (9) was employed with equal *a priori* probability, all the estimated Mahalanobis distance D^2 in Table 1 and Table 2 were improved to be within a $[0,1]$ bounded value. These findings confirmed that the unbounded $[0,+\infty)$ Mahalanobis distance D^2 can be improved to a $[0,1]$ bounded Mahalanobis distance $D^2{}_1$ using equation (9). The findings in Tables 3 and 4 were also found to be consistence with the findings in Table 1 and 2 in terms of ranking. Please refer to the ranking based on the number in the square bracket. For example, sensors closeness test for InSniff, PEN3 and Cyranose320 based on aromatic rice in the lower diagonal of Table 4 illustrates that all the three pairs are dissimilar to each other where the most disparate sensors were PEN3-Cyranose320, followed by InSniff-Cyranose320 and finally InSniff-PEN3. This finding was also congruent with the finding based on the estimated Mahalanobis distance D^2 for sensors closeness test based on aromatic rice in Table 2. Further comparison for sensors closeness test of all the sensors based on ordinary rice, glutinous rice and basmathi can be evaluated based on Table 3 and 4.

CONCLUSION

Testing closeness of sensors using the unbounded $[0,+\infty)$ Mahalanobis distance Δ^2 can be improved using $[0,1]$ bounded Mahalanobis distance Δ_1^2. Difficulties in making decision based on boundless magnitude of Δ^2 can be resolved using bounded Mahalanobis distance Δ_1^2. Besides, this concept can also be applied in sensors screening before adapting multi sensor data fusion model. Fundamental knowledge for the kind of sensors involve for fusion is very important for which level of data fusion model is relevant to be employed. Beneficially this paper contributes towards the aim.

ACKNOWLEDGMENTS

This research is part of the postgraduate research work. Special thanks to the Ministry of Higher Education Malaysia (MOHE) for providing the financial assistance, Center of Excellence for Advanced Sensor Technology (CEASTech),Universiti Malaysia Perlis (UniMAP) for granting good cooperation and data exchange, as well as College of Arts and Sciences, Universiti Utara Malaysia for such encouraging support. The involvement and cooperation among all the authors of this article is very much appreciated.

REFERENCES

1. G. V. S. Raju and H. Wang, *IEEE International Conference on Systems, Human, Information and Technology* **3**, 2605–2610 (1994).
2. S. Ray and L. F. Turner, *Information Sciences* **60**, 217-245 (1992).

3. P. C. Mahalanobis, *Proc. Natl. Inst. Sci. (India)* **12**, 49-55. (1936).
4. P. C. Mahalanobis, D. N. Majumdar, M. W. M. Yeatts and C. R. Rao, *Sankhyā: The Indian Journal of Statistics (1933-1960)* **9(2/3)**, 89-324 (1941).
5. M. J. Masnan, N. I. Mahat, A. Y. M. Shakaff, A. H. Abdullah, N. Z. I. Zakaria, N. Yusuf, N. Subari, A. Zakaria and A. H. Aziz, "Understanding Mahalanobis Distance Criterion for Feature Selection," in *International Conference on Mathematics, Engineering and Industrial Applications 2014 (ICoMIA 2014),* AIP Conference Proceedings 1660(2), edited by M. F. Ramli et al. (American Institute of Physics, Melville, NY, 2015), 050075.
6. R. C. Gonzalez and C. G. Wagner, *IEEE Transactions on Systems, Man, and Cybernatics* **13(6)**, 1135-1139 (1983).
7. C. R. O. Lawoko and C. Y. J. Fu, *Pattern Recognition* **26(6)**, 989-990 (1993).