Classification of HCI Datasets Using Information Fusion and Weighted Frechet Distance

Aleksandar Jeremic and Huaying Li

Department of Electrical and Computer Engineering McMaster University, Hamilton, ON, Canada

Keywords: Classification, Information Fusion, Bioinformatics, Biomedical Signal Processing.

Abstract: The identification and classification of human breast cancer cells (MCF7) undergoing various treatments are widely used for studies of tumour biology and drug mechanism action. The development of adequate detection/classification strategies that would meet clinical needs is currently a subject of significant research interest as the optimal techniques are application/cell/treatment dependent. In addition to commonly used machine learning techniques for classification/clustering there has been an effort to utilize deep learning techniques as well. However, due to the fact that different cancer cells and different treatments require different data sets these techniques had rather limited success. In this paper we propose an information fusion technique that utilizes Frechet distance measures by combining their decisions in an optimal way by minimizing the overall classification error. The applicability of our results is demonstrated using real data sets with ten different treatments.

1 INTRODUCTION

Information fusion techniques have been widely applied in many applications including clustering, classification, detection and etc. One of the major objectives is to improve the classification performance (i.e. minimize overall probability of error) by incorporating decisions from various sources individually and combining them into a global decision that is potentially more accurate. Classification techniques are commonly used tools in analytical chemistry due to high complexity and dimensionality of the chemical measurements and in recent years a large number of deep learning (DL) techniques have been successfully utilized in this field. (Debus et al., 2021). However in certain application the amount of data available may not be sufficient for DL techniques and hence certain feature reduction techniques may be required in order to achieve desirable performance.

In particular in bio-image analysis the increase in imaging throughput, new analytical frameworks and large computational resources created new research opportunities in drug discovery by enabling effect analysis of various treatments on similar cell types. Most of the current solutions still utilize machine learning techniques requiring feature extraction/reduction preprocessing due to the fact that the number of images generated for particular treatments/starvations may not be sufficient for deep learning (DL) methods. In our previous work we developed mathematical methods for calculating Frechet mean with respect to Riemannian distances and demonstrated their applicability to estimating sample mean of matrix ensemble (Jahromi et al., 2015). Furthermore, we demonstrated that Frechet mean can be used to classify HCI using covariance structure of random variations between various classes by focusing not only on the center of the cluster distance but by accounting for covariance structure of these classes as well.

In this paper, we extend our previous work by proposing weighted distance measures and information fusion algorithms to combine classification decisions of different classifiers. In Section 2 we introduce the Fréchet mean based on several Riemannian distances and present information fusion algorithm for making global classification decision. In Section 3 we illustrate the applicability of our techniques using a real data set. In Section 4 we discuss conclusions and directions for future research.

2 FRECHET MEAN

We use the notion of Fréchet mean to unify the method of finding the mean of positive definite ma-

Jeremic, A. and Li, H.
Classification of HCI Datasets Using Information Fusion and Weighted Frechet Distance.
DOI: 10.5220/0011798100003414
In *Proceedings of the 16th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2023) - Volume 4: BIOSIGNALS*, pages 340-343
ISBN: 978-989-758-631-6; ISSN: 2184-4305

trices. The Fréchet mean is given as the point which minimizes the sum of the squared distances (Barbaresco, 2008):

$$\hat{\mathcal{S}} = \text{argmin}_{\mathcal{S} \in \mathcal{M}} \sum_{i=1}^{n} d^2(\mathbf{S}_i, \mathcal{S}) \tag{1}$$

where $\{\mathbf{S}_i\}_{i=1}^{n}$ represents the symmetric positive definite matrices and $d(.,.)$ denotes the metric being used respectively.

To measure the distance between two $M \times M$ covariance matrices $\mathbf{A}$ and $\mathbf{B}$ on manifold of positive definite matrices $\mathcal{M}$, we consider the metrics which have been developed to measure distance between two points on the manifold itself. The following metrics will be considered throughout the remaining chapters.

The first metric is obtained when we lift the points $\mathbf{A}, \mathbf{B}$ to the horizontal subspace $\mathcal{U} \subset \mathcal{H}$ using the fibre and measure the distance between them(Li and Wong, 2013):

$$d_{R_1}(\mathbf{A}, \mathbf{B}) = \text{argmin}_{\tilde{\mathbf{U}}_1, \tilde{\mathbf{U}}_2 \in U(M)} \left\| \mathbf{A}^{\frac{1}{2}} \tilde{\mathbf{U}}_1 - \mathbf{B}^{\frac{1}{2}} \tilde{\mathbf{U}}_2 \right\|_2 \tag{2}$$

where $U(M)$ denotes the space of unitary matrices of size $M \times M$. Alternatively Eq.(2) can be rewritten as:

$$\sqrt{\text{Tr}(\mathbf{A}) + \text{Tr}(\mathbf{B}) - 2\text{Tr}(\mathbf{A}^{\frac{1}{2}} \mathbf{B} \mathbf{A}^{\frac{1}{2}})^{\frac{1}{2}}} \tag{3}$$

In general for any positive definite matrix $\mathbf{A}$ its square root is defined as $\mathbf{A}^{\frac{1}{2}} = \mathbf{S}\sqrt{\mathcal{L}}\mathbf{D}^H$; where $\mathbf{A} = \mathbf{S}\mathcal{L}\mathbf{D}^H$ is the eigenvalue value decomposition of matrix $\mathbf{A}$ with diagonal matrix $\mathcal{L}$ consisting of eigenvalues of $\mathbf{A}$.

The second distance measure we will use is given by

$$\begin{aligned} d_{R_2}(A,B) &= \|A^{\frac{1}{2}} - B^{\frac{1}{2}}\|_2 \\ &= \sqrt{\text{Tr}(A) + \text{Tr}(B) - 2\text{Tr}(A^{\frac{1}{2}} B^{\frac{1}{2}})} \end{aligned} \tag{4}$$

To define the last distance we will use, let the points $\mathbf{A}, \mathbf{B} \in \mathcal{M}$ and let $\mathbf{X}$ be a the point on the manifold at which we construct a tangent plane (it is usually denoted as $T_{\mathcal{M}}\mathbf{X}$). According to the inner-product $\langle \mathbf{A}, \mathbf{B} \rangle_{\mathbf{X}} = \text{Tr}(\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1}\mathbf{B})$ the log- Riemannian metric is given as (Moakher, 2005):

$$d_{R_3}(\mathbf{A}, \mathbf{B}) = \left\| \log(\mathbf{A}^{-\frac{1}{2}} \mathbf{B} \mathbf{A}^{-\frac{1}{2}}) \right\|_2 = \sqrt{\sum_{i=1}^{M} \log^2(\mathcal{L}_i)} \tag{5}$$

where the $\mathcal{L}_i$'s are the eigenvalues of the matrix $\mathbf{A}^{-1}\mathbf{B}$ (Absil et al., 2009). (Metric d_{R3} has been developed in various ways and has, for a long time, been used in theoretical physics).

In detection and classification process we can improve the performance of a classifier in discriminating between the features with similar properties resulting from same class by keeping them as close as possible and similarly keep different features as far as possible by utilizing images previously labelled by experts. This process can be performed using the concept of weighted distances presented in (Li et al., 2009) where the weighting matrix is calculated using

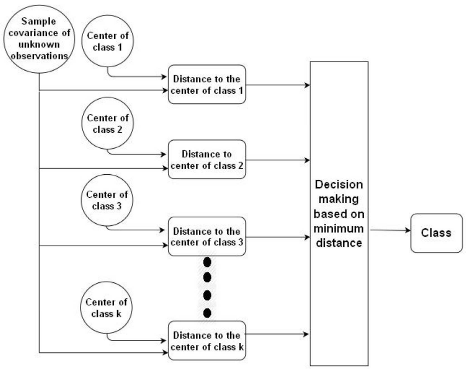
Sample covariance of unknown observations
Center of class 1
Distance to the center of class 1
Center of class 2
Distance to center of class 2
Center of class 3
Distance to the center of class 3
Center of class k
Distance to the center of class k
Decision making based on minimum distance
Class

Figure 1: Single Distance Local Classifier.

$$\text{argmax}_w \frac{d_w^2(S_{ik}, S_{jk})}{d_w^2(S_{ik}, S_{jk'})} \tag{6}$$

where W is positive definite Hermitian matrix. The summation in nominator of is performed over all covariance matrices in similar classes. On the other hand, the denominator in same equation is summation over the all possibilities of covariance matrices in the dissimilar classes.

2.1 Distributed Classification System

Consider a scenario in which each of k classes has a corresponding covariance matrix describing its randomness. Due to the fact that number of images is rather large, only small fraction of these cell images can be labelled by experts. These images are then used to create the corresponding covariance matrices that define particular clusters. The new, unlabelled data, i.e. cell images undergoing particular treatment can then be classified automatically without expert involvement. The graphic illustration of the system is presented in Figure 1. In Figure 2 we illustrate the overall schematic. Data set consisting of sample covariance matrices is classified using 3 local classifiers (based on three distances) and the results of these classifications are then transmitted to fusion center. (Liu et al., 2007).

The role of the local classifiers LC_n is to make local decision u_n based on their own distance measure. All the local decisions are then sent to the fusion center, where the global decision u_0 is made based on a fusion rule in order to minimize the overall probability of error. In this work, we only focus on the case of three local classifiers using three aforementioned

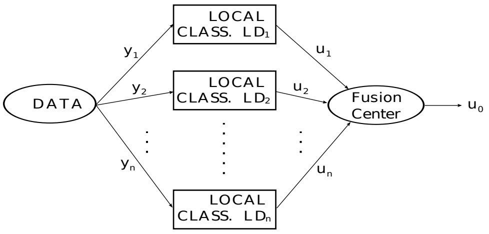

Figure 2: Fusion Classification.

distances.

Following the approach of (Liu et al., 2014) we formulate the above problem as M-ary classification problem with corresponding unknown anomalies ε_{ij}. Since the overall performance of each classifier is data dependent all of the anomalies (probabilities of incorrect classification i.e. picking class Ci when class Cj is a correct choice) are unknown. In (Liu et al., 2014) we derived maximum likelihood estimator of the unknown anomalies using multiple decision vectors and demonstrated that the algorithm converges to true values after certain number of global decisions assuming that the statistical model governing the phenomenon of interest is not changing.

In this application we assume that the statistical distribution modelling the covariance matrices resulting from cell images undergoing particular treatment is not changing. Once the anomalies are estimated the global decision can be made following the approach of (Varshney, 1986) by minimizing the overall probability of error. The global decision is given by

$$\begin{aligned} u_0 &= \arg\max_i P(C_i|u_0) \\ &= \arg\max_i P(C_i) \prod_{j\in S_0} \varepsilon^j_{i0} \cdot \prod_{jinS_{M-1}} \varepsilon^j_{iM-1} \end{aligned} \quad (7)$$

where

$$\begin{aligned} S_0 &= \{j|C_j = 0, \forall j = 1,2,3\} \\ &\vdots \\ S_{M-1} &= \{j|C_j = M-1, \forall j = 1,2,3\} \end{aligned}$$

correspond to partitions of local classifiers indices (class decisions).

3 EXPERIMENTAL RESULTS

The data set that we have for classification consists of 11 labels corresponding to 11 types of treatment as illustrated in Table 1 and was provided by Dr. David Andrews lab at Sunnybrook Hospital, Toronto, Ontario, Canada. The input data set consists of multiple images of breast cancer cells (MCF7) obtained using Opera High Content Screening System producing multichannel images. These images are then automatically segmented in order to obtain segments corresponding to a single cell. Then, feature extraction is then performed extracting 705 predefined features commonly used in analytical chemistry. Due to the nature of the data certain features can be correlated. In that case these could be removed following the approach of (Shawe-Taylor and Cristianini, 2004). In this paper we remove the correlated features using mutual correlation approach presented in (Shawe-Taylor and Cristianini, 2004). Then we construct sample covariances by calculating sample covariance of feature vectors corresponding to cells undergoing the same treatment. In order to generate multiple covariance matrices we divide the labelled images (training set) into smaller groups of vectors resulting in multiple covariance matrices per class.

Table 1: Medications and doses.

Treatment	Dose
DMSO	2.5%
Ethanol	6
BFA	10 g/ml
Rapamycin	25 M
Tamoxifen	30 m
Thapsigargin	40 nM
Tunicamycin	25M
TNFalpha	10 ng/m
Starvation24	24hours
Starvation72	72 hours

The most important effect of weighting algorithm is to keep the covariance matrices within the same classes sufficiently close to each other while separating covariance matrices belonging to the different classes as far as possible. In Figure 2 we illustrate the results of the weighted distance classification and for comparison purposed in Figure 3 we illustrate the same results without distance weights. These results indicate that for different cell treatments different distances have superior performance. To this purpose we examine applicability of the fusion techniques as mentioned before. In Table 3 we illustrate the classification performance. We can see that the performance improvement for all the treatment types increases between 1-5%.

4 CONCLUSIONS

In this paper we demonstrated ability to classify high content cell images that are commonly used in drug development using classifier based on the Frechet dis-

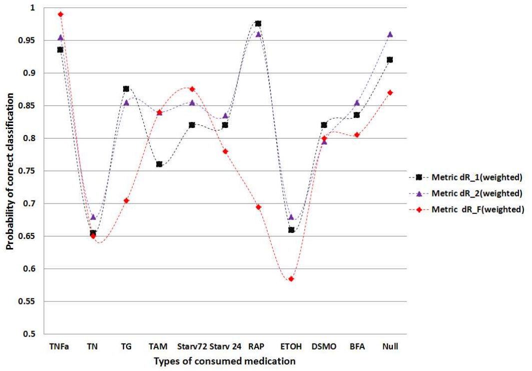

Figure 3: Probability of correct classification - weighted.

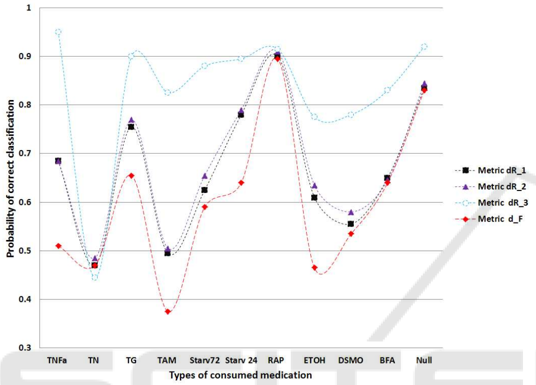

Figure 4: Probability of correct classification - non-weighted.

tance measure between sample covariance matrices calculated from the HCI real data set. In order to improve the performance of previously used distance measures in this paper we use weighted distance measure as it allows to reduce the probability of classification errors by increasing the cluster center distances. We evaluate the performance using dataset consisting of 11 different classes corresponding to different treatments. To further improve the classification results we perform maximum likelihood based classification fusion. Our results indicate that the results obtained perform similarly to classically used algorithms based on average based classification algorithms. In future work we plan to develop more efficient computational algorithms and evaluate performance as a function of training set size. In addition, the performance of the proposed algorithm may depend significantly on the algorithm used to construct a sample covariance set and therefore an effort should be made to investigate robustness/dependency of the proposed algorithm on the sampling process.

Table 2: Medications and doses.

Treatment	ML-fused improvement
DMSO	1%
Ethanol	3%
BFA	1%
Rapamycin	2%
Tamoxifen	2%
Thapsigargin	1%
Tunicamycin	2%
TNFalpha	2%
Starvation24	3%
Starvation72	5%

REFERENCES

Absil, P.-A., Mahony, R., and Sepulchre, R. (2009). *Optimization algorithms on matrix manifolds*. Princeton University Press.

Barbaresco, F. (2008). Innovative tools for radar signal processing based on Cartans geometry of SPD matrices & information geometry. *Radar Conference, 2008. RADAR'08. IEEE*, pages 1–6.

Debus, B., Parastar, H., Harrington, A., and Kirsanov, D. (2021). Deep learning in analytical chemistry. *Trends in Analytical Chemistry*, 145:116459.

Jahromi, M., Wong, K., and Jeremic, A. (2015). Estimating Positive Definite Matrices using Frechet Mean. In *Biosignal 2015*, pages 2021–2026. INSTIC.

Li, Y., Wong, K., and deBruin, H. (2009). Eeg signal classification based on a riemannian distance measure. *IEEE TIC-STH*, pages 225–230.

Li, Y. and Wong, K. M. (2013). Riemannian distances for EEG signal classification by power spectral density. *IEEE journal of selected selected topics in signal processing*.

Liu, B., Jeremic, A., and Wong, K. (2007). Blind adaptive algorithm for M-ary distributed detection. In *IEEE International Conference on Acoustics, Speech and Signal Processing, 2007. ICASSP 2007*, volume 2.

Liu, B., Jeremic, A., and Wong, K. (2014). Optimal distributed detection of multiple hypotheses using blind algorithm. *IEEE Trand. on Aerospace and Electronic Systems*, 50:1190–1203.

Moakher, M. (2005). A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747.

Shawe-Taylor, J. and Cristianini, N. (2004). *Kernel methods for pattern analysis*. Cambridge university press.

Varshney, P. (1986). Optimal data fusion in multiple sensor detection systems. *IEEE Trans. on Aerospace and Electronic Systems*, pages 98–101.