15th IAPR International Conference on Machine Vision Applications (MVA)
Nagoya University, Nagoya, Japan, May 8-12, 2017.

Asymmetric Locality Preserving Projection and Its Application to k-Nearest Neighbor Method

Yoshio Iwai Masashi Nishiyama Hiroki Yoshimura
Graduate School of Engineering, Tottori University
iwai@ike.tottori-u.ac.jp

Abstract

In recent years, many methods for data compression and structure extraction from various types of massive data using multivariate analysis have been proposed. The locality preserving projection, which uses a symmetric similarity matrix, is one of these data compression methods. However, the similarity matrix expressing the characteristic of data may often not be symmetric in real. In this study, we propose an asymmetric locality preserving projection that expands the locality preserving projection from a symmetric similarity matrix method to one that uses an asymmetric similarity matrix. We also show the experimental results of its application to the k-nearest neighbor method as an example.

1 Introduction

In recent years, a large amount of data has become available online, collected various sensors. The extraction of data structure from large-scale databases for mining various information and dimensional compression for reducing the data processing costs are tasks that often must be performed. Multivariate analysis is often used for dimensional reduction and structure extraction. Many of these multivariate analysis methods can be expressed using an extended pairwise representation [1, 2]. Similarly, as is well-known, the locality preserving projection (LPP) is also expressed using an extended pairwise representation. Because a covariance matrix or a Gramian matrix is used as a statistical index for linking multivariates, it is assumed that these matrices are symmetric in the extended pairwise representation.

An LPP using a symmetric similarity matrix is often applied to data compression. However, the similarity matrix expressing the characteristic of data may often not to be symmetric in real. For example, it is known that the Smith-Waterman and BLAST scores [3] for measuring the similarity between DNA and protein sequences in bioinformatics, and the tangent distant kernel [4] and Simpson score [5] used in image classification may not be symmetric.

In this research, we propose an extension of the LPP from a method that uses a symmetric similarity matrix to one that uses an asymmetric similarity matrix. We can apply the proposed asymmetric LPP (ALPP) to classification problems that have both label similarity and data proximity. We also present some experimental results to confirm the effectiveness of our method.

2 LPP [6]

In this section, we briefly explain the LPP [6]. Let $\{x_i\} \in R^m$ be the training data and let $X = (\ x_1 \quad x_2 \quad \cdots \quad x_n\ )$ be a data matrix. Further, let S^{LPP} be a symmetric similarity matrix. Then, the LPP is expressed as follows:

$$W_{opt} = \arg\min_W \left[\frac{1}{2} \sum_{i,j=1}^{n} S_{ij}^{\mathrm{LPP}} \|Wx_i - Wx_j\|^2 \right]$$
$$= \arg\min_W \mathrm{tr}\left(WXLX^TW^T\right) \ \text{ s.t. } WXDX^TW^T = I_m,$$

where $L = D - S^{\mathrm{LPP}}$ is called a graph Laplacian matrix and I_m is a m-dimensional identity matrix. In addition, W expresses a linear mapping that embeds training data x_i into the subspace.

Various similarity matrices can be considered; however, typical similarity matrices are as follows:

$$S_{ij}^{\mathrm{LPP}} = \begin{cases} \exp\left(-\|x_i - x_j\|^2 \big/ \gamma\right) & \|x_i - x_j\|^2 < \varepsilon \\ 0 & \text{otherwise} \end{cases},$$

or

$$S_{ij}^{\mathrm{LPP}} = \begin{cases} 1 & x_i \in N_\varepsilon(x_j) \cup x_j \in N_\varepsilon(x_i) \\ 0 & \text{otherwise} \end{cases},$$

where ϵ is a constant that determines the neighborhood and N_ϵ represents the neighborhood. The above optimization problem results in the following generalized eigenvalue problem:

$$XLX^TW^T = \lambda XDX^TW^T.$$

The optimized solution W is obtained by arranging the eigenvectors in ascending order of the magnitude of the eigenvalues of the generalized eigenvalue problem.

3 Representation of Asymmetric Similarity

As mentioned above, the LPP assumes that a similarity matrix is a symmetric matrix, but this is not always the case in the real world. As shown in Fig. 1, this often occurs when a data point x_j is not in its own neighborhood $N_k(x_i)$ in the k-nearest neighbor method. In order to apply an asymmetric matrix to a symmetric matrix method, a symmetric matrix series expansion of the asymmetric matrix can be used to approximate an asymmetric matrix. In this paper, we utilize the method using the Hermitian matrix proposed by Chino et al. [7].

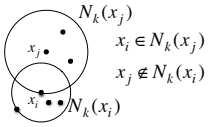


Figure 1. Asymmetric proximity of a three-nearest neighbor method

3.1 Hermitian Representation of Asymmetry [7]

Let S be a real asymmetric (so-called general) similarity matrix. Matrix S can be separated into real symmetric matrix $S^{\text{symm}}$and real alternative matrix S^{skew}as follows:

$$S = S^{\text{symm}} + S^{\text{skew}},$$
$$S^{\text{symm}} = \left(\frac{s_{ij} + s_{ji}}{2}\right), \quad S^{\text{skew}} = \left(\frac{s_{ij} - s_{ji}}{2}\right).$$

Here, we define the Hermitian matrix H as follows:

$$H = S^{\text{symm}} + iS^{\text{skew}} \quad (i^2 = -1).$$

Real symmetric matrix S^{symm} is used for the real part of H, and real alternative matrix S^{skew} is used for the imaginary part of H. Because H is an Hermitian matrix, it can be decomposed using a unitary matrix U and a real diagonal matrix Λ as follows:

$$H = U\Lambda U^*,$$

where U^* is the Hermitian conjugate matrix. Matrix H is complex, but the eigenvalue matrix is a real matrix. Letting $U = U_r + iU_c$, we obtain

$$H = (U_r\Lambda U_r^T + U_c\Lambda U_c^T) + i(U_c\Lambda U_r^T - U_r\Lambda U_c^T).$$

Because the imaginary part U_c is derived from the asymmetric component, the metric obtained from the symmetric component is corrected by the asymmetric component U_c.

If H is a positive (negative) definite matrix, concretely, if all the eigenvalues of the eigenvalue matrix Λ are the same sign, the norm and inner product can be defined by the following equation:

$$\|x\|_H = \sqrt{(x, Hx)} = \sqrt{x^T U\Lambda U^* x} = \sqrt{(\Lambda^{\frac{1}{2}}U^*x, \Lambda^{\frac{1}{2}}U^*x)}.$$

This equation is no different from the metric calculation used in machine learning, i.e., it is possible to construct a Hilbert space with H as the measure.

3.2 Extension of the Hermitian Representation of Asymmetry

As mentioned above, the Chino's method [7] equally adds symmetric components and asymmetric components to the Hermitian matrix. In order to control the effect of asymmetric components on its norm more freely, we propose the following representation to express asymmetric similarity:

$$H = (1-\alpha)S^{symm} + \alpha iS^{skew} \quad (i^2 = -1, 0 \leqq \alpha \leqq 1).$$

In this formulation, it is possible to adjust the effect of asymmetric components, such that only asymmetric components are considered when $\alpha = 1$, and only symmetric components are considered when $\alpha = 0$.

4 ALPP

Using H defined in the previous section, we can simply rewrite the objective function of the LPP as follows:

$$W_{\text{opt}} = \arg\min_W \left[\frac{1}{2}\sum_{i,j=1}^{n} H_{ij}\|Wx_i - Wx_j\|^2\right].$$

Here, as H is a complex matrix, this optimization is a minimization with complex numbers. At a glance, this objective function itself has no meaning because the complex numbers cannot be totally ordered[1]. Note that when this objective function is concretely calculated, we obtain

$$\begin{aligned}
&\frac{1}{2}\sum_{i,j=1}^{n} H_{ij}\|Wx_i - Wx_j\|^2 \\
&= \frac{1}{2}\sum_{i,j=1}^{n} H_{ij}\{(Wx_i,Wx_i) - 2(Wx_i,Wx_j) + (Wx_j,Wx_j)\} \\
&= \frac{1}{2}\sum_{i=1}^{n}(Wx_i, Wx_i)\sum_{j=1}^{n} H_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{n} H_{ij}\,(Wx_i, Wx_j) \\
&\quad + \frac{1}{2}\sum_{j=1}^{n}(Wx_j, Wx_j)\sum_{i=1}^{n} H_{ij} \\
&= \frac{1}{2}\sum_{i=1}^{n}(Wx_i, Wx_i)\sum_{j=1}^{n}(H_{ij} + H_{ji}) \\
&\quad - \sum_{i=1}^{n}\sum_{j=1}^{n} H_{ij}\,(Wx_i, Wx_j).
\end{aligned}$$

Because $S_{ij}^{symm} = S_{ji}^{symm}$, $S_{ij}^{skew} = -S_{ji}^{skew}$, if we set D as follows:

$$\begin{aligned}
D &= \text{diag}\left(\sum_{j=1}^{n}(H_{ij} + H_{ji})\right), \\
D_{ii} &= \sum_{j=1}^{n}(H_{ij} + H_{ji}) \\
&= \sum_{j=1}^{n} S_{ij}^{Symm} + iS_{ij}^{Skew} + S_{ji}^{Symm} + iS_{ji}^{skew} \\
&= 2\sum_{j=1}^{n} S_{ij}^{Symm} = 2D'_{ii}, \quad \left(D'_{ii} = \sum_{j=1}^{n} S_{ij}^{Symm}\right),
\end{aligned}$$

we can rewrite the objective function as follows:

$$\begin{aligned}
&\frac{1}{2}\sum_{i,j=1}^{n} H_{ij}\|Wx_i - Wx_j\|^2 \\
&= \frac{1}{2}\sum_{i=1}^{n} D_{ii}\,(Wx_i, Wx_i) - \sum_{i=1}^{n}\sum_{j=1}^{n} H_{ij}\,(Wx_i, Wx_j) \\
&= \text{tr}\left(WXD'X^TW^T\right) - \text{tr}\left(WXHX^TW^T\right) \\
&= \text{tr}\left(WXLX^TW^T\right) \quad (L = D' - H),
\end{aligned}$$

[1] This objective function has meaning if we take the absolute value of the term on the right side, but discussion will be necessary about the validity of the function used in applications.

where the graph Laplacian matrix L is a complex matrix with imaginary parts S^{skew}. In order to avoid a trivial solution $W = 0$, we introduce the same constraint as used in the LPP as follows:

$$WXD'X^TW^T = I_m.$$

In short, the optimization problem of the LPP with an asymmetric similarity matrix is formulated as follows:

$$W_{\text{opt}} = \arg\min_W \text{tr}\left(WXLX^TW^T\right) \text{ s.t. } WXD'X^TW^T = I_m.$$

Because L is a Hermitian matrix and its eigenvalues are real numbers, the trace of the objective function also becomes a real number. This optimization problem, therefore, becomes meaningful. Using the method of Lagrangian multipliers, we define the objective function $J(W)$ as follows:

$$J(W) = WXLX^TW^T - \lambda\left(WXD'X^TW^T - I_m\right).$$

Then, partial differentiation is performed with W^T, we obtain the following equation:

$$\frac{\partial}{\partial W^T}J(W) = XLX^TW^T - \lambda XD'X^TW^T = 0.$$
$$\therefore XLX^TW^T = \lambda XD'X^TW^T.$$

Solution W can be obtained by solving the generalized eigenvalue problem for complex matrices and arranging the eigenvectors in ascending order of the magnitude of the eigenvalues. Although D' is a real matrix, because L is a Hermitian matrix, the eigenvalues are real numbers, but the eigenvectors are complex vectors, so W is the projection of the complex subspace.

5 Experimental Results

As an application example of the ALPP, we implemented the k-nearest neighbor method and compared it with the other dimensional reduction methods. As comparative methods, we implemented principle component analysis (PCA) and the LPP.

5.1 Data Set

In this experiment, we used the USPS handwritten digit database[2] [8]. We used 7,291 images as the training data of a binary classification problem, of which we used 1,005 images of "1" as positive examples and we used 6,268 images of other digits as negative examples. We determined that the number of dimensions of the projected subspace was two using visual observation. For the classification, we used the k-nearest neighbor method in the projected subspace, and we conducted the experiment with k=3. We used 254 images as positive examples, 1,753 images as negative examples, and, in total, we used 2,007 images for the evaluation.

We used the following equation for the similarity matrix of the LPP so that classification became feasible:

$$S_{ij}^{\text{LPP}} = \begin{cases} 1 & l(x_i) = l(x_j) \\ 0 & \text{otherwise} \end{cases},$$

[2]Available from https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/

Table 1. Classification results of PCA

	positive	negative
true	251	13
false	3	1740

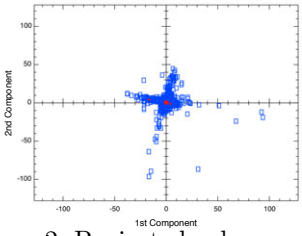


Figure 2. Projected subspace of PCA

where $l(.)$ is a class label. Data with the same class label is assigned 1, and 0 is given to data with other class labels. This similarity matrix is symmetric.

The similarity matrix of the ALPP adds the following neighborhood similarity to the symmetric similarity matrix of the LPP:

$$S_{ij}^{\text{asym}} = \begin{cases} 1 & x_i \in N_k(x_j) \\ 0 & \text{otherwise.} \end{cases}$$

As a result, the similarity matrix of the ALPP becomes asymmetric. In short, this similarity matrix has both the similarity of the identity of the learning labels and the similarity of the data proximity. In this experiment, we used $\alpha = 1/2$.

5.2 Results and Discussion

The PCA classification result is shown in Table 1, and the result of dimensional reduction is shown in Fig. 2. In PCA, the subspace is not divided into the positive and negative examples, and both data are grouped together. The classification result, however, is good.

The classification result of LPP is shown in Table 2, and the result of dimensional reduction is shown in Fig. 3. In LPP, the subspace separates the positive and negative examples because the similarity matrix is designed to gather nearby examples using the class label identities. The data, however, crosses at the boundary between the positive and negative examples, and the recognition rate of LPP is below that of PCA.

Table 3 shows the recognition rate when the subspace projected by ALPP is two dimensional space. The performance of ALPP is slightly improved in comparison to LPP. Because the subspace projected by ALPP is a complex subspace, the number of dimensions of the subspace is four. Figs. 4 (a) and (b) show graphs in which the horizontal axis is the first component and the vertical axis is the second component for the real and imaginary parts, respectively, of its components. As shown in Fig. 4 (a), the test data are separated into two classes in the real part, and the data is concentrated at the origin in the imaginary part, as shown in Fig. 4 (b). This means that there is almost

Table 2. Classification results of LPP

	positive	negative
true	248	16
false	9	1734

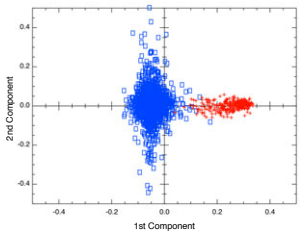


Figure 3. Projected subspace LPP

Table 3. Classification result of ALPP (two-dimensional complex space)

	positive	negative
true	248	16
false	6	1737

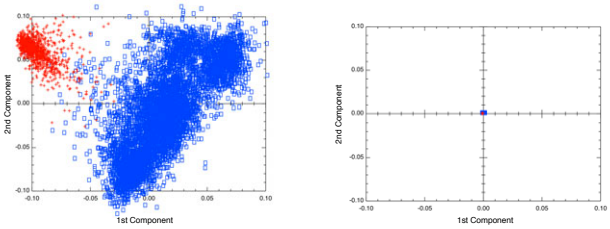
no component of the imaginary part of the projected data.

Just as for LPP, the data crosses at the boundary between positive and negative examples, and the recognition rate of ALPP is below that of PCA. Most data is on the real axis and the projected subspace is like one-dimensional space. This is because the similarity matrix itself is an asymmetric matrix but it is nearly a symmetric matrix. Because the size of the similarity matrix is $7,291 \times 7,291$, there are approximately 20 million symmetric pairs of elements. The number of asymmetric components in this experiment was 8,132 pairs, which was about 0.3% of the total. In order to evaluate the performance of ALPP, it seems necessary to conduct experiments using data with stronger asymmetry.

Finally, we show the examples that are correctly classified by the ALPP method but incorrectly classified by the LPP method in Fig. 5. The leftmost column shows the test image and other columns show the nearest training images. The upper row shows the training images selected by the LPP method and the lower row shows the training images selected by the ALPP method. As shown in this figure, the ALPP method gathers more similar images of positive examples than the LPP method.

6 Conclusion

We have proposed an ALPP that expands the LPP from a symmetric similarity matrix to an asymmetric similarity matrix using the Hermitian matrix for asymmetric representation. When we simply replace



(a) real part (b) imaginary part

Figure 4. Projected subspace of ALPP (two-dimensional complex space)

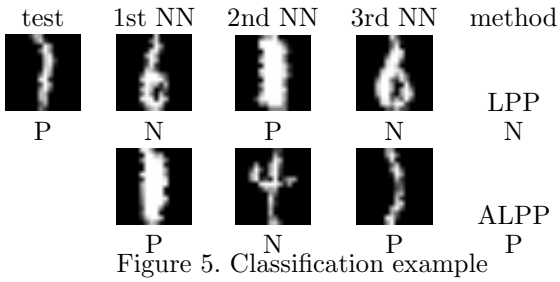


Figure 5. Classification example

the optimization problem with one using the Hermitian matrix, the optimization problem becomes complex, which cannot be totally ordered and sometimes does not make sense. We have, however, shown that it could be calculated well by devising a method to do so.

Although it is necessary to examine the meaning of the distance between two points on the complex subspace obtained by the ALPP, our method can be applied in many ways because there is an enormous amount of data with asymmetric similarly in the real world. In this paper, we have not compared and examined a wide variety of data. We will evaluate the performance of our method by increasing the variety of experimental data in future. We will also consider whether it can be applied to other methods that use the pairwise expression, e.g., sparse learning, or metric learning.

References

[1] M. Sugiyama, T. Ide, S. Nakajima, and J. Sense, "Semi-supervised local fisher discriminant analysis for dimensionality reduction," Machine Learning, Vol. 78, No. 1-2, pp. 35–61, 2010.

[2] A. Kimura, M. Sugiyama, H. Sakano, and H. Kameoka, "Designing various component analysis at will via generalized pairwise expression," IPSJ Trans. on Mathematical Modeling and Its Applications, Vol. 6, No. 1, pp. 136–145, 2013.

[3] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, "Basic local alignment search tool," Molecular Biology, Vol. 215, pp. 403–410, 1990.

[4] B. Haasdonk and D. Keysers, "Tangent distance kernels for support vector machines," Proc. of ICPR, Vol. 2, pp. 864–868, 2002.

[5] J. Laub and K.R. Müller, "Feature discovery in non-metric pairwise data," Machine Learning Research, Vol. 5, pp. 801–808, 2004.

[6] X. He and P. Niyogi, "Locality preserving projection," Advances in Neural Information Processing Systems, Vol. 16, 2003.

[7] N. Chino and K. Shiraiwa, "Geometrical structures of some non-distance models for asymmetric MDS," Behaviormetrika, Vol. 20, pp. 35–47, 1993.

[8] J.J. Hull, "A database for handwritten text recognition research," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 16, No. 5, pp. 550–554, 1994.