

NON-NEGATIVE MATRIX FACTORIZATION WITH MIXTURE OF ITAKURA-SAITO DIVERGENCE FOR SAR IMAGES

Chi Liu^{1,2}, Wenzhi Liao¹, Heng-Chao Li², Wilfried Philips¹

¹Ghent University, TELIN-IPI-IMEC, Sint-Pietersnieuwstraat 41, B-9000 Ghent, Belgium.

²Sichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University, Chengdu 611756, China

ABSTRACT

Synthetic aperture radar (SAR) data are becoming more and more accessible and have been widely used in many applications. To effectively and efficiently represent multiple SAR images, we propose the mixture of Itakura-Saito (IS) divergence for non-negative matrix factorization (NMF) to perform the dimension reduction. Our proposed method incorporates the unit-mean Gamma mixture model into the NMF to model the multiplicative noise. To obtain the closed-form update equations as much as possible, we approximate the log-likelihood function with its lower bound. Finally, we apply Expectation-Maximization (EM) algorithm to estimate the parameters, resulting in the closed-form multiplicative update rules for the two matrix factors. Experimental results on real SAR dataset demonstrate the effectiveness of the proposed method and its applicability to post-applications (e.g., classification) with improved performances over the conventional dimension reduction methods.

Index Terms— Non-negative matrix factorization (NMF), synthetic aperture radar (SAR), dimension reduction, Itakura-Saito divergence, mixture model.

1. INTRODUCTION

Synthetic aperture radar (SAR) is an active imaging system, which possess the capability of penetration and can operate all-day and under all-weather condition. It has been widely used in many applications such as urban planning, object detection and ocean surveillance. Recent advances in SAR technology have led to an increased availability of the multitemporal and multi-resolution SAR data, providing much richer information for Earth observation. However, representing and interpreting huge amount of SAR data remains challenging.

Many efforts have been devoted to dimension reduction and feature extraction of multiple SAR images. The implementations of the Laplacian eigenmaps and principle component analysis (PCA) are respectively reported in the classifi-

cation and change detection of SAR data to relax the computational burden and better represent the SAR data [1, 2]. Non-negative matrix factorization (NMF) [3] is an alternative technique for dimension reduction and data analysis, and found to be effective in the SAR data representation and feature extraction [4, 5]. Unlike PCA or independent component analysis (ICA), the elements in the resulting matrices by NMF are all guaranteed to be positive, which can provide intuitive and visual interpretation. Among the divergence measures used in NMF, the Itakura-Saito (IS) divergence [6, 7] is one of the most popular and appropriate methods for dimension reduction of SAR data, as the IS divergence and many successful distributions for SAR data are both derived with the help of multiplicative model [8]. However, the NMF with a single IS divergence is not sufficient to cope multiplicative noise (speckle) of multiple SAR images from homogeneous, heterogeneous and extremely heterogeneous areas.

In this paper, we propose a generalized IS divergence for NMF to represent multiple SAR images in a feature space. Our proposed method exploits the mixture of the IS divergence (MoIS), which is based on mixture of unit-mean Gamma distribution. In particular, the commonly used IS divergence is a special case of our proposed MoIS divergence. By considering multiplicative noise, our proposed method can better represent the multiple SAR images, and enables better performances for post-applications (e.g., classification) than conventional dimension reduction methods.

2. PRELIMINARIES

2.1. Non-negative Matrix Factorization

NMF, which is a popular technique for dimension reduction and data analysis, can approximate an observation matrix with the product of two matrix factors. Specifically, given a non-negative matrix $\mathbf{X} = [x_{nm}]_{N \times M} \in \mathbb{R}_+^{N \times M}$, NMF seeks to find the non-negative matrices $\mathbf{W} = [w_{nr}]_{N \times R} \in \mathbb{R}_+^{N \times R}$ and $\mathbf{H} = [h_{rm}]_{R \times M} \in \mathbb{R}_+^{R \times M}$, such that $\mathbf{X} \approx \mathbf{WH} = \hat{\mathbf{X}}$, where $\hat{\mathbf{X}} = [\hat{x}_{nm}]_{N \times M}$ is the approximated matrix. Generally, the rank R is selected such that $(M+N) \times R \ll M \times N$.

Thanks to the FWO project G037115N: Data fusion for image analysis in remote sensing and the National Natural Science Foundation of China under Grant 61371165.

Under some divergence measures, the matrix factors \mathbf{W} and \mathbf{H} can be obtained by minimizing the divergence between the observation matrix \mathbf{X} and the approximated matrix $\hat{\mathbf{X}}$ [3]. The resulting optimization problem turns out to be

$$\begin{aligned} \min_{\mathbf{W}, \mathbf{H}} \quad & \mathcal{D}(\mathbf{X}|\mathbf{W}\mathbf{H}) = \sum_{n,m} d(x_{nm}|\hat{x}_{nm}) \\ \text{s.t.} \quad & \mathbf{W} \geq 0, \mathbf{H} \geq 0, \end{aligned} \quad (1)$$

where $d(x_{nm}|\hat{x}_{nm})$ is the scalar divergence measure between two elements. The commonly used divergence includes Euclid (EUC) distance, Kullback-Leibler (KL) [3] and Itakura-Saito (IS) [6] divergences, which are respectively given by $d_{EUC}(x|y) = (x - y)^2$, $d_{KL}(x|y) = x \ln(\frac{x}{y}) - x + y$ and $d_{IS}(x|y) = \frac{x}{y} - \ln \frac{x}{y} - 1$. The NMFs based on these divergences are denoted as EUC-NMF, KL-NMF and IS-NMF, respectively.

2.2. Itakura-Saito Divergence NMF

IS divergence is first proposed in 1968 by Itakura and Saito and has been widely used in speech and music signal processing [6, 7]. Under the assumption of multiplicative Gamma noise, IS divergence can be derived from a generative model, which is used to estimate \mathbf{W} and \mathbf{H} [7]. Specifically, the observation matrix is modeled by the product of its NMF and noise matrix, which can be written as

$$\mathbf{X} = \mathbf{E} \odot (\mathbf{W}\mathbf{H}), \quad (2)$$

where the elements $[e_{nm}]_{N \times M}$ in \mathbf{E} are independently identical distributed (i.i.d) noises and the operator \odot is the element-wise product. The unit-mean noise is always assumed so that $\mathbf{E}[x_{nm}] = \sum_r w_{nr} h_{rm} = \hat{x}_{nm}$. Let every e_{nm} follow the unit-mean Gamma distribution defined by

$$Gam(x, \alpha) = \frac{\alpha^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\alpha x\}, \quad (3)$$

we have the log-likelihood function

$$L = \sum_{n,m} \ln \left(\frac{1}{\hat{x}_{nm}} \cdot Gam(x_{nm}|\hat{x}_{nm}, \alpha) \right). \quad (4)$$

On this basis, the IS divergence is the negative log-likelihood function when α is a constant and the constant terms are omitted.

3. PROPOSED MIXTURE OF IS DIVERGENCE NMF AND ALGORITHM

Mixture models can be interpreted as a weighted summation of distributions, which allows for great flexibility to model various signals. It has found wide applications such as image classification, segmentation and blind source separation. In view of its flexibility, we exploit the use of the mixture model in NMF for dimension reduction of multiple SAR images.

3.1. MoIS-NMF Model

In IS-NMF, the i.i.d noise is assumed to follow the unit-mean Gamma distribution. To improve the flexibility to model the i.i.d noise, we propose to use the unit-mean Gamma mixture model. Thus, in conjunction with equation (2), an observation follows the distribution of

$$f(x_{nm}|\mathbf{\Pi}, \boldsymbol{\alpha}, \mathbf{W}, \mathbf{H}) = \sum_{k=1}^K \frac{\pi_k}{\hat{x}_{nm}} \cdot Gam\left(\frac{x_{nm}}{\hat{x}_{nm}}, \alpha_k\right), \quad (5)$$

where K is the number of components and $\mathbf{\Pi} = \{\pi_1, \dots, \pi_K\}$ is the set of the positive mixing coefficients for the components, which ensures $\sum_{k=1}^K \pi_k = 1$. $\boldsymbol{\alpha} = \{\alpha_1, \dots, \alpha_K\}$ is the parameter set for the component distributions.

To formulate the generative model, the binary vector $\mathbf{z}_{nm} = [z_{nm1}, \dots, z_{nmK}]$ for each element in the observation matrix is introduced. There is only one element in \mathbf{z}_{nm} that is equal to 1 (i.e. $z_{nmk} = 1$ and $z_{nmj} = 0$ for $j \neq k$), which reveals that the element (n, m) belongs to the k -th component. The binary vector follows the multinomial distribution conditioned on $\mathbf{\Pi}$. Thus, the log-likelihood function can be obtained. However, in view of the terms $-\ln(\sum_l w_{nr} h_{rm})$ and $-\frac{1}{\sum_r w_{nr} h_{rm}}$, the update equations for w_{nr} and h_{rm} can not be obtained in closed form.

Since $-\ln(x)$ is convex and $-\frac{1}{x}$ concave, the lower bound of the log-likelihood function can be obtained by implementing the 1st-order Taylor approximation and Jensen's inequality, respectively. Thus, we find the approximated log-likelihood function $\tilde{\mathcal{L}}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{\Pi}, \boldsymbol{\alpha}, \mathbf{W}, \mathbf{H})$ as

$$\begin{aligned} L &\geq \tilde{\mathcal{L}}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{\Pi}, \boldsymbol{\alpha}, \mathbf{W}, \mathbf{H}) \\ &= \sum_{n,m,k} z_{nm} \left\{ \ln \pi_k + \alpha_k \ln \alpha_k - \ln \Gamma(\alpha_k) \right. \\ &\quad \left. + (\alpha_k - 1) \ln x_{nm} - \alpha_k x_{nm} \left(\sum_r \frac{\lambda_{nr}^2}{w_{nr} h_{rm}} \right) \right. \\ &\quad \left. - \alpha_k \left[\ln \beta_{nm} + \left(\sum_r w_{nr} h_{rm} - \beta_{nm} \right) / \beta_{nm} \right] \right\}. \end{aligned} \quad (6)$$

In the following, $\tilde{\mathcal{L}}(\boldsymbol{\beta}, \boldsymbol{\lambda}, \mathbf{\Pi}, \boldsymbol{\alpha}, \mathbf{W}, \mathbf{H})$ is denoted by $\tilde{\mathcal{L}}$ to keep the notation uncluttered. With the help of this approximation, our goal is to maximize the lower bound $\tilde{\mathcal{L}}$ with respect to (w.r.t.) all π_k , α_k , β_{nm} , λ_{nr} , w_{nr} and h_{rm} , which can lead to the tractable solutions for w_{nr} and h_{rm} .

3.2. Parameter Estimation

With the hidden variables \mathbf{z}_{nm} being introduced in the generative model, we apply Expectation-Maximization (EM) algorithm [9] for the parameter estimation. In the E-step, the expectation of z_{nmk} is given by

$$\gamma_{nmk} = \frac{\pi_k Gam\left(\frac{x_{nm}}{\hat{x}_{nm}}, \alpha_k\right)}{\sum_k \pi_k Gam\left(\frac{x_{nm}}{\hat{x}_{nm}}, \alpha_k\right)}. \quad (7)$$

In the subsequent M-step, the approximated lower bound $\tilde{\mathcal{L}}$ will be successively increased by alternatively finding the solutions to the parameters in $\tilde{\mathcal{L}}$. According to the property of 1st-order Taylor approximation and the Jensen's inequality, the approximated lower bound can be tightened w.r.t. β and λ by the following update rules

$$\beta_{nm}^{(t+1)} = \sum_r w_{nr}^{(t)} h_{rm}^{(t)}, \quad (8)$$

$$\lambda_{nrm}^{(t+1)} = \frac{w_{nr}^{(t)} h_{rm}^{(t)}}{\sum_r w_{nr}^{(t)} h_{rm}^{(t)}}. \quad (9)$$

The update equations of the mixing coefficients in Π can be obtained in element-wised form by the Lagrange multiplier method, which leads to

$$\pi_k^{(t+1)} = \frac{\sum_{n,m} \gamma_{nmk}}{N \times M}. \quad (10)$$

As for the update of α , we can find the equation by setting the 1st-order derivative of $\tilde{\mathcal{L}}$ w.r.t. each α_k to zero and substituting the obtained $\beta^{(t+1)}$ and $\lambda^{(t+1)}$. Subsequently, by implementing numerical methods such as bisection method, the update rule for each α_k can be found. Similarly, by solving the corresponding 1st-derivative equations, the update equations for the matrix factors \mathbf{W} and \mathbf{H} in the proposed NMF are given in element-wise form by

$$w_{nr}^{(t+1)} = w_{nr}^{(t)} \left(\frac{\sum_m \eta_{nm} \cdot h_{rm}^{(t)} x_{nm} / q_{nm}^2}{\sum_m \eta_{nm} \cdot h_{rm}^{(t)} / q_{nm}} \right)^{1/2}, \quad (11)$$

$$h_{rm}^{(t+1)} = h_{rm}^{(t)} \left(\frac{\sum_n \eta_{nm} \cdot w_{nr}^{(t+1)} x_{nm} / q_{nm}^2}{\sum_n \eta_{nm} \cdot w_{nr}^{(t+1)} / q_{nm}} \right)^{1/2}, \quad (12)$$

where $\eta_{nm} = \sum_k \gamma_{nmk} \cdot \alpha_k^{(t+1)}$ and $q_{nm} = \sum_r w_{nr}^{(t)} \cdot h_{rm}^{(t)}$. By alternatively evaluating update equations (7)-(12) until the criterion is reached, all the parameters can be estimated.

3.3. Relation to IS Divergence

From the update equations of \mathbf{W} and \mathbf{H} , we can see that the IS divergence is a special case of our method. When there is only one component in the mixture model (i.e. $K = 1$), both the π_k and γ_{nmk} are guaranteed to be one. Further, by setting each α_k to be constant, the resulting η_{nm} in equations (11) and (12) are equal and constant. Under this parameter setting, it is found that the update equations of each w_{nr} and h_{rm} are exactly the same as that for the IS divergence. Moreover, by using the mixture model to describe the noise, we extend the IS divergence to the weighted combination of multiple IS divergence.

4. EXPERIMENTS AND DISCUSSION

Experiments are conducted based on nine registered C-band ERS SAR images which were acquired between 1992 and 1995 over Pavia, northern Italy. The terrain types of water, buildings and vegetation areas are included.

4.1. Estimated Noise

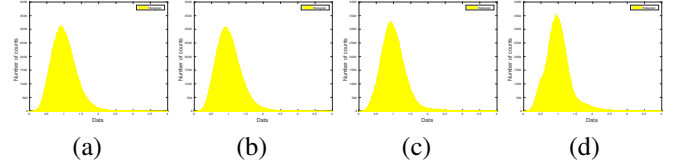


Fig. 1. The estimated noise distributions by (a) IS-NMF, (b) one-component MoIS-NMF, (c) two-component MoIS-NMF and (d) four-component MoIS-NMF.

To verify the effectiveness of the proposed MoIS-NMF, we will first investigate the estimated noise distribution to demonstrate the advantage of introducing the mixture model to the IS divergence. In view of the multiplicative noise model, the estimated noise for each pixel is evaluated in element-wise form by $e_{nm} = x_{nm} / \hat{x}_{nm}$. For the NMF with the rank of two, the estimated noise distributions with different number of components are illustrated in Fig. 1. The peak of the estimated noise distributions of the proposed MoIS-NMF is higher than IS-NMF. With more components, we adjust the distribution of the estimated noise, making them concentrated around the expected unitary mean better.

4.2. Application in Classification

To further validate the proposed MoIS-NMF, we apply the extracted features for classification. Specifically, the MoIS-NMF is first performed by setting the rank to one, and the resulting low rank matrix \mathbf{W} is used as the representation of the nine SAR images. On this basis, 100 labeled data in each class are randomly selected as the input to train the support vector machine (SVM) [10]. The classification maps are demonstrated in Fig. 2 and the classification accuracy for the terrain types is reported in Table 1.

According to Table 1, we get similar classification results for IS-NMF and one-component MoIS-NMF (1C-MoIS-NMF) and this can be explained by the relation between the IS divergence and the MoIS divergence as discussed in Section 3.3. In Fig 2, the vegetation, the roads, the urban areas and the water are correctly recognized. Most pixels in the background are classified as vegetation, which is consistent with the observation. The two-component MoIS-NMF can provide better classification result for the terrain types of water and vegetation. Although the classification based on the EUC-NMF performs better for the buildings, there are less miscellaneous

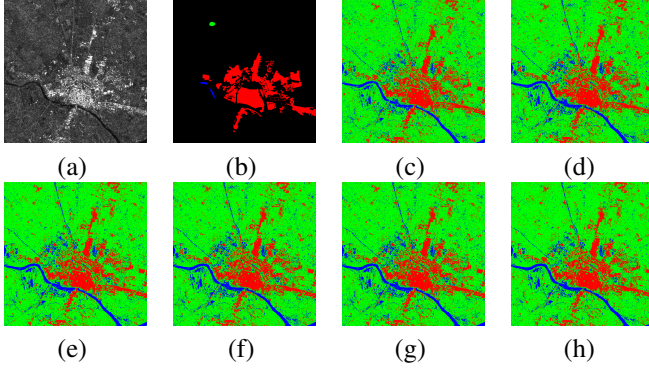


Fig. 2. (a) One of the SAR images. (b) The known labeled SAR data map, where the red pixels represent the buildings/urban areas, the blue ones for the water, the green ones for the vegetation and the black ones for the background. The resulting classification maps are respectively presented based on (c) the PCA, (d) EUC-NMF, (e) KL-NMF, (f) IS-NMF, (g) one-component MoIS-NMF (1C-MoIS-NMF) and (h) two-component MoIS-NMF (2C-MoIS-NMF).

pixels in the background for the two-component MoIS-NMF (2C-MoIS-NMF) by comparing Fig. 2(d) and (h), leading to better visual interpretation.

Table 1. Classification Accuracy.

Methods	Accuracy for terrain types		
	Buildings	Water	Vegetation
PCA	0.649	0.821	0.948
EUC-NMF	0.674	0.872	0.958
KL-NMF	0.651	0.871	0.973
IS-NMF	0.660	0.873	0.963
1C-MoIS-NMF	0.660	0.873	0.963
2C-MoIS-NMF	0.661	0.887	0.980

5. CONCLUSION

In this paper, we have proposed the mixture of IS divergence NMF (MoIS-NMF) for dimension reduction of multiple SAR images. Our proposed method incorporates the mixture of unit-mean Gamma distribution into NMF. Thus, the IS divergence is extended to the mixture of IS divergence. The algorithm of the MoIS-NMF is developed within the framework of the EM algorithm leading to the closed-form multiplicative updates for the matrix factors. It is demonstrated in the experiments that the proposed MoIS-NMF can adjust the distribution of the estimated noise to be concentrated around the unitary mean. Furthermore, the proposed MoIS-NMF can provide an appropriate representation for multiple SAR images and enables improved performances on post-applications (e.g., classification).

6. REFERENCES

- [1] S. T. Tu, J. Y. Chen, W. Yang, and H. Sun, "Laplacian eigenmaps-based polarimetric dimensionality reduction for SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 170–179, Jan. 2012.
- [2] O. Yousif and Y. Ban, "Improving urban change detection from multitemporal SAR images using PCA-NLM," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2032–2041, Apr. 2013.
- [3] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [4] M. Babae, X. Yu, G. Rigoll, and M. Datcu, "Immersive interactive SAR image representation using non-negative matrix factorization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 2844–2853, Jul. 2016.
- [5] J. Karvonen and A. Kaarna, "Sea ice SAR feature extraction by non-negative matrix and tensor factorization," in *Proc. IEEE IGARSS. 2008*, Jul. 2008, vol. 4, pp. IV–1093–IV–1096.
- [6] F. Itakura and S. Saito, "Analysis synthesis telephony based on the maximum likelihood method," in *Proceedings of the 6th International Congress on Acoustics*. pp. C17–C20, 1968, vol. 17, pp. C17–C20.
- [7] C. Févotte, N. Bertin, and J. Durrieu, "Nonnegative matrix factorization with the itakura-saito divergence: With application to music analysis," *Neural computation*, vol. 21, no. 3, pp. 793–830, Mar. 2009.
- [8] C. C. Freitas, A. C. Frery, and A. H. Correia, "The polarimetric \mathcal{G} distribution for SAR data analysis," *Environmetrics*, vol. 16, no. 1, pp. 13–31, Feb. 2005.
- [9] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *J. Royal Statist. Soc., Ser. B*, vol. 39, no. 1, pp. 1–38, 1977.
- [10] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.