Position-Aware Non-negative Matrix Factorization for Satellite Image Representation

Mohammadreza Babaee, Institute for Human-Machine Communication, Technische Universität München, Germany
Gerhard Rigoll, Institute for Human-Machine Communication, Technische Universität München, Germany
Mihai Datcu, German Aerospace Center (DLR), Wessling, Germany

Abstract

Satellite images clustering is a challenging problem in remote sensing and machine vision, where each image content is represented by a high-dimensional feature vector. However, the feature vectors might not be appropriate to express the semantic content of images, which eventually leads to poor results in clustering and classification. To tackle this problem, we propose a novel approach to generate compact and informative features from image content. To this end, we utilize geometrical information (as meta data accompanied with images) in the context of Non-negative Matrix Factorization (NMF) to generate new features. We assess the quality of new features by applying k-means clustering on the generated features and compare the obtained clustering results with those achieved by original features. We perform experiments on several satellite image data sets represented by different state-of-the-art features and demonstrate the effectiveness of the proposed method.

1 Introduction

Satellite image patches clustering has attracted strong attention in the recent years due to noisy and complex image content [1]. Several state-of-the-art techniques have been proposed to represent the content of images [2, 3]. For instance, these techniques are Scale Invariant Feature Transform (SIFT) descriptors [4], Gabor descriptors [5], and Weber descriptors [6] are used in a Bag-of-Words model [7] to represent each image with a high-dimensional feature vector. However, experimental results show that these techniques are not strong enough to prove discriminative representation. However, satellite images are normally accompanied with meta data that includes the position of places where they are taken from. Intuitively, we expect that those images that are captured from close regions should belong to the same class. For instance, if two image patches are taken from two regions with several meters far away from each other, then with high probability these two image patches belong to the same class.

In this paper, we aim to use the position of images and also the current feature vectors in a novel non-negative matrix factorization in order to generate new features. To this end, we compute two neighborhood graphs based on 1) the euclidean distances between feature vectors and 2) the real distance between the position of regions. These two graphs are used in two constraints coupled to the main objective function of non-negative matrix factorization to factorize the current features and generate new features. These constraints are controlled by two parameters. We claim that the new features are much more compact and also discriminative than current features.

2 Position-Aware Non-negative Matrix Factorization

NMF is an unsupervised learning algorithm which allows the user to represent the data in a low dimension space. It uses the feature information of the image to find the new representation. As a result, some images that are semantically similar may have features that are far away from each other in new representation. This problem will affect the clustering result seriously. For satellite images, we have not only the images, but also the meta data which contains geographical positions of these images. Intuitively, the images from close geographical positions may contain the similar content. Thus, the usage of geographical position may help the NMF to ease this problem. Based on the original objective function, new constraints can be added to improve the performance of NMF and impose the practical meanings for the new representations. In this section, a *Geography Constraint Non-negative Matrix Factorization (GeoNMF)* algorithm is introduced to find a better representation with practical meaning.

2.1 Objective Function

The input data is represented as a matrix $X = [x_1, ..., x_N] \in \mathbb{R}^{D\times N}$, $x_i \in \mathbb{R}^D$, where N denotes the number of samples and D represents the feature dimension.

In Graph Non-Negative Matrix Factorization, the locality property is used. Locality property infers that two images that are closed in the high dimensional space should also be close in the low dimensional space. Consider a graph G_e with N vertices that each vertex is mapped to a image point. For image point x_j, we find its k nearest neighbors and create edges between x_j and its neighbors. Thus, the graph will contains N vertices and a certain number of edges. We can represent the graph with an adjacency matrix W with the size $N * N$ and 0 and 1 are the two possible values in this matrix. If $W_(i,j) = 1$, it means x_i and x_j are neighbor points in high dimensional space. The first regularization term can be formulated as

$$\begin{aligned} R_1 &= \frac{1}{2}\sum_{i,j}^{N} \|x_i - x_j\|^2 W_{ij} \\ &= \sum_{i=1}^{N} x_i^T x_i D_{ii} - \sum_{i,j=1}^{N} x_i^T x_j W_{ij} \\ &= Tr(V^T DV) - Tr(V^T WV) \\ &= Tr(V^T LV) \end{aligned} \tag{1}$$

Where D is a diagonal matrix that each entry is the corresponding column sum of W, $D_{ii} = \sum_{j=1}^{N} W_{ji}$. In equation (1), a new matrix L that is equal to $D - W$ is introduced to facilitate the computation [15].

Similarly, for the geographical distance, another graph G_g is introduced to represent the geographical position relationship. Here, we introduce another parameter θ to define the radius of neighbor position. For example, take point x_i as center, all the points whose geographical distance to point x_i are less than θ are considered as x_i's geographical neighbors. The edges between x_i and these neighbors are created to link them together. We can convert this graph into another adjacency matrix Q with the size $N * N$. Q is created based on the geographical distance.

$$Q_{i,j} = \begin{cases} 1 & \text{if } \|x_i - x_j\| <= \theta \\ 0 & \text{else} \end{cases} \tag{2}$$

The second regularization term can be formulated as

$$\begin{aligned} R_2 &= \frac{1}{2}\sum_{i,j}^{N} \|x_i - x_j\|^2 Q_{ij} \\ &= \sum_{i=1}^{N} x_i^T x_i P_{ii} - \sum_{i,j=1}^{N} x_i^T x_j Q_{ij} \\ &= Tr(V^T PV) - Tr(V^T QV) \\ &= Tr(V^T ZV) \end{aligned} \tag{3}$$

where $Z = P - Q$.

Add these two regularization terms to the original cost function. The cost function can be written as

$$\begin{aligned} C &= \|X - UV^T\|_F^2 + \lambda_1 tr(V^T LV) + \lambda_2 tr(V^T ZV) \\ &= \sum_{i}^{N}\sum_{j}^{N}(x_{ij} - \sum_{k=1}^{K} u_{ik}v_{jk})^2 + \frac{1}{2}\sum_{i,j}^{N}\|x_i - x_j\|^2 W_{ij} \\ &+ \frac{1}{2}\sum_{i,j}^{N}\|x_i - x_j\|^2 Q_{ij} \end{aligned} \tag{4}$$

In this equation, parameter λ_1 and λ_2 control the contribution of Euclidean distance and geographical distance in the objective function.

2.2 Optimizing rules

To minimize the cost function, Equation (4), we first expand it to

$$\begin{aligned} C &= Tr((X - UV^T)(X - UV^T)^T) + \lambda_1 Tr(V^T LV) \\ &+ \lambda_2 Tr(V^T ZV) \\ &= Tr(XX^T) - 2Tr(XVU^T) + Tr(UV^T VU^T) \\ &+ \lambda_1 Tr(V^T LV) + \lambda_2 Tr(V^T ZV) \end{aligned} \tag{5}$$

We define Lagrange multiplier α_{ik} and β_{jk} for the constraints $u_{ik} \geq 0$ and $v_{jk} \geq 0$, respectively. Therefore, by defining $A = [\alpha_{ik}]$ and $B = [\beta_{jk}]$, the Lagrangian $\mathcal{L}$ is

$$\begin{aligned} \mathcal{L} &= Tr(XX^T) - 2Tr(XVU^T) + Tr(UV^T VU^T) \\ &+ \lambda_1 Tr(V^T LV) + \lambda_2 Tr(V^T ZV) + Tr(AU) + Tr(BV) \end{aligned} \tag{6}$$

The partial derivatives of $\mathcal{L}$ with respect to U, V are

$$\frac{\partial \mathcal{L}}{\partial U} = -2XV + 2UV^T V + A \tag{7}$$

$$\frac{\partial \mathcal{L}}{\partial V} = -2X^T U + 2VU^T U + 2\lambda_1 LV + 2\lambda_2 ZV + B \tag{8}$$

Using the Karush-Kuhn-Tucker (KKT) conditions [16], where $\alpha_{ij}u_{ij} = 0$ and $\beta_{ij}v_{ij} = 0$, the following equations are obtained:

$$-(XV)_{ik}u_{ik} + (UV^T V)_{ik}u_{ik} = 0 \tag{9}$$

$$[-X^T U + VU^T U + \lambda_1 LV + \lambda_2 ZV]_{jk}v_{jk} = 0 \tag{10}$$

The updating rules for U and V can be written as:

$$u_{ik} \leftarrow u_{ik}\frac{(XV)_{ik}}{(UV^T V)_{ik}} \tag{11}$$

$$v_{jk} \leftarrow v_{jk}\frac{X^T U + \lambda_1 WV_{jk} + \lambda_2 QV_{jk}}{VU^T U + \lambda_1 DV_{jk} + \lambda_2 PV_{jk}} \tag{12}$$

The convergence of updating rules can be proved using an auxiliary function similar to the one used in [17].

3 Experiments

In this section, we evaluate the performance of proposed algorithm for data clustering on different data set from three satellite images. We compare our algorithm with other methods like PCA, GNMF and so on.

3.1 Feature description

We create three data sets from three satellite images shown in Figure 1, first image is color image and the rest two are gray images. Among them, two data sets have 5 classes and each class contains 40 images. Another data set has 4 classes and each class contains 50 images. We use the raw data as the feature of each images for the experiments.

(a) (b) (c)

Figure 1: Three satellite image data sets used in our experiments.

3.2 Evaluation metrics

We use two metrics to evaluate the performance of the compared algorithms, namely accuracy (AC) and normalized Mutual Information (nMI) [18]. The accuracy represents the percentage of correctly predicted groups compared to the true labels. Given a dataset with N samples, where for each sample, c_i indicates its label given by the dataset and l_i the label predicted by the algorithm, the accuracy is defined as

$$AC = \frac{\sum_{i=1}^{N} \delta\left(c_i, \mathrm{map}(l_i)\right)}{N}, \tag{13}$$

where $\delta(x, y)$ is 1 if $x = y$ and 0 otherwise, and $\mathrm{map}(l_i)$ is a function that maps each label to the corresponding label in the dataset. The permutation mapping is determined using the Kuhn-Munkres algorithm [19].
The similarity of two clusters is determined by normalized mutual information. Given two sets of clusters $C = \{c_1, ..., c_k\}$ and $\acute{C} = \left\{\acute{c}_1, ..., \acute{c}_{\acute{k}}\right\}$, the mutual information metric is computed by

$$MI(C, \acute{C}) = \sum_{c_i \in C, \acute{c}_j \in \acute{C}} p(c_i, \acute{c}_j) \log \frac{p(c_i, \acute{c}_j)}{p(c_i)p(\acute{c}_j)}, \tag{14}$$

where $p(c_i), p(\acute{c}_j)$ represent the probability that an arbitrarily selected data point belongs to the clusters C or $\acute{C}_j$, respectively, and $p(c_i, \acute{c}_j)$ represents the joint probability that a point belongs to both clusters simultaneously. As the similarity of the two clusters increases, the mutual information $MI(C, \acute{C})$ increases from 0 to $\max\left\{H(C), H(\acute{C})\right\}$. $H(C), H(\acute{C})$ represent the entropy of the clusters $C, \acute{C}$ respectively.

$$nMI(C, \acute{C}) = \frac{MI(C, \acute{C})}{\max\left\{H(C), H(\acute{C})\right\}}. \tag{15}$$

3.3 Design

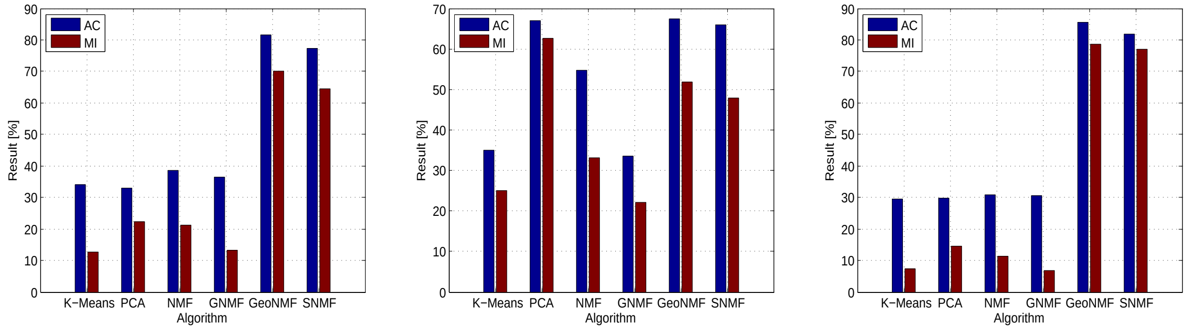
For the experiments, we use the raw data of image as feature information. While finding the neighbors of a data point on the graph, we choose the data point that has minimum Euclidean distance to current data point as nearest neighbor and build an edge between them in graph, namely set 1 in the W matrix. For the geographical distance, we choose $\theta = 200$ meter. For a data point, only the data points whose geographical distance to current data point are less than 200 meter are considered as candidate neighbors. Among these neighbors, we choose 2 closest neighbors and build edges between them in graph, namely set 1 in the Q matrix.

3.4 Result and Discussion

We compare the results of GeoNMF with applying K-Means algorithm directly on original high dimension data, PCA, NMF, GNMF, consider geographical distance only, which is called Spatial NMF. From the figures we can see that the performance of SNMF and GeoNMF outweighs the rest methods. The improvement is about 40%. When we consider the geographical positions only, we can get much better result than consider the features of images. This verifies our proposal that images from close geographical positions are prone to have comparing content that may not be easily distinguished by computer. Compare GeoNMF and SNMF, we can find that if we also consider the features of image, we can still get little improvement, about $2\% - 5\%$ in accuracy and mutual information. From all the results, we can draw a conclusion that considering the geographical distance will help the NMF with significant improvement. Compared with Euclidean distance which is measured in feature domain, the geographical distance dominant the influence of regularizer.

4 Conclusion

In this paper we have introduced a novel image representation technique to represent the content of satellite images content. Here, we used geometrical information as well as the original features as constraints in the process of non-negative matrix factorization to generate new features from image content. Experimental results show that the proposed method efficiently consider these both information to generate discriminative features that lead to higher accuracy in clustering applications. As future work, it would be interesting to also investigate geometric information in other feature learning process such as feature coding in the classification applications.



(a) (b) (c)

Figure 2: Accuracy and mutual information for three data sets, city image, SAR with MGB feature, SAR with EEC feature correspondingly.

References

[1] J. Curlander and R. McDonough, *Synthetic Aperture Radar: Systems and Signal Processing*, ser. Wiley Series in Remote Sensing and Image Processing. Wiley, 1991.

[2] M. Babaee, S. Tsoukalas, G. Rigoll, and M. Datcu, "Visualization-based active learning for the annotation of sar images," *Selected Topics in Applied Earth Observations and Remote Sensing (J-STARS), IEEE journal of*, 2015.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification (2Nd Edition)*. Wiley-Interscience, 2000.

[4] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[5] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *Image processing, IEEE Transactions on*, vol. 11, no. 4, pp. 467–476, 2002.

[6] J. Chen, S. Shan, G. Zhao, X. Chen, W. Gao, and M. Pietikainen, "A robust descriptor based on weber's law," in *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*. IEEE, 2008, pp. 1–7.

[7] M. Babaee, G. Rigoll, and M. Datcu, "Immersive interactive information mining with application to earth observation data retrieval," in *Availability, Reliability, and Security in Information Systems and HCI*. Springer, 2013, pp. 376–386.

[8] 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.

[9] M. Babaee, R. Bahmanyar, G. Rigoll, and M. Datcu, "Farness preserving non-negative matrix factorization," in *Image Processing (ICIP), IEEE International Conference On*, 2014, pp. 3023–3027.

[10] L. Haifeng, W. Zhaohui, L. Xuelong, C. Deng, and T. Huang, "Constrained non-negative matrix factorization for image representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 7, pp. 1299–1311, July 2012.

[11] S. Jia and Y. Qian, "Constrained non-negative matrix factorization for hyperspectral unmixing," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 47, no. 1, pp. 161–173, 2009.

[12] N. Gillis and R. J. Plemmons, "Sparse nonnegative matrix underapproximation and its application to hyperspectral image analysis," *Linear Algebra and its Applications*, vol. 438, no. 10, pp. 3991–4007, 2013.

[13] D. Seung and L. Lee, "Algorithms for non-negative matrix factorization," *Advances in neural information processing systems*, vol. 13, pp. 556–562, 2001.

[14] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd Ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996.

[15] D. Cai, X. He, J. Han, and T. Huang, "Graph regularized non-negative matrix factorization for data representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 8, pp. 1548–1560, 2011.

[16] S. P. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 1–38, 1977.

[18] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*. ACM, 2003, pp. 267–273.

[19] M. D. Plummer and L. Lovász, *Matching theory*. Access Online via Elsevier, 1986.