Multi-Instance Multi-Label with Application to High Resolution Remote Sensing Images

DJELLALI Mohamed1,*
1College of computer science and technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016 china
*mohameddjellali@nuaa.edu.cn

Abstract. In this paper, we seek to solve the problem of remote sensing image classification using Multi Instance Multi Label Learning (MIML) framework, where each image contain multiple regions (instances) corresponds to multiple objects (labels). MIML framework is more efficient than traditional learning framework for complicated objects such as satellite images. Existing MIML algorithms such as MIMLboost, MIMLsvm and MIMLfast have been found useful in scene classification and can provide multiple labels for each instance in complicated objects. The proposed approach was performed in two steps: In the first step, we process the dataset using segmentation and feature extraction. Images are ambiguous because they contain numerous objects so we consider that each image is a bag and various blocks in the image are instances. In the second step, we apply MIML to classify each block of images. According to the experimental results, the proposed method outperforms the state-of-the-art methods.

1. Introduction
Recently, MIML attracts much attention in machine learning community. The MIML formalization is learning the model function: \(2^X \rightarrow 2^Y\) for dataset \((X_1, Y_1), \ldots, (X_i, Y_i)\), where \(X_i \subseteq X\) is set of instances. \({x_{i1}, x_{i2}, \ldots, x_{in}}\), \(Y_i \subseteq Y\) is set of associated labels \({y_{i1}, y_{i2}, \ldots, y_{im}}\) such that \(n\) and \(m\) are sizes of \(X_i\) and \(Y_i\) respectively. \((X_1, Y_1), \ldots, (X_i, Y_i)\), \(X_i \subseteq X\) is set of instances \({x_{i1}, x_{i2}, \ldots, x_{in}}\), \(Y_i \subseteq Y\) with the development of satellite technology, the plentiful information provided in high resolution images totally changes the perspective of satellite image classification. HR image as one individual object with only a single semantic label. Actually, with more information the HR image contain numerous semantic meaning and multi-labels emerged from image segmented regions instead of the entire image. Feng [1] introduced an algorithm of multi-instance semi-supervised learning with hierarchical sparse representation. Liu and al [2] presented the key instance detection task and propose a new framework as solution based on the relationship between instances. T. Gärtner [3] used multi-instance bags after extraction of feature vector in order to prepare multi-instance kernel function. In order to predict one label of multi instance bags, MIML framework was proposed by Zhou et al. [3, 4], which has advanced toward becoming another exploration area. MIML framework provides a useful formulation for HR satellite images. In MIML HR images are described as bags, each bag contains many instances associated to segmented regions of the original image. Labelling a bag as negative required that all its instances are negative. We defined a bag as positive if at least one from its instances is positive. The rest of this paper is organized as follows: In section 2 we introduce the related works for both remote sensing image classification and other application of MIML framework.
In section 3 we discuss the experimental results with preparation of dataset and evaluation criteria. The experiments and result analysis are given in section 4. Section 5 offers our conclusion.

2. Related work

2.1 MIML
Multi Label Learning (MLL) [5, 6], can be used in classification problems, where each sample can be mapped to various classes at the same time. The main function of MLL is to learn the classifier $H_{Mll}: X \rightarrow 2^Y$. In a set of MLL training set: $\{(x_i, Y_i)\}_{1 \leq i \leq n}$ $x_i$ is the i-th instance and $Y_i$ is the label set of $x_i$. In [7], Chaudhuri B proposed semi-supervised graph-theoretic method related to the approach of multi-label image retrieval problem. This approach diffuses the labels in the training images using a label correlation propagation technique to the unlabelled image in order to map them with multiple labels. In order to recuperate the images out of archives with the same classes of regions as the image in question. MLL has been applied in many real world fields, such as medical, banking, text classification and ecological studies [8, 9]. Numerous methods are proposed in order to solve the task of MLL for instance: ML k-Nearest neighbour, ML decision tree and ranking SVM are proposed in [10].

Multi Instance Learning (MIL) model was formalized by Dietterich et al, in which they propose algorithm for dry activity prediction [11]. The main function of MIL is: $L_{MIL}: 2^x \rightarrow \{-1, +1\}$ From a set of MIL training set $\{(X_i, Y_i)\}_{1 \leq i \leq n}$; $x_i$ is a bag $MIL = \{x_1^i, x_2^i, ..., x_m^i\}$ and $y_i \in \{-1, +1\}$ is the label of the bag $x_i$. Jaume amores [12], present comparative studies about multi-label solutions of MIL.

Zhou et al [13, 14] introduce two MIML algorithms, MIMLboost and MIMLsvm algorithm respectively and their application in dataset scene classification. Li et al [15] proposed a Key Instances Sharing among Related labels (KISAR) algorithm which can find key-instance activating several labels in a bag. In [16] they propose the RankLossSIM approach to optimize with label ranking loss. To effectively deal with large dataset and minimize time complexity of MIML, MIMLfast algorithm is proposed with an approximated ranking-loss by using stochastic gradient descent (SGD).

2.2 Remote sensing Images
Remote sensing data is about the object or phenomenon through devises sensor without contact the objects physically. The use of satellite sensors is to detect and classify ground objects for the interest of numerous fields: including geography, geology, military, intelligence and humanitarian application.

In fact remote sensing is referred to satellite or aircraft sensor devises, according to the process of emission/reception of signals, remote sensing is divided into active (signal is emitted by satellite and reflection is used in the sensor) and passive (reflection of sunlight is detected by sensor).

2.3 Feature extraction methods
- Discrete Cosine Transform (DCT) is applied for each block after subdividing the image into regions of 8x8 pixels. The precision of classifiers on DCT features extraction is better than statistical and texture feature extraction techniques. The main idea of DCT is that the intensity of an image is often high in the low frequency elements. Generally the DCT only uses the coefficient with the largest values to reconstruct the original image [17].
- Local Binary Pattern (LBP) corresponds the functions of the descriptor, i.e. a local neighbourhood is started from gray scale of middle pixels into a binary Pattern [18].
- Gabor texture features is efficient for analysing remote sensed imagery [19]. They were shown to surpass other texture features in the images classification. Gabor method uses a set of scales and orientations filters to extract the texture features [20].

We summary the general process of object detection and mention the step of previous feature extraction methods in Figure 1.
3. Proposed method

Remote sensing images are complicated where each sample can correspond to multi-instances and multi-labels at the same time. For example, a satellite image contain several region, each one can be associated to an instance. Simultaneously the image can have the part with buildings and trees. In this paper we apply MIML into remote sensing images. When we learn the model by numerous bags; each bag has several instances and several labels as well.

MIML is the continuation of classical supervised learning (MIL and MLL). In MIL[21] an object presented by several instances is attached to one class label. The training dataset used under the form \( \{(i_1, l_1), (i_2, l_2), ..., (i_n, l_n)\} \), where \( l_1 \subseteq I \) is set of instances \( \{i_1, i_2, ..., i_n\} \) is and \( n \) is number of instances, \( l_1 \in \{-1, +1\} \) is the label of \( l_1 \). In MLL [22] an object presented by one instance is attached with numerous class labels. The training dataset used under the form \( \{(i_1, L_1), (i_2, L_2), ..., (i_n, L_n)\} \), where \( i_k \in I \) is an instance and \( L_4 \) is an array of labels in class label \( L \).

MIML framework provides the mapping between objects instances and corresponding labels as show in figure 2. The classification on label \( l \) is defined by the function \( f: f_l(X) = W_l^T W_0 X \), \( W_0 \) is a \( m \times d \) matrix, \( W_l \) is \( m \)-dimensional weight vector corresponding to label \( l \), \( d \) and \( m \) are the sizes of the feature space and the shared space successively. The prediction of instance \( x \) on label \( l \) as:

\[
\hat{f}_l(X) = \max_{k=1..K} f_{l,k}(X) = \max_{k=1..K} W_{l,k}^T W_0 X
\]

According to Dietterich, a bag is positive equivalently it contains anu minimum one positive instance, the prediction of a bag \( X \) on label \( l \) is defined by the following function:

\[
f_l(X) = \max_{X \in X} f_l(X) .
\]

The basics parameters of gradient descent are as follow:

\[
W_0^{t+1} = W_0^t - \gamma_t S_{Y,Y} (W_{Y,Y}^t X - W_{Y,Y}^t X) \quad (2)
\]

\[
W_{Y,k}^{t+1} = W_{Y,k}^t + \gamma_t S_{Y,Y} W_0^t X \quad (3)
\]

\[
W_{Y,k}^{t+1} = W_{Y,k}^t - \gamma_t S_{Y,Y} W_0^t \hat{X} \quad (4)
\]

Such that \( \gamma_t \) is the step size of SGD, \( W_{Y,k} \) and \( W_{Y,k} \) and all column are standardised to Norm L2.
Figure 2. Four different learning frameworks

Algorithm
1: **INPUT:** training images, parameters m, C, K and t
2: **TRAIN:** initialization $w_0$ and $w_{t,k}$ ($i=1...L$, $k=1...K$)
3: **repeat:** randomly sample a bag $X$ and one of its relevant label $y$
4: select key instance and sub-concept by $(x, k) = \arg \max_{x \in X, k \in \{1...K\}} f_{y,k}(x)$
5: **for** $i = 1: |\hat{y}|$ ; 6: sample an irrelevant label $\hat{y}$ from $\hat{y}$
7: select key instance and sub-concept by $(\hat{x}, \hat{k}) = \arg \max_{x \in X, k \in \{1...K\}} f_{\hat{y},\hat{k}}(x)$
8: **if** $f_{\hat{y}}(x) > f_{y}(x)$-1; 9: $v = i$
10: update $w_0, w_{y,k}$ and $w_{\hat{y},\hat{k}}$ as Eqs. 4 to 9, and perform normalization; 11: break
12: **until** stop criterion reached
13: **OUTPUT:** Relevant labels set for the test bag $X_{test}$ is: \{l | 1 + f_l(x_{test}) > f_y(x_{test}) \}

Figure 3 show the scenario of image classification and the illustration according to selected approaches (MIL, MLL, MIML). MLL Principe associate the labels (building, trees and water) to the input image by founding the class objects in the image and fill the matrix instance-label by 1(-1) if it exist (not). MIL approach link multi sub images with one label if and only if at least one class object exist in the numerous sub images. Finally MIMIL provide the opportunity to label the image linking each sub image with its conform label and this is the developed approach to classify the image with complicated objects.

Figure 3. Explicative examples clarify the difference between MIL, MLL and MIMIL.
4. Experiment

4.1 Data collection
The dataset we used contains 2000 high resolution images affiliated to the classes: ground, water, buildings, trees and roads. We collected those images from Google earth in the area of California USA, more than 90% belong to multiple classes in the same image. We obtained 2000 sub-images after segmentation of original images. Totally we use 1800 samples for training and 200 for test, for each sample we divide it into four regions and extract the features using dct features. This project required to have labelled dataset in order to train the model and evaluate its performance. The images are resized to be a uniform scale for the object extraction.

4.2 Multi label Performance criteria
Multi-label classification problem needs special criteria for evaluation different from traditional signal label classification that uses just accuracy as performance evaluation criteria [23, 24]. In the experiment part we employ five criteria Hamming loss (h.l), One-error (o.e), Coverage (co), Ranking loss (r.l) and Average precision (a.p) to check the performance of the proposed MIML classification method. These criteria are defined as follows:

- **Hamming loss:** calculate how frequently an instance-label pair is wrong classified.
  \[ HL(h) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{|Y_i|} |h(X_i) \Delta Y_i| \]
  \( \Delta \) is the difference between two sets. The less value, the improved achievement.

- **One-error:** calculate how frequently the object is mislabelled. The ideal performance is OE(h) = 0; the less value of OE(h), the improved achievement of h.
  \[ OE(h) = \frac{1}{p} \sum_{i=1}^{p} \left[ \left\lfloor \text{argmax}_{y \in Y_i} h(X_i, y) \right\rfloor \notin Y_i \right] \]

- **Coverage:** calculate the required value, on the normal, to check the array of labels and include the corresponding labels of the object. It is approximately linked to the accuracy at the level of perfect recall. The less value of coverage (h), the improved achievement of h.
  \[ \text{coverage}(x) = \frac{1}{p} \sum_{i=1}^{p} \max_{y \in Y_i} \text{rank}^h(X_i, y) - 1 \]

- **Ranking loss:** Calculate the medium portion of labels that are not correctly requested.
  \[ \text{Rank Loss}(h) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{|Y_i|} \left\lfloor \left\lfloor \left\lfloor (y_1, y_2) \right\rfloor h(X_i, y_1) \leq f(X_i, y_2), (y_1, y_2) \in Y_i \times \bar{Y}_i \right\rfloor \right\rfloor \]
  Where the \( \bar{Y}_i \) denotes the supplementary set of \( Y_i \) in C. The performance is ideal when RankLoss (f) = 0. The less value, the improved achievement.

- **Average Precision:** Calculate the medium portion of labels classified to a certain label \( y \in Y_i \) which actually is in \( Y_i \).
  \[ \text{AverPrec}(h, D_o) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{|Y_i|} \sum_{y \in Y_i} \frac{|\left\lfloor \left\lfloor \text{rank}^h(X_i, y') \right\rfloor \right\rfloor \text{rank}^h(X_i, y) \neq Y_i|}{\text{rank}^h(X_i, y)} \]

4.3 Efficiency Comparison
Table 1 describes the experimental results of many feature extraction methods in order to compare the evaluation performance.

|       | h.l   | o.e   | co    | r.l   | a.p   |
|-------|-------|-------|-------|-------|-------|
| Lbp   | .272±.041 | .146±.01 | .2842±.041 | .024±.006 | .843±.002 |
| Dct   | .368±.014 | .177±.008 | .3181±.019 | .291±.008 | .792±.005 |
| Lbp+dct | .402±.074 | .239±.311 | .294±.041 | .0328±.121 | .793±.082 |
| gabor | .469±.065 | .409±.203 | .3267±.234 | .471±.107 | .702±.068 |
It is essential to evaluate the effectiveness of the model trained with MIML approach, for this purpose we calculate the five MIML criteria as shown in Figure 4. Finally we conclude that MIMLfast give more accurate result than two other methods (MIMLsvm, KISAR) especially with LBP feature extraction method.

5. Conclusion

MIML approach is the appropriate for Remote sensing object recognition thus it is created for learning the complicated objects. MIML can be comprehended by transformation MIML problem into its corresponding in classical supervised learning using MIML or MLL as intermediate. Although, this transformation may miss information in ignore the relation between labels. When taking in consideration the relation between labels in common area and finding sub-ideas for complex labels. Comparison results demonstrate the efficiency of proposed method in terms of classification accuracy. Future work will cover the optimization of hamming loss function by implementing active learning in the process of training the model.

References

[1] S. H. Feng, W. H. Xiong, B. Li, C. Y. Lang, and X. K. Huang, “Hierarchical sparse representation based multi-instance semisupervised learning with application to image categorization,” Signal Processing, vol. 94, pp. 595–607, 2014.
[2] G. Liu, J. Wu, and Z. H. Zhou, “Key instance detection in Multi-instance learning,” in Proceedings of the 4th Asian Conference on Machine Learning, SIN, pp. 253–268, 2012.
[3] T. Gärtner, P. A. Flach, A. Kowalczyk, and A. J. Smola, “Multi-instance kernels,” in Proceedings of the 19th International Conference on Machine Learning, Sydney, AU, pp. 179–186, 2002.
[4] Z. H. Zhou and M. L. Zhang, “Multi-instance multi-label learning with application to scene classification,” Advances in Neural Information Processing Systems 19, Cambridge, MA: MIT Press, pp. 1609–1616, 2007.
[5] A. McCallum, “Multi-label text casserification by EM,” AAAI’99 Workshop on Text Learning., 1999.
[6] R. E. Schapire and Y. Singer, “BOOSTEXTER: a boosting-based system for text categorization,” Machine Learning, vol. 39, no. 2/3, pp. 135–168, 2000.
[7] Chaudhuri B, Demir B, Chaudhuri S, et al. Multilabel Remote Sensing Image Retrieval Using a Semisupervised Graph-Theoretic Method[J]. IEEE Transactions on Geoscience & Remote Sensing, 2017, PP(99):1-15.
[8] Kang F, Jin R, Sukthankar R. Correlated Label Propagation with Application to Multi-label Learning[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2006:1719-1726.
[9] Guo L, Jin B, Yu R, et al. Multi-label classification methods for green computing and application for mobile medical recommendations[J]. IEEE Access, 2016, 4:3201-3209.
[10] Zhang M L, Zhou Z H. A Review on Multi-Label Learning Algorithms[J]. IEEE Transactions on Knowledge & Data Engineering, 2014, 26(8):1819-1837.
[11] T. G. Dietterich, R. H. Lathrop, T. Lozano-Perez, Solving the multiple-instance problem with axis-parallel rectangles, Artificial Intelligence 89 (1997) 31–71.
[12] Amores J. Multiple instance classification: Review, taxonomy and comparative study[J]. Artificial Intelligence, 2013, 201(4):81-105.
[13] Z. H. Zhou and M. L. Zhang, “Multi-instance multi-label learning with application to scene classification,” Advances in Neural Information Processing Systems 19, Cambridge, MA: MIT Press, pp. 1609–1616, 2007.
[14] Z. H. Zhou, M. L. Zhang, S. J. Huang, and Y. F. Li, “Multi-instance multi-label learning,” Artificial Intelligence, vol. 176, pp. 2291–2320, 2012.
[15] Y. F. Li, J. H. Hu, Y. Jiang, and Z. H. Zhou, “Towards discovering what patterns trigger what labels,” in Proceedings of the 26th AAAI Conference on Artificial Intelligence, Toronto, Ontario, CA, pp. 1012–1018, 2012.
[16] F. Briggs, X. Z. Fern, and R. Raich, “Rank-loss support instance machines for MIML instance annotation,” in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, Beijing, CHN, pp. 534–542, 2012.
[17] Al-Shehhi S, Al-Muhairi M, Abuoeida D, et al. Robust and Fragile Watermarking Scheme Based on DCT and Hash Function for Color Satellite Images[C]// International Conference on Developments in Eystems Engineering. IEEE, 2013:253-258.
[18] T. Ojala, M. Pietikainen, and T. Maenpaa, “MultiresolutionGray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns”, IEEE Transactionon Pattern Analysis and Machine Intelligence – PAMI, vol. 24, no. 7, 2002, pp. 971-987.
[19] S. Newsam, L. Wang, S. Bhagavathy, and B. S. Manju-nath, “Using texture to analyze and manage large collec-tions of remote sensed image and video data,”Journal of Applied Optics: Information Processing, vol. 43, no. 2, pp. 210–217, 2004.
[20] B. S. Manjunath and W. Y. Ma, “Texture features forbrowsing and retrieval of image data,”.IEEE Trans.OnPattern Analysis and Machine Intelligence, vol. 18, no.8, pp. 837–842, 1996
[21] S. Andrews, I. Tsochantaridis, and T. Hofmann, “Support vectormachines for multiple-instance learning,” in The Neural Information Processing Systems (NIPS), 2002, pp. 561–568.
[22] Kong X, Ng M K, Zhou Z H. Transductive Multilabel Learning via Label Set Propagation[J]. IEEE Transactions on Knowledge & Data Engineering, 2013, 25(3):704-719.
[23] Zhang M L, Zhou Z H. Multilabel Neural Networks with Applications to Functional Genomics and Text Categorization [J]. IEEE Transactions on Knowledge & Data Engineering, 2006, 18(10):1338 1351.
[24] Zhou Z H, Zhang M L, Huang S J, et al. Multi-Instance Multi-Label Learning[J]. Artificial Intelligence, 2008, 176(1):2291-2320.