Multiple Instance Learning algorithm for medical image classification

(Discussion Paper)

Annabella Astorino[1] [0000-0002-3439-180X], Antonio Fuduli[2] [0000-0002-1657-0257], Manlio Gaudioso[3][0000-0002-3022-7041] and Eugenio Vocaturo[3][0000-0001-7457-7118]

[1]ICAR-CNR National Research Council, Rende, Italy
annabella.astorino@icar.cnr.it

[2]Department of Mathematics and Computer Science,
University of Calabria, Rende, Italy
antonio.fuduli@unical.it

[3]DIMES - Department of Computer Engineering, Modelling, Electronic and System
University of Calabria, Rende, Italy
{gaudioso, e.vocaturo}@dimes.unical.it

Abstract. After an overview on most relevant methods for image classification, we focus on a recently proposed Multiple Instance Learning (MIL) approach, suitable for image processing applications and based on a mixed integer nonlinear optimization problem. In particular, the algorithm has been preliminarily applied to a set of color images, with the aim to identify images containing some specific color pattern, and successively to a medical dataset, containing photos of melanoma and common nevi. Since the results appear promising, this technique could be at the basis of computer vision systems that act as a filter mechanism to support physicians in detecting melanomas cancer.

Keywords: Image Classification, Multiple Instance Learning, Lagrangian Relaxation.

1 Introduction

Nowadays, continuous development of digital technologies and ICT solutions implies a great availability of digital images in various fields, such as scientific, medical and industrial. Use of this information requires navigation, searching, indexing, classification and retrieval. Image classification is aimed at classifying images according to their visual contents: a particular application is in the medical field, where medical images classification can support the diagnosis of some diseases.
Interesting algorithms to evaluate image similarities have been proposed in [9, 13]: they work by extracting first the features vectors from the image and then by using the vector distances as a measure of similarity. Since these methods, based on object recognition techniques, are sensitive to the noise and do not satisfactorily work when there are no

distinct objects in the images, De Bonet and Viola [8] proposed an algorithm that works well with natural scenes and single object test sets. Ravela et al. [16] proposed a method to evaluate the similarity by means a correlation measure: the application of this system involves a collection of regions of interest, but possible limitations are connected to the difficulties in identifying such regions. A General Image Retrieval (GIR) model that calculates the similarities between query image and the images in the image database was proposed in [11], while R. da Silva Torres et al. [19] released an interesting overview on content based image retrieval and its applications on digital libraries, crime prevention, medicine, fingerprint identification and biodiversity information systems. P.N. Shinde and A.A. Manjrekar [17] proposed different schemes to check similarity between two images: the images are classified, re-ranked and retrieved using an algorithm based on score histogram computation.
In this paper, we face image classification by means of a Multiple Instance Learning approach [1, 6] based on a Lagrangian relaxation technique [3]. Multiple instance learning (MIL) is a recent paradigm of machine learning that performs well in the context of medical images and video analysis [15]. Manual segmentation, typically needed in the training phase, is no more necessary since the supervision is based on global labels, unlike traditional single-instance learning algorithms. The paper is organized as follows. In the next section, we recall the main concepts at the basis of the Multiple Instance Learning paradigm, focusing on the binary case. In Section 3, we shortly describe the algorithm proposed in [3], while in Section 4 we report some preliminary numerical results performed on the classification of some artificial color images and on a data set of dermoscopic images. Finally, some conclusions are drawn in Section 5.

2 Multiple Instance Learning

The objective of the Multiple Instance Learning (MIL) is to discriminate among sets of points: the sets to be classified are called bags, while the points inside the bags are named instances. In the image classification, an image is represented by a bag, whereas the sub-regions forming the images are the instances. In a MIL problem, during the learning phase only the labels of the bags are known, whereas the labels of the instances are unknown. This is different from a classical supervised classification approach, where the label of each point is known. In this work, we focus on the binary MIL classification, whose objective is to discriminate between positive and negative images. Two different approaches are mainly possible: one is in the instance-space and the other one is in the bag-space. In the instance-space, the learning process looks only at the characteristics of each individual instance, without considering the more global characteristics of the entire bag. Vice-versa, in the bag-space case, the learning phase is more global, because each bag is treated as a whole entity: the classification consists in separating directly the positive bags from the negative ones. There exists also a compromise between these two approaches. It consists in representing each bag by means of some of its instances containing maximal similarities and it is referred as the embedding-space approach [6]. An important issue in binary MIL problems is to specify what a positive bag is. The standard assumption is that a bag is negative if all its instances are negative and positive whenever at least one instance is positive: as a consequence, under such an assumption, only the labels of the instances inside the positive bags are to be predicted. An extensive survey on classification of medical images by means of MIL techniques is reported in [15]. The proposed classification method, described in the next section, is a heuristic optimization algorithm [3] based on the SVM (Support Vector Machine) type model introduced in [2].We recall that optimization techniques, aimed at obtaining separation surfaces, play a relevant role in every machine learning problem and in literature various mathematical programming models have been proposed to face a binary MIL classification problem, such as [5, 12].

3 The Classification Algorithm

In this section we describe the heuristics optimization MIL algorithm proposed in [3], which is suitable for image classification. This method is based on the SVM type model introduced by Andrews et al. in [2]. In particular, let J^+ and J^- the index sets corresponding to m positive bags and k negative ones. The two sets are therefore defined respectively as: $J^+ = \{J_1^+, \dots, J_m^+\}$ and $J^- = \{J_1^-, \dots, J_k^-\}$. Let p be the total number of points (the instances) inside the bags.
The j_{th} instance, $j = (1, \dots, p)$ is represented by a vector $x_j \in \mathbb{R}^n$ and its class label $y_j \in \{-1,1\}$ is supposed to be unknown. Differently from the class label of the instances, the class label of each bag is known: it is equal to +1 for each positive bag $J_i^+ = 1, \dots, m$ and it is equal to -1 for each negative bag $J_i^- = 1, \dots, k$.
We hypothesize to work under the MIL standard assumption: we suppose that a bag is positive if it contains at least a positive instance, while it is negative if it contains only negative instances. The objective is to determine a hyperplane:

$$H(w,b) \triangleq \{x \,\epsilon \mathbb{R}^n \; | w^T \, x + b = 0\}$$

separating the two classes of bags, where $w \in \mathbb{R}^n$ is the normal to the hyperplane and $b \in \mathbb{R}$ is the bias. Since a negative bag is defined containing only negative instances, only the class label $y_j \in \{-1,1\}$ of the instances inside the positive bags are actually unknown. If in addition we want to minimize a measure of the classification error of all the instances inside the bags, the optimization model [2] is the following:

$$P \begin{cases} z^* = min_{w,b,y} \, f(w,b,y) \\ \displaystyle\sum_{j \in J_i^+} \frac{y_j + 1}{2} \geq 1, \qquad i = 1, \dots, m \\ y_j \in \{-1,1\}, \qquad j \in J_i^+ \;\; i = 1, \dots, m \end{cases}$$

where

$$f(w,b,y) \triangleq \frac{1}{2} \parallel w \parallel^2 + C \sum_{i=1}^{k} \sum_{j \in J_i^-} max\{0, 1 + (w^T x_j + b)\} + C \sum_{i=1}^{m} \sum_{j \in J_i^+} max\{0, 1 - y_j(w^T x_j + b)\}.$$

Note that problem P is a constrained, nonlinear, nonconvex and mixed integer program. Moreover, if variables y_j were fixed, we would obtain a classical supervised SVM model [7]. The objective function in problem P is the sum of three terms: the first one is aimed at maximizing the margin, while the other two ones are aimed at minimizing a measure of the classification error of the points belonging to the negative bags and of the points belonging to the positive bags, respectively. Note that the presence of the unknown labels y_j in the third term allows the possibility to allocate some points of positive bags in the negative part with respect to the separating hyperplane. Finally, by means of the constraints:

$$\sum_{j\in J_i^+} \frac{y_j+1}{2} \geq 1, \qquad i = 1, \dots, m \qquad (1)$$

we impose that, for each positive bag, at least one point must be labelled as a positive point. In [2], the authors proposed two different heuristic techniques, based on solving approximately problem P. Our optimization heuristic algorithm (see [3]) is based on the Lagrangian relaxation [10] of problem P, obtained by relaxing the linear constraints (1), i.e.:

$$P_{LR}(\lambda) \begin{cases} z_{LR}(\lambda) \triangleq min_{w,b,y}\, f_\lambda(w,b,y) \\ y_j \in \{-1,1\}, \quad j \in J_i^+ \;\; i = 1, \dots, m \end{cases}$$

with

$$f_\lambda(w,b,y) \triangleq f(w,b,y) \sum_{i=1}^{m} \lambda_i \left(1 - \sum_{j\in J_i^+} \frac{y_j+1}{2} \right)$$

where $\lambda \geq 0$ is the vector of the Lagrangian multipliers in $\mathbb{R}^m$. Any optimal solution of $P_{LR}(\lambda)$, denoted by $(w(\lambda), b(\lambda), y(\lambda))$, provides a lower bound for problem P, that is $z_{LR} \leq z^*$for any $\lambda \geq 0$. The Lagrangian dual problem is defined as:

$$P_{LD} \begin{cases} z_{LD} \triangleq max_{\lambda\geq 0}\, z_{LR}(\lambda) = max_{\lambda \geq 0}\, min_{w,b,y}\, f_\lambda(w,b,y) \\ y_j \in \{-1,1\}, \quad j \in J_i^+ \quad i = 1, \dots, m \end{cases}$$

with, of course, $z_{LD} \leq z^*$. In [3] it has been shown that $z_{LD} = z^*$: as a consequence, in solving P_{LD} it is possible to obtain an optimal solution to P by using any dual ascent method. Since problem P_{LD} is nondifferentiable, the use of nonsmooth optimization techniques is in order. We adopt the subgradient method [18], that we stop after a prefixed maximum number of iterations. Note that, at each iteration, it is necessary to evaluate the objective function $z_{LR}(\lambda)$. This is a difficult task because it requires in turn to face problem $P_{LR}(\lambda)$, that we solve approximately by means of a Block Coordinate Descent (BCD) algorithm [20]. In fact we alternately fix the value of vector y and of the couple (w, b): for any $\lambda \geq 0$, when variables y_j are kept fixed, problem $P_{LR}(\lambda)$, reduces to solving a classical SVM quadratic program; vice-versa, when the couple (w, b) is fixed, it is possible to solve it, with respect to y, by inspection.
In case BCD provides an infeasible solution to the original problem, such solution may be easily modified to obtain feasibility, giving, in this way, an upper bound (the incumbent) on the optimal solution of P. A simplified scheme of the MIL algorithm is reported in Figure 1; further details can be found in [3].

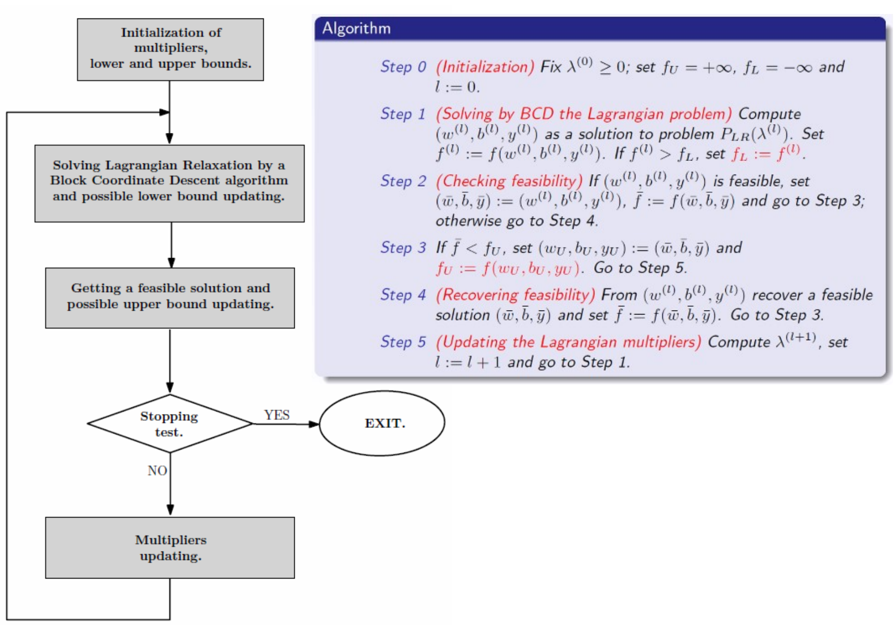


Fig. 1. The MIL Lagrangian relaxation approach.

4 Numerical Experiments

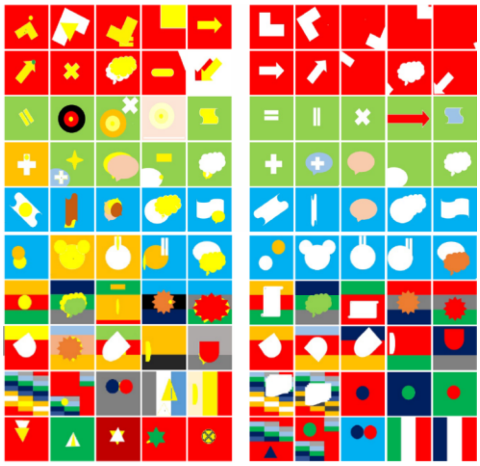
The algorithm presented in the previous section has been adopted in two types of numerical experimentations: preliminarily for classifying color images and successively for classifying dermoscopic images.

In the first numerical preliminary experiment (see [4]) we have generated one hundred color images of 128x128 pixels dimension, divided into two different classes, positive and negative. We have considered positive each image (bag) containing the yellow color and negative the images without yellow (Figure 2).

We have performed a segmentation process by means of some image processing standard Matlab routines. In particular, given a bitmap image, this image is read by the "*imread*" Matlab routine, which provides a 128x128 matrix, the indexed image, each element corresponding to a pixel and containing a triplet which represents the RGB (red, green, blue) scale. In fact, the first value of each triplet represents the red pixel intensity, the second one represents the green pixel intensity and the last one represents the blue pixel intensity. Once the indexed image has been generated, the successive step consists in converting each indexed image (and the corresponding color map) into an RGB image by means of the "*ind2rgb*" Matlab subroutine. After that, we have proceeded by grouping the pixels in square sub-regions of appropriate dimension: each sub-region forms the so called "*blob*". For each blob, we have computed the average of the RGB intensities and the differences between this value and the same quantity calculated for the adjacent blobs (up, down, left, right).

We recall that a crucial issue in a Support Vector Machine type approach is the choice of parameter C, which generally is performed among a grid of possible values.

We report the results obtained fixing C = 10, since in correspondence to this value the classification algorithm has exhibited the best performance. In discriminating between positive and negative color images, we have considered the following two cases. In the first one we have generated, for ten times, a testing set by choosing ten different images (five positive and five negative), so that each time the remaining ninety ones (forty-five positive and forty-five negative) have constituted the training set.
In the table reported in Figure 2, for each trial, we report the training and the testing correctness percentage, respectively, with the average for both of them in the last row.



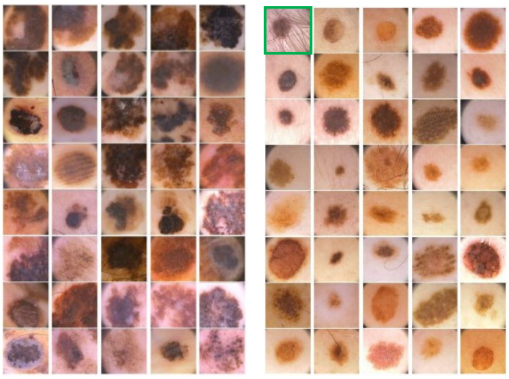
With Yellow   Without Yellow

#	Training correctness	Testing correctness
1	81.11	100
2	81.11	80
3	84.44	90
4	83.33	80
5	82.22	80
6	81.11	90
7	83.33	80
8	85.56	70
9	84.44	70
10	84,44	70
Average	83.11	81

Fig. 2. MIL-RL Results for Color Image classification

In the second case, we have performed a leave-one-out cross validation, obtaining 84.07% and 83% as training and testing correctness, respectively. We have noted that, among the seventeen failures in the testing set, fifteen are positive images and two are negative images. Considering the fifteen misclassified positive images it is worth noting that, apart two images, all of them contain a very small yellow colored region.
In the second type of numerical experiment, the MIL classification algorithm proposed in [3] has been applied to classification of some medical dermoscopic images drawn from the database PH_2 [14]. The image are in 8-bit RGB color with a resolution of 768x560 pixels. These images have been classified as common nevus or melanoma by expert dermatologists considering the manual segmentation of the skin lesion, the clinical and histological diagnosis and dermoscopic criteria (asymmetry, colors, pigment network, particular structures).
In our experiments, none of the features resulting from the manual analysis was used in the automatic classification process. We have only considered negative the images of common nevi and positive the images of melanomas. The dataset considered consists of 40 images of melanoma and 40 images of common nevi. To contain the number of instances, we first reduced the image resolution to 128x128 pixels and subsequently we have considered square blobs of 32x32 pixel size. We have adopted a be-level 10-fold cross-validation, listing the corresponding results in Figure 3, where we report, for each fold, the testing correctness, the testing sensitivity, the testing specificity and the CPU time. The results are very promising, even because they have been obtained from photos

with a small resolution (128x128 pixels): we note in fact that in three cases (folds n. 2, 4 e 7) the algorithm has been able to correctly classify all the images in the testing set. We conclude this section by noting that the healthy image with the green box (see Figure 3), characterized by a lot of hair, is the unique misclassified image belonging to the testing set of fold n. 10. This suggests that better results could be obtained by using preprocessing techniques aimed at eliminating the presence of possible noises.



Melanomas **Common Nevi**

Fold number	Testing correctness	Testing sensitivity	Testing specificity	CPU Time
1	87.50	83.33	100.00	0.27
2	100.00	100.00	100.00	0.37
3	75.00	50.00	83.33	0.47
4	100.00	100.00	100.00	0.38
5	87.50	80.00	100.00	0.32
6	62.50	50.00	75.00	0.43
7	100.00	100.00	100.00	0.45
8	75.00	100.00	66.67	0.44
9	62.50	50.00	100.00	0.22
10	87.50	100.00	83.33	0.39
Average	83.75	81.33	90.83	0.37

Fig. 3. MIL-RL Results for Melanoma Image classification

5 Conclusion

In this paper, we have presented an application of a Multiple Instance Learning approach on color and melanoma images, obtaining promising results, both in terms of classification accuracy and CPU times. MIL paradigm fits very well on image classifications because it is able to detect global information (bags) by working at the local level (instance-level). Future research could consist in designing more sophisticated segmentation techniques, aiming to improve the classification performance. Moreover, it would be useful also to test the performance of the proposed algorithm considering preprocessing steps for the optimization of medical images [22]. The inclusion of geometry and textures features may also enrich the classification performance of the proposed model [21]. As for a future extension, we plan to integrate the proposed method into a more general purpose medical information system in order to test and validate its effectiveness on larger datasets also related to different diseases [23, 24].

References

1. Amores, J.: *Multiple instance classification: Review, taxonomy and comparative study*. Artificial intelligence 201, 81-105 (2013)
2. Andrews, S., Tsochantaridis, I., Hofmann, T.: *Support vector machines for multiple-instance learning*. In: Advances in neural information processing systems. pp. 577-584 (2003)
3. Astorino, A., Fuduli, A., Gaudioso, M.: *A Lagrangian relaxation approach for binary multiple instance classification*. IEEE Transactions on Neural Networks and Learning Systems, DOI:10.1109/TNNLS.2018.2885852 (2019)

4. Astorino, A., Fuduli, A., Gaudioso, M., Vocaturo, E.: *A multiple instance learning algorithm for color images classification*. In: Proceedings of the 22nd International Database Engineering & Applications Symposium, IDEAS 2018, Villa San Giovanni, Italy, June 18-20, 2018. pp. 262-266 (2018)
5. Bergeron, C., Moore, G., Zaretzki, J., Breneman, C., Bennett, K.: *Fast bundle algorithm for multiple instance learning*. IEEE Transactions on Pattern Analysis and Machine Intelligence 34(6), 1068-1079 (2012)
6. Carbonneau, M.A., Cheplygina, V., Granger, E., Gagnon, G.: *Multiple instance learning: A survey of problem characteristics and applications*. Pattern Recognition 77, 329-353 (2018)
7. Cristianini, N., Shawe-Taylor, J., et al.: *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press (2000)
8. De Bonet, J.S., Viola, P.A.: *Structure driven image database retrieval*. In: Advances in Neural Information Processing Systems. pp. 866-872 (1998)
9. Grosky, W.I., Mehrotra, R.: *Index-based object recognition in pictorial data management.* Computer vision, Graphics, and Image processing 52(3), 416-436 (1990)
10. Guignard, M.: Lagrangean relaxation. Top 11(2), 151-200 (2003)
11. Li, X., Shou, L., Chen, G., Hu, T., Dong, J.: *Modeling image data for effective indexing and retrieval in large general image databases*. IEEE Transactions on Knowledge and Data Engineering 20(11), 1566-1580 (2008)
12. Mangasarian, O.L., Wild, E.W.: *Multiple instance classification via successive linear programming*. Journal of Optimization Theory and Applications 137(3), 555-568 (2008)
13. Mehrotra, R., Gary, J.E.: *Similar-shape retrieval in shape data management*. Computer (9), 57-62 (1995)
14. Mendonca, T., Celebi, M., Mendonca, T., Marques, J.: *Ph_2: A public database for the analysis of dermoscopic images*. In: Dermoscopy image analysis. CRC Press (2015)
15. Quellec, G., Cazuguel, G., Cochener, B., Lamard, M.: *Multiple-instance learning for medical image and video analysis*. IEEE reviews in biomedical engineering 10, 213-234 (2017)
16. Ravela, S., Manmatha, R., Riseman, E.M.: *Image retrieval using scale-space matching*. In: European Conference on Computer Vision. pp. 273-282. Springer (1996)
17. Shinde, P.N., Manjrekar, A.: *Retrieval of efficiently classified, re-ranked images using histogram based score computation algorithm extended with the elimination of duplicate images*. International Conference for Convergence for Technology 2014. pp. 1-7. IEEE (2014)
18. Shor, N.Z.: *Minimization methods for non-differentiable functions*, vol. 3. Springer Science & Business Media (2012)
19. da Silva Torres, R., Falcao, A.X.: *Content-based image retrieval: theory and applications*. RITA 13(2), 161-185 (2006)
20. Tseng, P.: *Convergence of a block coordinate descent method for non-differentiable minimization*. Journal of Optimization Theory and Applications (2001)
21. Vocaturo, E., Zumpano, E., Veltri, P.: *Features for melanoma lesions characterization in computer vision systems*. In: 9th International Conference on Information, Intelligence, Systems and Applications, IISA 2018, Zakynthos, Greece, July 23-25, 2018. pp. 1-8 (2018)
22. Vocaturo, E., Zumpano, E., Veltri, P.: *Image pre-processing in computer vision systems for melanoma detection*. In: IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018, Madrid, Spain, December 3-6, 2018. pp. 2117-2124 (2018)
23. Zumpano E. et al., *SIMPATICO 3D: A Medical Information System for Diagnostic Procedures*, BIBM 2018, pp. 2125-2128.
24. Iaquinta P. et al., *eIMES 3D: An innovative medical images analysis tool to support diagnostic and surgical intervention*. FNC/MobiSPC 2017: 459-464.