IJAMML 3:1 (2015) 41-49 September 2015
ISSN: 2394-2258
Available at http://scientificadvances.co.in
DOI: http://dx.doi.org/10.18642/ijamml_7100121537

MULTI-INSTANCE LEARNING ON INNER STRUCTURE OF BAGS VIA WEIGHTED MATRIX KERNEL

Haitao Xu, Liya Fan* and Hongxia Zheng

School of Mathematics Sciences, Liaocheng University, Liaocheng, 252059, P. R. China

Abstract

Most previous approaches on multiple instance learning (MIL) had focus on the structures between bags, such as positive instance clustering and bag similarity. In this paper, we proposed a novel method which called weighted matrix kernel support vector machine (WMKSVM) to solve the MIL problems. For WMKSVM, we consider the inner bag structure and assign each instance a weight based on a distance metric between each pair of instances in the same bag. Experiments on six data sets have shown that WMKSVM performs better than other key existing MIL algorithms.

Keywords: machine learning, multiple instance learning, support vector machine, instance weighting, weighted matrix kernel.

*Corresponding author.
E-mail address: fanliya63@126.com (Liya Fan).

2010 Mathematics Subject Classification: 68T10.
Submitted by Jianqiang Gao.
Received August 10, 2015

1. Introduction

Multiple instance learning (MIL) is a supervised machine learning problem. In an MIL problem, instances are considered to be contained in bags (a set of instances is termed as a bag). A bag is classified as a positive bag if one or more instances in that bag are positive, otherwise, it is classified as a negative bag. The main difference between multiple instance learning and traditional supervised learning is that in multiple-instance learning, class labels are related to bags and the goal is to predict unseen bags labels, whereas in traditional supervised learning, class labels are related to instances and the goal is to predict the unseen instances labels.

The multiple instance learning is firstly introduced by Dietterich et al. [1] on drug activity prediction, besides which, MIL has been applied in many other areas, such as image processing [2, 3], text classification [4, 5], computer aided diagnosis [6, 7], and it has been viewed as a new learning framework compared to the traditional supervised single instance learning. Since proposed, MIL has been drawn much attention and numerous of MIL algorithms have been proposed, it is difficult to list all existing MIL algorithms and we mainly focus on algorithms based on support vector machine (SVM) in this paper. Andrews et al. [4] reformulates the SVM with label and bag style constraints, and proposed two MIL algorithms, mi-SVM and MI-SVM. Many researchers have proposed plenty of methods applying SVM to MIL problem [8-23]. Zhou et al. [10] considered that the instances in the same bag are rarely independent in real tasks, they regarded each bag as a graph and each instance as a node in the graph. The basic idea in [10] is to regard every bag as an entity to be processed as a whole. With the same idea of this article, we also regard each bag as an entity and consider the inner structure of each bag.

The rest of the paper is organized as follows. In Section 2, some notations and related works are introduced. In Section 3, we propose our WMKSVM. Experiments and results analysis are performed in Section 4. Section 5 gives some concluding remarks.

2. Related Works

Many multi-instance learning methods have been developed during the past years. Andrews et al. [4] proposed mi-SVM and MI-SVM. mi-SVM tries to identify a maximal margin hyperplane for the instances with subject to the constraints that at least one instance of each positive bag locates in the positive half-space while all instances of negative bags locate in the negative half-space; MI-SVM tries to identify a maximal margin hyperplane for the bags by regarding margin of the "most positive instance" in a bag as the margin of that bag. Kwok and Cheung [18] designed marginalized multi-instance kernels by incorporating generative model into the kernel design. Chen and Wang [19] proposed the DDSVM method which employed diverse density [20] to learn a set of instances and then maps the bags to a feature space. Zhou and Xu [9] proposed the MissSVM method by regarding instances of negative bags as labelled examples while those of positive bags as unlabelled examples with positive constraints. In this paper, we proposed a novel method which called weighted matrix kernel support vector machine (WMKSVM) to solve the MIL problems. For WMKSVM, we consider the inner bag structure and assign each instance a weight based on a distance metric between each pair of instances in the same bag.

3. Weighted Matrix Kernel Support Vector Machine

In this paper, we propose a new multi-instance learning method based on inner structure of bags via weighted matrix kernel using in support vector machine (WMKSVM). We consider that each instance can affect the bag label by its "position" in the bag, which can be presented as the distance between it and the other instances in the same bag. With the thought of this, we assign each instance a weight, which can be used to show the contribution that it makes to the labels of the bag.

Before presenting the details, we give the formal definition of multi-instance learning as following. Let X denote the instance space. Given a data set $\{(X_1, y_1), \cdots, (X_i, y_i), \cdots, (X_m, y_m)\}$, where $X_i = \{x_{i1}, \cdots, x_{ij}, \cdots, x_{i,ni}\} \subseteq \mathbf{X}$ is called a *bag* and $y_i \in \mathbf{Y} = \{-1, +1\}$ is the label of X_i, the goal is to generate a learner to classify unseen bags. Here $x_{ij} \in \mathbf{X}$ is an instance $[x_{ij_1}, \cdots, x_{ij_l}, \cdots, x_{ij_d}]'$, x_{ij_l} is the value of x_{ij} at the l-th attribute, m is the number of training bags, n_i is the number of instances in X_i, and d is the number of attributes. If there exists $x_{ig} \in X_i$ is a positive instance X_i is a positive bag and thus $y_i = +1$; otherwise $y_i = -1$.

Proposition 1. *Given two multi-instance bags X_i and X_j, where n_i and n_j are the number of instances in X_i, X_j. The function k_w defined as*

$$k_w(X_i, X_j) = \sum_{k=1}^{n_i}\sum_{l=1}^{n_j} w_k w_l k(x_{ik}, x_{jl}), \tag{1}$$

is a kernel if and only if $k(x_{ik}, x_{jl})$ is a kernel. Here w_k, w_l denote the weight of x_{ik} and x_{jl}, which are instances in bag X_i and X_j, respectively.

There are many ways to define the weight of each instance, in this paper, we consider the distance metric. Given an arbitrary bag $X = \{x_1, x_2, \cdots, x_n\}$, let

$$d_{x_i} = \frac{1}{\sum_{k=1}^{n} \|x_i - x_k\|^2}, \quad i = 1, 2, \cdots, n, \tag{2}$$

we define the weight of x_i, $w_i = \dfrac{d_{x_i}}{\sum_{i=1}^{n} d_{x_i}}$.

Similar to SVM, the basic idea of WMKSVM is to construct a decision function $f(X) = \operatorname{sgn}(< W, \phi(X) > +b)$ by solving an optimization problem:

$$\min_{W, b, \xi} \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{m}\xi_i$$

$$\text{s.t. } y_i(< W, \phi(X_i) > +b) \geq 1 - \xi_i,$$

$$\xi_i \geq 0, \quad i = 1, 2, \cdots, m, \tag{3}$$

where $W \in \mathbb{R}^{\infty \times n_i}$, $b \in \mathbb{R}$ and $C > 0$. Considering the Lagrangian function of the problem (3)

$$L(W, b, \xi, \alpha, \beta) = \frac{1}{2}\|W\|^2 + C\sum_{i=1}^{m}\xi_i - \sum_{i=1}^{m}\alpha_i((y_i < W, \phi(X_i) > +b) + \xi_i - 1)$$

$$- \sum_{i=1}^{m}\beta_i\xi_i, \tag{4}$$

and letting $\frac{\partial L}{\partial W} = \frac{\partial L}{\partial b} = \frac{\partial L}{\partial \xi_i} = 0$, we can deduce that

$$\frac{\partial L}{\partial W} = W - \sum_{i=1}^{m}\alpha_i y_i \phi(X_i) = 0 \Rightarrow W = \sum_{i=1}^{m}\alpha_i y_i \phi(X_i), \tag{5}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{m}\alpha_i y_i = 0, \tag{6}$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \beta_i = 0, \quad i = 1, \cdots, m. \tag{7}$$

Substituting (5)-(7) into (4), we can get the Wolfe dual form of the problem (3):

$$\min_{\alpha} \frac{1}{2}\alpha^T H\alpha - e^T\alpha$$

$$\text{s.t.} \quad y^T\alpha = 0,$$

$$0 \leq \alpha \leq Ce, \tag{8}$$

which has the same form as the problem of SVM, but by Proposition 1, here $H = [H_{ij}] \in \mathbb{R}^{m\times m}$ and

$$H_{ij} = y_i y_j K(X_i, X_j) = y_i y_j \sum_{k=1}^{n_i}\sum_{l=1}^{n_j} w_k w_l k(x_{ik}, x_{jl}),$$

where $X_i = [x_{i_1}, \cdots, x_{i_{n_2}}]$ and $X_j = [x_{j_1}, \cdots, x_{j_{n_2}}]$. By solving the problem (8), we can obtain the decision function of MWKSVM

$$f(X) = \mathrm{sgn}(\sum_{i=1}^{m} \tilde{\alpha}_i y_i K(X_i, X) + b), \quad \forall X \in \mathbb{R}^{n_1\times n_2}, \tag{9}$$

where $\alpha \in \mathbb{R}^m$ is the optimal solution of the problem (8) and $b = y_j - \sum_{i=1}^{m} \alpha_i y_i K(X_i, X_j)$ for some $0 < \alpha_j < C$.

4. Experiments

In this section, in order to demonstrate the effectiveness of WMKSVM, we perform a series of comparative experiments with some other classifiers on 5 data sets of Musk1, Musk2, Elephant, Fox, Tiger, the data sets details can be found at [21, 4]. Classification accuracy of each classifier is estimated by the standard tenfold cross-validation (CV) methodology. All the classifiers are tested in MATLAB (2014b) [18] running on a PC with system configuration Intel Core2 Celeron (2.6GHz) with 2GB of RAM.

The comparison results on 5 data sets are shown in Table 1.

Table 1. Accuracy (%) on benchmark tasks

Data Sets	MI-SVM[2]	mi-SVM[2]	EM-DD[23]	MI-graph[10]	MI-kernel[22]	WMKSVM
Musk1	77.90	87.40	84.80	**90.00**	88.00	88.80
Musk2	84.30	83.60	84.90	90.00	89.30	**93.00**
Elept	81.40	82.00	78.30	85.10	84.30	**85.50**
Fox	59.40	58.20	56.10	61.20	60.30	**65.50**
Tiger	84.00	78.90	72.10	81.90	84.20	**86.50**

From Table 1, we can see that the classification accuracies of WMKSVM is higher than that of the others except on Musk1 data set, MI-graph get the highest classification accuracies on Musk1 data set.

In summary, we can conclude that the presented WMKSVM is an effective and competitive classifier on multi-instance learning.

5. Conclusion

In this paper, we proposed a novel method which called weighted matrix kernel support vector machine (WMKSVM) to solve the MIL problems. For WMKSVM, we consider the inner bag structure and assign each instance a weight based on a distance metric between each pair of instances in the same bag. Experiments on 5 data sets have shown that WMKSVM performs better than other key existing MIL algorithms.

There is also a lot of work to be done, such as generalization of modelling, improvement of algorithms. These are the next step of our work.

References

[1] T. Dietterich, R. Lathrop and T. Lozano-Pérez, Solving the multiple instance problem with axis-parallel rectangles, Artificial Intelligence 89 (1-2) (1997), 31-71.

[2] O. Maron and A. L. Ratan, Multiple-instance learning for natural scene classification, in: Proceedings of the 15th International Conference on Machine Learning (1998), 341-349.

[3] Q. Zhang, S. A. Goldman, W. Yu and J. Fritts, Content-based image retrieval using multiple-instance learning, in: Proceedings of the 19th International Conference on Machine Learning (2002), 682-689.

[4] S. Andrews, I. Tsochantaridis and T. Hofmann, Support vector machines for multiple-instance learning, Adv. Neural Inf. Process. Syst. (2003), 561-568.

[5] S. Ray and M. Craven, Supervised versus multiple instance learning: An empirical comparison, in: Proceedings of the 22nd International Conference on Machine Learning (2005), 697-704.

[6] J. Bi and J. Liang, Multiple instance learning of pulmonary embolism detection with geodesic distance along vascular structure, in: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2007), 1-8.

[7] G. Fung, M. Dundar, B. Krishnapuram and R. B. Rao, Multiple instance learning for computer aided diagnosis, Adv. Neural Inf. Process. Syst. (2007), 425-432

[8] F. Li and C. Sminchisescu, Convex multiple-instance learning by estimating likelihood ratio, in: J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel and A. Culotta (Eds.), NIPS, Curran Associates Inc. (2010), 1360-1368.

[9] Z. Zhou and J. Xu, On the relation between multi-instance learning and semisupervised learning, In: International Conference on Machine Learning (2007), 1167-1174.

[10] Z. Zhou, Y. Sun and F. Li, Multi-instance learning by treating instances as non I.I.D, In: International Conference on Machine Learning (2009), 1249-1256.

[11] Yanshan Xiao, Bo Liu and Zhifeng Hao, A similarity-based classification framework for multiple-instance learning, IEEE Transaction on Cybernetics 44(4) (2014), 500-515.

[12] Y. Shao, Z. Yang, X. Wang and N. Deng, Multiple instance twin support vector machines, ISORA (2010), 433-442.

[13] Zhiquan Qi, Yingjie Tian, Xiaodan Yu and Yong Shi, A multi-instance learning algorithm based on nonparallel classifier, Applied Mathematics and Computation 241 (2014), 233-241.

[14] Zhi-Hua Zhou, Min-Ling Zhang, Sheng-Jun Huang and Yu-Feng Li, Multi-instance multi-label learning, Artificial Intelligence 176 (2012), 2291-2320.

[15] Zhan Li, Guo-Hua Geng, Jun Feng, Jin-ye Peng, Chao Wen and Jun-li Liang, Multiple instance learning based on positive instance selection and bag structure construction, Pattern Recognition Letters 40 (2014), 19-26.

[16] Veronika Cheplygina, David M. J. Tax and Marco Loog, Multiple instance learning with bag dissimilarities, 48(1) (2015), 264-275.

[17] J. Chai, X. Ding, H. Chen and T. Li, Multiple-instance discriminant analysis, Pattern Recognition 47 (2014), 2517-2531.

[18] J. T. Kwok and P.-M. Cheung, Marginalized multi-instance kernels, In Proceedings of the 20th International Joint Conference on Artificial Intelligence, pages 901-906, Hydrabad, India, 2007.

[19] Y. Chen and J. Z. Wang, Image categorization by learning and reasoning with regions, Journal of Machine Learning Research 5 (2004), 913-939.

[20] O. Maron and T. Lozano-Pérez, A framework for multiple-instance learning, In M. I. Jordan, M. J. Kearns and S. A. Solla, editors, Advances in Neural Information Processing Systems 10, pages 570-576, MIT Press, Cambridge, MA, 1998.

[21] T. G. Dietterich, R. H. Lathrop and T. Lozano-Pérez, Solving the multiple instance problem with axis-parallel rectangles, Artificial Intelligence 89(1-2) (1997), 31-71.

[22] T. Gartner, P. A. Flach, A. Kowalczyk and A. J. Smola, Multi-instance kernels, In Proceedings of the 19th International Conference on Machine Learning, pages 179-186, Sydney, Australia, 2002.

[23] Q. Zhang and S. A. Goldman, EM-DD: An improved multi-instance learning technique, In T. G. Dietterich, S. Becker and Z. Ghahramani, editors, Advances in Neural Information Processing Systems 14, pages 1073-1080, MIT Press, Cambridge, MA, 2002.

■